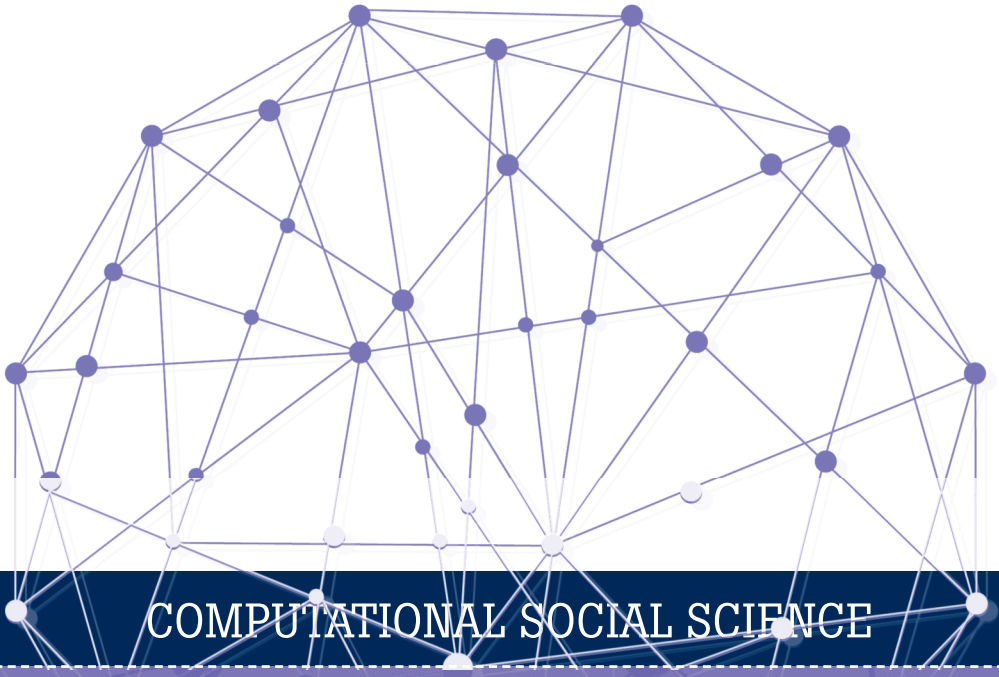
A cura di Giuseppe Giordano,
Mara Maretti, Michelangelo Misuraca,
Giuseppina Damiana Costanzo

ECOLOGIE DELL'ODIO

Razzismi e populismi nell'era digitale



COMPUTATIONAL SOCIAL SCIENCE

FrancoAngeli

COMPUTATIONAL SOCIAL SCIENCE

La collana accoglie contributi di carattere interdisciplinare relativi al dibattito sul campo derivate dall'applicazione di metodi innovativi di ricerca e pratiche di uso dei Big Data, con un'attenzione particolare alle tematiche epistemologiche, metodologiche e politiche di gestione dei contenuti digitali.

Secondo la letteratura internazionale è possibile definire la scienza sociale computazionale come una disciplina che sfrutta la capacità di vasti set di Big Data per analizzare le interazioni umane al fine di definire prospettive qualitativamente nuove sul comportamento collettivo, in un approccio interdisciplinare che comprende sociologia, statistica, informatica, psicologia, diritto, matematica e fisica teorica.

La ricerca sociale computazionale, basandosi sull'analisi delle tracce digitali delle attività online, l'analisi dei network sociali, le fonti aperte digitali, la simulazione sociale attraverso modelli computazionali, rappresenta uno strumento proficuo per l'analisi del mutamento sociale. In tale direzione essa ha già prodotto, negli ultimi dieci anni, moltissimi contributi che confermano la rivoluzione metodologica in atto.

All'interno di questa cornice e in considerazione della crescente consapevolezza della comunità scientifica internazionale di quanto la ricerca sociale debba passare necessariamente per un utilizzo attivo delle tecnologie dell'informazione, la collana ha quindi come obiettivo principale la costituzione di uno spazio di discussione epistemologica, ontologica e metodologica interdisciplinare nel quale poter raccogliere, valutare e catalogare i contributi specifici dell'analisi computazionale.

I volumi pubblicati, in lingua italiana o inglese, sono sottoposti alla valutazione anonima di almeno due referees esperti nei settori scientifico-disciplinari della matematica, della sociologia, della statistica, della fisica teorica, del diritto, dell'informatica e della psicologia.

ISBN Open Access: 9788835191971

ISBN Open Access: 9788835191971

A cura di Giuseppe Giordano,
Mara Maretti, Michelangelo Misuraca,
Giuseppina Damiana Costanzo

ECOLOGIE DELL'ODIO

Razzismi e populismi nell'era digitale

COMPUTATIONAL SOCIAL SCIENCE

FrancoAngeli

ISBN Open Access: 9788835191971

Questa pubblicazione è stata realizzata con il supporto finanziario del Piano Nazionale di Ripresa e Resilienza (PNRR), Missione 4 “Istruzione e Ricerca” – Componente 2 Investimento 1.1, “Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)”, nell’ambito del Progetto PRIN 2022 PNRR “*TOLERANT: Identification and Critical Analysis of Online Racism and Xenophobia against (Im)migrants and Roma people*” (Cod. progetto: P2022APKJL, CUP: D53D23020400001).

Isbn: 9788835185642

Isbn e-book Open Access: 9788835191971

ISBN Open Access: 9788835191971

Indice

Introduzione, *di Giuseppe Giordano, Mara Maretti, Michelangelo Misuraca, Giuseppina Damiana Costanzo* pag. 7

Razzismo e xenofobia online: un'analisi bibliometrica e una rassegna qualitativa della letteratura, *di Melissa Vassallo, Jai Jobe, Stefania Fensore, Annalina Sarra* » 17

Razzismo sistemico e disuguaglianze naturali. Le vecchie e nuove forme di razzismo, *di Alfredo Alietti, Dario Padovan* » 35

Jim Crow, il taccuino di McCarthy e la vittoria postuma di Hitler. Blocchi egemonici, sociologia delle emozioni e politiche della paura, *di Alfredo Agustoni* » 52

Xenofobia online: populismi e othering nella sfera pubblica digitale, *di Alessandra De Luca, Mara Maretti* » 80

Il governo dei discorsi: come cambiano i topic sulle migrazioni al variare delle legislazioni, *di Germano Nocera, Valeria Policastro, Giancarlo Ragozini* » 105

La rappresentazione dell'immigrazione nella stampa italiana: un'analisi statistica del linguaggio mediatico, *di Alex Cucco, Emiliano del Gobbo, Sara Fontanella, Lara Fontanella* » 124

Strumenti lessicali per la decostruzione dell'odio. Dizionari tematici e ontologie di dominio nell'analisi del razzismo e della xenofobia online, *di Alex Cucco, Lara Fontanella, Annalina Sarra, Mario Monteleone* » 143

ISBN Open Access: 9788835191971

Oltre il consenso: un corpus annotato in ottica prospettivista per il rilevamento di razzismo e xenofobia nei social media italiani, *di Lara Fontanella, Michelangelo Misuraca, Giuseppe Giordano, Alex Cucco, Emiliano del Gobbo, Elisa Ignazzi* pag. 160

Rappresentazioni causali dell'odio online: reti bayesiane per l'analisi dei discorsi razzisti, *di Franca Garreffa, Anthony Cossari, Paolo Carmelo Cozzucoli, Michelangelo Misuraca* » 175

Narrazioni d'odio e memoria collettiva: una prospettiva metodologica per lo studio dell'antisemitismo sui social media, *di Luca De Benedictis, Giuseppe Giordano, Maria Prosperina Vitale* » 195

L'altro che immaginiamo e la produzione sociale dell'alterità. Il caso delle comunità romanès in Italia, *di Maria Pia Franciosa, Valentina Isabella, Melissa Vassallo* » 216

Le comunità romanès in Italia dalla stigmatizzazione alle contronarrative negli spazi digitali, *di Stefania Bevilacqua, Fiore Manzo, Nadia Bevilacqua, Alex Cucco, Melissa Vassallo* » 232

Automazione e agency: i sistemi di counter-speech tra efficienza algoritmica e resistenza comunitaria, *di Mara Maretti, Clara Salvatori* » 252

ISBN Open Access: 9788835191971

Introduzione

di *Giuseppe Giordano, Mara Maretti, Michelangelo Misuraca, Giuseppina Damiana Costanzo*

1. Definire l'odio online: coordinate concettuali e tensioni normative

Le piattaforme digitali hanno ridisegnato in profondità il paesaggio della comunicazione pubblica. In poco più di due decenni, i social media non si sono limitati ad amplificare dinamiche preesistenti, ma hanno generato grammatiche espressive inedite, nuovi spazi di aggregazione e modalità di diffusione dei contenuti senza precedenti per velocità, portata e capacità di radicalizzazione (Phillips e Milner, 2018; Castells, 2009). Questa trasformazione ha portato con sé opportunità straordinarie per la partecipazione democratica e la costruzione di comunità, ma anche sfide urgenti: tra queste, la proliferazione del razzismo e della xenofobia online rappresenta una delle più complesse e insidiose per le società contemporanee.

Cosa intendiamo esattamente quando parliamo di hate speech? La domanda, apparentemente semplice, nasconde una notevole complessità. Non esiste una definizione universalmente accettata, e questa indeterminatezza non è un limite marginale, ma attraversa l'intero campo di studi (Strossen, 2018). Sellars (2016), mappando le diverse definizioni proposte in ambito accademico, giuridico e sulle stesse piattaforme, individua alcuni tratti ricorrenti: prendere di mira un gruppo o un individuo in quanto membro di un gruppo; esprimere odio, causare danno o incitare ad azioni nocive senza finalità che lo riscattino; manifestare l'intenzione di ferire; assumere carattere pubblico; collocarsi in un contesto che rende plausibile una risposta violenta. Questi elementi, tuttavia, non compongono una definizione unitaria, bensì un insieme di criteri la cui compresenza, in misura maggiore o minore, orienta il riconoscimento di un discorso come discorso d'odio.

Se il razzismo online designa pratiche e linguaggi discriminatori rivolti a gruppi etnici negli ambienti digitali, e la xenofobia digitale esprime atteggiamenti di rifiuto nei confronti di chi è percepito come "altro", entrambi i

ISBN Open Access: 9788835191971

fenomeni assumono nel web forme peculiari e qualitativamente distinte rispetto alle loro manifestazioni offline. Le piattaforme non sono contenitori neutri: come ha mostrato Matamoros-Fernández (2017) con il concetto di platformed racism, esse modellano attivamente i contenuti che ospitano, ne determinano la visibilità e la viralità, contribuiscono alla costruzione stessa dei significati. L'architettura algoritmica, le logiche della raccomandazione, la presunzione di anonimato e la velocità di diffusione creano un ecosistema in cui l'odio può propagarsi, sedimentarsi e radicalizzarsi con un'intensità prima impensabile.

Qui si innesta una tensione fondamentale che attraversa l'intero dibattito: quella tra tutela della dignità e salvaguardia della libertà di espressione (Brown e Sinclair, 2019; Lepoutre, 2017). La tradizione liberale, sintetizzata nella celebre formulazione del giudice Brandeis – "il rimedio al discorso dannoso è più discorso, non il silenzio forzato" (Whitney v. California, 1927) – affida al libero confronto delle idee il compito di far prevalere le posizioni migliori. Ma questa fiducia nel "mercato delle idee" regge davvero alla prova dei fatti? Chi subisce hate speech raramente dispone delle stesse risorse: tempo, visibilità, legittimazione sociale, per rispondere efficacemente. E chi odia difficilmente cambia idea perché qualcuno lo confuta con argomenti razionali.

Inoltre, come hanno mostrato i Critical Race Theorists nel volume *Words That Wound* (Matsuda *et al.*, 1993), le parole che feriscono producono danni reali e documentabili: conseguenze psicologiche, esclusione sociale, perpetuazione di stereotipi che alimentano pratiche discriminatorie concrete. L'hate speech non è dunque la semplice espressione di pregiudizi individuali, ma si inscrive in strutture più ampie di violenza simbolica che pervadono il tessuto sociale. In questa prospettiva, Gelber (2002) propone di ripensare il counter-speech non come mera aggiunta di voci al dibattito, bensì come pratica che richiede supporto istituzionale, materiale ed educativo per permettere ai gruppi marginalizzati di partecipare effettivamente alla sfera pubblica.

I discorsi d'odio che circolano sui social media sono dunque, al tempo stesso, specchio e motore delle trasformazioni in corso nelle società occidentali (Pasta, 2018). Le loro conseguenze si dispiegano su più piani: le vittime subiscono effetti psicologici significativi – ansia, depressione, isolamento –, mentre l'esposizione reiterata a contenuti discriminatori contribuisce a normalizzare stereotipi e pregiudizi, erodendo la coesione sociale e i fondamenti stessi del dialogo democratico. È a partire da questa consapevolezza che il presente volume si propone di offrire strumenti analitici e chiavi interpretative per comprendere un fenomeno che, nella sua complessità, richiede uno sguardo autenticamente interdisciplinare.

ISBN Open Access: 9788835191971

2. Obiettivi e struttura del volume

È in questo scenario – segnato dalla proliferazione del discorso d'odio, dalla sua ibridazione con le logiche delle piattaforme e dalla persistente difficoltà a tracciare confini netti tra espressione legittima e contenuto dannoso – che si colloca il presente volume. L'obiettivo è offrire una riflessione organica sul fenomeno del razzismo e della xenofobia negli ambienti digitali, adottando una prospettiva che intreccia sociologia, statistica, linguistica computazionale e metodologia della ricerca sociale.

Il volume nasce dalla convinzione che comprendere l'odio online richieda uno sguardo prismatico, capace di coniugare profondità teorica e rigore empirico. Il progetto si articola, infatti, in tre direttrici complementari. La prima è di natura teorico-interpretativa: diversi contributi ricostruiscono le grammatiche del razzismo contemporaneo, indagando come stereotipi, paure e costruzioni identitarie si sedimentino nel discorso pubblico e si riattualizzino negli spazi digitali. Vengono esplorate le radici storiche del pregiudizio, dalle politiche segregazioniste ai "demoni popolari" del panico morale, e le loro metamorfosi nell'epoca delle piattaforme algoritmiche, dove nazionalismi, teorie della sostituzione etnica e retoriche securitarie trovano nuove casse di risonanza.

La seconda direttrice è metodologico-strumentale. Il volume dedica ampio spazio allo sviluppo e alla validazione di risorse per l'analisi computazionale del linguaggio d'odio: dizionari tematici come HurtLex nella sua versione revisionata ed espansa, ontologie di dominio che formalizzano le relazioni semantiche sottostanti alle narrazioni ostili, corpora annotati con approcci prospettivisti che valorizzano la pluralità delle interpretazioni anziché appiattirla su etichette univoche. Questi strumenti – costruiti a partire da dati italiani e calibrati sulle specificità linguistiche e culturali del contesto nazionale – rispondono a un'esigenza largamente avvertita: disporre di risorse affidabili per il rilevamento automatico di contenuti tossici che non si limitino alla mera trasposizione di modelli anglofoni.

La terza direttrice è empirico-analitica: il volume presenta una serie di studi di caso che mettono in luce aspetti specifici del fenomeno. L'analisi della rappresentazione dell'immigrazione nella stampa italiana rivela come testate di orientamenti politici diversi costruiscano frame narrativi divergenti, contribuendo alla polarizzazione del dibattito pubblico. L'applicazione di tecniche di topic modeling consente di tracciare l'evoluzione dei temi migratori nel corso delle legislature, mostrando come il discorso mediatico si modelli sulle agende politiche. L'impiego di reti bayesiane consente di ricostruire le dipendenze probabilistiche tra repertori tematici e categorie discorsive dell'odio, identificando i meccanismi generativi del linguaggio osti-

ISBN Open Access: 9788835191971

le. Lo studio dell'antisemitismo online, condotto attraverso Social Network Analysis e Natural Language Processing, mostra come eventi commemorativi e conflitti geopolitici ridefiniscano le strutture relazionali e i domini semantici dell'odio. L'analisi delle comunità romanès – condotta sia attraverso ricerche qualitative basate su interviste, sia attraverso l'esame delle rappresentazioni digitali – documenta come l'antiziganismo si riproduca nella quotidianità e si amplifichi negli spazi virtuali, ma anche come emergano forme di autorappresentazione e contronarrazione.

Un filo rosso attraversa tutti i contributi: l'attenzione alle strategie di contrasto. Se l'odio online non può essere semplicemente silenziato – per ragioni normative, tecniche e di efficacia – diventa cruciale interrogarsi sulle possibilità del *counter-speech* e delle contronarrative. Il volume esplora tanto i sistemi automatizzati di risposta ai contenuti tossici quanto le pratiche comunitarie di resistenza simbolica, mostrando come l'agency dei gruppi marginalizzati possa trovare negli stessi ambienti digitali che amplificano l'odio spazi per la riaffermazione identitaria e la ricostruzione di narrazioni alternative.

Il risultato è un lavoro che ambisce a essere, insieme, una fotografia dello stato dell'arte e una cassetta degli attrezzi. Una fotografia, perché restituisce la complessità di un fenomeno in rapida evoluzione, cogliendone le articolazioni tematiche, i gruppi bersaglio, le dinamiche di produzione e di circolazione. Una cassetta degli attrezzi, perché mette a disposizione della comunità scientifica – ma anche di decisori politici, educatori, professionisti dell'informazione e operatori del terzo settore – risorse teoriche e metodologiche per comprendere, monitorare e contrastare il discorso d'odio. In un'epoca in cui le parole possono ferire con una velocità e una portata senza precedenti, dotarsi di strumenti adeguati non è un lusso accademico, ma una necessità civile.

3. Guida alla lettura

Il volume si articola in tredici contributi che, pur nella loro autonomia, compongono un percorso argomentativo coerente. L'architettura del testo procede in fasi successive: dai quadri teorici e dalle mappature sistematiche della letteratura si passa agli strumenti metodologici, per arrivare, infine, agli studi empirici su casi specifici e alle strategie di contrasto. Il lettore può naturalmente seguire percorsi di lettura selettivi, ma l'ordine proposto riflette una logica interna che procede dal generale al particolare, dalla cornice interpretativa all'applicazione operativa.

Il volume si apre con il contributo di Vassallo, Jobe, Fensore e Sarra (*Razzismo e xenofobia online: un'analisi bibliometrica e una rassegna qualitativa della letteratura*), che offre una mappatura sistematica del campo di

ISBN Open Access: 9788835191971

studi attraverso l'analisi di oltre tremila pubblicazioni indicizzate su Web of Science. L'approccio bibliometrico consente di ricostruire l'evoluzione temporale della ricerca – con una crescita esponenziale a partire dal 2015 – di identificare le aree disciplinari coinvolte e di tracciare le reti concettuali che strutturano il dibattito scientifico. La rassegna qualitativa approfondisce cinque cluster tematici emergenti: gli effetti sulla salute mentale, il ruolo delle piattaforme, il movimento Black Lives Matter, l'impatto della pandemia di COVID-19 e l'intersezione con le discriminazioni contro le persone LGBTQ+. Il capitolo fornisce così le coordinate fondamentali per orientarsi in un territorio di ricerca in rapida espansione.

Il secondo capitolo, firmato da Alietti e Padovan (*Razzismo sistemico e disuguaglianze naturali. Le vecchie e nuove forme di razzismo*), propone un inquadramento teorico di ampio respiro. Gli autori ricostruiscono la progressiva normalizzazione del discorso razzista nelle società contemporanee, analizzando il passaggio dal "razzismo sottile" delle retoriche democratiche all'esplicita rivendicazione identitaria dei movimenti xeno-populisti. Il contributo esplora le logiche di eterorazzizzazione e autorazzizzazione, mostrando come il razzismo operi simultaneamente come ideologia e come legame sociale, nutrendosi del rancore dei ceti vulnerabili e traducendosi in pratiche di esclusione che attraversano le istituzioni, le politiche e il quotidiano.

Con il terzo capitolo, Agustoni (*Jim Crow, il taccuino di McCarthy e la vittoria postuma di Hitler. Blocchi egemonici, sociologia delle emozioni e politiche della paura*) si colloca deliberatamente "a monte" del fenomeno digitale, recuperando l'analisi sociologica delle emozioni come chiave interpretativa del pregiudizio. Attraverso un percorso che intreccia la riflessione di Lippmann sugli stereotipi, la teoria gramsciana dell'egemonia e la sociologia delle emozioni di Illouz, il contributo ricostruisce alcuni casi storici paradigmatici: la costruzione dell'etnicità negli Stati Uniti del XIX secolo, il maccartismo, le politiche della paura urbana a partire dagli anni Sessanta. L'obiettivo è mostrare come le rappresentazioni razzializzanti non siano mai il prodotto neutrale di distorsioni cognitive, bensì il risultato dell'intreccio tra emozioni profonde, rapporti di potere e strategie egemoniche.

Il quarto capitolo, di De Luca e Maretti (*Xenofobia online: populismi e othering nella sfera pubblica digitale*), introduce la dimensione propriamente tecnologica dell'analisi. Attraverso un'analisi bibliometrica mirata, le autrici tracciano l'evoluzione del dibattito scientifico sull'hate speech xenofobo, identificando tre assi interpretativi: la transizione dal paradigma multiculturalista all'islamofobia, la mobilitazione identitaria della destra radicale attorno ai temi della whiteness e del nazionalismo, l'emergere delle piattaforme come ambienti specifici in cui il discorso ostile assume forme pecu-

ISBN Open Access: 9788835191971

liari. Il contributo sostiene che l'hate speech xenofobo online è l'esito della convergenza tra la crisi del multiculturalismo, la riconfigurazione etno-nazionale delle identità politiche e la trasformazione tecnologica della sfera pubblica.

I capitoli quinto e sesto inaugurano la sezione dedicata all'analisi empirica della rappresentazione mediatica dell'immigrazione in Italia. Nocera, Policastro e Ragozini (*Il governo dei discorsi: come cambiano i topic sulle migrazioni al variare delle legislazioni*) approfondiscono l'analisi diacronica, applicando tecniche di topic modeling a un corpus di articoli pubblicati negli ultimi tre anni delle legislature italiane. Il contributo mostra come i temi dominanti nella narrazione migratoria si modifichino in relazione al colore politico dei governi in carica: se le testate di destra tendono a enfatizzare argomenti legati alla sicurezza nazionale, alla critica alle ONG e alla regolamentazione dell'accoglienza, quelle di sinistra privilegiano i diritti dei migranti, le politiche europee e la solidarietà sociale. L'analisi evidenzia come il linguaggio stesso si polarizzi, con l'uso di termini come "clandestino" e "immigrato" nelle testate conservative, "migrante" e "persona" in quelle progressiste. Cucco, del Gobbo, S. Fontanella e L. Fontanella (*La rappresentazione dell'immigrazione nella stampa italiana: un'analisi statistica del linguaggio mediatico*) analizzano un corpus di quasi duemila articoli pubblicati su dodici testate di orientamenti politici diversi. Attraverso tecniche di analisi delle reti semantiche, il contributo mappa le strutture concettuali sottostanti alla narrazione giornalistica, identificando cluster tematici, pattern lessicali e differenze sistematiche tra la stampa progressista e quella conservatrice. I risultati confermano la persistenza di frame securitari e di meccanismi di othering, ma anche l'esistenza di variazioni significative nelle strategie di rappresentazione.

I capitoli settimo, ottavo e nono costituiscono il nucleo metodologico del volume, dedicato allo sviluppo di risorse computazionali per l'analisi del linguaggio d'odio. Cucco, Fontanella, Sarra e Monteleone (*Strumenti lessicali per la decostruzione dell'odio: Dizionari tematici e ontologie di dominio nell'analisi del razzismo e della xenofobia online*) presentano un percorso che muove dalla revisione sistematica del lessico HurtLex – con l'espansione del repertorio terminologico, la riorganizzazione categoriale e la graduazione dell'offensività – alla costruzione di un'ontologia di dominio che formalizza le relazioni semantiche e narrative del discorso ostile. Il contributo illustra inoltre le potenzialità applicative di tali risorse attraverso il software NooJ, mostrando come l'integrazione tra dizionario e ontologia possa supportare strategie di riconoscimento più robuste e interpretabili. Fontanella, Misuraca, Giordano, Cucco, del Gobbo e Ignazzi (*Oltre il consenso: un corpus annotato in ottica prospettivista per il rilevamento di razzismo e xenofobia nei*

ISBN Open Access: 9788835191971

social media italiani) affrontano una questione cruciale per l'addestramento dei sistemi di rilevamento automatico: la soggettività intrinseca della percezione dell'*hate speech*. Adottando l'approccio prospettivista, che riconosce il disaccordo tra gli annotatori non come rumore da eliminare, bensì come segnale informativo, il contributo presenta la costruzione di un nuovo corpus italiano per la rilevazione del razzismo e della xenofobia. Lo schema di annotazione multidimensionale – che include la presenza di contenuti razzisti, il livello di gravità, i gruppi bersaglio e le strategie retoriche – consente di catturare la complessità del fenomeno e valorizzare la pluralità delle prospettive interpretative. Garreffa, Cossari, Cozzucoli e Misuraca (*Rappresentazioni causali dell'odio online: reti bayesiane per l'analisi dei discorsi razzisti*) propongono un framework computazionale integrato che combina l'analisi di rete e la modellizzazione bayesiana. Applicato a un corpus di oltre ottomila tweet annotati nell'ambito dei task HaSpeeDe, l'approccio consente di ricostruire la struttura semantica del discorso ostile, identificando cluster tematici e formalizzando le dipendenze probabilistiche tra temi e categorie discorsive. Le reti bayesiane permettono di esplorare scenari controfattuali e configurazioni probabilistiche non immediatamente osservabili, offrendo una comprensione più profonda dei meccanismi generativi del linguaggio d'odio.

Il decimo capitolo di De Benedictis, Giordano e Vitale (*Narrazioni d'odio e memoria collettiva: una prospettiva metodologica per lo studio dell'antisemitismo sui social media*), sposta l'attenzione su una forma specifica di hate speech che intreccia memoria storica e attualità geopolitica. Attraverso l'integrazione di Social Network Analysis e di tecniche di Natural Language Processing, gli autori analizzano il dibattito sviluppatosi su *X* (ex Twitter) in occasione del Giorno della Memoria tra il 2022 e il 2025, nonché l'evoluzione delle narrazioni successive al 7 ottobre 2023. Lo studio mostra come eventi commemorativi e conflitti contemporanei contribuiscano a ridefinire le strutture relazionali e i domini semantici dell'antisemitismo online, con l'emergere di forme nuove che mescolano revisionismo storico, distorsione delle responsabilità della Shoah e ostilità verso gli ebrei mediata dalla critica allo Stato di Israele.

Gli ultimi tre capitoli sono dedicati alle comunità romanès, caso emblematico di minoranza storicamente marginalizzata e oggetto di forme specifiche di razzismo strutturale. Franciosa, Isabella e Vassallo (*L'altro che immaginiamo e la produzione sociale dell'alterità. Il caso delle comunità romanès in Italia*) adottano un approccio qualitativo, presentando i risultati di venti interviste condotte in cinque territori italiani, tra cui comunità rom e operatori del terzo settore. Il contributo esplora i meccanismi di costruzione dell'alterità – dalle categorizzazioni imposte alle barriere strutturali, dalla

ISBN Open Access: 9788835191971

segregazione spaziale dei "campi nomadi" alle discriminazioni quotidiane nel mercato del lavoro e dell'abitazione –, facendo emergere al contempo le forme di agency individuale e le strategie di resistenza. Bevilacqua, Manzo, N. Bevilacqua, Cucco e Vassallo (*Le comunità romanès in Italia dalla stigmatizzazione alle contronarrative negli spazi digitali*) completano il quadro analizzando le manifestazioni dell'antiziganismo negli ambienti digitali e le strategie di contrasto. Il contributo ricostruisce la lunga storia degli stereotipi sui gruppi rom, documentando come si riproducano e si amplifichino sui social media attraverso commenti denigratori, fake news e rappresentazioni caricaturali. Ma il capitolo esplora anche l'altra faccia della medaglia: le forme di autorappresentazione che stanno emergendo, la produzione letteraria e culturale degli autori rom italiani, le contronarrative e le pratiche di "detossificazione" dei contenuti ostili che trasformano gli stessi spazi digitali in luoghi di riaffermazione identitaria. Il volume si chiude con il contributo di Maretti e Salvatori (*Automazione e agency: i sistemi di counter-speech tra efficienza algoritmica e resistenza comunitaria*), che affronta frontalmente la questione delle strategie di risposta all'hate speech. Le autrici propongono un'analisi comparativa di quattro sistemi automatizzati di counter-speech – Perspective API, CONAN, ChatGPT e SELMA –, esaminandoli attraverso tre dimensioni: epistemologica (come "conoscono" l'hate speech), politica (chi definisce le regole del gioco) ed etica (quali valori incorporano). Il contributo sviluppa il concetto di "ortoprassi algoritmica" per descrivere i limiti di approcci che privilegiano l'efficienza scalabile rispetto alla comprensione contestuale, e propone una prospettiva "community-in-the-loop" che riconosca la centralità delle comunità colpite nella definizione delle strategie di contrasto. La riflessione teorica si nutre della tradizione performativa degli atti linguistici (Austin), della teoria butleriana della performatività e dell'analisi bourdieusiana del potere simbolico, mostrando come il counter-speech non sia semplicemente "più parole" da aggiungere al dibattito, bensì una pratica che richiede supporto istituzionale e partecipa alla ridefinizione dei confini del dicibile nella sfera pubblica.

Nel loro insieme, i tredici contributi compongono un mosaico che ambisce a restituire la complessità del fenomeno senza rinunciare alla profondità analitica. Il percorso dal primo all'ultimo capitolo disegna una traiettoria che parte dalla mappatura del campo, attraversa le cornici teoriche e gli strumenti metodologici, si immerge negli studi empirici e approda alle strategie di intervento. È un percorso che riflette la convinzione, condivisa da tutti gli autori, che comprendere l'odio sia il primo passo – necessario ma non sufficiente – per contrastarlo.

ISBN Open Access: 9788835191971

Riferimenti bibliografici

Brown, A. and Sinclair, A. (2019), *The Politics of Hate Speech Laws*, Routledge, New York.

Castells, M. (2009), *Communication Power*, Oxford University Press, Oxford.

Gagliardone, I., Gal, D., Alves, T. and Martinez, G. (2015), *Countering online hate speech*, UNESCO Publishing, Paris.

Gelber, K. (2002), *Speaking Back: The Free Speech Versus Hate Speech Debate*, John Benjamins Publishing, Amsterdam.

Lepoutre, M. (2017), Hate speech in public discourse: A pessimistic defense of counterspeech, *Social Theory and Practice*, 43, 4: 851-883.

Matamoros-Fernández, A. (2017), Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube, *Information, Communication & Society*, 20, 6: 930-946.

Matsuda, M., Lawrence, C.R., Delgado, R. and Crenshaw, K.W. (1993), *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*, Westview Press, Boulder.

Pasta, S. (2018), *Razzismi 2.0. Analisi socio-educativa dell'odio online*, Scholé-Morcelliana, Milano.

Phillips, W. and Milner, R.M. (2018), *The Ambivalent Internet: Mischief, Oddity, and Antagonism Online*, John Wiley & Sons, Hoboken.

Sellars, A. (2016), *Defining Hate Speech*, Berkman Klein Center Research Publication No. 2016-20, Harvard University, Cambridge MA, testo disponibile al sito: https://cyber.harvard.edu/publications/2016/DefiningHateSpeech.

Strossen, N. (2018), *HATE: Why We Should Resist It with Free Speech, Not Censorship*, Oxford University Press, New York.

ISBN Open Access: 9788835191971

ISBN Open Access: 9788835191971

Razzismo e xenofobia online: un'analisi bibliometrica e una rassegna qualitativa della letteratura

di *Melissa Vassallo*[*], *Jai Jobe*[**], *Stefania Fensore*[**], *Annalina Sarra*[**]

1. Introduzione

L'avvento e la diffusione capillare delle piattaforme digitali hanno profondamente trasformato le modalità di interazione sociale, generando nuovi spazi per la costruzione di comunità e la circolazione di idee e opinioni. Tuttavia, questo ecosistema comunicativo, caratterizzato da velocità, interconnessione e anonimato, ha favorito la proliferazione di contenuti tossici e fenomeni di intolleranza su scala globale (Phillips e Milner, 2018). Tra questi, il razzismo e la xenofobia online rappresentano alcune delle sfide più complesse e urgenti per le società contemporanee, fenomeni multidimensionali che intrecciano dimensioni tecnologiche, sociali, psicologiche e giuridiche. Il termine *razzismo online* designa l'insieme di pratiche e linguaggi discriminatori rivolti a gruppi etnici attraverso gli ambienti digitali, amplificati dall'anonimato e dalla viralità. Analogamente, la *xenofobia digitale* esprime atteggiamenti di rifiuto e ostilità nei confronti di individui percepiti come "altri" o "diversi", trovando nelle reti sociali un potente amplificatore. Questi fenomeni non rappresentano una semplice trasposizione digitale di dinamiche preesistenti, ma assumono forme peculiari per rapidità di diffusione, capacità di raggiungere pubblici vasti e diversificati e potenziale di radicalizzazione.

L'impatto sociale di tali manifestazioni è profondo. Da un lato, le vittime di *hate speech* online subiscono conseguenze psicologiche significative, tra cui ansia, depressione e isolamento sociale; dall'altro, l'esposizione reiterata a contenuti discriminatori può contribuire a normalizzare stereotipi e pregiudizi, erodendo la coesione sociale e minando i fondamenti del dialogo demo-

[*] Dipartimento di Scienze Giuridiche e Sociali, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, melissa.vassallo@unich.it

[**] Dipartimento di Studi Socio-Economici, Gestionali e Statistici, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, jai.jobe@phd.unich.it; stefania.fensore@unich.it; annalina.sarra@unich.it

ISBN Open Access: 9788835191971

cratico. Eventi recenti hanno evidenziato come i discorsi d'odio diffusi online possano tradursi in violenza fisica e discriminazioni nel mondo reale. Negli ultimi anni la comunità scientifica ha risposto con crescente attenzione a queste criticità, sviluppando approcci che spaziano dall'analisi qualitativa dei discorsi alle metodologie computazionali per il rilevamento automatico dell'hate speech, dagli studi sugli effetti psicosociali alle riflessioni di carattere giuridico e normativo. Tuttavia, la letteratura sul tema risulta ancora frammentata e dispersa tra differenti ambiti disciplinari. In questo contesto si colloca il presente contributo, che mira a fornire una mappatura sistematica e aggiornata del panorama scientifico sul razzismo e sulla xenofobia online attraverso l'applicazione di metodologie bibliometriche, con lo scopo di identificare tendenze, temi emergenti, reti di collaborazione e lacune conoscitive in ambiti di studio complessi e in rapida trasformazione. Attraverso l'analisi di un corpus di pubblicazioni indicizzate nella banca dati *Web of Science*, questo studio intende delineare le coordinate fondamentali di un campo di ricerca in crescita, offrendo una base solida per l'individuazione di priorità di ricerca future e per il rafforzamento della collaborazione interdisciplinare. I risultati potranno orientare decisori politici, educatori e professionisti nella creazione di strategie contro l'hate speech online e nella promozione di culture digitali inclusive.

Il resto del contributo è strutturato come segue. Nella sezione 2 vengono presentati gli strumenti e la metodologia per l'analisi bibliometrica, la sezione 3 approfondisce le aree tematiche della produzione scientifica e l'evoluzione temporale, mentre nella sezione 4 si riporta una rassegna qualitativa della letteratura. Infine, la sezione 5 contiene delle considerazioni finali.

2. Metodo di ricerca

2.1. Approccio bibliometrico

Il processo di raccolta e selezione della letteratura è avvenuto attraverso uno *scanning* del database Web of Science (WoS) Core Collection, scelto per la sua ampia copertura e l'affidabilità nelle analisi bibliometriche. La ricerca è stata condotta sui campi titolo, abstract e parole chiave, impiegando una query volta a identificare gli studi pertinenti. Il periodo di riferimento comprende gli anni dal 2000 al 2025, includendo esclusivamente articoli di ricerca e review in lingua inglese. Oltre a termini specifici ("online racism", "digital xenophobia", etc.), è stata utilizzata la seguente query:

(("racis*" OR "xenophob*") AND ("online" OR "social media" OR "Twitter" OR "Telegram" OR "Facebook" OR "Reddit" OR "TikTok")).

ISBN Open Access: 9788835191971

2.2. Strumenti e metodologie per l'analisi bibliometrica

L'analisi bibliometrica è stata sviluppata attraverso un approccio multi-strumento. Il workflow si è articolato sull'impiego combinato di Bibliometrix (Aria e Cuccurullo, 2017), CiteSpace (Chen, 2006) e VOSviewer (Van Eck e Waltman, 2010). La metodologia di ricerca ha integrato tre tecniche bibliometriche fondamentali. L'analisi delle citazioni è stata utilizzata per ricostruire l'evoluzione temporale del campo e misurare la produttività scientifica, identificando i contributi più influenti e i cluster di ricerca consolidati. L'analisi delle co-occorrenze, basata sulle parole-chiave, ha permesso di esplorare la dimensione concettuale del dominio, individuando temi emergenti.

3. Risultati

3.1. Andamento temporale delle pubblicazioni

L'analisi della produzione scientifica sulla ricerca del razzismo nei social media copre un arco temporale dal 2001 al 2025, con un corpus complessivo di 3094 documenti provenienti da 1545 fonti. L'evoluzione può essere distinta in tre fasi principali (Fig. 1). La prima fase (2001-2009) rappresenta il periodo in cui la ricerca è ancora agli albori, caratterizzato da una produzione discontinua e modesta. La seconda rappresenta una fase di consolidamento e crescita moderata (2010-2014). La terza fase (2015-2025) coincide con un'esplosione dell'interesse accademico, contraddistinta da una crescita esponenziale delle pubblicazioni, con picchi significativi nel 2023 e 2025. Questa accelerazione rispecchia la maturazione del campo di studi e la crescente rilevanza sociale del razzismo digitale, amplificata da fenomeni globali quali l'impatto della pandemia COVID-19 sulle dinamiche digitali, la polarizzazione politica nei social media, la diffusione di campagne d'odio online e il crescente impegno istituzionale nel contrasto ai discorsi d'odio.

La Figura 1 evidenzia che il numero di citazioni ha avuto una dinamica simile. L'impatto scientifico, misurato attraverso una media di 15,71 citazioni per articolo, attesta la rilevanza e visibilità internazionale dei contributi. Il settore appare, dunque, in una fase di piena maturità, caratterizzata da elevati livelli di collaborazione e consolidamento delle aree tematiche emergenti.

ISBN Open Access: 9788835191971

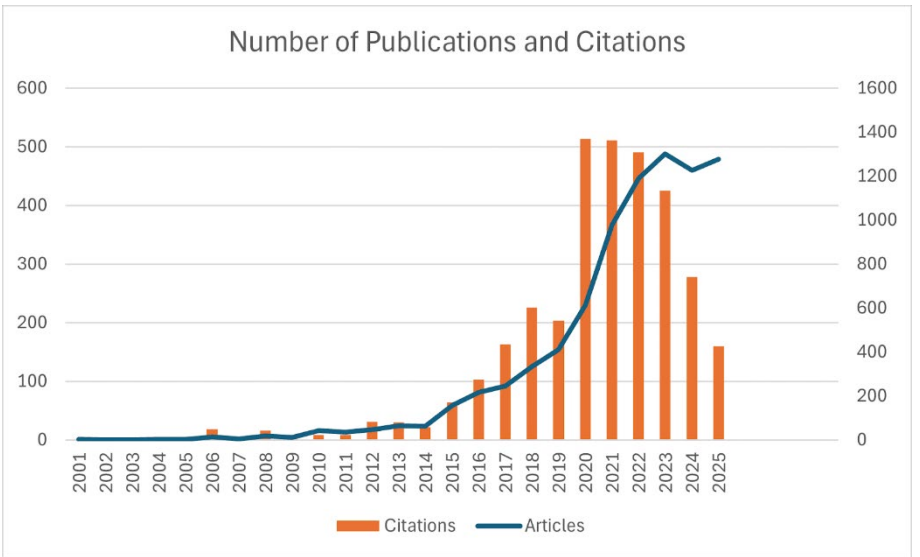


Fig. 1 – Produzione scientifica annuale: numero di pubblicazioni e citazioni

È stata, inoltre, confrontata la produttività scientifica dei diversi Paesi, considerando sia il numero di articoli pubblicati sia il volume complessivo di citazioni ricevute (tab. 1).

Tab. 1 – Produttività scientifica (Top 10); PSP: pubblicazioni di un singolo paese; PMP: pubblicazioni di più paesi; #art.: numero di articoli; #cit.: numero di citazioni; CMA: Citazioni medie per articolo

Rank	Paese	#art.	%	PSP	PMP	Rank	Paese	#cit.	CMA
1	USA	1547	51.4	679	868	1	USA	28283	18.28
2	UK	363	12.1	144	219	2	UK	5946	16.38
3	Canada	210	6.9	56	154	3	Australia	3774	18.32
4	Australia	206	6.9	83	123	4	Canada	2122	10.1
5	Germania	108	3.6	26	82	5	Svezia	1309	26.71
6	Cina	103	3.4	27	76	6	Germania	1260	11.67
7	Spagna	86	2.9	26	60	7	Cina	1257	12.2
8	Sud Africa	67	2.2	30	37	8	Spagna	1151	13.38
9	Paesi Bassi	52	1.7	16	36	9	Paesi Bassi	954	18.35
10	Svezia	49	1.6	22	27	10	Sud Africa	605	9.03

ISBN Open Access: 9788835191971

Per quanto riguarda la produzione scientifica, gli Stati Uniti si collocano al primo posto con 1547 articoli (51.43%), seguiti dal Regno Unito (363 articoli, 12.07%) e dal Canada (210 articoli, 6.98%). La frequenza di pubblicazione mostra una forte eterogeneità tra i Paesi analizzati, con valori che oscillano da poco più dell'1% fino a oltre il 50%.

Anche in termini di impatto misurato attraverso il numero totale di citazioni, gli Stati Uniti mantengono una posizione di leadership con 28283 citazioni complessive. Il Regno Unito si conferma al secondo posto con 5946 citazioni, mentre Australia e Canada seguono con 3774 e 2122 citazioni, rispettivamente. Circa le citazioni medie per articolo, spiccano invece Svezia e Paesi Bassi, che mostrano valori medi molto elevati (26.71 e 18.35), indicando una notevole influenza rispetto alla loro produzione complessiva.

3.2. Aree tematiche e multidisciplinarietà

L'analisi delle categorie tematiche conferma la forte vocazione interdisciplinare della ricerca sul razzismo e sulla xenofobia online, distribuita complessivamente su 45 diverse aree di classificazione Web of Science. La Tabella 2 presenta le *prime dieci* categorie per numerosità.

Tab. 2 – Distribuzione delle pubblicazioni per categoria Web of Science (Top 10)

Categoria WoS	# pubblicazioni	%
Communication	530	17.11%
Sociology	324	10.46%
Public Environmental Occupational Health	306	9.88%
Education & Educational Research	187	6.04%
Ethnic Studies	185	5.97%
Interdisciplinary Social Sciences	179	5.78%
Multidisciplinary Psychology	178	5.75%
Social Psychology	121	3.91%
Computer Science Information Systems	110	3.55%
Political Science	110	3.55%

Il settore prevalente è quello della Communication (17.11%), a testimonianza del fatto che il fenomeno viene studiato principalmente attraverso le sue dinamiche di circolazione e ricezione mediale. Seguono la Sociology (10.46%) e la Public Environmental & Occupational Health (9.88%), quest'ultima particolarmente significativa poiché riflette il crescente interesse verso gli effetti psico-fisici dell'esposizione al razzismo digitale. Risulta altrettanto rilevante la ricerca in ambito educativo (6.04%). Il nucleo delle scienze sociali si evince dai contributi provenienti da Ethnic Studies (5.97%),

ISBN Open Access: 9788835191971

Interdisciplinary Social Sciences (5.78%) e Political Science (3.55%). Parallelamente, la dimensione psicologica individuale e collettiva del fenomeno trova spazio nelle categorie di Psychology, Multidisciplinary (5.75%) e Social Psychology (3.91%). La presenza dell'area Computer Science – Information Systems (3.55%) indica il ruolo crescente degli approcci computazionali, fondamentali per l'analisi automatizzata dei contenuti d'odio e per lo sviluppo di modelli predittivi e sistemi di identificazione. Nel complesso, questa distribuzione testimonia la natura intrinsecamente ibrida del razzismo digitale, un fenomeno che integra prospettive socioculturali, psico-sanitarie e computazionali.

3.3. Mappa concettuale e analisi delle parole chiave

La rete delle parole-chiave maggiormente ricorrenti è raffigurata nella Figura 2. A partire dalle 7140 parole-chiave complessive, sono stati selezionati i termini con almeno 10 occorrenze, per un totale di 189 *keyword*, caratterizzate da 6629 connessioni e organizzate in sei cluster tematici distinti.

Il cluster più esteso (in rosso), composto da 47 termini, raccoglie parole come *discrimination*, *intersectionality*, *mental health*, *depression*, *sexism* e *black woman*, evidenziando una forte attenzione verso le dimensioni psicosociali e identitarie del razzismo, con particolare riferimento all'intreccio tra discriminazione, genere, salute mentale e condizioni di vulnerabilità. Il secondo cluster (in verde), costituito da 35 elementi, comprende invece termini quali *social media, Twitter, Facebook, islamophobia*, *refugees* e *political communication*, delineando un'area di ricerca centrata sui contesti digitali e sulle dinamiche sociopolitiche della comunicazione online. La rete mette in luce le connessioni più significative, tra cui spicca il forte legame tra *racism* e social media, coerente con il focus del presente studio. Altre connessioni rilevanti coinvolgono *COVID-19*, *hate speech*, *Twitter* e *discrimination*, indicando l'emergere di temi altrettanto interconnessi. Il terzo cluster (blu) è incentrato sul termine *COVID-19* ed è caratterizzato da un'elevata intensità di collegamenti (522). Tale raggruppamento segnala una traiettoria di ricerca che ha acquisito crescente rilevanza negli ultimi anni. Il termine *COVID-19* si associa a parole come *pandemic*, *coronavirus*, *health equity*, *cultural safety*, *health disparities* e *health care*, indicando una riflessione sugli intrecci tra crisi sanitaria, disuguaglianze e forme di razzismo strutturale. Dalla letteratura recente emerge che la pandemia ha agito come catalizzatore di nuove ondate di xenofobia e discriminazione mediate socialmente. Diversi studi hanno documentato un aumento significativo della sinofobia e dell'ostilità verso persone di origine asiatica a seguito dell'identificazione geografica del

ISBN Open Access: 9788835191971

virus (Reny e Barreto, 2022). Inoltre, nel corso dell'emergenza sanitaria si sono registrati episodi di stigmatizzazione nei confronti di alcune comunità nazionali e religiose, dagli italiani nelle prime fasi della pandemia in Europa, alle comunità musulmane in India, divenute bersaglio di accuse e stereotipi dopo focolai legati a raduni religiosi.

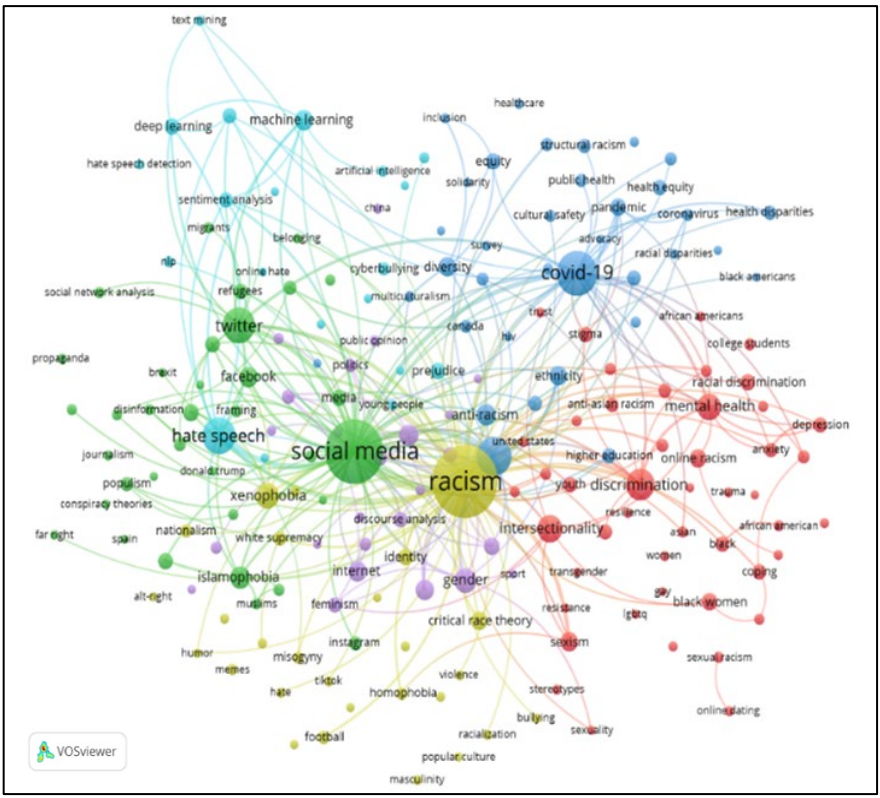


Fig. 2 – Rete delle parole-chiave più ricorrenti

3.4. Pubblicazioni fondanti e tuning point

L'analisi delle opere più citate (tab. 3) evidenzia i contributi che hanno segnato le principali svolte teoriche e metodologiche.

Lo studio di Matamoros-Fernández (2017) su *platformed racism* è stato tra i primi a mostrare come il razzismo sia modellato e amplificato dalle infrastrutture delle piattaforme digitali. A questo si affiancano contributi legati alla comprensione dei discorsi d'odio online: Awan (2016) ha presentato una

ISBN Open Access: 9788835191971

delle prime analisi sistematiche dell'islamofobia sui social media, mentre Farkas, Schou e Neumayer (2018) hanno indagato la natura strategica dei contenuti manipolativi attraverso lo studio delle *fake Facebook pages*. Dal lavoro di Litchfield *et al.* (2018), si evince come il sessismo, il razzismo e gli stereotipi si intersechino, costituendo un campo di battaglia per le politiche dell'identità. Recentemente, studi come quelli di Bouvier e Machin (2021) sulla *cancel culture* e di Carlson e Frazer (2020) sui social media indigeni hanno spostato l'attenzione verso forme di resistenza, contronarrazioni e pratiche di agency digitale che si contrappongono alle dinamiche discriminatorie.

Tab. 3 – Principali contributi metodologici e relativo numero di citazioni (Top 10)

N.	Autori	Anno	Titolo	#cit.
1	Matamoros-Fernández, A.	2017	Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube	247
2	Litchfield, C.; Kavanagh, E.; Osborne, J; Jones, I.	2018	Social media and the politics of gender, race and identity: the case of Serena Williams	86
3	Farkas, J.; Schou, J.; Neumayer, C.	2018	Cloaked Facebook pages: Exploring fake Islamist propaganda in social media	80
4	Bouvier, G.; Machin, D.	2021	What gets lost in Twitter 'cancel culture' hashtags? Calling out racists reveals some limitations of social justice campaigns	47
5	Merrill, S.; Åkerlund, M.	2018	Standing Up for Sweden? The Racist Discourses, Architectures and Affordances of an Anti-Immigration Facebook Group	44
6	Carlson, B.; Frazer, R.	2020	They Got Filters: Indigenous Social Media, the Settler Gaze, and a Politics of Hope	34
7	Kor-Sins, R.	2023	The alt-right digital migration: A heterogeneous engineering approach to social media platform branding	27
8	Miller, G.H.; Marquez-Velarde, G.; Williams, A.A.; Keith, V.M.	2021	Discrimination and Black Social Media Use: Sites of Oppression and Expression	27
9	Awan, I.	2016	Islamophobia on Social Media: A Qualitative Analysis of the Facebook's Walls of Hate	62
10	Matamoros-Fernández, A.; Rodriguez, A.; Wikström, P.	2022	Humor That Harms? Examining Racist Audio-Visual Memetic Media on TikTok During Covid-19	22

ISBN Open Access: 9788835191971

3.5. Struttura concettuale ed evoluzione dei temi

Al fine di ottenere una panoramica dettagliata delle relazioni tematiche nel campo del razzismo online, è stata realizzata una mappa tematica tramite l'algoritmo Louvain (Fig. 3).

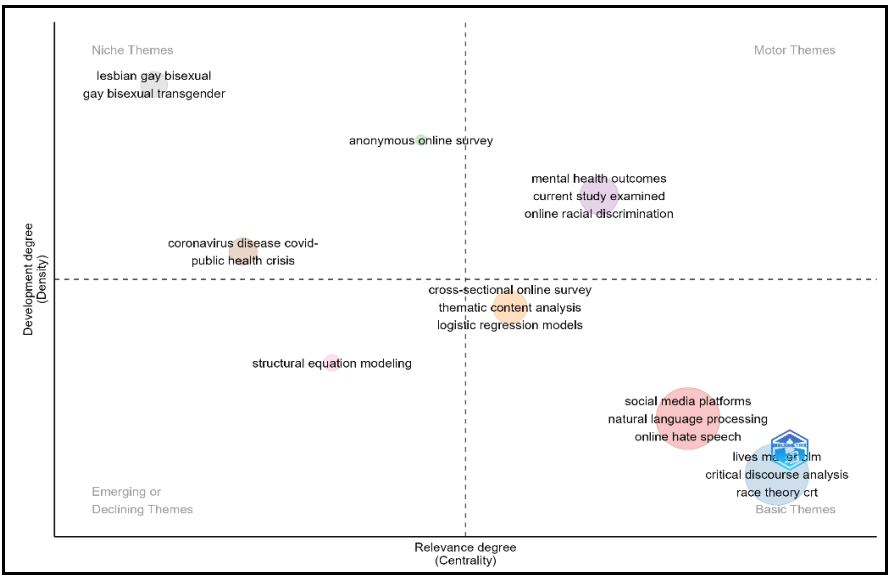


Fig. 3 – Mappa tematica

Essa indica, in base alla centralità, quanto un tema sia rilevante all'interno del campo di studi e, in base alla densità, quanto esso sia sviluppato e consolidato. Dalla mappa emerge, tra i *Motor Themes,* il cluster che collega "*mental health outcomes*" e "*online racial discrimination*", a indicare che gli studi più influenti e sviluppati si concentrano sull'impatto della discriminazione razziale online sulla salute mentale, un'area di ricerca matura e centrale. Tra i *Basic Themes* emergono i cluster su "*social media platforms / natural language processing / online hate speech*" e "*Black Lives Matter (BLM) / critical discourse analysis / Critical Race Theory (CRT)*", che rappresentano temi fondamentali ma ancora in fase di consolidamento. Questo suggerisce che l'analisi computazionale dell'odio online e lo studio dei movimenti antirazzisti sono aree con potenziale di crescita. I temi di nicchia sono quelli relativi a "*lesbian gay bisexual / gay bisexual transgender*" e "*coronavirus disease covid / public health crisis*" ed indicano filoni di ricerca specializzati che esplorano l'intersezione tra discriminazione online e comunità LGBTQ+, così come l'aumento dei fenomeni razzisti durante la pande-

ISBN Open Access: 9788835191971

mia. Il tema "*structural equation modeling*" si presenta come un topic emergente/in declino, suggerendo che alcune metodologie statistiche tradizionali stiano cedendo il passo ad approcci più innovativi, come il NLP. Infine, al centro della mappa troviamo il cluster "*cross-sectional online survey / thematic content analysis / logistic regression models*". Questo tema rappresenta il nucleo metodologico della ricerca sul razzismo e xenofobia online. La sua posizione centrale indica che questi approcci metodologici fungono da ponte tra tutti gli altri temi della mappa.

La Tabella 4 mostra l'evoluzione della ricerca in tre periodi.

Tab. 4 – Evoluzione dei termini nella ricerca su razzismo e xenofobia online (Top 10); F=frequenza; C=centralità

2001 – 2008	F	C	2009 – 2016	F	C	2017 – 2025	F	C
multicultural education	2	.05	social media	17	.03	social media	459	.3
cultural diversity	2	.05	discourse analysis	7	.05	hate speech	151	.14
hate groups	2	0	hate speech	5	.04	mental health	85	.1
coalition formation	1	0	public opinion	4	0	black lives matter	55	.11
disgust sensitivity	1	0	south Africa	4	0	machine learning	48	.04
social network analysis	2	0	sexual orientation	3	.01	critical race theory	39	.07
fear of death	1	0	sexual racism	3	0	online racism	38	.03
intergroup bias	1	0	computer-mediated communication	3	.06	black women	36	.16
web communities	1	0	gay men	3	0	racial discrimination	32	.03
web mining	2	0	young people	3	.01	critical discourse analysis	31	.04

Il primo periodo (2001-2008) rappresenta una fase embrionale, in cui la ricerca è ancora agli albori, con frequenze molto basse e centralità quasi nulle. I temi sono prevalentemente legati all'educazione multiculturale, alla diversità culturale e a concetti psicologici come il bias intergruppo e la sensibilità al disgusto. L'attenzione al web è marginale.

Il secondo periodo (2009-2016) rappresenta una fase di transizione. Emerge prepotentemente il termine "social media" (17 occorrenze), segnando il riconoscimento delle piattaforme digitali come spazio rilevante per studiare questi fenomeni. Compare "hate speech" e si affacciano temi legati alla discriminazione sessuale online.

ISBN Open Access: 9788835191971

Il terzo periodo (2017-2025) rappresenta la fase di esplosione e maturazione: "social media" raggiunge 459 occorrenze con centralità elevata (0.3) ed "hate speech" sale a 151. Emergono tre tendenze chiave: l'ingresso massiccio di metodi computazionali (machine learning, deep learning, NLP, sentiment analysis), l'attenzione agli impatti sulla salute mentale e la centralità di eventi sociopolitici come Black Lives Matter e la pandemia COVID-19.

4. Rassegna qualitativa della letteratura

Per approfondire i contenuti della produzione scientifica emersa dall'analisi bibliometrica, è stata condotta una rassegna qualitativa su un corpus selezionato di pubblicazioni. L'analisi ha riguardato i cinque cluster tematici individuati nella Figura 3 (*mental health outcomes*, *social media platforms*, *lives matter BLM*, *coronavirus disease COVID* e *lesbian gay bisexual*). Per ciascun cluster sono stati selezionati e analizzati gli articoli maggiormente rilevanti, esaminandone in modo integrale i focus tematici, gli approcci metodologici, i risultati principali e le lacune della letteratura.

Il cluster *mental health outcomes* riunisce contributi che mostrano come esperienze di discriminazione, nelle loro forme online e offline, producano effetti destabilizzanti sulla salute mentale di gruppi marginalizzati e razzializzati. Gli studi evidenziano che l'esposizione a razzismo, microaggressioni e discorsi d'odio costituisce un rilevante fattore di rischio spesso associato ad ansia, depressione, stress e peggioramento del benessere psicologico (Lee e Waters, 2021; Moody e Lewis, 2019; Wright e Lewis, 2020). Tali effetti risultano più marcati in condizioni di vulnerabilità strutturale o oppressione multipla, come nel caso delle giovani donne afroamericane soggette contemporaneamente a razzismo e sessismo (Jones *et al.*, 2021; Jones *et al.*, 2022). Il cluster mette inoltre in luce come l'ambiente digitale amplifichi tali dinamiche. Uno studio condotto tra adolescenti afroamericani sostiene che il razzismo online subito durante le tensioni sociali del 2020 ha prodotto un deterioramento immediato dello stato emotivo, con un impatto più elevato rispetto ai loro coetanei (Del Toro e Wang, 2023). Sebbene il supporto sociale e la costruzione identitaria rappresentino risorse importanti, essi possiedono una capacità soltanto parziale nel mitigare tali esiti, poiché non incidono sulle radici strutturali delle discriminazioni (Jones *et al.*, 2021; Lee e Waters, 2021). Questi studi convergono nell'identificare il razzismo come un fattore determinante centrale coinvolto nella salute mentale delle minoranze, generando effetti immediati e cumulativi che alimentano disuguaglianze preesistenti (Lewis *et al.*, 2017) e sottolineano la necessità di interventi non solo individuali ma anche strutturali.

ISBN Open Access: 9788835191971

Per quanto riguarda il cluster *social media platforms*, si evidenzia come le piattaforme costituiscano spazi complessi in cui comunicazione, politica e identità culturali trovano il loro intreccio. Esse non sono meri canali di trasmissione, ma attori che contribuiscono alla costruzione di significati, influenzando la visibilità e viralità dei contenuti e modellando le interazioni tra utenti (Lim, 2017; Matamoros-Fernández, 2017). Numerosi studi indagano la diffusione di discorsi d'odio online e mostrano come modelli automatizzati possano supportare la rilevazione di contenuti tossici, pur richiedendo un'attenta considerazione del contesto semantico (Badjatiya *et al.*, 2017; Founta *et al.*, 2019; Mozafari, Farahbakhsh e Crespi, 2019). Al tempo stesso, emerge come le piattaforme possano veicolare forme di attivismo digitale (Fischer, 2016; Bosch, 2017), facilitando la promozione di campagne di sensibilizzazione e la partecipazione giovanile nella costruzione di spazi pubblici critici. Tuttavia, le dinamiche di condivisione delle informazioni sui social possono generare bolle autoreferenziali (*echo chambers*) capaci di amplificare divisioni sociali, culturali e politiche, favorendo il rafforzamento di nazionalismi, identità tribali e discorsi settari (Criss *et al.*, 2021; DeCook, 2018; Lim, 2017). L'uso di contenuti visivi e la crescente polarizzazione contribuiscono a consolidare identità di gruppo nonché a facilitare la diffusione di discorsi di odio (*hate speech*), il cui impatto è amplificato dall'istantaneità della comunicazione online (Brown, 2018). Tutto ciò conferma come le piattaforme abbiano un ruolo di mediazione culturale e politica ben più significativo di quello di semplici canali di comunicazione (Matamoros-Fernández, 2017).

Il cluster *black lives matter* (BLM) mostra come il movimento (nato negli Stati Uniti a seguito di alcuni episodi di violenza razziale) abbia ridefinito forme contemporanee di protesta e comunicazione politica attraverso l'uso dei social media, rafforzando l'attenzione pubblica su razzismi, discriminazioni e critica al *white privilege*, generando cambiamenti duraturi nel dibattito sulla disuguaglianza razziale (Dunivin *et al.*, 2022). La struttura decentralizzata del movimento e la sua dimensione internazionale hanno trasformato esperienze individuali in narrazioni collettive capaci di raggiungere ampi segmenti sociali, riuscendo a spostare questioni relative allo sviluppo e al benessere delle comunità nere dai margini al centro del dibattito (Nartey, 2023).

Il movimento BLM adotta un approccio intersezionale, amplificando le voci di gruppi marginalizzati all'interno della stessa comunità nera (musulmani afroamericani, donne, persone LGBTQ+) (Auston, 2017; Estes, Straub e Leòn-Corwin, 2023; Nummi, Jennings e Feagin, 2019) combinando attivismo politico, spiritualità e pratiche culturali come forme di resistenza. È stato, inoltre, portato alla luce il ruolo dei media e degli attori istituzionali nella delegittimazione delle proteste, evidenziando come queste siano state

ISBN Open Access: 9788835191971

spesso rappresentate come violente o radicali (Banks, 2018), con un conseguente ridimensionamento delle richieste di riforme strutturali. Le rappresentazioni digitali della comunità evidenziano la persistenza di stereotipi e razzismi online, spesso amplificati da logiche commerciali e dalla spettacolarizzazione del dolore (Sobande, 2021), richiamando l'urgenza di includere la voce dei soggetti coinvolti nella produzione culturale digitale.

Il cluster *coronavirus disease* (COVID) evidenzia come la pandemia abbia generato forme specifiche di stigmatizzazione e discriminazione, amplificando disuguaglianze strutturali preesistenti (Jones, 2021), con effetti significativi online e offline. La diffusione sui social media di contenuti razzisti ha alimentato processi di alterizzazione nei confronti delle comunità asiatiche, rappresentate come potenziali vettori del virus, contribuendo alla propagazione di discorsi d'odio (Abd-Alrazaq *et al.*, 2020; Wang e Santos, 2022) e disinformazione online (Rovetta e Bhagavathula, 2020). Parallelamente, le comunità colpite hanno riportato elevati livelli di preoccupazione per episodi di discriminazione percepita e/o anticipata (Banerjee *et al.*, 2020), con impatti psicologici negativi su adulti e bambini, particolarmente evidenti nelle famiglie cinesi americane (Cheah *et al.*, 2020; Huynh, Raval e Freeman, 2022). La letteratura evidenzia inoltre come la sfiducia verso le istituzioni sanitarie, storicamente radicata in alcune comunità afroamericane, abbia influenzato la risposta pubblica all'emergenza e l'adozione di comportamenti preventivi, tanto che l'efficacia dei messaggi di prevenzione è risultata maggiore quando veicolati da personale medico con lo stesso background culturale (Alsan *et al.*, 2021). Le piattaforme digitali emergono come spazi centrali ma problematici per l'accesso alle informazioni sanitarie: se da un lato costituiscono una delle fonti più utilizzate per reperire informazioni legate alla pandemia, dall'altro alimentano confusione, incomprensione e sfiducia (Chandler *et al.*, 2021), a causa della circolazione di notizie poco accurate o contraddittorie (Rathore e Farooq, 2020; Sasidharan *et al.*, 2020).

La letteratura relativa al cluster *lesbian gay bisexual* evidenzia come le persone appartenenti a minoranze etniche e razzializzate vivano esperienze modellate dall'intreccio tra discriminazioni razziali, omofobia e transfobia. La costruzione identitaria e la capacità di resilienza dei giovani LGBTQ+ *of color*, così come vengono descritte in Singh (2013) si sviluppano in un contesto caratterizzato da queste molteplici forme di oppressione.

Tale pressione intersezionale si traduce in esiti concreti: le discriminazioni multiple sono spesso associate a un grado maggiore di disagio psicologico, ideazione suicidaria e vulnerabilità mentale (Sutter e Perrin, 2016; Pham e Borton, 2024). Allo stesso tempo, l'interiorizzazione dell'oppressione subita può contribuire allo sviluppo di comportamenti a rischio con esiti negativi sullo stato di salute generale degli individui (Drazdowski *et al.*,

ISBN Open Access: 9788835191971

2016). Nello specifico, microaggressioni razziali, rifiuto familiare e omofobia/transfobia interiorizzata emergono come predittori significativi di malessere psicologico tra i giovani appartenenti a questa comunità (Salerno *et al.*, 2023). Tale impatto si estende anche alle attitudini sanitarie, portando a una maggiore riluttanza a cercare aiuto psicologico e a paure relative a screening medici (James e Bowling, 2025). Emerge, dunque, l'urgenza di promuovere approcci clinici e di ricerca che tengano conto dell'intreccio tra differenti background culturali/etnici e sessualità/orientamento sessuale, evidenziando come le pressioni derivanti dallo status di minoranza e l'interiorizzazione dello stigma contribuiscano congiuntamente a plasmare le esperienze di questa popolazione.

Considerati nel loro insieme, i cinque cluster delineano un quadro interpretativo coerente e complesso. La dimensione digitale si configura come uno spazio intrinsecamente ambivalente: se da un lato amplifica dinamiche di potere e discriminazione preesistenti (come razzismo, hate speech e delegittimazione), dall'altro offre alle comunità marginalizzate opportunità inedite per esercitare forme di azione collettiva, costruire reti di solidarietà e produrre contronarrazioni efficaci. L'analisi dimostra come discriminazioni strutturali, progressi tecnologici e condizioni sociopolitiche operino in modo strettamente interconnesso. Questa interazione può generare vulnerabilità che si accumulano e si estendono attraverso gruppi sociali e contesti geografici differenti. Parallelamente, emerge il potenziale trasformativo delle pratiche collettive (dal mutuo aiuto comunitario alle mobilitazioni transnazionali) nel ridisegnare i confini della cittadinanza digitale e nel mettere in discussione gerarchie razziali e culturali consolidate. Questi risultati sottolineano l'urgenza di elaborare approcci analitici e interventi istituzionali che non solo integrino le dimensioni materiali, simboliche e tecnologiche dei fenomeni studiati, ma che riconoscano la natura strutturale e sistemica delle disuguaglianze e, al contempo, la centralità delle esperienze vissute dalle persone che ne sono bersaglio.

5. Conclusioni

Attraverso l'integrazione di analisi bibliometriche quantitative e di una rassegna qualitativa tematica, questo studio offre una mappatura sistematica e interdisciplinare della ricerca su razzismo e xenofobia online, ricostruendone l'evoluzione, la struttura concettuale e le principali linee di sviluppo.

I risultati mostrano un campo ormai maturo e in rapida espansione, con una crescita particolarmente marcata dal 2015 in poi. La produzione scientifica, prevalentemente anglofona e trainata dagli Stati Uniti, si distingue per un forte impatto accademico e per un'elevata interdisciplinarità, che coinvol-

ISBN Open Access: 9788835191971

ge comunicazione, sociologia, psicologia, salute pubblica e scienze informatiche. Questa varietà riflette la natura complessa del fenomeno, situato all'incrocio tra dinamiche socioculturali, effetti psico-sociali sulle persone e funzionamento delle infrastrutture digitali. Il nucleo concettuale della letteratura ruota attorno a temi centrali: l'impatto sulle comunità marginalizzate, l'ambivalenza delle piattaforme digitali, al tempo stesso amplificatrici di odio e spazi di empowerment, le nuove forme di mobilitazione antirazzista, il ruolo della pandemia nell'accentuare stigma e discriminazioni, e la prospettiva intersezionale che evidenzia la compresenza di più forme di oppressione. La fase attuale della ricerca è segnata dall'uso crescente di metodologie computazionali e da un forte legame con i cambiamenti sociopolitici globali. Emerge, dunque, un ecosistema digitale intrinsecamente ambivalente: un luogo che può riprodurre e amplificare le disuguaglianze, ma anche favorire pratiche di resistenza, solidarietà e costruzione di contronarrazioni.

La sfida principale consiste nel trasformare questa complessità in approcci integrati che combinino analisi tecnologiche, interventi psico-sociali e politiche inclusive, mettendo al centro le esperienze e le voci delle persone direttamente colpite. Solo così è possibile avanzare verso un ambiente digitale più equo e democratico.

Riferimenti bibliografici

Adb-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M. and Shah, Z. (2020), Top concerns of tweeters during the COVID-19 pandemic: infoveillance study, *Journal of Medical Internet Research*, 22, 4: e19016.

Alsan, M., Stanford, F.C., Banerjee, A., Breza, E., Chandrasekhar, A.G., Eichmeyer, S. and Duflo, E. (2021), Comparison of knowledge and information-seeking behavior after general COVID-19 public health messages and messages tailored for Black and Latinx communities: a randomized controlled trial, *Annals of Internal Medicine*, 174, 4: 484-492.

Aria, M. and Cuccurullo, C. (2017), *bibliometrix*: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11, 4: 959-975.

Auston, D. (2017), Prayer, protest, and police brutality: Black Muslim spiritual resistance in the Ferguson era, *Transforming Anthropology*, 25, 1: 11-22.

Awan, I. (2016), Islamophobia on Social Media: A Qualitative Analysis of the Facebook's Walls of Hate, *International Journal of Cyber Criminology*, 10, 1: 1.

Badjatiya, P., Gupta, S., Gupta, M. and Varma, V. (2017), Deep learning for hate speech detection in tweets. In *Proceedings of the 26th international conference on World Wide Web companion*, 759-760.

Banerjee, D., Vaishnav, M., Rao, T.S., Raju, M.S.V.K., Dalal, P.K., Javed, A. and Jagiwala, M.P. (2020), Impact of the COVID-19 pandemic on psychosocial health and well-being in South-Asian (World Psychiatric Association zone 16)

ISBN Open Access: 9788835191971

countries: A systematic and advocacy review from the Indian Psychiatric Society, *Indian Journal of Psychiatry*, 62, 3: S343-S353.
Banks, C. (2018), Disciplining Black activism: post-racial rhetoric, public memory and decorum in news media framing of the Black Lives Matter movement, *Continuum*, 32, 6: 709-720.
Bosch, T. (2017), Twitter activism and youth in South Africa: The case of# RhodesMustFall, Information, *Communication & Society*, 20, 2: 221-232.
Bouvier, G. and Machin, D. (2021), What gets lost in Twitter 'cancel culture' hashtags? Calling out racists reveals some limitations of social justice campaigns, *Discourse & Society*, 32, 3: 307-327.
Brown, A. (2018), What is so special about online (as compared to offline) hate speech?, *Ethnicities*, 18, 3: 297-326.
Carlson, B. and Frazer, R. (2020), They Got Filters: Indigenous Social Media, the Settler Gaze, and a Politics of Hope, *Social Media+ Society*, 6, 2: 2056305120925261.
Chandler, R., Guillaume, D., Parker, A.G., Mack, A., Hamilton, J., Dorsey, J. and Hernandez, N.D. (2021), The impact of COVID-19 among Black women: evaluating perspectives and sources of information, *Ethnicity & Health*, 26, 1: 80-93.
Cheah, C.S., Wang, C., Ren, H., Zong, X., Cho, H.S. and Xue, X. (2020), COVID-19 racism and mental health in Chinese American families, *Pediatrics*, 146, 5.
Chen, C. (2006), CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, *Journal of the American Society for information Science and Technology*, 57, 3: 359-377.
Criss, S., Michaels, E.K., Solomon, K., Allen, A.M. and Nguyen, T.T. (2021), Twitter fingers and echo chambers: Exploring expressions and experiences of online racism using twitter, *Journal of Racial and Ethnic Health Disparities*, 8, 5: 1322-1331.
DeCook, J.R. (2018), Memes and symbolic violence:# proudboys and the use of memes for propaganda and the construction of collective identity, *Learning, Media and Technology*, 43, 4: 485-504.
Del Toro, J. and Wang, M.T. (2023), Online racism and mental health among Black American adolescents in 2020, *Journal of the American Academy of Child & Adolescent Psychiatry*, 62, 1: 25-36.
Drazdowski, T.K., Perrin, P.B., Trujillo, M., Sutter, M., Benotsch, E.G. and Snipes, D.J. (2016), Structural equation modeling of the effects of racism, LGBTQ discrimination, and internalized oppression on illicit drug use in LGBTQ people of color*, Drug and Alcohol Dependence*, 159: 255-262.
Dunivin, Z.O., Yan, H.Y., Ince, J. and Rojas, F. (2022), Black Lives Matter protests shift public discourse, *Proceedings of the National Academy of Sciences*, 119, 10: e2117320119.
Estes, M.L., Straub, A.M. and León-Corwin, M. (2023), Making the invisible visible: examining Black women in Black Lives Matter, *Sociological Spectrum*, 43, 4-5: 127-146.
Farkas, J., Schou, J. and Neumayer, C. (2018), Cloaked Facebook pages: Exploring fake Islamist propaganda in social media, *New Media & Society*, 20, 5: 1850-1867.

ISBN Open Access: 9788835191971

Fischer, M. (2016), # Free_CeCe: the material convergence of social media activism, *Feminist Media Studies*, 16, 5: 755-771.
Founta, A.M., Chatzakou, D., Kourtellis, N., Blackburn, J., Vakali, A. and Leontiadis, I. (2019), A unified deep learning architecture for abuse detection, In *Proceedings of the 10th ACM conference on web science*, 105-114.
Huynh, V.W., Raval, V.V. and Freeman, M. (2022), Ethnic-racial discrimination towards Asian Americans amidst COVID-19, the so-called "China" virus and associations with mental health, *Asian American Journal of Psychology*, 13, 3: 259.
James, D. and Bowling, A.M. (2025), Intersectional discrimination, internalized heterosexist racism, and health attitudes among gay and bisexual Black American men: A path analysis approach, *Psychology of Sexual Orientation and Gender Diversity*.
Jones, J.M. (2021), The dual pandemics of COVID-19 and systemic racism: Navigating our path forward, *School Psychology*, 36, 5: 42
Jones, M.S., Womack, V., Jérémie-Brink, G. and Dickens, D.D. (2021), Gendered racism and mental health among young adult US Black women: The moderating roles of gendered racial identity centrality and identity shifting, *Sex Roles*, 85, 3: 221-231.
Jones, M.K., Leath, S., Settles, I.H., Doty, D. and Conner, K. (2022), Gendered racism and depression among Black women: Examining the roles of social support and identity, *Cultural Diversity & Ethnic Minority Psychology*, 28, 1: 39.
Kor-Sins, R. (2023), The alt-right digital migration: A heterogeneous engineering approach to social media platform branding, *New Media & Society*, 25, 9: 2321-2338.
Lee, S. and Waters, S.F. (2021), Asians and Asian Americans' experiences of racial discrimination during the COVID-19 pandemic: Impacts on health outcomes and the buffering role of social support, *Stigma and Health*, 6, 1: 70.
Lewis, J.A., Williams, M.G., Peppers, E.J. and Gadson, C.A. (2017), Applying intersectionality to explore the relations between gendered racism and health among Black women, *Journal of Counseling Psychology*, 64, 5: 475.
Lim, M. (2017), Freedom to hate: social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia, *Critical Asian Studies*, 49, 3: 411-427.
Litchfield, C., Kavanagh, E., Osborne, J. and Jones, I. (2018), Social media and the politics of gender, race and identity: the case of Serena Williams, *European Journal for Sport and Society*, 15, 2: 154-170.
Matamoros-Fernández, A. (2017), Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube, *Information, Communication & Society*, 20, 6: 930-946.
Matamoros-Fernández, A., Rodriguez, A. and Wikström, P. (2022), Humor That Harms? Examining Racist Audio-Visual Memetic Media on TikTok During Covid-19, *Media and Communication*, 10, 2: 180-191.
Merrill, S. and Åkerlund, M. (2018), Standing Up for Sweden? The Racist Discourses, Architectures and Affordances of an Anti-Immigration Facebook Group, *Journal of Computer-Mediated Communication*, 23, 6: 332-353.

ISBN Open Access: 9788835191971

Miller, G.H., Marquez-Velarde, G., Williams, A.A. and Keith, V.M. (2021), Discrimination and Black Social Media Use: Sites of Oppression and Expression, *Sociology of Race and Ethnicity*, 7, 2: 247-263.
Moody, A.T. and Lewis, J.A. (2019), Gendered racial microaggressions and traumatic stress symptoms among Black women, *Psychology of Women Quarterly*, 43, 2: 201-214.
Mozafari, M., Farahbakhsh, R. and Crespi, N. (2019), A BERT-based transfer learning approach for hate speech detection in online social media. In *International Conference on Complex Networks and their Applications* (pp. 928-940).
Nartey, M. (2023), Centering marginalized voices: A discourse analytic study of the Black Lives Matter movement on Twitter. In *Voice, agency and resistance* (pp. 65-80).
Nummi, J., Jennings, C. and Feagin, J. (2019), # BlackLivesMatter: Innovative black resistance, *Sociological Forum*, 34: 1042-1064.
Pham, M.D. and Borton, J.L. (2024), "Are you a homophobic racist?": Applying lay theory of generalized prejudice to the discrimination-distress link, *Cultural Diversity & Ethnic Minority Psychology*, 30, 2: 273.
Phillips, W. and Milner, R.M. (2018), *The ambivalent internet: Mischief, oddity, and antagonism online*, John Wiley & Sons, USA.
Rathore, F.A. and Farooq, F. (2020), Information overload and infodemic in the COVID-19 pandemic, *Journal of the Pakistan Medical Association*, 70, 5: S162-S165.
Reny, T.T. and Barreto, M.A. (2022), Xenophobia in the time of pandemic: Othering, anti-Asian attitudes, and COVID-19, *Politics, Groups, and Identities*, 10, 2: 209-232.
Rovetta, A. and Bhagavathula, A.S. (2020), COVID-19-related web search behaviors and infodemic attitudes in Italy: infodemiological study, *JMIR public health and surveillance*, 6, 2: e19374.
Salerno, J.P., Pease, M.V., Gattamorta, K.A., Fryer, C.S. and Fish, J.N. (2023), Impact of racist microaggressions and LGBTQ-related minority stressors: Effects on psychological distress among LGBTQ+ young people of color, *Preventing Chronic Disease*, 20: E63.
Sasidharan, S., Singh, D.H., Vijay, S. and Manalikuzhiyil, B. (2020), COVID-19: Pan (info) demic, *Turkish Journal of Anaesthesiology and Reanimation*, 48, 6: 438.
Singh, A.A. (2013), Transgender youth of color and resilience: Negotiating oppression and finding support, *Sex Roles*, 68, 11: 690-702.
Sobande, F. (2021), Spectacularized and branded digital (re) presentations of black people and blackness, *Television & New Media*, 22, 2: 131-146.
Sutter, M. and Perrin, P.B. (2016), Discrimination, mental health, and suicidal ideation among LGBTQ people of color, *Journal of Counseling Psychology*, 63, 1: 98.
Van Eck, N. and Waltman, L. (2010), Software survey: VOSviewer, a computer program for bibliometric mapping, *Scientometrics*, 84, 2: 523-538.
Wang, S.C. and Santos, B.M.C. (2022), "Go back to China with your (expletive)

ISBN Open Access: 9788835191971

virus": A revelatory case study of anti-Asian racism during COVID-19, *Asian American Journal of Psychology*, 13, 3: 220.
Wright, L.N. and Lewis, J.A. (2020), Is physical activity a buffer? Gendered racial microaggressions and anxiety among African American women, *Journal of Black Psychology*, 46, 2-3: 122-143.

ISBN Open Access: 9788835191971

Razzismo sistemico e disuguaglianze naturali. Le vecchie e nuove forme di razzismo

di *Alfredo Alietti*[*], *Dario Padovan*[**]

1. La prospettiva egemonica del razzismo

Nell'epoca contemporanea è oltremodo evidente quanto il razzismo sia diffuso e pervasivo nelle nostre società. Negli ultimi decenni tra le due sponde dell'Atlantico i discorsi e le rappresentazioni dello straniero nelle sue molteplici forme hanno assunto una loro centralità politica, sovente con una retorica aggressiva che nutre risentimento e odio sociale. Gli esempi di questa dinamica sono diversi e mostrano caratteri molto simili tra loro. Il successo elettorale dei vari partiti di estrema destra chiaramente xenofobi nei principali paesi europei, così come il rinnovato trionfo del trumpismo, raffigurano un quadro assai fosco. L'ostilità verso i soggetti migranti razzializzati non è certo una novità, nondimeno attualmente vi è stato un progressivo sdoganamento della logica razzista. Alla nota asserzione "*Io non sono razzista, ma*…" la quale, pur con tutta la sua ambiguità e ambivalenza, mostrava comunque il riconoscimento morale di un sentimento negativo, si è sostituita l'asserzione "*Io sono razzista, perciò*…" che liquida senza alcun compromesso retorico tale prospettiva. Ciò che fino a poco tempo fa veniva chiamato "razzismo sottile", "nascosto" o "democratico" attraverso cui si rilevava una premura contro l'aperta espressione di avversione è chiaramente superato dalla normalizzazione del razzismo. In una ricerca condotta nel contesto americano si conferma come i bianchi esprimano più apertamente il linguaggio razzista e comportamenti razzisti sia in contesti pubblici, sia nelle reti private mono-razziali (Picca e Feagin, 2007).

Tornando a ragionare sulla dimensione sociopolitica, è evidente che tutto

[*] Dipartimento di Studi Umanistici, Università degli Studi di Ferrara, alfredo.alietti @unife.it

[**] Dipartimento di Culture, Politica e Società, Università degli Studi di Torino, dario.padovan@unito.it

ISBN Open Access: 9788835191971

l'armamentario xenofobo promosso nel dibattito pubblico sia divenuto, anche in questo ambito una normalità, non più un'eccezione, che agevole e rinforza il linguaggio pregiudiziale e xenofobo. Le accuse rivolte verso i migranti e i rifugiati agisce su due piani che interagiscono tra loro definendo ciò che abbiamo chiamato xeno-populismo (Alietti e Padovan, 2020): da un lato, vi è il ricorso a una strategia discorsiva che paventa gli effetti nocivi della globalizzazione e delle nuove élite transnazionali che vanno contro gli interessi del popolo agevolando il multiculturalismo; dall'altro vi è una strategia politica tesa a difendere le nostre tradizioni, il nostro standard di vita e i nostri privilegi in quanto bianchi messi a repentaglio dalla presenza di individualità e gruppi aliene, ulteriore esito dei processi globali.

In questo caso, risulta determinante la costruzione di un nemico interno che implica, oltre alla sua demonizzazione in quanto estraneità, la sua disumanizzazione. Infatti, la fobia si trasforma in razzismo poiché trasforma il portatore di una potenziale minaccia in un "altro razzializzato" entro cui si delinea una continuità *in negativo* tra caratteri etnici e comportamenti morali e dove le differenze culturali sono cristallizzate una volta per tutte (Alietti e Padovan, 2020, p. 11).

Qui gioca la naturalizzazione della cultura che raffigura da tempo il fondamento di ciò che è descritto come "nuovo razzismo" per cui, a partire dalla distorsione dell'assioma relativista, le peculiarità etniche sono inassimilabili e quindi non in grado di dialogare ma solo di confliggere (Taguieff, 1994; Wieworka, 1993; Barker, 1981).

La trasformazione della cultura in una sorta di "natura" immodificabile e l'alchimia sociale alla quale sono stati sottoposti primariamente gli ebrei europei ora conosce una traslazione potente consegnando le differenze di origine e di condotte a un'ancestrale diversità irriducibile a qualsiasi mediazione sociale e istituzionale (Alietti, Padovan e Vercelli, 2014). Non è un caso che il linguaggio politico promosso dai movimenti xenopopulisti, ma anche di una parte delle sinistre liberali, si orienti decisamente verso una critica dura e insistita al multiculturalismo, al di là delle sue impostazioni teoriche e ricadute sulla cittadinanza (Kivisto, 2010).

In questo orizzonte è altresì significativo che la stessa democrazia pluralista diventi obiettivo di profonde critiche e alimenti, allo stesso tempo, tendenze autoritarie, le quali ripropongono una gerarchia etno-razziale all'interno e all'esterno in termini neocoloniali. L'ideologia razzista si rafforza nutrendosi, empiricamente, della persistente discriminazione di chi non è membro della comunità maggioritaria raffigurata come un corpo unico. Recentemente altri discorsi, per quanto minoritari, si affacciano per riprodurre tale sentimento collettivo, tra cui la cosiddetta sostituzione etnica, in primis, collegata al fantasma dell'islamico e alle sue peculiarità culturali antioccidentali.

ISBN Open Access: 9788835191971

2. Razzismo come legame sociale: eterorazzizzazione e autorazzizzazione

Il panorama fin qui descritto fornisce il quadro di una progressiva crisi della convivenza e l'avvento egemonico di un pensiero e di una pratica razzializzante. Sulle basi di una ricerca condotta sul tema dell'islamofobia e dell'antisemitismo nel 2010 emergeva come l'odio etno-razziale assumesse i connotati di un collante, di un legame sociale che preclude la differenza e la diversità ai diritti di cittadinanza e, come abbiamo già enunciato, all'umanità (Alietti e Padovan, 2010; 2024). Un legame sociale che si nutre di risentimento, in particolare nei ceti autoctoni vulnerabili, di un "rancore socializzato" nei confronti dello straniero, visto quale usurpatore di risorse esclusive e, di conseguenza, come nemico (Alietti, 2017, p. 10; vedi anche Castel, 2004).

L'arabo, l'ebreo, lo "zingaro", il richiedente asilo, il migrante, il rifugiato assumono i tratti di soggetti che mettono in crisi, e a rischio, la nostra identità e la nostra prerogativa nazionale quale esclusivo criterio di redistribuzione delle risorse sempre più rare. Potremmo affermare da tali premesse che l'ideologia e la prassi razzista sia un modello valido per tutte le soggettività razzializzate e per tutti i contesti analizzabili. Su questo punto fondamentale, valgono le riflessioni di Stuart Hall, il quale afferma che è possibile evidenziare aspetti generali del razzismo sebbene sia necessaria una ricognizione delle modalità con cui essi si modificano e si trasformano nelle specificità storiche del contesto e dell'ambiente in cui divengono attive (Hall, 2006, p. 6; vedi per una discussione Alietti, Padovan e Vercelli, 2014).

In effetti, se dovessimo visionare attentamente l'evoluzione storica del razzismo, ad esempio dal colonialismo classico alle forme neocoloniali attuali, oppure dai vecchi nazionalismi ai neonazionalismi emergerebbero chiaramente talune dissimilarità riguardo i gruppi colpiti, le narrazioni e le azioni messe all'opera. Inoltre, il caso degli afroamericani con il loro portato di schiavitù e di colonialismo interno accompagnato da meccanismi istituzionalizzati di segregazione e discriminazione contribuisce a delineare un quadro complesso e specifico nei suoi caratteri peculiari. Senza dimenticare l'antigiudaismo e il successivo antisemitismo il quale non soltanto ha rappresentato uno dei *topos* della civilizzazione europea e della sua modernità, ma anche è divenuto nel suo unicum il paradigma razziale e razzista.

Tuttavia, è ipotizzabile rintracciare il fatto che tutti i razzismi, a prescindere dalla loro collocazione storica e nazionale, hanno in comune forme di discorso essenzializzante, una comprensione dei fenomeni sociali e storici in termini innati – biologici o culturali. Operano secondo una strategia duale: da un lato, attribuiscono qualità inferiori agli altri, collocandoli così nella

ISBN Open Access: 9788835191971

classica relazione dominante/dominato, assicurando che solo l'altro sia una razza, e quindi inferiore. D'altra parte, affermano non solo che 'siamo i migliori' – coloro che detengono potere, democrazia, ricchezza, tecnologia – ma anche che 'siamo umanità', una non-razza superiore che incarna l'essenza dell'umanità, i suoi migliori e quindi i suoi tratti universali.

Questa attribuzione di uno status identitario inferiore alle minoranze razzializzate influisce sull'errata percezione dell'alterità e determina, lo si è già anticipato, la diseguale distribuzione delle risorse economiche e la partecipazione alla vita sociale. Qui risiede il problema della giustizia distributiva e di riconoscimento, così come è stato analizzato da Nancy Fraser (2007). Il modello è la supremazia bianca che ancora oggi si pone quale chiave di volta nella comprensione del razzismo e delle sue pertinenti declinazioni, a dispetto di chi sostiene di essere nell'epoca delle società post-razziali e "color-blind" (Picca e Feagin, 2007). Come sarà discusso, le citate forme autoritarie all'interno di un sistema economico e produttivo e socioculturale regressivo divengono essenziali nella progressiva affermazione di un mondo a parte, esclusivo la cui famosa metafora della cittadella assediata è parte integrante della retorica sovranista e identitaria bianca che denuncia le migrazioni quale vero e proprio colpo di stato etnico (Bracke e Aguilar, 2023). Nella dialettica tra identità incommensurabili diviene importante riprendere le riflessioni di Taguieff riguardo la presenza di due logiche di razzizzazione all'interno delle configurazioni storiche del razzismo: la prima si esplicita attraverso la serie autorazzizzazione, differenza, purificazione/epurazione, sterminio; la seconda attraverso la serie eterorazzizzazione, disuguaglianza, dominazione, sfruttamento. Queste due sequenze producono due tipi differenti di razzismo.

Il primo afferma la propria identità razziale in quanto gruppo, che solo secondariamente porta ad affermare la propria superiorità sugli altri gruppi. Il secondo è centrato sull'affermazione della differenza razziale basata sull'inferiorità dell'altro. Mentre quest'ultimo meccanismo di eterorazzizzazione è finalizzato alla costituzione di relazioni di dominio, oppressione e sfruttamento – normalmente di tipo economico e orientate all'interesse e al profitto, quello di autorazzizzazione è finalizzato alla costituzione di relazioni di esclusione, che raggiunge il paradosso nello sterminio dell'altro e nella distruzione della relazione di differenza (Taguieff 1987, 163-5).

Le due logiche, qui brevemente descritte, rimandano ovviamente a due razzismi diversi: l'autorazzizzazione genera l'antisemitismo, l'eterorazzizzazione genera il razzismo della schiavitù, coloniale e quello rivolto agli immigrati.

Questa distinzione operata da Tagiueff ci sembra ancora sostanzialmente valida. Aggiungiamo tuttavia che i razzismi del presente operano facendo interagire costantemente tali due logiche. Esse non sono separate o contrap-

ISBN Open Access: 9788835191971

poste, come a volte sembra credere Taguieff, ma operano in contemporanea fondendo logiche identitarie e suprematiste da un lato, e logiche della colonizzazione, dello sfruttamento e delle disuguaglianze dall'altro. In effetti il razzismo israeliano contro i palestinesi, così come quello europeo contro i russi, o quello statunitense contro il resto del mondo, o infine quello italiano contro immigrati e musulmani, è una combinazione di identità etnonazionale e suprematismo sociotecnico ed economico, di rivendicazione di status e di sfruttamento del non-bianco (che si tratti di maggioranza o minoranza razziale). Come vedremo, tale proiezione teorica ed epistemologica ha a che fare con una interessante versione del contratto sociale, quella del cosiddetto "contratto razziale".

3. Razzismo della personalità, istituzionale e sistemico

Questa premessa sintetica ci conduce a ricostruire i molteplici significati che danno forma e contenuto al razzismo, in particolare alla deriva istituzionale delle pratiche e politiche razziste e alla sistematicità di esse che ne favorisce la sua riproducibilità nel tempo.

Ragionare intorno al concetto di razzismo/razzismi fa emergere un punto assai rilevante sul quale la discussione si è protratta per lungo tempo, ovvero se esso si identifica con un atteggiamento individuale oppure se è esito delle strutture sociali. Nella famosa e classica ricerca sulla personalità autoritaria di Adorno pubblicata nel 1950 si delinea un insieme di caratteri dell'individuo che rilevano tendenze aggressive, rigide e potenzialmente antidemocratiche (Adorno *et al.*, 1950). La focalizzazione sui fattori psicologici individuali poneva, e pone tuttora, una sfida alle dimensioni sociali per quanto lo stesso Adorno e collaboratori ritenessero significativo valutare l'impatto di ciò che definiscono "clima culturale" sul pregiudizio etnico e/o antisemita il quale alimenta e amplifica le tendenze autoritarie del soggetto (Alietti e Padovan, 2023).

Nel solco dell'orizzonte psicosociale e in linea con la citazione precedente è assai interessante l'analisi svolta da Dovidio e Gartner (1998) i quali avanzano l'idea del "razzismo democratico" che comporta il mascheramento dei pregiudizi i quali, in realtà, sono accettati e interiorizzati dall'individuo anche in modo inconsapevole. In tal senso, si ripropone il tema dei modelli culturali prevalenti che riproducono relazioni, atteggiamenti discriminanti nei confronti di quei gruppi tradizionalmente collegati, o collegabili, a stereotipi negativi.

A dispetto di questa impostazione psicologica a cui concorre il sociale con le sue strutture significanti nei confronti delle minoranze etniche, il tema

ISBN Open Access: 9788835191971

della patologia razzista quale individualità si è rinnovato all'interno del dibattito intorno alla citata società color-blind e/o post razziale. L'idea di fondo, in riferimento soprattutto agli Stati Uniti, è che vi sia stato un superamento del razzismo come forma endemica e sociologicamente fondata, ovvero la razza non rappresenta più un fattore attivo che discrimina. Ne consegue una doppia individualizzazione che colpisce la vittima e il carnefice: da un lato s'individualizza la responsabilità e la volontà del singolo di integrarsi e non alle condizioni strutturali che producono e riproducono le discriminazioni; dall'altro, chi ideologicamente e praticamente agisce da razzista è frutto di una sua individualità patologica che non trova più ancoraggio in una società senza razze (per una discussione vedi Doane, 2006).

In questo panorama depotenziato è evidente la finalità di assolvere la maggioranza dalle sue prerogative escludenti e di garantire lo status quo delle politiche discriminanti e la persistenza delle disuguaglianze razziali (Bonilla-Silva, 2006). Riprendendo la celebre analisi di Du Bois la "linea del colore" è ancora decisiva nel configurare le relazioni interetniche e i suoi esiti segregativi, in particolare nel mercato del lavoro, abitativo e nell'accesso ai servizi (Du Bois, 2010).

A contrastare tale prefigurazione fin dalla fine degli anni sessanta si affaccia l'analisi del razzismo istituzionale, la cui prima definizione è enunciata nel testo di Carmichael e Hamilton, *Black Power*, in cui si evidenzia con un'ampia analisi di dati che il razzismo negli Stati Uniti contro gli afroamericani non era individuale ma esito di un sistematico pregiudizio che pervadeva tutti gli ambiti istituzionali: creando le condizioni per svantaggi posizionali rispetto ai bianchi dal sistema penale a quello repressivo, dall'accesso al mercato del lavoro e al sistema scolastico, dall'acquisto della casa alle scelte residenziali (Carmichael e Hamilton, 1967).

Interessante notare che questa prospettiva riprende essenzialmente l'intuizione di Herbert Blumer il quale, in un penetrante saggio incredibilmente poco citato, asseriva che la paura che le azioni delle minoranze etno-razziali potessero minacciare la posizione dominante della maggioranza bianca alimenta una rivendicazione e pressione politica per salvaguardare l'esclusività dei diritti su gran parte della vita sociale (Blumer, 1958; per una rivisitazione critica vedi Alietti e Padovan, 2000). Successivamente, il concetto è tornato al centro della discussione in Europa a partire dal famoso Rapporto McPherson pubblicato nel 1999 in seguito all'uccisione Il 22 aprile 1993 di Stephen Lawrence accoltellato a morte a una fermata dell'autobus nel sud di Londra in un attacco razzista non provocato. Nel rapporto si rilevava quanto l'indagine della polizia fosse segnata da incompetenze e, soprattutto, da razzismo istituzionale (House of Commons, 2009). In questo caso quest'ultimo veniva definito come "il fallimento collettivo di un'organizzazione nel fornire un

ISBN Open Access: 9788835191971

servizio adeguato e professionale alle persone a causa del loro colore, cultura o origine etnica. Può essere osservata o rilevata in processi, atteggiamenti e comportamenti che equivalgono a discriminazione attraverso pregiudizi inconsapevoli, ignoranza, insensibilità e stereotipi razziali" (McPherson Report, 1999).

Il fatto che questa forma di razzismo, poiché fa parte dei processi dell'istituzione e non richieda un'intenzione deliberata, merita la giusta attenzione. Come sottolineato da Howard Winant (2004), il razzismo deve essere inteso in termini di conseguenze, non come questione di intenzioni o convinzioni. In sintesi si può affermare come l'istituzionalizzazione chiarisca quei meccanismi sociali e politici che, prefigurando una differenza di trattamento giuridico tra gruppi su basi etniche, getta le fondamenta per politiche securitarie, per legittimare pratiche discriminatorie negli ambiti di vita collettiva, nei servizi favorendo in tal modo la diffusione di calunnie e falsità nei mass-media e social e la formazione di un razzismo popolare e quotidiano a rinforzo dell'esclusione (Alietti e Padovan, 2023).

La terza forma è il razzismo strutturale o sistemico: schemi di svantaggio che emergono dal funzionamento complessivo del sistema globale, spesso accumulati nel corso dei secoli che hanno definito una gerarchia etno-razziale e le logiche ad essa collegate di sfruttamento ed esclusione, come esemplificato dalle eredità coloniali e/o dello schiavismo. Tale prospettiva, come è stato sottolineato, è un'espansione del razzismo istituzionale che si focalizza accentuando la continuità delle profonde strutture razziste che sono a fondamento del sistema sociale (Feagin, 2006). Nelle parole di Bonilla-Silva, la prospettiva strutturalista indica quanto le società siano costituite da contesti politici, sociali, economici, culturali e ideologici strutturati dalla collocazione degli attori in categorie razziali (Bonilla-Silvia, 1997).

Richiama questa analisi, per quanto sia differente l'approccio teorico, il concetto di "progetto razziale" di Omi e Winant ritenuto in grado di alimentare rappresentazioni, interpretazioni e/o spiegazioni delle identità razziali a cui corrisponde uno sforzo per organizzare e distribuire risorse materiali e simboliche secondo queste stesse linee razziali (Omi e Winant, 1986). Ne deriva che l'ordine razziale deve essere prodotto e riprodotto all'interno della varietà delle istituzioni e delle conseguenti azioni messe in atto per garantire privilegi e dominio. Il razzismo sistemico, quindi, si delinea in quanto estensivo sistema gerarchico concepito dalla maggioranza bianca per controllare e soggiogare le minoranze subordinate etno-razziali che ingiustamente arricchisce materialmente e socialmente la prima e, altrettanto ingiustamente, impoverisce materialmente e socialmente le seconde (Elias e Feagin, 2016). Un ulteriore e fondamentale aspetto del razzismo sistemico, oltremodo evidenziato dal pensiero femminista black, è quello di analizzare in maniera appro-

ISBN Open Access: 9788835191971

fondita l'intersezione con la dimensione di genere e di classe quali possibili e operative configurazioni di oppressione e sfruttamento (Crenshaw, 1989; Hill Collins, 1986).

Infine, la questione del razzismo climatico, del quale ci occupiamo ormai da tempo (vedi Padovan e Alietti 2019; 2024), che rientra nella categoria di razzismo sistemico. Non esiste alcun comitato di bianchi che complotta per opprimere l'Africa alterando il clima, ma possiamo identificare il razzismo dai suoi risultati. Le ragioni di ciò sono strutturali e storiche, in quanto le radici dell'ingiustizia climatica sono la schiavitù, il colonialismo e l'impero. Il razzismo climatico è un problema di colonizzazione umana diseguale dell'atmosfera e quindi del sistema climatico da cui scaturiscono conseguenze che sono altrettanto disuguali e che seguono linee di frattura razziali. Il cambiamento climatico, che per alcuni manifesta la più radicale democraticità, può essere visto come un processo profondamente razzializzato poiché causato in modo sproporzionato da stati, organizzazioni e collettività a maggioranza bianca in Paesi a maggioranza bianca, con danni che si scaricano in modo preponderante sulle minoranze etniche e persone non bianche. La crisi climatica riflette e rafforza le ingiustizie razziali e si iscrive, come anticipato, all'interno dell'approccio del "razzismo sistemico". Il fatto è che il cambiamento climatico, in quanto crisi esistenziale globale, aggrava le disuguaglianze e le ingiustizie razziali ed è fondamentale che si capisca come e perché. Si tratta di spiegare chiaramente il legame tra la crisi climatica e il razzismo sistemico, di capire che il cambiamento climatico non solo coesiste con il razzismo, ma si interseca con la disuguaglianza razziale, di comprendere che il razzismo sistemico non solo renderà disuguale l'impatto del cambiamento climatico ma che è alla base della sua stessa manifestazione e fenomenologia (Williams, 2021). La crisi climatica ha genesi, riflette e rafforza le ingiustizie razziali, rivelando la profonda radice materiale dei processi di razzizzazione. Riteniamo tale prospettiva necessaria per andare oltre quelle teorie del razzismo – alcune delineate nei precedenti paragrafi di questo libro – che si concentrano esclusivamente sulle dinamiche e dimensioni sociali e psicologiche del razzismo escludendo le sue basi materiali o, meglio, la distribuzione razziale dei processi metabolici socio-ecologici. Come detto prima, il razzismo è sempre stato "più che un pregiudizio", ma gli scienziati sociali hanno per lo più inquadrato le questioni razziali come organizzate dalla logica del pregiudizio (Basso, 2010). Proponendo un approccio alle questioni razziali non convenzionale sottolineiamo la necessità di fondare materialmente la nostra analisi, ossia di comprendere che il razzismo è sistemico e radicato nelle differenze di potere tra le razze, e da questo orizzonte penetra nelle coscienze singolari, nelle soggettività razziste.

È evidente che la struttura sociale razzializzata opera, se non plasmando,

ISBN Open Access: 9788835191971

influenzando apertamente le credenze e le pratiche comuni che riproducono le posizioni sociali e la distribuzione della ricchezza. Come dimostrato dalle ricerche etnografiche di Philomena Essed, le categorie razziali e razziste vengono integrate nell'agire quotidiano proprio a partire da preesistenti asimmetrie nelle relazioni etnorazziali prevalenti in un determinato sistema sociale (Essed, 1990). Questo significa che tra struttura e agency vi è un rinforzo comune che alimenta costantemente la grammatica razzista e l'allargamento dello spazio del razzismo. Tale prospettiva ci permette di sostenere che l'odio razziale è radicato nel funzionamento del sistema sociale. L'odio sistemico e strutturale sono forme di odio profondamente radicate nelle istituzioni, nelle leggi, nelle politiche scritte o non scritte, nelle pratiche consolidate e nelle credenze e atteggiamenti collettivi che producono, tollerano e perpetuano un trattamento ingiusto nei confronti delle persone vittime di razzismo in virtù della loro diversità etno-razziale, o altre persone emarginate, come i poveri, i mendicanti o i tossicodipendenti. Poiché le reali fonti sistemiche di miseria, precarietà, alienazione e paura sono oscurate, coloro che provano questi timori sono fin troppo facilmente trasformati da imprenditori politici senza scrupoli in odi convenienti, spesso odi razziali cuciti nel tessuto della società dalla storia degli imperi. Ma la nostra domanda qui è: può esistere "l'odio senza odiatori"?

4. Contratto razziale

Ciò che serve è un quadro teorico globale per inquadrare le discussioni sulla razza e sul razzismo bianco, e quindi mettere in discussione i presupposti della filosofia politica bianca. In altre parole, ciò che serve è il riconoscimento che il razzismo (o la supremazia bianca globale) è di per sé un sistema politico, una particolare struttura di potere di regole formali o informali, privilegi socioeconomici e norme per la distribuzione differenziale della ricchezza materiale e delle opportunità, dei benefici e degli oneri, dei diritti e dei doveri.

Quando riflettiamo sulle disuguaglianze naturali, che sono state usate per millenni per giustificare i processi di differenziazioni e poi di dominio di alcune categorie e individui da parte di altri collettivi e individui, abbiamo in mente le differenze naturali che segnano le singolarità, ma allo stesso modo un processo che attribuisce tali disuguaglianze alla natura in quanto tale così da assolvere la società che di queste disuguaglianze ha fatto la propria matrice distributiva di ricchezze, potere e riconoscimento. Infine, quando parliamo di naturalizzazione e razzializzazione delle differenze ci riferiamo a uno stato di natura che le teorie del contratto sociale hanno ritenuto superato

ISBN Open Access: 9788835191971

nel processo di fondazione della società che segna l'inizio della modernità occidentale.

Nella sua definizione classica, tale mito fondativo della società e sovranità occidentale implica nuovi obblighi politici, sia nella forma di un contratto sociale là dove individui presociali escono dallo stato di natura e si costituiscono come membri di un corpo collettivo, sia nella forma di un contratto politico che fonda lo stato trasferendo così in modo più o meno definitivo i diritti e i poteri che gli individui detengono nello stato di natura a un'entità governativa sovrana. In breve, il "contratto sociale" indica il funzionamento in virtù del quale gli esseri umani, partendo da uno "stato di natura", decidono di istituire una società civile e un governo. Ciò che abbiamo, quindi, è una teoria che fonda il governo sul consenso popolare di individui considerati uguali. La metafora del contratto sociale della teoria politica occidentale, ripresa da Rawls a partire dagli anni Settanta, non è affatto uno strumento neutrale per rappresentare queste realtà, ma è tendenziosa e profondamente distorta dal punto di vista teorico.

Tale contratto sociale racchiude contemporaneamente la sua negatività che Charles W. Mills (2022) ha chiamato "contratto razziale", un contratto globale accuratamente occultato che per più di mezzo millennio ha normalizzato la supremazia bianca predicando l'uguaglianza mentre praticava la deumanizzazione e depersonalizzazione delle persone di colore. Secondo la logica del contratto, i colonizzati e le persone di colore sono "barbari" perché rimangono "naturali", mentre i bianchi sono "civili" e quindi colti e superiori, e solo loro meritano i diritti e i privilegi dell'uguaglianza costruita fuori dalla "natura".

Lo scopo generale del Contratto è sempre quello di privilegiare in modo differenziale i bianchi come gruppo rispetto ai non bianchi come gruppo, di sfruttare i loro corpi, le loro terre e le loro risorse, negando loro pari opportunità socioeconomiche. Il Contratto Razziale è quell'insieme di accordi formali o informali o meta-accordi tra i membri di un sottogruppo di esseri umani, di seguito designati con criteri "super-razziali" o "non razziali" (fenotipici/genealogici/culturali) come "bianchi" e coestensivi con la classe delle persone a pieno titolo, per classificare il restante sottogruppo di esseri umani come "non bianchi" e di uno status morale diverso e inferiore.

Come notato da George Caffentzis (2005), una sintomatica applicazione della normatività del contratto sociale che si manifesta nel suo negativo del contratto razziale è quella del colonialismo inglese. Il problema concettuale del colonialismo britannico dopo il 1689 – la rivoluzione gloriosa – era che, mentre si presentava come guidato dalla regola della proprietà privata, allo stesso tempo espropriava vaste aree di terra a persone che erano chiaramente in possesso del litorale atlantico del Nord America. In circostanze capitali-

ISBN Open Access: 9788835191971

stiche normali, gli agricoltori capitalisti del Massachusetts e della Virginia avrebbero dovuto pagare parte dei loro profitti come rendita ai proprietari indigeni della terra che coltivavano. Dopotutto, un agricoltore inglese che si recava nelle Highlands scozzesi per coltivare e vendere lino avrebbe dovuto acquistare la terra o pagare una rendita al proprietario terriero locale. Perché gli agricoltori inglesi non avrebbero dovuto pagare una rendita ai capi indigeni del Nord America sulle cui terre coltivavano il tabacco che scambiavano a Glasgow per trarne profitto? Il testo scritto per far quadrare questo particolare cerchio colonialista è *Due trattati sul governo* di John Locke. In esso troviamo una classica affermazione del razzismo della rendita. Locke sosteneva che la terra delle Americhe non è in primo luogo proprietà degli indigeni, poiché questi non sono "industriosi e razionali" e non utilizzano la terra per produrre al massimo livello. Locke misurava la mancanza di industria e ragione in modo molto "quantitativo", ritenendo che non fossero mai usciti dallo stato di natura e quindi non includibili nelle regole del contratto sociale. Il giusnaturalismo lockiano non può che farci condividere l'affermazione di Mills secondo la quale "Lungi dall'essere perso nelle nebbie dei secoli, (il contratto razziale) è chiaramente collocabile storicamente nella serie di eventi che hanno segnato la creazione del mondo moderno da parte del colonialismo europeo e dei viaggi di 'scoperta', ora sempre più e più appropriatamente definiti spedizioni di conquista" (Mills, 2022, p. 20). Questo contratto razziale crea il concetto di razza e le identità ad esso associate. Il potere dello Stato viene spesso utilizzato per far rispettare i termini dell'accordo e per sconfiggere le sfide poste dai subordinati razziali.

5. Guerre civili razziali

Recentemente, la diffusione e radicalizzazione dello scontro socio-razziale negli Stati Uniti ha spinto alcuni a rievocare il concetto di guerra civile – *civil war*. La proposta più interessante è quella di David Theo Goldberg (2020), secondo il quale gli stati scivolano nella guerra civile quando concezioni contrastanti della vita si confrontano e si attestano su posizioni irreconciliabili, quando la vita, per una parte considerevole degli abitanti dello stato, è resa insopportabile, e le richieste per cambiarla vengono combattute da altri abitanti dello Stato. Le guerre civili non sono solo la conseguenza imprevista del fallimento della lotta per la democrazia, sono lotte su modi competitivi di essere nel mondo, sulle loro concezioni sottostanti, sul controllo degli apparati politici ed economici dediti alla riproduzione, sul rapporto con la parte materiale della vita sociale, per il controllo della natura stessa.

La guerra civile – *stasis* in greco – offre un'analisi della contemporanea

ISBN Open Access: 9788835191971

condizione sociale dettata dal capitale globale, là dove la posta in gioco è la gestione delle relazioni tra la società e i suoi fondamenti bio-fisici, si tratti delle relazioni tra presunte "razze" e sessi, o dei fondamenti ecologici riproduttivi del sociale, il suo *oikeios*, ossia l'insieme delle relazioni tra il "sociale" e il "naturale". Hannah Arendt e poi Giorgio Agamben (2019) hanno messo in luce come la base di svolgimento delle guerre civili abbiano come oggetto del contendere la relazione tra vita biologica, riproduzione dell'*oikos* e politica. Anche l'approccio di Goldberg (2020) ha a che fare con tale difficoltà nel superare in termini accettabili le differenze biologiche, fenotipiche, culturali e politiche che segnano la popolazione di uno stato o di un territorio e che forniscono al contempo i termini per la differenziazione "razzializzata" dell'umano. Abbiamo erroneamente pensato che la "lunga pace" fra gli stati a livello globale potesse continuare indefinitamente. Ma se avessimo dato uno sguardo alle dinamiche della crisi socio-ecologica e alle mutevoli condizioni di numerose società del pianeta avremmo potuto notare come i conflitti interni al corpo sociale avrebbero proiettato l'ombra scura della guerra civile sulla lunga pace interstatale estendendosi al sistema globale degli stati in un mix caotico di guerre tra componenti civili e apparati statali, come è accaduto nel caso della Libia, della Siria, dell'Ucraina e ora dell'Iran.

La natura, la terra, Gaia viene sottoposta al duplice movimento di inclusione/esclusione, occultamento/sfruttamento, e politicizzazione/depoliticizzazione. Le guerre civili possono così essere considerate e analizzate come una delle conseguenze – ma forse anche una delle cause – del deterioramento delle relazioni tra il complesso politico/economico e il complesso ecologico/sociale, della crisi profonda del nesso capitalismo/natura. La Natura diventa quindi la posta in palio della guerra civile, diventa un obiettivo politico, entra nella politica.

In questo duplice movimento si annidano i potenziali delle guerre civili che si sono combattute, che si combattono e che si combatteranno tra chi include e chi esclude, tra chi sfrutta e chi cura, tra i presunti legittimi portatori di diritti statali e i globalmente segregati, si tratti di razze, sessi, classi, ceti, etnie, specie differenti. La Natura – o la Terra – nella sua genericità diventa così il fondamento di una guerra civile globalizzata intra- e interstatale che si combatte per decidere quale sia il modo in cui ci relazioniamo con la Terra, con Gaia. Non è difficile pensare che i perdenti della guerra civile, quelli al di fuori della linea territoriale, gli espulsi, i non appartenenti e immeritevoli, verranno scacciati come estranei, "incatramati con il pennello della differenza razziale" (Goldberg, 2020).

Lo scontro per l'accesso e il controllo di tali risorse e la protezione e conservazione degli habitat umani e non-umani – messi a repentaglio dai progetti di appropriazione di tali risorse – non è più ora confinata all'interno delle

ISBN Open Access: 9788835191971

società e degli stati dove prende la forma della guerra civile combattuta tra gruppi che si identificano per i loro differenti caratteri fenotipici, sociali, etnici, culturali, linguistici – spesso naturalizzati e resi permanenti dallo stesso conflitto, come nel caso della lotta tra indios amazzonici e imprese del legno e minerarie, con i primi a difendere la foresta amazzonica contro i lavoratori reclutati dalle imprese. Le guerre civili non rimangono «civili» a lungo. Nel 2015, venti dei cinquanta conflitti interni in atto, dall'Afghanistan allo Yemen alla Siria, sono state guerre civili internazionalizzate, che hanno coinvolto forze di paesi confinanti o comportato l'intervento di potenze esterne, di solito gli Usa. La guerra civile non rispetta i confini. La civiltà della merce, che continua a tenere in ostaggio l'intera umanità, si fonda sull'acquisizione di qualsiasi componente del mondo non umano per convertirlo in merce, mentre rimuove qualunque elemento che ostacoli tale appropriazione. Per estensione logica e pragmatica, la civiltà si fonda anche sull'annullamento di qualsiasi agente umano (culture o individui) che ostacola, per qualsiasi motivo, l'accesso alle risorse, com'è accaduto negli ultimi anni a più di un migliaio di attivisti e difensori dell'ambiente. Per la precisione, negli ultimi quindici anni sono stati uccisi 1.558 attivisti, in gran parte soggetti razzializzati come gli indigeni, 198 nel solo 2023[1]. Ma allo stesso tempo, all'orizzonte della *stasis*, della guerra tra natura e società e nella società tra *oikos* e polis, si profila la riconciliazione, ossia una civilizzazione ecologica che potrebbe fornire un orizzonte di ireniche relazioni tra il sociale e il naturale. Qui l'*oikeios* diventa la matrice che permette la riconciliazione di ciò che ha diviso, il fondamento sul quale si ricostituisce un orizzonte di cooperazione e mutualità tra umani, il fondamento della nuova comunità umana alternativa alla comunità-capitale.

6. Conclusioni

Da questa breve rassegna critica è del tutto evidente che il razzismo nelle sue variabili forme attraverso le quali prende sostanza non è un retaggio del passato ma è al centro delle contemporanee dinamiche globali.

Si conferma quanto questo fenomeno raffiguri, riprendendo la nota proposta di Marcel Mauss, un "fatto sociale totale" nel senso che intorno ad esso si coagulano l'insieme delle dinamiche sociali, politiche, economiche, psicologiche e simboliche che attraversano una determinata società e, al contempo, aggiungiamo noi, la interrogano sul suo passato, presente e futuro. Si è discusso di come il pensiero razzista, poiché introduce gerarchie di umanità

[1] Questi dati sono ripresi da Global Witness, che annualmente pubblica il rapporto su questa silente guerra con le popolazioni indigene (www.globalwitness.org).

ISBN Open Access: 9788835191971

e confini socio-spaziali della cittadinanza, sia sempre coevo al pensiero di Stato in ragione dei privilegi della maggioranza. Nell'epoca in cui è evidente la crisi degli assetti democratici e la tendenza verso politiche autoritarie in cui non vi è più posto per l'inclusione delle differenze, se non in una rinnovata subordinazione, il contratto razziale, così come lo si è discusso, risulta ancora potente dal punto di vista euristico e della realtà.

Siamo convinti che questa "volontà di potenza" sia rintracciabile, lo ribadiamo, dentro alla crisi ecologica e climatica il cui impatto principalmente ricade sulle maggioranze del Sud globale che subiscono le logiche estrattiviste del Nord, alimentando in parte quei flussi di migranti e rifugiati visti nella loro essenza di minaccia e di pericolo.

Ne consegue che tale ideologia non solo nutre il campo di battaglia della politica e delle politiche, ma entra nel quotidiano e nella strutturazione delle relazioni sociali. I discorsi d'odio nei social media di cui siamo tutti testimoni sono un riflesso di ciò che si è venuto a creare nelle basi societarie delle società occidentali (Pasta, 2018), in particolare tra i ceti popolari, pervase da rancore e ostilità nei confronti degli ultimi arrivati e di chi potenzialmente arriverà.

In tal senso, il riferimento al tema della guerra civile sorto nel contesto delle lotte afroamericane promosse dal Black Lives Matters è un elemento che s'insinua, non solo metaforicamente, nei diversi ambiti della vita collettiva producendo e riproducendo polarizzazione e marginalità. Da qui, come anticipato, si dovrà partire per ridefinire e innovare l'antirazzismo che comprenda tutte le istanze a livello mondiale di giustizia sociale e giustizia ambientale per ristabilire un nuovo contratto sociale, il quale non sarà in grado di annullare la forza del pregiudizio razziale, ma che sarà sempre più necessario promuovere e affermare.

Riferimenti bibliografici

Adorno, T. W. *et al.* (1973). *La personalità autoritaria*, Edizione di Comunità, Milano.

Agamben, G. (2019). *Stasis. La guerra civile come paradigma politico*, Bollati Boringhieri, Torino.

Alietti, A., a cura di (2017). *Razzismi, discriminazioni e disuguaglianze. Analisi e ricerche sull'Italia contemporanea*, Mimesis, Milano.

Alietti, A. e Padovan, D., a cura di (2025). *Antirazzismo: teoria e pratica per una società giusta. Riflessioni inattuali per una politica necessaria*, Mimesis, Milano.

Alietti, A. e Padovan, D. (2024). *Le grammatiche del razzismo. Un'introduzione teorica e un percorso di ricerca*, Edizioni Cà Foscari, Venezia.

ISBN Open Access: 9788835191971

Alietti, A. e Padovan, D. (2010). *Il razzismo come legame sociale nella società dell'eccezione giuridica. Alcune note su antisemitismo e anti-islamismo in Italia dopo l'11 settembre*, Fondazione San Paolo, Torino.
Alietti, A. e Padovan, D. (2000). *Sociologia del razzismo*, Carocci, Roma.
Alietti, A., Padovan, D. e Vercelli, C., a cura di (2014). *Antisemitismo, Islamofobia e Razzismo. Rappresentazioni, immaginari e pratiche nella società italiana*, FrancoAngeli, Milano.
Arendt, H. (2009). *Sulla rivoluzione*, Einaudi, Torino.
Barker, M. (1981). *The New Racism*, Junction Book, London.
Basso, P., a cura di (2010), *Razzismo di Stato*, FrancoAngeli, Milano.
Bonilla-Silva, E. (1997). Rethinking Racism. Toward a Structural Interpretation. *American Sociological Review*, 62: 465-480.
Bonilla-Silvia, E. (2006). *Racism Without Racists: Colour-Blid Racism and the Persistence of Racial Inequality*, Rowman&Littlefield Publisher, Lanham.
Bracke, S. and Aguilar, H. L. (2023). *The Politics of Replacement. Demographic Fears, Conspirancy Theory, and Race Wars*, Routledge, New York.
Caffentzis, G. (2005). *The Many Forms of Racism and the Politics of Oil*. Notes for a Talk at the 2005 Left Forum CUNY Graduate Center April 16.
Castel, R. (2004). *L'insicurezza sociale*, Einaudi, Torino.
Crenshaw, K. (1989). *Demarginalizing the Intersection of Race and Sex: a Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Anti-racist Politics*. University of Chicago Legal Forum.
Carmichael, S. and Hamilton, C. V. (1967). *Black Power*, Random House, New York.
Feagin, J. R. (2006). *Systemic Racism: A theory of Oppression*, Routledge, New York.
Doane, A. (2006). What is Racism? Racial Discourse and Racial Politics. *Critical Sociology*, 32(2-3): 255-274.
Du Bois, W. E. B. (2010). *Sulla linea del colore. Razza e democrazia negli Stati Uniti*, il Mulino, Bologna.
Elias, S. and Feagin, J. R. (2016). *Racial Theories in Social Science. A Systematic Racism Critique*, Routledge, New York.
Essed, P. (1990). *Everyday Racism*, Hunter House, Claremont.
Feagin, J. R. and O'Brien, E. (2010). Studying Race and Ethnicity: Dominant and Marginalised Discourses in the Critical North America Case. In: Collins P. H. and Solomos J., eds., *The Sage Book of Race and Ethnic Studies*, Sage, London.
Fraser, N. (2007). Giustizia sociale nell'era della politica dell'identità: redistribuzione, riconoscimento e partecipazione. In: Frase N., Honneth A., *Redistribuzione o riconoscimento. Lotte di genere e disuguaglianze economiche*, Meltemi, Roma.
Goldberg, D. T. (2020). "On Civil War". Critical Times September 9th. https://ctjournal.org/category/anti-racism/
Kivisto, P. (2010). Multiculturalism and Racial Democracy: State Policies and Social Practices. In Collins P. H. and Solomos J., eds., *The Sage Book of Race and Ethnic Studies*, Sage, London.

ISBN Open Access: 9788835191971

Hall, S. (2006). *Il soggetto e la differenza. Per un'archeologia degli studi culturali e post-coloniali*, Meltemi, Roma.
Hill Collins, P. (1986). Learning from the Outsider Within: The Sociological Significance of Black Feminist, *Social Problems*, 33, 6: 14-32.
McPherson Report (1999). *The Sthepen Lawrence Inquiry*. https://assets.publishing.service.gov.uk/media/5a7c2af540f0b645ba3c7202/4262.pdf
Mills, C. W. (2022). *The Racial Contract*, Cornell University Press, New York: Ithaca.
Omi, M. and Winant, H. (1986). *The Racial Formation in the United States*, Routledge, New York.
Padovan, D. and Alietti, A. (2019). Geo-capitalism and global racialization in the frame of Anthropocene. *International Review of Sociology*, 29, 2: 172-196.
Pasta, S. (2018). *Razzismi 2.0. Analisi socio-educativa dell'odio online*, Scholé-Morcelliana, Milano.
Pettigrew, T. F. and Meertens, R. W. (1993). Subtle and Blatant Prejudice in Western Europe, *European Journal of Social Psychology*, XXV, 1: 57-75.
Picca, L. H. and Feagin, J. R. (2007). *Two-faced Racism: Whites in the Backstage and Frontstage*, Routledge, New York.
Taguieff, P.-A. (1994). *La forza del pregiudizio*, il Mulino, Bologna.
Wieworka, M. (1993). *Lo spazio del razzismo*, Il Saggiatore, Milano.
Williams, J. (2021). *Climate Change is Racist. Race, Privilege and the Struggle for Climate Justice*, Icon Books, London.
Winant, H. (2004). *The New Politics of Race: Globalism, Difference, Justice*, University of Minnesota, Minneapolis.

ISBN Open Access: 9788835191971

Jim Crow, il taccuino di McCarthy e la vittoria postuma di Hitler. Blocchi egemonici, sociologia delle emozioni e politiche della paura

di *Alfredo Agustoni*[*]

1. Egemonie, stereotipi e politiche della paura. Considerazioni introduttive

Questo saggio compare in un volume dedicato al problema del pregiudizio e dei "discorsi di odio" in rete, ma lo fa ponendosi a monte e recuperando l'analisi sociologica di emozioni come la paura. La paura e il disprezzo sono forse, tra tutte, le emozioni più direttamente coinvolte nella costruzione di un immaginario razzista. È perfettamente chiaro che le emozioni non sono una realtà meramente psicologica, ma una realtà che ha un carattere intermediario tra la nostra vita psichica e il mondo che ci circonda, e che, nell'universo dei fenomeni culturali, si trasforma e si rielabora in una straordinaria varietà di sfumature del sentimento. Con le parole di Maurice Merleau-Ponty possiamo dire che la nostra vita psichica non prescinde mai dalla nostra natura di esseri biologici e che, nel contempo, la trascende di continuo. Non riusciamo mai, malgrado ogni sforzo, ad essere animali e neppure angeli. Ed è forse questo che ci definisce come uomini. Le nostre rappresentazioni sono profondamente radicate nelle emozioni che ci definiscono come mammiferi appartenenti ad un particolare genere, homo, e ad una particolare specie, cioè homo sapiens. Il sapiens ha un problema. Ha una natura storica che porta, nel quadro di contesti contingenti, all'interazione tra emozioni, rappresentazioni, tecnologie e relazioni di potere. La costruzione di mostri o di reietti, nell'immaginario umano, non è la semplice e diretta trasposizione delle emozioni nel pensiero, ma è il prodotto di un complesso insieme di fattori. Nel presente saggio prendiamo in considerazione alcuni casi storici, come la costruzione dell'etnicità negli Stati Uniti del XIX secolo, di fronte a fenomeni come l'immigrazione dall'Europa e la difficile gestione dell'abolizione della schiavitù,

[*] Dipartimento di Scienze Filosofiche, Pedagogiche e Sociali, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, alfredo.agustoni@unich.it

ISBN Open Access: 9788835191971

nonché la costruzione del "nemico escatologico" comunista, nell'ambiguità dei meccanismi democratici nei decenni della guerra fredda, e la costruzione della paura in ambito urbano, soprattutto a partire dagli anni Sessanta.

Potremmo partire dal lavoro di un famoso giornalista americano, Walter Lippmann, che stese circa cento anni fa uno dei capisaldi delle riflessioni sui rapporti tra media, opinione pubblica e democrazia. Lippmann, da liberale scettico nei confronti della democrazia, esprime la propria convinzione che il protagonismo delle masse sulla scena politica incontri il proprio limite strutturale nell'enorme complessità dei problemi. L'uomo della strada necessita di un'indispensabile riduzione di complessità, che non può esimerlo dal ricorso a "stereotipi". È interessante il fatto che, nel lavoro di Lippmann del 1922, il termine "stereotipo" è per la prima volta traghettato dal mondo delle tipografie a quello dei media studies. Uno stereotipo è una rappresentazione ipersemplificata, che si basa sull'associazione rigida e sistematica tra variabili, una delle quali può essere l'appartenenza ad un determinato gruppo etnico, religioso, politico ecc. Uno stereotipo può essere quello dell'irlandese ubriacone, dell'italiano pigro, focoso e familista, del musulmano terrorista, del nero aggressivo e indolente.

C'è da dire che, per Lippmann, la formazione e diffusione di stereotipi si fondano su di una tacita complicità tra media, politica ed opinione pubblica. L'uomo della strada, in qualche modo, deve semplificare una realtà la cui complessità trascende sistematicamente la sua capacità di comprensione. I media ricorrono in modo sistematico alla semplificazione della realtà. Questo dipende dal fatto che, da un lato, anche i giornalisti non sono in grado di gestire, su di un piano cognitivo, l'estrema complessità del mondo che devono descrivere e, dall'altro, dal fatto che sono costretti a ricorrere agli stereotipi dei loro interlocutori per comunicare con loro.

Ma non dobbiamo credere che lo stereotipo, nell'ottica del giornalista americano, sia il prodotto neutrale di una mera e sistematica distorsione e semplificazione cognitiva. Da un lato, Lippmann è estremamente attento al coevo sviluppo della psicoanalisi, che lo porta a confrontarsi con la profonda connotazione emotiva di simboli e rappresentazioni: gli stereotipi si colorano delle mie inquietudini più profonde, del disgusto che posso provare di fronte a determinati gruppi etnici e sociali, della paura che mi incutono prospettive di trasformazione sociale potenzialmente eversive di una normalità dai tratti rassicuranti (la dimensione heideggeriana dell'*heimlich*).

Lo stereotipo consente, in qualche modo, di riportare una realtà che sfugge alla normalità, alla dimensione del consueto. Di fronte ad una dimostrazione politica che è degenerata in una serie di scontri, le cronache dei giornali sono sempre pronte a chiamare in causa i "soliti" violenti, i "soliti" facinorosi che si sono infiltrati in una dimostrazione potenzialmente pacifica.

ISBN Open Access: 9788835191971

L'autore dell'articolo, in fondo, può fornire un accettabile resoconto dell'accaduto senza particolari approfondimenti, mentre l'anziano signore che legge il giornale seduto sulla panchina (scusate lo stereotipo!) scuote la testa in segno di disapprovazione, ma in fondo i conti gli tornano.

Così, sempre recuperando le parole di Lippmann, noi ci troviamo a vivere in uno "pseudo-ambiente", dove le pacifiche manifestazioni degenerano a causa dei "soliti" facinorosi e le donne sono vittime del "tipico" maschio che non sopportava la loro relativa autonomia. Un primo appunto che potremmo fare a questa visione lippmaniana è che si fonda sulla contrapposizione tra un presunto, quanto irreperibile, "ambiente reale" e uno "pseudo-ambiente", quello che noi esperiamo nella nostra esistenza quotidiana che si nutre necessariamente di stereotipi, simboli e costruzioni sociali.

Il nostro "pseudo-ambiente" è quello che uno dei padri dell'etologia, Jacob von Uexküll (1933), definiva *Umwelt* (il "mondo circostante", il mondo con cui entriamo in contatto grazie alle potenzialità del nostro corpo e dei nostri organi di senso). Salvo che, prosegue Uexküll, l'animale è chiuso nel suo *Umwelt*, mentre l'*Umwelt* umano è apertura al mondo, e quindi possiede una propria variabilità storica, capace di trascendere qualsiasi tipo di determinazione biologica[1]. Ma, se noi parliamo di contingenza storica, non possiamo prescindere da uno degli aspetti maggiormente significativi delle dinamiche storiche, che sono i rapporti di potere. Seguendo l'idea estremamente generale, espressa nel Manifesto dei comunisti, secondo cui "In ogni epoca, le idee delle classi dominanti, sono le idee dominanti", è evidente che alcuni gruppi di persone hanno maggiori capacità di altre di diffondere idee, rappresentazioni, stereotipi di grande potere. È quello che Weber prende in esame come "influenza", Gramsci come "egemonia" e, più di recente, Joseph Nye chiama "*soft power*". Chi è in grado di dotare la propria visione del mondo di un carattere egemonico, sarà in grado di cementare intorno a sé un "blocco storico", cioè una variegata e composita coalizione di forze sociali, che nelle rappresentazioni egemoniche trovano la propria sintesi politica.

Recuperando l'illuminante analisi di Eva Illouz (2024), una studiosa francese che si occupa di sociologia delle emozioni, possiamo dire che il rapporto tra paura e politica ha subito radicali trasformazioni con l'avvento della modernità e con la democratizzazione delle nostre società. Nelle società tradizionali, vale l'analisi formulata da Machiavelli e poi da Hobbes. Machiavelli ritiene che il principe debba sapersi far temere ed amare ad un tempo, ma aggiunge anche che farsi temere è più importante che farsi amare. Machiavelli, in fondo, è un uomo politico, acuto osservatore delle vicende umane,

[1] Aspetto, quest'ultimo, estremamente influente su numerosi filosofi e antropologi, da Max Scheler ad Arnold Gehlen.

ISBN Open Access: 9788835191971

che si propone di fornire saggi consigli a quegli stessi despoti che, da fervente repubblicano, aveva avversato. Hobbes si spinge oltre. Si propone infatti come sistematico e astratto studioso della natura delle cose, che ritiene che il potere abbia un duplice rapporto con la paura. La paura legittima il potere, perché gli uomini si assoggettano al potere stesso per essere tutelati dai loro simili. Nello stesso tempo, la paura è lo strumento per mezzo del quale il potere è in grado di assoggettare gli uomini. La rappresentazione che Hobbes fornisce della società umana si caratterizza per una fondamentale assenza di fiducia, che in qualche modo anticipa la teoria dei giochi. Io non posso fidarmi degli altri miei simili e, per questo motivo, mi assoggetto ad un potere del quale pure so di non potermi fidare, perché so che esercita su di me un arbitrario potere di vita e di morte. Ma sono parimenti consapevole del fatto che tutte le altre persone di cui potrei avere paura si sono comunque assoggettate all'arbitrio di quello stesso potere, che ha comunque promesso di proteggermi.

Ma, con l'avvento del liberalismo e della democrazia, il rapporto tra paura e potere ha cambiato natura. La paura costituisce pur sempre un fondamentale strumento di governo, ma secondo modalità differenti, spesso indirette. È quello che potremmo chiamare "governo della paura" (Simon, 2006), "politica della paura" (Agustoni, 2024) o "politica della vulnerabilità" (Illouz, 2024), nel quadro di quella che potremmo a sua volta definire una "cultura della paura" (Glassner, 2018).

Uno dei problemi del cittadino elettore, particolarmente in un contesto post-ideologico, consiste nel comprendere chi sia in grado di spiegargli quali siano le reali minacce, di che cosa occorra avere davvero paura. Devo avere più paura del mutamento climatico o della delinquenza, soprattutto se legata al fattore etnico e migratorio? Ma i cambiamenti climatici possono costituire un significativo driver migratorio. In questo caso, di cosa ci dobbiamo veramente preoccupare? Di popolazioni imprigionate in contesti nei quali la vita diventa sempre più impossibile, o di un presunto ed improbabile esodo, che porterebbe a una presunta e improbabile islamizzazione della vecchia Europa?

La paura fa parte del nostro corredo genetico di mammiferi, come tutte le altre emozioni fondamentali. Ma, elaborandole con gli strumenti simbolici e linguistici, introspettivi e artistici che fanno parte della nostra cultura, noi ci rendiamo capaci di trasformarle in sentimenti, e di trasformare le immediate reazioni emotive del mammifero in sofisticate pratiche introspettive. Il mammifero reagisce alla paura aggredendo o scappando, mentre noi possiamo elaborarla fino a negarla. Eppure, talora la paura ci assale, emergendo dalla nostra natura di mammiferi, fino a paralizzarci. La paura, come le altre emozioni fondamentali, collegano la nostra natura biologica con la nostra capacità di elaborazione culturale. Le nostre emozioni sono il retaggio dell'evolu-

ISBN Open Access: 9788835191971

zione biologica che ci ha portati a essere mammiferi. I nostri sentimenti sono il prodotto dell'elaborazione culturale delle elementari emozioni del mammifero. L'evoluzione culturale conferisce una grande quantità di sfumature alla paura che la preda potenziale prova di fronte al potenziale predatore.

Nel quadro dell'evoluzione biologica, la paura è abbastanza chiaramente regolata. Qualche tempo fa, mentre tagliavo il prato in una località rurale, sento dei rumori provenire dal cielo. Alzo lo sguardo e vedo un corvo che, rumorosamente, si butta contro un falchetto delle sue stesse dimensioni che, volando, disegna anelli nel cielo. Il corvo lo prende di mira, lo punta con il becco e gracchia rumorosamente. Il falco, apparentemente impassibile, continua a girare nello stesso modo, ma sempre più in là, a crescente distanza. Il corvo ha probabilmente un nido da difendere e, per questo, la paura si traduce in un comportamento aggressivo. Altrimenti, con ogni probabilità, si sarebbe semplicemente allontanato da un altro animale potenzialmente pericoloso come il falco.

Quando dall'evoluzione biologica emerge l'altro grande processo evolutivo, cioè lo sviluppo culturale (Elias, 1989), le cose si complicano. Noi possiamo abbandonare consapevolmente la nostra prole, per sottostare a determinate convenzioni sociali (è il caso di Evita Perón, disconosciuta dal padre che l'aveva concepita in un rapporto occasionale con la cuoca). Possiamo mantenere la nostra prole per numerosi anni dopo il raggiungimento della pubertà, per consentire il completamento di un cursus di studi. Noi possiamo imparare, se siamo membri di una società guerriera, di un'aristocrazia militare o di un gruppo religioso votato al martirio, che non dobbiamo conoscere la paura, e il nostro sistema culturale ci fornisce numerosi sistemi per domesticarla.

Le emozioni, naturalmente, non si limitano alla paura, ed Eva Illouz (2024) ci fornisce un quadro molto completo dei loro potenziali rapporti con la vita sociale e la politica. Le emozioni sono radicate, dicevamo, nella nostra natura di *sapiens*, che però trascendono, "condensando in sé strutture sociali, codici etici e identità di gruppo" e proponendosi come "momenti nei quali gli individui elaborano processi sociali di base come il dominio, la competizione, la sottomissione, la diseguaglianza, l'affetto, la giustizia" (Illouz, 2024, p. 8). Come la paura, la speranza si radica nel carattere incerto e indeterminato del futuro. L'evento negativo o positivo che si paventa o si auspica potrebbe realizzarsi o meno. Paura e speranza possono rivelarsi complementari, nella misura in cui si spera di poter evitare l'evento temuto (si pensi all'ansia ecologica e alla speranza di una società più sostenibile). Questa complementarità non è necessaria, perché la speranza può anche riguardare l'uscita da un'attuale condizione di desolazione, di sofferenza. La speranza è peraltro più elaborata della paura, perché implica una sviluppata capacità

ISBN Open Access: 9788835191971

di prefigurare una situazione futura, mentre la paura, nelle sue forme più elementari, può anche limitarsi a minacce imminenti. L'una e l'altra, spesso, soprattutto se impugnate da un leader, possono facilmente creare delle comunità politiche (o, con le parole di Benedict Anderson, comunità immaginate).

La paura e la speranza sono radicate nel futuro, così come la nostalgia e lo spaesamento sono radicati nel passato. Pensiamo allo spaesamento di fronte a un mondo che non riconosciamo più, a seguito di mutamenti repentini del nostro quadro di vita. È il caso del contadino di Marciana descritto da Ernesto De Martino, che quando sale per la prima volta a bordo di un'automobile e perde di vista il campanile sotto la cui ombra si era consumata la sua esistenza, è assalito da un sentimento di spaesamento, di incapacità di riconoscere la propria posizione, l'ordine dello spazio (è il sentimento che De Martino definisce "angoscia territoriale"). Un misto di spaesamento e nostalgia è anche la reazione che consegue al venir meno dei solidi riferimenti sociali al cui interno si era cresciuti. Sempre la Illouz ci fornisce l'esempio dell'anziano aristocratico del multinazionale Impero austro-ungarico che, di punto in bianco, finita la Prima guerra mondiale, si trova cittadino polacco, e con il suo sguardo cerca dappertutto segni della vecchia "kakania" che gli restituiscano il senso della sua identità[2].

Pensiamo quindi alla "nostalgia filistea" dei giovani aristocratici francesi dell'Ottocento che, nel grigiore di un mondo dove il potere era transitato dai castelli alle banche, con le parole del marchese De Lafayette, optavano per l'"emigrazione interiore", verso gli spazi sconfinati dell'anima e della contemplazione estetica. Spaesamento e nostalgia corrispondono a due modi differenti di sentirsi estranei al proprio mondo. La nostalgia, in particolare, presenta due volti differenti: il rimpianto di una realtà perduta per sempre (si pensi al ricordo di un "caro estinto"), o il richiamo ad una dimensione perduto che si spera di poter restaurare (è il caso del MAGA dei conservatori americani: rendere l'America grande di nuovo).

La delusione, a sua volta, collega un passato di speranza ad un presente nel quale la speranza non ha trovato realizzazione: "la vita che vorremmo aver vissuto, ma non si è realizzata, circonda di un alone di delusione la nostra esperienza. La vita non vissuta è il fantasma che infesta la cultura consumistica e l'autoaiuto" (Illouz, 2024, p. 77). La delusione, come la vergogna e la disperazione, per la Illouz sono incarnate dalla figura di Madame Bovary, ambiziosa figlia di agricoltori e avida lettrice di romanzetti d'appendice, dai quali apprende che un amore altolocato è la strada maestra verso il

[2] Così Robert Musil, ne *L'uomo senza qualità*, definisce ironicamente l'impero austro-ungarico, dove tutto era real-imperiale (königlich-kaiserlich).

ISBN Open Access: 9788835191971

riscatto. Soffocata dall'ambiente provinciale e dal borghesissimo matrimonio con un mediocre medico di provincia, cercherà la propria rivincita tra le lenzuola di un aristocratico, poi nell'amore quantomeno romantico con un giovane impiegato. Abbandonata da tutti e piena di debiti, temendo anche lo scandalo che si scatenerebbe se la sua situazione debitoria e i suoi amori illeciti venissero alla luce, si uccide con il veleno rubato dal farmacista del paese, borghese di provincia gretto e calcolatore, in qualche modo l'antitesi simbolica della protagonista.

Pensiamo poi al disgusto, al disprezzo, e al ruolo che esercitano nel regolare i rapporti etnici e di classe. Diceva un vecchio proverbio milanese che non c'è niente di peggio che puzzare di povero. E, prosegue a questo proposito Georg Simmel (1908), tra tutti i sensi l'olfatto è quello che maggiormente tradisce le differenze etniche e sociali, proprio perché è il meno consapevole, il meno facilmente controllabile. È più facile che l'odore di un reietto ci ispiri il disgusto, piuttosto che non la sua vista. È il caso, di nuovo, dell'ira, passione chiaramente ambivalente, perché può essere l'ira di Dio, l'ira del giusto di fronte all'ingiustizia, ma anche l'ira del borioso Achille o del protervo Agamennone, del mortale che non accetta di non essere Dio. Come la paura e la speranza, l'ira e l'indignazione si prestano facilmente alla creazione di comunità politiche, ma come la paura devono trovare un oggetto verso cui dirigersi, "vale a dire i veri responsabili di tanta insoddisfazione" (Illouz, 2024, p. 133).

Già questi pochi cenni lasciano spazio a numerose riflessioni. La gestione delle emozioni, in un contesto come quello contemporaneo, caratterizzato da una sfiducia diffusa e da un sistema mediatico più complesso che in ogni altra epoca, rivela tutta la sua importanza, ma anche la sua ambiguità, e lo spaesamento che ne consegue: persi dei chiari punti di riferimento, diventa assai più difficile "dosare le emozioni". Questo diventa perfettamente chiaro nel contesto del lock-down del 2020. Dobbiamo avere più paura della pandemia, e quindi accettare le restrizioni dello stesso lock-down, oppure dobbiamo avere più paura delle conseguenze economiche di queste stesse restrizioni, che vanno pertanto rigettate? La leadership liberale e popolare dell'Unione europea è abbastanza compatta sulla prima ipotesi, ma molti governi dell'estrema destra neoconservatrice e/o anarco-capitalista, come quello statunitense di Trump o quello brasiliano di Bolsonaro propendono piuttosto per la seconda strada. Dobbiamo avere più paura del loro negazionismo o di chi, come l'ungherese Victor Orban ha strumentalizzato la pandemia per chiedere poteri speciali e porre limiti alla libertà di informazioni? Dobbiamo essere più spaventati dal virus o da fantomatici complotti finalizzati a farci rimanere in casa come prova generale per l'instaurazione di governi d'eccezione, capaci di sospendere le garanzie democratiche e, sostanzialmente, di

ISBN Open Access: 9788835191971

instaurare regimi dittatoriali? Dobbiamo avere più paura della pandemia o del complotto che sta dietro ai vaccini, che coinvolge sicuramente il governo americano, i rettiliani provenienti da altri pianeti, i servizi segreti britannici, big pharma e quant'altro, finalizzato ad asservire e ridurre numericamente il genere umano? Dobbiamo avere paura delle migrazioni, perché animate da un chiaro e deliberato progetto "sostituzionista", o dobbiamo avere più paura dei sostituzionisti stessi, per le ambigue mire politiche che probabilmente li contraddistinguono?

Il sociologo Barry Glassner (2018), nel suo lavoro sulle "culture della paura", si domanda perché gli americani abbiano paura delle cose sbagliate. Nei termini in cui l'autore la pone, la sua domanda è sensata. Ma, immediatamente, sorge l'altra domanda: di cosa occorre avere veramente paura, in un'infosfera produttrice di mostri? Io prego chiunque possa farlo di spiegarmi di cosa devo davvero avere paura, ma dubito che qualcuno possa farlo in maniera convincente: tanti, in compenso, sapranno farlo in maniera strumentale.

Potrebbe, a questo punto, venire in nostro aiuto un sociologo britannico, Stanley Cohen (1972), che circa mezzo secolo fa studia il fenomeno da lui definito "panico morale", cui si associano, generalmente, quelli che Cohen definisce dei "diavoli popolari". I diavoli popolari sono stereotipi profondamente carichi di valenze emotive, spesso collegati ad un cambiamento sociale che provoca spaesamento ed ansia. Questo spaesamento provoca ansia piuttosto che paura: di cosa abbiamo veramente paura? Non dei MIG sovietici, ma del mutamento in quanto tale. Allora l'opinione pubblica stigmatizza volentieri figure stereotipate che appaiono sulla stampa, per cui le bande giovanili come i Mods e i rockers sembrano trasformarsi in autentiche emergenze, in simboli che sembrano condensare in se stessi tutte le inquietudini che si legano all'incertezza, al carattere indeterminato del futuro, allo spaesamento che ineluttabilmente accompagna il cambiamento sociale, alle minacce di cui un futuro incerto sembra necessariamente gravido.

2. Know Nothing, Jim Crow e il KuKlux Klan

Come dicevamo, utilizzando il linguaggio della vulgata psicoanalitica, rappresentazioni, simboli e stereotipi si prestano a condensare diversi aspetti, in un clima di fondamentale inquietudine. Potremmo prendere tre esempi che riguardano le problematiche dell'immigrazione nell'America del XIX secolo. L'immagine dell'irlandese ubriacone, subumano e cattolico, quella dell'afroamericano trasformato nell'immagine caricaturale di Jim Crow e quella dell'italiano sporco e indolente, mafioso o rivoluzionario. Si tratta di imma-

ISBN Open Access: 9788835191971

gini che nascono e si sviluppano in periodi differenti, anche se poi si prestano a transitare attraverso i decenni, assorbendo nuove ansie, nuove inquietudini, nuove issue politiche. Ancora in occasione delle presidenziali del 1960, il candidato repubblicano, Richard Nixon, cercò di sfruttare le origini irlandesi e la fede cattolica del proprio rivale, il democratico John F. Kennedy. Nixon osservò, in effetti, che Kennedy era cattolico, quindi non c'erano che due possibilità. Se era un buon cattolico, era tenuto a obbedire al Papa, ma una persona vincolata dall'obbedienza a un capo di stato straniero, soprattutto se si trattava del vertice di una confessione religiosa, non avrebbe potuto essere un buon presidente di un paese libero e protestante nelle sue radici. Ma magari Kennedy non era un buon cattolico. In questo caso, però, non sarebbe stato un buon cristiano, e non avrebbe quindi potuto essere un buon presidente per un paese profondamente religioso come gli Stati Uniti.

Il panico morale per l'invasione degli irlandesi cattolici attraversa la storia americana nel quindicennio che precede la guerra di secessione, ed è splendidamente esposta da Martin Scorsese nel film *Gangs of New York*, che racconta molto bene le vicissitudini che interessano la città tra la fine degli anni Quaranta e il principio degli anni Sessanta.

Per capire bene la vicenda, è però necessaria una premessa. Negli anni Quaranta del XIX secolo, gli Stati Uniti, in particolare il nord degli Stati Uniti, non sono ancora una meta migratoria appetibile come lo saranno a cavallo tra il XIX e il XX secolo. L'apertura della colonizzazione del territorio ad occidente dei monti Appalachi, soprattutto grazie alla navigazione a vapore sui grandi laghi e all'ancora embrionale espansione della rete ferroviaria (Hugill, 1999; Agustoni, 2022), ha contribuito a richiamare una certa quantità di immigrati provenienti dall'Europa centrale e settentrionale. Ma quello che, nella seconda metà degli anni Quaranta, spinge centinaia di migliaia di irlandesi verso la costa orientale degli Stati Uniti è piuttosto un fattore "push", cioè la spaventosa carestia delle patate che sconvolge l'Irlanda. Ma, nelle città del nord degli Stati Uniti, l'arrivo improvviso di tanta gente dall'aspetto miserabile, peraltro notoriamente praticante la religione cattolica non è sempre visto di buon grado.

Per raccontare meglio questa storia, possiamo recuperare la prima scena del film, dove una banda di irlandesi si sta preparando a sfidare una banda di "nativi"[3] per il controllo del quartiere di *Five Points*. Il capo degli irlandesi, Padre Vallon[4], è un prete cattolico, che ha al proprio fianco il figlioletto

[3] Va tenuto conto del fatto che, nell'America degli anni Quaranta dell'Ottocento, il termine "nativo" non è applicato ai pellirosse ma ai soggetti nati negli Stati Uniti da famiglie che vi abitavano già da prima dell'indipendenza. Di qui il termine "nativista", che si applica ai gruppi che in qualche modo contrastano i nuovi arrivi.

[4] Interpretato da Liam Neeson.

ISBN Open Access: 9788835191971

Amsterdam Vallon[5]. Potrebbe sorprenderci il fatto che un prete cattolico abbia al proprio fianco il figlio, ma il fatto che sia il capo di una banda di delinquenti potrebbe lasciarci molto più perplessi. Ma dal film si capisce che quel tipo di organizzazione è dettato dalle esigenze della sopravvivenza in un ambiente maledettamente ostile. Il capo dei "nativi" è un torvo personaggio dai lunghi baffi e dall'occhio di vetro[6], Bill the Butcher[7]. Nel corso del feroce combattimento, che avrebbe potuto ricordare uno scontro tra vichinghi e irlandesi o sassoni di mille anni prima, tutto a colpi di mazza, di scure e di coltello, Bill the Butcher ammazza il prete. Nel corso del combattimento, i "nativi" apostrofano i propri rivali come "pezzenti irlandesi" e "luridi papisti" (come fanno, nelle sequenze successive del film, quelli che vanno ad accogliere a sassate gli irlandesi che scendono dalle navi). Il figlio vede e piange, già pensa probabilmente ad una futura vendetta.

Sono passati numerosi anni, siamo ormai agli inizi della guerra civile contro i confederati, e Amsterdam Vallon torna in una New York dove le bande etniche controllano una grande quantità di attività ed esercitano un'indiscutibile influenza sulla politica cittadina. Accanto agli irlandesi, che ormai cominciano a essere più potenti e in qualche caso riescono a far eleggere i propri candidati, ecco i cinesi, che gestiscono bische e fumerie d'oppio e "odiano i nativi più di noi", come spiega ad Amsterdam un altro irlandese. Nel frattempo, la guerra non sembra volgere al meglio per l'Unione, e nel 1863 viene proclamata la leva generale. La popolazione di New York non sembra gradire e insorge, devastando le ville dei ricchi, che si possono riscattare dalla leva, e linciando tutti i pochi neri che trovano per le strade o nelle loro case (e qui troviamo il secondo "diavolo popolare", cioè il negretto Jim Crow). New York viene messa in stato d'assedio, con i soldati che sparano sulla folla in tumulto e le cannoniere che cominciano a scaricare colpi di cannone sulla città insorta. Nel fumo di una città dove divampa uno scontro feroce, anche il "nativo" Bill the Butcher e Amsterdam Vallon si scontrano all'ultimo sangue e Bill muore, non senza aver fatto in tempo a dire di essere orgoglioso di morire come il padre, caduto per difendere l'America dagli inglesi.

Bill cade per difendere l'America, ma da chi? Come ci spiega splendidamente un grande demografo, Peter Turchin (2024), l'America dove sbarcano gli irlandesi è tutt'altro che un luogo accogliente. È un contesto in crisi, dove, a fronte di una ristretta minoranza di persone che sta accumulando enormi fortune, il benessere della stragrande maggioranza della popolazione declina, e questo si riflette, per esempio, nel drastico incremento della mortalità

[5] Amsterdam Vallon adulto sarà interpretato da Leonardo Di Caprio.

[6] Solo verso la fine del film apprenderemo che a cavargli l'occhio era stato lo stesso Padre Vallon in un precedente scontro.

[7] Daniel Day-Light, nel film.

ISBN Open Access: 9788835191971

infantile, ma anche nel moltiplicarsi di disordini e sommosse urbane. Miseria e violenza aumentano, all'interno di città che vedono comunque aumentare sensibilmente il numero dei propri abitanti. Un contesto di questo genere produce facilmente conflittualità e inquietudine, e la comparsa di una nuova figura, l'immigrato irlandese, cattolico, cencioso e propenso all'abuso di bevande alcoliche, condensa facilmente su di sé molte inquietudini.

Si struttura rapidamente, a cavallo tra gli anni Quaranta e Cinquanta, un movimento di "nativi" protestanti, intenzionati a contrastare l'arrivo di irlandesi. Il movimento assume la denominazione di Know Nothing, in ragione della sua struttura segreta e del carattere omertoso, che per certi versi precede il Ku Klux Klan degli anni che seguiranno la guerra di secessione. Quali sono gli stereotipi e le inquietudini che forniscono brodo di coltura al "nativismo" e al Know Nothing? Forse poche illustrazioni, comparse su giornali statunitensi negli anni Cinquanta dell'Ottocento, rendono splendidamente l'idea. In una, due irlandesi girano per Boston. Entrambi sono vestiti con due botti tenute su da comode bretelle. Sulla prima c'è scritto "Irish Whisky" e sulla seconda "Beer". E fino a qui ci fermiamo allo scherzoso stereotipo, che pure tradisce la percezione dell'irlandese come portatore del vizio, particolarmente inviso ai gruppi evangelici. In altre illustrazioni, gli irlandesi sono rappresentati come scimmie, o comunque dotati di fattezze subumane, e la disumanizzazione precede spesso la messa in opera di comportamenti disumani (lo insegna la storia coloniale così come quella del regime nazista).

Ma forse più interessante di tutte è un'illustrazione dove un barcone di irlandesi cenciosi approda alle coste degli Stati Uniti. Tra i pezzenti sono nascosti piccoli chierici e vescovi, e su di tutti risalta il Papa, che si alza in piedi e annuncia agli americani protestanti di essere venuto lì per prendersi cura delle loro anime. Sulla costa, un ragazzetto protestante brandisce la Bibbia come uno scudo e, fondamentalmente, risponde che alla sua anima ci penserà da solo (da buon protestante). Alle sue costole sta seduto un signore, forse suo padre, che ricorda già un po' lo "Zio Sam", con l'aria tranquilla e una colt nella cintura … così, giusto se la Bibbia non dovesse bastare[8].

Ma, se torniamo al film di Scorsese, la gente di New York inferocita per la leva generale del 1863, se la prende con i ricchi, che possono evitare la leva con trecento dollari, e con i neri, per i quali dovrebbe andare a morire. Perché morire per Jim Crow? E siccome non voglio morire per Jim Crow, io, cittadino newyorkese, comincio ad uccidere tutti i Jim Crow che mi capitano

[8] Samuel Colt ha inventato la pistola a tamburo negli anni Trenta dell'Ottocento. La colt ha cominciato ad essere utilizzata dai rangers del Midwest contro gli indiani, e poi è stata proficuamente utilizzata nella guerra contro il Messico, negli anni Quaranta dello stesso secolo. Attorno al 1850, quando viene pubblicata l'illustrazione, è già uno dei simboli dell'identità nordamericana.

ISBN Open Access: 9788835191971

sotto tiro (ancora pochi, visto che la gran parte dei neri degli Stati Uniti vivono da schiavi nelle piantagioni del "Vecchio Sud").

Chi è Jim Crow? È una caricatura del nero, che emerge nella letteratura e nel teatro degli stati del "Vecchio Sud", e siccome i neri non possono comparire sulle scene del teatro, la parte di Jim Crow è recitata da bianchi che si tingono la faccia di nero. Inizialmente, Jim Crow è l'incarnazione del paternalistico disprezzo che i proprietari del "Vecchio Sud" hanno per i propri schiavi. Ma con l'abolizione della schiavitù il fantasma di Jim Crow si fa più minaccioso. Lo vediamo vestito di una casacca nordista tutta sdrucita, che lascia intendere il disprezzo che la gente del "Vecchio Sud" ha tanto per i neri che per la causa abolizionista, ma anche l'ansia, la rabbia e lo spaesamento provocati dal (sia pur giustificato) "genocidio culturale" del "Vecchio Sud".

Queste espressioni di ansia e di rabbia sono splendidamente rappresentate dal più famoso bandito americano, Jesse James, che una cospicua parte degli attori americani del dopoguerra hanno interpretato, non ultimo Brad Pitt. Jesse James è un bandito, svaligiatore di banche e di treni, ma si presenta come un patriota del "Vecchio Sud" libero, la cui libertà si fondava sulla schiavitù dei neri. Ulysses Grant, il generale che aveva guidato l'esercito nordista, metterà in stato d'assedio il Vecchio Sud. Ma non sarà lui, bensì il "vigliacco Robert Ford", a porre fine alla vita di Jesse James. Jesse James non c'è più, ma Jim Crow non è morto, e in fondo lo stesso Jesse James continua a vivere come figura "positiva" nella memoria storica degli Stati Uniti.

Dopo il "genocidio culturale" di una società schiavista, la "repubblica del nord" (Hughil, 1989) vuole tentare la strada della riconciliazione, anche come conseguenza della comparsa del primo KuKlux Klan negli anni Sessanta dell'Ottocento. Le leggi di segregazione che gli stati del "Vecchio Sud" cominciano a promulgare, con il consenso di Washington che vuole la pacificazione nazionale, assumono il nome di "leggi Jim Crow".

A seguito della promulgazione delle leggi di segregazione razziale, il KuKlux Klan si sopisce, per poi riprendere fiato negli anni Venti del Novecento, e qui di nuovo si vede la potenza dell'immaginario mediatico. Nel 1915, esce nelle sale cinematografiche degli Stati Uniti un film sulla guerra di secessione, *The Birth of a Nation* di David Griffith. Nel film prevale un'immagine edulcorata e arcadica del "vecchio Sud" schiavista, rappresentata dalla famiglia dei Cameron, dove anche il KuKlux Klan gode di una decisa trasfigurazione romantica e diventa espressione dell'indignazione dei bianchi verso la protervia che pervade i neri non appena conseguita l'emancipazione. L'immaginario del film si rivela straordinariamente contagioso, e di lì a poco, all'inizio degli anni Venti, vediamo organizzarsi un nuovo KuKlux Klan, che però non è più geograficamente circoscritto al vecchio Sud, diffondendosi al

ISBN Open Access: 9788835191971

contrario nelle città del Nord e nel Midwest conservatore. L'obiettivo non sono più soltanto gli afroamericani, "che devono rimanere al loro posto" e non devono votare: il secondo KuKlux Klan recepisce anche la tradizionale avversione "nativista" contro i cattolici e l'antisemitismo che nella vecchia Europa si sta facendo strada. In breve, rilanciato da un film, il nuovo KuKlux Klan raggiunge e supera i due milioni d'affiliati, trasformandosi in un indispensabile interlocutore per la politica americana degli anni Venti, per sgonfiarsi in maniera repentina negli anni della crisi (senza che la segregazione e l'avversione agli afroamericani ne sia minimamente intaccata).

Le leggi Jim Crow sono abolite definitivamente nel 1964, sotto la presidenza di Lyndon Johnson, ma la loro abolizione a colpi di sentenze della Corte Suprema e, poi, di leggi federali, costituirà uno dei punti d'appiglio della polemica conservatrice contro le istituzioni federali, di cui parleremo più diffusamente nei prossimi paragrafi.

3. La Guerra fredda, la democrazia e la politica della paura

Oltre a sdoganare il termine "stereotipo", Walter Lippmann, con il titolo di un suo libro, applicò per primo l'appellativo di "Guerra fredda" al decennale conflitto tra Stati Uniti e Unione Sovietica. Tale conflitto si stava inasprendo nella seconda metà degli anni Quaranta ed era destinato a costituire il principale driver della politica mondiale nei successivi quattro decenni. Bene. Nel quadro della Guerra fredda, vediamo consolidarsi alcuni dei più significativi movimenti e alcune delle più significative istituzioni al cui interno si costruisce un'autentica "cultura della paura" – o, sempre con le parole di Glassner, sulle tracce di Richard Hoftstadter (1952), si instaura una politica dal profilo paranoide.

Il fenomeno del maccartismo, nell'America dei primi anni Cinquanta, è ampiamente conosciuto. Il panico si diffonde negli Stati Uniti quando si apprende che l'Unione Sovietica possiede l'atomica e, ancora di più, quando scoppia la guerra di Corea. Quest'ultima si presta a facili fraintendimenti, perché gli ambienti politici conservatori hanno gioco facile a presentarla come manifestazione dell'espansionismo sovietico, dell'aggressività del mondo comunista, che va perciò stesso contrastato su di un piano militare. Sarebbe in realtà facile controbattere ad un'interpretazione di questo tipo, dal momento che la responsabilità della guerra non può essere unilateralmente attribuita al governo nordcoreano, che ha pure iniziato le ostilità.

Un caso molto simile a questo, una trentina di anni più tardi, può essere identificato nell'invasione sovietica dell'Afghanistan, quando Mosca si infila, senza volerlo, in uno spaventoso vespaio, ma la cosa viene utilizzata dai

ISBN Open Access: 9788835191971

media occidentali per sostenere la "seconda Guerra fredda" di Ronald Reagan, intenzionato a chiudere definitivamente la partita con il mondo comunista. I due casi presentano profonde analogie: si collocano alle radici di un'escalation internazionale, ma, nel medesimo tempo, si collocano alle radici di un'escalation della paura diffusa all'interno dell'opinione pubblica, ampiamente esasperata e strumentalizzata da media e forze politiche conservatrici.

Quando si diffonde la notizia che l'Unione Sovietica possiede l'atomica, e l'anno dopo scoppia il conflitto in Corea, è presidente degli Stati Uniti il democratico Harry Truman, rieletto nel 1948 dopo aver sostituito il defunto Roosevelt nell'aprile del 1945. Truman ha fatto propri, a tutti gli effetti, i suggerimenti di un diplomatico americano, George Kennan, che nel 1947 gli scrive dall'Unione Sovietica. Kennan è dell'opinione che il sistema sovietico sia fondamentalmente sclerotico, e quindi condannato al collasso. Ciò nondimeno, ritiene che abbia una certa vocazione espansionista, legata soprattutto all'esigenza di tutelare la sicurezza del cuore dell'impero, l'Unione Sovietica[9]. Questo significa che il comunismo va fondamentalmente contenuto, ne va impedita l'espansione. Quella che abbiamo esposta è la "dottrina Kennan", che, come dicevamo, ispira poi la "dottrina Truman".

Con tutti i suoi detti e i suoi "non detti", la dottrina Truman ispirerà la politica internazionale degli Stati Uniti dal 1947 ai primi anni Sessanta. La dottrina Truman si fonda sull'assunto esplicito che gli Stati Uniti interverranno in difesa della libertà dei popoli, laddove quest'ultima sia messa a repentaglio da minoranze sediziose[10]. Con la dottrina Truman, gli Stati Uniti si fanno alfieri mondiali del "Mondo Libero", ma non escludono, nell'espletamento della propria missione, di interfacciarsi con regimi dittatoriali, che nel subcontinente latino-americano certo non mancano.

La democrazia va difesa con il "contenimento" del comunismo. Ma quando viene alla ribalta il fatto che i sovietici dispongono dell'arma atomica e scoppia la guerra di Corea, l'opposizione conservatrice ha gioco facile nell'accusare Harry Truman di essere troppo cauto nel contrastare il "pericolo rosso". La voce che gira è che la Casa Bianca, il Pentagono, La Corte Suprema, così come Broadway e Hollywood siano pieni di "criptocomunisti" che insidiano la nazione. È un intraprendente e ambizioso senatore repubblicano, Joseph McCarthy, a farsi portavoce delle paure dell'America profonda. Rimane iconica la scena di un comizio repubblicano svoltosi a Wheeling, in Virginia, nel feb-

[9] La Seconda guerra mondiale, dove circa metà dei cinquanta milioni di morti erano cittadini sovietici, enfatizzava effettivamente l'attenzione che l'élite sovietica attribuiva ai problemi della sicurezza internazionale.

[10] C'è un assunto implicito: ammesso e non concesso che queste minoranze siano d'ispirazione comunista.

ISBN Open Access: 9788835191971

braio 1950, in cui McCarthy agita un taccuino, insinuando che contenga la lista di numerosi dipendenti del Dipartimento di Stato coinvolti in attività antiamericane e di chiara fede comunista. Il problema è che nessuno ha mai aperto il taccuino che McCarthy agitava: c'era scritto qualcosa e, se sì che cosa? Era in bianco o conteneva le ricette di cucina della mamma del senatore? Conteneva il nome di persone effettivamente affiliate ad attività di carattere comunista o solo nomi di politici progressisti da infangare?

Mc Carthy si portò nella tomba il segreto nel 1957, dopo qualche giorno di coma alcolico, dovuto anche all'ostracismo cui lo aveva condannato la politica americana dopo il 1954, a seguito del suo scontro con l'esercito degli Stati Uniti – incluso il nuovo presidente, il pragmatico Dwight "Ike" Eisenhower, che pure gli doveva in buona parte la vittoria sui democratici, troppo deboli verso i "rossi". Ma Eisenhower continuò ad attenersi alla "dottrina Kennan-Truman": del resto, da comandante in capo delle truppe alleate in Europa durante la Seconda guerra mondiale, non poteva non capire le potenzialità catastrofiche di un conflitto più diretto contro l'Unione Sovietica.

Dietro a figure come quella di McCarthy, stanno anche alcuni intellettuali, come James Burnham, vecchio collaboratore di Trockij a Città del Messico, rapidamente convertitosi a posizioni conservatrici dopo la morte del suo maestro rivoluzionario. Burnham è spaventosamente popolare in quel momento. La rivoluzione manageriale, che ha scritto durante la Seconda guerra mondiale, è uno dei *best seller* negli Stati Uniti negli anni che seguono la fine del conflitto. Il libro si propone come un manifesto profondamente impolitico, che prevede l'affermazione di una "casta" manageriale e burocratica, a prescindere dalla natura democratica o dittatoriale, comunista o fascista, dei differenti regimi. Questa "casta", dal suo punto di vista, avrebbe posto fine al dominio della vecchia borghesia proprietaria, ma avrebbe anche frustrato le aspirazioni delle classi subalterne. Questo, dal punto di vista di Burnham, non è qualcosa di desiderabile o deprecabile, ma semplicemente una tendenza storica incontrovertibile. Nel bene o nel male, secondo l'autore, la proprietà sarebbe stata sostituita dalla competenza e dal merito.

Ma l'impolitico Burnham, nell'immediato dopoguerra, diventa uno dei più influenti opinionisti di parte, e nella fattispecie di parte conservatrice. Al *containment* della dottrina Kennan-Truman, preferisce il "*roll back*": dal suo punto di vista, cioè, gli Stati Uniti avrebbero dovuto mostrare il coraggio di affrontare militarmente il blocco sovietico, per farlo retrocedere. Chi ha un minimo di ragione capisce molto bene che un'opzione di questo tipo significa la Terza guerra mondiale. Ma, in fondo, a Burnham, non serve dimostrare di essere un uomo ragionevole. Gli serve solo dimostrarsi persuasivo presso l'opinione pubblica conservatrice, affaristica o evangelica. In sostanza, il suo punto di vista è funzionale alla costruzione di un "blocco storico".

ISBN Open Access: 9788835191971

In Corea, nel 1950, le cose non vanno molto bene, tanti soldati americani sono fatti prigionieri dai cinesi (e viceversa). Il generale McArthur, già comandante delle truppe americane nel Pacifico durante la Seconda guerra mondiale, chiede che vengano usate armi atomiche contro la Corea del Nord e la Cina. Forse a McArthur sfuggiva che un simile gesto avrebbe potuto significare la fine del mondo, ma forse la sua richiesta era del tutto propagandistica.

Il presidente Truman lo rimuove dal comando, e l'opposizione conservatrice ha buon gioco, dal momento che non gestisce il potere, per scatenare una feroce campagna contro una presidenza che non avrebbe il coraggio di portare avanti la lotta contro il "pericolo rosso", forse anche per le simpatie cripto-comuniste che l'avrebbero caratterizzata: non bisogna dimenticare che Henry Dexter-White, rappresentante degli Stati Uniti a Bretton Woods, era morto nel 1948 sotto interrogatorio, in quanto sospetto di attività anti-patriottica e di simpatie criptocomuniste, come non bisogna dimenticare il caso di Julius ed Ethel Rosemberg, finiti sulla sedia elettrica nel 1953, con la falsa accusa di aver trasferito ai sovietici i segreti per la realizzazione dell'atomica, è ampiamente conosciuto.

La Nuova Inghilterra della fine del XVII secolo era forse l'ultimo paese dove si conduceva la caccia alle streghe. Un celebre episodio è costituito dall'eccidio di Salem, vicino a Boston, dove, nel fosco fin del secolo morente, diverse decine di donne vengono impiccate e poi arse sulla base di ridicoli indizi che partono da alcuni giochi che una servetta di origine africana faceva fare ai bambini. La caccia alle streghe era profondamente radicata nella memoria storica americana. Soprattutto, l'opposizione conservatrice, in quello scorcio di decennio, ha bisogno di costruire alcune streghe da ardere in piazza, per convincere l'opinione pubblica che l'amministrazione cui sarebbe demandato il compito di difendere il paese è infiltrata da nemici del paese stesso, e questo spiegherebbe l'azione troppo poco incisiva dell'amministrazione stessa. Quanto più il "capro espiatorio" è visibile (è il caso di Dexter-White) o la minaccia temibile (il caso dell'atomica sovietica), tanto maggiore è l'efficacia pubblicitaria dell'evento, ma anche del panico che si diffonde. Per cui, chi ricorre, alla "politica della paura", può godere di un ritorno immediato, ma produce conseguenze ampiamente imprevedibili, che molto facilmente sfuggono al suo controllo.

Tornando alla guerra di Corea, McArthur viene quindi rimosso dall'incarico, la situazione è in fase di stallo e nel 1953 si arriva ad un trattato che restaura lo status quo, dopo circa tre milioni di morti, tra nordcoreani, sudcoreani, americani e cinesi. La guerra si risolve in un pareggio, ma a vincerla sono i conservatori americani, che nel 1952 riescono a insediare alla presidenza degli Stati Uniti Ike Eisenhower, già comandante in capo delle truppe

ISBN Open Access: 9788835191971

alleate in Europa. Divenuto presidente, come dicevamo, Eisenhower si guarderà bene dall'applicare l'idea del "*roll back*", attenendosi invece rigorosamente al "*containment*" della dottrina Truman (anche perché, da esperto militare, aveva sviluppato una sua chiara percezione dei rischi di un conflitto atomico).

A questo punto, ci riesce comodo tornare al lavoro di Walter Lippmann, che era abbastanza critico verso la gestione della Guerra fredda da parte dell'establishment degli Stati Uniti. E, nella critica che Lippmann porta avanti nei confronti di tale gestione, si ritrovano alcuni punti fermi del suo pensiero di liberale non democratico e di conservatore illuminato. In primo luogo, Lippmann vi ritrova piena conferma del suo discorso di vent'anni prima sui rapporti estremamente problematici tra politica democratica, opinione pubblica e media. Dal punto di vista del nostro giornalista, infatti, gli Stati Uniti non sono in grado di assumersi il ruolo di arbitro degli equilibri globali, se non a condizione di una profonda enfatizzazione delle funzioni centrali di governo, a discapito della tradizionale diffidenza dell'opinione pubblica americana, soprattutto di quella conservatrice, nei confronti delle istituzioni federali.

Ma la guerra di Corea, tre anni dopo, dimostra chiaramente che gli Stati Uniti stanno assumendo il ruolo di paladino planetario del "mondo libero", tanto che un'operazione bellica delle Nazioni Unite è combattuta da soldati americani e comandata da un generale americano. Tra l'altro, nel 1947 nasce la CIA (*Central Intelligence Agency*), che sostituisce l'OSS (*Office of Stategic Services*), che era stato in funzione negli anni della Seconda guerra mondiale (fondamentalmente dal 1942 al 1945). Fino ad allora, gli Stati Uniti, scarsamente coinvolti nelle controversie internazionali, non esprimevano il bisogno di servizi di *intelligence* su scala globale, da affiancare a quelli prevalentemente interni dell'FBI (*Federal Bureau of Intelligence*). Ma ora del 1947, trovandosi in misura crescente nel ruolo di arbitri della politica globale, gli Stati Uniti capiscono di non poterne più fare a meno. Nel 1946 viene così istituita la *School of Americas* o *Escuela de Americas* (SoA o EdA)[11]. La EdA è ubicata a Panama, aperta a militari di tutti gli eserciti delle Americhe, anche se gestita dal Pentagono e centrata sull'insegnamento delle tecniche della "*contro-insurgencia*". Tra gli anni Quaranta e gli anni Ottanta diplomerà sessantamila studenti, tra i quali molti dei peggiori golpisti e torturatori dell'America Latina, e per questo sarà ribattezzata *Escuela de Asesinos* (rispettando comunque l'acronimo EdA).

Lippmann ritiene che l'unica soluzione ragionevole consisterebbe in una distensione diplomatica, previo reciproco riconoscimento delle rispettive

[11] Il nome originario è Latin American Trading Center. La denominazione definitiva è assunta nel 1963.

ISBN Open Access: 9788835191971

sfere d'influenza. Ma si rende ben conto del fatto che è molto difficile che il suo punto di vista trovi ascolto presso la classe politica americana, non malgrado, ma proprio in virtù del carattere democratico della politica americana. L'amministrazione Truman, che Lippmann critica per le ragioni sopra spiegate, è comunque attaccata dall'opposizione di destra che, in un clima di panico diffuso, chiede una politica molto più aggressiva (che poi non mette in atto una volta conquistata la presidenza).

Ma qui si colloca un'ulteriore contraddizione. L'opinione pubblica conservatrice ed evangelica richiede una politica internazionale molto aggressiva, che però porta ad enfatizzare il ruolo delle istituzioni federali. L'avversione nei confronti del livello federale è uno degli aspetti più caratteristici della cultura politica del conservatorismo americano, dal Midwest al "Vecchio Sud" (Pally, 2022). Ma questa stessa contraddizione si ritrova anche con riferimento alla sicurezza interna. Come ricorda un grande politologo americano, Jonathan Simon, negli anni Trenta, Roosevelt utilizzò la lotta al crimine per enfatizzare il ruolo delle istituzioni politiche federali, dal momento che le reti criminali non erano quasi mai circoscrivibili ai singoli stati. L'implementazione del ruolo delle istituzioni federali serviva, in realtà, per attuare i programmi del New Deal. Lippmann (1937) è profondamente critico nei confronti del New Deal, anche perché, da buon conservatore, diffida del livello federale, oltre che della democrazia più in generale.

L'avvento della Guerra fredda è quindi un'ulteriore riprova delle fragilità della politica democratica e del centralismo politico. Non è un caso che la Guerra fredda sia fondamentalmente nata dalla diffidenza degli Stati Uniti e della Gran Bretagna nei confronti del vecchio alleato sovietico, molto più che non dall'atteggiamento della cauta nomenclatura sovietica, appena uscita da una guerra dove aveva perso venticinque milioni di persone sul totale dei cinquanta milioni di vittime della Seconda guerra mondiale. La nomenclatura sovietica, in un contesto che si caratterizza per l'assenza di un'opposizione che non sia dissidenza, è in grado di gestire la propria opinione pubblica. Non è assente una qualche forma di "politica della paura", ma di carattere profondamente diverso rispetto ad un regime democratico. L'avversario imperialista è certo agguerrito, occorre difendersi da lui, e i dissidenti sono il suo braccio armato. Rivolte come quelle di Berlino del 1952 e di Budapest nel 1956 sono crepe che rischiano di far crollare il muro di difesa. Ma, in fondo, si evita di generare un'escalation della paura, come quella che caratterizza i democratici Stati Uniti.

Al contrario, in un contesto democratico, caratterizzato dalla presenza di una competizione elettorale, l'opposizione è portata a dimostrare una serie di cose, e la presenza di una stampa libera la aiuta in questo suo compito. In particolare, deve dimostrare all'uomo della strada che qualcuno lo minaccia,

ISBN Open Access: 9788835191971

e che chi dovrebbe difenderlo non lo fa. Deve dimostrare all'uomo della strada che qualcuno vive alle sue spalle, e che chi potrebbe impedirlo non lo fa. L'auspicabile libertà d'opinione e d'opposizione rimane uno dei beni che vanno difesi, ma si presta facilmente a dei loop, come quelli dell'America degli anni Cinquanta.

4. Jim Crow a Detroit. Politiche della paura, immaginario neoconservatore e genesi del neoliberalismo

Il già citato Jonathan Simon, con estrema sottigliezza, evidenzia come non ci sia soltanto un "governo del crimine", ma anche un "governo attraverso il crimine", cioè una strumentalizzazione della questione criminale, da parte del governo in carica, per il perseguimento di determinati obiettivi politici. Alle considerazioni del politologo americano, si potrebbe aggiungere che, ancora più efficace, come nei casi sopra mostrati, esiste un'opposizione della paura. La paura è, cioè, più facilmente strumentalizzabile da parte dell'opposizione che non del governo, che si trova sempre, in qualche modo, a rispondere dello stato di cose presente.

La cosa si ripete, infatti, una quindicina d'anni dopo, quando l'opinione pubblica americana si scopre coinvolta nella guerra del Vietnam. Arrivano le presidenziali del 1964, e il candidato repubblicano, il conservatore Barry Goldwater, nuovamente attacca l'uscente Lyndon Johnson, accusandolo di un atteggiamento troppo morbido nei confronti dell'avversario escatologico. Goldwater non escluderebbe, per suo conto, l'utilizzo di ordigni nucleari a bassa intensità per interrompere la linea di rifornimento che collega il nord e il sud del paese. Goldwater non ha chiesto l'utilizzo dell'atomica contro le città vietnamite, ma il suo discorso è facilmente utilizzabile per scatenare il panico. Alla televisione americana viene trasmesso, sia pure un'unica volta, il "*Daisy Ad*", uno spot elettorale in cui si vede una bambina che gioca con una margherita, si innesca un conto alla rovescia e si scatena alla fine un'esplosione nucleare. Si sente la voce di Johnson che ammonisce che se non ci amiamo tra uomini siamo destinati a morire. Trasmesso una sola volta, lo spot comincia a essere discusso, elogiato o criticato su tutti gli altri canali mediatici, trasformandosi in uno dei più significativi successi pubblicitari della storia americana. Per quanto il coinvolgimento americano nel Vietnam fosse stato opera di un'amministrazione democratica, i democratici colgono la palla al balzo: alla paura suscitata dal "pericolo rosso", rispondono soffiando sulla paura suscitata "pericolo atomico". Questo fu forse cruciale per garantire la vittoria di Lyndon Johnson, come il taccuino di McCarthy fu forse cruciale per garantire quella di Eisenhower.

ISBN Open Access: 9788835191971

Un sociologo italiano scomparso una trentina d'anni fa, Roberto Guiducci (1986), in un suo scritto di metà anni Ottanta, ragionando sulla Guerra fredda, parla di una "vittoria postuma di Hitler". Il suo acuto ragionamento può essere così riassunto: gli americani sviluppano l'arma atomica, con il "progetto Manhattan", per impedire che Hitler possa arrivare a possederla per primo. Quando l'ordigno è pronto, Hitler è già morto suicida nel bunker di Berlino, ma l'atomica viene utilizzata contro il Giappone, refrattario alla resa, anche come monito all'Unione Sovietica, quando la Guerra fredda è ormai in gestazione.

I sovietici, sentendosi con le spalle al muro, sviluppano a loro volta l'arma nucleare e, quando negli Stati Uniti si diffonde la notizia, si scatena il panico e ha inizio il fenomeno noto come "caccia alle streghe". Una crescente quantità di denaro comincia a essere investita nel settore bellico, tanto che, alla scadenza del suo mandato, Dwight Eisenhower lamenta il fatto che gli Stati Uniti sono ormai nelle mani di un "complesso militar-industriale". Le commesse belliche durante la Seconda guerra mondiale, e i successivi investimenti tecnologici negli anni della Guerra fredda, hanno un ruolo di primo piano nel traghettare gli Stati Uniti fuori dalla recessione dei primi anni Trenta. È sufficiente ricordare che, senza la Guerra fredda, l'uomo non sarebbe arrivato sulla Luna, le sonde americane e sovietiche non sarebbero arrivate rispettivamente su Marte e Venere e oggi noi non perderemmo le giornate a chattare sui social o a leggere utili messaggi nella nostra casella di e-mail.

Guiducci ragiona su tutto questo, a metà anni Ottanta, nella fase finale della Guerra fredda, in un mondo percorso dalle speculari paure della dominazione sovietica e dell'annientamento nucleare: "*better dead than red or better red than dead?*", ci si domanda all'epoca. Guiducci osserva come l'assurdità dell'alternativa tra la perdita della vita pur di non rinunciare alla libertà o la perdita della libertà pur di non rinunciare alla vita sia il prodotto del clima patologico creato dal circolo vizioso paura-riarmo. Ci troviamo in un mondo sul cui capo pendono numerose spade di Damocle, che sono i Pershing II americani, gli SS-20 e i MiG sovietici… poco importa che uno "scudo spaziale" sia lì a parare i colpi della spada di Damocle: chi ha memoria di quel tempo, ricorda bene il fatto che lo "scudo spaziale" di Reagan contribuì ampiamente alla tensione internazionale, rendendo quindi più probabile l'ipotesi di un conflitto.

Il messaggio di Guiducci è chiaro. L'escalation militare è stata in primo luogo innescata da Hitler, autentico maestro nell'utilizzo politico della paura. L'hanno raccolta americani e sovietici all'inizio della Guerra fredda. Nell'arco di quindici anni, è divenuta tale da mettere a repentaglio l'esistenza stessa del mondo, come si capisce perfettamente in occasione della crisi dei missili di Cuba. Se veramente Stati Uniti e Unione Sovietica avessero dovuto arri-

ISBN Open Access: 9788835191971

vare ad annientarsi reciprocamente in un conflitto atomico, sostiene Guiducci, questa sarebbe stata la "vittoria postuma di Hitler" contro le due potenze che lo avevano sconfitto e che avevano processato e messo a morte i suoi gerarchi corresponsabili della Shoah ebraica. L'eventuale "vittoria postuma di Hitler" avrebbe potuto produrre molte più morti dei sei milioni di vittime dei lager, molte più morti dei cinquanta milioni di vittime della Seconda guerra mondiale, metà delle quali sovietici.

Ma torniamo a Barry Goldwater. Costui non riuscirà a farsi eleggere presidente degli Stati Uniti, ma non si può dire che non abbia ottenuto comunque una sua vittoria, forse più significativa e duratura di quanto non lo sarebbe stata una mera vittoria elettorale. Attorno alla sua figura comincia a strutturarsi un significativo "blocco storico", un movimento conservatore, cementato da un fervente anticomunismo e da un forte indirizzo neoliberale in economia, che comunque si collega ad un altrettanto forte conservatorismo religioso, capace di porre sullo stesso tavolo il fondamentalismo evangelico e il tradizionalismo cattolico. Sono lontani i tempi rappresentati da *Gangs of New York*, ma in fondo anche quelli del Secondo KuKlux Klan. Già negli anni Quaranta diverse autorità religiose, "fondamentaliste" evangeliche e "tradizionaliste" cattoliche, avevano auspicato l'unità di tutte le chiese in una crociata contro il comune "pericolo rosso", a partire dal pastore presbiteriano Carl McIntire.

All'inizio degli anni Sessanta, questa prospettiva neoconservatrice sembrava aprire le braccia a tutti quelli che univano un forte sentimento religioso con un altrettanto forte sentimento dell'identità nazionale, ovvero che univano l'una o l'altra cosa con un'indistruttibile fede nel libero mercato e con un'elevata avversione nei confronti delle istituzioni federali, nonché, chiaramente, del comunismo.

Magari non è facile capire il tutto per gli estranei, ma in fondo lo può diventare. Le istituzioni federali, negli anni di Johnson, proclamano la guerra alla povertà, cioè il più ardito sistema di welfare mai concepito negli Stati Uniti. La guerra alla povertà è facilmente attaccabile dalle opposizioni di destra, che biasimano il fatto che gruppi di persone pigre, refrattarie al lavoro e sessualmente lascive gravino sul bilancio di sane famiglie bianche e della middle class.

La guerra del Vietnam è facilmente attaccabile su tre fronti: i contestatori di Berkley domandano perché dovrebbero morire in una guerra contro un pacifico popolo di agricoltori che chiedono solo il diritto all'autodeterminazione, e i conservatori chiedono perché la guerra non sia portata avanti con maggiore determinazione, i neri non vogliono partire per la guerra dei bianchi, e dopo il 1964 i ghetti cominciano a esplodere. L'esplosione conferma i conservatori bianchi nelle proprie convinzioni. Perché era stato possibile

ISBN Open Access: 9788835191971

sconfiggere la titanica potenza di Hitler e del Giappone, ai tempi della Seconda guerra mondiale, e ora non si riesce ad avere la meglio di un miserabile popolo di agricoltori, guidato da un leader molto determinato? Evidentemente l'America non è più quella di una volta, perché gli studenti di Berkeley, che appartengono a un'élite, dicono che non vogliono andare a morire per il Vietnam, e la stessa cosa fanno i paria del ghetto nero. Neanche loro vogliono andare a morire per il Vietnam. Il rifiuto di andare in guerra è un chiaro sintomo della decadenza della nazione.

Apparentemente, il tema è privo di nessi con l'adesione ad orientamenti di carattere neoliberale, a meno che non si attribuisca il declino spirituale della nazione al carattere deresponsabilizzante delle politiche di welfare, inaugurate dal New Deal rooseveltiano e implementate, negli anni Sessanta, dalla "war on poverty" di Lyndon Johnson (Cartosio, 1998). Nell'ottica neoconservatrice, queste ultime non porterebbero ad altro che alla formazione di un'underclass, specializzata nel ricorrere a qualsiasi sotterfugio, pur di vivere alle spalle del contribuente, cioè dell'americano bianco middle class che lavora e conduce una tranquilla vita familiare in un villino con piccolo giardino nello sconfinato suburbio (qualcuno forse ricorda "American Beauty", uno splendido film che ritrae il "tranquillo" universo suburbano di fine anni Novanta). Complementare a questa underclass, nella retorica neoconservatrice, è la "casta" dei dipendenti del sistema del welfare[12], che comunque gravano sulle spalle del contribuente produttivo e la cui sola ragione di esistenza è il presunto opportunismo di chi decide di vivere alle spalle degli altri.

Qui potremmo chiamare in causa un significativo studioso tedesco di politica sociale, Claus Offe (1984). Offe si occupa dell'emergere delle politiche neoliberali e, con una certa arguzia, osserva la specularità tra la critica radicale di fine anni Sessanta ai sistemi di welfare e la critica neoconservatrice di fine anni Settanta. Negli anni Sessanta, la "nuova sinistra" accusa il welfare di "narcotizzare" le masse popolari con le briciole che elargisce, mentre, una decina di anni dopo, il neoconservatorismo accusa gli stessi sistemi di welfare di "deresponsabilizzare" gli individui. Negli anni Sessanta, la sinistra rivoluzionaria accusa la socialdemocrazia di essere insostenibile, perché fondata su di un'estrazione dai profitti in un'economia basata su di una "caduta tendenziale del saggio di profitto" che caratterizzerebbe il capitalismo secondo Marx. Negli anni Settanta, la nuova destra accusa i sistemi di welfare di essere insostenibili, perché basati sulla spoliazione di chi produce veramente ricchezza, a vantaggio di indolenti parassiti. Così l'immaginario neoconservatore degli anni Settanta costruisce nuovi "diavoli popolari", mostri immaginari, rici-

[12] Tipico, nella retorica neoconservatrice, lo stereotipo della assistente sociale, specie se di colore, che vive delle presunte disgrazie dei suoi assistiti.

ISBN Open Access: 9788835191971

clando molti dei luoghi comuni già presenti nell'immaginario del Know Nothing, del primo e del secondo KuKlux Klan, della caccia alle streghe.

Recuperando Freud, i topos dell'immaginario condensano al proprio interno differenti aspetti, numerosi dei quali provengono da più arcaici prototipi. Vi è un particolare demone popolare, particolarmente caro ad uno dei più solerti interpreti della cultura politica neoconservatrice dell'ultimo quarto del XX secolo, Newt Gingrich, che continua a chiamarlo in causa nei suoi discorsi pubblici, esattamente come Joseph McCarthy aveva agitato il proprio taccuino, presumibilmente vuoto. Gingrich addita all'opinione pubblica l'immagine della minorenne, presumibilmente appartenente ad una minoranza etnica, che si fa mettere incinta in un rapporto occasionale, per ottenere il pubblico sussidio in quanto madre single. Questo demone popolare condensa numerosi delle ossessioni della cultura politica neoconservatrice e si presta egregiamente ad evidenziare gli stretti legami che esistono tra dottrine economiche neoliberali e destra religiosa. Da un lato c'è il declino dei valori e il degrado dei costumi e dall'altro la propensione di certi individui, ma soprattutto di determinate categorie sociali a vivere alle spalle degli altri.

Il declino dei valori, la crisi della famiglia, il degrado dei modi di vita sono tematiche particolarmente care alla destra conservatrice. Ma si coniugano splendidamente ad altre tematiche caratteristiche del neoliberalismo del tempo, propenso a descrivere i sistemi di welfare come ambigui e torbidi dispositivi che consentono a soggetti pigri, indolenti, incapaci e disonesti di vivere alle spalle di chi lavora (e forse fuori luogo un riferimento al recente dibattito sul reddito di cittadinanza, che ha interessato il nostro paese dal 2017 all'avvento del governo Meloni nel 2023?). Fanno questo alimentando, come fenomeno complementare, una casta di burocrati del servizio sociale, parimenti mantenuto da iniqui sistemi fiscali alle spalle dei ceti produttivi. Di nuovo, si coglie la specularità di cui parlava Claus Offe: da una parte, la "vecchia nuova sinistra" (mi si perdoni l'ossimoro!) teorizza un'alleanza strutturale tra i detentori del capitale e le burocrazie di stato, alle spalle dei veri produttori di ricchezza, cioè i soggetti costretti a vendere la propria forza lavoro. Dall'altra, la destra conservatrice e liberale (scusate, di nuovo, l'ossimoro!) denuncia l'alleanza tra la casta parassitaria dei burocrati di stato e dei professionisti del welfare e i pidocchi sociali che aspirano a vivere sulle spalle degli altri. Questi ultimi, guarda che caso, sono membri di minoranze etniche, figli della cultura del ghetto, figli di quella che l'antropologo Oscar Lewis, negli anni Sessanta, aveva definito la "cultura della povertà".

Alla malefica intesa tra fannulloni e professionisti del sociale non può che contrapporsi la "santa alleanza" tra i ceti produttivi, quella "rivolta di Atlante" che negli anni Cinquanta era stata auspicata dalla scrittrice anarcocapitalista Ayn Rand. La Rand immagina, in qualche modo, una seconda

ISBN Open Access: 9788835191971

secessione, che non è quella degli stati schiavisti contro un presunto nord abolizionista, ma di imprenditori, tecnici e altri creatori di ricchezza, che rivendicano la propria indipendenza contro i vincoli imposti e i latrocini perpetrati dal parassitismo delle burocrazie, dei sindacati, del governo federale, dei nullafacenti. Atlante porta sulle proprie spalle il peso del mondo, per cui il dibattito tra la "vecchia nuova sinistra" e la "destra conservatrice e liberale" potrebbe anche essere così sintetizzata: chi è davvero Atlante? La classe operaia alienata e sfruttata dal capitalista parassita (secondo la versione marxista), oppure l'eroico imprenditore, vittima di un'opprimente burocrazia statale che tutela a sue spese schiere di immorali nullafacenti (secondo il punto di vista anarco-capitalista della stessa Rand)?

Stiamo bene attenti. Lo spettro della procace parassita adolescente, proposto da Newt Gingrich, condensa al proprio interno anche due dei più caratteristici elementi dell'immaginario conservatore: le minoranze etniche, soprattutto afroamericane, da un lato; il governo federale dall'altro, che imperterrito interferisce con la vita delle comunità locali, imponendo l'abolizione della schiavitù, misure contrarie alla segregazione razziale, il diritto all'aborto, vincoli al diritto di portare armi e, a un certo punto, tra il 1972 e il 1976, persino l'abolizione della pena di morte. Dal livello federale, ai tempi di Roosevelt e poi di Johnson, sono venuti il New Deal e poi la War on Poverty, finalizzati alla costruzione di un welfare state e, per ciò, esecrati dai gruppi più conservatori. Non è un caso che, ai tempi della "caccia alle streghe", a essere presi di mira fossero soprattutto membri delle istituzioni federali, oltre che dell'industria culturale e cinematografica.

Quella proposta da Gingrich è una rappresentazione, e qui dobbiamo stare di nuovo attenti. Come ci insegnano Antonio Gramsci, e sulla sua scorta Stuart Hall, con buona pace di Durkheim le rappresentazioni non sono il prodotto di un'indefinibile società, ma di coalizioni egemoniche, che devono il proprio carattere egemonico alla propria capacità di affermare uno specifico punto di vista attraverso efficaci rappresentazioni. La furba e lasciva adolescente nera o portoricana di Gingrich ci racconta tante cose su diversi segmenti di un'America conservatrice, evangelica, allergica al "*big government*" federale, diffidente verso le minoranze etniche – in generale diffidente, al punto da essere convinta che qualsiasi intervento sociale a livello federale non possa fare altro che scatenare gli appetiti dei "peggiori" (peraltro appartenenti a minoranze etniche e probabilmente non affiliati ad alcuna chiesa) contro i "migliori".

Una rappresentazione contiene già un programma di governo. In occasione della campagna del 1964, prende la parola a favore di Goldwater un politico non ancora particolarmente conosciuto, un certo Ronald Reagan, che sarebbe poi diventato governatore della California e infine presidente degli

ISBN Open Access: 9788835191971

Stati Uniti. L'intervento si intitola "*Time for Choosing*", e si rivela particolarmente efficace nella costruzione di un blocco conservatore, nei successivi decenni. L'immagine base è quella dell'"appuntamento con la Storia", che ci chiama a difendere la libertà o a sprofondare nella barbarie. Reagan dice che, con Goldwater, il programma politico conservatore ha fatto un salto di qualità. Non è più soltanto cauta opposizione alle trasformazioni volute dalla sinistra, che è per ciò stesso egemone, ma si rivela capace di lavorare attorno ad un proprio specifico nucleo di convinzioni.

Lo stesso Reagan lo ritroviamo, nel 1980, all'atto dell'accettazione della nomination repubblicana per la presidenza degli Stati Uniti, che lo vedrà poi trionfare su Jimmy Carter. Il congresso repubblicano del 1980 ha luogo a Detroit, la città che ospita l'industria dell'automobile in crisi. Quello di Detroit è un caso molto particolare. Negli anni Quaranta, Detroit è una città prevalentemente abitata da esponenti della working class bianca. All'inizio degli anni Settanta è ormai una città a stragrande maggioranza nera (circa il 75% degli abitanti). Detroit non ha un ghetto, come possono esserlo Harlem e il Bronx di New York o Bronzeville a Chicago. Detroit *è* un ghetto. Ormai i bianchi abitano nel suburbio, e forse molti dei bianchi che sono lì ad acclamare Reagan vengono dal suburbio di Detroit. E Reagan fa un discorso finalizzato a catturare la fiducia di un'America demoralizzata dalla guerra del Vietnam e terrorizzata dalla crisi economica, dallo shock petrolifero e, di recente, dalla rivoluzione iraniana e dall'invasione sovietica dell'Afghanistan, nonché dalle rivoluzioni in Nicaragua e nel Salvador. Somma le paure planetarie della Guerra fredda a quelle urbane, molto diffuse in quegli anni Settanta che volgono al termine e che non a caso un grande storico americano, Philip Jenkins, definì nel titolo di un suo libro "il decennio degli incubi". Crisi internazionale, crisi urbana e degrado morale convergono, comunque, nella narrazione neoconservatrice, dove l'incapacità di contrastare l'espansione del comunismo e il degrado dei valori della nazione americana sono i due rovesci della stessa medaglia.

C'è da stare attenti, a questo punto, a qualche sottile differenza. Tutte le questioni legate alla Guerra fredda servono per accattivarsi il consenso di una platea che evidentemente non aveva gradito il relativo disimpegno di Carter, che nel 1977 dichiarava conclusa la Guerra fredda. Non hanno però niente a che vedere con la rivoluzione neoliberale, ma sono comunque uno specchietto per le allodole a uso di chi le voglia promuovere. La Guerra fredda ha enfatizzato il ruolo delle aborrite istituzioni federali, che tipicamente controllano le forze armate e la diplomazia e che, soprattutto a partire dagli anni di Roosevelt, hanno cominciato a implementare un sia pur blando controllo dell'economia e delle politiche sociali. Ogni riferimento alla Guerra fredda ha una sua indiscutibile presa su di un elettorato conservatore. Ma

ISBN Open Access: 9788835191971

quello che può fare la differenza è una nuova guerra, che non è più una guerra alla povertà (come quella di Johnson), ma una guerra ai poveri.

Ronald Reagan, quella sera, sta parlando della situazione internazionale, dove negli ultimi quattro anni[13] "il nemico" (cioè quello che più avanti definirà l'"impero del male") è avanzato[14]. Passa alla situazione economica, parimenti negativa[15], e incalza i presenti: "Quanti di voi, quattro anni fa, hanno votato Jimmy Carter e ora sono senza lavoro?", "Quanti di voi oggi sono più felici di quattro anni fa?". I presenti, in cappello texano, esultano e fanno suonare le trombette. A questo punto, il futuro presidente passa alla parte propositiva. Aumenteremo la spesa militare, ma nello stesso tempo ridurremo le imposte (non lo dice, ma è evidente l'intento di ridurre le spese sociali). Ma l'aspetto più interessante è il seguito, dove veramente noi troviamo lo stigma urbano corrispondente alla madre single minorenne di Gingrich. Reagan dice che "nessuno sarà lasciato indietro". Anzi, ognuno dovrà avere modo di vivere in modo dignitoso, "anche nella inner city in cui abita".

Potremmo intanto domandarci cosa significhi vivere in modo dignitoso, ma ancora più interessante è interrogarsi sul significato del riferimento alla inner city. Vivere in modo dignitoso, evidentemente, nell'ottica di un neoliberalismo intriso di valori evangelici, vuol dire vivere del sudore della propria fronte, senza pesare in alcun modo sulla collettività: il contrario, cioè, della giovane e procace parassita di Gingrich, cui sembra rivolto l'invito. Ma quest'ultima, dal discorso di Reagan, sembra avere dimora in una inner city. Il discorso di Reagan, dicevamo, è stato pronunciato a Detroit, una città che è l'incarnazione del discorso del candidato presidente.

Autentica "Mecca" dell'automobile, sede della Ford e della General Motors, negli anni Quaranta Detroit era fondamentalmente abitata da una popolazione bianca working class. Alla fine degli anni Settanta, ormai, la presenza bianca è esigua, e oltre il 75% degli abitanti sono di colore, mentre la popolazione bianca, come nella gran parte delle grandi città americane, si è spostata verso il suburbio. A Detroit, pochi anni prima, era stato eletto il primo sindaco di colore di una grande città, Coleman Young, un democratico di sinistra particolarmente attento alle problematiche delle minoranze e dei settori più marginali della popolazione, e per questo profondamente inviso all'elettorato bianco, conservatore ed evangelico. Il declino del settore automobilistico, naturalmente, ha profondamente inciso sulle possibilità di reddito della popolazione, quindi la concentrazione di popolazione dalla pelle nera si sovrappone con la concentrazione di popolazione povera. Il senso d'in-

[13] Gli anni della presidenza del suo sfidante, l'uscente Jimmy Carter.

[14] Il riferimento è all'invasione sovietica dell'Afghanistan, alla rivoluzione sandinista in Nicaragua, all'ancora difficilmente interpretabile rivoluzione iraniana, ecc.

[15] Siamo nel pieno del secondo shock petrolifero.

ISBN Open Access: 9788835191971

sicurezza, come in tutte le altre città americane del tempo, è estremamente diffuso, tanto che una popolare battuta dell'epoca dice che la "Motor city" si è trasformata nella "Murder city". In breve, Detroit sembra condensare al proprio interno buona parte degli spettri che nell'inquieto decennio degli anni settanta si agita davanti agli occhi dell'opinione pubblica conservatrice: il declino della vecchia nazione protestante, che trova espressione nell'aumento della criminalità, nel declino dei costumi, nell'emergere di un'underclass di parassiti e di una complementare, e altrettanto parassitaria, casta di professionisti del welfare; l'affermarsi negli scenari della politica di minoranze etniche, che nel contempo sono spesso protagoniste di forti conflitti urbani.

Una nazione in declino sarà in grado di portare avanti la missione, che la storia sembrava averle assegnato a partire dal 1945, di contrastare l'espansione del comunismo? Jimmy Carter, appena arrivato alla presidenza, nel 1977, aveva affermato che la Guerra fredda poteva considerarsi ormai conclusa, e quest'affermazione, osservano alcuni storici (Jenkins, 2006), era destinata a costargli particolarmente cara, perché gli aliena tutto il voto evangelico (ma anche cattolico tradizionalista) che nel novembre 1976 aveva ampiamente contribuito alla vittoria di quest'uomo estremamente devoto, impegnato a livello ecclesiastico. Il solo problema è che, evidentemente, non condivide i "demoni popolari" di quest'elettorato religioso e conservatore, non vive nello stesso "pseudo-ambiente", nello stesso Umwelt. Al contrario, l'attore Ronald Reagan è capace di captare le inquietudini del proprio interlocutore e di sintetizzarle in poche battute; è capace di captare le emozioni e di parlare alle emozioni. Con poche battute, è capace di proporsi come cardine di un blocco storico che unisce gli interessi del grande capitale industriale e finanziario, delle nuove classi professionali, delle destre religiose, evangeliche o cattoliche, di segmenti di ceti popolari conservatori, soprattutto nel Vecchio Sud e nel Midwest. Questa sua abilità "egemonica" contribuì non poco al suo trionfo nel novembre del 1980.

Riferimenti bibliografici

Agustoni, A. (2022), *Città e sistemi mondo*, Carocci, Roma.

Agustoni, A. (2024), *Città e politiche della paura*, Carocci, Roma.

Castells, M. (1997), *La forza dell'identità*, ed. it. Egea, Milano.

Cohen, S. (1972), *Demoni popolari e panico morale*, ed. it. Mimesis, Bari, 2019.

Downs, A. (1957), An Economic Theory of Political Action, *The Journal of Political Economy,* 2: 135-150.

Elias, N. (1989), *Che cos'è la sociologia*? Rosenberg&Sellier, Torino.

Gans, H. (1962), *The Urban Villagers*, Free Press, NY.

ISBN Open Access: 9788835191971

Glassner, B. (2018), *The Culture of Fear. Why Americans are Afraid of the Wrong Things?*, Basic Books, NY.
Guiducci, R. (1986), *Ti uccido come un cane. Violenza e magia nera in età nucleare*, Rizzoli, Milano.
Hannerz, U. (1996), *La diversità culturale*, il Mulino, Bologna.
Hugill, P. (1993), *World Trade Since 1431*, John Opkins U.P., Baltimore.
Illouz, E. (2024), *Modernità esplosiva,* Einaudi, Torino.
Jenkins, Ph. (2006), *Decade of Nightmares,* Oxford University Press, Oxford.
Jenkins, Ph. (2021), *A Global History of the Cold War*, Palgrave MacMillan, NY.
Lippmann, W. (1922), *Public Opinion*, ed. it. Donzelli, Roma, 2000
Lippmann, W. (1947), *The Cold War. A Study in US Foreign Policy*, True Oak Books, Highland, NY.
Nye, J. (2004), *Soft Power*, ed. it. Einaudi, Torino, 2005.
Offe, C. (1984), *Modernity and the State. East and West*, Polity Press, Cambridge.
Pally, M. (2022), *White Evangelicals and Right-Wing Populism*, Routledge, New York.
Simmel, G. (1908), *Sociologia*, ed. it. Comunità, Milano, 2008.
Simon, J. (2006), *Il governo della paura*, ed. it. Cortina, Milano, 2008.
Turchin, P. (2024), *End Times. Elites, Counter Elites and the Path of Political Disintegration,* Penguin, NY.
Uexküll, J. (1933), *Ambienti umani e ambienti animali*, ed. it. Quodlibet, Roma, 2010.

ISBN Open Access: 9788835191971

Xenofobia online: populismi e othering nella sfera pubblica digitale

di *Alessandra De Luca*[*], *Mara Maretti*[**]

1. Introduzione

Le trasformazioni della sfera pubblica sul tema dell'immigrazione sono inscindibilmente legate all'infrastruttura tecnologico-comunicativa in cui si sviluppano. Negli ultimi vent'anni i social media non hanno semplicemente amplificato atteggiamenti xenofobi preesistenti, ma hanno contribuito a riconfigurarli, offrendo nuove grammatiche espressive, nuovi spazi di aggregazione, nuove dinamiche di diffusione e nuove narrazioni (Castells, 2009; Gagliardone *et al.*, 2015). Comprendere la xenofobia contemporanea richiede pertanto un'analisi che tenga insieme la dimensione ideologica e quella infrastrutturale della comunicazione digitale, interrogandosi su come contenuti, piattaforme e ideologie politiche si influenzano reciprocamente.

Il presente contributo si propone di tracciare una mappa interpretativa dell'odio razziale online, muovendo da un'analisi della letteratura scientifica internazionale e sviluppando una riflessione teorica sui principali nuclei tematici emergenti, in grado di analizzare tale fenomeno sociale in modo dinamico. Il mapping tematico, condotto su un corpus di 348 documenti pubblicati tra il 2002 e il 2025, ha rappresentato un valido punto di partenza per meglio identificare le traiettorie evolutive della riflessione sociologica su tale tema. Ne emerge un percorso interpretativo che può essere articolato lungo tre assi principali, tra loro concatenati.

Il primo asse riguarda la transizione, che si evince dal mapping tematico, dal paradigma multiculturalista a quello dell'islamofobia. Infatti, se fino alla metà degli anni Duemiladieci la letteratura si concentrava prevalentemente

[*] Dipartimento di Tecnologie Innovative in Medicina & Odontoiatria, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, alessandradeluca1197@gmail.com.

[**] Dipartimento di Scienze Filosofiche, Pedagogiche e Sociali, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, mara.maretti@unich.it.

ISBN Open Access: 9788835191971

sulle sfide della convivenza in società plurali, a partire dal 2015-2016 si osserva una focalizzazione crescente sulla figura del migrante musulmano come incarnazione dell'alterità minacciosa. Questa svolta non è casuale, perché, come documentato da Vertovec e Wessendorf (2010), si era manifestato, in quegli anni, un backlash contro il multiculturalismo nelle retoriche politiche europee, con leader come Merkel, Cameron e Sarkozy che ne avevano proclamato il "fallimento". Tale delegittimazione discorsiva ha creato lo spazio per narrazioni alternative, incentrate sulla presunta incompatibilità culturale dell'Islam con i valori occidentali. La cosiddetta "crisi dei rifugiati" del 2015, l'intensificarsi della minaccia terroristica legata all'ISIS e l'ascesa di retoriche politiche esplicitamente antislamiche (di cui la campagna presidenziale di Donald Trump rappresenta l'esempio più eclatante) hanno accelerato questo processo (Chavez *et al.*, 2023).

Il secondo asse del percorso argomentativo emergente della letteratura concerne la mobilitazione identitaria della destra radicale. L'emersione simultanea, nel 2018, di temi quali il nazionalismo e la whiteness segnala l'affermarsi di una narrativa che definisce la paura dello straniero in termini esplicitamente razziali e identitari. Non si tratta più soltanto di resistenza culturale all'immigrazione, ma della rivendicazione di un'identità bianca percepita come minacciata. Questo processo trova la propria cornice narrativa nella teoria della "grande sostituzione" (grand remplacement), elaborata da Renaud Camus (2011) e rapidamente diffusasi negli ambienti della destra radicale transnazionale (Mudde, 2007). Come ha mostrato Bonilla-Silva (2013), il razzismo contemporaneo opera sempre più frequentemente attraverso forme "color-blind" che negano la propria natura discriminatoria pur producendo effetti di esclusione sistematica. La mobilitazione identitaria bianca rappresenta un ritorno a forme più esplicite di rivendicazione etnonazionale. La destra radicale populista, definita da Mudde (2007) attraverso la triade ideologica di nativismo, autoritarismo e populismo, emerge come attore centrale nella produzione e circolazione di questo discorso xenofobo.

Il terzo asse introduce la dimensione propriamente tecnologica. A partire dal 2020, la letteratura inizia a interrogarsi sistematicamente sul ruolo delle singole piattaforme, Facebook e Twitter in particolare, come ambienti specifici in cui il discorso xenofobo assume forme e dinamiche peculiari (Yarchi, Baden e Kligler-Vilenchik, 2021; Kakavand, 2024). Le affordance tecnologiche, gli algoritmi di raccomandazione, le logiche della viralità non sono elementi neutri, ma contribuiscono attivamente a modellare i contenuti che ospitano (Siegel, 2020) e a ridefinire i processi di othering. Come teorizzato da Castells (1996) nella sua analisi della "società in rete", le tecnologie digitali non costituiscono un mero strumento di comunicazione, ma ridefiniscono le coordinate stesse della sfera pubblica. La preoccupazione per le echo

ISBN Open Access: 9788835191971

chambers (Sunstein, 2017) e le filter bubbles (Pariser, 2011), sebbene empiricamente contestata nella sua portata (Ross Arguedas *et al.*, 2022), sembra oramai acclarata poiché risulta ampiamente riconosciuto come l'architettura delle piattaforme possa favorire dinamiche di polarizzazione e radicalizzazione. Emerge inoltre una differenziazione geografica e tematica: Facebook appare maggiormente, anche se non esclusivamente, connesso al populismo di destra europeo e al dibattito sulla Brexit (Russo e Maretti, 2025; Hall, 2021), mentre Twitter/X si configura come arena privilegiata del discorso politico statunitense (Tillery, 2019; Goodman, Perkins e Windel, 2024; Corsi, 2024). I social media sembrano quindi ridefinire i meccanismi di costruzione dell'alterità e il loro intreccio con la disinformazione. Il concetto di othering, la demarcazione simbolica tra un "noi" e un "loro" (Bonacchi e Krzyzanska, 2021), si rivela, in questo contesto, centrale per comprendere come la xenofobia operi discorsivamente negli ambienti digitali. A questo si aggiunge, negli anni più recenti, una crescente attenzione al ruolo delle campagne di disinformazione nel veicolare e radicalizzare atteggiamenti anti-immigrazione, in un circuito che lega fake news, polarizzazione politica e ostilità verso i migranti (Obreja, 2022).

La tesi è che l'hate speech xenofobo costituisca l'esito di una convergenza storica tra tre driver di mutamento sociale apparentemente concettualmente distinti che si sono reciprocamente alimentati. Il primo è la crisi del modello multiculturale: per almeno tre decenni, le democrazie occidentali hanno gestito la diversità culturale attraverso politiche di riconoscimento delle differenze; questo modello è entrato in crisi sul piano della legittimazione pubblica, creando uno spazio discorsivo in cui è divenuto legittimo proporre alternative identitarie (Kymlicka, 2010; Modood, 2007). Il secondo processo riguarda la riconfigurazione delle identità politiche attorno a nuove linee etno-nazionali: in questo vuoto si è inserita una proposta identitaria alternativa, che mobilita ansie demografiche, culturali e securitarie, e che ha conquistato partiti di governo in diversi paesi, ridefinendo il vocabolario politico mainstream (Mudde, 2010). Il terzo processo è la trasformazione tecnologica della sfera pubblica: i social media hanno modificato in modo sostanziale le condizioni di produzione e circolazione del discorso pubblico, attraverso la disintermediazione, la viralità, le filter bubbles e la presunzione di anonimato, offrendo all'hate speech xenofobo un ambiente particolarmente favorevole (Gagliardone *et al.*, 2015).

Questi tre processi non sono semplicemente paralleli, ma si alimentano reciprocamente: la crisi del multiculturalismo crea la domanda di discorsi alternativi; la riconfigurazione etno-nazionale fornisce l'offerta ideologica; le piattaforme digitali offrono l'infrastruttura di distribuzione. L'hate speech xenofobo online è il prodotto di questa triangolazione. Se questa tesi è cor-

ISBN Open Access: 9788835191971

retta, ne derivano conseguenze significative: gli interventi puramente tecnologici possono contenere il fenomeno ma non risolverlo; la xenofobia online non è separabile dalla xenofobia offline, ne è piuttosto un'intensificazione e un'accelerazione; comprendere il fenomeno richiede strumenti analitici che vengono da tradizioni diverse, dalla sociologia delle migrazioni agli studi sui nazionalismi, dai media studies all'analisi del discorso politico.

Il capitolo è strutturato come segue: il paragrafo successivo presenta un'analisi mirata dei trend topic emersi dai dati, volta a individuare i principali ambiti di interesse della letteratura scientifica e la loro evoluzione nel tempo; su questa base discuteremo poi le tre traiettorie delineate nell'introduzione, per arrivare infine a una sintesi conclusiva e alla formulazione di proposte orientate al contrasto delle narrative populiste di matrice xenofoba.

2. Trend topic della produzione scientifica internazionale

Dall'analisi bibliometrica della letteratura sull'hate speech xenofobo online emerge una struttura tematica articolata. I suoi sviluppi nel tempo rispecchiano le trasformazioni del fenomeno[1].

In base alle keyword selezionate dagli autori dei documenti contenuti nel corpus, emergono 16 temi che possono essere rappresentati dalle seguenti etichette: multiculturalism; islamophobia; nationalism; whiteness; race; far-right; gender; Facebook; right-wing populism; Twitter; hate speech; othering; disinformation; politics; Black Lives Matter; Europe.

Nella Tabella 1, la colonna "Frequenza" indica il numero di occorrenze di ciascun tema nel corpus. "Anno_Q1" rappresenta il primo quartile di comparsa, cioè l'anno in cui il tema si è manifestato per la prima volta nel campione di letteratura analizzato. "Anno_med" indica l'anno mediano di occorrenza, vale a dire il punto centrale della distribuzione delle frequenze del tema. Infine "Anno_Q3", il terzo quartile di occorrenza, segnala invece il periodo più recente in cui il tema è rimasto stabilmente presente nella letteratura. La Tabella 1 riporta i trend topic individuati disposti in ordine crescente per anno di prima occorrenza nel corpus. La Figura 1 rappresenta

[1] L'analisi bibliometrica è stata condotta utilizzando il pacchetto Bibliometrix (Aria e Cuccurullo, 2017) in ambiente R. I dati sono stati estratti da Scopus (19 novembre 2025) con la seguente query: ("conservatism" OR "populism*" OR "right*" OR "alt-right") AND ("racism" OR "xenophobia" OR "migration*" OR "migrant*") AND ("social media" OR "social network*" OR "online"). Limitando i risultati all'area "Social Sciences" e alla lingua inglese, sono stati esportati 863 documenti. Dopo la pulizia del dataset, che ha escluso 515 contributi non pertinenti, il corpus finale comprende 348 documenti pubblicati tra il 2002 e il 2025.

ISBN Open Access: 9788835191971

meglio graficamente i dati contenuti nella Tabella mostrando chiaramente la tendenza dei topic[2].

Tab. 1 – Trend topic

Topic	Frequenza	Anno_Q1	Anno_med	Anno_Q3
Multiculturalism	5	2012	2015	2016
Islamophobia	14	2016	2020	2022
Nationalism	12	2018	2020	2022
Whiteness	9	2018	2021	2022
Race	8	2019	2020	2022
Far-right	30	2019	2022	2024
Gender	9	2019	2024	2024
Facebook	19	2020	2021	2023
Right-wing populism	9	2020	2021	2023
Twitter	24	2020	2022	2024
Hate speech	16	2020	2022	2024
Othering	5	2021	2023	2023
Disinformation	7	2022	2023	2024
Politics	5	2022	2023	2024
Black Lives Matter	6	2022	2024	2025
Europe	8	2023	2024	2024

Analizzando i risultati dei trend topic, è possibile osservare da subito come il topic "multiculturalism", emerso nel 2012, risulti attivo nel corpus in esame, per l'ultima volta, nel 2016, anno di prima occorrenza di "islamophobia". Questa "staffetta" tra keyword ci permette di ipotizzare uno spostamento del focus dal tema della società multiculturale al più ristretto e specifico argomento dell'intolleranza nei riguardi della comunità musulmana. L'ipotesi appare significativa soprattutto se si considera che la letteratura sull'hate speech online in Germania risulta consistente: nel campione di pubblicazioni ricavate da Scopus ben 48 documenti derivano da università o enti di ricerca tedeschi, dopo gli Stati Uniti (158 documenti) e il Regno Unito (91 documenti). Nel 2015, infatti, l'allora Cancelliera Angela Merkel sospese temporaneamente il regolamento di Dublino per consentire l'ingresso in territorio tedesco di un elevato numero di rifugiati, soprattutto siriani; ma vale anche la pena menzionare come nello stesso anno l'Europa abbia assistito a diversi attentati terroristici, in un contesto in cui l'ISIS rappresentava una minaccia crescente (Simpson, 2016; Pickel e Yendell, 2016). D'altra parte, negli Stati Uniti Donald Trump annunciò la sua candidatura alle elezioni

[2] Il grafico rappresenta l'evoluzione temporale delle keyword mediante un diagramma a dispersione, con il tempo sull'asse orizzontale e i temi su quello verticale. Per ciascuna keyword, l'anno di riferimento è determinato dalla mediana delle occorrenze nel periodo analizzato (Aria e Cuccurullo, 2025). È stata utilizzata l'impostazione standard di Biblioshiny, identificando le keyword con una frequenza minima di 5 nel corpus e con 3 termini per ciascun anno.

ISBN Open Access: 9788835191971

presidenziali del 2016, promuovendo, sia in campagna elettorale che durante il mandato, una linea politica fortemente critica nei riguardi dell'immigrazione illegale e in particolare degli immigrati musulmani, accusati di incrementare il rischio di attacchi terroristici (Chavez *et al.*, 2023; Khan *et al.*, 2021).

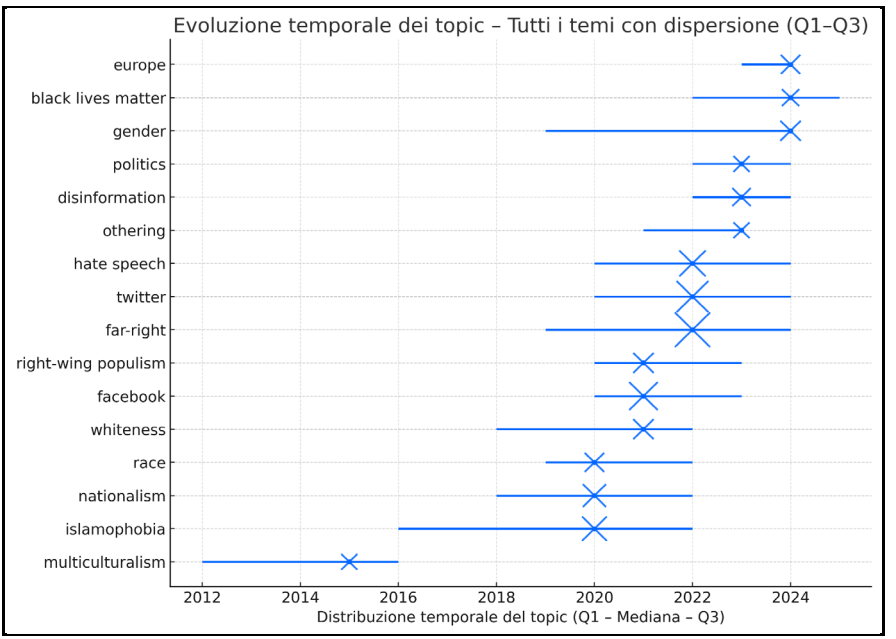


Fig. 1 - Evoluzione temporale dei topic

L'emergere simultaneo dei topic "nationalism" e "whiteness" nel 2018, in piena amministrazione Trump, non appare casuale: riflette l'intensificarsi del dibattito accademico sulla dimensione etno-razziale del populismo identitario statunitense e sulle sue ripercussioni nel discorso pubblico globale. Con "whiteness" si intende il costrutto identitario normativo rispetto al quale vengono definite tutte le altre identità etnico-culturali, una posizione strutturale di vantaggio razziale che si presenta come neutra e universale, rispetto alla quale le altre vengono definite come "altre" (Dyer, 1997). Negli ultimi anni, i movimenti nazionalisti bianchi hanno reinterpretato la whiteness in chiave biologica, utilizzando i test genetici ancestrali per rivendicare una presunta "purezza razziale" europea (Panofsky, Dasgupta e Iturriaga, 2021).

L'emersione del topic "far-right" nel 2019, accompagnato da "gender" e "race", sembra suggerire ulteriormente questa interpretazione: è interessante notare come l'estrema destra sembri di conseguenza raggruppare da un lato

ISBN Open Access: 9788835191971

il tema della razza e dall'altro quello del genere, spesso dipingendo il fenomeno migratorio in termini di minaccia alla sicurezza e alla tutela dei diritti delle donne (Adlung, Lünenborg e Raetzsch, 2021), anche nel contesto italiano (Lasio *et al.*, 2026).

Un maggiore focus su piattaforme social specifiche si ha a partire dal 2020, con "Facebook" e "Twitter" come topic a sé stanti e la presenza di "right-wing populism" e "hate speech". Si può ipotizzare che l'emergenza sanitaria costituita dal COVID-19 abbia favorito un maggiore utilizzo dei social media per ovviare alle misure di contenimento del virus, ma bisogna anche tenere in conto l'influenza degli eventi connessi al movimento Black Lives Matter (sebbene il relativo topic emerga solo nel 2022). In quest'ottica, Twitter e Facebook potrebbero essersi distinti come piattaforme privilegiate per l'osservazione di discorsi di odio razziale politicamente connotati.

L'emergere del topic "othering" nel 2021 riflette la crescente attenzione accademica verso i processi di costruzione antagonistica dell'alterità sui social media, attraverso i quali viene tracciato un confine simbolico tra "noi" e "loro", spesso declinato in chiave biologico-razziale (Bonacchi e Krzyzanska, 2021).

Altro aspetto rilevante è che nel 2022 emergono anche i topic "disinformation" e "politics", che suggeriscono una crescente attenzione da parte della letteratura in merito alla relazione fra le campagne di disinformazione online connesse a eventi e decisioni di natura politica e gli atteggiamenti xenofobi e contrari all'immigrazione: esempi sono la comunicazione social di Trump nella campagna elettorale del 2016 e il dibattito relativo alla Brexit (Obreja, 2022). Infine, lo specifico topic "Europe" emerso nel 2023 da un lato segna l'influenza di tendenze provenienti dal contesto americano, ma anche il riagganciarsi alla serie di crisi di vasta portata – come il cambiamento climatico, il COVID-19 e il conflitto russo-ucraino – che si sono affiancate alla gestione dei flussi migratori nell'ottica delle elezioni europee del 2024, in un clima di elevatissima polarizzazione dell'opinione pubblica (Haßler *et al.*, 2025).

Per comprendere meglio come questi trend topic si colleghino fra loro, la Figura 2 mostra il grafo ottenuto dall'analisi delle co-occorrenze che esamina la frequenza con cui le keyword selezionate dagli autori compaiono insieme all'interno degli stessi documenti. Questa tecnica permette di identificare cluster tematici e di visualizzare, attraverso rappresentazioni a rete, le connessioni strutturali tra i diversi filoni di ricerca[3].

[3] L'algoritmo di clusterizzazione utilizzato fra le alternative proposte da Biblioshiny è "Walktrap", selezionato per motivi relativi a una migliore visibilità e comprensione immediate del network.

ISBN Open Access: 9788835191971

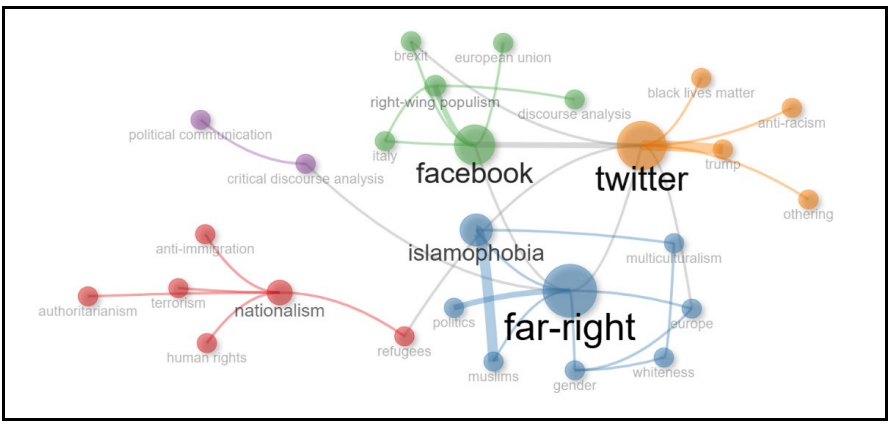


Fig. 2 - Network delle keyword degli autori

Si può notare dal grafo come i nodi raffiguranti i trend topic appena discussi siano affiancati da altre keyword che ne integrano il contesto e ne restituiscono una connotazione più dettagliata che va ad arricchire l'interpretazione precedente. È importante notare come nello stesso cluster siano raggruppati i trend topic "multiculturalism", "islamophobia", "far-right", "whiteness", "gender", "politics" ed "Europe", suggerendo una stretta relazione fra questioni identitarie, culturali e di genere e l'estrema destra. Rilevante è anche la presenza del nodo "nationalism" accanto a espressioni quali terrorismo e autoritarismo, richiamando possibili derive autoritarie e violente del fenomeno, ma anche la probabile connessione fra l'incremento degli atteggiamenti nazionalisti-identitari e la percezione della minaccia terroristica. "Right-wing populism" sembra essere particolarmente collegato a "Facebook", e la presenza di riferimenti alla Brexit e all'Unione Europea potrebbe indicare che su questa piattaforma tali manifestazioni populiste siano particolarmente osservabili. Twitter, invece, sembra per lo più ricollegabile al discorso statunitense, come suggerito dal collegamento con "Black Lives Matter" e "othering", ma anche dal riferimento a Trump che ha consistentemente utilizzato questa piattaforma per la propria comunicazione politica.

In sintesi, la successione cronologica dei topic delinea una traiettoria coerente: dal "multiculturalism" (2012-2016) all'"islamophobia", fino all'emergere congiunto di "far-right", "whiteness" e "nationalism" (2018-2019). Tale sequenza rispecchia il passaggio dalla crisi del paradigma multiculturale all'affermazione di narrative identitarie etno-nazionali. Al contempo, la comparsa di "Facebook" e "Twitter" come topic autonomi, ciascuno con specifiche connotazioni tematiche, evidenzia come le piattaforme digitali non costituiscano meri contenitori neutrali, ma operino come infrastrutture che modellano attivamente forme e traiettorie del discorso d'odio xenofobo.

ISBN Open Access: 9788835191971

3. Crisi del multiculturalismo e islamofobia

La crisi della società multiculturale ha rappresentato un'esperienza comune nel mondo occidentale. Precisamente, con multiculturalismo liberale si intende un approccio che mira a gestire una popolazione composta da gruppi culturali diversi garantendo equità, senza imporre l'assimilazione al gruppo maggioritario né limitarsi a una semplice coesistenza fra comunità. Per favorire il conseguimento di questo obiettivo, la tutela dei diritti delle minoranze immigrate e delle loro specificità culturali diventa cruciale: per questo motivo ci si orienta verso politiche volte, da un lato, a favorire l'espressione delle differenze culturali, dall'altro a compensare gli svantaggi dovuti alla condizione minoritaria (Eisenberg, 2025).

Nel 1971 il Canada ha inaugurato l'inizio di tali politiche nelle democrazie occidentali, ma a partire dalla metà degli anni novanta il supporto per il multiculturalismo ha subito una battuta d'arresto: sia per l'emergere di partiti populisti connessi alla destra, sia per le accuse, da parte della sinistra, di fallimento del multiculturalismo in quanto non si erano prese in adeguata considerazione tutte le componenti economiche e sociali necessarie a perseguirlo in modo efficace (Kymlicka, 2012).

In particolare, i discorsi di matrice identitaria, rafforzatisi soprattutto in correlazione all'aumento dei flussi migratori da paesi non occidentali, insieme ai dibattiti sulla tutela dei diritti femminili nelle comunità minoritarie tradizionali, alle accuse di intolleranza religiosa nei riguardi dei musulmani e al timore di attentati terroristici di matrice religiosa hanno comportato un progressivo ritiro delle democrazie occidentali dall'ideale di società multiculturale (Gozdecka, Ercan e Kmak, 2014).

L'attacco al World Trade Center di New York del 2001 e i successivi episodi in varie metropoli europee, come Londra e Parigi, hanno contribuito a un radicale mutamento dell'opinione pubblica nei confronti dell'Islam, con conseguente tendenza alla segregazione (Mikelatou e Arvanitis, 2019) e al rafforzamento della percezione di alterità nei confronti dei musulmani (Kaya, 2015).

Centrale è stato anche il ruolo dei media e della propaganda politica (Ramadhan *et al.*, 2025), soprattutto afferente al populismo di destra, che verrà discusso più nel dettaglio nel paragrafo successivo. In questa linea può essere fatta rientrare una serie di misure, sia in Europa che negli Stati Uniti, incardinate in una logica prettamente securitaria: per esempio, molti paesi europei hanno limitato o del tutto proibito l'utilizzo di capi d'abbigliamento che coprano il viso nei luoghi pubblici, sebbene l'applicazione della norma, di per sé neutrale, sembri indirizzarsi per lo più ai veli femminili tipicamente connessi all'Islam come il burqa e il niqab (Cox, 2022).

ISBN Open Access: 9788835191971

Negli Stati Uniti, la stretta sugli ingressi nel paese posta da Donald Trump è stata concretizzata anche per mezzo del cosiddetto Muslim Ban, inizialmente avviato durante il primo mandato presidenziale, nel 2017, e reintrodotto nel 2025. Si tratta di una serie di ordini esecutivi formalmente emanati per ragioni di sicurezza nazionale e per prevenire l'ingresso di terroristi, con l'obiettivo di bloccare per 90 giorni l'ingresso dei cittadini provenienti da sette paesi a maggioranza musulmana: Iran, Iraq, Libia, Somalia, Sudan, Siria e Yemen. L'ordine esecutivo 13769 ha anche sospeso illimitatamente l'accesso al paese dei rifugiati siriani (Mohn, 2023).

Nella versione del 4 giugno 2025, invece, la portata delle restrizioni si è indirizzata a un più elevato numero di paesi, diciannove in totale. Dodici di questi, e precisamente Afghanistan, Birmania, Ciad, Repubblica del Congo, Guinea Equatoriale, Eritrea, Haiti, Iran, Libia, Somalia, Sudan e Yemen, sono stati sottoposti a un divieto totale di ingresso, mentre altri sette, ossia Burundi, Cuba, Laos, Sierra Leone, Togo, Turkmenistan e Venezuela, hanno subito restrizioni parziali (Wilson, 2025). È importante rilevare come, sebbene nella versione attuale la portata securitaria del travel ban si sia espansa anche a Paesi percepiti come ostili agli Stati Uniti e non più a soli paesi a maggioranza musulmana, questi ultimi costituiscano comunque la maggior parte di quelli soggetti a restrizioni, sempre per motivi connessi ufficialmente alla tutela della sicurezza nazionale.

4. Populismo e identitarismo

Come indicato nel paragrafo introduttivo e accennato in quello precedente, la reazione alla società multiculturale può essere inscritta nella più ampia rilevanza acquisita da controversie essenzialmente riconducibili a una natura culturale e identitaria.

La tendenza a interpretare il conflitto in questa prospettiva affonda le proprie radici nel contesto successivo alla fine della Guerra fredda: con il prevalere del modello liberale occidentale, e in particolare statunitense, su quello sovietico, i conflitti hanno perso la loro natura ideologica. Le loro motivazioni hanno piuttosto iniziato a essere ricollegate a questioni di matrice etnico-culturale, fino a quel momento apparentemente sopite dallo scontro fra le ideologie dei due blocchi (Huntington, 1996).

In questo contesto, la globalizzazione ha rivestito un ruolo fondamentale per via dell'incremento dei flussi migratori diretti verso l'Europa e gli Stati Uniti. Come riportato dal report 2025 sulle migrazioni internazionali dell'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE, 2025), nel 2024, nonostante un declino del 4% rispetto all'aumento comportato dal

ISBN Open Access: 9788835191971

COVID-19 nei tre anni precedenti, la componente di nuova popolazione con background migratorio permanentemente residente nei paesi OCSE si è collocata ai massimi storici, 6,2 milioni, cioè il 15% in più rispetto al 2019; anche i flussi temporanei per motivi lavorativi sono stati i più elevati mai registrati, del 26% più alti rispetto al 2019. Un dato significativo è che il 23% degli ingressi è costituito da migrazioni umanitarie, con più della metà dei tre milioni di domande di asilo presentate nei paesi OCSE nel 2024 registrata negli Stati Uniti, principalmente da richiedenti provenienti da Venezuela, Colombia, Siria, Afghanistan e India.

Le sfide connesse a questo nuovo assetto, alle sue proporzioni e alle circostanze geopolitiche che contribuiscono a motivarlo hanno favorito l'emergere di movimenti populisti, spesso connessi all'estrema destra, accomunati da un approccio nazionalista e contrario agli ingressi nei propri confini territoriali. Si tratta infatti di una prospettiva fortemente radicata in un orientamento appunto emergente da istanze di matrice culturale e identitaria, in quanto una società multiculturale è percepita, oltre che come un pericolo sul piano della sicurezza interna e per i cittadini del Paese, anche come una minaccia concreta al tessuto sociale e culturale dei Paesi ospitanti, la cui omogeneità risulterebbe intaccata da culture percepite come assolutamente aliene ed, eventualmente, prima o poi sostituita visto l'elevato numero di arrivi.

Questo atteggiamento costituisce un tratto fondamentale e distintivo degli attuali populismi, ossia il dichiarato anti-internazionalismo. I processi di globalizzazione sono infatti percepiti come causa della società multiculturale e, di conseguenza, dell'erosione dell'unicità culturale del popolo. Ne consegue dunque anche il rifiuto delle istituzioni sovranazionali, accusate di favorire questa minaccia alle singole identità nazionali (Mudde e Rovira Kaltwasser, 2017). Si tratta di un approccio che discende da un altro tratto distintivo dei populismi descritti dai due autori, cioè il concetto di popolo come unica entità morale espressione di un bene comune collettivo e intrinsecamente in grado di distinguere fra ciò che è bene e ciò che è male, proprio in virtù di tali caratteristiche culturali uniche che acquisiscono quindi una vera e propria connotazione mitologica. Da ciò deriva dunque un rifiuto di questi movimenti populisti e anti-internazionalisti di collocarsi nel continuum politico destra-sinistra e la loro tendenza a riferirsi a leader carismatici: in quanto espressione del popolo e della sua volontà, si vuole infatti evitare la creazione di divisioni che andrebbero a danneggiare questa comunità unica.

I movimenti populisti contemporanei possono infatti essere letti nella prospettiva della reazione culturale descritta da Norris e Inglehart (2019), in base alla quale l'identità e i valori tradizionali vengono mobilitati in risposta a una divisione culturale e generazionale legata ai cambiamenti nelle norme sociali. In questo quadro, il populismo emerge come una forma di resistenza

ISBN Open Access: 9788835191971

da parte di gruppi che si sentono minacciati dall'evoluzione dei valori moderni, cercando di riaffermare valori considerati tradizionali e identitari contro idee percepite come progressiste o globaliste e ricollegandosi pertanto a un generale rifiuto dei fenomeni e delle istituzioni connesse ai fenomeni della globalizzazione.

Anche il timore di una sostituzione etnico-culturale può essere letto secondo una teoria, quella, già accennata nel paragrafo introduttivo, del cosiddetto great replacement (Camus, 2011). Secondo i suoi sostenitori, la popolazione nativa degli Stati Uniti, e adesso anche dell'Europa, finirà per essere sostituita dagli immigrati di prima e seconda generazione provenienti da paesi esteri, in particolare da quelli ispanici per gli States e da quelli a maggioranza musulmana per l'Europa. Si tratta dunque di una teoria intimamente connessa alle istanze del suprematismo bianco e motivata dagli alti tassi di natalità registrati fra la popolazione immigrata (Sedgwick, 2024). Altra fondamentale caratteristica è che questo approccio ricade nella narrazione, tipicamente populista, della contrapposizione fra élites corrotte e onesti cittadini (Chapelan, 2020): il great replacement non sarebbe infatti che il prodotto di interessi economici e politici delle élite liberali globali, che, anche attraverso politiche a supporto della comunità LGBT, della parità di genere e di diritti quali l'aborto, avrebbero come obiettivo la riduzione dei tassi di natalità e l'indebolimento della razza bianca in vista della sua sostituzione (Dennison e Kustov, 2025).

Questo generale impianto di pensiero è supportato significativamente dall'alt-right statunitense (in seguito diffusasi anche in Europa), che, emersa online fin dai primi anni Duemila con l'esplicito obiettivo di riportare il tema della razza all'interno del dibattito politico del paese, ha fornito un'espressione politica alle correnti connesse al nazionalismo e al suprematismo bianco e significativo supporto elettorale per Trump (Banda e Cluverius, 2023). Il cosiddetto "genocidio bianco" è difatti un fondamentale elemento ideologico dell'alt-right, proprio in riferimento al timore della minaccia esistenziale per i bianchi rappresentata dai flussi migratori in arrivo da paesi non occidentali, in aperta opposizione al multiculturalismo (Deem, 2019). Data la componente demografica prevalentemente giovane, si nota come i social media e le piattaforme online costituiscano un veicolo cruciale per la diffusione delle idee connesse a questo orientamento politico (Colley e Moore, 2020).

5. L'othering come dispositivo: dalle radici sociologiche alla strumentalizzazione populista negli ambienti digitali

L'emergere del topic "othering" nell'analisi bibliometrica (Fig. 1), in stretta connessione con il discorso politico statunitense (Fig. 2), segnala la

ISBN Open Access: 9788835191971

centralità di questo concetto per comprendere le dinamiche discorsive della xenofobia online. Con il termine "othering" s'intende la costruzione simbolica di un confine tra "noi" e "loro" che trasforma la differenza in estraneità e l'estraneità in minaccia (Bonacchi e Krzyżanowska, 2021). Il processo non riguarda soltanto la rappresentazione dell'altro, ma investe simultaneamente la definizione del sé: l'identità del gruppo si costituisce e si rafforza proprio attraverso la demarcazione rispetto a ciò che viene posto come radicalmente diverso. Considerare la circolarità tra costruzione dell'alterità e affermazione identitaria è particolarmente calzante quando si analizzano le narrative populiste contemporanee, dove la figura del migrante diviene il termine di contrasto privilegiato per definire i confini della comunità nazionale. In tale ambito la riflessione delle scienze sociali sulla figura dello straniero offre gli strumenti concettuali per comprendere le radici di questa dinamica.

5.1. Lo straniero, l'ordine sociale e il capro espiatorio

La figura dello Straniero teorizzata da Georg Simmel (1908), definito come colui "che oggi viene e domani rimane", definisce una sintesi paradossale di vicinanza e lontananza, appartenenza ed estraneità, che incarna l'ambivalenza costitutiva di ogni relazione con l'altro. Lo straniero simmeliano non è semplicemente chi proviene da altrove, ma rappresenta una posizione strutturale all'interno del gruppo, quella di chi, pur essendo presente, non condivide pienamente le qualità particolari che definiscono l'appartenenza. Questa tensione tra prossimità spaziale e distanza sociale genera una forma peculiare di oggettività che, se da un lato può risultare funzionale al gruppo, dall'altro espone lo straniero a divenire bersaglio privilegiato di proiezioni ostili nei momenti di crisi. In termini antropologici l'incarnazione del cosiddetto "capro espiatorio".

Alfred Schütz (1944) approfondisce questa dinamica in chiave fenomenologica. Secondo l'autore lo straniero, per il gruppo ospitante, rappresenta un'anomalia categoriale che mette in discussione la naturalità dell'ordine sociale. Norbert Elias e John Scotson (1965), nel loro studio sulla comunità di Winston Parva, dimostrano come la distinzione tra *established* e *outsiders* operi indipendentemente da differenze etniche o di classe, attraverso meccanismi di *group charisma* e *group disgrace* che attribuiscono ai nuovi arrivati le caratteristiche della "minoranza anomica peggiore". Con questa espressione, gli autori designano quella frazione marginale di un gruppo che non si conforma alle norme sociali condivise dalla comunità, individui i cui comportamenti vengono percepiti come disordinati, devianti o minacciosi per l'ordine costituito. Il meccanismo di stigmatizzazione opera attraverso una duplice distorsione: il

ISBN Open Access: 9788835191971

gruppo dominante costruisce la propria immagine collettiva sulla base dei suoi membri più esemplari, mentre proietta sull'intero gruppo subordinato i tratti negativi della sua componente più marginale. In questo modo, le caratteristiche di pochi vengono generalizzate a tutti, legittimando l'esclusione e la squalificazione simbolica dell'intero gruppo *outsider*. Arrivando ad autori contemporanei, è Zygmunt Bauman (1991, 2016) a sistematizzare queste intuizioni in una teoria dell'alterità radicale. Lo straniero baumaniano è l'inclassificabile, colui che sfida le tassonomie attraverso cui l'ordine moderno organizza il mondo: "né amico né nemico, né vicino né lontano, né membro né estraneo". Questa indeterminatezza categoriale genera un'"incongruenza cognitiva" che minaccia la stabilità simbolica dell'ordine sociale, suscitando reazioni che oscillano tra assimilazione forzata ed esclusione violenta. Nei suoi lavori più recenti, Bauman (2016) ha analizzato come nel contesto contemporaneo delle migrazioni globali questa dinamica si traduca in una securitizzazione dello straniero: il migrante viene costruito discorsivamente come minaccia esistenziale, legittimando politiche di esclusione sempre più radicali.

La dimensione biopolitica di questi processi emerge con particolare chiarezza nell'opera di Michel Foucault. Nel corso al Collège de France del 1975-76, *Bisogna difendere la società*, Foucault (2009) analizza la trasformazione storica del potere sovrano: il diritto di "far morire o lasciar vivere" in un potere biopolitico orientato alla gestione della vita delle persone. È in questo quadro che introduce il concetto di "razzismo di Stato": un dispositivo che opera una cesura nel continuum biologico della popolazione, distinguendo "ciò che deve vivere e ciò che deve morire". Lungi dall'essere un residuo arcaico, il razzismo diviene, secondo tale prospettiva teorica, una tecnologia moderna di governo, funzionale a legittimare l'esclusione di alcuni segmenti della popolazione in nome della salute e della sicurezza del corpo sociale. Questa visione biopolitica non si esaurisce nell'eliminazione fisica, ma opera anche attraverso forme di "morte indiretta": l'esposizione sistematica al rischio, la relegazione ai margini della società, la sottrazione di diritti e risorse. È qui che l'analisi foucaultiana illumina i meccanismi più profondi dell'othering contemporaneo: la costruzione discorsiva dello straniero come minaccia non mira semplicemente a tracciare un confine simbolico tra "noi" e "loro", ma a giustificare un trattamento differenziale che può tradursi in esclusione giuridica, precarizzazione economica o abbandono istituzionale. La figura dello straniero-nemico internalizza questo dispositivo: non più il nemico esterno delle società di sovranità, ma l'elemento contaminante che minaccia dall'interno l'integrità del corpo sociale e che, proprio per questo, deve essere isolato, neutralizzato, espulso.

In questa prospettiva, l'hate speech xenofobo sui social media può essere letto come una forma di micro-biopolitica diffusa: un discorso che, nel

ISBN Open Access: 9788835191971

designare ripetutamente determinati gruppi come pericolosi, parassitari o incompatibili, contribuisce a legittimare la loro esclusione dalla sfera dei diritti e della protezione sociale.

Stuart Hall (1997), nel capitolo "The Spectacle of the 'Other'", analizza lo stereotipo come pratica significante centrale nei processi di costruzione dell'alterità. Lo stereotipo opera come esercizio di potere simbolico: riduce la complessità dell'altro a pochi tratti semplificati, li essenzializza e li naturalizza, fissandoli come caratteristiche immutabili. Hall individua due strategie fondamentali attraverso cui lo stereotipo agisce: la *scissione* (*splitting*), che separa il normale dall'anormale, l'accettabile dall'inaccettabile; e la *chiusura* (*closure*), che traccia confini simbolici ed esclude tutto ciò che non si conforma all'ordine dominante, relegando l'altro in una posizione di subordinazione.

René Girard (1982) aggiunge a questo quadro la teoria del già citato "capro espiatorio", affermando che, in situazioni di crisi sociale, la violenza diffusa viene canalizzata su una vittima designata, accusata di essere responsabile del disordine. Il capro espiatorio è tipicamente un individuo o gruppo marginale, riconoscibile per "segni vittimari" che lo distinguono dalla comunità: lo straniero, l'eretico, il diverso. Il suo sacrificio restaura simbolicamente l'unità del gruppo, trasformando la violenza mimetica in violenza fondatrice. Questa dinamica arcaica, lungi dall'essere superata dalla modernità, si ripropone nelle forme secolarizzate della stigmatizzazione e della persecuzione dei gruppi minoritari.

5.2. La trasformazione digitale: online othering e dispositivi algoritmici

L'avvento delle piattaforme digitali ha amplificato le dinamiche dell'othering ma senza produrre, in sostanza, una discontinuità radicale rispetto ai meccanismi classici individuati dalla teoria psicoanalitica e postcoloniale. La struttura fondamentale di questi processi affonda le radici nella dialettica hegeliana del riconoscimento (Hegel, 1807), che Kojève (1947) ha riletto in chiave antropologica e Sartre (1943) ha tradotto nella fenomenologia dello sguardo come forma di oggettivazione dell'Altro. Muovendo da questo quadro, de Beauvoir (1949) ha mostrato come la donna sia concepita come Altro nel dominio patriarcale, mentre Fanon (1952) ha declinato la stessa dinamica nel contesto coloniale, dove lo sguardo del colonizzatore fissa il soggetto nero in un'identità alienata. Lacan (1966) ha conferito a questa dinamica una formalizzazione simbolica, distinguendo tra il piccolo altro immaginario e il "grande Altro" (grand Autre) come ordine del linguaggio che precede e struttura la soggettività, e mostrando così come l'accesso stesso alla parola e al desiderio passi necessariamente attraverso questa alterità costitutiva.

ISBN Open Access: 9788835191971

L'analisi di Said (1978) ha poi esteso lo schema alla costruzione discorsiva dell'Oriente, confermandone la portata euristica anche sul piano culturale e geopolitico. L'altro è quindi strumento essenziale nella definizione dell'identità individuale e collettiva, strumento di ordine sociale.

Karen Lumsden ed Emily Harmer (2019), nel volume collettaneo *Online Othering*, propongono questo concetto come chiave interpretativa per analizzare come le tecnologie dell'informazione e della comunicazione facilitino e/o esacerbino le offese che avvengono offline, problematizzando al contempo la dicotomia tra reale e virtuale che ha a lungo caratterizzato gli studi sui media digitali. L'othering online non costituisce un fenomeno separato dalle dinamiche sociali "offline", ma ne rappresenta un'estensione, un'intensificazione.

Un contributo significativo all'analisi di queste dinamiche viene dalla teoria dell'Othering Online Discourse (OOD), sviluppata a partire dall'analisi del forum finlandese Suomi24 (Vaahensalo, 2021). Questo approccio distingue tra othering interno, rivolto verso gruppi minoritari già presenti nella società, e othering esterno, orientato verso popolazioni straniere percepite come potenziali invasori. L'analisi enfatizza l'aspetto sociale e relazionale dei discorsi di esclusione: l'othering online non è mai un atto puramente individuale, ma si sviluppa attraverso dinamiche di gruppo, rinforzo reciproco, costruzione collaborativa di narrative stigmatizzanti. Su questa stessa linea, Silvia Bonacchi e Marta Krzyżańska (2021) analizzano i discorsi su ascendenza e patrimonio genetico in Twitter, introducendo il concetto di heritage-based tribalism, una forma di othering antagonistico che sviluppa narrazioni sull'origine e l'ancestralità per costruire confini esclusivi di appartenenza. Lo studio identifica un assemblaggio complesso di concezioni biologizzanti della razza, sfiducia verso l'expertise scientifica e orientamento politico che si intrecciano in configurazioni narrative sull'altro, mostrando come l'othering digitale operi attraverso l'articolazione di elementi eterogenei integrati in modo dinamico piuttosto che attraverso ideologie monolitiche e statiche.

La specificità delle piattaforme emerge con particolare chiarezza se si analizzano delle dinamiche conflittuali che si sviluppano nelle pieghe dei *reply to comment* (nel caso ad esempio di FB). Orian Harel *et al.* (2020), applicando il modello del conflitto intrattabile di Northrup alle interazioni su Facebook, identificano tre fasi nella spirale dell'othering: la minaccia (percezione dell'altro come pericolo per la propria identità), la distorsione (costruzione di rappresentazioni stereotipate e demonizzanti) e la rigidificazione (cristallizzazione di posizioni deumanizzanti che rendono impossibile il dialogo). Le affordance delle piattaforme, la velocità di diffusione, la possibilità di aggregazione identitaria, i meccanismi di profilazione e rinforzo algoritmico accelerano e intensificano questi processi. Tale meccanismo di

ISBN Open Access: 9788835191971

escalation è dimostrato dallo studio sistematico sull'hate speech online di Paasch-Colberg *et al.* (2021), i quali, analizzando commenti di utenti su siti di news tedeschi, blog, pagine Facebook e canali YouTube nel contesto del dibattito sull'immigrazione, distinguono cinque modalità di discorso d'odio: *racist othering*, *racist criminalization*, *dehumanization*, *raging hate* e *call for hate crimes.* Questa tipologizzazione mostra come l'othering costituisca la base su cui si costruiscono forme progressivamente più intense di violenza simbolica, fino all'incitamento esplicito alla violenza fisica.

5.3. Piattaforme social, affordance e geografie dell'odio: Facebook, Twitter/X e la differenziazione del discorso xenofobo

A partire dal 2020 (Fig. 1), la letteratura scientifica inizia a interrogarsi sistematicamente sul ruolo delle singole piattaforme social come ambienti specifici in cui il discorso xenofobo assume forme e dinamiche peculiari. Questo spostamento di focus, che nell'indagine bibliometrica si manifesta con l'emergere di "Facebook" e "Twitter" come topic autonomi, affiancati da "right-wing populism" e "hate speech" (Fig. 2), riflette una crescente consapevolezza del fatto che le piattaforme digitali non costituiscono un medium neutro, ma modellano attivamente i contenuti che ospitano attraverso le proprie affordance tecnologiche. Gli algoritmi di raccomandazione, progettati per massimizzare la partecipazione degli utenti, tendono sistematicamente a favorire contenuti polarizzanti ed emotivamente carichi, creando filter bubbles (Pariser, 2011) ed echo chambers che intensificano l'esposizione a discorsi di othering e riducono le occasioni di imbattersi in narrative diverse. Come noto, l'anonimato, o la pseudonimia, rimuove alcune delle barriere sociali che normalmente inibiscono l'espressione di pregiudizi espliciti, mentre la viralità consente la rapida diffusione di contenuti stigmatizzanti. Si assiste così a un'amplificazione della spirale del silenzio teorizzata da Noelle-Neumann (1974): chi percepisce le proprie opinioni come minoritarie tende a tacere, mentre chi si riconosce in posizioni dominanti si esprime con crescente sicurezza. Nei contesti digitali, questo meccanismo opera non solo rispetto all'opinione pubblica generale, ma anche all'interno delle singole echo chambers, dove la percezione di consenso locale può rafforzare ulteriormente l'espressione di posizioni altrimenti marginali nel dibattito mainstream (Hampton *et al.*, 2014).

Questa configurazione può essere concettualizzata, in termini foucaultiani, come un dispositivo digitale dell'othering: un assemblaggio eterogeneo di algoritmi, interfacce, politiche di moderazione, culture di piattaforma, pratiche degli utenti e modelli di business che produce sistematicamente effetti

ISBN Open Access: 9788835191971

di esclusione, stigmatizzazione e disumanizzazione dell'altro generalizzato. Il dispositivo non opera attraverso un'intenzionalità centralizzata, ma attraverso l'articolazione di elementi disparati che convergono nel produrre effetti di potere: determinare chi può parlare e chi deve tacere, chi è visibile e chi invisibile, quali voci vengono amplificate e quali silenziate dal gruppo. Le pratiche di moderazione dei contenuti, lo shadowban, la sospensione degli account funzionano come nuove forme di censura biopolitica, tracciando confini tra discorso legittimo e illegittimo, presenza e assenza nello spazio pubblico digitale.

Volendo analizzare più in profondità il ruolo delle piattaforme social, è utile distinguere come le loro caratteristiche tecnologiche influenzano l'othering. Un dato significativo che emerge dalla letteratura è la differenziazione geografica e tematica tra le principali piattaforme (Fig. 2). Facebook appare maggiormente connesso al populismo di destra europeo e, in particolare, al dibattito sulla Brexit. Studi condotti sulle pagine pro-Leave attive durante e dopo il referendum del 2016 hanno mostrato come la piattaforma abbia funzionato da catalizzatore per la mobilitazione di identità politiche antagoniste, creando "universi di contenuto" distinti e polarizzati. Hall (2023), nella sua ricerca etnografica con utenti pro-Brexit su Facebook, ha documentato come l'engagement prolungato con gruppi e pagine della piattaforma abbia contribuito a consolidare identità politiche transnazionali di matrice populista.

Più in generale, Facebook si è rivelato un ambiente particolarmente ospitale per la comunicazione dei partiti populisti di destra europei. Questi partiti utilizzano i social media in modo strategico per aggirare i gatekeepers dei media tradizionali e comunicare direttamente con i propri elettori (Schroeder, 2019), costruendo reti di pagine Facebook che funzionano come ecosistemi propagandistici autonomi (Russo e Maretti, 2025). Uno studio su larga scala (Törnberg e Chueri, 2025), basato sull'analisi di 32 milioni di tweet di parlamentari in 26 paesi, ha evidenziato che la diffusione di disinformazione non è associata al populismo in quanto tale, né all'orientamento di destra in generale, ma specificamente al populismo radicale di destra – suggerendo che la disinformazione costituisca una strategia comunicativa distintiva di questa famiglia politica.

Twitter/X rappresenta un caso emblematico di trasformazione politica di una piattaforma. Nella sua prima fase, fino al 2022, il social network si distingueva per politiche di moderazione relativamente rigorose che, pur non impedendo il confronto aspro, consentivano la coesistenza di movimenti e contromovimenti sulla stessa arena digitale: è il caso del movimento Black Lives Matter e delle contromobilitazioni #AllLivesMatter e #BlueLivesMatter, che si confrontarono apertamente sulla piattaforma (Tillery, 2019; Ince, Rojas e Davis, 2017; Goodman, Perkins e Windel, 2024).

ISBN Open Access: 9788835191971

Con l'acquisizione da parte di Elon Musk della piattaforma, questo equilibrio tra posizioni diverse si è notevolmente ridotto: la rimozione dei vincoli alla moderazione e la riconfigurazione dell'algoritmo (Corsi, 2024) hanno prodotto un vero e proprio rovesciamento dell'ecosistema della piattaforma. I dati del Pew Research Center (Faverio e Anderson, 2025) documentano un rilevante ribaltamento delle percezioni partisan: la quota di utenti repubblicani che considera X "buono per la democrazia" è triplicata dal 17% (2021) al 58% (2025), mentre quella dei democratici è crollata dal 47% al 17%, mentre il 55% degli utenti democratici ritiene oggi che la piattaforma favorisca sistematicamente le posizioni conservative.

6. Conclusioni

L'hate speech xenofobo online non può essere considerato un semplice prodotto della tecnologia, bensì l'esito di una convergenza storica tra processi distinti che si sono reciprocamente alimentati. La crisi del modello multiculturale, documentata dalla scomparsa del topic "multiculturalism" e dalla sua sostituzione con "islamophobia", ha difatti creato lo spazio discorsivo per narrazioni alternative, supportate da una mobilitazione in chiave esplicitamente identitaria della destra radicale: l'emersione simultanea di "nationalism", "whiteness" e "far-right" costituisce infatti una rilevante conferma della centralità di istanze di natura etnico-culturale, da inquadrare entro la più ampia prospettiva di una reazione a cambiamenti sociali percepiti come troppo rapidi. Le piattaforme digitali, con le loro affordance specifiche e i loro algoritmi di raccomandazione, hanno offerto un'ideale infrastruttura di distribuzione di queste idee, in particolare della costruzione dell'alterità "othering".

La costruzione simbolica dell'alterità, radicata nella tradizione socio antropologica, da Simmel a Bauman, da Schütz a Girard, assume infatti negli ambienti online forme peculiari, amplificate dalla logica algoritmica e dalle dinamiche di gruppo che caratterizzano le echo chambers. È inoltre significativo rilevare come l'othering digitale non sostituisca quello offline, ma piuttosto lo intensifichi e lo acceleri.

Questa triangolazione tra crisi del multiculturalismo, mobilitazione identitaria e infrastrutture digitali produce conseguenze significative per la comprensione e il contrasto del fenomeno. In primo luogo, sebbene gli interventi puramente tecnologici, dalla moderazione dei contenuti alla rimozione degli account, così come tutte le strategie di online counter speech, possano contenere le manifestazioni più estreme, non risolvono il problema alla radice, in quanto esso affonda in dinamiche sociali, culturali e politiche più profonde.

ISBN Open Access: 9788835191971

In secondo luogo, la xenofobia online non è separabile dalla xenofobia offline: come già evidenziato, ne costituisce piuttosto un'intensificazione e un'accelerazione, grazie alle dinamiche di viralità, anonimato e rinforzo algoritmico proprie degli ambienti digitali.

In terzo luogo, comprendere la complessità del fenomeno richiede strumenti analitici provenienti da tradizioni disciplinari diverse: dalla sociologia delle migrazioni agli studi sui nazionalismi, dai media studies all'analisi del discorso politico, dalla teoria critica all'analisi delle piattaforme.

La differenziazione emersa tra Facebook e Twitter/X suggerisce inoltre che le piattaforme non sono intercambiabili, ma costituiscono ambienti con affordance, culture e geografie politiche specifiche. La trasformazione di Twitter/X dopo l'acquisizione da parte di Elon Musk rappresenta un caso emblematico di come le scelte di governance delle piattaforme possano alterare radicalmente l'ecosistema discorsivo, con effetti che travalicano i confini della singola piattaforma per investire la sfera pubblica nel suo complesso.

In ultima analisi, l'hate speech xenofobo online si configura come un dispositivo in senso foucaultiano: un assemblaggio eterogeneo di discorsi, pratiche, tecnologie e istituzioni che produce effetti di potere. Contrastarlo efficacemente richiede quindi un approccio altrettanto multilivello, capace di intervenire simultaneamente sulle condizioni sociali che alimentano la domanda di capri espiatori, sulle offerte ideologiche che la intercettano e sulle infrastrutture che ne amplificano la circolazione.

Riferimenti bibliografici

Adlung, S., Lünenborg, M. and Raetzsch, C. (2021), Pitching Gender in a Racist Tune: The Affective Publics of the #120decibel Campaign, *Media and Communication*, 9, 2: 16-26. DOI: https://doi.org/10.17645/mac.v9i2.3749.

Aria, M. and Cuccurullo, C. (2017), bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11, 4: 959-975. DOI: https://doi.org/10.1016/j.joi. 2017.08.007.

Aria, M. and Cuccurullo, C. (2025), Focus on Domain. 4 Levels of analysis, *Bibliometrix*, testo disponibile al sito: https://bibliometrix.org/biblioshiny/biblioshiny2.html (consultato in data 19/12/2025).

Banda, K.K. and Cluverius, J. (2023), White Americans' Evaluations of the Alt-Right, *American Politics Research*, 51, 4: 435-442. DOI: https://doi.org/10.1177/1532673X231157398.

Bauman, Z. (1991), *Modernity and Ambivalence*, Polity Press, Cambridge.

Bauman, Z. (2016), *Strangers at Our Door*, Polity Press, Cambridge.

Bonacchi, C. and Krzyzanska, M. (2021), Heritage-based tribalism in Big Data ecologies: Deploying origin myths for antagonistic othering, *Big Data & Society*, 8,

ISBN Open Access: 9788835191971

1: 1-16. DOI: https://doi.org/10.1177/20539517211003310.
Bonilla-Silva, E. (2013), *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*, Rowman & Littlefield, Lanham, 4th ed.
Camus, R. (2011), *Le Grand Remplacement*, David Reinharc, Paris.
Castells, M. (1996), *The Rise of the Network Society*, Blackwell Publishers Ltd, Oxford.
Castells, M. (2009), *Communication Power*, Oxford University Press, Oxford.
Chapelan, A. (2020), Populist conservatism or conservative populism? The Republican Party, the new Conservative Revolution and the political education of Donald Trump, *Perspective Politice*, 13, 1-2: 8-19.
Colley, T.P. and Moore, M. (2020), The challenges of studying 4chan and the Alt-Right: 'Come on in the water's fine', *New Media and Society*, 1-26. DOI: https://doi.org/ 10.1177/1461444820948803.
Corsi, G. (2024), Evaluating Twitter's Algorithmic Amplification of Low-Credibility Content: An Observational Study, *EPJ Data Science*, 13: 18. DOI: https://doi.org/10.1140/ epjds/s13688-024-00456-3.
Cox, N. (2022), Pejorative Assertions, Human Rights Evaluation, and European Veiling Laws, *The American Journal of Comparative Law*, 70, 4: 695-735. DOI: https://doi.org/ 10.1093/ajcl/avad012.
de Beauvoir, S. (1949), *Le deuxième sexe* (2 voll.), Gallimard, Paris.
Deem, A. (2019), Extreme Speech| The Digital Traces of #whitegenocide and Alt-Right Affective Economies of Transgression, *International Journal of Communication*, 13: 3183-3202.
Dennison, J. and Kustov, A. (2025), Public Belief in the "Great Replacement Theory", *International Migration Review*, 0, 0.
DOI: https://doi.org/10.1177/01979183251343877.
Dyer, R. (1997), *White: Essays on Race and Culture,* Routledge, London.
Eisenberg, A. (2025), Integration Before Multiculturalism, *Nations and Nationalism*. DOI: https://doi.org/10.1111/nana.13129.
Elias, N. and Scotson, J.L. (1965), *The Established and the Outsiders: A Sociological Enquiry into Community Problems*, Frank Cass, London.
Fanon, F. (1952), *Peau noire, masques blancs,* Éditions du Seuil, Paris.
Faverio, M. and Anderson, M. (2025), Republicans and Democrats on X differ over the site's politics and their experiences, *Pew Research Center*, Washington, DC, testo disponibile al sito: https://www.pewresearch.org/short-reads/2025/06/05/republicans-and-democrats-on-x-differ-over-the-sites-politics-and-their-experiences/ (consultato in data 16/12/2025).
Foucault, M. (2009), *Bisogna difendere la società. Corso al Collège de France 1975-1976*, Feltrinelli, Milano, ed. or. 1997.
Gagliardone, I., Gal, D., Alves, T. and Martinez, G. (2015), *Countering Online Hate Speech*, UNESCO Publishing, Paris.
Girard, R. (1982), *Le bouc émissaire*, Grasset, Paris.
Goodman, S., Perkins, K.M. and Windel, F. (2024), All Lives Matter discussions on Twitter: Varied use, prevalence, and interpretive repertoires, *Journal of Community & Applied Social Psychology*, 34, 2: e2767.

ISBN Open Access: 9788835191971

DOI: https://doi.org/10.1002/casp.2767.
Gozdecka, D.A., Ercan, S.A. and Kmak, M. (2014), From multiculturalism to post-multiculturalism: Trends and paradoxes, *Journal of Sociology*, 50, 1: 51-64. DOI: https://doi.org/10.1177/1440783314522191.
Haßler, J., Magin, M., Russmann, U., Wurst, A., Balaban, D., Baranowski, P., Jensen, J., Kruschinski, S., Lappas, G., Machado, S., Novotná, M., Marcos-García, S., Petridis, I., Rožukalne, A., Sebestyén, A. and von Nostitz, F. (2025), Weaponizing Wedge Issues: Strategies of Populism and Illiberalism in European Election Campaigning on Facebook, *Media and Communication*, 13: 10718. DOI: https://doi.org/10.17645/mac.10718.
Hall, N.A. (2021), Understanding Brexit on Facebook: Developing Close-up, Qualitative Methodologies for Social Media Research, *Sociological Research Online*, 27, 3: 707-723. DOI: https://doi.org/10.1177/13607804211037356.
Hall, N.A. (2023), *Brexit, Facebook, and transnational right-wing populism*, Bloomsbury Publishing, New York.
Hall, S. (1997), The Spectacle of the 'Other'. In Hall S., ed., *Representation: Cultural Representations and Signifying Practices* (pp. 223-290), Sage, London.
Hampton, K.N., Rainie, L., Lu, W., Dwyer, M., Shin, I. and Purcell, K. (2014), Social Media and the 'Spiral of Silence', *Pew Research Center*, Washington, DC, testo disponibile al sito: http://www.pewinternet.org/2014/08/26/social-media-and-the-spiral-of-silence/ (consultato in data 15/12/2025).
Harel, T.O., Jameson, J.K. and Maoz, I. (2020), The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict, *Social Media + Society*, 6, 2.
DOI: https://doi.org/10.1177/2056305120913983.
Hegel, G.W.F. (1807), *Phänomenologie des Geistes*, Joseph Anton Goebhardt, Bamberg und Würzburg.
Huntington, S.P. (1996), *The Clash of Civilizations and the Remaking of World Order*, Simon & Schuster, New York.
Ince, J., Rojas, F. and Davis, C.A. (2017), The social media response to Black Lives Matter: How Twitter users interact with Black Lives Matter through hashtag use, *Ethnic and Racial Studies*, 40, 11: 1814-1830.
DOI: https://doi.org/10.1080/01419870.2017.1334931.
Kakavand, A.E. (2024), Far-Right Social Media Communication in the Light of Technology Affordances: A Systematic Literature Review, *Annals of the International Communication Association*, 48, 1: 37-56.
DOI: https://doi.org/10.1080/23808985.2023.2280824.
Kaya, S. (2015), Islamophobia in Western Europe: A Comparative, Multilevel Study, *Journal of Muslim Minority Affairs*, 35, 3: 450-465. DOI: https://doi.org/10.1080/13602004. 2015.1080952.
Khan, M.H., Qazalbash, F., Adnan, H.M., Yaqin, L.N. and Khuhro, R.A. (2021), Trump and Muslims: A Critical Discourse Analysis of Islamophobic Rhetoric in Donald Trump's Selected Tweets, *SAGE Open*, 11, 1: 1-16. DOI: https://doi.org/10.1177/21582440211004172.
Kojève, A. (1947), *Introduction à la lecture de Hegel: Leçons sur la*

ISBN Open Access: 9788835191971

Phénoménologie de l'Esprit, Gallimard, Paris.
Kymlicka, W. (2010), The Rise and Fall of Multiculturalism? New Debates on Inclusion and Accommodation in Diverse Societies. In Vertovec S. and Wessendorf S., eds., *The Multiculturalism Backlash: European Discourses, Policies and Practices* (pp. 32-49), Routledge, London.
Kymlicka, W. (2012), *Multiculturalism: Success, Failure, and the Future*, Migration Policy Institute, Washington, DC.
Lacan, J. (1966), *Écrits*, Éditions du Seuil, Paris.
Lasio, D., Girei, E., de Oliveira, J.M., Piras, L. and Serri, F. (2026), Employing women's rights as a racist weapon: The case of Giorgia Meloni in Italy's radical right, *Women's Studies International Forum*, 114: 103237.
DOI: https://doi.org/10.1016/j.wsif.2025. 103237.
Lumsden, K. and Harmer, E., eds. (2019), *Online Othering: Exploring Digital Violence and Discrimination on the Web*, Palgrave Macmillan, Cham.
Maretti, M., Tontodimamma, A. and Biermann, P. (2019), Environmental and climate migrations: an overview of scientific literature using a bibliometric analysis, *International Review of Sociology*.
DOI: https://doi.org/10.1080/03906701.2019.1641270.
Mikelatou, A. and Arvanitis, E. (2019), Multiculturalism in the European Union: A Failure beyond Redemption?, *The International Journal of Diversity in Organizations Communities and Nations Annual Review*, 19, 1:1-18. DOI: https://doi.org/10.18848/1447-9532/CGP/v19i01/1-18.
Modood, T. (2007), *Multiculturalism: A Civic Idea*, Polity Press, Cambridge.
Mohn, E. (2023), Trump travel ban (Executive Order 13769), *EBSCO*, testo disponibile al sito: https://www.ebsco.com/research-starters/history/trump-travel-ban-executive-order-13769 (consultato in data 09/12/2025).

Mudde, C. (2007), *Populist Radical Right Parties in Europe*, Cambridge University Press, Cambridge.
Mudde, C. (2010), The Populist Radical Right: A Pathological Normalcy, *West European Politics*, 33, 6: 1167-1186.
DOI: https://doi.org/10.1080/01402382.2010.508901.
Mudde, C. and Rovira Kaltwasser, C. (2017), *Populism: A Very Short Introduction*, Oxford University Press, Oxford.
Noelle-Neumann, E. (1974), The Spiral of Silence: A Theory of Public Opinion, *Journal of Communication*, 24, 2: 43-51. DOI: https://doi.org/10.1111/j.1460-2466.1974.tb00367.x.
Norris, P. and Inglehart, R. (2019), *Cultural Backlash. Trump, Brexit, and Authoritarian Populism*, Cambridge University Press, Cambridge.
Obreja, D.M. (2022), Mapping the Political Landscape on Social Media Using Bibliometrics: A Longitudinal Co-Word Analysis on Twitter and Facebook Publications Published Between 2012 and 2021, *Social Science Computer Review*, 41, 5: 1712-1728. DOI: https://doi.org/10.1177/08944393221117749.
OCSE (2025), *International Migration Outlook 2025*, OECD Publishing, Paris. DOI: https://doi.org/10.1787/ae26c893-en.

ISBN Open Access: 9788835191971

Paasch-Colberg, S., Strippel, C., Trebbe, J. and Emmer, M. (2021), From Insult to Hate Speech: Mapping Offensive Language in German User Comments on Immigration, *Media and Communication*, 9, 1: 171-180. DOI: https://doi.org/10.17645/mac.v9i1.3399.

Panofsky, A., Dasgupta, K. and Iturriaga, N. (2021), How White nationalists mobilize genetics: From genetic ancestry and human biodiversity to counterscience and metapolitics, *American Journal of Physical Anthropology*, 175, 2: 387-398. DOI: https://doi.org/10.1002/ ajpa.24150.

Pariser, E. (2011), *The Filter Bubble: What the Internet Is Hiding from You*, Penguin Press, New York.

Pickel, G. and Yendell, A. (2016), "Islam als Bedrohung?, *Zeitschrift für Vergleichende Politikwissenschaft*, 10: 273-309. DOI: https://doi.org/10.1007/s12286-016-0309-6.

R Core Team (2021), R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, testo disponibile al sito: https://www.R-project.org/ (consultato in data 09/12/2025).

Ramadhan, J., Widianingsih, K., Zulfa, E.A. and Hayatullah, I.K. (2025), Media and Islamophobia in Europe: A Literature-Based Analysis of Reports 2015–2023, *Religions*, 16, 5: 584. DOI: https://doi.org/10.3390/rel16050584.

Ross Arguedas, R., Robertson, C.T., Fletcher, R. and Nielsen, R.K. (2022), Echo chambers, filter bubbles, and polarisation: A literature review, *Reuters Institute for the Study of Journalism*. DOI: 10.60625/risj-etxj-7k60.

Russo, V. and Maretti, M. (2025), Electoral propaganda in digital space: profiling Facebook use in the Italian general election of 2018, *Quality & Quantity*, 59: 4439-4459. DOI: https://doi.org/10.1007/s11135-025-02152-4.

Said, E.W. (1978), *Orientalism*, Pantheon Books, New York.

Sartre, J.-P. (1943), *L'Être et le néant: Essai d'ontologie phénoménologique*, Gallimard, Paris.

Schroeder, R. (2019), Digital Media and the Entrenchment of Right-Wing Populist Agendas, *Social Media + Society*, 5, 4. DOI: https://doi.org/10.1177/2056305119885328.

Schütz, A. (1944), The Stranger: An Essay in Social Psychology, *The American Journal of Sociology*, 49, 6: 499-507.

Sedgwick, M. (2024), The great replacement narrative: fear, anxiety and loathing across the West, *Politics, Religion & Ideology*, 25, 4: 548-562. DOI: https://doi.org/10.1080/ 21567689.2024.2424790.

Siegel, A.A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J. and Tucker, J.A. (2021), Trumping Hate on Twitter? Online Hate Speech in the 2016 U.S. Election Campaign and its Aftermath, *Quarterly Journal of Political Science*, 16, 1: 71-104. DOI: http://dx.doi.org/10.1561/100.00019045.

Simmel, G. (1908), Der Fremde. In *Soziologie: Untersuchungen über die Formen der Vergesellschaftung* (pp. 509-512), Duncker & Humblot, Leipzig.

Simpson, P.A. (2016), Mobilizing Meanings. Translocal Identities of the Far Right Web, *German Politics and Society*, 34, 4: 34-53. DOI: https://doi.org/10.3167/gps.

ISBN Open Access: 9788835191971

2016.340403.
Sunstein, C.R. (2017), *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, Princeton.
Tillery, A.B. (2019), What Kind of Movement is Black Lives Matter? The View from Twitter, *The Journal of Race, Ethnicity, and Politics*, 4, 2: 297-323. DOI: https://doi.org/ 10.1017/rep.2019.17.
Törnberg, P. and Chueri, J. (2025), When Do Parties Lie? Misinformation and Radical-Right Populism Across 26 Countries, *The International Journal of Press/Politics*, 0, 0. DOI: https://doi.org/10.1177/19401612241311886.
Vaahensalo, E. (2021), Creating the Other in Online Interaction: Othering Online Discourse Theory. In Bailey J., Flynn A. and Henry N., eds., *The Emerald International Handbook of Technology-Facilitated Violence and Abuse (Emerald Studies in Digital Crime, Technology and Social Harms)* (pp. 227-246), Emerald Publishing Limited, Leeds.
Vertovec, S. and Wessendorf, S., eds. (2010), *The Multiculturalism Backlash: European Discourses, Policies and Practices*, Routledge, London.
Wilson, J.H. (2025), Presidential Proclamation of June 4, 2025, Restricting the Entry of Certain Foreign Nationals, *Congress.gov*, testo disponibile al sito: https://www.congress.gov/crs-product/IN12561 (consultato in data 09/12/2025).
Yarchi, M., Baden, C. and Kligler-Vilenchik, N. (2021), Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media, *Political Communication*, 38, 1-2: 98-139. DOI: https://doi.org/10.1080/10584609.2020.1785067.

ISBN Open Access: 9788835191971

Il governo dei discorsi: come cambiano i topic sulle migrazioni al variare delle legislazioni

di *Germano Nocera*[*], *Valeria Policastro*[*], *Giancarlo Ragozini*[*]

1. Introduzione

L'analisi dei contenuti testuali ha assunto un ruolo centrale nello studio dei fenomeni sociali contemporanei, in particolare nei contesti in cui la produzione informativa è ampia, dinamica e politicamente connotata. Nell'ecosistema mediatico digitale, l'immigrazione è tra i temi più discussi e polarizzanti, e la sua narrazione è fortemente influenzata sia dagli orientamenti ideologici delle testate giornalistiche sia dal contesto politico-istituzionale in cui emergono. In questo quadro, il *text mining* offre strumenti quantitativi essenziali per investigare in modo sistematico la struttura latente dei discorsi pubblici.

Il presente capitolo si propone di analizzare l'evoluzione semantica della rappresentazione mediatica dell'immigrazione in Italia nel corso delle ultime tre legislature, attraverso l'elaborazione statistica dei testi pubblicati da quotidiani e mensili politicamente orientati. La domanda di ricerca mira a individuare analogie e differenze tra le testate esaminate, con l'obiettivo di comprendere come il tema migratorio venga discusso, quali argomenti vengano enfatizzati e quali scelte lessicali ricorrano nella costruzione delle narrazioni giornalistiche. Particolare attenzione è dedicata al ruolo congiunto del susseguirsi dei governi e dell'orientamento politico delle fonti nel modellare la rappresentazione dei flussi migratori, consentendo di osservare ricorrenze, divergenze e sfumature linguistiche in grado di influenzare la percezione pubblica.

Nel paragrafo 2 verrà introdotto il *text mining* come strumento di studio dei fenomeni sociali. Nel paragrafo 3 verranno descritti i metodi di *topic modeling*, con particolare attenzione al metodo adottato per l'analisi. Nel paragrafo 4 si discuterà dei dati e delle analisi effettuate, a seguire i paragrafi sui risultati e sulle conclusioni.

[*] Dipartimento di Scienze Politiche, Università degli Studi di Napoli Federico II, germano.nocera@unina.it; valeria.policastro@unina.it; giragoz@unina.it

ISBN Open Access: 9788835191971

2. Il *text mining* come strumento di studio dei fenomeni sociali

La rivoluzione digitale iniziata nella seconda metà del secolo scorso ha avuto, tra le sue diverse conseguenze, anche l'incremento enorme della disponibilità di dati in tutte le forme, tra cui quella testuale. Attualmente, la maggior parte dell'informazione disponibile nel mondo è racchiusa in testi digitalizzati, anche in forma strutturata come articoli di giornale, libri, etc.

Ed è proprio in questo contesto che si sviluppa il *text mining*, che negli anni recenti è emersa come disciplina che risponde all'esigenza di analizzare ed estrarre informazioni utili a partire da dati testuali (Berry e Kogan, 2010). Può dunque essere definito come un processo di trasformazione di un testo più o meno strutturato in informazione strutturata e significativa, che, come sottolineato da alcuni autori, è il risultato multidisciplinare della convergenza di tre materie: informatica, linguistica e statistica (Hobbs, Walker, Amsler, 1982). La possibilità di digitalizzare testi analogici, unita alla grande quantità di testi digitali già disponibili sia sul World Wide Web che nei documenti istituzionali, ha favorito lo sviluppo di un approccio al *text mining* basato sull'applicazione di tecniche informatiche per la consultazione, l'analisi e l'elaborazione di grandi quantità di informazione testuale in tempi brevi. Il supporto tecnologico alla risoluzione di problemi come il rilevamento, il monitoraggio e la sintesi degli argomenti di una raccolta di testi consente l'automatizzazione di procedure lunghe e complesse, rendendo il *text mining* uno degli strumenti più efficaci per l'analisi dei fenomeni sociali in contesti caratterizzati da una crescente produzione di dati testuali. Social media, forum online, articoli di giornale, documenti istituzionali e archivi digitali costituiscono una fonte estremamente ricca di informazioni sulle dinamiche sociali, culturali e politiche contemporanee. Le tecniche di *text mining* consentono l'estrazione di pattern, temi ricorrenti, sentimenti e relazioni latenti all'interno dei testi, permettendo di individuare tendenze emergenti, cambiamenti nel discorso pubblico e modalità di costruzione delle opinioni collettive. La relazione tra il discorso politico e la rappresentazione mediatica dei temi dell'attualità costituisce un tema centrale nelle scienze sociali e politiche. Entman (1993) introduce il concetto di "*framing*", secondo cui i media selezionano e strutturano le informazioni per plasmare l'opinione pubblica e, indirettamente, influenzare le politiche governative. Al tempo stesso, le strategie comunicative dei governi in carica si articolano su questioni particolarmente sensibili, come l'economia, la sicurezza e i conflitti internazionali, e la selezione strategica delle informazioni consente ai governi di costruire narrazioni mirate per rafforzare il consenso politico (Soroka, 2002).

L'immigrazione rappresenta uno dei punti focali, tanto nelle campagne elettorali e nell'agenda politica quanto nella narrazione dei media. Sul fronte

ISBN Open Access: 9788835191971

politico, la copertura mediatica dell'immigrazione varia in funzione della posizione politica del governo, ad esempio con una maggiore enfasi sugli aspetti securitari durante le amministrazioni conservatrici (Boomgaarden e Vliegenthart, 2009). Sul fronte mediatico, è evidente come la stampa eserciti un'influenza determinante sulla percezione pubblica dei fenomeni migratori (Berry *et al.*, 2016). Diversi sistemi mediatici europei presentano variazioni significative nel rapporto tra politica e stampa, con l'Italia che si distingue per un'elevata polarizzazione del discorso mediatico (Hallin e Mancini, 2004). Nel nostro paese, il tema dell'immigrazione ha assunto una rilevanza crescente nel dibattito politico e mediatico, soprattutto negli ultimi decenni, caratterizzati da flussi migratori irregolari e da un'accentuata polarizzazione dell'opinione pubblica (Ambrosini, 2020). Geograficamente, l'Italia grazie alla sua posizione strategica è crocevia di flussi migratori. Il suo estendersi al centro del Mediterraneo fra l'Africa e l'Europa la rende un punto di approdo naturale in particolare per chi proviene da Paesi dell'Africa e del Medio Oriente. La gestione del fenomeno migratorio diventa sempre più complessa, soprattutto se osservata nella cornice di una comunità europea che negli ultimi anni non si è rivelata particolarmente efficace nell'individuare soluzioni condivise utili a contrastare quanto quotidianamente accade nel Mediterraneo nel quale sono scomparse, solo nell'ultimo anno, più di tremila persone (IOM, 2024).

Un fenomeno tanto complesso e delicato si presta a una forte polarizzazione, e chi si occupa di veicolare informazioni su larga scala ha una responsabilità importante. L'ideale di una stampa neutrale è solo questo: un ideale (Cook, 2005), e dunque una narrazione equilibrata e accurata può contribuire a ridurre i pregiudizi, promuovere la comprensione e favorire l'inclusione, mentre, al contrario, una rappresentazione distorta o sensazionalistica può alimentare paure e discriminazioni.

3. Metodi di *topic modeling*

I *topic models* (Blei, 2012) rappresentano una classe di tecniche statistiche che mirano a far emergere le strutture latenti (gli argomenti) dalla collezione di documenti oggetto di studio. Tali argomenti vengono denominati *topic*. Si tratta di modelli estremamente potenti che aiutano nella ricerca e nella determinazione di contenuti testuali in diversi ambiti d'applicazione, come, ad esempio, le librerie digitali, le interfacce di ricerca e le pagine web. Il loro funzionamento è legato all'idea di etichettare le parole e i documenti con appositi tag, così da poter visualizzare, organizzare, sintetizzare, ricercare e raggruppare il contenuto informativo di ciascun testo. È così possibile

ISBN Open Access: 9788835191971

identificare gli argomenti trattati nei documenti analizzati attraverso le parole che li caratterizzano. L'innovazione introdotta dal *topic modeling*, rispetto alle precedenti tecniche di *text mining*, risiede nel fatto di averne ribaltato la filosofia di base, assumendo che i testi siano generati a partire dagli argomenti di cui trattano. Si tratta dunque di un approccio gerarchico, che considera le parole presenti in un testo come espressione del topic a partire dal quale è stato redatto quel testo. Chiaramente, il discorso inizia a complicarsi quando si considera che un singolo documento è in realtà il frutto di più topic e che, tipicamente, il numero di documenti da analizzare è molto ampio. La sfida è dunque quella di riuscire, partendo dalla collezione di documenti a disposizione, a far emergere i topic latenti alla base della generazione di tali documenti, eliminando il rumore rappresentato dalle parole prive di utilità informativa.

3.1. Modelli

I primi modelli per la rappresentazione probabilistica dei documenti testuali erano caratterizzati da una struttura estremamente semplice, come nel caso dell'Unigram Model, in cui un documento è descritto come una realizzazione di un processo di campionamento indipendente di parole secondo una distribuzione di probabilità sul vocabolario, secondo l'ipotesi del bag of words (Harris, 1951), che implica l'indipendenza tra i termini e la perdita dell'informazione sull'ordine. Un'estensione naturale di tale impostazione è il Mixture of Unigrams (Nigam *et al.*, 2000), che introduce distribuzioni multinomiali condizionate alla base del processo di generazione dei documenti che compongono il *corpus*, modellando dunque i topic come componenti di una mistura di distribuzioni multinomiali ed esplicitando la forte assunzione che ciascun documento esibisca un unico topic, ovviamente molto limitante in ampie collezioni di documenti. In ambito deterministico, il Vector Space Model (Salton *et al.*, 1975; Baeza-Yates e Ribeiro-Neto, 1999) rappresenta i documenti come vettori in uno spazio dei termini ad alta dimensionalità, consentendo il calcolo della similarità tramite misure angolari, ma al prezzo di un'assunzione di indipendenza tra i termini e di un'elevata complessità computazionale. Il Latent Semantic Indexing (Deerwester *et al.*, 1990; Papadimitriou *et al.*, 1998) supera parzialmente tali limiti proiettando la matrice termine-documento in un sottospazio di dimensione ridotta ottenuto tramite decomposizione in valori singolari, facendo emergere fattori latenti interpretabili come topic ortogonali, pur rimanendo privo di una formulazione probabilistica esplicita del processo generativo. Il Probabilistic Latent Semantic Indexing (Hofmann, 1999) fornisce una reinterpretazione statistica di LSI,

ISBN Open Access: 9788835191971

modellando le co-occorrenze parola-documento mediante un modello a mistura con variabili latenti, in cui ciascuna parola è generata condizionatamente a un topic e ciascun documento è rappresentato da una distribuzione sui topic. Tuttavia, l'assenza di un modello probabilistico sui documenti e il numero di parametri, che cresce linearmente con la dimensione del corpus, rendono il pLSI poco parsimonioso e limitato nella capacità di generalizzazione, motivando lo sviluppo di modelli generativi gerarchici più completi, come la Latent Dirichlet Allocation (Blei *et al.*, 2003).

3.2. Latent Dirichlet Allocation

La Latent Dirichlet Allocation è definita come un modello probabilistico generativo bayesiano a tre livelli che descrive il processo di generazione di una collezione di dati discreti, ad esempio i *corpora* testuali. Ogni item della collezione viene modellato come una mistura finita di un insieme di k topic latenti. Ogni topic, a sua volta, è modellato come una mistura infinita basata su un set di probabilità ad esso associate. Da tali probabilità derivano le rappresentazioni esplicite di ciascun documento del *corpus*. L'obiettivo del modello è descrivere i membri della collezione considerata e consentire il processamento efficiente anche di grandi moli di dati, preservando le relazioni essenziali sottostanti al rilevamento, alla classificazione e alla sintesi dell'informazione.

Partendo dalla considerazione che la parola rappresenta l'unità testuale di base definita come un *item* di un vocabolario indicizzato dall'insieme $\{1, ..., V\}$, nella notazione della LDA un documento è definito come una sequenza di N parole $\mathbf{w} = \{w_1, w_2, ..., w_N\}$ e un *corpus* è definito come una collezione di M documenti $D = \{\mathbf{w_1}, \mathbf{w_2}, ..., \mathbf{w_M}\}$. I documenti sono rappresentati come misture casuali di topic latenti, in cui ogni topic avrà una certa distribuzione tra le parole. I topic sono dunque variabili latenti multinomiali che rappresentano distribuzioni di probabilità su un insieme di parole.

Il processo di generazione per ciascun documento **w** del *corpus D* avviene tramite i seguenti passaggi:

- Scelta di $N \sim Poisson(\lambda)$.
- Scelta di $\theta \sim Dir(\alpha)$.
- Per ciascuna delle N parole w_n.
- Scelta di un topic $z_n \sim Multinomiale(\theta)$.
- Scelta di una parola w_n da $p(w_n|z_n,\beta)$, una probabilità multinomiale condizionata al topic z_n.

Questo modello prevede una serie di assunzioni semplificatrici. Innanzi-

ISBN Open Access: 9788835191971

tutto, la dimensione *k* della distribuzione Dirichlet è nota e fissa e, di conseguenza, anche la dimensione della variabile *z*, che rappresenta i topic. Inoltre, le probabilità delle parole sono parametrizzate da una matrice β_{kxV}, dove $\beta_{ij} = p(w_j = 1 \mid z_i = 1)$, che viene considerata una quantità fissa da stimare.

L'iperparametro β influenza il modo in cui le parole vengono assegnate ai topic, regolandone la sparsità. Nella sua forma matriciale come riportata in Fig. 1, β è una matrice in cui, in riga, sono indicati i topic e, in colonna, le *V* parole che compongono il vocabolario del *corpus*. Il generico elemento della matrice rappresenterà la probabilità che la singola parola sia generata dalla distribuzione di quel topic, come riportato nella Tabella 1.

Tab. 1 – Distribuzione di probabilità delle parole nei topic

	w_1	w_2	...	w_v
z_1	$p(w_1)$	$p(w_2)$	...	$p(w_v)$
z_2	$p(w_1)$	$p(w_2)$	...	$p(w_v)$
...	...	...	...	...
z_k	...	...	...	...

Al crescere del valore di β, la distribuzione di probabilità delle parole per topic tende a essere più uniforme, con meno distinzione tra i topic. Al diminuire di β, la distribuzione di probabilità delle parole sui topic è invece più sparsa, con una maggiore distinzione tra i topic (che saranno caratterizzati da meno parole e probabilità più alte).

L'assunzione relativa alla Poisson non è critica; anzi, è realistico ipotizzare che la lunghezza *N* del documento segua tale distribuzione, notando inoltre che è indipendente dalle altre variabili generatrici dei dati (θ e *z*) e, dunque, possa essere considerata una variabile casuale ausiliaria, la cui casualità potrebbe non essere presa in considerazione negli sviluppi successivi.

La variabile casuale θ, una *Dirichlet k*-dimensionale (che è a sua volta parametro della distribuzione multinomiale che governa la proporzione di topic nel documento), assume valori nel simplesso *(k-1)*-dimensionale (un vettore *k*-dimensionale θ giace nel simplesso *(k-1)*-dimensionale se $\theta_i \geq 0$ e $\sum_{i=1}^{k} \theta_i = 1$) e su di esso avrà la seguente funzione di densità di probabilità:

$$p(\theta|\alpha) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} {\theta_1}^{\alpha_1 - 1} \dots {\theta_k}^{\alpha_k - 1}$$

dove il parametro *k*-dimensionale α ha componenti $\alpha_i > 0$ e $\Gamma(.)$ è la funzione Gamma.

ISBN Open Access: 9788835191971

La Dirichlet è una distribuzione conveniente da utilizzare sul simplesso, poiché appartiene alla famiglia esponenziale ed è la distribuzione a priori coniugata della distribuzione multinomiale. Le sue proprietà facilitano l'elaborazione degli algoritmi di inferenza e le stime dei parametri del modello LDA. Dati gli iperparametri α e β, un vettore di k topic $\mathbf{z}$ e un vettore documento $\mathbf{w}$ composto da N parole, la distribuzione di probabilità congiunta di una mistura di topic θ è data da

$$p(\theta, \mathbf{z}, \mathbf{w}\,|\alpha, \beta) = p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta)$$

dove $p(z_n|\theta)$ è semplicemente θ_i, per i tale che $\sum_i p(z_i) = 1$. Avendo dotato θ di una Dirichlet di parametro α, integrando rispetto a θ e sommando su z si ottiene la distribuzione marginale di un documento:

$$p(\mathbf{w}\,|\alpha, \beta) = \int \left(p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \right) d\theta$$

Infine, effettuando il prodotto delle probabilità marginali dei singoli documenti, si ottiene la probabilità del *corpus*

$$p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N} \sum_{z_{dn}} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn}, \beta) \right) d\theta_d$$

Il modello LDA è illustrato nella Fig. 1.

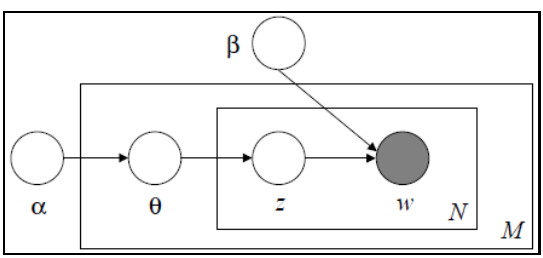


Fig. 1- Rappresentazione schematica del LDA topic model

Lo schema in Fig. 1 mostra chiaramente i tre livelli della rappresentazione LDA. Gli iperparametri α e β agiscono a livello di corpus, con campionati un'unica volta durante il processo di generazione. Le variabili θ_d agiscono a livello di documento, campionate una volta per ciascun documento. Infine,

ISBN Open Access: 9788835191971

le variabili z_{dn} e w_{dn} agiscono al livello della parola e sono entrambe campionate una volta per ciascuna parola in ciascun documento. Nella LDA si assume che le parole siano generate dai topic (attraverso una distribuzione condizionata fissata), con θ che svolge il ruolo di parametro casuale (estratto da una Dirichlet di parametro α) e che governa i parametri della multinomiale che determinano le proporzioni dei topic all'interno del corpus.

La potenza del modello LDA con k topic risiede nella possibilità di considerare un documento capace di esporre più topic a diversi gradi di intensità, contemplando al tempo stesso la presenza di un processo generativo dei documenti e coniugando il tutto con la parsimonia del numero di parametri da stimare, pari a $k + kV$ e dunque indipendente dalla dimensione M del corpus.

Dal momento che il modello LDA assume noto e fisso il numero k dei topic da considerare computazionalmente, si sfrutta la *perplexity* (Jelinek, Mercer, Bahl e Baker, 1977). Essa è una metrica basata sull'entropia, generalmente utilizzata per valutare la qualità di un modello. Nell'ambito del *text mining*, misura quanto bene un modello riesce a predire una sequenza di parole. Nello specifico, la *perplexity* di una distribuzione di probabilità discreta p è una misura dell'incertezza che il modello ha nel predire il testo, ed è data da

$$PP(p) = 2^{H(p)} = 2^{-\sum_x p(x)\log_2 p(x)} = \prod_x p(x)^{-p(x)}$$

con *H(p)* che rappresenta l'entropia della distribuzione. Minore il valore della *perplexity,* minore l'incertezza legata alla previsione.

3.3. Pretrattamento

La sfida, che si rinnova di fronte a ogni nuova applicazione delle tecniche di text mining, consiste nel comprendere, tra i tanti elementi testuali a disposizione, quali mantenere e in che modo eventualmente trasformarli, al fine di estrarre informazione ed evitare di perderla. Non esiste un procedimento univoco, e tutto dipende dalla natura dei *corpora* e dagli obiettivi dello studio. È possibile suddividere le tecniche di pretrattamento del testo in fasi che si susseguono in modo ordinato (Tuzzi, 2003).

La prima fase è solitamente quella della *tokenizzazione* del corpus. Essa consiste nel ripulire una sequenza di caratteri, ottenendo unità testuali definite token che non è detto siano necessariamente parole significative. Ha dunque l'obiettivo di delimitare le unità di testo da considerare e di eliminare eventuali caratteri che potrebbero interferire con il processo di analisi. Ad esempio, rimuovendo certi caratteri, come i segni di punteggiatura, i simboli, i numeri o

ISBN Open Access: 9788835191971

le lettere dell'alfabeto di uno specifico linguaggio, e mantenendone invece altri, si possono ottenere anche sequenze di caratteri prive di un vero significato o di quello originale. La seconda fase, che solitamente va di pari passo con la precedente, è quella della *normalizzazione* del *corpus.* Essa mira a identificare, come simili o differenti, le unità di testo considerate, al fine di uniformarle. Prevede, ad esempio, la rimozione delle lettere maiuscole, la gestione degli acronimi, la correzione degli errori, la riconciliazione dello spelling e degli stili differenti e, in generale, una riduzione della variabilità della massa di dati testuali a disposizione. La terza fase è quella del *filtering, transformation and enrichment*, che consiste nel filtrare, trasformare e arricchire l'*output* delle fasi precedenti. L'operazione più importante, solitamente eseguita in questa fase, consiste nell'identificazione delle cosiddette *stopwords* del linguaggio di riferimento. Si tratta di vere e proprie liste di parole ritenute prive di utilità, poiché non contengono informazioni significative. Tendenzialmente, si tratta di parole più comuni in un certo linguaggio (anche se non esiste un accordo generale), come preposizioni, articoli e congiunzioni, ma potrebbero esserne considerate ulteriori sulla base del *corpus* oggetto di analisi e dell'*output* delle fasi precedenti. Come quarta fase è possibile considerare due procedimenti tra loro alternativi, ovvero lo *stemming* e la *lemmatizzazione* (Balakrishnan ed E. Lloyd-Yemoh, 2014). Lo *stemming* consiste nella riduzione della variabilità introdotta dall'uso di alcuni vocaboli, come le forme plurali, i diversi tempi verbali e le situazioni simili. In sostanza, gli algoritmi di *stemming* (Porter, 1980) riducono i termini alla loro radice, tagliandone la coda. Infine, la *lemmatizzazione* consiste in una riduzione della variabilità morfologica del testo attraverso un processo che dipende in gran parte dal linguaggio del documento, con l'obiettivo di ricondurre ciascun termine al lemma di riferimento, solitamente contrassegnandolo con un tag che ne consenta l'individuazione esatta della categoria grammaticale (verbo, aggettivo, ecc.).

Tali operazioni di pretrattamento del testo sono necessarie alla creazione della *Term-Document Matrix* (TDM), una tabella di contingenza che contiene le frequenze (eventualmente opportunamente pesate) con cui ciascuna parola si presenta in ciascun documento. Sarà tipicamente una matrice di grandi dimensioni e, in alcuni casi, estremamente sparsa. La creazione della TDM genera la struttura di dati necessaria alle operazioni di analisi statistica e rappresenta il momento in cui l'informazione non strutturata contenuta nel *corpus* assume la forma strutturata che ne consente l'estrazione mediante le tecniche di *text mining*.

ISBN Open Access: 9788835191971

4. Dati, pretrattamento e analisi

Volendo indagare gli articoli riguardanti il fenomeno dei flussi migratori in ingresso pubblicati da testate giornalistiche italiane dichiaratamente orientate politicamente, con l'obiettivo di evidenziare analogie e differenze rispetto alle tematiche trattate, ai termini utilizzati per descriverle e ai governi che si sono susseguiti, sono stati necessari alcuni passaggi preliminari all'analisi testuale. Partendo dagli articoli pubblicati nel periodo compreso tra settembre 2016 e marzo 2024 raccolti dal PRIN denominato "*TOLERANT: Identification and Critical Analysis of Online Racism and Xenophobia against (Im)migrants and Roma People*", si è dovuto in primis dividere le testate giornalistiche per il loro orientamento politico come descritto nella Tabella 2 riportata di seguito:

Tab. 2 – Suddivisione riviste per orientamento politico

Destra	**Sinistra**
Destra.it	*Left.it*
Libero	*Il Manifesto*
Il Giornale	*Il Riformista*
La voce del patriota	*L'Unità*
Il Secolo d'Italia	*Città Futura*
Il Primato Nazionale	*Contropiano*

Come secondo step si è dovuto suddividere il corpus composto da 1997 documenti, in periodi temporali che corrispondessero ad inizio e fine delle legislature come descritto nella Tabella 3.

Tab. 3 – Suddivisione documenti per legislatura e orientamento politico

Periodo		**Governo**	**Documenti**		
Inizio	*Fine*		*Fonte SX*	*Fonte DX*	*totale*
22/10/2022	in corso	Meloni	*885*	*682*	*1567*
13/02/2021	22/10/2022	Draghi	*110*	*263*	*373*
05/09/2019	13/02/2021	Conte II	*16*	*25*	*41*
01/06/2018	04/09/2019	Conte I	*5*	*2*	*7*
12/12/2016	01/06/2018	Gentiloni	*8*	*0*	*8*
22/02/2014	12/12/2016	Renzi	*1*	*0*	*1*

ISBN Open Access: 9788835191971

Dato lo sbilanciamento dei documenti rispetto alle fonti di destra o di sinistra delle prime tre legislature, si è deciso di includere nell'analisi solo il periodo degli ultimi tre governi italiani.

A tal punto è stato necessario procedere a un pretrattamento del testo per ricondurlo a una forma strutturata, consentendo l'analisi con la metodologia di *Latent Dirichlet Allocation.*

Nel caso specifico sono state dunque applicate, per ciascuno dei *subcorpora* considerati, le operazioni di pretrattamento del testo necessarie alla creazione della *Term-Document Matrix* (TDM).

Le operazioni di normalizzazione e tokenizzazione effettuate hanno previsto la rimozione della punteggiatura, dei numeri, dei caratteri speciali, delle URL e delle *stopwords* italiane presenti nei documenti. Una volta ripulito il testo, trattandosi di articoli di giornale scritti in lingua italiana, si è scelto di ridurre la variabilità dei termini mediante una lemmatizzazione, particolarmente efficace per una lingua morfologicamente complessa come quella italiana, che consente di ricondurre ciascun termine al suo lemma. Trattandosi di un'operazione effettuata in automatico da un apposito algoritmo basato sulla combinazione di un processo di POS *tagging* con l'utilizzo di uno specifico dizionario morfologico (scelto in funzione della lingua in cui sono espressi i testi da analizzare) continuamente aggiornato, non sempre si rivela efficace nel trattare correttamente tutte le parole che incontra nel *corpus*. Si è dunque reso necessario un ulteriore passaggio di verifica e di correzione delle parole per le quali la lemmatizzazione non ha restituito un risultato soddisfacente. Una volta conclusa questa fase, sfruttando la legge di Zipf (1949), sono state individuate e rimosse le parole con frequenza più elevata e gli *hapax legomena* (parole che compaiono nel *corpus* un'unica volta), al fine di far emergere più facilmente l'informazione latente nei documenti.

Si è dunque scelto di eseguire più esperimenti di LDA sui diversi *subcorpora*, incrementando di volta in volta il numero di topic considerati, per poi selezionare il numero corrispondente al livello di *perplexity* ritenuto più adatto, sulla base del minimo raggiunto o del cosiddetto elbow point, ovvero il punto in cui il guadagno ottenuto non giustifica l'aggiunta di un nuovo topic. Di seguito (Fig. 2) viene riportato il numero di topic*,* scelto sulla base della variazione della *perplexity,* per ciascuno dei *subcorpora* considerati.

ISBN Open Access: 9788835191971

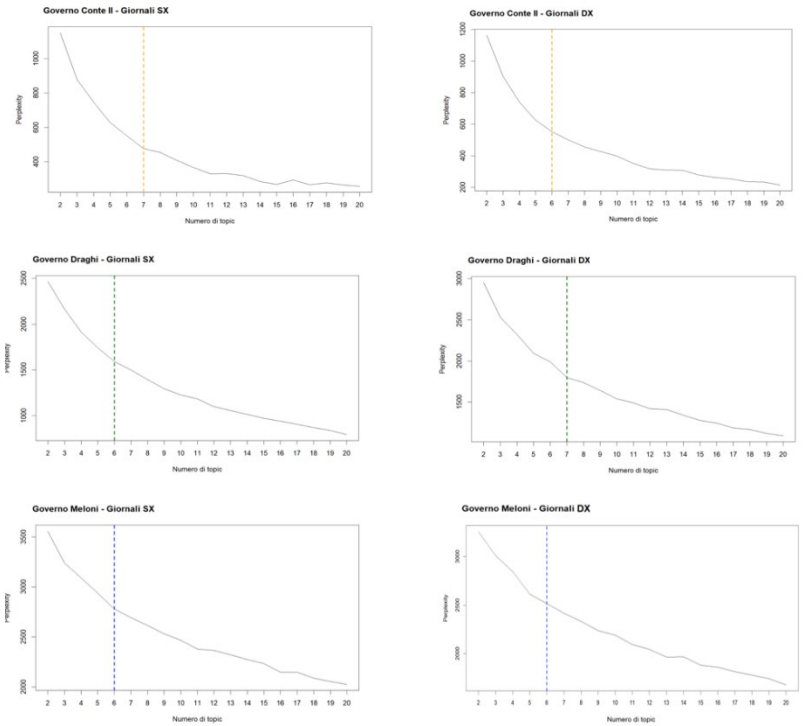


Fig. 2 – Perplexity *per ciascuno dei subcorpora*

5. Risultati

Il modello di *Latent Dirichlet Allocation* si è dimostrato efficace nel fornire una risposta soddisfacente ai nostri obiettivi, mettendo chiaramente in evidenza affinità e divergenze sia in funzione dell'ideologia sottostante sia degli eventi che hanno caratterizzato il periodo storico della legislatura, rivelando informazioni latenti all'interno del corpus analizzato.

Nel *subcorpus* relativo al governo Conte II i topic individuati come riportato in Figura 3 sono stati:

1. Immigrazione e richieste di asilo 2.
2. Immigrazione in Italia, Openarms.
3. Emergenza sbarchi: profughi, clandestini e ONG.
4. Immigrazione, Italia e PIL.
5. ONG, OceanViking, OpenArms e Libia.
6. Immigrazione irregolare, sicurezza, sanità ed economia (post-Covid).
7. Elezioni americane, Biden vs Trump.

ISBN Open Access: 9788835191971

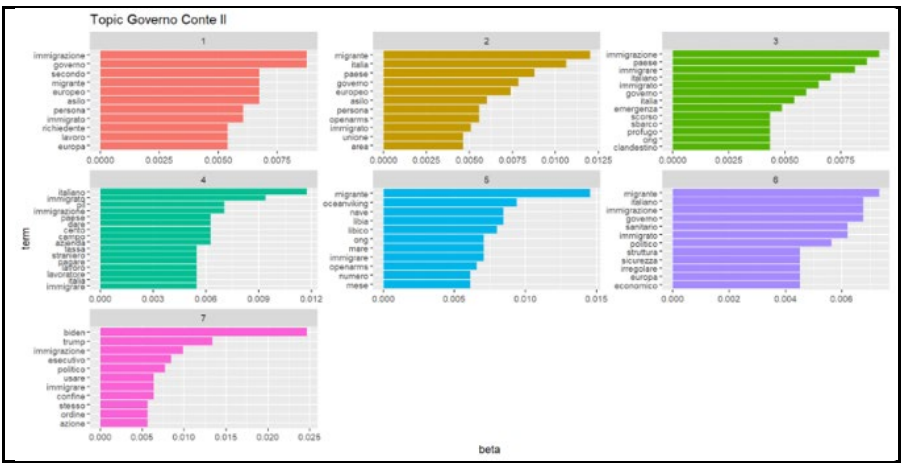


Fig. 3 – LDA topic governo Conte II

I temi predominanti hanno riguardato le richieste di asilo, gli sbarchi, il ruolo delle ONG (con particolare riferimento ai casi Open Arms e Ocean Viking), la sanità e l'economia post-pandemica, nonché le elezioni presidenziali statunitensi

Invece, analizzando separatamente i giornali di destra e di sinistra, si notano differenze. I giornali di destra, come riportato in Figura 4, hanno posto maggiore enfasi sull'emergenza degli sbarchi in Sicilia e sulle ONG percepite come vettori di ingresso degli stranieri.

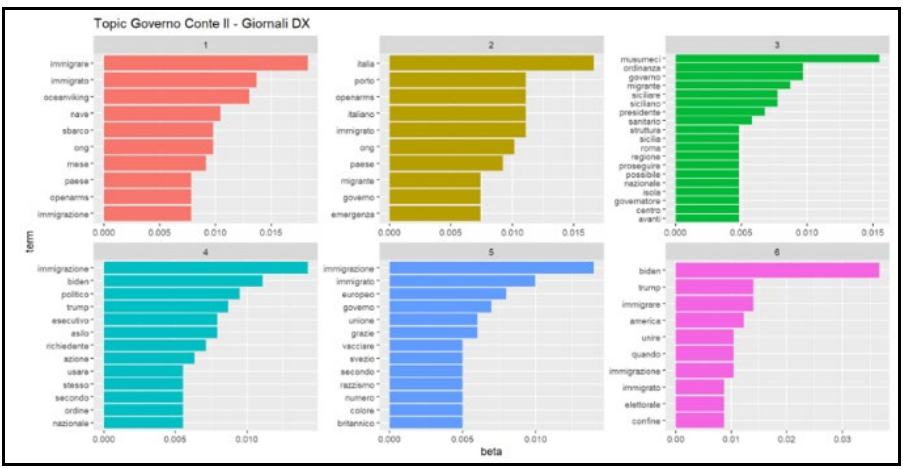


Fig. 4 – LDA topic governo Conte II giornali di destra

ISBN Open Access: 9788835191971

I giornali di sinistra, riportati in Figura 5, hanno invece enfatizzato le politiche nazionali di gestione dell'immigrazione (in particolare i decreti Salvini), il legame tra immigrazione e lavoro, con particolare attenzione al lavoro irregolare, e i conflitti in Libia e in Siria.

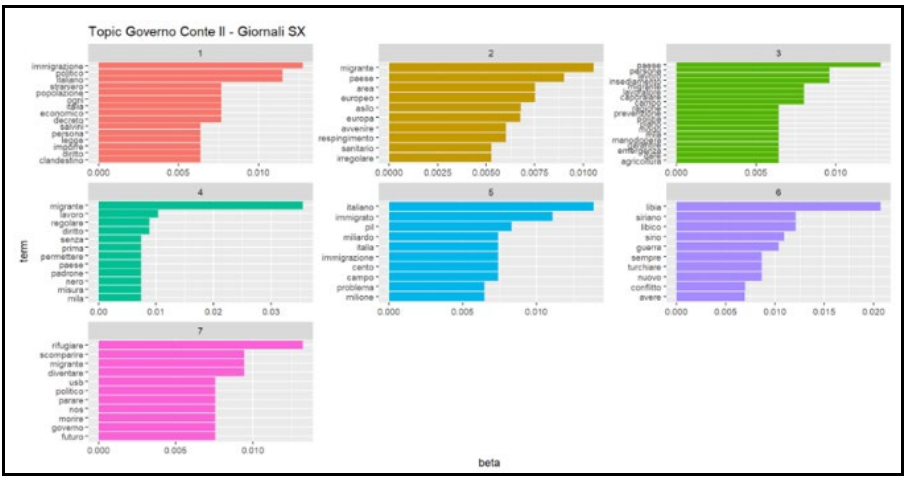


Fig. 5 – LDA topic governo Conte II giornali di sinistra.

Invece, per quanto riguarda il *subcorpus* relativo al governo Draghi, i topic individuati, come riportato in Figura 6, sono stati:

1. Politiche europee.
2. ONG e sbarchi clandestini in Italia.
3. Trafficanti e ONG.
4. Migranti, Italia ed Europa.
5. Immigrazione, persone e diritti.
6. Immigrazione dall'Africa.
7. Politiche interne.
8. I migranti sono persone.
9. Condanna al favoreggiamento dell'immigrazione clandestina.

ISBN Open Access: 9788835191971

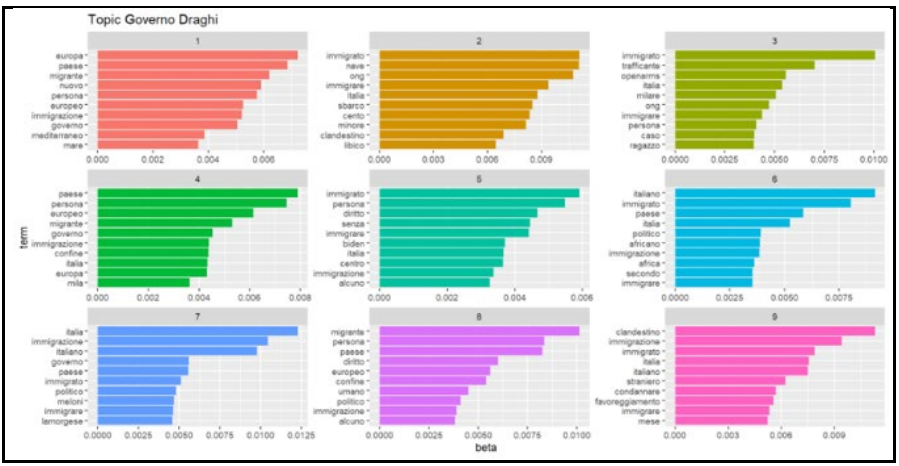


Fig. 6 – LDA topics governo Draghi.

In generale, i temi del governo Draghi hanno riguardato le politiche migratorie nazionali ed europee, il ruolo dei trafficanti e delle ONG, nonché la sicurezza interna. I giornali di destra (Fig. 7) hanno attribuito maggiore rilevanza alle questioni legate alla cittadinanza, all'immigrazione irregolare, alla sicurezza interna, alle ONG percepite come vettori dell'immigrazione e alle politiche interne.

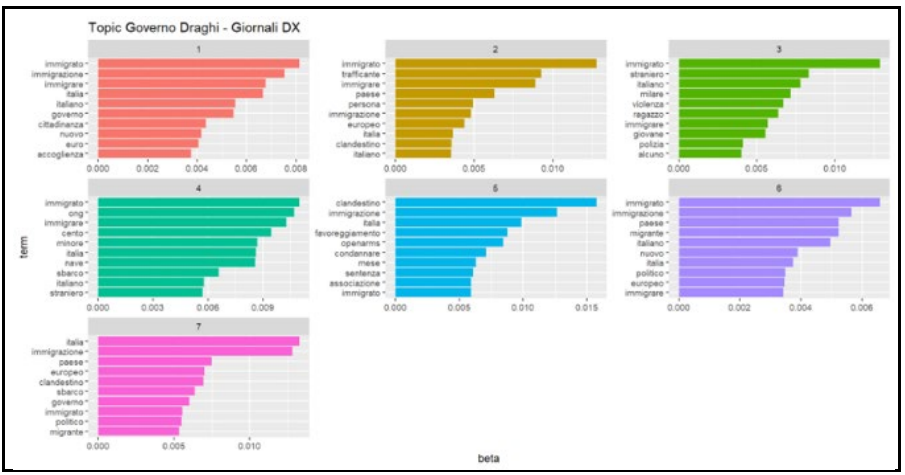


Fig. 7 – LDA topic governo Draghi giornali di destra

I giornali di sinistra (Fig. 8) hanno invece enfatizzato i diritti dei migranti, il lavoro e la guerra in Libia.

ISBN Open Access: 9788835191971

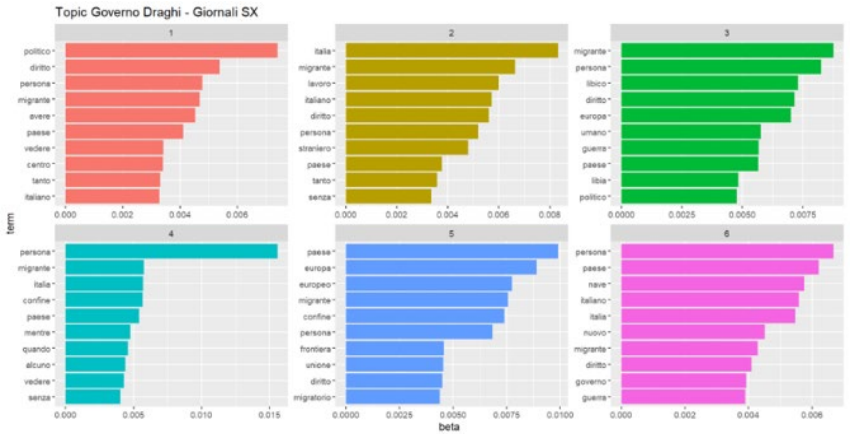


Fig. 8 – LDA topic governo Draghi giornali di sinistra

Infine, per quanto riguarda il *subcorpus* relativo al governo Meloni, i topic individuati, come riportato in Figura 9, sono stati:

1. Lavoro e Immigrazione.
2. Ricerca di accordi politici europei.
3. Persone migranti e la morte nel Mediterraneo.
4. Le ONG, gli sbarchi nei porti italiani e il recupero di persone in mare.
5. Immigrati, stranieri e violenza: la sicurezza interna.
6. Il diritto all'accoglienza.
7. Rapporti con Tunisia e Libia e rotte migratorie.
8. Politica interna.
9. Garanzia dei diritti dei richiedenti asilo.

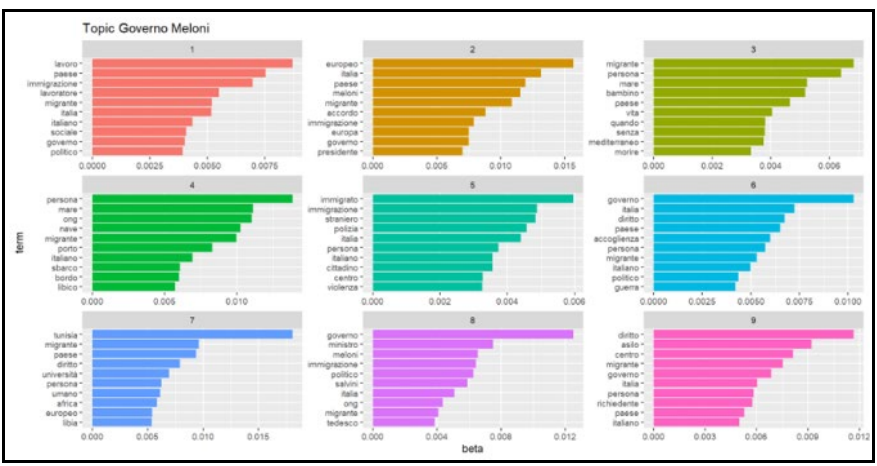


Fig. 9 – LDA topic governo Meloni

ISBN Open Access: 9788835191971

Quindi i temi predominanti hanno riguardato le politiche migratorie nazionali ed europee, il ruolo delle ONG, la sicurezza interna e il lavoro. Analizzando solo i giornali di destra (Fig. 10), hanno posto maggiore enfasi sulla diplomazia internazionale, sulla regolamentazione dell'accoglienza, sulla sicurezza interna e sulle ONG, percepite come responsabili dell'arrivo degli stranieri.

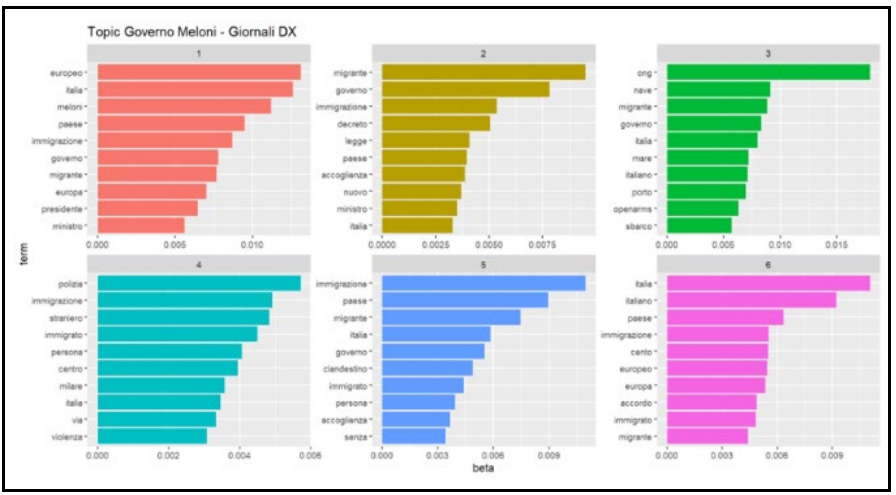


Fig. 10 – LDA topic governo Meloni giornali di destra

Nel corso dell'analisi dei giornali di sinistra (Fig. 11), questi hanno enfatizzato i diritti dei migranti, i Centri di Permanenza per il Rimpatrio (CPR), gli accordi persi dal governo con l'Albania, i rapporti con la Libia e la Tunisia e gli sbarchi a Lampedusa.

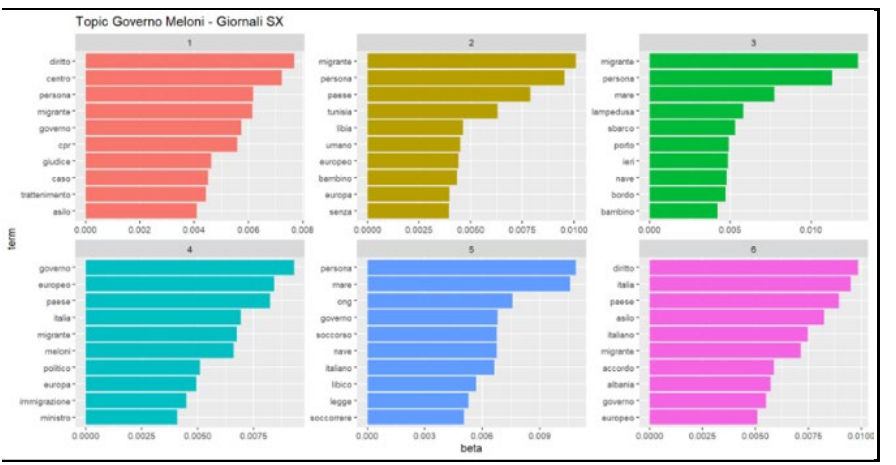


Fig. 11 – LDA topic governo Meloni giornali di sinistra

ISBN Open Access: 9788835191971

6. Conclusioni

Il presente capitolo si è posto l'obiettivo di rivolgere l'attenzione, in particolare, ai termini utilizzati e ai temi trattati nell'evoluzione temporale della narrazione relativa al fenomeno dei flussi migratori in ingresso, con l'idea non di esprimere giudizi di valore sui media coinvolti, bensì di evidenziarne analogie e differenze rispetto alle modalità del racconto e alla scelta delle tematiche da evidenziare rispetto al contesto comune. L'obiettivo è stato comprendere come le diverse testate, a seconda del loro orientamento politico e della legislatura vigente al momento della stesura dell'articolo, affrontano e descrivono il fenomeno dei flussi migratori in entrata nel nostro Paese e le tematiche annesse, costruendo la narrazione che alimenta l'opinione pubblica.

Per quanto riguarda la terminologia utilizzata dai giornali esaminati, oltre a riflettere l'attualità e il momento storico a cui sono riferiti, sono evidentemente influenzate dall'orientamento politico dei giornali e dal diverso governo in carica. Negli articoli analizzati sono emersi sia l'uso di termini specifici sia quello di termini comuni a entrambe le parti.

In generale, le testate di destra hanno mostrato, indipendentemente dalla legislatura, un linguaggio più polarizzato (*immigrato, clandestino*), una tendenza a enfatizzare argomenti legati alla sicurezza nazionale, alle politiche migratorie e alla critica alle ONG. Al contrario, le testate di sinistra hanno mostrato, indipendentemente dalla legislatura, un linguaggio più inclusivo (*migrante, persona*), con una costante attenzione a temi quali i diritti civili, la solidarietà sociale e le politiche europee.

Quanto emerso riflette non solo le priorità politiche e strategiche dei diversi esecutivi, ma anche il modo in cui i media, con orientamenti ideologici differenti, interpretano e presentano le questioni di attualità scegliendo quale aspetto evidenziare.

Riferimenti bibliografici

Balakrishnan, V. and Lloyd-Yemoh, E. (2014), Stemming and lemmatization: A comparison of retrieval performances, *Lecture Notes on Software Engineering,* 2(3): 174-179.

Berry, M. W. and Kogan, J., eds. (2010), *Text mining: applications and theory*, John Wiley & Sons, New Jersey.

Blei, D. M. (2012), Probabilistic topic models, *Communications of the ACM*, 55.4: 77-84.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), Latent Dirichlet Allocation, *Journal of machine Learning research*, 993-1022.

ISBN Open Access: 9788835191971

Cook, T. E. (2005), *Governing with the News: The News Media as a Political Institution* (2nd ed.), University of Chicago Press, Chicago.
Entman, R. M. (1993), Framing: Toward Clarification of a Fractured Paradigm, *Journal of Communication*, 43(4): 51-58.
Hallin, D. C. and Mancini, P. (2004), *Comparing Media Systems: Three Models of Media and Politics*, Cambridge University Press, Cambridge.
Harris, Z. S. (1951), *Methods in structural linguistics.* Chicago: University of Chicago Press.
Hobbs, J. R., Walker, D. E. and Amsler, R. A. (1982), Natural language access to structured text. *Coling 1982: Proceedings of the Ninth International Conference on Computational Linguistics.*
IOM (2024), "World Migration Report 2024".
Jelinek, F., Mercer, R. L., Bahl, L. R. and Baker, J. K. (1977), Perplexity – a measure of the difficulty of speech recognition tasks, *The Journal of the Acoustical Society of America*, 62(S1): S63-S63.
Porter, M. F. (1980), An algorithm for suffix stripping, *Program*, 14(3): 130-137.
Soroka, S. N. (2002), *Agenda-Setting Dynamics in Canada*, UBC Press, Vancouver, BC.
Tuzzi, A. (2003), *L'analisi del contenuto. Introduzione ai metodi e alle tecniche di ricerca*, Carocci, Roma.
Zipf, G. K. (1949), *Human Behaviour and the Principle of Least Effort,* Addison-Wesley Press, Cambridge (MA).

ISBN Open Access: 9788835191971

La rappresentazione dell'immigrazione nella stampa italiana: un'analisi statistica del linguaggio mediatico

di *Alex Cucco*[*], *Emiliano del Gobbo*[**], *Sara Fontanella*[***], *Lara Fontanella*[*]

1. Il ruolo dei media nella costruzione sociale della migrazione

Un consolidato insieme di ricerche ha evidenziato la funzione centrale svolta dai mezzi di comunicazione nella diffusione e costruzione di informazioni nella società contemporanea (Bennett ed Entman, 2001; Chong e Druckman, 2007; Koopmans e Statham, 2010). L'influenza esercitata dalla rappresentazione mediatica sul pubblico può essere analizzata attraverso due principali paradigmi teorici: l'agenda setting e il framing (Eberl *et al.*, 2018). Laddove il primo paradigma si basa sul presupposto che l'informazione giornalistica possa orientare l'attenzione del pubblico verso determinate questioni ritenute rilevanti (McCombs, 2004), il secondo sostiene che i mezzi di informazione forniscono gli schemi interpretativi attraverso cui elaborare cognitivamente tali tematiche (Entman, 1993). Secondo la teoria dell'agenda setting, gli organi di informazione assumono pertanto il ruolo di selezionatori informativi (*gatekeeper*): definiscono quali contenuti accedono allo spazio pubblico ed attuano un potere di definizione dell'agenda attraverso la cernita delle informazioni da divulgare (Bleich *et al.*, 2015). Le tematiche di interesse pubblico possono essere inoltre oggetto di rappresentazioni differenziate: seguendo quest'ottica le fonti informative quali gli organi di stampa farebbero ricorso alle cornici interpretative o frame (Goffman, 1974) al fine di strutturare la presentazione dei contenuti informativi. Il processo di framing può essere concettualizzato come la procedura mediante la quale una

[*] Dipartimento di Studi Socio-Economici, Gestionali e Statistici, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, alex.cucco@unich.it; lara.fontanella@unich.it

[**] Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, emiliano.delgobbo@unina.it

[***] National Heart and Lung Institute, Imperial College, London, s.fontanella@imperial.ac.uk

ISBN Open Access: 9788835191971

entità comunicativa, quale un'organizzazione giornalistica, circoscrive e configura una problematica di natura politica o una controversia di rilevanza pubblica. Gli "effetti di framing" si riferiscono al fenomeno per cui le modalità attraverso cui una tematica viene descritta o categorizzata esercitano un'influenza sulla costruzione dell'opinione pubblica: dal momento che la maggior parte delle cognizioni politiche dei cittadini è veicolata mediaticamente, la comprensione e persino gli orientamenti valutativi relativi a questioni di natura politica possono essere condizionati in modo sostanziale dalle scelte di selezione e dalle strategie di presentazione dei contenuti informativi (Nelson *et al.*, 1997). Di conseguenza, un approccio metodologico ricorrente negli studi sulla rappresentazione mediatica di determinate tematiche è rappresentato dalla *frame analysis* (Goffman, 1974), che esamina il quadro simbolico ed ermeneutico attraverso cui si articola il discorso mediale e l'idea organizzativa centrale mediante la quale si attribuisce significato agli eventi rilevanti, definendo la questione cruciale (Gamson e Modigliani, 1989).

Il presente lavoro concentra l'analisi sulla rappresentazione del fenomeno migratorio nei mezzi di comunicazione di massa. La rassegna della letteratura scientifica (Eberl *et al.*, 2018; Fuller, 2024) evidenzia come l'esame del discorso mediatico possa fornire un apporto rilevante allo studio della visibilità e della rappresentazione di migranti e minoranze nella sfera pubblica, contribuendo alla comprensione tanto dei fattori che determinano la copertura giornalistica, quanto degli effetti di quest'ultima sugli orientamenti dell'opinione pubblica, sulle dinamiche sociali e sui processi di policy-making. La rappresentazione dell'immigrazione nel discorso mediatico non costituisce un processo neutrale: essa concorre in maniera sostanziale alla costruzione sociale della percezione di tale fenomeno complesso (Valenzuela-Vergara, 2019). Nonostante le differenze tra contesti nazionali e sebbene le strategie di inquadramento possano divergere in funzione dei gruppi migranti analizzati, la copertura mediatica della migrazione nel continente europeo tende ad assumere connotazioni prevalentemente negative e a privilegiare la dimensione del conflitto (Fuller, 2024). Un'esposizione prolungata a tali narrazioni favorirebbe lo sviluppo di orientamenti avversi verso la migrazione, potendo innescare l'attivazione di rappresentazioni stereotipate dei gruppi migranti e condizionare i comportamenti di voto (Eberl *et al.*, 2018). Concentrando l'analisi sul periodo post-2015 nell'Europa occidentale e settentrionale, Fuller (2024) identifica discorsi mediatici dominati da meccanismi di costruzione dell'alterità (*othering*), dalla retorica della minaccia e dalla questione della meritevolezza. Emergono evidenze di una differenziazione parziale tra tipologie di soggetti con background migratorio, pur persistendo una forte inclinazione alla stereotipizzazione e all'essenzializzazione, a

ISBN Open Access: 9788835191971

prescindere dal reale profilo dei migranti o dei loro discendenti. I migranti vengono rappresentati come un elemento di minaccia per la società ospitante, in alcuni casi attraverso enunciazioni dirette ed esplicite. I discorsi di securitizzazione nel Regno Unito, in Francia ed in Italia costruiscono i migranti come una minaccia alla sicurezza nazionale e come un elemento di vulnerabilità per la stabilità economica dello Stato (Caviedes, 2015). Il discorso relativo ai rifugiati tende generalmente ad assumere connotazioni maggiormente empatiche. Nei media svedesi, i rifugiati vengono rappresentati come "vittime", i migranti come "intrusi illegali" e i migranti comunitari come "rom mendicanti" (Wojahn, 2023). Per contro, nei media austriaci risulta assente una netta distinzione tra le categorie lessicali impiegate (Schroter, 2023). Anche nelle ricerche sui media polacchi emerge un utilizzo intercambiabile dei termini "*rifugiato*" e "*migrante*" (Troszyński ed El-Ghamari, 2022). Attraverso l'analisi di tre testate giornalistiche italiane, Flinz e Leonardi (2023) evidenziano come il lessema più frequente, "*migranti*", funzioni quale termine onnicomprensivo (passe-partout), subendo un processo di desemantizzazione parziale; viceversa, i termini "*rifugiati*" e "*immigrati*" acquisiscono una marcata valenza politica: laddove le occorrenze di "*rifugiati*" diminuiscono al crescere dell'orientamento politico conservatore delle testate, quelle di "*immigrati*" aumentano nelle pubblicazioni maggiormente orientate verso destra. Lo studio fornisce inoltre una chiara demarcazione concettuale tra tali termini.

La presente ricerca si propone di analizzare la presenza e la distribuzione dei frame impiegati nella rappresentazione dell'immigrazione nella stampa italiana. Poiché l'orientamento politico può rappresentare una variabile significativa che condiziona la copertura giornalistica (Mancini *et al.*, 2021), l'analisi si sviluppa attraverso un confronto tra testate giornalistiche con posizionamenti ideologici differenziati: quotidiani e periodici di area progressista/sinistra e quotidiani e periodici di area conservatrice/destra.

Il presente lavoro propone un approccio metodologico basato sull'analisi statistica dei contenuti testuali, finalizzato a caratterizzare le scelte lessicali ed i contenuti semantici attraverso strategie quantitative. In particolare, si fa uso di reti semantiche per mappare le relazioni tra termini, individuando cluster concettuali, schemi ricorrenti e pattern di significato nei testi. Questo approccio consente di descrivere in maniera strutturata i legami tra parole, offrendo una rappresentazione delle strutture semantiche sottostanti senza fare assunzioni a priori sui contenuti. La metodologia sarà applicata al caso della rappresentazione del fenomeno migratorio nei mezzi di comunicazione di massa, con l'obiettivo di mostrare come le reti semantiche possano evidenziare differenze lessicali, pattern concettuali e modalità di rappresentazione di tematiche complesse.

ISBN Open Access: 9788835191971

2. Analisi testuale esplorativa della copertura migratoria nella stampa italiana

Il corpus analizzato comprende 1.997 articoli giornalistici, identificati mediante parole chiave specifiche (immigrazione, immigrati, migranti, rifugiati) e pubblicati su 12 testate italiane: 6 di area progressista/sinistra e 6 di area conservatrice/destra (tab. 1), con una ripartizione degli articoli sostanzialmente equilibrata (1.025 vs 972 articoli) tra i due orientamenti. La classificazione delle testate si fonda sul concetto di parallelismo politico (Hallin e Mancini, 2004), caratteristica strutturale del sistema mediatico italiano in cui le testate giornalistiche tendono a riflettere orientamenti ideologici distinti. Tale parallelismo, documentato empiricamente anche nella copertura del tema migratorio (Roncarolo e Mancini, 2018), permane nel contesto italiano contemporaneo nonostante la scomparsa della stampa di partito in senso stretto (Mancini, 2013). La classificazione adottata tiene conto dell'autodichiarazione editoriale, dei legami storici con aree politiche e del posizionamento riconosciuto nel dibattito pubblico. Pur trattandosi di due insiemi relativamente coerenti, si riconosce la presenza di una certa eterogeneità interna ai due gruppi, con orientamenti che spaziano dalla sinistra radicale al riformismo progressista per il primo gruppo, e dal conservatorismo liberale alla destra identitaria per il secondo. Tale classificazione è stata successivamente validata empiricamente mediante l'analisi delle corrispondenze (Fig. 1), che ha confermato la separazione lessicale tra i due gruppi.

Tab. 1 – Testate giornalistiche incluse nel corpus

Area progressista/sinistra	N. articoli	Area conservatrice/destra	N. articoli
Il Manifesto	374	*Il Primato Nazionale*	335
L'Unità	275	*Secolo d'Italia*	212
Left.it	153	*Il Giornale*	187
Il Riformista	102	*Libero*	125
Contropiano	93	*La Voce del Patriota*	68
La Città Futura	28	*Destra.it*	45

L'obiettivo dell'analisi è quello di rilevare convergenze e divergenze nel lessico associato al fenomeno migratorio e di verificare in che misura le scelte linguistiche possano riflettere orientamenti ideologici distinti.

Le procedure di pre-elaborazione e analisi dei dati testuali sono state implementate mediante il pacchetto quanteda per R (Benoit *et al.*, 2018).

In una fase preliminare, il corpus è stato sottoposto a pretrattamento me-

ISBN Open Access: 9788835191971

diante rimozione delle parole funzionali (*stopwords*), applicazione di un filtro di frequenza minima (20 occorrenze complessive) ed eliminazione dei termini presenti in meno di cinque testate. La Figura 1a, risultante dall'applicazione dell'analisi delle corrispondenze semplici al corpus testuale, evidenzia una marcata contrapposizione lungo l'asse orizzontale: le testate giornalistiche di orientamento progressista si distribuiscono sul semiasse negativo, mentre quelle di orientamento conservatore occupano il semiasse positivo. La dispersione all'interno di ciascun gruppo rispetto alla seconda dimensione suggerisce l'esistenza di uno spazio discorsivo bidimensionale in cui convergono sia l'opposizione ideologica sia differenze interne a ciascun campo relative all'intensità della connotazione politica e alla relazione con le tradizioni editoriali consolidate.

Le coordinate delle testate giornalistiche rispetto alle due dimensioni principali sono state successivamente utilizzate per identificare, tramite un algoritmo di clustering gerarchico basato su distanza euclidea, sottogruppi caratterizzati da similarità nella struttura lessicale (Fig. 1b).

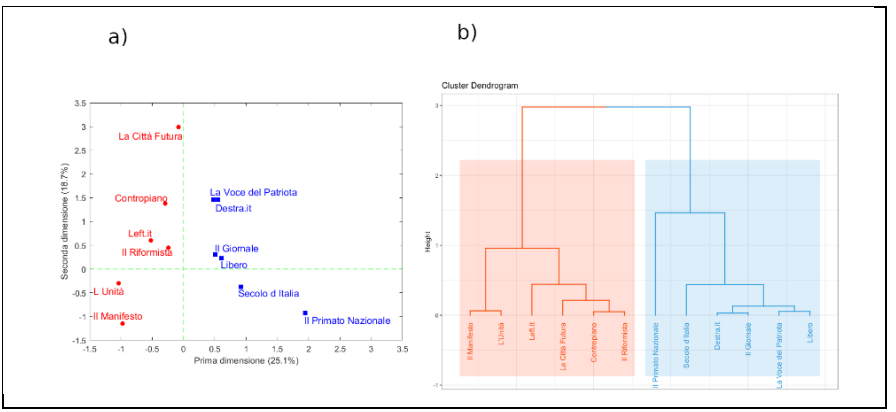


Fig. 1 – Analisi delle corrispondenze semplici e clustering gerarchico delle testate giornalistiche. a) Proiezione delle testate sul piano fattoriale, ottenuta tramite analisi delle corrispondenze sugli articoli raggruppati per giornale. b) Clustering gerarchico delle testate giornalistiche sulla base delle due dimensioni ottenute dall'analisi delle corrispondenze

I risultati della riduzione dimensionale e del clustering gerarchico confermano la classificazione adottata nella distinzione fra area progressista/sinistra e area conservatrice/destra, fornendo un riscontro empirico alla bipartizione del corpus, ma al tempo stesso evidenziano una significativa eterogeneità intragruppo. Nel gruppo progressista, *Il Manifesto* e *L'Unità* formano un nucleo relativamente coeso, mentre *Left.it* occupa una posizione più peri-

ISBN Open Access: 9788835191971

ferica; nel gruppo conservatore, *Il Primato Nazionale* ed *Secolo d'Italia* si collocano a maggiore distanza dalle restanti testate. Tale variabilità lessicale riflette differenze negli orientamenti specifici all'interno di ciascuna area ideologica. I risultati che seguono vanno pertanto interpretati come tendenze prevalenti, riconoscendo l'eterogeneità intragruppo qui documentata.

La wordcloud presentata in Figura 2 consente di osservare con immediatezza la distanza lessicale e concettuale che caratterizza il discorso sulla migrazione nelle testate di orientamento conservatore e progressista. La contrapposizione nell'utilizzo dei termini configura cornici interpretative profondamente divergenti.

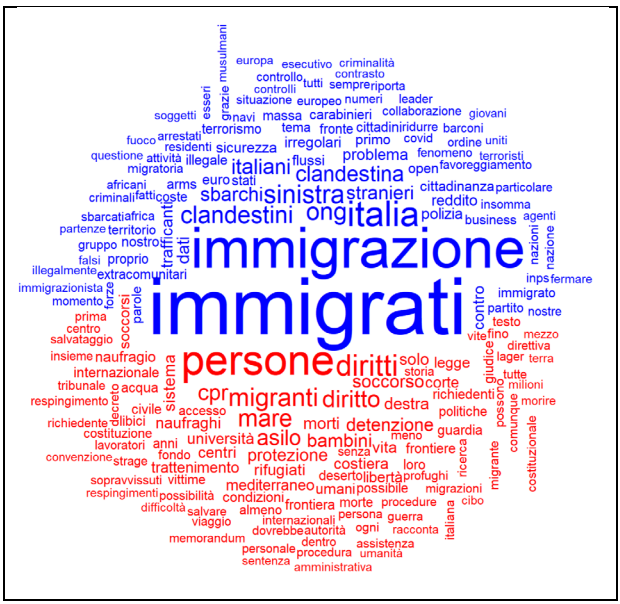


Fig. 2 – Rappresentazione dei termini caratteristici degli articoli della stampa con orientamento politico conservatore/destra (in blu) e progressista/sinistra (in rosso)

Una prima divergenza fondamentale emerge nella denominazione stessa dei soggetti coinvolti: le testate di area conservatrice/destra ricorrono sistematicamente a *immigrati* e *clandestini*, enfatizzando la dimensione amministrativa e la questione della legalità; le testate di area progressista/sinistra privilegiano invece *persone* e *migranti*, termini che sottraggono il fenomeno alla logica dell'irregolarità per riportarlo alla dimensione dell'umanità comune e della neutralità descrittiva.

Questa differenza lessicale di base si inscrive in repertori terminologici più ampi che configurano frame interpretativi radicalmente opposti. Nel lessico delle testate conservatrici emerge un repertorio riconducibile a una

ISBN Open Access: 9788835191971

cornice securitaria ed emergenziale. Vocaboli quali *illegalità*, *criminalità*, *sicurezza*, *controllo*, *frontiere*, *sbarchi*, unitamente a riferimenti a *terrorismo*, *polizia*, *carabinieri* e *respingimenti*, concorrono a costruire l'immigrazione come una minaccia per l'ordine pubblico e per la sicurezza nazionale. La ricorrenza di *clandestini* si combina così con un lessico della repressione e del controllo, rafforzando una rappresentazione del fenomeno in termini di emergenza da contrastare.

Al contrario, il lessico delle testate progressiste si colloca prevalentemente all'interno di una cornice umanitaria e garantista. Termini come *diritti, asilo, bambini, mare, soccorso, protezione, detenzione, naufragio* e *Mediterraneo* evidenziano una focalizzazione sulle condizioni di vulnerabilità dei migranti, sui rischi connessi alle traversate e sulle responsabilità istituzionali nei processi di accoglienza e tutela. In questa prospettiva, la centralità di *persone* si accompagna alla costruzione del migrante come soggetto vulnerabile, necessitante di protezione e di riconoscimento dei propri diritti fondamentali.

3. Mappatura della struttura semantica nel discorso migratorio

Per esaminare sistematicamente le differenze nelle cornici interpretative utilizzate dai due sottogruppi nella rappresentazione del fenomeno migratorio, si è fatto ricorso a un approccio di *Textual Network Analysis* (Segev, 2020; Pronello *et al.*, 2024), una metodologia che integra tecniche di estrazione testuale con strumenti di analisi delle reti per analizzare simultaneamente la struttura e la dimensione semantica di corpora testuali di considerevoli dimensioni. All'interno di questo framework analitico, il testo viene modellizzato mediante *reti semantiche*, ossia grafi in cui i nodi rappresentano unità linguistiche (quali parole o lemmi) e gli archi codificano relazioni di co-occorrenza o di contiguità semantica, permettendo di quantificare la *semantic relatedness* presente nel corpus.

Nel presente studio è stata costruita, per ciascuna testata, una rete semantica $G(k) = \big(V, E(k)\big)$. L'insieme dei nodi V è stato definito come un vocabolario comune di $n = 500$ termini più frequenti presenti in tutte le testate. Tale soglia rappresenta un compromesso tra copertura semantica e parsimonia computazionale: un vocabolario più ampio avrebbe introdotto termini a bassa frequenza con co-occorrenze instabili, mentre un vocabolario più ristretto avrebbe escluso relazioni semantiche rilevanti. L'insieme degli archi $E(k)$ è stato determinato sulla base delle co-occorrenze lessicali rilevate entro una finestra di contesto di 10 parole, valore comunemente adottato nella letteratura sulla Textual Network Analysis (Danowski, 1993; Segev, 2020)

ISBN Open Access: 9788835191971

in quanto consente di catturare relazioni semantiche locali senza introdurre eccessivo rumore derivante da associazioni spurie. Questa procedura consente di costruire una rappresentazione strutturale del lessico caratterizzante ciascuna testata e di effettuare un confronto sistematico tra i due gruppi di grafi, uno corrispondente alle testate di sinistra e uno a quelle di destra, mediante le rispettive reti semantiche medie (*mean semantic networks*), ottenute attraverso l'aggregazione delle proprietà strutturali e informative delle singole reti. Ciascuna rete semantica $G(k), k = 1, \ldots, K$, può essere rappresentata attraverso una matrice Laplaciana definita come $\boldsymbol{L}^{(k)} = \boldsymbol{D}^{(k)} - \boldsymbol{A}^{(k)}$, dove $\boldsymbol{A}^{(k)}$ è la matrice di adiacenza avente come elementi i pesi degli archi (i.e., le co-occorrenze fra le parole) e $\boldsymbol{D}^{(k)}$ è la matrice diagonale che riporta sugli elementi diagonali i gradi dei nodi.

Le matrici $\{\boldsymbol{L}^{(1)}, \boldsymbol{L}^{(2)}, \ldots, \boldsymbol{L}^{(k)}\}$, normalizzate per la traccia, appartengono allo spazio $\mathcal{L}_n$ delle matrici Laplaciane di grafi simmetrici, definito come

$$\mathcal{L}_n = \{\boldsymbol{L} = (l_{i,j}) : \boldsymbol{L} = \boldsymbol{L}^T, l_{i,j} \leq 0 \; \forall i \neq j, \boldsymbol{L}\mathbf{1}_n = \mathbf{0}_n\},$$

dove $\mathbf{1}_n$ e $\mathbf{0}_n$ sono rispettivamente il vettore di n elementi tutti pari a uno ed il vettore nullo. Poiché $\mathcal{L}_n$ è uno spazio non euclideo (Ginestet *et al.*, 2017), l'analisi richiede l'impiego di metriche e procedure geometriche specifiche per l'applicazione di analisi statistiche e la derivazione della rete media. Per tali finalità, nella nostra analisi abbiamo adottato il framework proposto da Severn *et al.* (2022). In particolare, abbiamo considerato la Procustes Power metric che definisce la distanza tra due matrici Laplaciane, $\boldsymbol{L}^{(k)}$ e $\boldsymbol{L}^{(k')}$, come

$$d_{\alpha_y,S}(\boldsymbol{L}^{(k)}, \boldsymbol{L}^{(k')}) = \inf_{R \in \mathcal{O}_n} \left|\left| L^{(k)^\alpha} - L^{(k')^\alpha} R \right|\right|_F$$

dove $\mathcal{O}_n$ è l'insieme delle matrici ortogonali di rotazione, $\|\cdot\|_F$ denota la norma di Frobenius, e $F_\alpha(\boldsymbol{L}) = \boldsymbol{L}_\alpha$ rappresenta l'applicazione che proietta $\mathbf{L}$ nello spazio di embedding. Questa metrica costituisce il punto di partenza per la proiezione delle reti nello spazio tangente, necessaria per il calcolo della media delle matrici Laplaciane in ognuno dei due gruppi di testate giornalistiche. Tali medie sono state poi ritrasformate nello spazio delle Laplaciane e convertite nelle corrispondenti reti semantiche medie rappresentate nelle Figure 3 e 4.

ISBN Open Access: 9788835191971

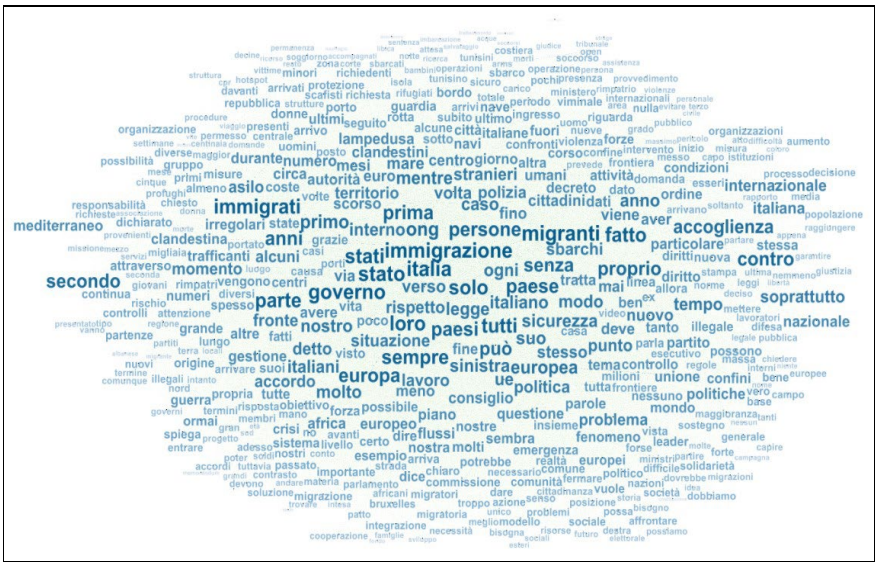


Fig. 3 – Rete semantica media delle testate giornalistiche con orientamento conservatore/destra

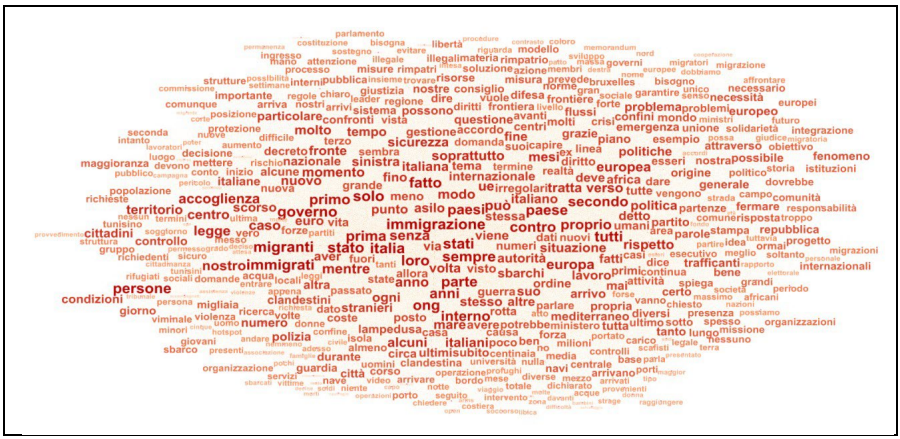


Fig. 4 – Rete semantica delle testate giornalistiche con orientamento progressista/sinistra

Per quantificare le differenze puntuali nella forza delle associazioni semantiche tra i due insiemi di giornali, è stata applicata la z-statistic proposta da Severn *et al.* (2022) che consente di identificare le coppie di parole che mostrano connessioni espresse diversamente nelle reti delle testate giornalistiche con orientamento conservatore/destra e progressista/sinistra.

ISBN Open Access: 9788835191971

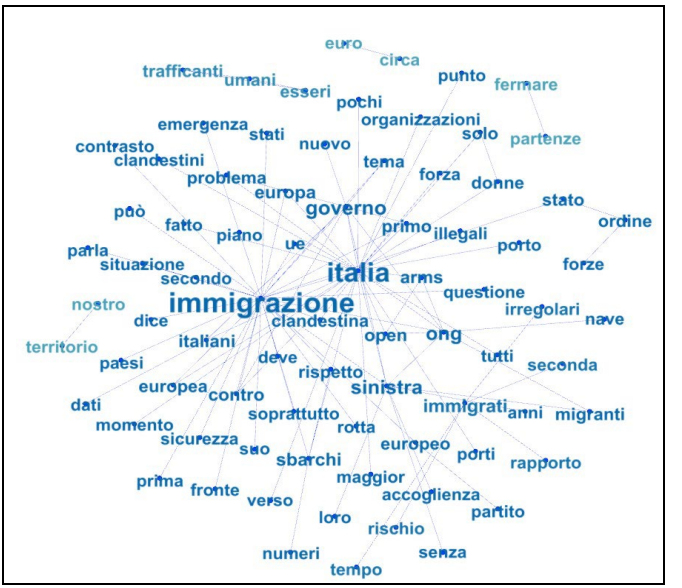


Fig. 5 – Principali associazioni lessicali nelle testate di area conservatrice/destra. La rete mostra le 100 coppie di termini con maggiore significatività statistica (valori più elevati della statistica z)

La rete semantica delle relazioni maggiormente espresse nei quotidiani di area conservatrice (Fig. 5) evidenzia una struttura discorsiva fortemente coesa e organizzata intorno a tre nuclei relazionali principali. Il primo nucleo comprende l'asse *immigrazione–Italia–governo*, da cui si irradiano connessioni verso termini quali *clandestina/i, sbarchi, illegali, irregolari*, delineando un cluster che costruisce socialmente l'immigrazione come *fenomeno* da gestire attraverso *controllo, ordine* e *sicurezza*. La centralità di Italia come nodo connettore rivela una rappresentazione nazionale del fenomeno, rafforzata dalle connessioni con *stato, nostro* e *territorio*, che pone al centro l'identità territoriale. Le co-occorrenze tra *porto, forze e nave* completano questo frame securitario, enfatizzando la dimensione del controllo delle frontiere marittime. Un secondo nucleo è costituito dal triangolo *immigrazione–sinistra–ong*, dove le co-occorrenze mostrano un intreccio sistematico tra immigrazione e attori politici o organizzazioni interpretate come facilitatori del fenomeno. Il legame frequente tra *sinistra* e termini come *accoglienza, open, arms* (riferiti a Open Arms) e *porti* segnala una presentazione della tematica in diretta contrapposizione ad alcune scelte dell'opposizione. Il terzo nucleo configura una costellazione di relazioni che collegano *sbarchi, numeri, emergenza, territorio, sicurezza* e *maggior*, creando un cluster emergenziale-quantitativo in cui l'immigrazione viene descritta come un processo

ISBN Open Access: 9788835191971

in espansione che richiede interventi. La presenza di *traffico, trafficanti e umani* (in "traffico di esseri umani") introduce inoltre la dimensione criminale del fenomeno, mentre termini come *rischio, problema e contrasto* rafforzano la descrizione dell'immigrazione come minaccia.

La densità delle connessioni all'interno di questo cluster e la loro prossimità al nucleo principale rivelano la coesistenza strutturale tra narrazione del rischio e narrazione del controllo. Trasversalmente ai tre nuclei emerge una forte politicizzazione del discorso, evidenziata dalla presenza di termini quali *partito, governo e questione*, che costruiscono socialmente l'immigrazione come arena di conflitto politico-istituzionale. La dimensione europea, rappresentata da termini come *europa/eo, ue e paesi*, colloca il fenomeno in una prospettiva continentale.

Fig. 6 – Principali associazioni lessicali nelle testate di area progressista/sinistra. La rete mostra le 100 coppie di termini con maggiore significatività statistica (valori più elevati della statistica z)

La rete semantica delle relazioni maggiormente espresse nei quotidiani di area progressista (Fig. 6) presenta una configurazione relazionale differente rispetto a quella osservata nelle testate di destra.

Il nodo *persone* funge da centro organizzatore di una densa costellazione di termini, generando un cluster umanitario in cui concetti come *diritti, protezione, minori, bambini, asilo e soccorso* risultano strettamente interconnessi. Le co-occorrenze formano una rete che costruisce l'immigrazione co-

ISBN Open Access: 9788835191971

me un'esperienza individuale e collettiva caratterizzata da bisogni, vulnerabilità e garanzie di tutela. La presenza di termini quali *cittadinanza, sociali, lavoratori e membri* estende il discorso oltre l'emergenza umanitaria verso una prospettiva di integrazione ed inclusione sociale. Un secondo nucleo di legami emerge attorno al tema della detenzione amministrativa e dei CPR (Centri di Permanenza per il Rimpatrio), con connessioni verso termini come *diritti, costituzione, libertà, legale e assistenza*. Questa configurazione indica che la discussione intorno ai centri di permanenza per il rimpatrio è inserita in un contesto semantico fortemente normativo ed orientato alla valutazione della loro compatibilità con principi costituzionali e standard internazionali. La co-occorrenza tra *detenzione* e termini del campo semantico dei diritti costruisce i CPR non come strumenti necessari di controllo, ma come dispositivi problematici che sollevano questioni di legittimità costituzionale. Un ulteriore cluster relazionale ruota intorno al tema del *mare* e delle operazioni di ricerca e soccorso, collegandosi a *salvataggio, soccorso, morti, emergenza, acqua, difficolta, famiglie e responsabilità*. La struttura dei legami suggerisce una narrazione che interpreta la migrazione via mare non come una minaccia o un flusso da bloccare, ma come un processo che coinvolge soggetti vulnerabili ed impone doveri di protezione e intervento. Il Mediterraneo è così rappresentato come spazio di tragedia umanitaria, dove la presenza di termini come *morti* e *naufragio* (implicito nelle connessioni) richiama l'attenzione sul costo umano delle traversate. La connessione tra *mare* e *ong* si colloca in questo frame di soccorso e salvataggio. Un quarto nucleo evidenzia la dimensione politico istituzionale attraverso connessioni tra *stato, governi, paese, politica e termini* quali *responsabilità, garantire, chiedere e rispetto*. Questo cluster rappresenta le istituzioni come soggetti che hanno responsabilità di tutela e protezione. Nel complesso, la configurazione delle relazioni mostra una rete semantica in cui l'immigrazione è concettualizzata come fenomeno umano e sociale, intrecciato a diritti e tutela dei migranti e responsabilità istituzionali, più che come problema di sicurezza o ordine pubblico. L'opposizione strutturale con la rete delle testate denominate di destra è evidente: dove quest'ultima costruisce cluster coesi intorno a *clandestini-sicurezza-contrasto*, la rete di sinistra organizza il discorso intorno a *persone-diritti-protezione*. Significativamente, termini centrali nel discorso conservatore come *clandestini, illegali, sbarchi* (come evento da contrastare) e *numeri* (enfasi quantitativa) risultano assenti o estremamente marginali, mentre concetti assenti nella rete di destra, come *libertà e rifugiati*, occupano qui posizioni centrali, confermando una divergenza radicale nei frame interpretativi impiegati nella rappresentazione del fenomeno migratorio.

ISBN Open Access: 9788835191971

4. Clustering tematico delle reti semantiche

Per identificare i principali nuclei tematici (topic) presenti nelle due aree ideologiche, è stata applicata una fattorizzazione non negativa simmetrica (*Symmetric Nonnegative Matrix Factorization*, *SymNMF*; Kuang *et al.*, 2015) alle matrici di adiacenza medie di ciascun gruppo. Tale metodologia consente di effettuare simultaneamente una riduzione di dimensionalità ed un clustering dei nodi della rete, individuando così gruppi di termini co-occorrenti che costituiscono cluster semantici coerenti (Pronello *et al.*,2024). Data la matrice di adiacenza media normalizzata $\widetilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, la SymNMF effettua un'approssimazione di rango inferiore mediante una matrice non negativa U, risolvendo il seguente problema di minimizzazione:

$$\min_{\mathbf{U}\geq 0} \|\widetilde{\mathbf{A}} - \mathbf{U}\mathbf{U}^{\top}\|_{F}.$$

Il problema di ottimizzazione è stato risolto mediante l'algoritmo ANLS (Alternating Nonnegative Least Squares) (Kim e Park, 2008). Grazie al vincolo di non-negatività, l'elemento maggiore nella k-esima riga di U indica l'assegnazione del k-esimo nodo (termine) al corrispondente cluster, permettendo così di identificare i principali topic che strutturano il discorso sull'immigrazione in ciascun gruppo di testate.

L'analisi ha consentito di individuare 6 topic distinti per ciascun gruppo di testate, rivelando le principali aree tematiche attraverso cui viene costruita la narrazione del fenomeno migratorio nella stampa italiana.

L'analisi dei topic delle testate di area conservatrice (Fig. 7) conferma sistematicamente quanto emerso dalle precedenti analisi basate sulla frequenza delle parole e sulle reti semantiche. Il **Topic 1** (*Dibattito politico sull'immigrazione*) aggrega la macro-narrazione politica in cui il fenomeno è inscritto nello scontro politico interno e collocato sul piano delle responsabilità istituzionali, confermando l'asse *immigrazione-Italia-governo* già osservato come centro organizzatore della rete semantica. Il **Topic 2** (*Flussi, numeri e gestione dell'accoglienza*) enfatizza la dimensione quantitativa ed emergenziale, con particolare attenzione alla pressione sul *territorio* italiano, rispecchiando il cluster emergenziale-quantitativo della rete semantica. Il **Topic 3** (*Politiche europee e patto migratorio*) riguarda la gestione europea dei flussi in termini prevalentemente burocratici e di responsabilità istituzionali. *Richiedenti e asilo* sono inquadrati in un contesto procedurale, confermando la rappresentazione europea come questione di distribuzione e controllo. Il **Topic 4** (*Frame securitario ed emergenza*) cristallizza la dimensione normativa e repressiva che richiede azione legislativa e di contrasto, sintetizzando il nucleo securitario

ISBN Open Access: 9788835191971

centrale della rete semantica. Il **Topic 5** (*Soccorso in mare e critica alle ONG*) rappresenta *ong, navi e porti* come elementi problematici del fenomeno migratorio, confermando il nucleo relazionale *immigrazione-sinistra-ong*. Il **Topic 6** (*Traffico di esseri umani*) rafforza il frame criminale articolato attorno a *trafficanti, tratta e scafisti*, dove *esseri umani* compare principalmente in un contesto di criminalità organizzata.

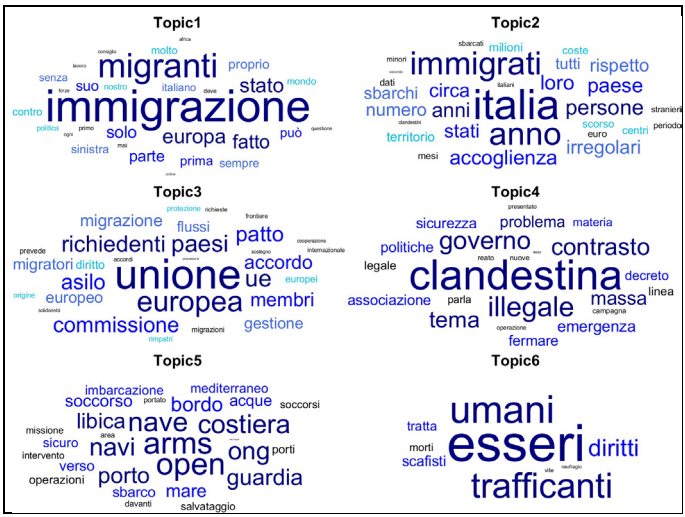


Fig. 7 - Temi associati alle testate giornalistiche di area conservatrice/destra

Anche i topic delle testate di area progressista (Fig. 8) confermano quanto emerso dalle analisi precedenti: il discorso costruisce cornici interpretative orientate alla tutela dei diritti, alla protezione umanitaria e alla dimensione sociale del fenomeno migratorio. Il **Topic 1** (*Dignità umana e diritti fondamentali*) esprime l'impianto valoriale della narrazione, centrata su *umani, persone, diritti e migranti*, operando un'umanizzazione del fenomeno che lo sottrae alla logica classificatoria. Il **Topic 2** (*Governance europea*) colloca l'immigrazione nello scenario europeo in termini di gestione. Il **Topic 3** (*Emergenza umanitaria in mare*) enfatizza la vulnerabilità di *bambini, donne e minori*, insieme alla dimensione tragica del *mare, mediterraneo, morti*, confermando il frame umanitario già osservato nella rete semantica. Il **Topic 4** (*Sistema di accoglienza e CPR*) tratta il sistema di accoglienza e gestione amministrativa. La compresenza di *cpr* e *detenzione* con termini normativi rispecchia il nucleo CPR-diritti-costituzione-libertà. Il **Topic 5** (*Diritto d'asilo e protezione internazionale*) sviluppa la dimensione giuridica dei diritti fondamentali attraverso termini quali *asilo, protezione, rifugiati, costituzione* e *garantire*, confermando un vocabolario normativo orientato alla garanzia. Il **Topic 6** (*Soccorso marittimo e ruolo*

ISBN Open Access: 9788835191971

delle ONG) rappresenta le operazioni di soccorso in mare, dove soccorso, *salvataggio* e *ong* costruiscono le organizzazioni umanitarie come attori necessari, confermando il nucleo *mare-soccorso-salvataggio-responsabilità.*

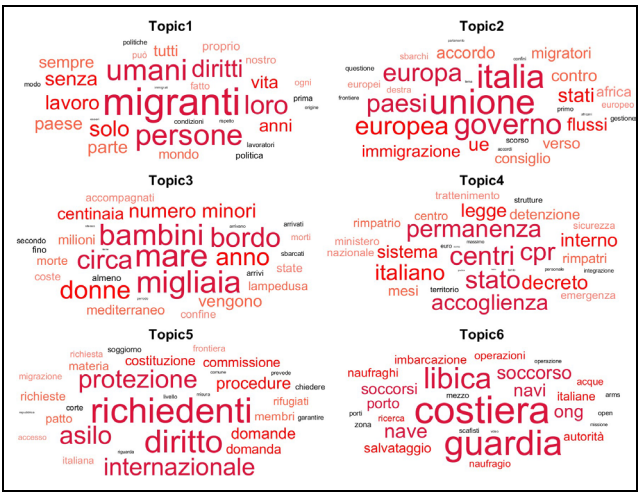


Fig. 8- Temi associati alle testate giornalistiche di area progressista/sinistra

5. Discussione

Le analisi svolte permettono di tracciare un confronto chiaro e sistematico dei frame interpretativi utilizzati nella rappresentazione del fenomeno dell'immigrazione nelle testate giornalistiche qui categorizzate come testate di destra e di sinistra in accordo a quanto descritto in precedenza. La Tabella 2 riporta una sintesi comparativa delle principali dimensioni discorsive che caratterizzano i due orientamenti.

Tab. 2 – Sintesi comparativa dei frame migratori: testate di destra vs testate di sinistra

Dimensione	Area conservatrice/ destra	Area progressista/ sinistra
Soggetti	Immigrati, clandestini, irregolari	Persone, migranti, umani, rifugiati
Frame dominante	Securitario-emergenziale	Umanitario-garantista
Mare	Spazio di controllo, invasione	Spazio di tragedia, morte
ONG	Facilitatori	Attori di soccorso
Vulnerabilità	Assente/marginale	Centrale (bambini, donne, minori)
CPR	Strumenti necessari	Questione di legittimità costituzionale
Asilo	Procedurale/marginale	Diritto fondamentale
Numeri	Invasione quantitativa	Tragedia umanitaria misurabile

ISBN Open Access: 9788835191971

6. Conclusioni

L'insieme delle procedure statistiche applicate al corpus delle notizie apparse sulle testate giornalistiche ha consentito di ricostruire e confrontare le cornici interpretative dell'immigrazione nei due universi discorsivi, evidenziando non solo le parole chiave, ma anche le relazioni semantiche e le strutture profonde che organizzano il discorso mediatico.

I risultati confermano l'ipotesi di partenza: l'orientamento ideologico costituisce un fattore determinante nella costruzione delle cornici interpretative attraverso cui viene rappresentato il fenomeno migratorio. Le analisi condotte, dall'esame del lessico distintivo all'analisi delle reti semantiche, fino alla decomposizione dei topic, convergono nel delineare due narrazioni divergenti e strutturalmente opposte. Le testate conservatrici costruiscono l'immigrazione attraverso frame securitari ed emergenziali. La stessa denominazione dei soggetti coinvolti enfatizza status amministrativo e (il)legalità.

Le testate progressiste presentano frame umanitari, garantisti, orientati ai diritti. La scelta di *persone* e *migranti* opera un'umanizzazione che sottrae il fenomeno alla logica securitaria.

Tali risultati confermano quanto emerso dalla letteratura internazionale sulla polarizzazione dei frame migratori (Benson, 2013; Fuller, 2024), e forniscono evidenze sulla divergenza nel contesto italiano caratterizzato da opposizioni strutturali nei modi stessi di concettualizzare il fenomeno. Le opposizioni fra i due processi di framing, infatti, attraversano tutti i livelli: numeri/invasione vs. vite/tragedia; clandestini/contrasto vs. persone/protezione; minaccia vs. vulnerabilità; controllo vs. soccorso; facilitazione problematica vs. azione salvavita; distribuzione forzata vs. solidarietà.

La divergenza osservata solleva questioni sulla dinamica tra frame dominanti e strategie di contronarrazione. Seguendo la tipologia proposta da Binotto (2022), il discorso progressista può essere interpretato non solo come narrazione alternativa ma come tentativo di *reframing* strategico: piuttosto che limitarsi a opporre al frame securitario componenti contrarie mantenendone le strutture di base (counter-frame), le testate progressiste tentano di riformulare radicalmente i *frames* sostituendo le metafore fondamentali, dalla "invasione" alla "tragedia umanitaria", dai "clandestini" alle "persone", dal "controllo" alla "protezione".

6.1. Limiti e sviluppi futuri

La ricerca presenta alcuni limiti che suggeriscono direzioni per futuri approfondimenti. L'analisi, pur evidenziando divergenze sistematiche tra i due

ISBN Open Access: 9788835191971

orientamenti, rileva anche una significativa eterogeneità interna a ciascun gruppo. La riduzione dimensionale e il clustering gerarchico hanno documentato variabilità lessicali che riflettono differenze all'interno di ciascuno schieramento. La scelta metodologica di operare confronti attraverso medie aggregate, reti semantiche medie e topic modeling su corpus raggruppati, ha consentito di identificare tendenze prevalenti ma ha comportato necessariamente una semplificazione della complessità interna. I risultati vanno pertanto interpretati come pattern distintivi dei due orientamenti ideologici nel loro complesso, riconoscendo che le singole testate possono presentare specificità non pienamente catturate dall'analisi aggregata. Questa impostazione fornisce un quadro metodologico per lo studio comparativo dei frame nella stampa italiana, lasciando al contempo spazio ad approfondimenti dedicati alle particolarità editoriali delle singole pubblicazioni.

L'analisi si concentra inoltre esclusivamente sulla stampa scritta, con un campione di 12 testate selezionate in base all'orientamento ideologico e alla disponibilità di articoli sulle parole chiave prescelte. Restano esclusi media televisivi, piattaforme digitali native e social media, che potrebbero presentare dinamiche discorsive, strategie di framing e modalità di coinvolgimento del pubblico significativamente diverse. Un'estensione dell'analisi a questi contesti mediali e ad altre testate giornalistiche consentirebbe una comprensione più completa dell'ecosistema informativo sul tema migratorio.

Infine, un'estensione necessaria riguarda l'analisi delle voci minoritarie e delle strategie di contronarrazione sviluppate da attori sociali non mainstream, associazioni, movimenti, organizzazioni della società civile e di migranti, per valutare in che misura e attraverso quali modalità riescano a penetrare lo spazio pubblico mediale, a contrastare i frame dominanti o a proporre reframing alternativi (Binotto, 2022).

Riferimenti bibliografici

Bennett, W.L. and Entman, R.M. (2001), *Mediated Politics: Communication in the Future of Democracy*, Cambridge University Press, Cambridge.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S. and Matsuo, A. (2018), quanteda: An R package for the quantitative analysis of textual data, *Journal of Open Source Software*, 3(30): 774.

Benson, R. (2013), *Shaping Immigration News: A French American Comparison,* Cambridge University Press, Cambridge.

Binotto, M. (2022), Countering or Reframing Migrations. Frames, Definitions, Strategies to Imagine New Metaphors and Narrative for the Media Agenda. *Comunicazioni Sociali*, p. 15.

Bleich, E., Bloemraad, I. and de Graauw, E. (2015), Migrants, Minorities and the

ISBN Open Access: 9788835191971

Media: Information, Representations and Participation in the Public Sphere, *Journal of Ethnic and Migration Studies*, 41(6): 857–873.
Caviedes, A. (2015), An Emerging 'European' News Portrayal of Immigration? *Journal of Ethnic and Migration Studies*, 41(6): 897–917.
Chong, D. and Druckman, J. N. (2007), Framing Theory, *Annual Review of Political Science*, 10: 103–126.
Danowski J. A. (1993), Network analysis of message content, *Progress in Communication Sciences*, 12: 198–221.
Eberl, J., Meltzer, C. E., Heidenreich, T., Herrero, B., Theorin, N., Lind, F., Berganza, R., Boomgaarden, H. G., Schemer, C. and Strömbäck, J. (2018), The European Media Discourse on Immigration and its Effects: A Literature Review, *Annals of the International Communication Association*, 42(3): 207–223.
Entman, R. M. (1993), Framing: Toward Clarification of a Fractured Paradigm, *Journal of Communication*, 43(4): 51–58.
Flinz, C. and Leonardi, S. (2023), The Representation of Refugees and Migrants in the Italian Media Discourse. In Fabián, A., ed., *The Representation of Refugees and Migrants in European National Media Discourses from 2015 to 2017*, Linguistik in Empirie und Theorie / Empirical and Theoretical Linguistics. J.B. Metzler, Berlin, Heidelberg.
Fuller, J. M. (2024), Media discourses of migration: A focus on Europe, *Language and Linguistics Compass*, 18(4): e12526.
Gamson, W. A. and Modigliani, A. (1989), Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach, *American Journal of Sociology*, 95(1): 1–37.
Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S. and Kolaczyk, E. D. (2017), Hypothesis testing for network data in functional neuroimaging, *The Annals of Applied Statistics*, 725–750.
Goffman, E. (1974), *Frame Analysis: An Essay on the Organization of Experience*, Harvard University Press, Cambridge, MA.
Hallin, D.C. and Mancini, P. (2004), *Comparing Media Systems: Three Models of Media and Politics*, Cambridge University Press, Cambridge.
Kim, H. and Park, H. (2008), Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method, *SIAM Journal on Matrix Analysis and Applications*, 30(2): 713–730.
Koopmans, R. and Statham, P., eds. (2010), *The Making of a European Public Sphere,* Cambridge University Press, Cambridge.
Kuang, D., Yun, S. and Park, H. (2015), SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering, *Journal of Global Optimization*, 62(3): 545–574.
Mancini, P. (2013), The Italian public sphere: a case of dramatized polarization, *Journal of Modern Italian Studies*, 18(3): 335–347.
Mancini, P., Mazzoni, M., Barbieri, G., Damiani, M. and Gerli, M. (2021), What shapes the coverage of immigration, *Journalism*, 22(4): 845–866.
McCombs, M. (2004), *Setting the Agenda: The Mass Media and Public Opinion*, Polity Press, Cambridge.

ISBN Open Access: 9788835191971

Nelson, T. E., Oxley, Z. M. and Clawson, R. A. (1997), Toward a psychology of framing effects. *Political Behavior*, 19(3): 221-246.

Pronello, N., Cucco, A., del Gobbo, E., Fontanella, S. and Fontanella, L. (2024), Dynamics of Online Debates: Insights from Textual Network Analysis, *Annals of Operations Research*.

Roncarolo, F. and Mancini, P. (2018), The traditional media, political parallelism and public opinion on contentious issues in the 2018 Italian election campaign. *Contemporary Italian Politics*, 10(3): 243–266.

Schroter, J. (2023), The Austrian press discourse on refugees, migrants, and migration: A corpus linguistic approach. In *The Representation of Refugees and Migrants in European National Media Discourses from 2015 to 2017: A Contrastive Approach Corpus Linguistics* (pp. 23–66). Springer, Berlino.

Segev, E. (2020), Textual network analysis: Detecting prevailing themes and biases in international news and social media, *Sociology Compass*, 14(4).

Severn, K. E., Dryden, I. L. and Preston, S. P. (2022), Manifold valued data analysis of samples of networks, with applications in corpus linguistics, *The Annals of Applied Statistics*, 16(1): 368-390.

Troszyński, M. and El-Ghamari, M. (2022), A Great Divide: Polish media discourse on migration, 2015–2018, *Humanities and Social Sciences Communications*, 9: 27.

Valenzuela-Vergara, E. M. (2019), Media Representations of Immigration in the Chilean Press: To a Different Narrative of Immigration? *Journal of Communication Inquiry*, 43(2): 129-151.

Wojahn, D. (2023), Refugees, migrants and asylum seekers in the Swedish media discourse: The discursive construction of mobile humans during the so-called European refugee crisis, 2015–2017. In The *Representation of Refugees and Migrants in European National Media Discourses from 2015 to 2017: A Contrastive Approach Corpus Linguistics* (pp. 307–346). Springer, Berlino.

ISBN Open Access: 9788835191971

Strumenti lessicali per la decostruzione dell'odio. Dizionari tematici e ontologie di dominio nell'analisi del razzismo e della xenofobia online

di *Alex Cucco*[*], *Lara Fontanella*[*], *Annalina Sarra*[*], *Mario Monteleone*[**]

1. Introduzione

L'identificazione di razzismo e xenofobia in contenuti testuali online costituisce oggi una delle sfide socio-computazionali più complesse, data la crescente sofisticazione delle strategie linguistiche attraverso cui tali atteggiamenti ostili si articolano. A differenza dei discorsi di odio espliciti e immediatamente riconoscibili, le forme contemporanee di ostilità si diffondono attraverso un linguaggio ibrido, allusivo e strategicamente mimetizzato, che sfrutta narrazioni implicite e un lessico in continua trasformazione (Agudelo e Olbrych, 2022; Rubio-Carbonero, 2020).

In questo scenario, gli approcci puramente lessicali basati su liste statiche di termini, sebbene rappresentino un punto di partenza essenziale, mostrano i propri limiti. Risorse consolidate come HurtLex (Basile, 2019) e la sua revisione recente (Tontodimamma *et al.*, 2023) necessitano di aggiornamenti ed integrazioni per adeguarsi alla polisemia, all'evoluzione diacronica dei lemmi e, soprattutto, alle complesse relazioni semantiche che strutturano le narrative ostili, spesso riconducibili al cosiddetto *ambient racism* (razzismo diffuso), inteso come forma di ostilità pervasiva e normalizzata nel discorso quotidiano (Sharma, 2018).

Alla luce di tali criticità, Cucco *et al.* (2025) propongono un percorso metodologico che integra risorse complementari, quali un dizionario tematico e un'ontologia di dominio, trasformandole in un ecosistema analitico coerente, capace di ampliare il repertorio lessicale, contestualizzarlo empiricamente e formalizzarlo entro una struttura concettuale esplicita.

[*] Dipartimento di Studi Socio-economici, Gestionali e Statistici, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, alex.cucco@unich.it; lara.fontanella@unich.it; annalina.sarra@unich.it

[**] Dipartimento di Scienze Politiche e della Comunicazione, Università di Salerno, mmonteleone@unisa.it

ISBN Open Access: 9788835191971

L'articolazione del presente contributo segue una progressione metodologica che muove dal livello strettamente lessicale fino alla formalizzazione concettuale e all'applicazione computazionale. Il punto di partenza è rappresentato dalle metodologie per la costruzione di risorse lessicali finalizzate all'identificazione del discorso d'odio, con particolare attenzione alle risorse disponibili per la lingua italiana. Si procede quindi con la revisione sistematica di HurtLex, una risorsa lessicale italiana, evidenziandone i criteri di espansione, categorizzazione e graduazione dell'offensività, per poi presentare un'estensione data-driven del lessico, orientata all'individuazione di contenuti razzisti e xenofobi attraverso un approccio probabilistico basato su topic models. Il percorso prosegue introducendo l'ontologia di dominio come strumento per strutturare le relazioni semantiche e narrative che organizzano il discorso ostile, mostrando il passaggio dalla semplice lista di termini a un modello concettuale esplicito. Infine, vengono illustrate le potenzialità applicative di tali risorse nell'identificazione computazionale del linguaggio razzista tramite NooJ, evidenziando come l'integrazione tra dizionario e ontologia possa supportare strategie di riconoscimento più robuste e interpretabili.

2. I dizionari tematici per l'individuazione di ostilità nei contenuti testuali

Il dizionario tematico costituisce uno dei fondamenti empirici essenziali delle analisi computazionali del discorso d'odio.

Nel contesto più ampio dell'analisi del *sentiment* e dell'*opinion mining*, esistono tre approcci principali per la costruzione di lessici: elaborazione manuale, metodi basati su dizionari e metodi basati su corpora (Almatarneh e Gamallo, 2018). I metodi basati su dizionari partono da una lista iniziale di *seeds* (termini-seme) ed effettuano un'espansione automatica utilizzando termini tratti da risorse esterne quali i dizionari dei sinonimi. Gli approcci basati su corpora utilizzano anch'essi liste di *seeds* per identificare altri termini e la loro polarità all'interno di un corpus dato, costruendo così lessici specifici di dominio.

Nell'ambito della creazione di risorse lessicali computazionali finalizzate all'identificazione di *hate speech* o di contenuti tossici sono stati utilizzati diversi approcci. Alcuni studi ricorrono direttamente a liste di termini denigratori e offensivi disponibili online (cfr. la rassegna di Schmidt e Wiegand, 2017 e Poletto *et al.*, 2021). Altri creano liste ad hoc, come *l'Insulting and Abusing Language Dictionary* di Razavi *et al.* (2010), successivamente affinate tramite apprendimento adattivo. Gitari *et al.* (2015) hanno compilato

ISBN Open Access: 9788835191971

una lista di "*hate verbs*" che incitano alla violenza, mentre Wiegand *et al.* (2018) hanno affrontato il problema di distinguere parole offensive da espressioni semplicemente negative, partendo dal *Subjectivity Lexicon* (Wilson, Wiebe e Hoffmann, 2005) ed espandendolo tramite annotazione c*rowdsourced* e classificazione automatica. Un contributo significativo è rappresentato dal *Multilingual Offensive Lexicon* (MOL) di Vargas, Carvalho e Rodrigues de Góes (2021) e Vargas *et al.* (2024), un lessico contestuale e cross-linguistico composto da 1.000 espressioni offensive e blasfeme esplicite ed implicite, estratte manualmente da corpora di hate speech e commenti offensivi sui social media e annotate da tre annotatori indipendenti. Il lessico è disponibile in sei lingue, incluso l'italiano, ed è caratterizzato dall'annotazione di informazioni contestuali che distinguono termini dipendenti e indipendenti dal contesto.

Maronikolakis, Köksal e Schütze (2024) hanno introdotto il lessico multilingue *HATELEXICON*. Per la costruzione di questo lessico, gli autori hanno coinvolto annotatori madrelingua con il compito di fornire epiteti denigratori (*slur*), denominazioni di gruppi bersaglio e parole neutre frequenti in contesti d'odio online. La raccolta dei termini è stata affidata alla conoscenza socioculturale degli annotatori, che potevano attingere liberamente da diverse fonti, inclusi i social media. I termini, redatti in portoghese brasiliano, inglese, tedesco, hindi o swahili, comprendono: (i) epiteti denigratori riferiti all'identità (etnia, religione, orientamento sessuale); (ii) denominazioni di gruppi bersaglio, come comunità religiose o marginalizzate; (iii) parole neutre ricorrenti in contesti d'odio.

Ibrahim *et al.* (2024) hanno sviluppato lessici multilingue per lingue africane, dimostrando l'importanza di risorse linguistiche specifiche per contesti culturali diversi. Ali, Blackburn e Stringhini (2025) hanno proposto un framework adattivo per l'identificazione di hate speech in tempo reale. A differenza dei metodi tradizionali basati su lessici statici e precostruiti, il loro approccio ibrido proposto, che combina BERT con tecniche lessicali, utilizza *word embeddings* per aggiornare dinamicamente i lessici, consentendo di identificare *slur* emergenti e varianti ortografiche intenzionali usate per eludere i sistemi di rilevamento.

Per quanto riguarda l'italiano, Maisto *et al.* (2017) hanno proposto una lista di parole ed espressioni tabù per il rilevamento di linguaggio offensivo sui social media. Una risorsa più ampia è rappresentata da HurtLex (Bassignana, Basile e Patti, 2018; Basile 2019), un lessico multilingue costruito semi-automaticamente a partire dal lessico italiano "*Le parole per ferire*" compilato manualmente dal linguista Tullio De Mauro (2016). HurtLex è stato successivamente tradotto in oltre 50 lingue e utilizzato in numerosi studi (Monnar, Perez Rojas e Labra, 2024). L'architettura di *HurtLex*, artico-

ISBN Open Access: 9788835191971

lata in diciassette categorie gerarchiche, include insulti espliciti, riferimenti a gruppi vulnerabili, discriminazioni basate su caratteristiche personali e termini legati alla violenza fisica o psicologica.

2.1. La revisione sistematica di HurtLex: espansione, categorizzazione e graduazione dell'offensività

La revisione di Tontodimamma *et al.* (2023) sul lessico in italiano ha proposto un aggiornamento di HurtLex attraverso una procedura articolata in quattro fasi. Nella prima fase, il lessico è stato ripulito eliminando termini non offensivi, parole non utilizzate in italiano e voci prive di significato. Sono stati inoltre aggiunti nuovi termini individuati attraverso fonti multiple: sinonimi dei termini già presenti, liste di insulti disponibili online e un'analisi sistematica del vocabolario di tre corpora (*Italian Hate Speech Corpus*, Sanguinetti *et al.*, 2018; *AMI Corpus on Misogyny*, Fersini, Nozza e Rosso, 2020; e un corpus proprietario di 433.003 commenti estratti da Facebook, YouTube e Twitter). Nella seconda fase, ogni termine è stato espanso includendo tutte le forme grammaticali rilevanti: singolare e plurale (maschile e femminile) per nomi e aggettivi, e le forme verbali più frequentemente utilizzate in contesti offensivi. La terza fase ha comportato un raffinamento categoriale, con l'aggiunta di quattro nuove categorie alle classificazioni esistenti di HurtLex: termini sessisti, parole intimidatorie, insulti politici e riferimenti a malattie. Infine, nella quarta fase, l'offensività di ciascun termine è stata graduata attraverso uno strumento di misurazione somministrato ad un gruppo selezionato di valutatori. Il lessico finale comprende 7920 forme flesse (7024 nomi/aggettivi, 576 verbi, 320 interiezioni), validate empiricamente tramite annotazione umana su larga scala con 81 valutatori.

L'elemento metodologico più rilevante proposto da Tontodimamma *et al.* (2023) consiste nell'applicazione di modelli di risposta all'item (*Item Response Theory*, de Ayala, 2009) per derivare un punteggio di offensività per ogni lemma inserito nel dizionario finale. In particolare, gli autori modellano esplicitamente la probabilità di risposta agli item attraverso un *Graded Response Model* unidimensionale (Samejima, 1969), nel quale il tratto latente è associato alle parole mentre i parametri degli item sono riferiti ai valutatori. Più specificamente, il modello assume che il punteggio assegnato ad un dato lemma dipenda congiuntamente dal suo livello latente intrinseco di offensività e da alcuni parametri specifici del valutatore. Questi parametri rappresentano, da un lato, il peso della valutazione del singolo annotatore nell'indice composito di offensività e, dall'altro, i valori di soglia tra categorie consecutive lungo la sua scala continua di valutazione.

ISBN Open Access: 9788835191971

Questo approccio ha permesso di: 1) valutare la coerenza intra- e inter-annotatore, identificando ed escludendo i valutatori meno affidabili; 2) assegnare un punteggio di offensività quantitativo a ciascun termine, con valori più elevati che corrispondono prevalentemente a ingiurie sessiste; 3) validare il lessico su corpora annotati, dimostrando una correlazione positiva tra punteggio di offensività e presenza effettiva di hate speech. La risorsa risultante, pubblicamente disponibile[1], è stata testata con successo in esperimenti di classificazione supervisionata, dove l'integrazione delle feature del Revised HurtLex con modelli BERT ha consentito di migliorare le performance nella rilevazione dell'hate speech (Tontodimamma *et al.*, 2023).

2.2. L'estensione data-driven del dizionario tematico per l'identificazione di contenuti razzisti e xenofobi

Focalizzandosi sull'identificazione di contenuti strettamente razzisti e xenofobi, Cucco *et al.* (2025) propongono un approccio metodologico ibrido per la derivazione di un lessico tematico che integra la risorsa lessicale Revised HurtLex con tecniche probabilistiche data-driven. Seguendo le linee indicate da Colace *et al.* (2016) e Huang *et al.* (2021), gli autori utilizzano la Latent Dirichlet Allocation (LDA, Blei, Ng e Jordan, 2003) per estrarre termini rilevanti di dominio. Nello specifico, adottano il modello Seeded Latent Dirichlet Allocation (seededLDA, Jagarlamudi, Daumé e Udupa, 2012), una variante di LDA che incorpora conoscenza a priori attraverso l'uso di seeds predefiniti associati a concetti chiave. A differenza del *topic modeling* tradizionale, Seeded LDA guida attivamente il processo di formazione dei topic verso cluster semanticamente coerenti.

Cucco *et al.* (2025) applicano questo modello ad un ampio corpus di commenti sui migranti (n=185.734) raccolti da Facebook, Instagram e YouTube. Come seed, utilizzano i termini estratti dalla categoria "*slur etnici con stereotipi negativi*" del Revised HurtLex, guidando così il modello nell'identificazione semi-supervisionata di nuovi termini razzisti e xenofobi semanticamente correlati, ampliando progressivamente il repertorio lessicale disponibile. I termini-seed guidano l'estrazione di un topic mirato all'individuazione di contenuti ostili nei confronti dei migranti, a cui si affianca un topic residuale non guidato che cattura il discorso generale non direttamente riconducibile allo spazio concettuale predefinito.

In questo approccio, ciascun documento è modellato come una mistura dei due topic: il topic guidato, orientato verso l'insieme predefinito di termini

[1] https://github.com/valeriobasile/hurtlex/tree/master/revised_hurtlex/IT

ISBN Open Access: 9788835191971

associati al discorso offensivo anti-migranti, e il topic residuale, che cattura contenuti generali o non correlati. Per classificare i commenti come pertinenti al topic offensivo anti-migranti, è stato adottato un approccio probabilistico basato su una soglia: un documento viene assegnato al topic anti-migranti se la sua probabilità a posteriori per quel topic supera 0,5, altrimenti è assegnato al topic residuale, coerentemente con la natura binaria della classificazione. Sulla base di questo criterio, circa il 44,4% dei commenti è stato classificato nel topic guidato. La Figura 1 (riadattata da Cucco *et al.*, 2025) illustra la composizione lessicale di entrambi i topic, mostrando i 100 termini più fortemente associati a ciascuno di essi.

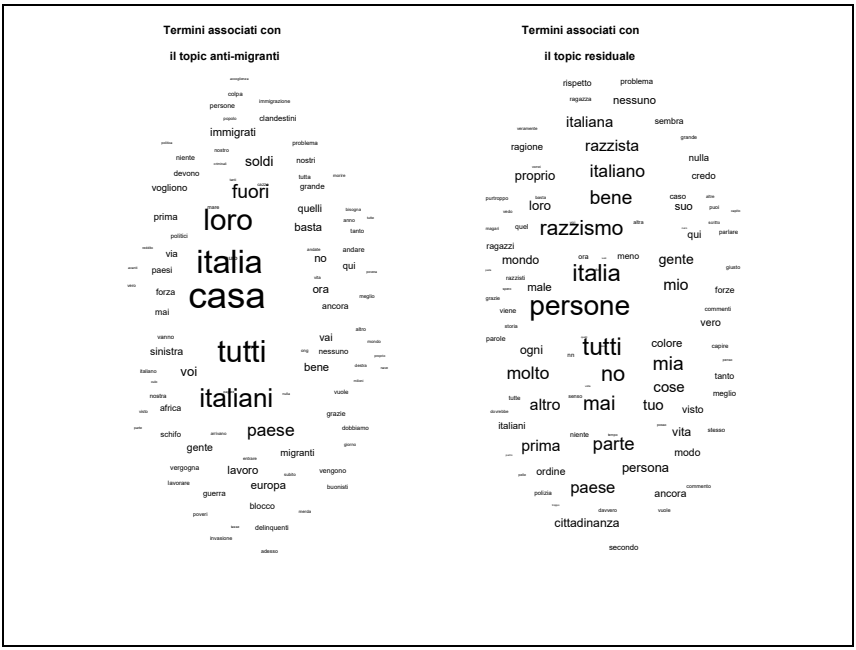


Fig. 1 – Word cloud che mostrano i 100 termini principali associati al topic anti-migrante e al topic residuale. La dimensione delle parole riflette le probabilità parola-topic stimate tramite il modello Seeded LDA

Per approfondire il contenuto lessicale associato al topic anti-migranti, i commenti corrispondenti sono stati categorizzati in cinque classi sulla base della probabilità del topic stesso di essere presente nel documento: (0.5-0.6), (0.6-0.7), (0.7-0.8), (0.8-0.9), (0.9-1). Sulla base di tale classificazione sono state individuate le parole chiave (Gabrielatos, 2018) di ogni sottogruppo di documenti. All'aumentare della probabilità di assegnazione dei documenti al topic anti-migranti (dalle classi con probabilità inferiore a quelle con proba-

ISBN Open Access: 9788835191971

bilità più elevata), si osserva una chiara intensificazione sia nella focalizzazione lessicale sia nella carica emotiva (Fig. 2; riadattata da Cucco *et al.*, 2025). La Classe 1 presenta un linguaggio neutrale, focalizzato su aspetti strutturali e geopolitici della migrazione. La Classe 2 introduce preoccupazioni economico-politiche, inquadrando la migrazione come questione di governance e risorse pubbliche, pur mantenendo un tono non apertamente ostile. Nelle Classi 3-5, il discorso si intensifica progressivamente con l'emergere di termini esplicitamente ostili ed escludenti (*clandestini, parassiti, mandateli fuori, lanciafiamme*), che enfatizzano sovranità nazionale, competizione per le risorse e percezione di minaccia, culminando nella Classe 5 con espressioni fortemente avversative (*schifo, blocco*).

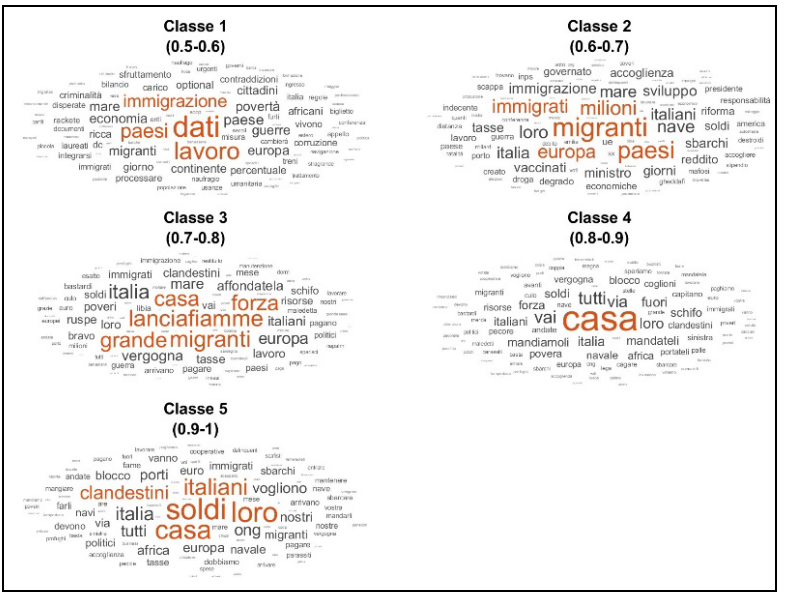


Fig. 2 – Word cloud che illustrano le parole chiave all'interno dei sottogruppi di commenti, segmentati in base a diversi livelli di soglia delle probabilità documento-topic per il topic offensivo anti-migranti.

Questo approccio metodologico consente di arricchire sistematicamente la risorsa lessicale originaria attraverso l'identificazione di termini emergenti e contestualmente rilevanti, catturando le sfumature e le gradazioni progressive del linguaggio razzista e xenofobo così come si manifesta nei contesti comunicativi reali dei social media.

ISBN Open Access: 9788835191971

3. L'ontologia di dominio: dalla dimensione lessicale alla struttura concettuale del discorso razzista

Per strutturare e formalizzare efficacemente la conoscenza lessicale estratta è necessario integrare un quadro definitorio che comprenda concetti intelligibili e relazioni esplicitamente delineate tra di essi. L'*ontology learning* è il processo attraverso cui si deriva una rappresentazione strutturata della conoscenza di dominio estraendo i concetti chiave, le loro definizioni e le relazioni che li legano dai dati disponibili. Un sottoinsieme rilevante di questo campo è l'*Ontology Construction from Texts* (OCT) (Tissaoui *et al.*, 2022), che si concentra specificamente sull'estrazione di elementi ontologici da fonti testuali non strutturate. Questo processo comporta l'identificazione di termini, concetti, relazioni e assiomi dal testo, utilizzandoli per costruire o aggiornare ontologie. Gli approcci statistici all'OCT (Wong, Liu e Bennamoun, 2012), operanti principalmente a livello sintattico e basati sul presupposto che forme sistematiche di co-occorrenza delle parole offrano informazioni significative sulle relazioni semantiche, risultano particolarmente utili nelle fasi di identificazione dei termini e di formazione delle gerarchie.

Concentrandosi sulle parole chiave caratterizzanti i sotto-corpora di commenti associati alle cinque classi identificate attraverso la procedura descritta nella Sezione 2.2, Cucco *et al.* (2025) hanno costruito una rete semantica basata su co-occorrenze per esplorare le relazioni tra i termini salienti. L'applicazione dell'algoritmo di *community detection* di Louvain (Blondel *et al.*, 2008) ha consentito di individuare cluster tematici coerenti che rivelano l'architettura concettuale del discorso razzista: dalla disumanizzazione alle narrative economiche, fino ai frame geopolitici dell'invasione e della sicurezza nazionale.

Le principali categorie di termini offensivi che caratterizzano tali cluster sono sintetizzate nella Tabella 1, insieme a esempi rappresentativi che ne illustrano le funzioni semantiche. Questi raggruppamenti forniscono la base empirica per una formalizzazione ontologica in cui i cluster tematici identificati diventano classi ontologiche organizzate in una tassonomia gerarchica. Le relazioni semantiche si codificano come proprietà, esprimendo connessioni quali la rappresentazione dei migranti attraverso metafore di contaminazione e degrado (*immondizia, luridi*), la retorica parassitaria che li dipinge come sfruttatori delle risorse nazionali (*parassiti, sanguisughe*), o la retorica invasiva e identitaria che li inquadra come minaccia all'identità nazionale (*invasori, contaminati*).

L'introduzione di assiomi e vincoli logici modella quindi la struttura argomentativa delle narrative ostili. Termini come *parassita* o *invasore* cessano di essere semplici etichette denigratorie per diventare concetti formaliz-

ISBN Open Access: 9788835191971

zati con proprietà e implicazioni precise: il primo si associa allo sfruttamento improduttivo delle risorse pubbliche e al peso economico, il secondo alla violazione della sovranità nazionale e alla contaminazione identitaria. Il discorso razzista e xenofobo emerge così come sistema concettuale strutturato, dotato di logica interna e capace di supportare inferenze che giustificano atteggiamenti escludenti e ostili.

Tab. 1 – Categorie di termini offensivi identificati nel discorso offensivo anti-migrante, con esempi che illustrano le loro funzioni semantiche ed espressioni tipiche

Categorie di termini	Esempi
Incitamento alla violenza fisica – esortano esplicitamente a compiere atti violenti o eliminare i migranti	sparategli, bruciamoli, fucilateli
Metafore di contaminazione e degrado – rappresentano i migranti come sporcizia, rifiuti o elementi inquinanti	immondizia, luridi, vomitevoli
Disumanizzazione animalesca – negano la dignità umana equiparando i migranti ad animali	bestie, scimmie, scarafaggio
Stigmatizzazione criminale – associano i migranti a devianza, criminalità e minaccia all'ordine sociale	ladri, malviventi, molestatori
Imperativi espulsivi – esortano all'allontanamento forzato o all'espulsione dei migranti	cacciateli, rimandateli, buttiamoli
Retorica parassitaria – dipingono i migranti come sfruttatori improduttivi delle risorse nazionali	parassiti, sanguisughe, fannulloni
Slur etnico-razziali – epiteti denigratori basati su caratteristiche etniche o razziali	negri, negretti, negracci, zulù (usato in senso dispregiativo)
Incitamento genocida – evocano violenza collettiva o sterminio attraverso immaginario genocida	forni, lanciafiamme, uccidetevi, massacrateli
Retorica invasiva e identitaria – rappresentano i migranti come invasori che minacciano l'identità nazionale	invasori, invaso, contaminati
Insulti generici disprezzanti – espressioni denigratorie generali che manifestano disprezzo	bastardi, feccia, gentaglia

Riadattata da Cucco et al., 2025

4. Le potenzialità di NooJ per l'identificazione computazionale del linguaggio razzista

Le risorse lessicali precedentemente descritte possono essere utilizzate per l'identificazione computazionale di contenuti offensivi nei confronti dei migranti tramite procedure di linguistica computazionale. In particolare, l'impiego di *NooJ* (Silberztein, 2015) nell'identificazione del linguaggio razzista trascende la semplice categorizzazione dei termini, configurandosi come un sistema di *Sentiment Analysis* avanzata capace di mappare l'architettura della conoscenza ostile attraverso la formalizzazione di relazioni logiche e semantiche.

NooJ è un ambiente di *Natural Language Processing* (NLP) di tipo rule-

ISBN Open Access: 9788835191971

based che effettua l'analisi automatica di testi e corpora grazie alla formalizzazione e applicazione delle proprietà morfologiche, grammaticali e sintatticosemantiche di una data lingua. Va sottolineato che NooJ rappresenta l'applicazione pratica di una specifica metodologia linguistica descrittiva, ovvero il Lessico-Grammatica, elaborato da Maurice Gross (1975, 1984) negli anni settanta all'Université Paris 7 (Parigi, Francia) con lo scopo di descrivere dettagliatamente tutti i meccanismi combinatori di governance e dipendenza morfosintattiche che, data una lingua, permettono di assemblare le entrate di un lessico in frasi e discorsi grammaticalmente accettabili in senso saussuriano. Per il Lessico-Grammatica, tali meccanismi combinatori vanno sotto il nome di regole di co-occorrenza e restrizioni di selezione. In NooJ, tali proprietà sono immagazzinate e disponibili in due specifici tools, i dizionari elettronici e le grammatiche locali[2]. I dizionari elettronici di NooJ vengono elaborati e applicati in forma di automi e trasduttori a stati finiti (Gross, 1989a), ossia modelli computazionali costituiti da un insieme finito di stati e transizioni che permettono di riconoscere e generare sequenze linguistiche (Gross, 1989b), e includono in modo sincronico e tassonomico tutte le unità linguistiche atomiche (ULA) di una data lingua, ossia parole semplici e composte corredate da metadati, cioè informazioni strutturate che descrivono le proprietà delle entrate lessicali, sotto forma di etichette univoche e non ambigue, riportanti indicazioni di tipo grammaticale, morfosintattico e semantico per ogni singola entrata. Le grammatiche di NooJ sono invece set di istruzioni formali che riutilizzano le informazioni dei dizionari elettronici per effettuare l'analisi automatica e la disambiguazione di testi e corpora. Sono dette locali perché vengono elaborate per descrivere singoli schemi strutturali e profili morfosintattici di una data lingua, come per esempio quelli relativi ai verbi dativi, di movimento, con costruzione a specchio (standard incrociata). Le grammatiche locali costruibili con NooJ sono quelle classiche della tipologia formale: regolari, indipendenti dal contesto, dipendenti dal contesto, e non ristrette. Oltre che come automi a stati finiti, esse vengono elaborate e applicate come trasduttori a stati finiti, ovvero come grammatiche in grado di localizzare specifiche parti di testo e trascriverle in base alle istruzioni immagazzinate.

L'applicazione di tali grammatiche a testi e corpora può avvenire in modo deterministico, non deterministico, ricorsivo, a cascata, e come automi di automi. Inoltre, in NooJ è possibile creare grammatiche ontologizzate (Monteleone, 2019), in cui uno specifico concetto viene rappresentato in forma di *Linguistic Linked Data* (Chiarcos, Nordhoff e Hellmann, 2012), ovvero da ULA fra loro connesse in termini di prossimità semantica e/o terminologica, di

[2] Non essendo qui possibile, per motivi di spazio a disposizione, esporre nella sua completezza tutte le funzioni e le modalità di analisi offerte da NooJ, per una completa descrizione di questo software si rimanda alla lettura di Silberztein (2015).

ISBN Open Access: 9788835191971

rapporto logico-deduttivo e attinenza concettuale. Questo ultimo tipo di grammatiche è usato per individuare uno o più concetti trattati in testi e corpora, e in ultima analisi permette di effettuare sia una accurata Sentiment Analysis che l'estrazione e la rappresentazione automatiche della conoscenza. Per tale motivo, NooJ è stato spesso usato come *corpus processor* in diversi settori della conoscenza, ovvero, oltre quello della Linguistica, in Storia, Psicologia, Letteratura, Data Mining nonché per l'interpretazione delle frasi musicali.

Un ulteriore aspetto importante della Sentiment Analysis effettuabile con le grammatiche di NooJ è la ricorsività applicativa. Infatti, combinando in grammatiche e trasduttori sia espressioni razionali che istruzioni morfosintattiche contestualizzate, ovvero combinando i quattro tipi di analisi formali precedentemente indicati, è possibile creare grammatiche ad ampia copertura, applicabili a testi diversi, ottenendo risultati simili se non identici.

L'identificazione del linguaggio razzista con NooJ può di fatto essere definita sia come tipologia di Sentiment Analysis sia come estrazione e rappresentazione automatiche della conoscenza. A tale scopo, il corpus utilizzato per impostare e portare a conclusione le procedure di Sentiment Analysis con NooJ è una porzione normalizzata del corpus già utilizzato per le altre analisi NLP, dalla quale sono state eliminate tutte le forme non alfabetiche di notazione e commento, inclusi emoji e altri tipi di immagini. Una volta importato in NooJ, il corpus viene *tokenizzato*, producendo una lista di occorrenze di tutte le parole formali del testo. I risultati della *tokenizzazione* vengono usati principalmente per delineare e localizzare i concetti più utilizzati nel corpus analizzato. In aggiunta, e a margine della tokenizzazione, NooJ produce anche una lista dei *digrams* del corpus analizzato, elencando le coppie di parole più ricorrenti, favorendo quindi l'avvio di una prima analisi morfosintattica manuale dei contenuti.

Il secondo passaggio previsto dalle operazioni NLP di NooJ è la *Linguistic Analysis* del corpus, ovvero la sua lettura e analisi tramite l'applicazione di dizionari elettronici. Come già accennato, i dizionari elettronici di NooJ hanno la funzione di motori linguistici e sono di due tipi, classificabili in base alle ULA che elencano:

- le parole semplici (dizionari DELAS-DELAF), che includono tutte quelle parole semanticamente autonome e composte da sequenze di lettere non interrotte e delimitate da spazi bianchi (ad esempio le parole *casa*, *battello*);
- le parole composte (dizionari DELAC-DELACF), che includono tutte quelle sequenze formate da due o più parole e che costruiscono congiuntamente singole unità di significato.

La Linguistic Analysis di NooJ produce un file denominato *Annotation*, in cui vengono elencate e commentate morfosintatticamente tutte le ULA

ISBN Open Access: 9788835191971

reperite nel testo. Insieme alla Tokenization e ai Digrams, il file Annotation e i metadati morfo-sintattici e semantici da esso contenuti rappresentano la base su cui vengono strutturate le grammatiche locali di NooJ. Va sottolineato che non tutti i risultati delle precedenti fasi di analisi vengono utilizzati contemporaneamente. La costruzione di una grammatica, infatti, si effettua a partire dal fenomeno o dalla struttura logico-linguistica che si intende evidenziare e descrivere formalmente. Pertanto, per ciò che riguarda l'identificazione di contenuto offensivo, prima di procedere all'analisi sintattico-semantica del corpus, sarà utile localizzare e descrivere non solo le sequenze testuali in cui vengono prodotti discorsi d'odio, ma anche le varie scelte lessicali impiegate per essi (Scoppetta *et al.*, 2018).

Dopo la Linguistic Analysis, tramite il Locate Pattern[3], NooJ prevede la creazione di *Concordance*[4], che permettono di localizzare e visualizzare specifiche parole del testo all'interno del loro contesto di occorrenza. Grazie alle informazioni fornite dalla Concordance, è possibile sviluppare due ulteriori fasi di analisi e visualizzazione dei risultati, ovvero:

- l'individuazione di strutture ricorrenti morfosintattiche ricorsive all'interno del testo e che possono, in quanto tali, essere formalizzati tramite specifiche grammatiche locali;
- la creazione e visualizzazione di un grafico *Standard Score* con cui è possibile localizzare i punti specifici di un testo in cui una ULA, un concetto, o uno schema linguistico hanno maggior occorrenza. Il grafico consente inoltre di combinare e confrontare i risultati provenienti da un massimo di quattro grammatiche. Qualora generato a partire da analisi sintattico-semantiche mirate, lo Standard Score permette altresì di tracciare l'andamento di specifiche catene di significazione lungo l'estensione del testo.

5. Conclusioni

Questo contributo mostra come l'impiego combinato di dizionari tematici, tecniche di espansione data-driven e ontologie di dominio costituisca un percorso metodologico solido per affrontare la complessità del linguaggio razzista e xenofobo online. Gli approcci puramente lessicali rappresentano un punto di partenza indispensabile: essi sono pienamente non supervi-

[3] Sia nel Locate Pattern che per la creazione e applicazione delle grammatiche locali, NooJ usa specifici comandi formali e logici basati sintatticamente e concettualmente sugli operatori booleani. Per maggiori informazioni sulle modalità formali di query con NooJ, si veda il manuale di NooJ, disponibile per download su https://nooj.univ-fcomte.fr/files/NooJManual.pdf.

[4] Per la storia e i primi usi delle concordanze, si veda https://lhs.unb.br/cliomatica/index.php?title=Index_Thomisticus.

ISBN Open Access: 9788835191971

sionati, trasferibili a nuovi domini senza costosi *training set* annotati e capaci di individuare un ampio spettro di termini denigratori.

Tuttavia, tali risorse mostrano limiti intrinseci. Da un lato, la rilevazione basata esclusivamente su *slur* espliciti porta a sistemi ad alta *precision* (proporzioni di contenuti correttamente identificati come offensivi rispetto al totale) ma basso *recall* (proporzione di contenuti offensivi effettivamente identificati rispetto al totale dei contenuti offensivi presenti); dall'altro, l'inclusione di lemmi potenzialmente offensivi ma polisemici produce un numero eccessivo di falsi positivi, compromettendo la *precision* complessiva (Davidson *et al.*, 2019). Inoltre, gli approcci puramente lessicali non riescono ad identificare forme linguistiche più sottili, quali metafore, sarcasmo o codici comunicativi (*dog-whistles*), né riescono a disambiguare l'intento comunicativo alla base dell'uso di un termine (Ali, Blackburn e Stringhini, 2025). A ciò si aggiunge la necessità di aggiornare continuamente tali risorse, in un contesto in cui il linguaggio tossico evolve rapidamente e nuovi *slur* emergono con cadenza costante (Waseem e Hovy, 2016; Nobata *et al.*, 2016).

Recentemente, nel contesto dell'identificazione di contenuti tossici, approcci ibridi hanno mostrato il loro potenziale integrando caratteristiche estratte da lessici in classificatori supervisionati (Koufakou *et al.*, 2020; Polignano *et al.*, 2022). Ali, Blackburn e Stringhini (2025) hanno dimostrato che un modello ibrido che combina BERT con tecniche lessicali raggiunge un'accuratezza del 95% su diversi dataset benchmark, superando sia gli approcci puramente lessicali sia quelli basati esclusivamente su apprendimento supervisionato. Polignano *et al.* (2022) hanno sviluppato un sistema che arricchisce il rilevamento di hate speech con spiegazioni human-centered, integrando feature lessicali con modelli di *deep learning*.

Tuttavia, è l'ontologia di dominio che consente il passaggio decisivo dalla dimensione lessicale alla struttura concettuale del discorso ostile. Formalizzando cluster semantici, relazioni logiche e narrative ricorrenti, dalla disumanizzazione alle retoriche dell'invasione, dai frame securitari ai processi di contaminazione simbolica, l'ontologia permette di ricostruire l'architettura profonda del discorso razzista e xenofobo. In tal modo, essa non solo supporta la classificazione automatica, ma diventa un elemento cruciale nei sistemi di *Explainable AI* (XAI), consentendo di motivare la classificazione non attraverso correlazioni opache, superando l'opacità dei modelli *black-box*, bensì richiamando relazioni concettuali e narrative riconoscibili e interpretativamente fondate. In linea con quanto discusso da Confalonieri *et al.* (2021) e Schwalbe e Finzel (2024), l'ontologia abilita forme di *explanation-by-design* (Adadi e Berrada, 2018), nelle quali la trasparenza esplicativa è integrata nel sistema sin dall'origine.

ISBN Open Access: 9788835191971

Riferimenti bibliografici

Adadi, A. and Berrada, M. (2018), Peeking inside the black-box: A survey on explainable artificial intelligence (XAI), *IEEE Access*, 6: 52138-52160.

Agudelo, F. I. and Olbrych, N. (2022). It's not how you say it, it's what you say: Ambient digital racism and racial narratives on Twitter, *Social Media + Society*, 8(3).

Ali, S., Blackburn, J. and Stringhini, G. (2025), Evolving hate speech online: An adaptive framework for detection and mitigation, *arXiv preprint, arXiv*:2502.10921, 2025.

Almatarneh, S. and Gamallo, P. (2018), A lexicon based method to search for extreme opinions, *PLoS ONE* 13(5): 1–19.

Basile, V. (2019), HurtLex: A Multilingual Lexicon for Hate Speech Detection. In *Proceedings of the 6th Italian Conference on Computational Linguistics* (pp. 1-6), CEUR-WS.

Bassignana, E., Basile, V. and Patti, V. (2018), Hurtlex: A multilingual lexicon of words to hurt. In *Proceedings of the 5th Italian Conference on Computational Linguistics*, CLiC-it 2018 (vol. 2253, pp. 1-6). CEUR-WS.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3: 993-1022.

Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. (2008), Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, (10), P10008.

Chiarcos, C., Nordhoff, S. and Hellmann, S., eds. (2012). *Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata*. Heidelberg: Springer.

Colace, F., De Santo, M., Greco, L., Moscato, V. and Picariello, A. (2016), Probabilistic approaches for sentiment analysis: Latent Dirichlet allocation for ontology building and sentiment extraction. In Pedrycz W. and Chen S.-M., eds., *Sentiment analysis and ontology engineering Studies in Computational Intelligence* (vol. 639, pp. 57–76), Springer, Cham, Switzerland.

Confalonieri, R., Weyde, T., Besold, T.R. and Martín, F.M. (2021), Using ontologies to enhance explainability in AI, *AI Magazine*, 42(3): 69-81.

Cucco, A., Fontanella, L., Sarra, A. and Fontanella, S. (2025) (manoscritto sottomesso), *Apprendimento di ontologie dei contenuti abusivi online: Seeded LDA e strutturazione semantica delle narrazioni offensive anti-migranti in italiano basata su reti*.

Davidson, T., Warmsley, D., Macy, M. and Weber, I. (2019), Automated hate speech detection and the problem of offensive language, *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1): 512-515.

de Ayala, R. J. (2009), *The theory and practice of item response theory*, The Guilford Press, New York.

De Mauro, T. (2016), *Le parole per ferire*. Internazionale. 27 settembre 2016, Compiled for the "Jo Cox" Committee on intolerance, xenophobia, racism and hate phenomena, of the Italian Chamber of Deputies (2016).

ISBN Open Access: 9788835191971

Fersini, E., Nozza, D. and Rosso, P. (2020), AMI @ EVALITA2020: automatic misogyny identification. In: Basile, V., D. C, Di Maro, M., *et al.*, eds., *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*. Final Workshop (EVALITA 2020), Online event, 17 Dec 2020, CEUR Workshop Proceedings, vol. 2765. CEUR-WS.org (2020).

Gabrielatos, C. (2018), Keyness analysis: Nature, metrics and techniques. In Taylor C. and Marchi A., eds., *Corpus approaches to discourse* (pp. 225–258). Routledge, London.

Gitari, N.D, Zuping, Z., Damien, H. and Long, J. (2015), A lexicon-based approach for hate speech detection, *International Journal of Multimedia and Ubiquitous Engineering*, 10: 215–230.

Gross, M. (1975). *Méthodes en syntaxe*, Hermann, Paris.

Gross, M. (1984). Lexicon-grammar and the syntactic analysis of French. In 22nd Annual Meeting of the Association for Computational Linguistics, ACL (pp. 275–282).

Gross, M. (1989a). The construction of electronic dictionaries [La construction de dictionnaires électroniques]. *Annales Des Télécommunications*, 44(1-2): 4–19.

Gross, M. (1989b). The use of finite automata in the lexical representation of natural language. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 377 LNCS, 34–50.

Huang, H., Harzallah, M., Guillet, F. and Xu, Z. (2021), Core-concept-seeded LDA for ontology learning, *Procedia Computer Science*, 192: 222–231.

Ibrahim, N., Mulford, F., Lawrence, M. and Batista-Navarro, R. (2024), Resources for annotating hate speech in social media platforms used in Ethiopia: A novel lexicon and labelling scheme. In *Proceedings of the Fifth Workshop on Resources for African Indigenous Languages @ LREC-COLING 2024* (pp. 115–123).

Jagarlamudi, J., Daumé III, H. and Udupa, R. (2012), Incorporating lexical priors into topic models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 204-213), Association for Computational Linguistics.

Koufakou, A., Pamungkas, E.W., Basile, V. and Patti, V. (2020), HurtBERT: Incorporating lexical features with BERT for the detection of abusive language. In *Proceedings of the Fourth Workshop on Online Abuse and Harms* (pp. 34-43).

Maisto, A., Pelosi, S., Vietri, S. and Vitale, P. (2017), Mining offensive language on social media. *Proceedings of the Fourth Italian Conference on Computational Linguistics CLiC-it 2017* (pp. 252–256).

Maronikolakis, A., Köksal, A. and Schuetze, H. (2024), Sociocultural knowledge is needed for selection of shots in hate speech detection tasks, In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion* (pp. 1–13).

Monnar, A.A., Perez Rojas, J. and Labra, B.P. (2024), Cross-lingual hate speech detection using domain-specific word embeddings, *PLOS ONE*, 19(7): e0306521.

ISBN Open Access: 9788835191971

Monteleone, M. (2019). NooJ Grammars and Ethical Algorithms: Tackling On-Line Hate Speech. In Mirto, I., Monteleone, M. and Silberztein, M., eds., *Formalizing Natural Languages with NooJ 2018 and Its Natural Language Processing Applications*. NooJ 2018. Communications in Computer and Information Science, vol. 987. Springer, Cham.

Monteleone, M. (2021). NooJ for Artificial Intelligence: An Anthropic Approach. In Bekavac, B., Kocijan, K., Silberztein, M. and Šojat, K., eds., *Formalising Natural Languages: Applications to Natural Language Processing and Digital Humanities*. NooJ 2020. Communications in Computer and Information Science, Springer, Cham (vol. 1389, pp. 173-184).

Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y. and Chang, Y. (2016), Abusive language detection in online user content. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 145-153).

Poletto, F., Basile, V., Sanguinetti, M., Bosco, C. and Patti, V. (2021), Resources and benchmark corpora for hate speech detection: a systematic review, *Language Resources and Evaluation*, 55: 477–523.

Polignano, M., Colavito, G., Musto, C., de Gemmis, M. and Semeraro, G. (2022), Lexicon enriched hybrid hate speech detection with human-centered explanations. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 88-93).

Razavi, A.H., Inkpen, D., Uritsky, S. and Matwin, S. 2010: Offensive language detection using multi-level classification. In: Farzindar, A., Kešelj, V., eds., *Advances in Artificial Intelligence Canadian AI 2010. Lecture Notes in Computer Science* (pp. 16–27). Springer, Berlin.

Rubio-Carbonero, G. (2020), Hybrid language and strategic camouflage in contemporary hate speech, *Discourse & Society*, 31(4): 456-478.

Samejima, F. (1969), Estimation of latent ability using a response pattern of graded scores, *Psychometrika Monograph Supplement*, 34(4, Pt. 2): 100.

Sanguinetti, M., Poletto, F., Bosco, C., Patti, V. and Stranisci, M. (2018), An Italian Twitter corpus of hate speech against immigrants. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), Miyazaki, Japan (2018).

Schmidt, A. and Wiegand, M. (2017), A survey on hate speech detection using natural language processing, In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media* (pp. 1–10). Association for Computational Linguistics, Valencia, Spain.

Schwalbe, G. and Finzel, B. (2024), A comprehensive taxonomy for explainable artificial intelligence, *ACM Computing Surveys*, 56(5), 1-38.

Scoppetta, C., Alfieri, A., Merenda, F., Lay, S., Colasanto, A. and Manna, R. (2017): From language to social perception of immigration. In Mbarki, S., Mourchid, M., Silberztein, M., eds., NooJ 2017 (pp. 213–224). Springer, Cham, CCIS, vol. 811.

Sharma, N. (2018), *Environmental racism and discursive violence*, Routledge, London.

Silberztein, M. (2015), *La formalisation des langues*, L'approche de NooJ. ISTE, London.

ISBN Open Access: 9788835191971

Tissaoui, A., Sassi, S., Chbeir, R. and Mechergui, A. (2022), A top-down enriching approach for ontology learning from text, *Concurrency and Computation: Practice and Experience*, 34(19): e7036.

Tontodimamma, A., Fontanella, L., Anzani, S. and Basile, V. (2023), An Italian lexical resource for incivility detection in online discourses, *Quality & Quantity*, 57(4), 3019-3037.

Vargas, F., Carvalho, I., Pardo, T.A.S. and Benevenuto, F. (2024), Context-aware and expert data resources for Brazilian Portuguese hate speech detection, *Natural Language Processing Journal*, 1-22.

Vargas, F.A., Carvalho, I. and Rodrigues de Góes, F. (2021), Identifying offensive expressions of opinion in context, *arXiv preprint, arXiv:2104.12227.*

Waseem, Z. and Hovy, D. (2016), Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. In *Proceedings of the NAACL Student Research Workshop* (pp. 88-93).

Wiegand, M., Ruppenhofer, J., Schmidt, A. and Greenber, C. (2018), Inducing a lexicon of abusive words—a feature-based approach. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1 (Long Papers)* (pp. 1046–1056). Association for Computational Linguistics, New Orleans, LO (2018).

Wilson, T., Wiebe, J. and Hoffmann, P. (2005), Recognizing contextual polarity in phrase-level sentiment analysis, In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05* (pp. 347–354). Association for Computational Linguistics (2005).

Wong, W., Liu, W. and Bennamoun, M. (2012), Ontology learning from text: A look back and into the future, *ACM Computing Surveys*, 44(4): 20, 1–20, 36.

ISBN Open Access: 9788835191971

Oltre il consenso: un corpus annotato in ottica prospettivista per il rilevamento di razzismo e xenofobia nei social media italiani

di *Lara Fontanella* [*], *Michelangelo Misuraca* [**], *Giuseppe Giordano* [***], *Alex Cucco* [*], *Emiliano del Gobbo* [****], *Elisa Ignazzi* [*****]

1. Il ruolo di corpora annotati nella rilevazione di contenuti razzisti

Nell'era della comunicazione digitale, in cui i social media amplificano ed accelerano la diffusione di contenuti discriminatori, i corpora annotati costituiscono uno strumento essenziale per comprendere e contrastare l'hate speech in generale, e il razzismo e la xenofobia in particolare. L'utilizzo di tali corpora annotati rappresenta il fondamento metodologico per la rilevazione automatica di discorsi d'odio attraverso l'applicazione di tecniche di machine learning e deep learning (Schmidt e Wiegand, 2017; Fortuna e Nunes, 2018). Gli approcci di machine learning tradizionali, basati su feature engineering, hanno dimostrato la loro efficacia quando addestrati su dataset annotati di qualità (Davidson *et al.*, 2017; Waseem e Hovy, 2016). Tuttavia, l'avvento delle tecniche di deep learning ha rivoluzionato il campo, permettendo l'apprendimento automatico di rappresentazioni complesse direttamente dai dati annotati (Badjatiya *et al.*, 2017) e, più recentemente, modelli transformer-based come BERT (Mozafari *et al.*, 2020) hanno raggiunto performance state-of-the-art proprio grazie alla disponibilità di corpora annotati su cui effettuare il fine-tuning.

[*] Dipartimento di Studi Socio-Economici, Gestionali e Statistici, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, lara.fontanella@unich.it; alex.cucco@unich.it

[**] Dipartimento di Scienze Aziendali – Management & Innovation Systems, Università di Salerno, mmisuraca@unisa.it

[***] Dipartimento di Studi Politici e Sociali, Università di Salerno, ggiordano@unisa.it

[****] Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, emiliano.delgobbo@unina.it

[*****] Dipartimento di Neuroscienze, Imaging e Scienze Cliniche, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, elisa.ignazzi@unich.it

ISBN Open Access: 9788835191971

La qualità e la rappresentatività dei corpora annotati utilizzati per il fine-tuning influenzano direttamente le performance e, criticamente, i potenziali bias dei modelli risultanti (Davidson *et al.*, 2019; Sap *et al.*, 2019). Parallelamente, l'emergere dell'Explainable Artificial Intelligence (XAI) ha posto nuove richieste ai corpora annotati. Non è più sufficiente avere etichette di classificazione binarie o multi-classe; dataset come HateXplain (Mathew *et al.*, 2021) hanno introdotto annotazioni a livello di token che forniscono "*rationales*" esplicativi, permettendo lo sviluppo di modelli interpretabili che non solo identificano contenuti tossici ma spiegano anche le ragioni delle loro predizioni.

Questo aspetto è particolarmente rilevante per applicazioni di moderazione dei contenuti, dove la trasparenza delle decisioni automatiche è essenziale sia per ragioni etiche che legali (Lepri *et al.*, 2018; Binns, 2018). La disponibilità di corpora annotati con metadata demografici sugli annotatori e sui target del discorso d'odio consente analisi sofisticate sulla percezione soggettiva della tossicità e sulla variabilità delle interpretazioni (Aroyo e Welty, 2015), aspetti sempre più riconosciuti come centrali nell'annotazione di fenomeni complessi e socialmente situati. Questa crescente consapevolezza della percezione intrinsecamente soggettiva dell'hate speech ha portato allo sviluppo degli approcci prospettivisti (*perspectivist*) all'annotazione (Plank, 2022; Davani *et al.*, 2022; Frenda *et al.*, 2025), che riconoscono il disaccordo tra annotatori non come rumore da eliminare attraverso il majority voting, ma come segnale informativo che riflette la legittima diversità di prospettive ed interpretazioni (Uma *et al.*, 2021b; Cabitza *et al*., 2023).

L'approccio prospettivista si distacca dal paradigma tradizionale della "ground truth" unica (Aroyo e Welty, 2015; Gordon *et al.*, 2022), preservando e valorizzando le annotazioni individuali per catturare la molteplicità di punti di vista esistenti. Tale approccio è particolarmente rilevante per fenomeni come il razzismo, dove fattori culturali e sociali, esperienze personali e background demografici degli annotatori influenzano significativamente il giudizio su cosa costituisca contenuto discriminatorio (Sap *et al.*, 2022).

La rilevazione del razzismo e della xenofobia richiede dataset particolarmente curati, capaci di riflettere la natura multidimensionale di questi fenomeni e di catturare le diverse modalità attraverso cui il pregiudizio razziale si manifesta nel linguaggio online, dalla discriminazione esplicita alle forme più velate di esclusione e stigmatizzazione. Per la lingua italiana esistono diversi corpora annotati per l'hate speech a sfondo razzista e xenofobo, in particolare l'Italian Hate Speech Corpus di Sanguinetti *et al.* (2018) e i dataset sviluppati nell'ambito del progetto "Contro l'Odio" (Capozzi *et al.*, 2020) e delle shared task HaSpeeDe (Bosco *et al.*, 2018; Sanguinetti *et al.*, 2020). In particolare, il DisaggregHate-It Corpus (Madeddu *et al.*, 2023) adotta e-

ISBN Open Access: 9788835191971

splicitamente un approccio prospettivista, conservando le annotazioni individuali e modellando il disaccordo inter-annotatore come segnale informativo sulla soggettività della percezione del razzismo. Il DisaggregHate-It Corpus è composto da 1.100 tweet estratti dal corpus "Contro l'Odio" e annotati secondo uno schema multidimensionale: (i) presenza di hate speech (binaria), (ii) presenza di ironia (binaria), e (iii) *stance* dell'autore verso le questioni migratorie (positiva, neutra, negativa). Il protocollo di annotazione ha coinvolto studenti universitari organizzati in gruppi di almeno 5 componenti, con l'obiettivo di ottenere molteplici prospettive per ciascun tweet. A ogni annotatore è stato chiesto di annotare almeno 100 tweet, risultando in una distribuzione variabile di annotazioni per tweet (1-13 annotazioni).

Il presente lavoro si ispira all'impianto teorico e metodologico proposto da Madeddu *et al.* (2023), applicando l'approccio *perspectivist* alla costruzione di un nuovo corpus italiano per la rilevazione del razzismo e della xenofobia. Il contributo è strutturato seguendo le fasi del workflow di costruzione del corpus: raccolta dei dati e selezione dei commenti da annotare, progettazione dello schema di annotazione, implementazione del processo di annotazione, ed analisi dei risultati preliminari con particolare attenzione ai pattern di disaccordo inter-annotatore.

2. Costruzione del corpus e selezione dei commenti da annotare

Per la costruzione del corpus sono stati raccolti commenti pubblicati nell'ultimo decennio su Facebook, Instagram e YouTube, selezionando i contenuti relativi a post sul tema dell'immigrazione. Il dataset iniziale è composto da 185.734 documenti. Applicando un approccio di filtraggio basato su parole chiave, che comprende un'ampia gamma di sinonimi rappresentativi sia della retorica pro-immigrazione sia di quella anti-immigrazione (Fontanella *et al.*, 2024), sono stati individuati 39.570 commenti pertinenti al dibattito migratorio. A partire da questa selezione è stato successivamente estratto il campione destinato al processo di annotazione.

Per garantire un'adeguata copertura della diversità tematica ed una rappresentatività equilibrata dello spettro di sentiment e di contenuti offensivi presenti, è stata adottata una strategia di campionamento basata su grafo che tenesse conto anche della direzione del sentiment (Cucco *et al.*, 2025a). Sul corpus di 39.570 commenti è stata applicata la Latent Dirichlet Allocation (Blei *et al.*, 2003) per derivare 20 topic e assegnarli ai documenti, fissando a 0,1 la soglia di probabilità di presenza del topic nel documento. Il grafo tematico è stato quindi costruito connettendo i documenti che condividono almeno un topic. Il sentiment è stato impiegato come variabile ausiliaria di

ISBN Open Access: 9788835191971

stratificazione. Le classi di stratificazione sono state definite combinando due dimensioni: (i) presenza o assenza di linguaggio offensivo, identificato mediante il lessico Revised HurtLex (Tontodimamma *et al.*, 2023), e (ii) concordanza del sentiment misurato attraverso tre lessici specifici in lingua italiana, Sentix (Basile e Nissim, 2013), Mal (Vassallo *et al.*, 2019) e W-Mal (Vassallo *et al.*, 2020), classificando i commenti come concordemente positivi (tutte e tre le misure positive), concordemente negativi (tutte e tre negative) o discordanti (disaccordo tra le misure). Questa combinazione ha prodotto sei classi di stratificazione. L'applicazione del campionamento su rete con stratificazione per queste sei classi (Cucco *et al.*, 2025b) ha consentito la selezione di 3.000 commenti destinati all'annotazione, garantendo una rappresentazione bilanciata con 500 commenti per classe di stratificazione.

3. Schema di annotazione ed implementazione del processo di annotazione

Lo schema di annotazione è strutturato su più livelli per catturare diverse dimensioni del contenuto razzista e xenofobo. Per ciascun commento, gli annotatori sono chiamati a valutare la presenza di razzismo/xenofobia attraverso un'etichetta binaria (sì/no) che indica se il commento contiene contenuto razzista o xenofobo. In caso di risposta affermativa, viene richiesta anche la valutazione del livello di gravità su una scala ordinale da 1 (lieve) a 5 (grave), che misura l'intensità del contenuto offensivo, nonché la selezione dei *rationales*, ovvero le porzioni di testo che motivano la classificazione, permettendo di identificare gli elementi linguistici specifici alla base del giudizio. Lo schema prevede inoltre tre ulteriori annotazioni binarie: la presenza di ironia o sarcasmo, l'uso di stereotipi razziali o etnici, e l'eventuale incitamento alla violenza e all'odio. Il processo di annotazione è stato condotto attraverso la piattaforma Label Studio (Tkachenko *et al.*, 2025). Sono stati reclutati 10 annotatori, tutti studenti universitari, appositamente formati sullo schema di annotazione.

Seguendo l'approccio perspectivist, ciascun annotatore ha annotato l'intero campione di 3.000 commenti in modo indipendente, senza discussione o riconciliazione delle divergenze. Questa procedura ha prodotto annotazioni individuali, preservando la molteplicità delle prospettive e permettendo l'analisi del disaccordo inter-annotatore come informazione rilevante sulla soggettività della percezione del razzismo.

ISBN Open Access: 9788835191971

4. Risultati del processo di annotazione

Di seguito si presentano i risultati del processo di annotazione, con specifico focus sulle due dimensioni principali dello schema: la valutazione della presenza di contenuto razzista e la valutazione del suo livello di gravità. L'analisi considera sia le etichette attribuite (etichetta binaria per la presenza di razzismo e scala 1–5 per il livello di gravità) dagli annotatori sia le porzioni di testo selezionate come giustificazione (*rationales*), permettendo di valutare l'accordo inter-annotatore sia sul giudizio complessivo sia sull'identificazione degli elementi linguistici specifici che motivano tale giudizio.

4.1. Variabilità inter-annotatore nella valutazione del razzismo: presenza e intensità

Per quantificare l'accordo inter-annotatore abbiamo adottato l'indice α di Krippendorff (2013), preferito rispetto ad altre metriche per le sue proprietà statistiche, fra cui rilevanti per il nostro corpus annotato la correzione per l'accordo casuale, l'applicabilità a qualsiasi numero di annotatori e l'adattabilità a diverse scale di misurazione. In particolare, per le annotazioni binarie (presenza/assenza di razzismo) abbiamo utilizzato pesi nominali, mentre per le valutazioni ordinali (livello di gravità 1–5) abbiamo applicato pesi ordinali. L'analisi ha prodotto valori di $\alpha = 0{,}389$ (*IC* 95%: 0,373 – 0.404) per la presenza di razzismo e $\alpha = 0{,}390$ (*IC* 95%: 0,373 – 0.406) per il livello di gravità. Entrambi i valori sono significativamente inferiori alle soglie di riferimento proposte da Krippendorff: $\alpha \geq 0{,}800$: accordo affidabile; $0{,}667 \leq \alpha < 0{,}800$: accordo accettabile; $\alpha < 0{,}667$: accordo insoddisfacente. I nostri risultati si collocano nella terza categoria, evidenziando elevata variabilità inter-annotatore. È importante sottolineare che le soglie interpretative proposte da Krippendorff sono state formulate nell'ambito di paradigmi che assumono l'esistenza di una ground truth unica e condivisa. In contesti prospettivisti, dove il disaccordo tra annotatori è concettualizzato come segnale informativo piuttosto che come errore di misurazione, valori di α inferiori alle soglie convenzionali non indicano necessariamente carenze nella qualità del processo di annotazione, ma riflettono piuttosto la legittima pluralità di interpretazioni di fronte a fenomeni intrinsecamente soggettivi come la percezione del razzismo.

Per entrare maggiormente nel dettaglio, la Figura 1a riporta la distribuzione percentuale dei commenti in funzione del numero di annotatori che hanno identificato la presenza di contenuti razzisti. La modalità più frequente corrisponde ai casi in cui nessun annotatore ha rilevato contenuto pro-

ISBN Open Access: 9788835191971

blematico (23,8% del corpus). La Figura 1b mostra una heatmap in cui ogni riga rappresenta un commento e ogni colonna rappresenta un annotatore specifico.

I colori codificano il livello di razzismo rilevato nel commento da ciascun annotatore, con valori più bassi (blu) indicativi di assenza o bassa intensità di razzismo e valori più elevati (rosso) indicativi di livelli più alti. Le Figure 1a e 1b evidenziano congiuntamente l'elevata variabilità inter-annotatore.

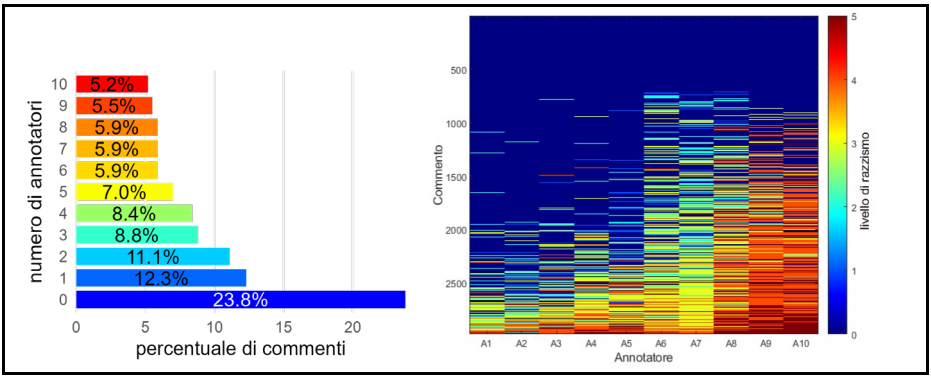


Fig. 1 – Annotazioni per presenza e livello di razzismo. (a) Distribuzione percentuale dei commenti rispetto al numero di annotatori che hanno rilevato la presenza di razzismo. (b) Heatmap delle annotazioni sul livello di razzismo: ogni riga rappresenta un commento, ogni colonna un annotatore; i colori codificano il livello di razzismo attribuito (da 0, blu, a 5, rosso)

Considerando la presenza di contenuto razzista, la distribuzione mostra bassa densità di casi ad alta concordanza: solo una minoranza di commenti è stata classificata come razzista da 8–10 annotatori, mentre la maggioranza si colloca in una zona di ambiguità interpretativa caratterizzata da concordanza parziale o disaccordo. La heatmap disaggrega questa variabilità a livello individuale, rivelando che per i commenti classificati come razzisti emergono divergenze sostanziali: alcuni annotatori attribuiscono sistematicamente livelli elevati di razzismo, mentre altri forniscono valutazioni moderate o assenti per gli stessi contenuti.

Per approfondire ulteriormente l'analisi della variabilità inter-annotatore, è possibile applicare un modello fattoriale che consenta di: (i) valutare le differenze nella propensione dei singoli annotatori a riconoscere contenuti razzisti attraverso la stima di parametri individuali, e (ii) derivare contestualmente un punteggio sintetico di razzismo per ciascun commento. Data la natura dicotomica dell'annotazione per la presenza di razzismo e la natura ordinale della scala che misura il livello di gravità, adottiamo un modello che rientra nell'ambito della Item Response Theory (IRT). In questa applicazio-

ISBN Open Access: 9788835191971

ne, la struttura tipica dei modelli IRT viene adattata al contesto dell'annotazione: i commenti assumono il ruolo usualmente riservato ai soggetti rispondenti, mentre gli annotatori fungono da item. In questo modo, il tratto latente stimato rappresenta il livello intrinseco di razzismo del commento, e i parametri degli item catturano le caratteristiche individuali degli annotatori, quali la loro severità e capacità discriminativa.

Seguendo la proposta di Tontodimamma *et al.* (2023), per derivare il punteggio sintetico per ogni commento, modelliamo esplicitamente la probabilità di fornire una valutazione in una data categoria (0/1 per la presenza di razzismo, da 1 a 5 per il livello di gravità) attraverso un Graded Response Model (GRM) unidimensionale (Samejima, 1969). Più specificamente, assumiamo che l'annotazione assegnata ad un commento dipenda dal livello intrinseco di razzismo del commento stesso e da alcuni parametri dell'annotatore che rappresentano il peso della sua valutazione nell'indice composito di razzismo ed i valori soglia tra categorie consecutive lungo la sua scala di valutazione continua.

Formalmente, indicando con X_{ij} il punteggio assegnato dall'annotatore j al commento i, secondo la formulazione 2PNO (*Two-Parameter Normal Ogive*), la probabilità che il punteggio ricada nella categoria $c = 1, \ldots, C$ della scala di risposta è data da:

$$P(X_{ij} = c|\theta_i, \lambda_j, \gamma_j) = \Phi\left(\lambda_j\theta_i - \gamma_{j,c-1}\right) - \Phi\left(\lambda_j\theta_i - \gamma_{j,c}\right)$$

dove Φ è la funzione di distribuzione normale standard, θ_i denota il livello di razzismo del commento i, λ_j è il peso fattoriale (*factor loading*), o parametro di discriminazione, per il valutatore j, e $\boldsymbol{\gamma}_j = \left(\gamma_{j,1}, \ldots, \gamma_{j,c-1}\right)$ è il vettore dei parametri soglia ordinati $-\infty \leq \gamma_{j,1} \leq \cdots \leq \gamma_{j,C-1} \leq \infty$. Per la stima dei parametri, adottiamo tecniche di simulazione Markov Chain Monte Carlo (MCMC) (Tontodimamma *et al.*, 2023).

Il modello è stato applicato separatamente alle annotazioni relative alla presenza di razzismo ed a quelle riguardanti il livello di gravità. Le Figure 2 e 3 riportano le stime dei parametri degli annotatori: il factor loading (λ_j, pannello sinistro), che riflette la coerenza e la capacità informativa delle valutazioni dell'annotatore j, e il parametro di localizzazione (pannello destro), che sintetizza la sua propensione a riconoscere contenuti come razzisti. Nel caso delle annotazioni binarie, il parametro di localizzazione coincide con l'unica soglia (γ_j) stimata dal modello per ciascun annotatore. Per il livello di gravità, invece, tale parametro è definito come la media delle soglie corrispondenti alle quattro transizioni tra categorie consecutive ($\bar{\gamma}_j =$

ISBN Open Access: 9788835191971

$\frac{1}{4}\sum_{c-1}^{4}\gamma_{j,c}$) stimate per ciascun annotatore. Tale parametro costituisce un indicatore sintetico della severità di valutazione: valori più elevati riflettono una maggiore cautela nel classificare contenuti come razzisti, ovvero la tendenza a richiedere livelli più marcati di evidenza prima di assegnare valutazioni di razzismo o gravità elevata. Al contrario, valori più bassi indicano una maggiore sensibilità nel riconoscere contenuti problematici. Le due figure evidenziano una chiara variabilità inter-annotatore sia nei pesi fattoriali sia nei parametri di severità. In entrambi i modelli, gli annotatori mostrano livelli differenti di discriminazione e soglie decisionali non sovrapponibili, indicando che non adottano criteri omogenei nella valutazione dei commenti.

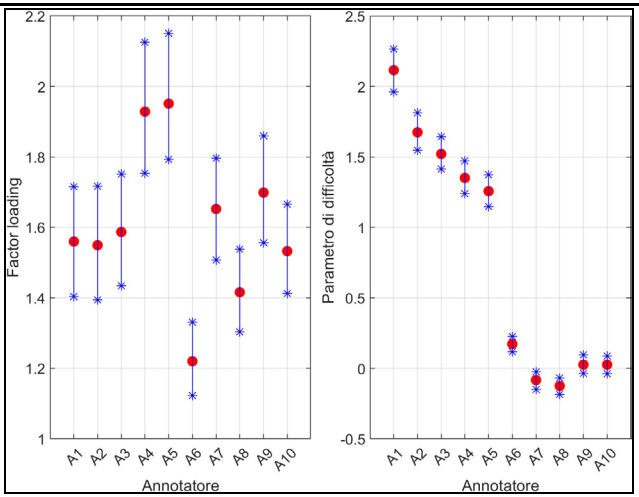
La separazione sistematica dei punti e l'assenza di sovrapposizione sostanziale tra molti intervalli di credibilità suggeriscono che tali differenze non sono attribuibili al caso, ma risultano statisticamente significative. In particolare, alcuni annotatori appaiono più severi o più sensibili rispetto ad altri, producendo parametri chiaramente distinti e quindi interpretabili come differenze reali nei criteri di giudizio.

Per l'annotazione per livelli di razzismo, la Figura 3 mostra, per ciascun annotatore, la struttura delle soglie ordinali stimate dal modello GRM lungo la scala del tratto latente. Ogni segmento colorato rappresenta l'intervallo della variabile latente associato a ciascuna categoria osservata (1–5), mentre i punti neri corrispondono alle soglie stimate ($\gamma_{j,c}$), ossia i punti di demarcazione lungo il continuum latente nei quali la probabilità di scegliere una categoria superiore eguaglia quella di scegliere la categoria immediatamente inferiore. Si evidenzia una marcata eterogeneità inter-annotatore: le soglie non sono allineate tra annotatori e mostrano traslazioni sistematiche lungo la scala latente.

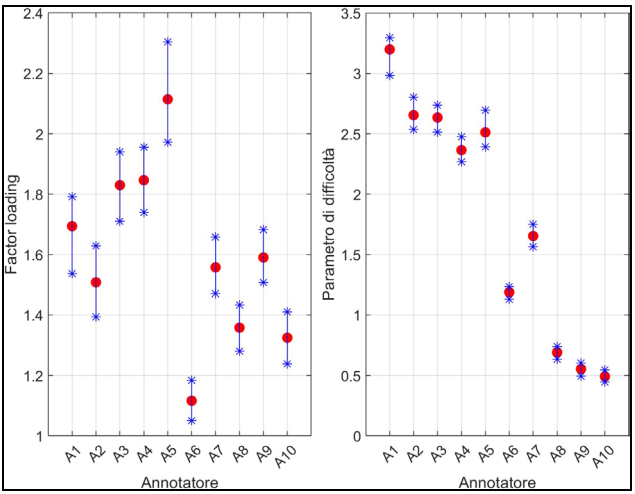
Alcuni annotatori collocano le soglie più a sinistra, segnalando maggiore sensibilità nel riconoscere contenuti problematici ed una propensione ad attribuire valutazioni di gravità più elevate a parità di livello latente di razzismo. Per altri, le soglie sono posizionate verso destra sulla scala latente, indice di una maggiore severità e della necessità di livelli più elevati di razzismo intrinseco prima di assegnare categorie superiori. Oltre alla traslazione sistematica delle soglie, emerge variabilità anche nella loro spaziatura. Annotatori con soglie più distanziate dimostrano maggiore capacità di discriminare tra livelli contigui del costrutto, mentre annotatori con soglie più ravvicinate mostrano minore differenziazione tra categorie adiacenti.

Nel complesso, l'analisi basata sul modello IRT mette in luce differenze strutturate e sistematiche nei criteri decisionali degli annotatori, confermando la natura soggettiva e complessa del processo di valutazione della presenza e della gravità del razzismo e sottolineando la rilevanza di approcci che modellino esplicitamente tale eterogeneità.

ISBN Open Access: 9788835191971



Annotazione per presenza di razzismo: medie a posteriori e intervalli di credibilità al 95% dei parametri degli annotatori



Annotazione per livello di razzismo: medie a posteriori e intervalli di credibilità al 95% dei parametri degli annotatori.

Fig. 2 – Stime dei parametri degli annotatori con il 2PNO Graded Response model unidimensionale

ISBN Open Access: 9788835191971

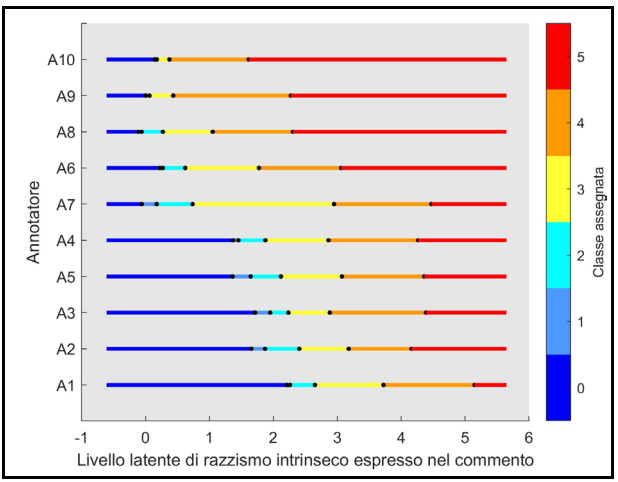


Fig. 3 – Annotazione per livello di razzismo: struttura delle soglie ordinali stimate dal modello GRM lungo la scala del tratto latente di razzismo

4.2. Variabilità inter-annotatore nella selezione dei rationales

L'accordo tra annotatori nell'identificazione degli span (rationales) testuali è stato misurato tramite una valutazione a livello di carattere. Per ciascun testo, le annotazioni individuali sono state convertite in vettori binari, in cui ogni carattere è codificato come annotato (1) o non annotato (0). L'accordo tra coppie di annotatori è stato calcolato utilizzando la F1-score sulla classe (1) inerente ai caratteri considerati parte del rational. L'indice F1-score effettua una valutazione combinando in modo bilanciato accordo sui positivi tra le due annotazioni, rispetto a tutti i positivi assegnati ed accordo sui positivi tra le due annotazioni, rispetto a tutti i positivi presenti. Questa metrica è particolarmente adatta al compito, data la potenziale forte asimmetria tra caratteri annotati e non annotati. I testi per i quali almeno uno degli annotatori non ha selezionato alcuno span sono stati esclusi dal calcolo in quanto non informativi rispetto alla capacità di localizzare contenuto rilevante. Per ciascuna coppia di annotatori, la F1-score è stata calcolata su tutti i testi rilevanti e successivamente aggregata tramite la media aritmetica. I risultati sono riportati in forma di matrice di F1 medio e relative standard deviation, visualizzata graficamente per facilitare il confronto tra annotatori (Fig. 4).

Le matrici riportano gli F1-score per coppie di annotatori medi (fFg. 4.a) e le rispettive deviazioni standard (Fig. 4.b) tra i 10 annotatori. Nonostante le medie siano generalmente alte (range 0.61–0.77), le deviazioni standard mostrano una variabilità elevata (range 0.22 0.28), indicando che le

ISBN Open Access: 9788835191971

performance differiscono notevolmente tra le coppie di annotatori. Questo sottolinea come l'F1-score, calcolato solo sul sottoinsieme di commenti ritenuti da entrambi gli annotatori come razzisti, possa essere influenzato dalla variabilità individuale degli annotatori presente anche in questo tipo di annotazione.

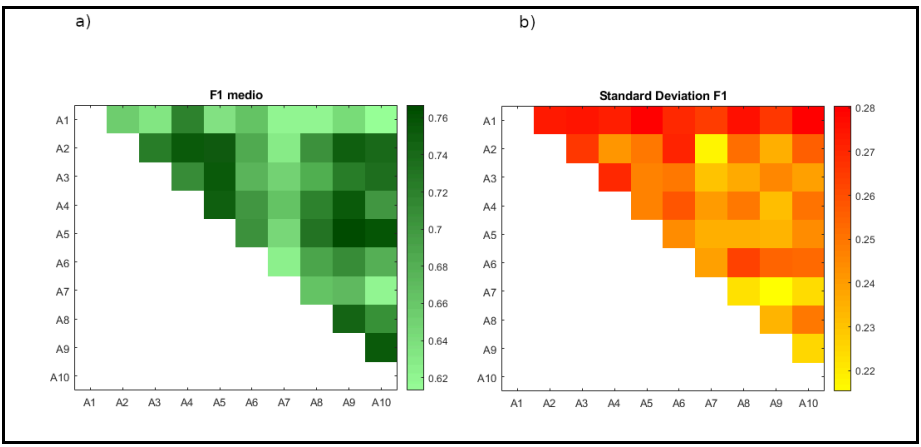


Fig. 4 – Matrice dei valori F1 ottenuti per coppie di annotatori. a) I valori riportati rappresentano l'F1 score medio per ciascuna coppia. b) I valori riportati rappresentano le deviazioni standard (SD) corrispondenti

5. Conclusioni

Il presente capitolo ha presentato la costruzione di un corpus annotato per la rilevazione del razzismo e della xenofobia nei social media italiani, adottando un approccio *perspectivist* che riconosce la natura intrinsecamente soggettiva della percezione del contenuto offensivo.

L'analisi del processo di annotazione ha rivelato elevata variabilità inter-annotatore. Secondo l'approccio prospettivista, i contenuti valori dell'indice di Krippendorff non indicano carenza nella qualità delle annotazioni, ma riflettono la complessità intrinseca del fenomeno. La sostanziale equivalenza dell'accordo tra dimensione binaria e scala ordinale è particolarmente informativa: il disaccordo emerge già nell'identificazione della presenza di contenuto problematico, non solo nella valutazione della sua intensità. Questo pattern è coerente con la natura contestuale e culturalmente situata della percezione del razzismo, dove ciò che per alcuni annotatori costituisce contenuto offensivo può essere interpretato da altri come espressione legittima di opinione o contenuto ambiguo.

L'applicazione di modelli Item Response Theory ha permesso di quanti-

ISBN Open Access: 9788835191971

ficare sistematicamente l'eterogeneità tra annotatori. I parametri di localizzazione stimati mostrano differenze statisticamente significative, riflettendo diversità individuali nelle soglie di sensibilità, nell'interpretazione del contesto e nei frame culturali di riferimento.

Da una prospettiva metodologica, questi risultati hanno implicazioni rilevanti per approcci futuri. L'aggregazione semplice mediante majority voting risulterebbe inadeguata, trattando il disaccordo come errore anziché come segnale informativo sulla soggettività della percezione. L'adozione di framework che preservino e modellino esplicitamente la distribuzione delle annotazioni individuali appare necessaria per catturare la complessità del fenomeno e per sviluppare sistemi di rilevazione automatica che tengano conto della molteplicità delle prospettive presenti nel discorso pubblico.

Il corpus così costruito costituisce una risorsa per lo sviluppo di modelli computazionali informati dalla variabilità interpretativa, aprendo prospettive di ricerca sulla modellazione della soggettività nella percezione del razzismo online e sulla costruzione di sistemi che riflettano la pluralità di prospettive presenti nella società. La recente letteratura in ambito Natural Language Processing (NLP) ha iniziato a confrontarsi in modo sistematico con la pluralità di opinioni (e.g. Cabitza *et al.*, 2023), recuperando temi da tempo centrali nella statistica, come la modellazione della variabilità, dell'incertezza e dell'eterogeneità tra annotatori. In particolare, si possono distinguere due approcci complementari alla base della valutazione di modelli basati sul disaccordo (Uma *et al.*, 2021a, b): l'approccio del soft-label e l'approccio prospettivista (Leonardelli *et al.*, 2025).

Nel primo caso, il disaccordo tra annotatori viene interpretato come una distribuzione di giudizi a livello di popolazione, ed il modello è chiamato a stimare tale distribuzione. Nell'approccio prospettivista (Frenda *et al.*, 2025; Lo *et al.*, 2025), invece, il disaccordo viene modellato a livello individuale, e l'obiettivo diventa la previsione delle interpretazioni dei singoli annotatori. Sebbene questi approcci siano stati introdotti in ambito della computer science, essi si inseriscono naturalmente in una prospettiva statistica, ancora poco sviluppata ma necessaria, in quanto riconducono il disaccordo a una fonte strutturata di variabilità e richiedono strumenti di modellazione probabilistica per essere trattati in modo coerente e informativo.

Riferimenti bibliografici

Aroyo, L. and Welty, C. (2015), Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation, *AI Magazine*, 36(1): 15–24.

Badjatiya, P., Gupta, S., Gupta, M. and Varma, V. (2017), Deep learning for hate

ISBN Open Access: 9788835191971

speech detection in tweets. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 759–760). International World Wide Web Conferences Steering Committee.

Basile, V. and Nissim, M. (2013), Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2013)* (pp. 100–107).

Binns, R. (2018), Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency, volume 81 of Proceedings of Machine Learning Research* (pp. 149–159). PMLR.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), Latent dirichlet allocation, *Journal of Machine Learning Research*, 3: 993–1022.

Bosco, C., Dell'Orletta, F., Poletto, F., Sanguinetti, M. and Tesconi, M. (2018), EVALITA evaluation of NLP and speech tools for Italian. In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018)*, Turin, Italy. Accademia University Press.

Cabitza, F., Campagner, A. and Basile, V. (2023), Toward a perspectivist turn in ground truthing for predictive computing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6): 6860–6868.

Capozzi, A. T., Lai, M., Basile, V., Poletto, F., Sanguinetti, M., Bosco, C., Patti, V., Ruffo, G., Musto, C., Polignano, M., Semeraro, G. and Stranisci, M. (2020), "Contro L'odio": A platform for detecting, monitoring and visualizing hate speech against immigrants in Italian social media. *Italian Journal of Computational Linguistics*, 6(6–1): 39–66.

Cucco, A., del Gobbo, E., Fontanella, L., Fontanella, S. and Ippoliti, L. (2025a), Constructing fair and comprehensive dataset of online opinions for human annotations: considerations and challenges dealing with Italian content. In Boccuzzo, G., Bovo, E., Manisera, M. and Salmaso, L., eds., *Innovation & Society: Statistics and Data Science for Evaluation and Quality.* Book of Short Papers IES 2025, Brixen–Bressanone, Italy. CLEUP.

Cucco, A., del Gobbo, E., Fontanella, L., Fontanella, S. and Ippoliti, L. (2025b), *Covering the online spectrum of opinion in social context: The benefit of network node sampling through an Italian case study*. In International Conference on Computational Science (pp. 60–67). Springer, Berlino.

Davani, A. M., Díaz, M. and Prabhakaran, V. (2022), *Dealing with disagreements: Looking beyond the majority vote in subjective annotations* (vol. 10, pp. 92–110). MIT Press, Cambridge (USA).

Davidson, T., Bhattacharya, D. and Weber, I. (2019), Racial bias in hate speech and abusive language detection datasets. In *Proceedings of the Third Workshop on Abusive Language Online* (pp. 25–35), Florence, Italy. Association for Computational Linguistics.

Davidson, T., Warmsley, D., Macy, M. and Weber, I. (2017), Automated hate speech detection and the problem of offensive language. In *Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM)* (pp. 512–515).

Fontanella, L., Sarra, A., Del Gobbo, E., Cucco, A. and Fontanella, S. (2024),

ISBN Open Access: 9788835191971

Exploring anti-migrant rhetoric on Italian social media. In Plaia, A., Egidi, L. and Abbruzzo, A., eds., *Proceedings of the SDS 2024 Conference*, Palermo, Italy. Università degli Studi di Palermo.

Fortuna, P. and Nunes, S. (2018), A survey on automatic detection of hate speech in text, *ACM Computing Surveys (CSUR)*, 51(4): 1–30.

Frenda, S., Abercrombie, G., Basile, V., Pedrani, A., Panizzon, R., Cignarella, A. T., Marco, C. and Bernardi, D. (2025), Perspectivist approaches to natural language processing: a survey, *Language Resources and Evaluation*, 59: 1719–1746.

Gordon, M. L., Zhou, K., Patel, K., Hashimoto, T. and Bernstein, M. S. (2022), Jury learning: Integrating dissenting voices into machine learning models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). ACM.

Krippendorff, K. (2013), *Content Analysis: An Introduction to Its Methodology*. Sage Publications, Thousand Oaks, CA, 3 ed.

Leonardelli, E., Casola, S., Peng, S., Rizzi, G., Basile, V., Fersini, E. and Poesio, M. (2025), Lewidi-2025 at nlperspectives: The third edition of the learning with disagreements shared task. In *Proceedings of the 4th Workshop on Perspectivist Approaches to NLP* (pp. 182–195).

Lepri, B., Oliver, N., Letouze, E., Pentland, A. and Vinck, P. (2018), Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4): 611–627.

Lo, S. M., Casola, S., Sezerer, E., Basile, V., Sansonetti, F., Uva, A. and Bernardi, D. (2025), Perseval: A framework for perspectivist classification evaluation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing* (pp. 22345–22370).

Madeddu, M., Frenda, S., Lai, M., Patti, V. and Basile, V. (2023), Disaggreghate-It: Introducing a novel resource for hate speech detection in Italian leveraging annotators' disagreement. In *Proceedings of the 9th Italian Conference on Computational Linguistics (CLiC-it 2023)* (pp. 1–12), Venice, Italy. CEUR-WS.org.

Mathew, B., Saha, P., Yimam, S. M., Biemann, C., Goyal, P. and Mukherjee, A. (2021), HateXplain: A benchmark dataset for explainable hate speech detection. In *Proceedings of the AAAI Conference on Artificial Intelligence,* 35: 14867–14875.

Mozafari, M., Farahbakhsh, R. and Crespi, N. (2020), Hate speech detection and racial bias mitigation in social media based on BERT model, *PLOS ONE*, 15(8): e0237861.

Plank, B. (2022), The "problem" of human label variation: On ground truth in data, modeling and evaluation, *arXiv preprint arXiv:2211.02570*.

Samejima, F. (1969), Estimation of latent ability using a response pattern of graded scores, *Psychometrika*, 34(Suppl. 1): 1–97.

Sanguinetti, M., Comandini, G., Di Nuovo, E., Frenda, S., Stranisci, M., Bosco, C., Caselli, T., Patti, V. and Russo, I. (2020), Overview of the EVALITA 2020 Second hate speech detection task (HaSpeeDe 2). In *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020), Online. CEUR-WS.org.*

ISBN Open Access: 9788835191971

Sanguinetti, M., Poletto, F., Bosco, C., Patti, V. and Stranisci, M. (2018), An Italian Twitter corpus of hate speech against immigrants. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 2798–2805), Miyazaki, Japan. European Language Resources Association (ELRA).

Sap, M., Card, D., Gabriel, S., Choi, Y. and Smith, N. A. (2019), The risk of racial bias in hate speech detection. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1668–1678).

Sap, M., Swayamdipta, S., Vianna, L., Zhou, X., Choi, Y. and Smith, N. A. (2022), Annotators with attitudes: How annotator beliefs and identities bias toxic language detection. *In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 5884–5906), Seattle, United States. Association for Computational Linguistics.

Schmidt, A. and Wiegand, M. (2017), A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media* (pp. 1–10).

Tkachenko, M., Malyuk, M., Holmanyuk, A. and Liubimov, N. (2020-2025), *Label Studio: Data labeling software*. Open source software available from https://github.com/HumanSignal/label-studio.

Tontodimamma, A., Fontanella, L., Anzani, S. and Basile, V. (2023), An Italian lexical resource for incivility detection in online discourses, *Quality & Quantity*, 57(4): 3019–3037.

Uma, A., Fornaciari, T., Dumitrache, A., Miller, T., Chamberlain, J., Plank, B., Simpson, E. and Poesio, M. (2021a), Semeval-2021 task 12: Learning with disagreements. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)* (pp. 338–347), Online. Association for Computational Linguistics.

Uma, A., Fornaciari, T., Hovy, D., Paun, S., Plank, B. and Poesio, M. (2021b), Learning from disagreement: A survey, *Journal of Artificial Intelligence Research (JAIR)*, 72:1385–1470.

Vassallo, M., Gabrieli, G., Basile, V. and Bosco, C. (2019), The tenuousness of lemmatization in lexicon-based sentiment analysis. In *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)* (vol. 2481, pp. 1–6).

Vassallo, M., Gabrieli, G., Basile, V. and Bosco, C. (2020), Polarity imbalance in lexicon-based sentiment analysis. In *Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020)* (pp. 1–7).

Waseem, Z. and Hovy, D. (2016), Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. In Proceedings of the NAACL Student Research Workshop (pp. 88–93), San Diego, California. Association for Computational Linguistics.

ISBN Open Access: 9788835191971

Rappresentazioni causali dell'odio online: reti bayesiane per l'analisi dei discorsi razzisti

di *Franca Garreffa*[*], *Anthony Cossari*[**], *Paolo Carmelo Cozzucoli*[**], *Michelangelo Misuraca*[***]

1. Introduzione

Il discorso ostile rivolto a migranti e minoranze etniche ha assunto, negli ultimi anni, una centralità crescente nel dibattito pubblico, grazie alla diffusione capillare delle piattaforme digitali e all'intensificazione dei processi di polarizzazione online. La natura profondamente dinamica e frammentaria della comunicazione nei social media rende tuttavia complesso comprendere come dimensioni tematiche, strategie retoriche ed elementi emotivo-stilistici concorrano alla costruzione di forme discorsive riconoscibilmente razziste, xenofobe o discriminatorie. In questo quadro, l'analisi empirica delle strutture semantiche e dei meccanismi linguistico-pragmatici che sostengono tali fenomeni rappresenta non solo una sfida metodologica, ma anche un contributo scientifico rilevante per l'interpretazione delle dinamiche socioculturali che plasmano l'immaginario collettivo sulle minoranze.

Il presente studio si propone di intervenire in tale ambito attraverso l'elaborazione di un framework computazionale integrato, volto a identificare i temi ricorrenti nel discorso ostile e a modellare le loro dipendenze probabilistiche con categorie discorsive connesse al linguaggio d'odio. L'obiettivo è duplice: da un lato, si intende ricostruire la struttura semantica del *corpus* mediante tecniche di analisi di rete, evidenziando i cluster tematici che emergono dalle co-occorrenze lessicali; dall'altro, attraverso l'impiego di reti bayesiane, si mira a formalizzare le dinamiche inferenziali che presiedono

[*] Dipartimento di Scienze Politiche e Sociali, Università della Calabria, f.garreffa@unical.it

[**] Dipartimento di Economia, Statistica e Finanza "G. Anania", Università della Calabria, a.cossari@unical.it; cozzucoli@unical.it

[***] Dipartimento di Scienze Aziendali – Management & Innovation Systems, Università di Salerno, mmisuraca@unisa.it

ISBN Open Access: 9788835191971

alla coattivazione dei temi e delle categorie discorsive, consentendo di esplorare scenari controfattuali e configurazioni probabilistiche non immediatamente osservabili a livello testuale.

La struttura del capitolo riflette questa articolazione metodologica. Dopo aver illustrato il quadro teorico, si descrivono il *corpus* e le procedure di costruzione del dataset. Seguono la presentazione dell'analisi di rete, l'individuazione delle tematiche e la modellizzazione bayesiana, con particolare attenzione alle procedure inferenziali e alla loro interpretazione. Il lavoro si conclude discutendo i risultati ottenuti, le implicazioni teoriche e metodologiche dello studio e le possibili direzioni per ricerche future.

2. Quadro teorico e letteratura di riferimento

L'analisi del discorso razzista e xenofobo nei contesti digitali richiede un quadro teorico capace di integrare prospettive linguistiche, sociologiche e computazionali. Nella dimensione *online*, l'ostilità verbale si manifesta attraverso forme espressive concise, ridotte e fortemente valutative; nominalizzazioni stigmatizzanti, categorizzazioni rigide e polarizzazioni semantiche contribuiscono alla costruzione di un'immagine dell'alterità fondata su semplificazioni e giudizi negativi. Tale dinamica è coerente con quanto descritto da Allan e Burridge (2007), secondo cui il linguaggio del tabù opera mediante etichette che incorporano significati culturali sedimentati e strutture stereotipiche difficili da scardinare. Il funzionamento dei social media introduce ulteriori dimensioni interpretative. Da un lato, la letteratura sulla polarizzazione *online* ha mostrato come le piattaforme digitali favoriscano la formazione di ambienti comunicativi ideologicamente omogenei, in cui le opinioni tendono a radicalizzarsi e a rinforzarsi reciprocamente (Sunstein, 2018; Cinelli *et al.*, 2021). Dall'altro, le teorie delle rappresentazioni sociali (Moscovici, 1961; Joffe, 2003) offrono una cornice interpretativa per comprendere la persistenza e la trasformazione delle immagini collettive associate ai gruppi stigmatizzati, mostrando come contenuti emotivamente saturi – paura, disgusto, ostilità – si coagulino in narrazioni socialmente condivise, spesso riattivate e ricombinate nei contesti digitali. In questa prospettiva, l'odio *online* non è un mero epifenomeno dell'interazione digitale, ma un prodotto discorsivo radicato in strutture cognitive e culturali già presenti nello spazio sociale. A livello empirico, diversi studi italiani e internazionali hanno documentato come l'hate speech online si articoli attorno a repertori tematici relativamente stabili, connessi a frame quali la minaccia alla sicurezza, lo sfruttamento del welfare, il degrado urbano o la presunta incompatibilità culturale (Bosco *et al.*, 2023). Le ricerche sulla violenza verbale e sulla radicalizza-

ISBN Open Access: 9788835191971

zione discorsiva (Humprecht, Hellmueller e Lischka, 2020) mostrano che tali repertori tendono a rafforzarsi all'interno di comunità digitali dense, dove i processi di omofilia e selezione algoritmica contribuiscono a stabilizzare schemi retorici ostili e a ridurre l'esposizione a prospettive alternative.

Diversi studi hanno approfondito il funzionamento del linguaggio d'odio razziale nei contesti digitali, rilevando come esso si configuri come pratica discorsiva altamente adattiva. Davidson *et al.* (2017) hanno mostrato come razzismo, xenofobia e insulti generalizzati tendano a sovrapporsi in modo non lineare, rendendo complessa la distinzione automatica fra aggressione, offensività e discriminazione. Studi sulle piattaforme di microblogging (Chung *et al.*, 2019; Ekman, 2019; Udanor e Anyanwu, 2019) hanno inoltre mostrato che le comunità *online* che producono contenuti razzisti tendono a sviluppare repertori espressivi riconoscibili, caratterizzati da lessico ricorrente, *pattern* di amplificazione intragruppo e dinamiche di radicalizzazione progressiva. In questo quadro teorico, il presente studio si propone di contribuire alla comprensione delle dinamiche attraverso cui il discorso razzista e xenofobo si struttura e si diffonde *online*, integrando l'analisi dei contenuti con una modellizzazione probabilistica orientata a cogliere la complessità delle relazioni che legano temi, forme stilistiche ed espressioni di ostilità.

3. Caratteristiche del *corpus* analizzato

Il *corpus* analizzato è costituito da una raccolta di messaggi pubblicati su Twitter, selezionati in funzione della loro rilevanza rispetto alle forme di ostilità rivolte a due specifici gruppi sociali: i migranti (con particolare riferimento a quelli di religione musulmana) e la comunità rom. Il set di messaggi è derivato dalle risorse rese disponibili nell'ambito dei task *HaSpeeDe 2018* (Bosco *et al.*, 2018) e *HaSpeeDe 2020* (Sanguinetti *et al.*, 2020) di *EVALITA*, campagna di valutazione periodica di strumenti per il trattamento del linguaggio naturale per la lingua italiana promossa dal 2007 dall'AILC (*Associazione Italiana per la Linguistica Computazionale*). Tali risorse rappresentano, allo stato attuale, uno dei principali riferimenti per lo studio computazionale dell'odio razziale e xenofobo nel contesto linguistico italiano.

Il dataset originale è costituito da 8.001 tweet annotati manualmente per la presenza o assenza di contenuti d'odio e di stereotipi razzisti e xenofobi. Oltre alle categorie principali, il *corpus* comprende una ricca stratificazione di metadati e marcatori linguistici (aggressività, ironia e sarcasmo), tra cui la standardizzazione di menzioni e URL e l'identificativo univoco di ciascun messaggio. Tale struttura consente non solo un'esplorazione quantitativa dei contenuti ostili, ma anche l'analisi dei legami concettuali e semantici tra for-

ISBN Open Access: 9788835191971

me di discriminazione, costrutti interculturali e marcatori emotivo-stilistici, fornendo un quadro ricco e articolato per l'indagine delle dinamiche del discorso razzista e xenofobo nel contesto italiano contemporaneo. La collezione integra due insiemi distinti: il primo sottototale è composto da 3.900 tweet raccolti tra ottobre 2016 e aprile 2017 per *HaSpeeDe 2018*, mentre il secondo sottototale è composto da 4.101 tweet provenienti dal progetto *Contro l'Odio* (Capozzi *et al.*, 2019), sviluppato tra settembre 2018 e maggio 2019. La dimensione annotativa riveste un ruolo centrale. La collezione è caratterizzata da una forte variabilità interna, riconducibile alla natura del mezzo di comunicazione e al periodo temporale di raccolta: i contenuti sono spesso legati a eventi contingenti, variazioni di agenda mediatica e cicli di attenzione pubblica, caratteristiche che contribuiscono a rendere il *corpus* particolarmente adatto allo studio della dinamica discorsiva e dei suoi mutamenti diacronici.

Dal punto di vista quantitativo, la distribuzione delle categorie annotate riflette una presenza significativa ma non predominante di contenuti d'odio: nei tweet inclusi nei compiti di classificazione binaria, circa il 41,81% dei testi è etichettato come messaggio d'odio, mentre il 44,17% presenta forme stereotipate. Relativamente ai marcatori linguistici, si rilevano invece in circa il 35,65% dei messaggi un tono aggressivo, nel 18,75% un tono ironico e nel 11,47% un tono sarcastico. Oltre alle dimensioni quantitative e annotative, il corpus presenta una serie di caratteristiche sociolinguistiche e pragmatiche che ne definiscono la specificità e ne accrescono l'interesse per un'analisi articolata dei discorsi razzisti e xenofobi. Dal punto di vista sociolinguistico, la natura digitale dei testi raccolti su Twitter comporta un'elevata densità di tratti tipici della comunicazione mediata da computer, quali abbreviazioni, marcatori discorsivi ridotti, emoticon, grafie deviate intenzionalmente e forme di variazione diastratica e diafasica che riflettono l'identità linguistica degli utenti.

La compresenza di registri diversi, dal colloquiale informale all'ipersintetico tipico delle interazioni rapide, costituisce un elemento strutturale della comunicazione ostile sui social media e richiede un trattamento specifico nelle successive fasi di analisi. Sul piano pragmatico, il *corpus* documenta una notevole varietà di atti linguistici ostili, che vanno dall'insulto diretto alla denigrazione attraverso insinuazioni e presupposizioni. Particolarmente rilevante è la presenza di enunciati nominali, una forma espressiva che riduce la struttura sintattica al minimo indispensabile, incrementando la forza illocutiva del messaggio attraverso l'immediatezza categoriale ("animali", "parassiti", "clandestini"). Tali enunciati, già evidenziati nella letteratura come indicatori tipici della retorica ostile, svolgono una funzione classificatoria che permette al parlante di attribuire in modo rapido e implicito tratti

ISBN Open Access: 9788835191971

stigmatizzanti al target, riducendo lo spazio interpretativo dell'interlocutore. Un ulteriore asse interpretativo è costituito dal legame dei contenuti del *corpus* agli eventi sociopolitici del periodo di raccolta.

L'andamento temporale dei tweet riflette infatti cicli di attenzione mediatica legati a temi come sbarchi, attacchi terroristici, provvedimenti normativi, campagne politiche o fatti di cronaca. Questa dimensione contestuale produce condensazioni discorsive in cui l'ostilità verso i gruppi target si intreccia con frame narrativi più ampi (insicurezza urbana, crisi del welfare, conflitti culturali), configurando un terreno fertile per l'analisi delle rappresentazioni sociali che strutturano l'odio online. La combinazione di tali aspetti rende il *corpus* particolarmente adatto a un'analisi multilivello: da un lato permette di esplorare la materialità del testo e i meccanismi retorici attraverso cui si costruisce l'ostilità, dall'altro consente di indagare le strutture concettuali e ideologiche che emergono dall'organizzazione tematica dei contenuti. La ricchezza stratificata del dataset, unita alla sua eterogeneità interna, lo qualifica come una fonte privilegiata per modellare le dinamiche discorsive dell'odio in ambiente digitale.

4. Metodologia e strategia analitica

L'impianto metodologico del presente studio si struttura come un *workflow* computazionale integrato, volto a identificare e modellare in termini probabilistico-causali le dinamiche interne al discorso razzista e xenofobo rivolto a migranti e comunità rom in un contesto digitale.

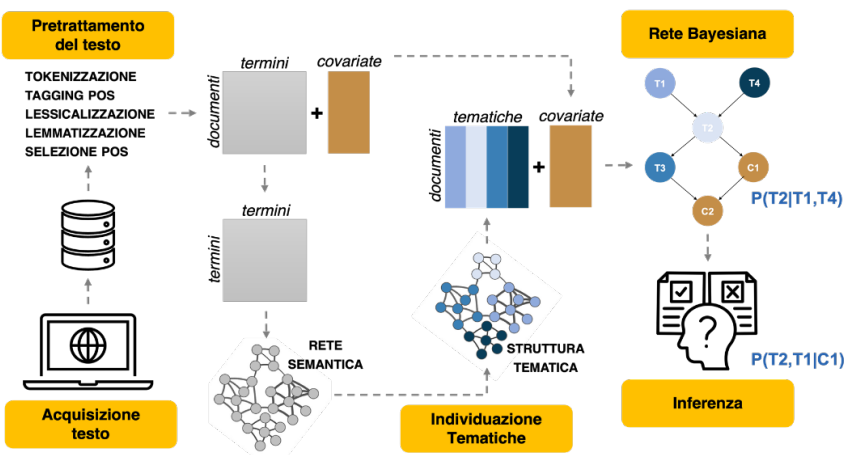


Fig. 1- Architettura della strategia analitica utilizzata nello studio

ISBN Open Access: 9788835191971

La strategia analitica è articolata in quattro fasi principali: (1) pretrattamento linguistico, (2) analisi di rete del contenuto, (3) individuazione delle tematiche tramite algoritmi di community detection e (4) modellizzazione del discorso tramite reti bayesiane (Fig. 1). Il punto di partenza, dopo aver acquisito e archiviato i testi, consiste nell'elaborazione preliminare del *corpus*, attraverso un trattamento linguistico sistematico volto a garantire coerenza interna e adeguatezza formale dei dati (Bolasco, 2013). Tale trattamento comprende operazioni canoniche della linguistica computazionale, quali la segmentazione del testo in unità elementari (*token*), l'assegnazione delle categorie morfosintattiche (*parti del discorso*, POS), la normalizzazione morfologica mediante lemmatizzazione e la selezione degli elementi lessicali più informativi sulla base di criteri sintattici e frequenziali (Manning e Schütze, 1999).

L'esito di tali operazioni è la costruzione di una rappresentazione matriciale documenti-termini, utile a stabilire una base stabile per l'analisi successiva delle relazioni tra unità lessicali. A partire da tale rappresentazione è possibile derivare una rete semantica fondata sulle co-occorrenze dei termini nei documenti (Misuraca e Spano, 2020). In questa rete, i nodi rappresentano le unità lessicali e gli archi indicano la loro compresenza contestuale, con ponderazioni che riflettono la frequenza o la rilevanza statistica delle associazioni.

La struttura risultante è tipicamente caratterizzata da fenomeni di aggregazione concettuale, che emergono come sottogruppi densamente connessi. L'identificazione di tali sottostrutture avviene mediante algoritmi di *community detection* (Fortunato, 2010), capaci di mettere in evidenza cluster tematici senza imporre *a priori* categorie semantiche. I cluster individuati definiscono una struttura tematica che consente di costruire una nuova rappresentazione matriciale, questa volta riferita alla distribuzione dei temi all'interno del *corpus*, ottenuta attraverso l'assegnazione dei documenti alle comunità linguistiche rilevate (Misuraca, Scepi e Spano, 2023).

La rappresentazione tematica così ottenuta viene integrata con eventuali informazioni aggiuntive – provenienti ad esempio da covariate discorsive o emotive – concatenate alla matrice documenti-tematiche per costituire una struttura analitica unica. Tale integrazione consente di modellare simultaneamente dimensioni semantiche, linguistiche e paralinguistiche in un unico spazio informativo. Su questa base viene quindi identificata, tramite addestramento automatico e successiva correzione secondo l'approccio system expert, una rete bayesiana, ossia un modello probabilistico capace di rappresentare le dipendenze condizionali tra variabili attraverso una struttura diretta aciclica (Heckerman, 1997).

Una rete bayesiana è un grafo aciclico orientato (DAG) in cui i nodi

ISBN Open Access: 9788835191971

rappresentano variabili aleatorie e gli archi rappresentano dipendenze probabilistiche dirette; a ciascuna variabile è associata una distribuzione di probabilità condizionata ai suoi genitori nel grafo, che specifica come essa dipende probabilisticamente da questi (Pearl, 1988). L'apprendimento automatico della struttura della rete può essere fatto tramite algoritmi *score-based*, che ricercano la configurazione più plausibile sulla base di criteri informativi capaci di bilanciare qualità di adattamento e penalizzazione della complessità. L'introduzione di vincoli sulla struttura delle dipendenze probabilistiche, sotto forma di *blacklist*, permette di escludere dipendenze causalmente implausibili o contrarie alla logica del disegno di ricerca (ad esempio l'insorgenza retroattiva di covariate come determinanti diretti dei temi), assicurando che il modello rifletta adeguatamente le ipotesi epistemologiche di partenza (Koller e Friedman, 2009). La validazione finale degli archi presenti nel modello appreso viene condotta tramite test di indipendenza condizionale con correzione *Monte Carlo* (Scutari e Denis, 2021), al fine di mitigare il rischio di falsi positivi e garantire robustezza inferenziale anche in presenza di distribuzioni sbilanciate o variabili poco rappresentate.

La rete bayesiana consente di descrivere in modo compatto l'intreccio delle dipendenze che caratterizzano l'insieme delle variabili considerate e permette di condurre interrogazioni inferenziali, sia di tipo prognostico che diagnostico, finalizzate alla stima di probabilità condizionate, alla valutazione di scenari controfattuali e all'analisi delle configurazioni più rilevanti dal punto di vista interpretativo. La strategia analitica complessiva risulta così orientata a produrre un modello coerente, formalmente interpretabile e sensibile alla complessità delle dinamiche discorsive, capace di coniugare l'indagine strutturale delle occorrenze linguistiche con la modellizzazione probabilistica delle loro relazioni interne e, soprattutto, utile a ottenere risposte, in termini probabilistici, a domande conoscitive complesse.

4.1. Struttura del modello e meccanismi di inferenza

La rete bayesiana adottata nel presente studio è costruita a partire dalle variabili tematiche derivate dall'analisi di rete e da quelle emotivo-stilistiche usate per annotare il *corpus*. Formalmente, la rete è definita come $\mathcal{B} = (\mathcal{G}, \boldsymbol{\Theta})$, dove il grafo diretto aciclico $\mathcal{G} = (\mathrm{V}, \mathrm{A})$ è costruito sul vettore di variabili $\mathbf{X} = (\mathrm{T}_1, \ldots, \mathrm{T}_\mathrm{k}, \mathrm{C}_1, \ldots, \mathrm{C}_\mathrm{m})$, comprendente i k temi emergenti e le m covariate discorsive. Ogni nodo $\mathrm{X}_\mathrm{i} \in \mathrm{V}$ rappresenta una variabile discreta ottenuta dalla discretizzazione delle intensità tematiche o dalla codifica categoriale delle covariate, mentre gli archi A riflettono dipendenze probabilistiche compatibili con il dominio empirico e con i vincoli epistemologici

ISBN Open Access: 9788835191971

imposti tramite *blacklist* per evitare relazioni invertite o teoricamente implausibili. La componente quantitativa del modello, ottenuta tramite la fattorizzazione della distribuzione di probabilità globale delle variabili del dominio, definita da $\boldsymbol{\Theta} = (\theta_1, \ldots, \theta_p)$, specifica le distribuzioni di probabilità locali: per ogni variabile X_i con insieme dei genitori $Pa(X_i)$, la distribuzione locale assume la forma $P(X_i \mid Pa(X_i); \theta_i)$.

La struttura del grafo è appresa automaticamente mediante l'algoritmo *Hill-Climbing* (Tsamardinos, Brown e Aliferis, 2006) con funzione obiettivo BIC (*Bayesian Information Criterion*), che permette di identificare configurazioni capaci di rappresentare in modo parsimonioso e interpretabile le relazioni emerse. Definito il grafo $\mathcal{G}$, la rete, tramite fattorizzazione, decompone la distribuzione congiunta di variabili tematiche e covariate:

$$P(\mathbf{X}; \boldsymbol{\Theta}) = \prod_{i=1}^{p} P(X_i \mid Pa(X_i); \theta_i),$$

riducendo così una distribuzione multivariata ad alta dimensionalità a un insieme di distribuzioni univariate condizionate. Tale decomposizione costituisce un passaggio cruciale nel *workflow*, in quanto consente di trasformare la complessità semantica ottenuta dalle reti di co-occorrenza e dagli algoritmi di *community detection* in un modello probabilistico compatto, capace di fornire una rappresentazione formalizzata delle interrelazioni discorsive.

Le distribuzioni locali sono stimate a partire dai dati campionari mediante il criterio bayesiano BDeu (*Bayesian Dirichlet Equivalent Uniform*), che in generale è da preferire, mentre la significatività statistica degli archi del grafo è verificata attraverso dei test di indipendenza condizionale con correzione Monte Carlo, passaggio essenziale per garantire che le dipendenze identificate riflettano pattern statistici genuini e non artefatti della combinazione tra frequenze tematiche e covariate emotivo-stilistiche. Una volta appresa e validata, la rete bayesiana consente di condurre inferenze di tipo prognostico, diagnostico, intercausali e combinate (Korb e Nicholson, 2023) del tipo $P(Z \mid Y = y)$, dove Y può rappresentare, ad esempio, una configurazione tematica multipla e Z una covariata discorsiva. Grazie a questa struttura, la rete permette di esplorare non solo la probabilità dei fenomeni osservati, ma anche scenari controfattuali e configurazioni ipotetiche, integrando in un quadro unitario l'informazione estratta dalle reti semantiche e dalla classificazione tematica con la modellizzazione probabilistica delle dipendenze.

5. Risultati empirici

ISBN Open Access: 9788835191971

Il dataset analizzato, dopo una fase di screening e pulizia, comprende 7.994 messaggi. Sono state considerate, unitamente al testo, le categorie "odio", "stereotipia", "aggressività" e "ironia". La categoria "sarcasmo" è stata invece esclusa dall'analisi, alla luce delle difficoltà ampiamente documentate dalla letteratura linguistica e computazionale nel distinguere in modo univoco tali fenomeni e nel procedere a una loro annotazione affidabile. Numerosi studi hanno infatti mostrato come il confine tra ironia e sarcasmo sia teoricamente sfumato ed empiricamente instabile, con conseguenti livelli di accordo tra annotatori spesso insufficienti a garantire una classificazione coerente (Filatova, 2012).

Dopo la fase di pretrattamento e la selezione delle sole categorie morfosintattiche rilevanti – nomi propri, sostantivi e aggettivi – è stata ottenuta una matrice documenti-termini di dimensione 7.994 × 2.953. A partire da tale matrice, mediante una sua trasformazione binaria che registra esclusivamente la presenza o l'assenza dei termini nei singoli messaggi, è stata costruita una matrice di co-occorrenza, finalizzata a rilevare la compresenza delle 2.953 unità lessicali all'interno dei messaggi. Tale struttura consente di quantificare le associazioni tra coppie di termini e costituisce la base per la successiva rappresentazione reticolare delle relazioni semantiche. La rete semantica ottenuta, costruita eliminando gli archi con peso inferiore a 7 compresenze, presenta una struttura complessivamente rada ma fortemente organizzata. A seguito di tale soglia, sono stati inoltre rimossi i nodi isolati, ossia i termini che non presentavano più connessioni significative, riducendo la rete ai 649 nodi effettivamente coinvolti in relazioni ricorrenti. I 649 nodi e i 647 archi evidenziano che solo una parte del lessico dà luogo a combinazioni sistematiche, come confermato dalla bassa densità (0,067). Nonostante tale rarefazione, il valore della *closeness* (0,270) indica che le distanze medie tra i termini rimangono contenute, mentre la *betweenness* (0,203) segnala la presenza di nodi che svolgono una funzione di mediazione tra aree semantiche altrimenti poco connesse. Il coefficiente di clustering (0,188) evidenzia la formazione di piccoli nuclei di termini che tendono a co-occorrere tra loro con maggiore frequenza, suggerendo la presenza di combinazioni lessicali relativamente stabili. L'elemento più rilevante è tuttavia l'elevata modularità (0,725), che indica una marcata suddivisione in comunità interne coese e debolmente collegate tra loro. Tale configurazione riflette un'organizzazione tematica tipica di domini discorsivi in cui i contenuti si articolano in gruppi concettuali distinti e poco intercomunicanti.

La fase di *community detection* è stata condotta applicando l'algoritmo *Walktrap* (Pons e Latapy, 2005), una procedura basata sull'idea che brevi cammini casuali all'interno di una rete tendano a rimanere confinati in regioni fortemente connesse. Questo principio consente di individuare gruppi

ISBN Open Access: 9788835191971

di nodi che condividono pattern di co-occorrenza più intensi rispetto al resto della rete, riflettendo quindi affinità semantiche e ricorrenze discorsive coerenti. L'impiego del Walktrap presenta diversi vantaggi in un contesto come quello studiato. Anzitutto, l'algoritmo è particolarmente efficace nell'identificare comunità anche in reti relativamente rade, come le reti lessicali, nelle quali le connessioni significative sono distribuite in maniera non uniforme. Inoltre, la sua struttura gerarchica consente di ottenere partizioni della rete a diversi livelli di granularità, offrendo una rappresentazione flessibile e adattabile delle relazioni semantiche. Il metodo risulta altresì meno sensibile a oscillazioni locali rispetto ad altri algoritmi basati esclusivamente sulla massimizzazione della modularità, restituendo così comunità più stabili e coerenti dal punto di vista interpretativo. L'applicazione del Walktrap al grafo di co-occorrenza ha portato all'identificazione di 24 comunità, ciascuna interpretabile come un nucleo tematico distinto (Fig. 2).

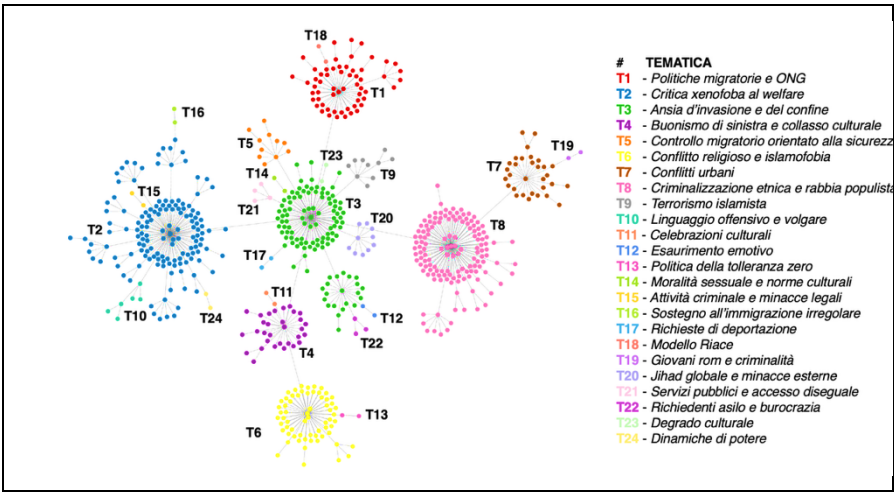


Fig. 2 – Struttura tematica del discorso razzista e xenofobo

La dimensione delle community, calcolata sul numero di termini effettivamente assegnati a ciascun cluster, riflette una struttura semantica fortemente eterogenea, articolata in poli tematici maggiormente estesi e in insiemi più compatti caratterizzati da una coesione lessicale elevata. La community più ampia è T2 (144 nodi), incentrata sulla critica xenofoba al welfare e fortemente caratterizzata da un lessico polarizzato sulla rappresentazione dei migranti come beneficiari indebiti delle risorse pubbliche, come attestano termini quali *finto profugo*, *tasse* e *accoglienza*. Seguono T8 (139 nodi), relativa alla criminalizzazione etnica e alla rabbia populista, che include espressioni come *rom*, *delinquente* e *quartiere*, e T3 (113 nodi), che articola un frame allarmistico

ISBN Open Access: 9788835191971

incentrato su invasioni, minacce al confine e insicurezze demografiche, con parole quali *invasione*, *barcone*, *clandestini* e *sbarco*.

A queste si affiancano cluster tematici di medie dimensioni, come T6 (64 nodi), legato alla retorica islamofoba e al conflitto religioso, in cui ricorrono termini come *Islam*, *moschea*, *Sharia* e *cristiano*; e T1 (59 nodi), dedicato alle politiche migratorie e al ruolo delle ONG nelle operazioni di soccorso, dove compaiono parole quali *nave*, *Lampedusa*, *Mediterraneo* e *soccorsi*. Altre comunità presentano una struttura più ridotta ma non meno significativa dal punto di vista semantico. T4 (31 nodi) rappresenta il tema del "buonismo" di sinistra e del presunto collasso culturale, con termini quali *buonista*, *idiota* e *ignorante*, mentre T7 (31 nodi) concentra il lessico relativo ai conflitti urbani, agli sgomberi e ai problemi abitativi, come *campo rom* e *degrado*. Tematiche più specifiche emergono in cluster ancora più compatti, come T5 (11 nodi), centrato sul controllo migratorio orientato alla sicurezza (*decreto, confine*), o T9 (10 nodi), relativo al terrorismo islamista (*attentato, strage, terrorista*). Una serie di community minori – T10 (6 nodi), T11-T19 (2 nodi ciascuna), T20 (10 nodi), T21 (4 nodi), T22 (3 nodi), T23 (2 nodi) e T24 (2 nodi) – aggregano porzioni di lessico altamente specializzate, spesso riferibili a frame discorsivi circoscritti: dal linguaggio offensivo e volgare (T10, con termini quali *cazzo*, *merda*, *vaffanculo*) alle pratiche burocratiche dei richiedenti asilo (T22, con termini come *asilo*), fino alle narrazioni del degrado urbano e culturale (T23, *cesso*). Questa articolazione multilivello consente di cogliere la complessità del discorso ostile, evidenziando tanto i nodi tematici centrali quanto le microaree semantiche in cui si condensano forme più sottili di ostilità, stereotipizzazione e costruzione ideologica.

La matrice documenti-tematiche è stata costruita a partire dalla matrice documenti-termini binarizzata, ottenuta registrando esclusivamente la presenza o assenza delle unità lessicali all'interno di ciascun messaggio. Una volta identificati i termini appartenenti alle ventiquattro comunità tematiche, è stata estratta la porzione della matrice che include esclusivamente tali unità lessicali, così da limitare l'analisi alle componenti semantiche rilevanti. In un secondo passaggio, i termini sono stati aggregati secondo la loro appartenenza comunitaria: per ogni messaggio, è stato calcolato il numero di termini presenti riconducibili alla medesima tematica, ottenendo una misura dell'intensità tematica del messaggio rispetto a ciascuno dei cluster identificati. Il risultato finale è una matrice dove ogni riga rappresenta un messaggio e ogni colonna una delle ventiquattro tematiche; le celle quantificano il contributo tematico del messaggio sulla base della frequenza di termini appartenenti a ciascuna community. Per garantire una maggiore stabilità statistica nelle fasi successive della modellizzazione, tali valori sono stati discretizzati in quattro livelli ordinali, riflettendo gradazioni crescenti di attivazione tematica.

ISBN Open Access: 9788835191971

La matrice così costruita è stata infine integrata con le covariate discorsive ed emotive estratte in precedenza, dando luogo a un dataset finale strutturato, pronto per la fase di apprendimento della rete bayesiana.

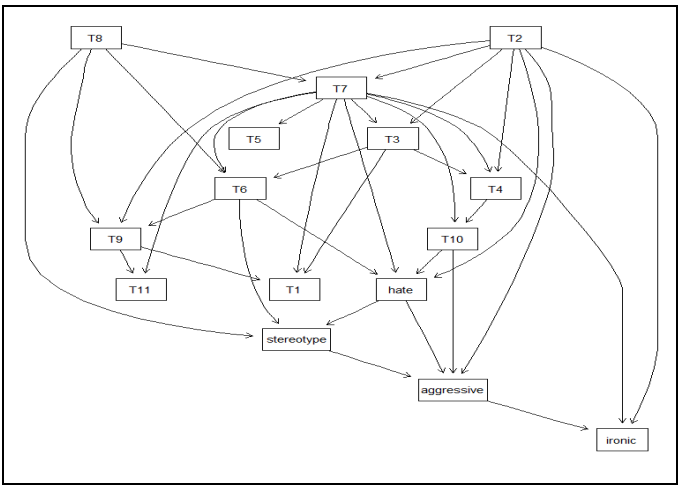


Fig. 3 – Struttura della rete bayesiana stimata, con relazioni dirette tra tematiche e covariate discorsive nel corpus analizzato

La costruzione della rete ha impiegato l'algoritmo *Hill-Climbing* (HC) con criterio di score BIC. L'algoritmo HC, uno dei metodi più usati per l'apprendimento automatico delle reti bayesiane, consiste nell'esplorazione dello spazio dei grafi possibili mediante tre operazioni: aggiungere, rimuovere o invertire un arco, accettando solo le modifiche che migliorano lo *score* scelto in fase di addestramento. L'addestramento automatico ha previsto l'uso di una *blacklist* volta a impedire relazioni metodologicamente implausibili tra tematiche e variabili discorsive. Una volta selezionati i soli temi con una presenza pari ad almeno il 4% dei messaggi – al fine di evitare la distorsione introdotta da comunità marginali o legate a eventi contingenti – è stata addestrata una struttura diretta aciclica capace di rappresentare le dipendenze condizionali tra dimensioni tematiche e covariate emotivo-stilistiche (Fig. 3).

La significatività statistica degli archi nella rete è stata valutata con un test di indipendenza condizionale con correzione Monte Carlo, garantendo robustezza inferenziale anche con distribuzioni asimmetriche. L'analisi condotta sul modello ha consentito di esplorare come la presenza/assenza di specifiche configurazioni tematiche influenzi la probabilità di osservare categorie discorsive rilevanti, quali "odio", "aggressività" o "stereotipia".

Tab. 1 – Probabilità condizionate dell'odio, dati i temi T2, T6, T7, T10 (valori >0,50)

ISBN Open Access: 9788835191971

Nodo	Stato	Condizione	Probabilità
odio	1	*T2*=1, *T6*=1, *T7*=1, *T10*=1	0,91
odio	1	*T2*=1, *T6*=1, *T7*=0, *T10*=1	0,78
odio	1	*T2*=1, *T6*=0, *T7*=0, *T10*=1	0,72
odio	1	*T2*=0, *T6*=1, *T7*=1, *T10*=1	0,69
odio	1	*T2*=0, *T6*=0, *T7*=0, *T10*=1	0,68
odio	1	*T2*=1, *T6*=1, *T7*=0, *T10*=0	0,61
odio	1	*T2*=0, *T6*=1, *T7*=0, *T10*=0	0,55
odio	1	*T2*=1, *T6*=0, *T7*=1, *T10*=1	0,53

Un primo esempio (tab. 1), riguarda la stima della probabilità di esprimere contenuti d'odio data l'attivazione congiunta di quattro temi ad alta intensità discriminatoria: *Critica xenofoba al welfare* (T2), *Conflitto religioso e islamofobia* (T6), *Conflitti urbani* (T7) e *Linguaggio offensivo e volgare* (T10). I valori riportati indicano come la probabilità di osservare espressioni d'odio sia fortemente modulata dall'attivazione selettiva delle quattro tematiche coinvolte. Il confronto fra configurazioni differenziate dalla presenza o assenza di una sola variabile mostra con chiarezza che alcuni temi esercitano un impatto più incisivo di altri. La presenza congiunta di tutte le tematiche produce la probabilità massima (0,91), configurandosi come uno scenario discorsivo altamente favorevole all'emergere di contenuti ostili. L'eliminazione di un singolo tema consente di stimarne il contributo marginale. La rimozione di T7 determina una riduzione sensibile della probabilità, che scende da 0,91 a 0,78. Ciò suggerisce che T7 (*Conflitti urbani*) contribuisce in modo rilevante all'intensificazione del *frame* ostile nelle configurazioni in cui è attivato insieme a T2 e T10. La rimozione di T6 (*Conflitto religioso e islamofobia*) comporta un'ulteriore diminuzione (da 0,78 a 0,72), indicando che tale tema opera come amplificatore della retorica d'odio in combinazione con T2 e T10, ma con impatto inferiore rispetto a T7. L'assenza di T2 (*Critica xenofoba al welfare*), a parità di presenza degli altri temi, produce una riduzione più marcata (da 0,91 a 0,69), segnalando che T2 mantiene un ruolo strutturale nel sostenere contenuti ostili: si tratta infatti di un tema sistematicamente associato a narrative di colpevolizzazione e risentimento economico, spesso centrali nella costruzione dell'*hate speech*. Infine, il confronto fra condizioni che differiscono solo per l'attivazione di T10 (*Linguaggio offensivo e volgare*) evidenzia che la sua presenza determina un incremento di probabilità rispetto alle configurazioni analoghe prive di tale tema. Sebbene l'effetto non sia sempre di ampiezza elevata, T10 agisce come catalizzatore stilistico che stabilizza il registro espressivo aggressivo, facilitando l'emergere di contenuti di odio anche quando altri temi discriminatori sono assenti o debolmente attivati.

Tab. 2 – Probabilità condizionate dell'aggressività verbale, dati linguaggio d'odio, stereotipia, e i temi T2, T10 (valori >0,50)

ISBN Open Access: 9788835191971

Nodo	Stato	Condizione	Probabilità
aggressività	1	*odio*=1, *stereotipia*=1, *T2*=1, *T10*=1	0,90
aggressività	1	*odio*=1, *stereotipia*=1, *T2*=0, *T10*=1	0,85
aggressività	1	*odio*=1, *stereotipia*=0, *T2*=1, *T10*=1	0,83
aggressività	1	*odio*=1, *stereotipia*=0, *T2*=0, *T10*=1	0,75
aggressività	1	*odio*=1, *stereotipia*=1, *T2*=1, *T10*=0	0,70
aggressività	1	*odio*=1, *stereotipia*=1, *T2*=0, *T10*=0	0,65
aggressività	1	*odio*=1, *stereotipia*=0, *T2*=1, *T10*=0	0,63
aggressività	1	*odio*=1, *stereotipia*=0, *T2*=0, *T10*=0	0,57
aggressività	1	*odio*=0, *stereotipia*=1, *T2*=1, *T10*=1	0,54

Un secondo esempio (tab. 2) riguarda la probabilità di osservare aggressività verbale dato un insieme di condizioni tematiche e stilistiche, nello specifico, la presenza simultanea di "odio", "stereotipia", T2 e T10. Anche in questo caso, gli enunciati esemplificativi mostrano come l'intreccio fra criminalizzazione etnica, insulto diretto e retoriche globaliste produca configurazioni discorsive particolarmente aggressive. I risultati relativi alla probabilità di produrre enunciati aggressivi mostrano una struttura influenzata in modo significativo dalla combinazione dei quattro predittori considerati: odio, stereotipia, T2 (*Critica xenofoba al welfare*) e T10 (*Linguaggio offensivo e volgare*). La configurazione completa, in cui tutte le variabili sono attive, genera la probabilità massima (0,90), delineando un contesto discorsivo in cui la dimensione aggressiva si inserisce con elevata coerenza all'interno di un registro polemico dominato da ostilità esplicita, stereotipizzazione e lessico insultante. L'eliminazione progressiva di una singola variabile consente di stimare il contributo specifico di ciascun predittore. La rimozione di T2 comporta una riduzione della probabilità da 0,90 a 0,85, indicando che la cornice redistributiva e accusatoria tipica di T2 contribuisce in modo rilevante a sostenere un tono aggressivo, pur non essendo il fattore determinante. Un effetto analogo si osserva per la stereotipia: la sua assenza porta a una diminuzione ulteriore (da 0,85 a 0,83), suggerendo che gli stereotipi operino come fattori di amplificazione emotiva, rafforzando le configurazioni ostili già orientate verso l'aggressività. L'impatto dell'assenza simultanea di stereotipia e T2 (pur mantenendo attivi l'odio e T10) risulta più marcato (da 0,83 a 0,75), mostrando che tali elementi agiscono sinergicamente nel sostenere un registro aggressivo, poiché contribuiscono alla costruzione di giudizi generalizzanti e delegittimanti. Il ruolo di T10 appare particolarmente rilevante nella seconda metà della distribuzione: quando T10 è disattivato, la probabilità si riduce sistematicamente, con un decremento significativo nelle configurazioni corrispondenti (ad esempio da 0,70 a 0,65, e poi a 0,63 e 0,57 al variare degli altri predittori). Questo indica che il linguaggio volgare e insultante costituisce un tratto stilistico che stabilizza l'aggressività, contribuendo alla sua espressione anche in assenza di alcuni

ISBN Open Access: 9788835191971

contenuti tematici. Il confronto tra condizioni con odio disattivato ma con gli altri predittori attivi (es. 0,54 in presenza di stereotipia, T2 e T10) mostra che l'aggressività può emergere anche in contesti non esplicitamente marcati da hate speech. Ciò suggerisce che la dimensione aggressiva non è semplicemente un sottoprodotto dell'odio, ma può essere sostenuta autonomamente da configurazioni stilistico-tematiche caratterizzate da stereotipia, attacchi al welfare e linguaggio offensivo.

Tab. 3 – Esempi di inferenza sulla rete bayesiana addestrata

a) P(• |odio)				***b) P(• |stereotipia)***				***c) P(• |aggressività)***			
		T8				**odio**				**stereotipia**	
		0	1			0	1			0	1
T6	0	0,26	0,18	**T7**	0	0,24	0,07	**T8**	0	0.13	0,29
	1	***0,40***	0,16		1	***0,64***	0,05		1	***0,16***	0,42

In sintesi, la rete bayesiana consente di formalizzare e quantificare le relazioni probabilistiche fra dimensioni tematiche e categorie discorsive, offrendo una rappresentazione interpretabile delle dinamiche interne del discorso ostile. Attraverso l'inferenza condizionata è possibile esplorare scenari controfattuali, individuare pattern ricorrenti e mettere in evidenza i meccanismi che conducono alla coalescenza di specifici temi attorno a nuclei di ostilità, aggressività o stereotipizzazione. Gli esempi in tab. 3 mostrano alcune interrogazioni possibili all'interno della rete, utili per mostrare come la struttura probabilistica del modello consenta di stimare in modo formalizzato la dipendenza tra temi e categorie discorsive. Condizionando sulla presenza di odio (tab. 3a), si osserva ad esempio che la probabilità di attivare il tema T6 (*Conflitto religioso e islamofobia*) raggiunge 0,40 quando T8 (*Criminalizzazione etnica e rabbia populista*) è assente, mentre scende a 0,16 quando T8 è presente; un pattern speculare si riscontra per T8, suggerendo una parziale competizione fra i due frame all'interno del discorso ostile. Analogamente (tab. 3b), la probabilità di utilizzare T7 (*Conflitti urbani*) in presenza di stereotipi ma in assenza di hate speech è pari a 0,64, mentre diminuisce drasticamente (0,05) quando è attivo l'odio, evidenziando come gli stereotipi rafforzino la tematizzazione dei conflitti urbani in contesti non apertamente ostili. Un ulteriore esempio (tab. 3c) riguarda la relazione tra aggressività e T8: in presenza di enunciati aggressivi e stereotipici la probabilità di ricorrere al frame della criminalizzazione etnica raggiunge 0,42, mentre è più contenuta (0,16) quando la stereotipia è assente. Questo gradiente probabilistico indica che aggressività e stereotipia agiscono come fattori congiunti che facilitano l'attivazione di temi fortemente stigmatizzanti.

È importante sottolineare che tali interrogazioni non esauriscono le potenzialità del modello: la struttura della rete bayesiana consente infatti al

ISBN Open Access: 9788835191971

ricercatore di formulare domande inferenziali personalizzate, variando a piacimento le condizioni osservate e analizzando la probabilità di attivazione di qualunque tema o categoria discorsiva di interesse. A seconda degli obiettivi di ricerca, è possibile esplorare scenari controfattuali, valutare l'effetto marginale di singole variabili, o identificare configurazioni tematiche indicative di specifiche dinamiche discorsive. In questo modo, la rete bayesiana diventa uno strumento versatile che permette sia di testare ipotesi teoriche sia di scoprire pattern emergenti all'interno di corpora caratterizzati da elevata complessità semantica e stilistica.

6. Discussione e riflessioni conclusive

La ricerca qui presentata ha l'obiettivo di proporre un *framework* per analizzare e modellare in senso probabilistico le strutture relazionali dei fattori che possono influenzare il discorso d'odio *online*, nonché il linguaggio ostile verso migranti e comunità rom, nel contesto italiano. L'approccio metodologico adottato integra l'analisi di rete e la modellizzazione tramite reti bayesiane, permettendo di superare i limiti degli studi puramente descrittivi o classificatori e offrendo un quadro analitico che non si limita a catalogare i temi o rilevare la presenza di *hate speech*, bensì ne esplora le interdipendenze e le dinamiche di coattivazione. I risultati confermano e approfondiscono quanto emerso nella letteratura di riferimento. Il discorso razzista e xenofobo si configura come un costrutto semioticamente denso e tematicamente articolato, organizzato intorno a cluster narrativi distinti ma interconnessi. La mappatura delle tematiche ha rivelato una costellazione di *frame* ricorrenti, dalla *Critica xenofoba al welfare* (T2) alla *Criminalizzazione etnica e la rabbia populista* (T8); dal *Conflitto religioso e islamofobia* (T6) ai *Conflitti urbani* (T7), riflettendo narrazioni sociali profonde e ampiamente documentate nel discorso pubblico. Il valore di questo studio risiede nella capacità di mostrare come questi *frame* non operino in modo isolato ma interagiscano piuttosto in configurazioni specifiche tali da produrre, facilitare o intensificare le espressioni di ostilità, aggressività e stereotipia.

La rete bayesiana ha funzionato come strumento utile a spiegare le interazioni, in senso probabilistico, di tipo diretto e indiretto. I risultati inferenziali, come quelli illustrati nelle tab. 1 e 2, mostrano in modo quantitativo come la probabilità di osservare contenuti d'odio o di aggressività verbale cambi in modo significativo a seconda delle combinazioni tematiche attivate. Ad esempio, il ruolo del linguaggio volgare (T10) come catalizzatore stilistico o l'effetto amplificatore della tematizzazione dei conflitti urbani (T7) forniscono una conferma empirica della natura performativa ed

ISBN Open Access: 9788835191971

emotivamente satura del discorso ostile. La capacità del modello di suggerire possibili effetti marginali, quali, ad esempio, il contributo specifico della critica al welfare (T2) anche in assenza di altri temi fortemente discriminatori, aiuta a decostruire il fenomeno identificando quali nuclei narrativi siano più strutturalmente legati alla produzione di odio e quali agiscano invece come amplificatori in contesti già ostili.

Questa ricerca dialoga con il dibattito sulla natura contestuale e adattiva dell'*hate speech online*: i risultati ottenuti supportano l'ipotesi che l'odio non sia un attributo fisso di certi lessemi ma emerga dall'interazione situata tra contenuti tematici, strategie retoriche e dimensioni pragmatiche. La parziale "competizione" tra alcuni fattori (ad esempio, i temi T6 e T8 nella tab. 3a) suggerisce, inoltre, che all'interno dell'ecosistema discorsivo ostile, esistano percorsi narrativi diversi che possono essere selettivamente attivati a seconda del contesto o del target. Si tratta di un aspetto coerente con la teoria delle rappresentazioni sociali (Jovchelovitch, 2007) e la loro riattivazione strategica (Gamson e Modigliani, 1989; Abric, 2001).

Sul piano metodologico, il lavoro contribuisce al crescente campo della *computational social science* applicata all'analisi del discorso, proponendo un *framework* rigoroso e replicabile. L'uso di reti bayesiane, in particolare, risponde all'esigenza di modelli interpretabili che, a differenza di molti approcci "black-box" tipici del *machine learning*, consentono di formulare e testare ipotesi relazionali esplicite. La procedura integrata – dalla costruzione della rete semantica alla *community detection*, fino all'apprendimento e validazione del modello bayesiano – fornisce una struttura analitica solida per trasformare dati testuali complessi in una rappresentazione probabilistica formale.

Nonostante questo tipo di avanzamento, lo studio presenta limiti che suggeriscono ipotesi di ricerca future: in primo luogo, il modello attuale cattura relazioni di dipendenza probabilistica la cui interpretazione causale richiede ulteriori assunzioni teoriche e validazioni. Includere variabili temporali o di rete (ad esempio la sequenza dei testi o le relazioni tra utenti) consentirebbe di distinguere meglio tra mere correlazioni e dinamiche di influenza. In secondo luogo, l'analisi si è concentrata su covariate discorsive interne al testo (odio, stereotipia, aggressività).

L'integrazione di metadati esterni come il profilo degli utenti, l'affiliazione politica, o eventi macro-sociali contemporanei alla raccolta dei dati, arricchirebbe il modello permettendo di contestualizzare le configurazioni discorsive all'interno di specifici eventi sociali, politici, mediatici, momenti storici (attentati, elezioni, crisi sanitarie, conflitti), mostrando quanto essi attivino picchi di ostilità, *hate speech* e aggressività discorsiva, soprattutto *online*, come suggerito dagli studi sull'ostilità guidata da eventi (Bliuc *et al.* 2018). Infine, la generalizzabilità dei risultati è legata al *corpus* specifico

ISBN Open Access: 9788835191971

(Twitter italiano), mentre applicazioni dello stesso framework ad altre piattaforme (forum o spazi di messaggistica) o ad altri contesti linguistico-culturali sarebbero preziose per verificare la stabilità dei pattern osservati e la trasferibilità del metodo.

In sintesi, questa ricerca dimostra che l'analisi computazionale del discorso d'odio può andare oltre la mera descrizione, muovendosi verso una comprensione più profonda dei suoi meccanismi generativi interni. La modellizzazione tramite reti bayesiane delle relazioni tra temi e stili offre uno strumento utile a mappare l'architettura probabilistica del razzismo digitale, rendendo espliciti i percorsi discorsivi che portano più frequentemente a espressioni di ostilità. Oltre al valore scientifico, questo approccio ha implicazioni operative promettenti: la capacità di identificare le combinazioni tematiche e stilistiche più "rischiose" potrebbe informare lo sviluppo di sistemi di rilevazione più sofisticati e contestualmente sensibili, oltre a fornire indicazioni ai vari media per ideare campagne di contrasto e di educazione mirate e finalizzate a smontare non singole parole quanto piuttosto strutture narrative e logiche retoriche che sostengono e alimentano l'odio online.

Riferimenti bibliografici

Abric, J. C. (2001), A structural approach to social representations. In Deaux K. and Philogène G., eds., *Representations of the social: Bridging theoretical traditions* (pp. 42–47), Blackwell, New Jersey.

Allan, K. and Burridge, K. (2007), *Forbidden Words. Taboo and the Censoring of Language*, Cambridge University Press, Cambridge.

Bliuc, A. M., Faulkner, N., Jakubowicz, A. and McGarty, C. (2018), Online networks of racial hate: A systematic review of 10 years of research on cyber-racism, *Computers in Human Behavior*, 87: 75–86.

Bolasco, S. (2013), *L'analisi automatica dei testi. Fare ricerca con il text mining*, Carocci, Roma.

Bosco, C., Dell'Orletta, F., Poletto, F., Sanguinetti, M. and Tesconi, M. (2018), Overview of the EVALITA 2018 Hate Speech Detection Task. In Caselli T., Novielli N., Patti V. and Rosso P., eds., *EVALITA Evaluation of NLP and Speech Tools for Italian*, Accademia University Press, 67–74.

Bosco, C., Patti, V., Frenda, S., Cignarella, A.T., Paciello, M. and D'Errico, F. (2023), Detecting racial stereotypes: An Italian social media corpus where psychology meets NLP. *Information Processing & Management*, 60 (1): 103118.

Capozzi, A.T.E., Lai, M., Basile, V., Musto, C., Polignano, M., Poletto, F., Sanguinetti, M., Bosco, C., Patti, V., Ruffo, G., Semeraro, G. and Stranisci, M. (2019), Computational Linguistics Against Hate: Hate Speech Detection and Visualization on Social Media in the "Contro L'Odio" Project. In *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)*,

ISBN Open Access: 9788835191971

CEUR Workshop Proceedings, 442–447.
Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. and Starnini, M. (2021), The echo chamber effect on social media. *PNAS*, 118 (9), e2023301118.
Chung, Y.-L., Kuzmenko, E., Tekiroglu, S.S. and Guerini, M. (2019), CONAN – COunter NArratives through Nichesourcing: a Multilingual Dataset of Responses to Fight Online Hate Speech. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, ACL, 2819–2829.
Davidson, T., Warmsley, D., Macy, M. and Weber, I. (2017), Automated Hate Speech Detection and the Problem of Offensive Language. In *Proceedings of the International AAAI Conference on Web and Social Media*, AAAI, 512–515.
Ekman, M. (2019), Anti-immigration and racist discourse in social media. *European Journal of Communication*, 34 (6), 606–618.
Filatova, E. (2012), Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, ELRA, 392–398.
Fortunato, S. (2010), Community detection in graphs. *Physics Reports*, 486 (3–5), 75–174.
Gamson, W. A. and Modigliani, A. (1989), Media discourse and public opinion on nuclear power. *American Journal of Sociology*, 95 (1), 1–37.
Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2014), *Bayesian Data Analysis*. CRC Press
Heckerman, D. (1997), Bayesian Networks for Data Mining. *Data Mining and Knowledge Discovery*, 1: 79–119.
Humprecht, E., Hellmueller, L. and Lischka, J.A. (2020), Hostile Emotions in News-Comments: A Cross-National Analysis of Facebook Discussions. *Social Media + Society*, 6 (1), 2056305120912481.
Joffe, H. (2003), Risk: From Perception to Social Representation. *British Journal of Social Psychology*, 42 (1): 55–73.
Jovchelovitch, S. (2007), *Knowledge in context: Representations, community and culture*, Routledge, London.
Koller, D. and Friedman, N. (2009), *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, Cambridge (MA).
Korb, K.B. and Nicholson, A. E. (2023), *Bayesian Artificial Intelligence*, CRC Press, Florida.
Manning, C.D. and Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge (MA).
Misuraca, M., Scepi, G. and Spano, M. (2023), Network-based dimensionality reduction for textual datasets. In Brentari E., Chiodi M. and Wit E.-J.C., eds., *Models for Data Analysis* (pp. 175–190), Springer, Berlino.
Misuraca, M. and Spano, M. (2020), Unsupervised analytic strategies to explore large document collections. In Iezzi D.F., Mayaffre D. and Misuraca M., eds., *Text Analytics. Advances and Challenges* (pp. 17–28), Springer, Berlino.
Moscovici, S. (1961), *La psychanalyse, son image et son public*, Presses Universitaires de France, Paris.

ISBN Open Access: 9788835191971

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inferenc*e. Morgan Kaufmann, USA.

Pons, P. and Latapy, M. (2005), Computing communities in large networks using random walks. In Yolum P., Güngör T., Gürgen F. and Özturan C., *Computer and Information Sciences – ISCIS 2005* (pp. 284–293), Springer, Berlino.

Popping, R. (2003), Knowledge graphs and network text analysis. *Social Science Information*, 42 (1): 91–106

Sanguinetti, M., Comandini, G., di Nuovo, E., Frenda, S., Stranisci, M., Bosco, C., Caselli, T., Patti, V. and Russo, I. (2020), HaSpeeDe 2 @ EVALITA2020: Overview of the EVALITA 2020 Hate Speech Detection Task. In Basile V., Croce D., Di Maro M. and Passaro L.C., eds., *EVALITA Evaluation of NLP and Speech Tools for Italian* (pp. 93–101), Accademia University Press, Torino.

Scutari, M. and Denis, J.B. (2021), *Bayesian Networks*, CRC Press, USA.

Sunstein, C.R. (2017), *#Republic: Divided Democracy in the Age of Social Media,* Princeton University Press, USA.

Tsamardinos, I., Brown, L.E. and Aliferis, C.F. (2006), The max-min hill-climbing Bayesian network structure learning algorithm, *Machine Learning*, 65: 31–78.

Udanor, C. and Anyanwu, C.C. (2019), Combating the challenges of social media hate speech in a polarized society: A Twitter ego lexalytics approach. *Data Technologies and Applications*, 53 (4): 501–527.

ISBN Open Access: 9788835191971

Narrazioni d'odio e memoria collettiva: una prospettiva metodologica per lo studio dell'antisemitismo sui social media

di *Luca De Benedictis*[*], *Giuseppe Giordano*[**], *Maria Prosperina Vitale*[**]

1. Premessa

La diffusione dei *social media*, come uso integrato nella vita quotidiana, da oltre dieci anni, ha profondamente trasformato le modalità con cui si costruisce e si articola il discorso pubblico (Fuchs, 2021). Le piattaforme digitali sono diventate, infatti, spazi centrali di partecipazione politica, informazione e memoria collettiva, ma anche ambienti privilegiati in cui si amplificano fenomeni di polarizzazione e di *hate speech* (Siegel, 2020; Vandebosch e Rothmund, 2024). In questo contesto, l'odio online nei confronti di gruppi minoritari assume caratteristiche specifiche, legate sia alle opportunità offerte dall'infrastruttura tecnologica sia alla rapidità con cui eventi esterni vengono rielaborati simbolicamente nel dibattito pubblico. L'antisemitismo costituisce un caso emblematico di tali dinamiche. Definito come "discriminazione", "ostilità" o "violenza" nei confronti degli ebrei[1], mostra una forte capacità di rielaborazione rispetto ai contesti storici e comunicativi. Studi recenti evidenziano come l'antisemitismo online non si limiti alla riproduzio-

[*] Dipartimento di Economia, Statistica e Impresa, Universitas Mercatorum, luca.debenedictis@unimercatorum.it

[**] Dipartimento di Studi Politici e Sociali, Università di Salerno, ggiordano@unisa.it; mvitale@unisa.it

[1] The Jerusalem Declaration on Antisemitism (2020): https://jerusalemdeclaration.org. Per altre definizioni, simili ma differenti, si vedano quella della International Holocaust Remembrance Alliance (IHRA): https://holocaustremembrance.com/resources/la-definizione-di-antisemitismo-dellalleanza-internazionale-per-la-memoria-dellolocausto. La definizione di antisemitismo dell'IHRA è oggetto di dibattito accademico ma anche politico, in quanto indica alcuni esempi concreti di "antisemitismo contemporaneo" che includono il diritto all'esistenza dello Stato di Israele: ad esempio, è definito antisemita fare paragoni tra la politica israeliana contemporanea e quella della Germania nazista.

ISBN Open Access: 9788835191971

ne di stereotipi storici, ma assuma forme nuove che intrecciano la memoria della Shoah, l'attualità geopolitica e i conflitti identitari (Bjola e Manor, 2020; Weimann, 2024). Questo rende necessario un cambio di paradigma negli studi sull'antisemitismo, e in quelli sull'antisemitismo *online* in particolare. Si ritiene che, utilizzando in modo sistematico un'analisi interdisciplinare di tale fenomeno nei social media, introducendo metodologie di ricerca miste che coniughino l'analisi qualitativa con quella quantitativa, fornendo dati e illustrando scenari utili a studiosi, decisori politici, piattaforme digitali e attori della società civile, si creino le condizioni per lo sviluppo di efficaci strategie di contrasto al fenomeno oggetto d'indagine (Hübscher e Von Mering, 2022). Inoltre, la metodologia di analisi utilizzata può essere applicata allo studio delle manifestazioni di odio online rispetto a categorie specifiche della popolazione, a fenomeni di razzismo, omofobia o misoginia e a contesti analoghi a quelli di questo studio, quali l'islamofobia, l'avversione per gli stranieri, soprattutto immigrati, e le minoranze etnico-religiose.

Il presente capitolo analizza le narrazioni online sull'antisemitismo nel contesto italiano attraverso un approccio metodologico che integra la prospettiva metodologica della *Social Network Analysis* (SNA, Wasserman, Faust; 1994) con *Natural Language Processing* (NLP – Mikolov *et al.*, 2013), applicato a due ambiti strettamente connessi ma analiticamente distinti: il dibattito online sviluppatosi sulla piattaforma 𝕏 (ex *Twitter*) in occasione del Giorno della Memoria, negli anni tra il 2022 e il 2025 inclusi, e l'evoluzione delle narrazioni successive all'inizio del conflitto Israele–Palestina a partire dal 7 ottobre 2023. L'obiettivo è comprendere come eventi commemorativi e conflitti contemporanei contribuiscano a ridefinire le strutture relazionali e domini semantici dell'odio online in tema di antisemitismo.

A seguito di una rassegna della letteratura di riferimento e della presentazione dei dati online analizzati secondo una metodologia mista, il capitolo sintetizza i principali risultati della ricerca. Da un lato, sono ricostruite le reti di interazione tra utenti, analizzate tramite misure di centralità (Freeman, 1978) e algoritmi di *community detection* (Clauset *et al.*, 2004), al fine di individuare attori influenti, comunità e potenziali *hotspot* comunicativi in cui circolano contenuti, narrazioni o discorsi specifici. Dall'altro lato, l'analisi semantica dei contenuti testuali è stata condotta mediante tecniche di *Text Mining* (Welbers *et al.*, 2017) e *Structural Topic Models* (STM – Roberts *et al.*, 2016), che hanno consentito di identificare i temi latenti del dibattito e di analizzarne i cambiamenti nel tempo.

ISBN Open Access: 9788835191971

2. Hate speech e antisemitismo online

Negli ultimi anni, la crescente diffusione di discorsi d'odio sui *social media* ha suscitato notevoli preoccupazioni per il loro impatto su individui e società, un fenomeno complesso e in aumento a seguito della proliferazione delle piattaforme digitali (Alkomah e Ma, 2022). L'*hate speech* è comunemente inteso come un insieme di pratiche discorsive che mirano a denigrare, intimidire o incitare alla violenza contro individui o gruppi vulnerabili e minoritari, sulla base di caratteristiche identitarie quali razza, religione, origine etnica, orientamento sessuale, disabilità o genere (Siegel, 2020; de la Fuente *et al.*, 2023; Vandebosch e Rothmund, 2024).

Tra i gruppi minoritari, gli ebrei sono stati storicamente bersaglio di attacchi fisici e discorsi d'odio (Langmuir, 1990). Recenti studi (Bjola e Manor, 2020; Hübscher e Von Mering, 2022; Weimann, 2024) e rapporti sulla situazione in Italia (CDEC, 2024, 2025) evidenziano un aumento significativo dei discorsi d'odio online antisemiti[2]. Dopo il 7 ottobre 2023, il conflitto tra Hamas e Israele ha favorito l'emergere di nuove forme di antisemitismo online, caratterizzate anche dalla distorsione delle responsabilità storiche della Shoah, attenuandone la tragicità attraverso il revisionismo storico, negando il valore sociale della memoria e mescolando eventi contemporanei al passato, generando un clima di ostilità e minaccia generalizzata verso gli ebrei. L'antisemitismo online che questo studio intende evidenziare considera tale duplice dimensione temporale. Da un lato, esso richiama elementi storici consolidati, legati alla Shoah e alla persecuzione degli ebrei; dall'altro, incorpora anche, in questo contesto specifico, narrazioni contemporanee, in particolare quelle connesse allo Stato di Israele e al conflitto medio-orientale (Weimann, 2024). L'estensione simbolica delle azioni e delle responsabilità del governo israeliano e dei suoi leader politici agli ebrei rappresenta uno dei principali meccanismi attraverso cui l'antisemitismo viene riformulato in chiave apparentemente politica, di sostegno e di impegno umanitario a favore della popolazione palestinese (Bjola e Manor, 2020).

L'ambiente digitale favorisce la diffusione di discorsi grazie a meccanismi di echo chambers, bassa soglia di partecipazione e interazioni reticolari (Himelboim *et al.*, 2017); al contempo la necessità di analizzare dinamiche

[2] Operativamente, nelle sue indagini e nei suoi rapporti periodici, il Centro di Documentazione Ebraica Contemporanea (CDEC) utilizza la definizione di antisemitismo dell'HIRA, che considera antisemita anche alcune forme specifiche di critica allo Stato di Israele. Si rimanda a CDEC (2024, 2025) per la descrizione puntuale delle azioni antisemite, incluse anche alcune forme di boicottaggio nei confronti dello Stato di Israele, di imprese di proprietà di imprenditori israeliani o di istituzioni israeliane, quali ad esempio le Università, nonché di individui ebrei non necessariamente israeliani o di soggetti definiti "sionisti".

ISBN Open Access: 9788835191971

discorsive ed evidenziare la presenza di *hate speech* online richiede l'utilizzo di approcci metodologici misti, che combinano la SNA con studi qualitativi (Fonseca *et al.*, 2024; Pontes *et al.*, 2024), di NLP[3] e *Text Mining* (Rawat *et al.*, 2024; Wong, 2024) al fine di individuare la presenza di attori influenti e di gruppi all'interno delle strutture reticolari, analizzare le dinamiche di polarizzazione nei discorsi ed esplorare l'evoluzione dei temi e dei frame discorsivi. L'integrazione di questi approcci risulta, a nostro avviso, particolarmente adatta allo studio dell'antisemitismo online.

3. Caso studio e obiettivi della ricerca

La necessità di individuare una finestra temporale per l'analisi comparativa dei fenomeni di antisemitismo online ha portato alla scelta di una specifica ricorrenza pubblica come momento focale, in quanto legata alla memoria della persecuzione degli ebrei durante il fascismo. Il Giorno della Memoria, istituito in Italia con la Legge n. 211 del 2000, celebrato il 27 gennaio, ha l'obiettivo di preservare la memoria della Shoah e delle persecuzioni razziali, promuovendo una riflessione collettiva sui rischi dell'odio, dell'intolleranza e della discriminazione. Oltre alla sua dimensione istituzionale ed educativa, questa ricorrenza genera un'intensa attività comunicativa online, che coinvolge istituzioni, media, attori politici e cittadini. Le commemorazioni pubbliche rappresentano, infatti, momenti di elevata densità discorsiva, in cui coesistono narrazioni ampiamente condivise e conflitti simbolici, rendendo visibili tensioni latenti nel dibattito pubblico (Siegel, 2020). In tale contesto, il periodo temporale intorno al Giorno della Memoria può diventare anche uno spazio in cui emergono forme di antisemitismo implicito o latente, spesso mascherate da polemiche politiche, riletture strumentali dell'attualità o revisionismi storici.

Questo contesto si è ulteriormente trasformato a seguito dell'attacco terroristico di Hamas del 7 ottobre 2023 e della successiva risposta militare israeliana a Gaza, eventi che hanno inciso in modo significativo sul dibattito pubblico globale. In tale quadro, nel discorso pubblico e sui *social media* si è affermata una narrazione che tende a sovrapporre il governo israeliano agli ebrei *in quanto tali*, estendendo la critica alle azioni politiche e militari di Israele fino a generare forme di ostilità generalizzata nei confronti degli ebrei. Questa dinamica tende così a sovrapporre in uno spazio astorico il

[3] Tali studi mostrano un crescente ricorso a tecniche di NLP per l'individuazione automatica dell'hate speech, ma evidenziano anche i limiti degli approcci puramente testuali, soprattutto nel riconoscimento di forme implicite o contestuali di odio.

ISBN Open Access: 9788835191971

presente (Israele e la diaspora ebraica) al passato (la Shoah), generando narrazioni che alterano le responsabilità storiche, attenuano la portata tragica dell'Olocausto e mettono in discussione il valore sociale della memoria, in contrasto con il principio fondativo del "non deve accadere mai più" sancito a livello nazionale e internazionale proprio con l'istituzione della Giornata della Memoria in Italia, in particolare, e internazionalmente, sotto l'egida delle Nazioni Unite[4].

Nel contesto italiano, tali dinamiche hanno avuto un impatto diretto anche sulle narrazioni legate al Giorno della Memoria. L'analisi delle tendenze di ricerca online, effettuata tramite Google Trends (Fig. 1), mostra, infatti, una discontinuità nella ciclicità dell'attenzione pubblica intorno al 27 gennaio, con un abbassamento dei picchi di ricerca online successivi al 2023. Nello specifico, l'analisi riguarda il volume relativo delle ricerche effettuate nel periodo di sette giorni a cavallo del Giorno della Memoria (da tre giorni prima a tre giorni dopo) tra il 2020 e il 2025. Confrontando i periodi precedenti al 7 ottobre 2023 con quelli successivi, l'obiettivo è verificare e valutare come l'attenzione pubblica verso temi legati alla memoria della Shoah si sia modificata in relazione a ciò che abbiamo ipotizzato possa costituire un evento di rottura nel discorso pubblico online. In particolare, nella Figura 1 sono riportati i trend di ricerca relativi ai termini "Giorno della Memoria" e "Ebrei", osservandone l'andamento temporale e le variazioni di intensità. L'analisi evidenzia una periodicità marcata dell'attenzione pubblica verso il Giorno della Memoria, caratterizzata da picchi ricorrenti in corrispondenza del 27 gennaio, confermando il ruolo della commemorazione come momento privilegiato di attivazione dell'interesse collettivo per tale ricorrenza. Tuttavia, il confronto tra le due serie temporali mostra, prima del 2023, una non perfetta sovrapposizione della serie relativa al termine "Ebrei" (linea arancione) con quella della "Giornata della Memoria" (linea blu): questa parziale sovrapposizione è giustificabile sia con il fatto che la Giornata della Memoria commemora l'Olocausto ebraico in Europa, la Shoah per l'appunto, ma non solo, ricordando tutte le vittime della persecuzione nazi-fascista, dagli omosessuali, alla popolazione rom, ai prigionieri politici e militari. Inoltre, la sovrapposizione parziale può essere dovuta alla valenza della Giornata della

[4] L'art. 2 della Legge del 20 luglio 2000, n. 211 – Istituzione del «Giorno della Memoria» in ricordo dello sterminio e delle persecuzioni del popolo ebraico e dei deportati militari e politici italiani nei campi nazisti – recita: "In occasione del «Giorno della Memoria» di cui all'articolo 1, sono organizzati cerimonie, iniziative, incontri e momenti comuni di narrazione dei fatti e di riflessione, in modo particolare nelle scuole di ogni ordine e grado, su quanto è accaduto al popolo ebraico e ai deportati militari e politici italiani nei campi nazisti in modo da conservare nel futuro dell'Italia la memoria di un tragico ed oscuro periodo della storia nel nostro Paese e in Europa, e affinché simili eventi non possano mai più accadere".

ISBN Open Access: 9788835191971

Memoria come momento condiviso per favorire il ricordo da parte dell'intera cittadinanza, e non solo dei cittadini ebrei. Il ricordo della Shoah ha quindi una funzione di monito per tutti e tutte e, – come ricordava Primo Levi – se è avvenuto, può ancora succedere agli ebrei, ai deportati nei campi di concentramento, ma potenzialmente a chiunque sia visto come diverso e sacrificabile da un regime totalitario.

La serie relativa al termine "Ebrei" ha un'improvvisa impennata in prossimità del 7 ottobre 2023, evidenziando un aumento significativo delle ricerche online sugli ebrei, in coincidenza con l'esplosione del conflitto israelo-palestinese. L'analisi comparativa dei livelli medi dei picchi prima e dopo il 7 ottobre 2023 indica, da una parte, una pressocché totale sovrapposizione delle due serie, evidenziando un'identità tra "Giornata della Memoria" e "Ebrei" che esclude dall'immaginario pubblico ogni altra vittima della persecuzione nazi-fascista; dall'altra parte, l'intensità media delle oscillazioni di ricerca risulta meno elevata nel secondo periodo, dopo il 2023, suggerendo una ridefinizione delle modalità di partecipazione simbolica alla commemorazione della Giornata della Memoria. Questa dinamica può essere interpretata come uno slittamento semantico, in cui la memoria della Shoah viene progressivamente riletta alla luce del conflitto contemporaneo in Medio Oriente. Tale processo potrebbe, in principio, favorire l'emergere di forme di nuovo antisemitismo, nelle quali l'ostilità verso Israele, l'antisionismo e i discorsi di odio contro gli Ebrei tendono a sovrapporsi e a confondersi, producendo nuove configurazioni discorsive che caratterizzano il rapporto tra memoria storica, attualità politica e identità collettive nello spazio digitale.

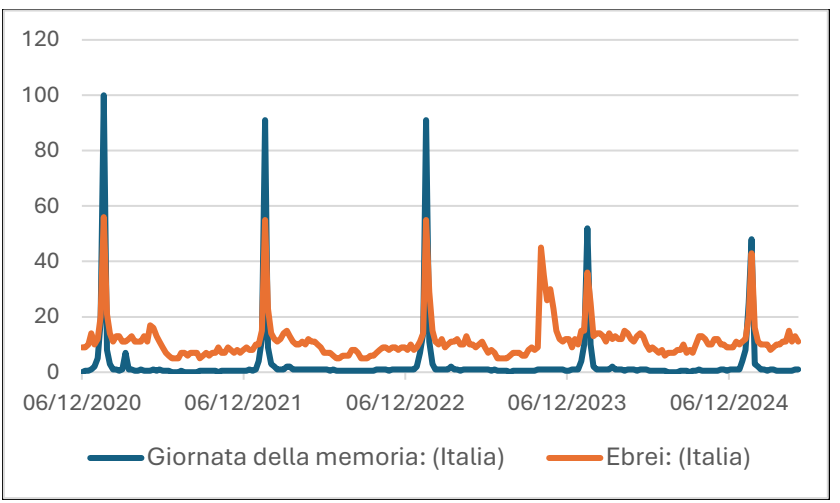


Fig. 1 – Google Trends sui temi "Giornata della Memoria" e "Ebrei" negli ultimi 5 anni in Italia [rilevazione aggiornata al 06-12-2025].

ISBN Open Access: 9788835191971

Alla luce di tali elementi, l'obiettivo principale dello studio è comprendere le modalità con cui l'antisemitismo circola nelle reti online, integrando l'analisi delle strutture di interazione con quella dei contenuti semantici. Dal punto di vista relazionale, vengono esaminate diverse forme di interazione tra gli utenti della piattaforma 𝕏, quali ricondivisioni, menzioni, citazioni e risposte, al fine di ricostruirne la rete sociale sottostante e identificare utenti influenti e potenziali *hotspot* di diffusione del discorso d'odio. Gli attori e le aree di elevata attività vengono individuati attraverso opportuni indici di centralità della rete (Freeman, 1978), che consentono di cogliere diverse dimensioni dell'influenza, della visibilità e della mediazione informativa.

Un ulteriore obiettivo consiste nell'individuare gruppi di utenti caratterizzati da pattern di interazione e da comportamenti discorsivi simili. A tal fine, si applicano algoritmi di *community detection* per identificare gruppi coesi e cluster comunicativi, secondo l'approccio proposto da Clauset *et al.* (2004). Tali comunità rappresentano unità semantiche e strutturali rilevanti attraverso le quali l'ostilità nei confronti degli ebrei può essere amplificata e rielaborata. La considerazione è che i *social media* non sono semplicemente archivi di testi né mere reti di relazioni tra utenti: sono sistemi complessi in cui interazioni sociali e produzione discorsiva si co-determinano. Ogni contenuto testuale nasce all'interno di una rete di relazioni, e ogni relazione è mediata – rafforzata o indebolita – da scambi linguistici. Per cogliere questa complessità è necessario adottare una prospettiva metodologica integrata.

4. Approccio metodologico integrato

L'integrazione tra SNA e Text Mining consente di affrontare simultaneamente due dimensioni fondamentali: la struttura sociale, ovvero come gli utenti si aggregano, interagiscono e formano community; la dimensione semantica, ovvero quali temi, narrazioni e concetti latenti emergono dai testi prodotti. Tradotto in termini dei nostri obiettivi, tale approccio metodologico mira a offrire una comprensione più approfondita delle dinamiche del dibattito online sulla Giornata della Memoria. L'interesse è rivolto ai contenuti generati dagli utenti e alle interazioni che si sviluppano sulle piattaforme digitali, dove le narrazioni di odio emergono, si diffondono e si trasformano attraverso strutture relazionali complesse. I dati raccolti dai *social media* possono essere utilizzati per costruire diverse tipologie di reti: reti di utenti, basate sulle interazioni tra account, e reti di parole, derivate dai contenuti testuali dei messaggi, attraverso la mappatura dei temi emersi dall'analisi dei *tweet*. Questa doppia prospettiva consente, in generale, di analizzare congiuntamente *chi interagisce con chi* e *quali contenuti circolano*, offrendo una

ISBN Open Access: 9788835191971

visione integrata dei processi di diffusione e di strutturazione dei discorsi sulla commemorazione delle vittime del nazifascismo, inclusi quelli che potrebbero rappresentare una nuova forma di antisemitismo.

Dal punto di vista dell'analisi dei contenuti, il contributo mira a caratterizzare i temi dominanti e le narrazioni ricorrenti del discorso d'odio online. Dopo una fase di normalizzazione e *pre-processing* dei testi, vengono stimati gli STM per estrarre dai diversi scambi testuali tra utenti i temi principali, analizzarne la prevalenza e studiarne le correlazioni nel tempo e all'interno delle diverse comunità di utenti (Roberts *et al.,* 2019). Questo consente di cogliere come specifici frame discorsivi coesistano, si sovrappongano o divergano tra i diversi segmenti della rete. L'estrazione dei temi dominanti restituisce dunque una rappresentazione della dimensione cognitiva e simbolica del dibattito.

Combinando indicatori strutturali della rete e le rappresentazioni tematiche del discorso, l'approccio proposto offre una prospettiva analitica integrata per identificare non solo chi guida la diffusione del discorso d'odio online, ma anche quali narrazioni vengono veicolate e come esse siano incorporate nelle strutture della comunicazione digitale. Questo approccio integrato ha una forte vocazione esplorativa. Non mira a verificare ipotesi rigide, ma a far emergere pattern, configurazioni e regolarità che possono poi essere interpretate alla luce di quadri teorici, sociologici, comunicativi o cognitivi. L'obiettivo non è, infatti, ridurre la complessità del discorso sociale, ma renderla interpretabile, offrendo strumenti per leggere i social media come spazi in cui strutture relazionali e di senso si intrecciano dinamicamente.

5. I principali risultati

Come anticipato, i dati oggetto di approfondimento sono stati raccolti dalla piattaforma durante la settimana intorno al 27 gennaio, per cinque anni consecutivi dal 2021 al 2025, mediante il software *NodeXL* (Smith *et al.*, 2010). La raccolta è stata effettuata tramite una query limitata alla lingua italiana e basata su parole chiave tematiche ["Giornata della Memoria" OR Olocausto OR Shoah OR Nazismo OR Fascismo OR Ebrei OR Antisemitismo OR Sionismo OR Antisionismo OR 27gennaio OR lilianasegre OR primolevi]. I dati estratti includono contenuti testuali, caratteristiche degli utenti e interazioni (*retweet*, *mention*, *reply to*).

Dopo una fase di pulizia e *preprocessing* testuale (normalizzazione, rimo-

ISBN Open Access: 9788835191971

zione delle *stop-word*, TF-IDF)[5], sono state costruite reti di utenti basate sulle interazioni online. L'analisi di rete ha incluso indicatori strutturali (densità, reciprocità, componenti) e misure di centralità, nonché l'applicazione di algoritmi di *community detection*. Parallelamente, l'analisi semantica è stata condotta tramite modelli STM, che ha consentito di individuare temi latenti e analizzarne le sovrapposizioni, integrando l'analisi lungo la dimensione temporale.

5.1. Utenti influenti e gruppi tematici nella rete

I risultati dell'analisi delle reti definite a partire dai legami tra gli utenti della piattaforma mostrano cambiamenti nel dibattito online nel periodo post-2023. La Tabella 1 presenta un confronto delle principali misure strutturali di rete nelle 4 settimane considerate, dal 2022 al 2025, prima e dopo il 7 ottobre 2023, evidenziando cambiamenti rilevanti nella configurazione del dibattito online. In primo luogo, si osserva una riduzione del numero di vertici nel 2024, seguita da una nuova crescita nel 2025, che suggerisce una temporanea contrazione del numero di utenti coinvolti, probabilmente associata a una riorganizzazione del discorso pubblico dopo l'avvio del conflitto. Parallelamente, il numero di archi mostra un incremento complessivo nel tempo, indicando una maggiore intensità delle interazioni e una crescita delle discussioni, anche in presenza di una base di utenti inizialmente più ridotta. Questo andamento suggerisce una maggiore attività comunicativa e una circolazione più frequente dei contenuti all'interno della rete. Il livello di reciprocità dei legami rimane sostanzialmente stabile tra i due periodi, segnalando una continuità nelle modalità di interazione tra gli utenti, nonostante le trasformazioni strutturali complessive della rete. Al contrario, le misure di

[5] Le analisi di Text Mining sono precedute da una serie di operazioni di preparazione delle strutture di dati finali. In particolare, si parla di Tokenizzazione come quel processo di scomposizione di un testo in frasi e parole chiamate "token". Questi vengono utilizzati nei modelli, come bag-of-words, per il raggruppamento di testi e per le attività di abbinamento dei documenti, e di Stemming**,** come processo di separazione dei prefissi e dei suffissi dalle parole per ricavare la forma e il significato della parola radice. Tale tecnica migliora il recupero delle informazioni ed aiuta la disambiguazione del significato dei lemmi. Altre operazioni caratteristiche sono l'eliminazione delle forme grafiche prive di significato proprio (stop words) e la trasformazione delle maiuscole in minuscole per omogeneizzare le forme grafiche (e ottenere distribuzioni di frequenze più consistenti). Le distribuzioni di frequenza vengono solitamente relativizzate rispetto a diversi criteri, tra questi, il principio del TF-IDF consente di tener conto della rilevanza del termine (token) non solo in base alla sua frequenza nel documento (TF) ma anche in misura inversa alla specificità di appartenenza al documento (IDF). L'idea alla base è di dare maggiore importanza ai termini che compaiono nel documento, pur essendo in generale poco frequenti.

ISBN Open Access: 9788835191971

connettività evidenziano un rafforzamento della struttura reticolare nel periodo 2024-2025, con un aumento del numero massimo di vertici in una componente connessa e una riduzione della distanza geodetica media, indicativa di una maggiore raggiungibilità tra gli attori. La densità della rete mostra oscillazioni tra i due periodi analizzati, riflettendo una dinamica di espansione e contrazione dei legami che accompagna la riorganizzazione della discussione online. Allo stesso tempo, si osserva (tab. 1) una crescente concentrazione delle interazioni in cluster specifici, accompagnata da un aumento del numero di voci isolate, che suggeriscono una polarizzazione del discorso e una segmentazione più marcata dello spazio comunicativo. Nel complesso, il confronto evidenzia una rete che, pur diventando più connessa e interattiva, appare anche più frammentata, con una coesistenza di nuclei altamente attivi e di partecipazioni marginali, delineando una configurazione discorsiva più complessa e diseguale nel periodo dopo il conflitto.

Tab. 1 – Misure di rete nelle settimane considerate tra il 2022 e 2025

	2022	**2023**	**2024**	**2025**
Vertici	8895	9405	8231	10168
Archi	12060	12420	14822	16398
Rapporto di Coppie di Vertici Reciproche	0,019	0,0211	0,0256	0,0233
Numero Massimo di Vertici in una Componente Connessa	6303	7423	8231	8903
Numero Massimo di Archi in una Componente Connessa	10512	11257	14822	15617
Distanza Geodetica Massima (Diametro)	18	20	16	14
Distanza Geodetica Media	6,2	6,0	5,3	5,3
Densità (numero indice base 2022)	100,0	92,1	143,6	104,1
Gruppi	979 gruppi (42 gruppi > 15 vertici)	765 gruppi (47 gruppi > 15 vertici)	560 gruppi (40 gruppi > 15 vertici)	576 gruppi (35 gruppi > 15 vertici)

Nella settimana del 2022 (Fig. 2 e tab. 2), l'analisi della rete su 𝕏 ha evidenziato tra gli utenti più rilevanti i politici e i partiti, come *Giorgia Meloni, Matteo Salvini e Fratelli d'Italia*, e giornalisti e testate giornalistiche, tra cui *Paolo Berizzi, Alberto Angela, Repubblica* e *La Stampa*. Altri attori importanti, come la *Rete Italiana Antifascista* e *l'A.N.P.I.* svolgono un ruolo significativo nella comunicazione e nella diffusione di contenuti informativi. *YouTube* rappresenta la piattaforma principale di amplificazione dei contenuti, connessa sia a media che a figure politiche. La visualizzazione del grafo mostra gruppi tematici distinti: i gruppi 1, 6 e 11 aggregano giornalisti e testate

ISBN Open Access: 9788835191971

giornalistiche, il gruppo 4 unisce politici e formazioni politiche, mentre i gruppi 2 e 17 rappresentano programmi televisivi e *broadcaster*, punti chiave per la diffusione delle informazioni.

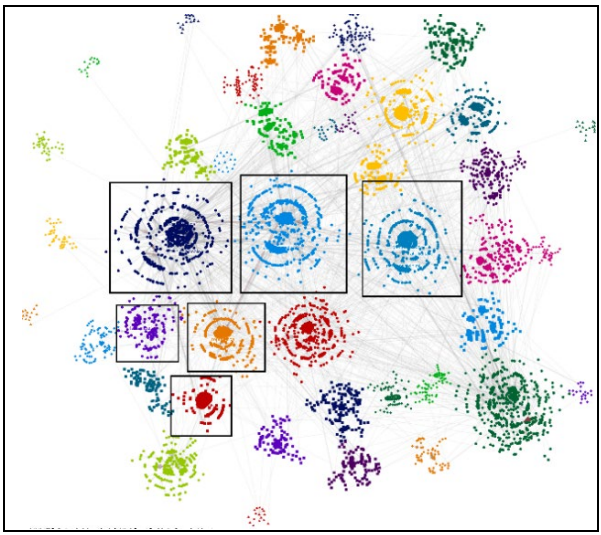


Fig. 2 – Visualizzazione della rete di interazioni tra utenti di 𝕏 raggruppati in comunità –2022. I riquadri rappresentano i gruppi di utenti evidenziati nella Tabella 2

Tab. 2 – Utenti di 𝕏 più influenti in base agli indici di centralità (IC_1 Indegree, IC_2 Betweenness, IC_3 Closeness, IC_4 Eigenvector) e per gruppo di appartenenza – 2022

Profilo 𝕏	Nome utente 𝕏	IC_1	IC_2	IC_3	IC_4	Gruppo
Pberizzi	*Paolo Berizzi*	291	6233824,32	0,19	0,26	1
Italiantifa	*Rete Italiana Anti-fascista*	189	362510,05	0,17	0,23	1
Anpinazionale	*A.N.P.I. Nazionale*	82	1448855,38	0,18	0,08	1
Romaebraica	*Comunità Ebraica di Roma*	43	2205073,89	0,18	0,01	1
Albertoangela	*Alberto Angela*	107	2301093,85	0,17	0,00	2
Giorgiameloni	*Giorgia Meloni*	182	3961733,50	0,17	0,00	4
matteosalvinimi	*Matteo Salvini*	70	1586573,84	0,17	0,00	4
Fratelliditalia	*Fratelli d'Italia* 𝕀𝕋	38	589181,64	0,15	0,00	4
Repubblica	*Repubblica*	96	4005061,07	0,18	0,01	6
Lastampa	*La Stampa*	44	974884,65	0,16	0,00	11
Youtube	*YouTube*	80	1177690,10	0,14	0,00	17

ISBN Open Access: 9788835191971

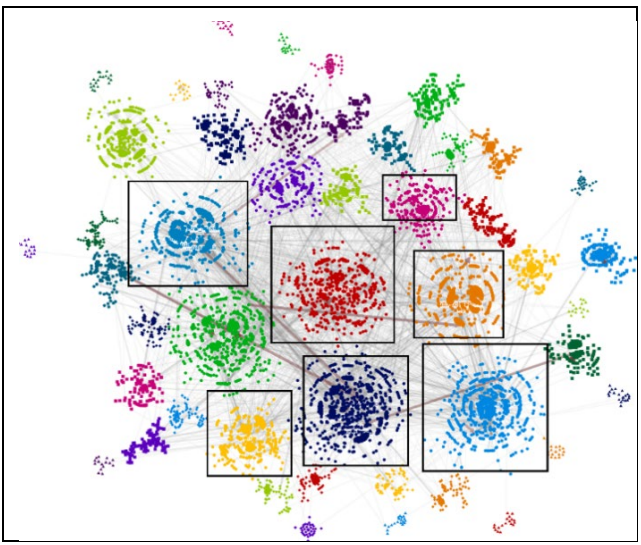


Fig. 3 – Visualizzazione della rete di interazioni tra utenti di 𝕏 raggruppati in comunità –2023. I riquadri rappresentano i gruppi di utenti evidenziati nella Tabella 3

Tab. 3 – Utenti di 𝕏 più influenti in base agli indici di centralità (IC_1 Indegree, IC_2 Betweenness, IC_3 Closeness, IC_4 Eigenvector) e per gruppo di appartenenza –2023

Profilo 𝕏	Nome utente 𝕏	IC_1	IC_2	IC_3	IC_4	Gruppo
senatostampa	*Senato Repubblica*	64	1943300,79	0,20	0,02	1
guidocrosetto	*Guido Crosetto*	64	1584722,13	0,19	0,03	1
fabfazio	*Fabio Fazio*	151	2311865,35	0,19	0,08	2
raiuno	*Rai1*	110	1865134,06	0,19	0,10	2
chetempochefa	*Che Tempo Che Fa*	71	1131949,04	0,18	0,03	2
giorgiameloni	*Giorgia Meloni*	208	4879836,92	0,21	0,47	3
quirinale	*Quirinale*	180	5275201,27	0,21	0,41	3
fratelliditalia	*Fratelli d'Italia* 𝐼𝑇	134	2899140,42	0,20	0,14	3
ignazio_larussa	*Ignazio La Russa*	118	3374750,17	0,20	0,07	3
jacopo_iacoboni	*jacopo iacoboni*	85	1662404,10	0,18	0,01	5
pberizzi	*Paolo Berizzi*	131	3847028,69	0,20	0,05	6
Repubblica	*Repubblica*	102	2498891,49	0,20	0,03	6
fattoquotidiano	*Il Fatto Quotidiano*	92	3230787,84	0,20	0,02	6
linkiesta	*Linkiesta*	52	1586817,70	0,19	0,01	7
marcofattorini	*Marco Fattorini*	109	3573758,98	0,20	0,02	9
lastampa	*La Stampa*	45	1546579,09	0,19	0,02	19

ISBN Open Access: 9788835191971

Nel periodo 2023 (Fig. 3 e tab. 3) emergono nel dibattito istituzioni politiche come il Senato della Repubblica e il Quirinale, politici e partiti quali *Guido Crosetto, Giorgia Meloni, Fratelli d'Italia e Ignazio La Russa*, nonché giornalisti e testate giornalistiche come *Jacopo Iacoboni, Paolo Berizzi, Repubblica, Il Fatto Quotidiano, Linkiesta, Marco Fattorini e La Stampa*. Inoltre, programmi televisivi e *broadcaster*, tra cui *Fabio Fazio, Rai1* e il programma televisivo *Che Tempo Che Fa*, fungono il ruolo di amplificatori dei contenuti. La struttura del grafo mostra una chiara segmentazione: i gruppi 5, 6, 7, 9 e 19 uniscono i giornalisti e le testate giornalistiche, i gruppi 1 e 3 comprendono politici e partiti politici, mentre il gruppo 2 rappresenta i programmi televisivi e *broadcaster*.

Dopo l'avvio del conflitto (Fig. 4 e tab. 4), tra gli utenti principali emergono nel 2024 giornalisti e testate giornalistiche come *Marco Fattorini*, *Federico Rampini*, *Paolo Berizzi*, *Il Fatto Quotidiano*, *Corriere della Sera*, *La Repubblica*, *Il Foglio* e *l'Agenzia ANSA*, nonché istituzioni come la *Farnesina* e il *Quirinale*. Sul versante politico, si distinguono *Antonio Tajani, Giorgia Meloni, il Partito Democratico, Ignazio La Russa, Matteo Salvini e Fratelli d'Italia*. I gruppi 1, 4 e 9 comprendono giornalisti, associazioni e testate giornalistiche, mentre i gruppi 2 e 3 aggregano partiti e singoli politici. Il gruppo 4, inoltre, rappresenta le agenzie di stampa, fulcro per la diffusione dei contenuti informativi.

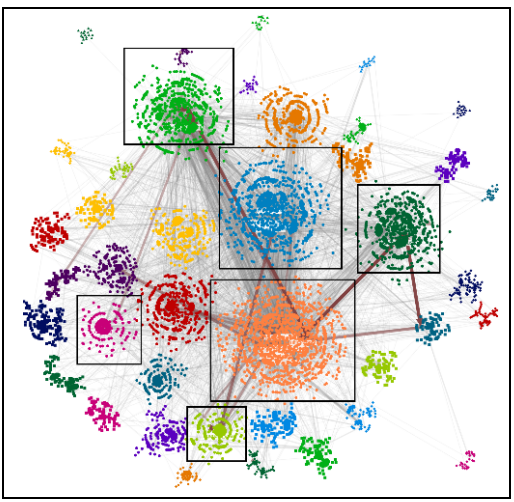


Fig. 4 – Visualizzazione della rete di interazioni tra utenti di 𝕏 raggruppati in comunità – settimana 2024. I riquadri rappresentano i gruppi di utenti evidenziati nella Tabella 4

ISBN Open Access: 9788835191971

Tab. 4 – Utenti di 𝕏 più influenti in base agli indici di centralità (IC_1 Indegree, IC_2 Betweenness, IC3 Closeness, IC_4 Eigenvector) e per gruppo di appartenenza – 2024

Profilo 𝕏	Nome utente 𝕏	IC_1	IC_2	IC_3	IC_4	Gruppo
antonio_tajani	*Antonio Tajani*	117	2445191,18	0,28	0,03	1
ilfoglio_it	*Il Foglio*	117	2445191,18	0,28	0,03	1
marcofattorini	*Marco Fattorini*	117	2445191,18	0,28	0,03	1
israelinitaly	*Israele in Italia*	117	2445191,18	0,28	0,03	1
federicorampini	*Federico Rampini*	117	2445191,18	0,28	0,03	1
italymfa	*Farnesina IT*	117	2445191,18	0,28	0,03	1
ultimoranet	*Ultimora.net*	117	2445191,18	0,28	0,03	1
repubblica	*Repubblica*	224	5591780,95	0,24	0,04	2
pberizzi	*Paolo Berizzi*	114	2395146,38	0,23	0,02	2
giorgiameloni	*Giorgia Meloni*	109	3434433,32	0,24	0,01	2
pdnetwork	*Partito Democratico IT EU*	101	2616213,50	0,23	0,01	2
quirinale	*Quirinale*	56	1476283,27	0,23	0,01	2
ignazio_larussa	*Ignazio La Russa*	148	3193240,80	0,23	0,01	3
matteosalvinimi	*Matteo Salvini*	107	1901610,45	0,22	0,01	3
fratelliditalia	*Fratelli d'Italia IT*	79	1325963,72	0,22	0,01	3
fattoquotidiano	*Il Fatto Quotidiano*	129	2900221,27	0,24	0,08	4
agenzia_ansa	*Agenzia ANSA*	181	3786771,20	0,24	0,02	8
corriere	*Corriere della Sera*	187	4329217,70	0,23	0,03	9

Infine, nel 2025 (Fig. 5 e tab. 5) tra i nodi principali si distinguono associazioni e realtà civiche come *A.N.P.I.*, *Roma Ebraica* e la *Rete Italiana Antifascista*, giornalisti e testate come *Paolo Berizzi*, *La Stampa*, *Corriere della Sera* e *La Repubblica*, programmi televisivi come *Che Tempo Che Fa*, politici e partiti quali *Giorgia Meloni*, *Matteo Salvini*, *Ignazio La Russa*, *Fratelli d'Italia* e il *Quirinale*. La segmentazione per gruppi mostra come i gruppi 1, 2 e 6 sono caratterizzati dalla presenza di giornalisti, associazioni e testate giornalistiche, i gruppi 4 e 11 da politici e partiti politici, mentre i gruppi 10, 18 e 21 da programmi televisivi e *broadcaster*.

ISBN Open Access: 9788835191971

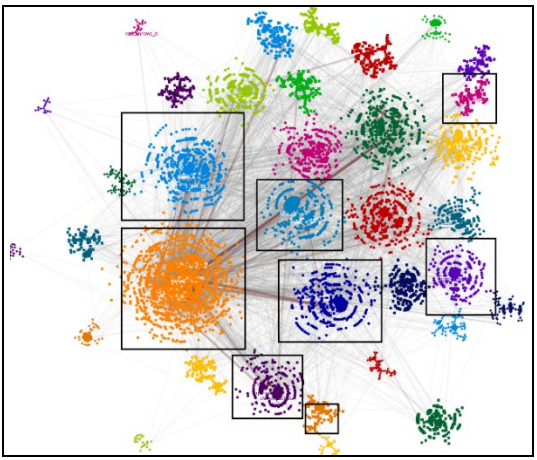


Fig. 5 – Visualizzazione della rete di interazioni tra utenti di 𝕏 raggruppati in comunità –2025. I riquadri rappresentano i gruppi di utenti evidenziati nella Tabella 5

Tab. 5 – Utenti di 𝕏 più influenti in base agli indici di centralità (IC_1 Indegree, IC_2 Betweenness, IC_3 Closeness, IC_4 Eigenvector) e per gruppo di appartenenza – 2025

Profilo 𝕏	Nome utente 𝕏	IC_1	IC_2	IC_3	IC_4	Gruppo
Anpinazionale	*A.N.P.I. Nazionale*	160	3429650,60	0,24	0,35	1
Corriere	*Corriere della Sera*	152	5153600,37	0,25	0,16	1
ilfoglio_it	*Il Foglio*	123	3324997,03	0,24	0,13	1
Gadlernertweet	*Gad Lerner*	105	1220109,10	0,24	0,26	1
Marcofattorini	*Marco Fattorini*	57	1266765,06	0,22	0,02	1
Bettafiorito	*Elisabetta Fiorito*	50	738638,61	0,22	0,05	1
Romaebraica	*Comunità Ebraica di Roma*	49	1185311,94	0,23	0,05	1
Ilriformista	*Il Riformista*	35	256315,01	0,22	0,06	1
Lastampa	*La Stampa*	160	5140227,04	0,25	0,15	2
ignazio_larussa	*Ignazio La Russa*	73	1237761,27	0,22	0,03	2
italiantifa	*Rete Italiana Antifascista*	57	1171157,14	0,22	0,07	2
fratelliditalia	*Fratelli d'Italia*	241	4841705,31	0,23	0,15	4
giorgiameloni	*Giorgia Meloni*	106	3083423,51	0,24	0,07	4
pberizzi	*Paolo Berizzi*	218	5123969,81	0,24	0,22	6
Repubblica	*Repubblica*	112	2772462,70	0,24	0,13	6
elonmusk	*Elon Musk*	38	1234464,96	0,21	0,02	6
antonio_tajani	*Antonio Tajani*	29	566023,31	0,21	0,01	6
chetempochefa	*Che Tempo Che Fa*	150	3568818,45	0,23	0,09	10
quirinale	*Quirinale*	83	2400053,44	0,23	0,06	10
matteosalvinimi	*Matteo Salvini*	118	2328012,48	0,22	0,02	11
youtube	*YouTube*	49	893208,60	0,21	0,02	18
skytg24	*Sky tg24*	26	321404,87	0,20	0,01	21

ISBN Open Access: 9788835191971

5.2. Trasformazioni del discorso online nel 2023 e nel 2024

L'analisi dei contenuti dei *tweet* attraverso le *word cloud* (Figure 6 e 7) mostra un cambiamento significativo tra le settimane del 2023 e del 2024. Prima del 16 ottobre 2023 (Fig. 6), i termini predominanti riguardano la memoria storica e la lotta contro l'antisemitismo, con parole chiave quali Auschwitz, Antisemitismo, Olocausto e Liliana Segre, evidenziando un focus sull'educazione, sulla commemorazione e sulla sensibilizzazione civica. Nel 2024 (figura 7), invece, la discussione si concentra su temi geopolitici attuali e conflitti, con termini quali Palestinesi, Palestina, Gaza, Hamas, Israele, Sionismo, Genocidio, accompagnati da un incremento del linguaggio offensivo ed espressioni di ostilità. Questo passaggio evidenzia uno spostamento dai temi della memoria storica e della riflessione civile verso dibattiti più polarizzati e conflittuali, in cui le tensioni geopolitiche dominano il discorso online.

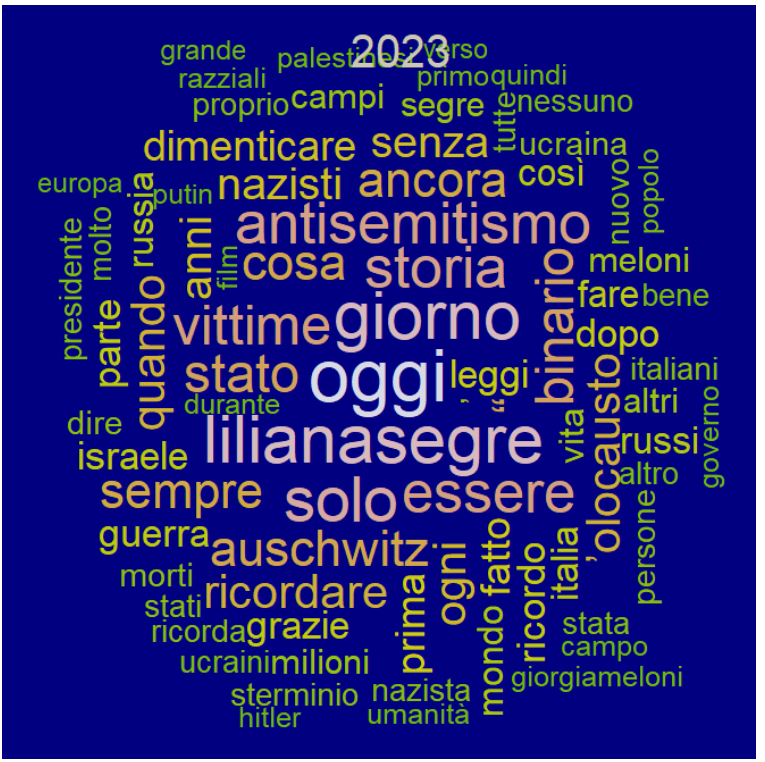


Fig. 6 – Word Cloud delle parole estratte da 𝕏: 2023

ISBN Open Access: 9788835191971

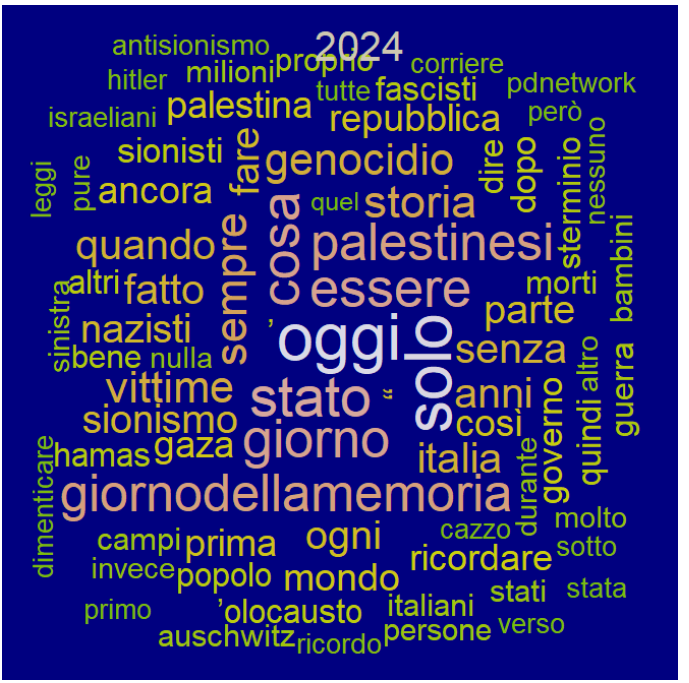


Fig. 7 – Word Cloud delle parole estratte da 𝕏: 2024

Risultati analoghi emergono dall'estrazione dei temi dominanti mediante la tecnica del *Topic Modeling*. L'analisi è stata effettuata mediante *STM* (Roberts *et al.*, 2019). Sono stati analizzati separatamente i *corpora* relativi al 2023 e al 2024. L'analisi dei discorsi sui *social media* nel 2023 evidenzia una forte concentrazione tematica attorno alla memoria della *Shoah* e al *Giorno della Memoria*, che emerge come il nucleo centrale del dibattito. I temi più rilevanti richiamano in modo ricorrente il *nazismo*, il *fascismo*, *l'Olocausto*, *Auschwitz* e la *persecuzione degli ebrei*, spesso in connessione con pratiche commemorative, cerimonie ufficiali e la ricorrenza istituzionale del Giorno della Memoria.

Accanto alla dimensione memoriale, si osserva una significativa presenza di temi legati all'antisemitismo contemporaneo, al razzismo e alla loro declinazione nel contesto politico e geopolitico attuale, inclusi riferimenti a *Israele, Palestina, Russia e Ucraina*. Tali elementi indicano una sovrapposizione tra memoria storica e interpretazioni del presente.

Un ulteriore filone riguarda il ruolo delle istituzioni, dei media e delle figure pubbliche (presidenti, giornalisti, intellettuali, programmi televisivi), nonché l'ambito *educativo*, con riferimenti alla *scuola*, agli *studenti*, ai *libri* e alla trasmissione intergenerazionale della memoria.

ISBN Open Access: 9788835191971

Nel complesso, i temi delineano un discorso pubblico in cui la commemorazione della Shoah funge da spazio privilegiato per riflessioni identitarie, politiche e morali, spesso attraversate da tensioni ideologiche e da un linguaggio emotivamente polarizzato.

Nel 2024 il discorso sui *social media* continua a essere fortemente strutturato attorno alla *Shoah* e al *Giorno della Memoria*, ma mostra una marcata riorientazione verso il presente politico e geopolitico. I temi più rilevanti combinano infatti riferimenti alla memoria storica (*Shoah, Olocausto, Auschwitz, campi di sterminio*) con un lessico fortemente ancorato all'attualità, in particolare al conflitto israelo-palestinese, all'antisemitismo contemporaneo e alle sue declinazioni nel dibattito pubblico. Accanto ai temi commemorativi, emergono con forza i riferimenti a *Israele, Gaza, Palestina, Hamas, genocidio*, spesso associati a valutazioni morali, accuse e prese di posizione polarizzate. Si evidenzia una presenza significativa del fascismo italiano, del neofascismo e delle relative controversie politiche, con richiami a *Mussolini*, alla *Repubblica* e a figure istituzionali; il ruolo dei media emerge ancora con i riferimenti ai social network e agli attori politici, con citazioni di testate giornalistiche, programmi televisivi, leader di partito e ministri.

Nel complesso, il discorso del 2024 appare più conflittuale, politicizzato e polarizzato, con un uso della memoria storica come strumento interpretativo e retorico per leggere il presente.

In sintesi, se nel 2023 la memoria della Shoah sui social media appare principalmente come uno spazio di commemorazione e riflessione storica, nel 2024 si configura sempre più come un campo di contesa simbolica e politica, in cui il passato viene mobilitato per interpretare, legittimare o criticare eventi e conflitti del presente. Questa evoluzione segnala un rafforzamento della dimensione conflittuale del discorso pubblico digitale e una crescente politicizzazione della memoria collettiva.

6. Discussione e conclusioni

I risultati dell'analisi forniscono un'evidenza, seppur preliminare, di una trasformazione significativa sia nelle strutture di rete sia nei domini tematici: mentre nell'intorno del Giorno della Memoria precedente al 7 ottobre 2023 (27 gennaio 2023) appare maggiormente centrato sulla memoria della Shoah e sulla commemorazione istituzionale, in quello successivo del 2024 (27 gennaio 2024), emerge una crescente centralità di temi legati al conflitto israelo-palestinese, accompagnata da una frammentazione reticolare e da forme di antisemitismo associate alla proiezione delle azioni del Governo israeliano sulla rappresentazione semantica dell'ebreo *in quanto tale*.

ISBN Open Access: 9788835191971

L'integrazione tra analisi di rete e analisi testuale è risultata quindi efficace nel cogliere la complessità delle dinamiche dell'odio online, offrendo un quadro interpretativo utile anche per lo studio di altri fenomeni di razzismo e xenofobia digitale.

L'analisi condotta riconferma il Giorno della Memoria come un momento chiave per la rielaborazione del discorso pubblico in Italia, ma evidenzia anche come il suo significato venga progressivamente rinegoziato nel contesto digitale, soprattutto in presenza di eventi geopolitici di forte impatto pubblico. Il riesplodere del conflitto mediorientale, dopo il 7 ottobre 2023, e le vittime civili a Gaza hanno contribuito a ridefinire le narrazioni online, alimentando il dibattito pubblico su questioni storiche quali la Nakba arabo-palestinese[6], ma anche facendo emergere le forme di ostilità generalizzata verso gli ebrei che sono presenti nelle discussioni online.

Dal punto di vista reticolare, si osserva una maggiore frammentazione e polarizzazione delle posizioni, con l'emergere di cluster distinti legati a media, istituzioni e attori politici. Sul piano semantico, nel 2023 i temi tradizionali della Shoah risultano caratterizzanti e distintivi. Nel 2024, invece, tali temi perdono centralità ed emergono tematiche associate al conflitto israelo-palestinese; tema caratterizzato da una minore coesione dei termini e da un linguaggio maggiormente oppositivo. Questo cambiamento segnala un indebolimento della funzione commemorativa e una crescente strumentalizzazione della memoria storica della Shoah.

Infine, dal punto di vista metodologico, l'identificazione di una finestra temporale pubblicamente rilevante per la raccolta dei dati online, associata a date istituzionali rilevanti come le commemorazioni o le elezioni, consente di individuare degli shock esogeni che possano determinare un mutamento nel dibattito pubblico. L'integrazione tra *Social Network Analysis* e *Natural Language Processing*, *Text Mining* e *Structural Topic Models,* si dimostra efficace per cogliere i mutamenti semantici nei dibattiti online, offrendo un approccio utile anche per l'analisi di altri fenomeni di razzismo e xenofobia online. In prospettiva futura, e come riflessione su possibili sviluppi di ricerca, incrociando comunità di utenti e temi latenti sarà possibile rispondere a domande di ricerca che indaghino su come gli stessi temi assumano significati diversi in gruppi differenti, se esistono gruppi definiti più dai contenuti che dalle relazioni e, infine, come la struttura della rete possa influire sulla circolazione e sulla stabilizzazione dei concetti.

[6] L'esodo palestinese del 1948, noto come Nakba ("catastrofe"), indica l'espulsione e la fuga forzata di una larga parte della popolazione araba palestinese durante la guerra civile del 1947-'48, alla fine del Mandato britannico, e della successiva guerra arabo-israeliana seguita alla proclamazione dello Stato di Israele.

ISBN Open Access: 9788835191971

Riferimenti bibliografici

Alkomah, F. and Ma, X. (2022). A literature review of textual hate speech detection methods and datasets. *Information*, 13(6): 273.

Bjola, C. and Manor, I. (2020). Combating Online Hate Speech and Anti-Semitism. *Oxford Department of International Development*, University of Oxford (DigDiploROx Working Paper No 4).

Clauset, A., Newman, M. E. J. and Moore, C. (2004). Finding community structure in very large networks. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*, 70(6): 066111.

CDEC – Centro di Documentazione Ebraica Contemporanea. (2024). *Relazione sull'antisemitismo in Italia*. Milano.

CDEC – Centro di Documentazione Ebraica Contemporanea. (2025). *Antisemitismo e discorso d'odio online: monitoraggio e analisi*. Milano.

de la Fuente, O. P., Tsesis, A. and Skrzypczak, J., eds. (2023). *Minorities, Free Speech and the Internet*. Taylor & Francis, London.

Fonseca, A., Pontes, C., Moro, S., Batista, F., Ribeiro, R., Guerra, R. and Silva, C. (2024). Analyzing hate speech dynamics on Twitter/X: Insights from conversational data and the impact of user interaction patterns. *Heliyon*, 10(11): e32246.

Freeman, L.C. (1978-79). Centrality in social networks conceptual clarification, *Social Networks*, 1, 3: 215-239.

Fuchs, C. (2021). *Social Media: A Critical Introduction*. Sage, London, 2021.

Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B. and Espina, C. (2017). Classifying Twitter topic-networks using social network analysis. *Social media + society*, 3(1): 2056305117691545.

Hübscher, M. and Von Mering, S., eds. (2022). *Antisemitism on social media*. Routledge, London.

Langmuir, G. I. (1990). *History, religion, and antisemitism*. Univ of California Press, USA.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Pontes, C., Fonseca, A., Moro, S., Batista, F., Ribeiro, R., Marques, C. and Guerra, R. (2024). Unveiling Patterns of Hate Speech in the Portuguese Sphere: A Social Network Analysis Approach. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 70-81). Springer Nature Switzerland, Cham.

Rawat, A., Kumar, S. and Samant, S. S. (2024). Hate speech detection in social media: Techniques, recent trends, and future challenges. Wiley Interdisciplinary Reviews: *Computational Statistics*, 16(2): e1648.

Roberts, M. E., Stewart, B. M. and Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2): 1–40.

Siegel, A.A. (2020). Online hate speech. In N. Persily J.A. Tucker, eds., *Social Media and Democracy: The State of the Field, Prospects for Reform* (pp. 56–88). Cambridge University Press, Cambridge

ISBN Open Access: 9788835191971

Smith, M., Ceni A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J. and Dunne, C. (2010). *NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016* from the Social Media Research Foundation.

Vandebosch, H. and Rothmund, T. (2024). Online hate: A European communication perspective. *Communications*, 49(3): 371–377.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge University Press, Cambridge.

Weimann, G. (2024). New trends in online antisemitism. In Hartleb F., ed., *Research Study.Antisemitism on the Rise New Ideological Dynamics* (pp. 43-53). European Institute for Counter Terrorism and Conflict Prevention (EICTP), Vienna.

Welbers, K., Van Atteveldt, W. and Benoit, K. (2017). Text analysis in R. *Communication methods and measures*, 11(4): 245-265.

Wong, S. G. J. (2024). What is the social benefit of hate speech detection research? A Systematic Review. In *Proceedings of the Third Workshop on NLP for Positive Impac*t (pp. 1-12).

ISBN Open Access: 9788835191971

L'altro che immaginiamo e la produzione sociale dell'alterità. Il caso delle comunità romanès in Italia

di *Maria Pia Franciosa*[*], *Valentina Isabella*[*], *Melissa Vassallo*[**]

1. Introduzione

L'antiziganismo come forma specifica di razzismo strutturale, non si configura come mero pregiudizio individuale, ma come sistema di esclusione radicato nelle politiche pubbliche, nelle pratiche istituzionali nonché nell'immaginario collettivo (Parekh e Rose, 2011; Piasere, 2015; Vitale, 2009). Tale meccanismo di costruzione dell'alterità, lungi dall'essere un retaggio del passato, continua nel contesto italiano a produrre marginalizzazione contribuendo allo sviluppo di stereotipi nei confronti delle comunità romanès (Menghini, 2017; Pasta e Vitale, 2018; Villano, 2017). L'antiziganismo si manifesta attraverso l'intreccio tra categorizzazioni identitarie imposte, barriere strutturali e forme di segregazione che assumono tanto una dimensione materiale quanto simbolica. In questo quadro, le traiettorie di vita delle persone coinvolte risultano profondamente eterogenee, segnate da storie, condizioni giuridiche e livelli di stabilità economica ed abitativa differenti. Elementi che, tuttavia, vengono spesso oscurati da rappresentazioni pubbliche omogeneizzanti, incapaci di cogliere la pluralità interna alle comunità e la complessità dei posizionamenti individuali (McGarry, 2014).

Un ruolo centrale in questi processi è svolto dalle rappresentazioni mediatiche, che contribuiscono a rafforzare la stigmatizzazione all'interno di un più ampio sistema di produzione dell'antiziganismo (Tremlett *et al.*, 2017). La narrazione pubblica si inserisce infatti in un circuito circolare di produzione del senso comune, nel quale immagini selettive e narrazioni stereotipate, non solo riflettono narrazioni categorizzanti preesistenti, ma ne alimentano la reiterazione. Tali narrazioni interagiscono con dispositivi normativi e

[*] Università G. d'Annunzio, mariapia.franciosa@studenti.unich.it; valentina.isabella001@studenti.unich.it

[**] Dipartimento di Scienze Giuridiche e Sociali, Università G. d'Annunzio, melissa.vassallo@unich.it

ISBN Open Access: 9788835191971

pratiche politiche, incidendo sulle modalità attraverso cui le istituzioni e la sfera pubblica delineano il ruolo delle comunità romanès nella società, spesso attraverso generalizzazioni talvolta legate a presunti comportamenti criminali (Ramadan, 2025). Queste forme di discriminazione, frequentemente latenti e sottostimate incidono in modo significativo sulla vita sociale, limitando le opportunità di mobilità socio-lavorativa e contribuendo a riprodurre condizioni di subalternità strutturale (Hellgren e Zapata-Barrero, 2025).

Per comprendere come questi processi operino nella vita quotidiana, l'analisi si avvale di venti interviste qualitative realizzate in cinque territori italiani ad appartenenti alle comunità romanès e ad operatori del terzo settore. Le narrazioni raccolte permettono di osservare da vicino l'intreccio tra categorizzazioni esterne, barriere strutturali e forme di agency individuale. Attraverso tali prospettive si intende far emergere zone di continuità, contraddizione e resistenza spesso poco valorizzate nel discorso pubblico.

1.1. La classificazione, stigma e produzione dell'alterità

Come evidenziato dalla Commissione straordinaria per la tutela e la promozione dei diritti umani nel Rapporto conclusivo dell'indagine sulla condizione di Rom, Sinti e Caminanti in Italia del 2023 "*mancano criteri precisi per classificare una persona o un gruppo come appartenente alle minoranze rom, sinti o caminanti*". Tale imprecisione categoriale non è neutra: essa riflette e perpetua una violenza simbolica (Bourdieu e Passeron, 1970) che nega l'autodeterminazione identitaria dei gruppi minoritari.

L'omogeneizzazione sotto un'unica etichetta produce quello che Becker (1963) definisce "*labeling*": un processo attraverso cui le categorie imposte dall'esterno generano conseguenze sociali e materiali sulla vita delle persone così classificate. In tale contesto, il *labeling* assume un ruolo cruciale nella definizione di "identità burocratiche", meccanismo fondamentale attraverso cui il sistema politico e le pratiche amministrative gestiscono e categorizzano l'alterità (Sigona, 2002).

Come discusso in un altro contributo di questo volume (Bevilacqua *et al.*, 2026), i processi di categorizzazione costituiscono uno degli strumenti principali attraverso cui si produce l'alterità. Nel caso delle popolazioni romanì, la classificazione esterna ha una lunga storia: denominazioni come "zingari", "nomadi", "gitani", "*gipsy*" o "*tsiganes*" (Piasere, 2009) e gli stereotipi che le accompagnano hanno contribuito a costruire un'immagine pubblica omogeneizzante, alla base delle forme contemporanee di antiziganismo (Piasere, 2012; 2015).

Di particolare conflittualità risulta il termine "zingaro", eteronimo dispre-

ISBN Open Access: 9788835191971

giativo che affonda le radici nell'Impero Bizantino, il vocabolo deriva da "*athinganos*" o "*atsinganos*", il cui significato è "intoccabile" o, in senso più esteso, "coloro che non devono essere toccati" (Spinelli, 2022). Originariamente riferito a una setta ascetica del IX secolo, il termine venne successivamente trasferito ai gruppi rom (Kármán, 2018).

L'uso contemporaneo del termine "zingaro" non è neutrale, come indicato all'interno del Glossario della Carta di Roma[1], che ne segnala dunque la natura stigmatizzante e l'inadeguatezza dell'utilizzo all'interno del discorso pubblico e giornalistico.

1.2. Antiziganismo istituzionale e segregazione spaziale

La discriminazione nei confronti delle comunità romanès conta oltre cinque secoli in Europa (Perocco *et al.*, 2016). Solo in tempi relativamente recenti, tuttavia, le istituzioni europee hanno riconosciuto formalmente la specificità di questo fenomeno: tra il 2005 e il 2006, in occasione delle conferenze di Varsavia e Bucarest, viene ufficializzato il termine "*antigypsyism*", sancendo il riconoscimento dell'antiziganismo come forma specifica di xenofobia e razzismo strutturale (Spinelli, 2022). Un importante contributo alla concettualizzazione di tale fenomeno è stato fornito da Nicolae (2007).

La profondità storica di tale forma di razzismo è testimoniata da una lunga sequenza di persecuzioni sistemiche, che spaziano dalla legislazione ghettizzante fino al genocidio perpetrato dal regime nazifascista, noto come "*Porrajmos*" (il divoramento) o "*Samudaripen*" (l'uccisione di tutti) (Perocco *et al.*, 2016; Bassoli e Bravi, 2013).

Nel contesto italiano, a partire dalla metà degli anni Settanta e con una progressiva formalizzazione normativa dal 1985, numerose amministrazioni locali hanno iniziato a governare la presenza delle comunità romanès attraverso la realizzazione di insediamenti separati, comunemente e impropriamente definiti "campi nomadi". La genesi di tali politiche è riconducibile a una lettura fortemente culturalista e astorica della mobilità, interpretata come tratto identitario volontario piuttosto che come esito di condizioni sociali, economiche e politiche specifiche, quali conflitti, persecuzioni o l'esclusione sistematica dall'accesso agli alloggi ordinari (Picker, 2015). Tale impostazione ha dato luogo a un sistema di governance fondato su dispositivi ambi-

[1] La Carta di Roma è un protocollo deontologico per l'informazione giornalistica sui temi delle migrazioni, dell'asilo e delle minoranze, adottato in Italia nel 2008 e sottoscritto dal Consiglio Nazionale dell'Ordine dei Giornalisti e dalla Federazione Nazionale della Stampa Italiana, in collaborazione con l'UNHCR. Il documento fornisce linee guida e glossari volti a prevenire stereotipi, linguaggi discriminatori e rappresentazioni distorte nei media.

ISBN Open Access: 9788835191971

valenti, in cui interventi presentati come inclusivi o solidali hanno finito per tradursi in modalità di gestione separata. Nel caso di Roma, Clough Marinaro (2003) mostra come tali politiche si basassero su assunzioni che rappresentavano le comunità romanès come intrinsecamente nomadi e incapaci di integrarsi, legittimando così soluzioni abitative eccezionali e segregate. All'interno di tale assetto, la segregazione non può essere intesa come una condizione statica o come un mero esito spaziale, ma come un processo dinamico di concentrazione e separazione di gruppi omogeneamente categorizzati come "altri". Tale processo si sviluppa e si rinnova nel tempo attraverso pratiche amministrative, scelte politiche e modalità ordinarie di governo della differenza, assumendo forme variabili, talvolta intermittenti, ma profondamente radicate nelle storie locali e nelle interazioni quotidiane (Picker, 2017). Un effetto concreto di tale configurazione istituzionale è stato, ad esempio, il trattamento riservato alle persone fuggite dai conflitti dell'ex Jugoslavia, cui è stata consentita la permanenza sul territorio italiano in quanto classificate come "nomadi", senza il riconoscimento dello status di rifugiati di guerra previsto dalla normativa vigente (Legge n. 390/1992). In questo modo, si sono consolidati dispositivi di segregazione urbana presentati come forme di tutela culturale, ma funzionali alla gestione amministrativa della marginalità e alla riproduzione della separazione spaziale (Picker, 2015; Sigona; 2005).

Come riportato in Manzoni (2017), le stime più accreditate indicano un intervallo compreso tra 170.000 e le 180.000 persone rom e sinti in Italia, pari a circa lo 0,2% della popolazione totale, una delle percentuali più basse d'Europa. Di queste, secondo il report "Bagliori di speranza" prodotto dall'Associazione 21 luglio (2024), solo 10.000-15.000 persone (meno del 10%) vivono in insediamenti/campi formali o informali[2]. Ciò significa che almeno 160.000 persone sono residenti in abitazioni ordinarie e in possesso della cittadinanza italiana o di Stati membri dell'UE. Questa realtà demografica contraddice radicalmente l'immaginario collettivo che associa tali comunità ai campi "nomadi".

[2] In questo lavoro si adotta la distinzione proposta dall'Associazione 21 Luglio tra insediamenti formali e informali. I primi comprendono strutture abitative progettate e gestite da enti pubblici, caratterizzate da regolamentazioni specifiche e da una logica di concentrazione spaziale su base etnica (baraccopoli, microaree, macroaree, edilizia residenziale pubblica monoetnica e centri di raccolta). Gli insediamenti informali includono invece soluzioni abitative auto-costruite o occupazioni temporanee non riconosciute formalmente, come microinsediamenti e aree di transito, spesso collocate in aree marginali e soggette a sgomberi.

ISBN Open Access: 9788835191971

1.3. Spunti di riflessione

Partendo da questo quadro, il presente contributo propone l'analisi di alcune testimonianze raccolte intorno a nuclei tematici ricorrenti, emersi nel corso delle interviste. Un primo ambito riguarda il modo in cui stereotipi e rappresentazioni sedimentate entrano nelle esperienze quotidiane delle persone, influenzando le interazioni sociali, l'accesso al lavoro e all'abitazione, nonché il vissuto dei contesti educativi formali. Un secondo insieme di riflessioni riguarda le trasformazioni delle pratiche e delle abitudini familiari, così come vengono raccontate dagli stessi interlocutori, mettendo in luce letture differenziate e talvolta ambivalenti del cambiamento. Infine, alcune testimonianze consentono di osservare le modalità attraverso cui si costruisce e si riproduce la distanza sociale, senza ricondurla a differenze essenzializzate, ma alle dinamiche relazionali e ai contesti in cui queste prendono forma.

Questo studio si propone di valorizzare le voci dirette delle persone appartenenti alle comunità romanès, integrando al contempo la prospettiva degli operatori del terzo settore senza farne il punto di vista predominante. Inoltre, il lavoro mira a sottolineare l'eterogeneità interna di queste comunità spesso ignorata dalle politiche pubbliche, mostrando come differenze di cittadinanza, territorio e storia migratoria producano condizioni di vita radicalmente diverse.

2. Quadro sociodemografico delle comunità romanès in Italia

Come ricostruito all'interno del report "*Italy – Promoting Social Inclusion of Roma – A Study of National Policies*" (Strati, 2011) la presenza di popolazioni romanès in Italia può essere descritta tramite quattro principali ondate migratorie. La prima risale a circa seicento anni fa, con l'arrivo dei primi gruppi nell'Italia meridionale e centrale. La seconda ondata si verificò negli anni successivi alla Prima Guerra Mondiale. La terza, negli anni Sessanta-Settanta del Novecento, coinvolse principalmente rom provenienti dall'ex Jugoslavia. L'ultima, iniziata tra il 1989 e il 1991 dopo il crollo dei regimi comunisti dell'Europa orientale, ha richiamato l'attenzione pubblica sulla "questione rom", in quanto i nuovi arrivati costruirono insediamenti informali duraturi. Tra gli anni Settanta e il 1992, si stima che circa 35.000 rom jugoslavi fossero arrivati in Italia; altri 10.000 giunsero successivamente durante la guerra in Bosnia-Erzegovina (1992-1995). I gruppi di più antico insediamento (XIV-XV secolo) includono le popolazioni romanès presenti nelle regioni centro-meridionali e i caminanti in Sicilia (circa 30.000 persone), oltre ai sinti stanziati prevalentemente nelle regioni settentrionali. Le

ISBN Open Access: 9788835191971

ondate migratorie successive hanno riguardato i *rom* Khorakhanè, Dasikhanè e rom Rumeni dall'Est Europa e dai Balcani.

Un quadro aggiornato delle condizioni abitative emerge invece dal già citato report "Bagliori di speranza" prodotto dall'Associazione 21 luglio (2024). Mentre, come già enfatizzato circa il 90% degli appartenenti alle comunità rom e sinti vive in abitazioni ordinarie, il restante 10% si collocherebbe presso 119 insediamenti abitati, di cui 106 formali con circa 11.100 residenti e 13 informali con circa 2.000 persone. La distribuzione geografica di tali insediamenti risulta fortemente disomogenea. Il Nord Italia concentra il maggior numero di insediamenti formali (oltre 5.000 persone), prevalentemente macroaree stabili destinate a famiglie sinte di cittadinanza italiana. Il Centro Italia ospita circa 2.700 persone, con prevalenza di baraccopoli (2.117 abitanti) rispetto a macroaree (590 abitanti). Il Lazio presenta la maggiore concentrazione, specialmente Roma, che conta sei baraccopoli attive con circa 1.826 residenti, in gran parte di origine balcanica (70%), italiana (25%) e rumena (5%). Il Sud Italia si caratterizza per l'assenza totale di macroaree e la presenza esclusiva di 20 baraccopoli che ospitano 2.786 persone (Campania, Puglia, Calabria, Sardegna). All'interno della totalità di persone che si identificano come rom in Italia, esistono almeno tre categorie giuridiche distinte: persone in possesso di cittadinanza italiana; cittadini di Stati membri dell'UE (principalmente rumeni, bulgari, polacchi); extracomunitari, categoria che include persone con permesso di soggiorno regolare, richiedenti asilo e persone in situazione di irregolarità (Loy, 2009). Particolarmente complessa è la condizione dei gruppi provenienti dall'ex Jugoslavia, molti dei quali sono apolidi o a rischio di apolidia. La maggiore precarietà giuridica e abitativa si rileva al Sud, con una presenza significativa di minori a rischio apolidia, emblematici i casi di Giugliano in Campania e dell'area metropolitana di Napoli (Senato della Repubblica, 2014).

Questa eterogeneità di status giuridici è spesso legata a conseguenze materiali rilevanti: chi possiede cittadinanza italiana o europea ha accesso formale a diritti (lavoro, istruzione, sanità, abitazione) che rimangono preclusi o fortemente limitati per extracomunitari e apolidi. Tale frammentazione giuridica si intreccia con la segregazione spaziale, producendo condizioni di vita radicalmente differenziate.

3. Metodologia

La ricerca adotta un approccio qualitativo di tipo esplorativo, finalizzato a comprendere in profondità le esperienze, le percezioni e le interpretazioni soggettive relative alle dinamiche di marginalizzazione e integrazione delle comunità rom, sinti e caminanti in Italia.

ISBN Open Access: 9788835191971

La definizione del campione dei soggetti da intervistare è avvenuta combinando contatti diretti preesistenti e il coinvolgimento di associazioni del terzo settore attive nel campo dell'inclusione sociale: INCONTRA (Puglia), On the Road (Marche/Abruzzo), AIZO (Piemonte), Voci di Dentro (Abruzzo), Karibu (Campania). Queste organizzazioni hanno agevolato l'accesso a potenziali partecipanti, operando come facilitatori (*gatekeepers*). Complessivamente sono state realizzate venti interviste in profondità, di cui undici rivolte a operatori del terzo settore e nove a persone appartenenti alle comunità rom e sinti, di età compresa tra 25 e 75 anni. La ricerca ha interessato cinque regioni italiane: Piemonte (Torino), Abruzzo (Pescara, Montesilvano, Ortona, San Salvo), Campania (Scampia, Napoli), Puglia (Bari Japigia, Lucera), Calabria (Lamezia Terme). La scelta di questi territori risponde a una logica comparativa Nord-Sud e alla presenza di diverse tipologie insediative (case popolari, campi formali, baraccopoli).

Sono state predisposte due tracce di intervista differenziate. La prima, rivolta a persone appartenenti alle comunità esplorava dimensioni identitarie, esperienze di discriminazione, difficoltà in ambito abitativo, lavorativo ed educativo, aspetti relazionali, tradizioni culturali e cambiamenti generazionali. La seconda traccia, destinata agli operatori, ha indagato la percezione del loro lavoro da parte delle comunità, le principali difficoltà incontrate dalle comunità nell'accesso a servizi, stereotipi diffusi e dinamiche di integrazione. Le interviste sono state condotte sia in presenza sia online e, previo consenso informato, registrate e successivamente trascritte integralmente, al fine di preservare fedelmente le parole dei partecipanti e ridurre il rischio di distorsioni interpretative.

Un limite significativo è rappresentato dalle difficoltà di accesso: la segregazione spaziale dei campi, spesso situati in aree periferiche, è stata in parte mitigata grazie al supporto delle associazioni quali soggetti intermediari, che hanno facilitato il contatto con gli intervistati. Infine, la dimensione ridotta del campione (20 interviste) e la copertura limitata a cinque regioni impediscono generalizzazioni. Tuttavia, l'approccio qualitativo privilegia la profondità sull'estensione, consentendo di cogliere sfumature e complessità che ricerche quantitative su larga scala potrebbero non rilevare.

4. Risultati

L'analisi qualitativa delle venti interviste ha fatto emergere quattro macrotemi interconnessi: (4.1) stereotipi, discriminazione e dinamiche di stigmatizzazione; (4.2) barriere strutturali all'accesso a diritti fondamentali; (4.3) cambiamento generazionale e trasformazioni interne; (4.4) diffidenza reciproca e distanza sociale.

ISBN Open Access: 9788835191971

4.1. Stereotipi, discriminazione e stigmatizzazione

Stereotipi e pregiudizi continuano a modellare in profondità le relazioni sociali, generando forme di stigmatizzazione che incidono sulle opportunità di vita e sulle interazioni quotidiane. La distanza simbolica e sociale che ne deriva si manifesta in ambiti diversi, dal lavoro all'abitare, fino ai rapporti interpersonali, contribuendo a riprodurre confini relazionali e trattamenti differenziati che non trovano giustificazione in caratteristiche individuali, ma in rappresentazioni sedimentate nel tempo.

Un esempio ricorrente riguarda l'accesso all'occupazione, dove la valutazione non si basa sulle competenze, ma su marcatori sociali attribuiti all'origine. Un uomo del campo di Japigia racconta: "Quando parlo al telefono e qualcuno ha bisogno di un trasloco, mi presento, mi vedono in faccia e in barese mi dicono 'ma tu chi sei?' Io dico che vengo dalla Romania e il lavoro magicamente sparisce". La perdita improvvisa di opportunità evidenzia come lo stigma agisca a livello immediato, condizionando la possibilità stessa di essere considerati come lavoratori legittimi.

In altri casi, il pregiudizio si attiva attraverso il cognome, che diventa un marcatore identitario riduttivo: "Succede che tu vai, bussi, ma come dici il cognome cambiano, se prima erano cordiali e gentili, dopo cambia tutto". Questa dinamica mostra come l'identificazione etnica attribuita dall'esterno possa produrre trattamenti differenziati e ostacolare percorsi di inclusione.

La stigmatizzazione si estende anche agli spazi economici e commerciali. Un operatore osserva: "In Abruzzo molte persone non frequentano locali o negozi dei rom". Questa esclusione indiretta contribuisce a riprodurre marginalità economica e limita la possibilità per alcune famiglie di sviluppare attività autonome.

Le narrazioni raccolte mostrano inoltre che lo stigma non opera solo dall'esterno verso le comunità romanès, ma può intrecciarsi con vissuti di vulnerabilità o di sfiducia, che prendono forma in risposta a esperienze ripetute di discriminazione. Tuttavia, quanto riportato, non si traduce mai nelle interviste in visioni generalizzate sugli italiani come gruppo, ma si riferisce a comportamenti specifici o a contesti relazionali circoscritti.

Nel complesso, il rapporto "noi/loro" che emerge non è una contrapposizione rigida, bensì un effetto di strutture sociali e simboliche che incidono sull'accesso a risorse materiali e sul riconoscimento sociale. La discriminazione non appare come un episodio isolato, ma come un insieme di pratiche diffuse che influenzano, in modo differenziato, mobilità economica, accesso al lavoro e possibilità di costruire relazioni paritarie.

ISBN Open Access: 9788835191971

4.2. Barriere strutturali: lavoro, abitazione, istruzione

Le barriere strutturali incidono in modo significativo sull'accesso all'abitazione, al lavoro e all'istruzione, delineando un quadro in cui le disparità non dipendono da caratteristiche individuali, ma da condizioni sociali e amministrative che limitano concretamente le opportunità di vita. Ciò che emerge è un sistema di vincoli che si riproduce nei processi quotidiani di selezione, nei mercati abitativi e lavorativi e negli spazi educativi, contribuendo a configurare percorsi di inclusione fortemente differenziati.

L'ambito abitativo emerge come uno dei più problematici: l'accesso può essere negato indipendentemente dalla disponibilità economica ma può scaturire da una semplice identificazione percepita. Come racconta una donna di Ortona: "una volta sono andata in affitto e dopo un mese mi hanno subito tolto perché hanno visto che noi siamo zingari. Mi hanno ridato indietro la caparra e mi hanno mandata via; avevo anche pagato in anticipo perché il proprietario richiedeva due caparre e io gliene ho lasciate tre". Qui, non sono condizioni contrattuali o comportamenti specifici a determinare il diniego, ma un sospetto generalizzato che opera come criterio di esclusione dal mercato immobiliare.

Difficoltà altrettanto rilevanti emergono nel lavoro. In alcuni casi, sono nuovamente i cognomi ad assumere una dimensione percepita come rilevante: "Quando vai a fare un colloquio di lavoro ti squadrano dalla testa ai piedi appena sentono il tuo cognome e poi è finita. Ti dicono che ti faranno sapere ma poi quella chiamata, quel messaggio non ti arriverà mai". Queste testimonianze mostrano come la possibilità di entrare nel mercato del lavoro dipenda da condizioni che eccedono la volontà individuale e si radicano in dinamiche sociali più ampie.

L'istruzione costituisce per molte famiglie un elemento centrale nei percorsi di mobilità sociale, ma non è esente da forme di distanziamento. Una madre afferma: "I figli vanno a scuola per non essere ignoranti come noi", sottolineando l'importanza crescente della scolarità delle nuove generazioni. Tuttavia, la scuola può diventare anche un luogo in cui emergono dinamiche di esclusione: "[…] L'altra mia figlia fa la terza elementare e una sua compagna l'ha chiamata 'zingara'".

Episodi simili mostrano come la frequenza scolastica non elimini automaticamente le distanze simboliche che attraversano le relazioni tra pari.

Nel complesso, queste esperienze indicano che le barriere strutturali non riguardano solo singole situazioni, ma dimensioni sistemiche che influenzano l'accesso ai diritti fondamentali. Le difficoltà abitative, lavorative ed educative descritte dagli intervistati delineano un quadro in cui l'inclusione non dipende esclusivamente dalle risorse individuali, ma dall'insieme di vin-

ISBN Open Access: 9788835191971

coli sociali e istituzionali che regolano l'appartenenza e la partecipazione alla vita pubblica.

4.3. Cambiamento generazionale e trasformazioni interne

Lungi dall'essere un insieme fisso di norme e pratiche, le modalità di vita e di relazione si trasformano nel tempo e assumono forme diverse da famiglia a famiglia. Non emergono processi lineari o cambiamenti collettivi univoci, ma adattamenti quotidiani che si intrecciano ai percorsi individuali, alle condizioni abitative e alle relazioni con il territorio. Le trasformazioni non vengono descritte come perdita o progresso, bensì come parte di un divenire che accompagna le esperienze personali e comunitarie.

In molti racconti emerge l'idea che ciò che si vive oggi non coincide del tutto con ciò che si viveva in passato. Un intervistato osserva semplicemente che "La nostra cultura oggi non è più come una volta. È molto cambiata", mentre un'altra persona sottolinea che alcune pratiche della vita quotidiana non sono più le stesse: "Ci sono cose che fino a 10 anni fa non erano possibili… Oggi è all'ordine del giorno". Queste affermazioni non delineano una direzione specifica del cambiamento, ma segnalano la consapevolezza che le pratiche sociali non rimangono identiche nel tempo.

In altre narrazioni, i cambiamenti vengono descritti come parte di un percorso personale o familiare. Una donna racconta ad esempio che "Le tradizioni non sono più come prima", segnalando il carattere dinamico delle consuetudini. In un'altra testimonianza, si riconosce la presenza di fattori nuovi che si intrecciano a elementi portanti, garantendo una continuità, senza rinunciare alle trasformazioni in atto: "No no, io le ho trasmesse (le tradizioni) perché è bello sapere [...] Adesso è cambiato. È bello così, perché la libertà è bella". In questo caso, il cambiamento non è presentato come totale, ma come una parte della propria esperienza quotidiana.

Nel complesso, le testimonianze raccolte confermano che il cambiamento non è percepito come un processo uniforme o comunitario, ma come una realtà plurale, fatta di adattamenti, negoziazioni e continuità diversificate. Le narrazioni degli intervistati mostrano che la cultura, come avviene in qualsiasi gruppo umano, si trasforma nel tempo attraverso le pratiche e le relazioni quotidiane, senza seguire traiettorie univoche e senza produrre una scissione tra ciò che è stato e ciò che è.

ISBN Open Access: 9788835191971

4.4. Diffidenza reciproca e distanza sociale

La distanza percepita che emerge dalle testimonianze dei soggetti intervistati tra appartenenti e non appartenenti alle comunità, non viene ricondotta a presunte differenze "culturali", ma a dinamiche che si sviluppano nelle interazioni quotidiane: aspettative non corrisposte, cautela reciproca e occasioni di dialogo che non si realizzano. Il confine che si produce è quindi relazionale e prende forma nell'incontro, o nella sua assenza, più che in caratteristiche attribuite ai gruppi. Alcune persone riferiscono dei pregiudizi percepiti nei loro confronti. Una donna afferma: "Pensano che siamo zingari che rubiamo. E qualcuno purtroppo lo fa. Ma anch'io posso pensare che anche tra loro ci siano persone che si comportano male, non siamo solo noi". Lo stralcio non rimanda a una contrapposizione essenziale, ma alla richiesta che le valutazioni non siano unilaterali.

In più narrazioni emerge l'idea che il riconoscimento debba essere reciproco. Una persona intervistata lo esprime così: "Noi ci siamo ambientati verso di loro e anche loro devono avere questa cosa verso di noi […] e il buono e il cattivo c'è dappertutto". Un'altra aggiunge: "Dobbiamo essere tutti alla pari, altrimenti come faccio a dire di essere pari a te se tu non mi ci fai sentire pari?". In entrambi i casi, la distanza sociale è descritta come il risultato di relazioni non equilibrate.

Vi sono inoltre casi in cui la diffidenza nasce prima dell'interazione stessa. Un referente racconta: "Ci sono italiani che non hanno mai visto uno zingaro, non hanno mai parlato con lui, ma hanno paura". La paura non è quindi necessariamente fondata sull'esperienza, ma su rappresentazioni sociali che precedono l'incontro.

Nel complesso, le testimonianze raccolte testimoniano come la distanza sociale sia il risultato di configurazioni relazionali che possono consolidarsi o attenuarsi a seconda dei contesti. Il confine tra "noi" e "loro" appare come qualcosa che si produce e si modifica nelle interazioni.

5. Conclusioni

L'analisi condotta mostra come la discriminazione verso le comunità rom non sia una semplice questione di disparità materiali, ma affondi le sue radici in una logica di alterizzazione che opera attraverso categorie, linguaggi, pratiche istituzionali e forme di visibilità selettiva, influenzando in modo diretto sia le rappresentazioni pubbliche sia l'elaborazione delle politiche locali (Vitale, 2011). Tale dinamica si inserisce in un quadro più ampio in cui, a livello europeo, le persone rom risultano tra i gruppi maggiormente esposti a stig-

ISBN Open Access: 9788835191971

matizzazione e ostilità sociale, frequentemente costruiti come "altro" rispetto alle società nazionali (Sam Nariman *et al.*, 2020).

La distanza costruita tra un "noi" e un "loro" non descrive una differenza oggettiva: la produce. È attraverso etichette, azioni amministrative, rappresentazioni mediatiche e discorsi quotidiani che l'identità rom viene continuamente ridefinita come problema, eccezione o anomalia rispetto alla cittadinanza maggioritaria, come mostrano i processi di razzializzazione e categorizzazione analizzati in letteratura (Baciu, 2020; Erjavec, 2001; Sigona, 2011).

I risultati confermano come tali processi si traducono in forme concrete di esclusione che attraversano ambiti centrali della vita sociale. L'accesso all'abitazione, al lavoro e all'istruzione emerge come uno dei principali terreni di riproduzione delle disuguaglianze: le difficoltà abitative, spesso legate a soluzioni segreganti e all'esclusione dal mercato ordinario; le discriminazioni occupazionali fondate su marcatori simbolici come il cognome o l'origine percepita; le esperienze di stigmatizzazione in ambito scolastico che incidono precocemente sui percorsi educativi.

Tali evidenze si collocano in continuità con una vasta letteratura che mostra come la marginalizzazione delle persone rom non operi in modo episodico, ma attraverso una trama coerente di vincoli istituzionali e simbolici. Numerosi studi hanno infatti documentato come l'esclusione abitativa sia strettamente intrecciata a processi più ampi di segregazione e disinteresse istituzionale (Berescu *et al.*, 2012; Váradi *et al.*, 2023), come la discriminazione nel mercato del lavoro persista indipendentemente dal livello di istruzione o dalle competenze individuali (Milcher e Fischer, 2011; O'Higgins, 2010), e come il sistema educativo, pur configurandosi come spazio potenziale di mobilità sociale, continui a riprodurre forme di separazione e stigmatizzazione precoce (O'Hanlon, 2016). Nel loro insieme, questi contributi confermano che le disuguaglianze osservate sono l'esito di barriere strutturali alla mobilità sociale, che agiscono simultaneamente sul piano formale e informale e limitano l'accesso effettivo ai diritti fondamentali (Ciaian e Kancs, 2016).

In questa prospettiva, contrastare l'antiziganismo implica riconoscere che la marginalizzazione derivi principalmente da un insieme di processi sociali che trasformano la differenza in stigmatizzazione. Le narrazioni raccolte mostrano come le persone elaborino posizionamenti molteplici attraverso pratiche di negoziazione, reinterpretazione e contestazione delle categorie imposte, evidenziando la natura processuale e relazionale dell'appartenenza.

Accanto alle dinamiche osservate nei contesti di vita quotidiana, logiche analoghe di alterizzazione risultano oggi particolarmente attive negli spazi digitali, dove il linguaggio naturale contribuisce a riprodurre ed amplificare

ISBN Open Access: 9788835191971

categorie stigmatizzanti (Chulvi 2022). In tale scenario, la possibilità di rilevare e analizzare automaticamente queste forme di esclusione assume un rilievo crescente, nella misura in cui gli strumenti digitali potrebbero contribuire non solo alla moderazione dei contenuti, ma anche a una possibile trasformazione delle rappresentazioni, rendendole più sensibili alla complessità dei vissuti. Infine, è essenziale riconoscere la parzialità di quanto emerso da questa ricerca. Le testimonianze riportate sono frammenti situati, legati a specifici contesti relazionali e non possono essere assunte come espressione univoca della pluralità delle esperienze delle comunità romanès. Altre voci, non ascoltate, non accessibili o non presenti in questo elaborato, potrebbero delineare un panorama ancora più complesso che sfugge a qualsiasi tentativo di totalizzazione.

Nel loro insieme, i risultati invitano a considerare l'antiziganismo come un processo dinamico e situato, che si produce nell'intreccio tra pratiche istituzionali, rappresentazioni sociali e interazioni quotidiane. In questa prospettiva, ulteriori ricerche potranno contribuire ad approfondire come tali processi si ridefiniscano nel tempo e nei diversi contesti, nonché le modalità attraverso cui vengono negoziati, contestati o trasformati dalle persone coinvolte.

Ringraziamenti

Si desidera ringraziare tutte le persone che hanno partecipato alla ricerca per la disponibilità e la collaborazione dimostrate nel corso dello studio. Un ringraziamento va alle persone intervistate, che hanno condiviso esperienze e riflessioni rilevanti per la comprensione dei fenomeni analizzati. Si ringraziano inoltre gli enti e le organizzazioni del terzo settore coinvolti, il cui contributo è stato fondamentale per l'accesso al campo e per la realizzazione delle interviste.

Un ulteriore ringraziamento è rivolto a tutte le persone che hanno collaborato alla progettazione, alla conduzione e alla trascrizione delle interviste. In particolare, oltre a Maria Pia Franciosa e Valentina Isabella, co-autrici del presente lavoro, le interviste sono state condotte da Giuseppe Gargiulo, Iris Tusino e Andrea Venditti, il cui apporto è stato essenziale allo sviluppo della ricerca.

ISBN Open Access: 9788835191971

Riferimenti bibliografici

Baciu, L. (2020), Reading Beyond the Label. Implications of the Critical Race Theory for the Social Work Practice with Roma People, *Revista de Asistenţă Socială*, 19, 3: 79-98.

Bassoli, M. and Bravi, L. (2013*), Il Porrajmos in Italia. La persecuzione di Rom e Sinti durante il fascismo*, I libri di Emil, Bologna.

Becker, H.S. (1963), *Outsiders. Studies in the Sociology of Deviance*, Free Press, New York.

Berescu, C., Petrović, M. and Teller, N. (2012), Housing exclusion of the Roma: Living on the edge. In Hegedüs J., Teller N. e Lux M., eds., *Social Housing in Transition Countries* (pp. 98-113), Routledge, London.

Bourdieu, P. and Passeron, J.C. (1970), *La reproduction. Éléments pour une théorie du système d'enseignement*, Minuit, Paris.

Ciaian, P. and Kancs, D.A. (2016), Causes of the Social and Economic Marginalisation: The Role of Social Mobility Barriers for Roma, EERI Research Paper Series, 03/2016.

Clough Marinaro, I. (2003), Integration or marginalization? The failures of social policy for the Roma in Rome, *Modern Italy*, 8, 2: 203-218.

Erjavec, K. (2001), Media representation of the discrimination against the Roma in Eastern Europe: The case of Slovenia, *Discourse & Society*, 12, 6: 699-727.

Hellgren, Z. and Zapata-Barrero, R. (2025), Discrimination meets interculturalism in theory, policy and practice, *International Migration*, 63, 1: 4-5.

Loy, G. (2009), *Violino tzigano*. La condizione dei rom in Italia tra stereotipi e diritti negati. In Cherchi R. e Loy G., a cura di, *Rom e Sinti in Italia. Tra stereotipi e diritti negati* (pp. 13-47), Ediesse, Roma.

Manzoni, C. (2017), Should I stay or should I go? Why Roma migrants leave or remain in nomad camps, *Ethnic and Racial Studies*, 40, 10: 1605-1622.

McGarry, A. (2014), Roma as a political identity: Exploring representations of Roma in Europe, *Ethnicities*, 14, 6: 756-774.

Meneghini, A.M. (2017), Stereotipi e paure degli italiani nei confronti degli zingari: una rassegna degli studi psicosociali condotti in Italia, *Psicologia Sociale*, 12, 1: 3-32.

Milcher, S. and Fischer, M.M. (2011), On labour market discrimination against Roma in South East Europe, *Papers in Regional Science*, 90, 4: 773-789.

Nicolae, V. (2007), *Towards a definition of anti-Gypsyism*, Roma Diplomacy, 21-30.

O'Hanlon, C. (2016), The European struggle to educate and include Roma people: A critique of differences in policy and practice in Western and Eastern EU countries, *Social Inclusion*, 4, 1: 1-10.

O'Higgins, N. (2010), "It's not that I'm a racist, it's that they are Roma". Roma discrimination and returns to education in South Eastern Europe, *International Journal of Manpower*, 31, 2: 163-187.

Parekh, N. and Rose, T. (2011), Health inequalities of the Roma in Europe: A literature review, *Central European Journal of Public Health*, 19, 3: 139-142.

ISBN Open Access: 9788835191971

Pasta, S. and Vitale, T. (2018), "Mi guardano male, ma io non guardo". Come i rom e i sinti in Italia reagiscono allo stigma. In *Razzismi, discriminazioni e diseguaglianze. Analisi e ricerche sull'Italia contemporanea* (pp. 217-241), FrancoAngeli, Milano.
Perocco, F., Basso, P. and Di Noia, L. (2016), *La condizione dei Rom in Italia*, in ARCA (Università Ca' Foscari Venezia), Rapporto (vol. 4, pp. 7-17).
Piasere, L. (2009), *I rom d'Europa. Una storia moderna*, Laterza, Roma-Bari.
Piasere, L. (2012), *Scenari dell'antiziganismo. Tra Europa e Italia, tra antropologia e politica*, SEID Editori, Firenze.
Piasere, L. (2015). *L'antiziganismo*, Laterza, Roma-Bari.
Picker, G. (2015), Sedentarizzazione e diritto al nomadismo: la genesi dei campi nomadi in Italia, *Historia Magistra. Rivista di storia critica*, 18, 2: 73-84.
Picker, G. (2017), *Racial Cities. Governance and the Segregation of Romani People in Urban Europe*, Routledge, London.
Ramadan, M.C.A.H.M. (2025), Sopravvivenza e resistenza di gruppi rom tra antiziganismo e questioni giuridiche, Educazione Interculturale. *Teorie, Ricerche, Pratiche*, 23, 1: 125-137.
Sam Nariman, H., Hadarics, M., Kende, A., Lášticová, B., Poslon, X.D., Popper, M., Minescu, A. *et al.* (2020), Anti-Roma bias (stereotypes, prejudice, behavioral tendencies): A network approach toward attitude strength, *Frontiers in Psychology*, 11: 2071.
Senato della Repubblica (2014), Rapporto conclusivo dell'indagine sulla condizione di Rom, Sinti e Caminanti in Italia, XVI Legislatura, Roma.
Sigona, N. (2002), *Figli del ghetto. Gli italiani, i campi nomadi e l'invenzione degli zingari*, Nonluoghi Libere Edizioni, Civezzano.
Sigona, N. (2005), Locating "the Gypsy problem". The Roma in Italy: Stereotyping, labelling and "nomad camps", *Journal of Ethnic and Migration Studies*, 31, 4: 741-756.
Sigona, N. (2011), The governance of Romani people in Italy: Discourse, policy and practice, *Journal of Modern Italian Studies*, 16, 5: 590-606.
Spinelli, G. (2022). *Rom e sinti. Dieci cose che dovresti sapere*, People, Milano
Strati, F. (2011), *Italy – Promoting Social Inclusion of Roma. A Study of National Policies*, Social Research Study (SRS).
Szabóné Kármán, J. (2018), The Church and the Gypsies, *Studia. Debreceni Teológiai Tanulmányok*, 10, 1-2: 76-84.
Tremlett, A., Messing, V. and Kóczé, A. (2017). Romaphobia and the media: mechanisms of power and the politics of representations. *Identities*, 24, 6: 641-649.
Váradi, L., Szilasi, B., Kende, A., Braverman, J., Simonovits, G. and Simonovits, B. (2023), "Personally, I feel sorry, but professionally, I don't have a choice". Understanding the drivers of anti-Roma discrimination on the rental housing market, *Frontiers in Sociology*, 8: 1223205.
Villano, P., Fontanella, L., Fontanella, S. and Di Donato, M. (2017), Stereotyping Roma people in Italy: IRT models for ambivalent prejudice measurement, *International Journal of Intercultural Relations*, 57: 30-41

ISBN Open Access: 9788835191971

Vitale, T. (2009). Da sempre perseguitati? Effetti di irreversibilità della credenza nella continuità storica dell'antiziganismo. *Zapruder*, (19): 46-61.
Vitale, T. (2011). *Gli stereotipi che ingombrano politiche e rappresentazioni. La condizione giuridica di Rom e Sinti in Italia* (pp. 255-272), FrancoAngeli, Milano.

ISBN Open Access: 9788835191971

Le comunità romanès in Italia dalla stigmatizzazione alle contronarrative negli spazi digitali

di *Stefania Bevilacqua*[*], *Fiore Manzo* [**], *Nadia Bevilacqua*[*], *Alex Cucco*[***], *Melissa Vassallo*[****]

1. Introduzione

Le comunità romanès rappresentano una delle minoranze storicamente più marginalizzate in Europa ed in Italia, soggette a persecuzioni, discriminazioni e narrazioni stereotipizzanti che hanno costruito la loro immagine pubblica in termini spesso negativi e omogeneizzanti. Con l'espansione degli spazi digitali, tali rappresentazioni si sono riprodotte e amplificate, dando vita a forme di discorso d'odio online che influenzano percezioni, relazioni sociali e opportunità concrete per i membri delle comunità. L'ambiente digitale, lungi dall'essere uno spazio neutro, tende infatti a rafforzare polarizzazioni preesistenti, favorendo la circolazione virale di contenuti tossici e agevolando la diffusione di narrazioni semplicistiche che cancellano la complessità delle storie e delle identità romanès.

Analizzare questi fenomeni richiede non solo di comprendere le radici storiche e culturali degli stereotipi, ma anche di osservare come essi si manifestino e si trasformino negli spazi digitali. Se da un lato è fondamentale riconoscere come tali rappresentazioni si siano consolidate nel tempo, dall'altro diventa imprescindibile considerare le nuove forme di ostilità che si sviluppano online e che contribuiscono a rinnovare e amplificare l'antiziganismo. In questo scenario, gli ambienti digitali non sono solo luoghi in cui si produce odio, ma anche spazi potenziali per promuovere narrazioni alterna-

[*] Ricercatore indipendente, stefy050778@gmail.com; nadiabevilacqua08@gmail.com
[**] Dipartimento di Scienza Politiche e Sociali, Università della Calabria, fioremanzo92@gmail.com
[***] Dipartimento di Studi Socio-Economici, Gestionali e Statistici, Università G. d'Annunzio, alex.cucco@unich.it
[****] Dipartimento di Scienze Giuridiche e Sociali, Università G. d'Annunzio, melissa.vassallo@unich.it

ISBN Open Access: 9788835191971

tive e percorsi di contrasto. Diventa quindi centrale la possibilità di sviluppare strumenti in grado di contenere l'odio online e mitigarne la diffusione: le contronarrative e la detossificazione dei contenuti tossici emergono come strategie chiave per trasformare i discorsi ostili, ricostruire rappresentazioni più accurate e creare spazi digitali più inclusivi. Tali approcci consentono non solo di rispondere all'odio, ma anche di promuovere forme attive di empowerment e partecipazione delle comunità stesse.

2. Gruppi rom in Italia: un esempio di comunità etero-costruita[1]

La minoranza romaní[2], comunemente definita attraverso termini quali zingari, nomadi, gitani, gipsy e tsiganes (Piasere, 2009), è da secoli rappresentata mediante un insieme eterogeneo di stereotipi, sia positivi (i "figli del vento", la "zingara ammaliatrice", lo "zingaro artista") sia negativi (lo "zingaro ladro", il delinquente, il bugiardo, il soggetto da civilizzare). Queste narrazioni, consolidate sin dal basso Medioevo (Giuffrè, 2014, pp. 43–82), contribuiscono ancora oggi alla costruzione sociale di un'immagine fortemente omogeneizzante dei gruppi rom (Giuffrè, 2014, p. 31). Gli stereotipi necessitano pertanto di essere decostruiti, poiché si manifestano attraverso forme di antiziganismo[3], che influenzano concretamente la vita delle persone appartenenti alla minoranza (Piasere, 2012; 2015).

I gruppi rom rappresentano un caso emblematico di comunità etero-costruita, la cui rappresentazione pubblica è stata definita per lungo tempo da narrazioni maggioritarie stereotipate. Solo recentemente, anche in Italia, sono emerse forme di autorappresentazione più strutturate, capaci di mettere in discussione tali immagini semplificate. La comparsa delle prime pubblicazioni letterarie di autori rom italiani è un fenomeno relativamente recente, legato anche al fatto che la comunità presentava storicamente tassi di analfabetismo più elevati rispetto a oggi, fattore che ha limitato a lungo la produzione scritta (Rizzin e Bravi, 2024).

Il primo libro è attribuito al musicista e docente universitario Santino Spinelli, che nel 1994 pubblica un'opera sulle tradizioni dei rom abruzzesi (Spi-

[1] Si utilizzerà il termine rom in riferimento alle comunità rom di antico insediamento o di recente migrazione, e la dizione comunità romanès o gruppi rom come termine ombrello che include, in generale, tutti gli etnonimi con cui tali comunità si autodefiniscono.

[2] Per una ricostruzione storica dei rom in Italia si veda Spinelli (2016; 2018); Pontrandolfo (2013); Piasere (2002; 2016); Saletti Salza, Leonardo Piasere (2004e).

[3] Con il designante antiziganismo s'intende un «fenomeno sociale, psicologico, culturale e storico che vede in quelli che individua come "zingari" un oggetto di pregiudizi e stereotipi negativi, di discriminazione, di violenza diretta o di violenza indiretta» (Piasere 2015, p. 11).

ISBN Open Access: 9788835191971

nelli, 1994). In seguito, Spinelli amplia il panorama editoriale italiano con diverse opere sulla storia romaní, introducendo studi di accademici rom provenienti da varie aree del mondo. A distanza di pochi anni, anche Bruno Morelli dedica un volume alle tradizioni dei rom abruzzesi (Morelli, 1997), mentre Nazzareno Guarnieri pubblica prima un libro sulla cultura romaní e sugli stereotipi e poi uno sui pregiudizi e la mediazione culturale (Guarnieri, 2000, 2023). Nel 1998, per Fatatrac, esce *Strada, patria sinta. Cento anni di storia nel racconto di un saltimbanco* di Gnugo De Bar (De Bar, 1998).

Un ulteriore contributo fondamentale è quello di Giorgio Bezzecchi, rom harvato residente a Milano, che nel 2004 pubblica un testo sul Porrajmos o samudaripen[4] (Bezzecchi, 2004). Molti di questi autori possono essere considerati tra i primi attivisti rom italiani. Tra le opere più recenti si ricordano il lavoro di Eva Rizzin sul samudaripen e la ricostruzione delle classi speciali lacio drom (Rizzin, 2020, 2024). Nel 2022 vengono pubblicati due volumi: uno di Spinelli (2022), dedicato alla decostruzione di dieci stereotipi sui rom e sui sinti, e uno di Manzo (2022) sulla ghettizzazione di una parte della comunità romaní di Cosenza. Nel 2023 Aldo De Ragna pubblica *Vite in cammino Storia di una famiglia rom di Milano* (De Ragna, 2023), mentre nel 2025 l'attrice e attivista Dijana Pavlovic firma *Irriducibili. Alterità nell'anima zingara* (Pavlovic, 2025).

Dal punto di vista storico, gli antenati delle comunità romanès presenti in Italia discendono da gruppi emigrati dall'India nord-occidentale, che attraversarono la Persia, l'Armenia e successivamente la Grecia. L'avanzata ottomana costrinse molte comunità, incluse quelle rom e quelle Arbëreshë, alla fuga. Le fonti attestano la presenza di gruppi rom in Italia almeno dal 1422: il 18 luglio di quell'anno un gruppo di circa cento persone, guidato da Duca Andrea, giunge nei pressi di Bologna.

A questo lungo percorso storico si affianca l'importante lavoro di Campigotto, Aresu, Bianchetti e Piasere (2020), nel quale si raccolgono testi di filosofia, diritto, teologia, politica e scienze pubblicati tra il 1422 e il 1812. I testi raccolti e discussi in Aresu, Bianchetti e Piasere (2020), come riconosciuto dagli autori, hanno fortemente contribuito alla costruzione e al consolidamento degli stereotipi sui gruppi rom, molti dei quali ancora attivi oggi (ladri, vagabondi, marginali, irreligiosi) (Campigotto *et al.*, 2020, p. 419).

Nell'Italia contemporanea, l'etnia romaní risulta la più stigmatizzata (Faloppa, 2011, p. 92). L'ostilità nei confronti dei gruppi rom si manifesta nella quotidianità e negli spazi digitali (Pasta 2018), e trova eco nei discorsi poli-

[4] Il termine samudaripen, che in lingua romanès significa "tutti morti" e che fu introdotto dal linguista Marcel Courthiade, designa lo sterminio di oltre 500.000 rom e sinti durante la Seconda guerra mondiale. Per approfondimenti si vedano Bravi (2013), Fings (2018), Trevisan (2024).

ISBN Open Access: 9788835191971

tici (Rizzin e Pontrandolfo, 2020; 2021), spesso diffusi attraverso i social media, contribuendo ad amplificare stereotipi radicati. La visione dominante delle comunità romanès rimane polarizzata: da un lato vengono considerate una minaccia sociale, associate a criminalità, sporcizia, immoralità e al pregiudizio dei "rapitori di bambini" (Fondazione Romaní Italia, 2014); dall'altro lato emergono narrazioni più romantiche che le rappresentano come gli "ultimi uomini liberi" (Manzo, Cosentino, 2025). L'analisi di Giovanni Agresti (2018) sui termini "zingar*" nella stampa conferma la persistenza di una rappresentazione stigmatizzante.

Alcune indagini recenti confermano un elevato livello di pregiudizio: l'83% degli italiani dichiara un'opinione sfavorevole verso le comunità romanès (Pew Research Center, 2019); in Calabria risultano percepiti come il gruppo più pericoloso (Elia e Fantozzi, 2016); l'84% degli italiani sostiene l'idea del "nomadismo culturale" dei rom, mentre il 47% esprime un'opinione esplicitamente negativa (Faloppa, 2011, p. 93). Le persone rom vengono associate a mancanza di calore, educazione, onestà e vengono percepite attraverso categorie legate alla delinquenza o alla furbizia (Meneghini, 2017). Tali rappresentazioni generano emozioni negative, in particolare paura e senso di minaccia (Albarello e Rubini, 2011, p. 360), nonostante la maggioranza della popolazione conosca poco la realtà rom (Faloppa, 2011). Il pregiudizio risulta trasversale a fasce d'età diverse e si trasmette ai bambini attraverso i processi educativi familiari. Il rom viene così sottoposto a processi di disumanizzazione (Giuffrè, 2014), spesso tramite l'uso di aggettivi riferiti alla sfera animale (Meneghini, 2017, p. 13). La sua presenza fisica viene percepita come minacciosa e affrontata attraverso strategie segregative, come i campi nomadi spesso localizzati nelle periferie urbane.

3. Antiziganismo digitale: manifestazioni, rischi e cornice europea

Negli ultimi anni, la comunità dei rom e dei sinti, storicamente soggetta a persecuzioni, discriminazioni e marginalizzazione, si confronta con una crescente ondata di odio e stereotipi anche nello spazio digitale. Commenti denigratori, fake news, discorsi d'odio e rappresentazioni caricaturali diffusi sui social media alimentano pregiudizi, esclusione e forme di minaccia reale. Questa sezione intende offrire una panoramica dei principali studi e report europei che documentano il fenomeno, analizzandone le caratteristiche e discutendo l'impatto dell'"antigypsyism digitale" sulle comunità romanès e sulla società.

I termini antigypsyism, romofobia e sintifobia (Corradi, 2018) descrivono forme specifiche di razzismo e discriminazione rivolte verso rom, sinti e

ISBN Open Access: 9788835191971

persone percepite come "zingare". Il termine "antiziganismo", riconosciuto in ambito accademico, indica il quadro più ampio di stereotipi, stigmatizzazione, esclusione, discriminazione istituzionale, rappresentazioni negative e violenza. Nello spazio online, tali dinamiche si traducono in contenuti d'odio, commenti denigratori, narrazioni tossiche, incitamento alla violenza e diffusione di stereotipi reiterati, osservabili nei commenti dei social media, nei forum, in gruppi pubblici o privati.

A livello europeo, i discorsi contro rom e sinti risultano tra i più frequenti nel contesto dell'hate speech online. L'European Roma Rights Centre (ERRC), attraverso il progetto Roma Rights Defenders (Albania, Serbia, Turchia, Ucraina, 2020-2021), ha realizzato la prima raccolta "data-driven" di casi di discorsi d'odio antirom. I contenuti monitorati spaziano dalla "commedia stereotipata" fino ad appelli espliciti alla violenza. Il report raccomanda una definizione legislativa chiara di hate speech, sanzioni adeguate, maggiore collaborazione istituzionale e un ruolo attivo delle piattaforme tecnologiche nella moderazione.

La European Union Agency for Fundamental Rights (FRA), nel rapporto del 2023 dedicato alla moderazione dei contenuti online, identifica rom e sinti tra i gruppi più frequentemente colpiti da discorsi d'odio, insieme a donne, persone di origine africana e gruppi religiosi. La FRA evidenzia criticità strutturali nella moderazione: errori umani, limiti degli algoritmi, sistemi di segnalazione inefficaci, che rendono l'odio persistente e spesso impunito.

Un'analisi transnazionale condotta in dieci Paesi europei mostra come le narrazioni d'odio antigypsyist rievocano stereotipi storici: i rom vengono rappresentati come "nomadi", "non integrabili", "criminali", "parassiti sociali", nonostante la realtà – fatta di persone che vivono stabilmente, lavorano e partecipano alla vita sociale – contraddica tali immagini. Secondo la FRA, una persona rom/traveller su quattro denuncia discriminazioni in almeno uno dei seguenti ambiti: istruzione, lavoro, sanità, abitazione, rapporti con lo Stato.

In Italia, l'Osservatorio Nazionale sull'Antiziganismo (ONA) svolge un ruolo essenziale nella documentazione dei casi di intolleranza e discriminazione verso rom e sinti, attraverso la raccolta sistematica di articoli, segnalazioni e materiali relativi al pregiudizio antizingaro. La costruzione del primo archivio dell'Osservatorio rappresenta un passo rilevante per monitorare e analizzare i discorsi d'odio promossi da media, attori politico-istituzionali e altri soggetti. L'odio online risulta particolarmente pericoloso per diverse ragioni. Innanzitutto, la ripetizione continua di stereotipi e insulti contribuisce alla normalizzazione dell'ostilità: ciò rende socialmente accettabile un linguaggio discriminatorio e favorisce la sedimentazione di pregiudizi duraturi.

ISBN Open Access: 9788835191971

A questo si aggiungono le conseguenze concrete sulla vita delle persone, poiché l'hate speech genera insicurezza, esclusione e limita opportunità sociali, lavorative ed educative, amplificando discriminazioni già esistenti. Un ulteriore elemento critico riguarda l'impunibilità e la scarsa moderazione: la mancanza di interventi efficaci da parte delle piattaforme digitali, unita alla sfiducia delle vittime nel segnalare gli abusi, rende il fenomeno in larga parte invisibile e difficilmente controllabile. L'odio digitale contribuisce anche alla riproduzione di stereotipi storici, perpetuando narrazioni sedimentate nei secoli, come quelle sul nomadismo, sulla criminalità o sulla devianza, con effetti negativi anche sulle nuove generazioni rom e sinte.

Per quanto riguarda le azioni di contrasto, risulta fondamentale rafforzare la moderazione dei contenuti, introducendo procedure più rapide, trasparenti ed efficaci. Parallelamente, è necessario investire in campagne educative e iniziative di sensibilizzazione che valorizzino contronarrazioni positive e realistiche, capaci di scardinare stereotipi consolidati. Un ruolo essenziale andrebbe inoltre attribuito alla rappresentanza attiva delle comunità rom e sinte nei processi decisionali e nei progetti di monitoraggio e advocacy, affinché le politiche risultino realmente inclusive. Infine, occorre potenziare la raccolta dati e sostenere lo sviluppo di osservatori nazionali e internazionali sull'antiziganismo, seguendo l'esempio dell'Osservatorio Nazionale sull'Antiziganismo (ONA), così da garantire strumenti più solidi per l'analisi, la prevenzione e la tutela dei diritti.

3.1. Ruolo della comunità online e promozione di contronarrative

Le modalità di rappresentazione e autorappresentazione delle comunità nello spazio digitale possono costituire ad oggi un nodo analitico cruciale nel comprendere come l'odio online non sia un semplice fenomeno comunicativo, ma un dispositivo socioculturale che riproduce e rielabora gerarchie preesistenti. Le interazioni mediate dalle piattaforme, pertanto, non operano in una situazione di vuoto sociale, ma si innestano su immaginari storicamente sedimentati su relazioni di potere che contribuiscono a definire chi può parlare, con quale legittimità e in quali termini.

Allo stesso tempo, una parte della letteratura sulle minoranze etniche e culturali evidenzia come lo spazio digitale diventi un luogo di possibilità, in cui gruppi altamente stigmatizzati e vittime di discriminazione, promuovo contronarrative capaci di sfidare la narrazione dominante (Correa e Jeong, 2011; Nakamura, 2008; Topidi e Metcalfe, 2024).

Le piattaforme digitali come i social media, quali Facebook, Instagram, TikTok e quant'altro diventano dunque arene nelle quali tali comunità cerca-

ISBN Open Access: 9788835191971

no di proporre una propria narrazione, riappropriandosi dell'agency a lungo negata. Queste dinamiche introducono non solo strategie di autorappresentazione, ma anche la condivisione di simboli culturali, intesi come elementi attraverso cui le comunità definiscono, negoziano e difendono la propria identità (Somers, 1994).

Le contronarrative assumono dunque un valore politico e relazionale, esprimendo forme di resistenza che mirano a sovvertire rappresentazioni ostili e stigmatizzanti attraverso l'uso di strumenti propri del discorso online (Bamberg, 2008; McKenzie-Mohr e Lafrance, 2017). La capacità delle comunità di attivare tali risorse è rafforzata dalla natura interconnessa degli spazi digitali moderni, che facilita la circolazione di tematiche condivise e strategie retoriche anche a livello transnazionale.

Nel caso delle comunità romanès, queste dinamiche assumono una rilevanza particolare. Come analizzato in precedenza, un'ampia porzione della letteratura esistente ha rilevato come l'esposizione ai discorsi d'odio alimenti e allo stesso tempo sia alimentata da stereotipi e forme di pregiudizio radicate, rilevato anche nel contesto italiano (Buturoiu e Corbu, 2020; Villano *et al.*, 2017).

In egual misura, si dovrebbe però cercare di rivolgere lo sguardo alle iniziative, portate avanti dalla stessa comunità, di autorappresentazione e contronarrazione come strumenti efficaci nella lotta al fenomeno dell'antiziganismo (Mišković *et al.*, 2020; Sabiescu, 2005). Gli spazi digitali si configurano come laboratori di immaginazione collettiva, strumenti di coordinamento orizzontale nei quali le comunità rom/sinti Rom, prese in esame, non si configurano soltanto come oggetti del discorso ma come soggetti produttori di significati.

4. Contronarrative e detossificazione

Le contronarrative si distinguono da altre forme di intertestualità poiché prevedono l'assunzione di una posizione rispetto ad altri discorsi (Lundholt *et al.*, 2018) pur non implicando necessariamente un rapporto oppositivo. Come osservano Bamberg e Andrews (2004), le contronarrative rafforzano il loro significato solo se messe in relazione a ciò che contestano, andandosi a identificare come categoria "posizionale" in tensione con un'altra categoria. Definite da Andrews (2004) come strategie implicite o esplicite, che si sviluppano tramite il racconto di storie e vissuti, esse possono agire come meccanismi di resistenza a narrazioni culturali dominanti. Le contronarrative, pertanto, operano tipicamente in opposizione o resistenza rispetto alle "*master narratives*", narrazioni socialmente e culturalmente condivise che

ISBN Open Access: 9788835191971

fanno spesso riferimento a categorie quali genere, sessualità, etnia, classe, professione, e che possono risultare normative, oppressive o escludenti rispetto a prospettive ed esperienze divergenti. La contronarrativa, inoltre, non vuole essere un resoconto esaustivo ma rappresenta una costruzione di senso costituita da esperienze ed elementi intenzionalmente ed accuratamente selezionati da chi la propone, ritenute da Nelson (2001) come risultato di una selezione moralmente connotata e guidata, operata al fine di porsi in contrasto a specifiche narrazioni dominanti. Ciò definisce un ulteriore elemento caratterizzante delle contronarrative inerente alla selezione delle tematiche e dei simboli eseguita al momento della costruzione della stessa.

Nell'ambito dei social media, le contronarrative sono, negli anni, emerse come strumento efficace per contrastare il fenomeno dell'hate speech online. Tale fenomeno, risulta un campo in forte espansione che ha attirato l'attenzione tanto delle scienze sociali quanto di quelle statistiche e computazionali. La ricerca in quest'ultimo ambito si è ampiamente concentrata sul riconoscimento automatico di linguaggio tossico e di altre forme di comunicazione problematiche, comprese microaggressioni e registri di carattere paternalistico (Zampieri *et al.*, 2019; Breitfeller *et al.*, 2019; Perez Almendros *et al.*, 2020) o nell'identificazione automatica di profili potenzialmente attivi nella diffusione d'odio (del Gobbo *et al.*, 2025). Tuttavia, come osservano Logacheva *et al.* (2022), la mera identificazione di messaggi dannosi, in tal senso, non offre modalità proattive per contrastarli oltre alla cancellazione degli stessi.

Da questo retroterra, emerge il concetto di detossificazione "*detoxification*": ovvero la riscrittura (automatica in ambito computazionale) di testo considerato tossico nell'ottica di preservare il contenuto eliminando il contenuto tossico. La detossificazione viene generalmente inquadrata come una particolare applicazione delle tecniche di "*style transfer*", che mirano a riformulare un testo mantenendone il contenuto ma modificandone la dimensione stilistica, intesa, ad esempio, in termini di tono emotivo, grado di formalità o registro (Logacheva *et al.*, 2022). Tale prospettiva è stata sviluppata da diversi contributi nel campo del Natural Language Processing, i quali hanno proposto modelli capaci di trasformare messaggi tossici in versioni non offensive preservandone l'informazione essenziale (Nogueira dos Santos *et al.*, 2018; Minh Tran *et al.*, 2020), aprendo una direzione alternativa rispetto alle strategie basate unicamente sulla rimozione o sul filtraggio del contenuto.

L'impiego di modelli di riscrittura automatica dei testi, inclusi quelli precedentemente illustrati, sta aprendo nuove possibilità per il contrasto al fenomeno dell'hate speech negli spazi digitali. Tuttavia, la loro applicazione non è priva di criticità: l'operazione di riscrittura rischia infatti di alterare o appiattire le sfumature del discorso date da gruppi sociali marginalizzati, producendo interpretazioni parziali o distorte. Per evitare che tali strumenti

ISBN Open Access: 9788835191971

riproducano forme di esclusione sistemica, è dunque essenziale garantire anche, e fin dalle fasi preliminari, (es. dalla costruzione dei dataset alla definizione degli obiettivi di training, fino alla valutazione) il coinvolgimento delle comunità direttamente coinvolte. Una progettazione partecipata può infatti favorire maggiore sensibilità culturale, aumentare la trasparenza del processo e garantire risultati più equi e rappresentativi

4.1. Strumenti deontologici e progetti partecipativi per il contrasto all'hate speech

Parallelamente allo sviluppo di tecniche automatiche di detossificazione e sviluppo di contronarrative, negli ultimi anni si è consolidata una mobilitazione a più livelli, interna alle comunità romanès, istituzionale e transnazionale, volta a contrastare l'hate speech e l'antiziganismo e a promuovere forme di comunicazione più eque e rispettose. Tali iniziative rivelano una visione integrata del problema: la produzione di contronarrative efficaci non può dunque essere delegata unicamente allo sviluppo tecnologico, ma richiede un lavoro culturale, di informazione e partecipazione capace di incidere sulle condizioni strutturali che alimentano stigmatizzazione e discriminazione.

Un primo ambito di intervento è rappresentato dagli strumenti deontologici rivolti ai professionisti dell'informazione. Documenti quali "Guidelines for Media Coverage of Roma Community – A Practical Guide for Journalists ed Equal Treatment, the Media and Roma Community" propongono criteri per evitare essenzializzazioni, errori terminologici e narrazioni sensazionalistiche, richiamando la responsabilità dei media nella riproduzione di stereotipi o nell'amplificazione della discriminazione. Nel contesto italiano, il "Glossario dell'Associazione Carta di Roma" dedicato alle terminologie relative alle comunità rom e sinte costituisce un ulteriore strumento di orientamento linguistico e culturale, utile tanto alla stampa quanto agli utenti attivi nelle piattaforme digitali per favorire una comunicazione più accurata e rispettosa.

Accanto ai dispositivi normativi e formativi, si colloca un ampio insieme di progetti finanziati a livello europeo, che affrontano l'hate speech tenendo conto di prospettive complementari. Il progetto "Freedom from Hate", sostenuto dal programma Rights, Equality and Citizenship dell'Unione Europea, ha esaminato l'impatto di cinque campagne digitali contro l'odio online in Bulgaria, Croazia, Repubblica Ceca, Ungheria e Slovacchia, evidenziando l'efficacia di narrazioni positive e contestualizzate nella riduzione dell'ostilità pubblica. In Italia, il progetto "C.O.N.T.R.O.", sviluppato in collaborazione con UNAR, ha invece mostrato come le contronarrative possano servire da contrasto all'hate speech senza limitare la libertà di espressione,

ISBN Open Access: 9788835191971

mettendo in luce le motivazioni psicologiche e sociali che spingono alcuni utenti a produrre contenuti discriminatori e delineando metodologie comunicative capaci di ridurre conflitto e polarizzazione.

A livello europeo, ulteriori iniziative evidenziano la crescente attenzione verso approcci interdisciplinari e partecipativi. Il progetto portoghese "KNOWHATE" integra approcci derivanti dalle scienze sociali, dalla linguistica e dall'informatica per sviluppare strumenti di rilevazione di contenuti tossici sensibili ai contesti culturali, dialogando al tempo stesso con le comunità interessate. A esso si affianca l'esperienza dei Roma Cultural Influencers, che attraverso la promozione di un percorso di formazione si è rivolto a giovani membri della comunità rom ungherese per promuovere la creazione di contenuti digitali, trasformando la comunicazione online in un luogo di riaffermazione identitaria e autodeterminazione culturale. In parallelo, i programmi finanziati nell'ambito del CERV – Citizens, Equality, Rights and Values Programme hanno promosso la costruzione di ambienti digitali più sicuri, sostenendo la diffusione di nuove forme di advocacy e la produzione di contenuti provenienti dalle comunità stesse.

Nel loro insieme, queste iniziative convergono su alcuni assunti chiave, ovvero, la necessità di coinvolgere direttamente le popolazioni rom nella progettazione delle strategie comunicative; l'importanza di un approccio preventivo che combini educazione, democratizzazione e strumenti tecnologici avanzati; la valorizzazione di figure interne alla comunità, attivisti, influencer, mediatori, come portavoce credibili; e l'attenzione alla gestione delle reazioni ostili, considerata parte integrante di una strategia di contronarrativa.

Nel loro insieme, politiche culturali, iniziative comunitarie e innovazione tecnologica indicano con chiarezza che il contrasto all'antiziganismo non può essere affidato a interventi isolati, ma richiede un ecosistema collaborativo e multilivello. Solo la sinergia tra istituzioni, media, piattaforme digitali e le comunità direttamente interessate permette infatti di generare contronarrative solide, autorevoli e capaci di incidere realmente sulla persistenza dei discorsi d'odio online.

In questo quadro, il coinvolgimento diretto delle comunità romanès non rappresenta soltanto una scelta etica o partecipativa, ma costituisce una condizione imprescindibile per assicurare efficacia, autenticità e sensibilità culturale agli interventi pensati per il contrasto di tale fenomeno. La progettazione partecipata consente infatti di evitare distorsioni interpretative, aumentare la trasparenza dei processi e produrre strumenti narrativi che rispecchino le esperienze e le priorità dei gruppi coinvolti.

ISBN Open Access: 9788835191971

5. Produzione di contronarrative e contenuti detossificati: costruzione di risorse lessicali

5.1. Descrizione del task

Per procedere con la formulazione delle contronarrative e l'elaborazione dei contenuti detossificati, sono stati selezionati manualmente 450 commenti a partire da un totale di 15.129 testi estratti dalla piattaforma Facebook tramite exportcomments.com. La selezione di questo sottoinsieme è stata effettuata con l'obiettivo di ridurre il più possibile la ridondanza dei contenuti. La selezione si è focalizzata sull'individuazione di commenti offensivi, per cui le tasks di detossificazione e contronarrative presentano il maggiore potenziale applicativo. In Figura 1 vengono riportate le 50 parole più usate nel corpus selezionato privato delle stopwords da cui emerge chiaramente la centralità ricoperta da parole quali "rom" e "zingari".

Fig. 1 – Wordcloud: la dimensione delle parole cresce al crescere della loro frequenza

ISBN Open Access: 9788835191971

Per ottenere una panoramica delle principali configurazioni lessicali ricorrenti nei commenti raccolti, è stato applicato un modello Latent Dirichlet Allocation (LDA) (Blei 2003), una tecnica di topic modeling ampiamente utilizzata per individuare strutture tematiche latenti all'interno di collezioni testuali. LDA assume che ciascun documento (in questo caso, ogni commento) sia generato come combinazione di uno o più temi e che ogni tema sia caratterizzato da un insieme di parole con differenti probabilità di occorrenza. Tale approccio risulta particolarmente utile quando si vuole esplorare in maniera non supervisionata la distribuzione degli argomenti trattati senza imporre categorie predefinite. Nel nostro caso, al fine di cogliere le principali linee discorsive presenti nel corpus e mantenere un livello di granularità adeguato, è stato impostato un modello con tre topic. Questa scelta consente di sintetizzare le macroaree tematiche evitando al contempo una frammentazione eccessiva, che sarebbe poco utile ai fini interpretativi. I tre topic risultanti sono stati successivamente interpretati sulla base delle 30 parole più rilevanti per ciascuno. L'analisi dei tre topic tramite la valutazione delle parole maggiormente rappresentative, mostra tre dimensioni complementari del discorso d'odio verso le comunità romanès (Fig. 2). Il primo topic è dominato dalla contrapposizione tra "italiani" e "rom", incentrata su temi come campi, casa e tasse, insieme a riferimenti a "razza" ed etnicità. Questo insieme di termini indica una narrativa fortemente ostile che costruisce i rom come gruppo esterno e percepito come economicamente parassitario o non integrato, con una forte struttura "noi contro loro". Il secondo topic riguarda invece la dimensione identitaria e culturale dello scontro: compaiono termini come "zingari", "nomadi", "etnia", insieme a riferimenti al razzismo e a giudizi morali. Qui il discorso oltre ad essere marcatamente insultante, riflette un livello più discorsivo, in cui si negoziano e si contestano categorie identitarie, attribuzioni culturali e accuse reciproche di razzismo. Il terzo topic si concentra su narrazioni episodiche e su un linguaggio fortemente aggressivo: emergono riferimenti ad azioni come rubare o fare elemosina, termini volgari e richiami a situazioni quotidiane ("visto", "giorno", "quando"), che funzionano come presunte prove a sostegno di generalizzazioni ostili. Presi insieme, i topic descrivono un repertorio di discorsi d'odio: un livello strutturale legato a conflitti socioeconomici, uno identitario basato su stereotipi culturali e uno aneddotico legato a narrazioni di sospetto e criminalizzazione quotidiana.

ISBN Open Access: 9788835191971

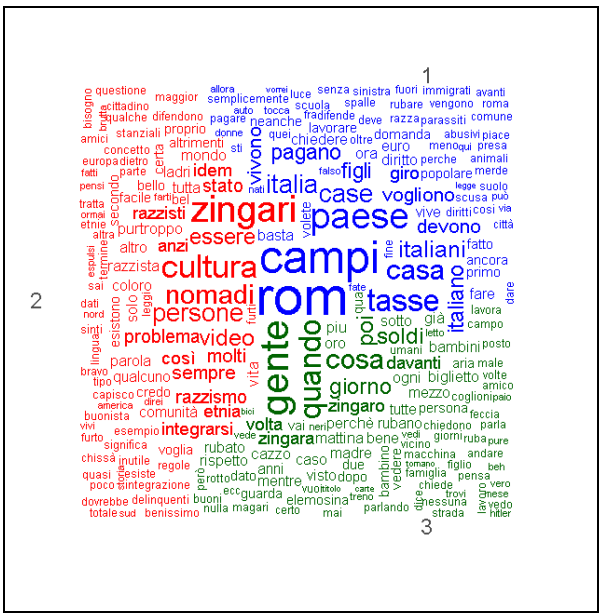


Fig. 2 – Rappresentazione dei tre topic mediante le parole maggiormente rappresentative, mostrando tre dimensioni del discorso d'odio rivolto alle comunità romanès

Nella fase di annotazione sono stati coinvolti tre annotatori appartenenti alle comunità romanès italiane (co-autori di questo contributo), che hanno operato in modo indipendente. I commenti destinati alle fasi di annotazione sono stati attribuiti in modo casuale dal corpus. Per verificare che tale selezione casuale non avesse introdotto associazioni tra la distribuzione dei cluster lessicali e l'assegnazione degli annotatori, è stato condotto un test del chi-quadro di indipendenza. Il test non ha evidenziato associazioni statisticamente significative (p-value: 0.467), confermando che i tre annotatori hanno ricevuto commenti distribuiti in modo equilibrato rispetto ai cluster individuati dal modello. Inoltre, per verificare che la selezione casuale non generasse aggregazioni semantiche involontarie, è stata costruita una rete semantica basata sulla cosine similarity tra le rappresentazioni vettoriali dei commenti ottenuta tramite "*distiluse-base-multilingual-cased-v2*"[5]. Data la rete rappresentata in Figura 3 è stato calcolato il coefficiente di assortatività rispetto agli annotatori, con l'obiettivo di misurare se i commenti assegnati allo stesso annotatore

[5] Reimers, N., Gurevych, I.: Sentence-BERT: Sentence embeddings using Siamese BERT-Networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, pp. 3982–3992. Association for Computational Linguistics (2019). doi: 10.48550/arXiv.1908.10084

ISBN Open Access: 9788835191971

tendessero a essere più simili tra loro rispetto a quelli assegnati ad annotatori diversi. Il valore di assortatività ottenuto è risultato prossimo allo zero, indicando assenza di strutture assortative (-0.002).

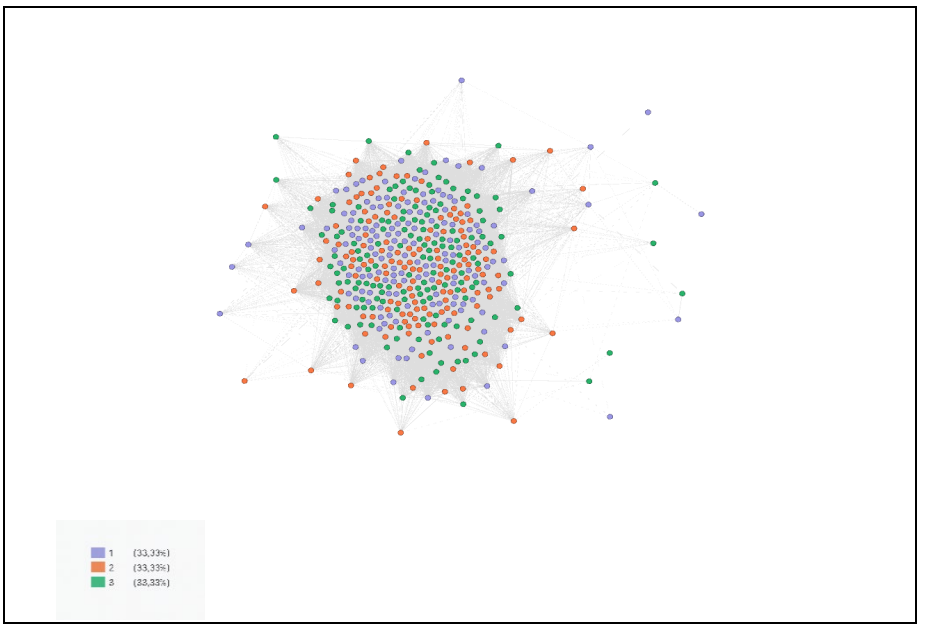


Fig. 3 - Visualizzazione della rete semantica dei commenti, con nodi colorati secondo l'annotatore

4.2. Analisi qualitativa preliminare

L'analisi qualitativa condotta durante la fase di annotazione ha messo in evidenza alcuni aspetti rilevanti del lavoro sulle controretoriche. Un primo risultato riguarda la presenza di commenti per i quali gli annotatori hanno giudicato impossibile sia la detossificazione sia la formulazione di una controretorica. Si tratta di enunciati composti quasi esclusivamente da insulti etnici o da sequenze di aggressioni verbali prive di contenuto inferenziale recuperabile. In altri casi, la struttura stessa del commento rendeva impossibile estrarre un nucleo semantico minimamente riformulabile. L'impossibilità di intervenire su tali segmenti è emersa come un risultato in sé: questi casi rappresentano forme di forte tossicità, non suscettibili di trasformazione. Esempi in cui non è stato possibile proporre una contronarrazione sono casi come:

Certo che i rom sono proprio la loro stessa disgrazia.

ISBN Open Access: 9788835191971

Ai rom piace l'oro, a noi non piacciono loro. Sottili differenze.

Per quanto concerne la detossificazione invece, si è ritenuto impossibile procedere in alcuni casi tra i quali:

Come si chiamano? ZINGARI e nn ce n'è uno buono. Ti piacciono? Trasferisciti in campo nomadi (che poi nomadi nn sono…) e poi facci un bel video... ma occhio che ti rubano anche le mutande che hai addosso oltre alla videocamera.

Voi non siete Italiani, siete e sarete per sempre Rom (per non dire altro), non vi vuole nessuno per questo siete nomadi. Ricordatevi che se un cane nasce in una stalla non può dire di essere un cavallo! Siete solo uno spreco di ossigeno....

Parallelamente, il lavoro di annotazione ha permesso di osservare un ventaglio di strategie adottate dagli annotatori nei casi in cui la costruzione di una controretorica è risultata praticabile. Una prima tipologia di intervento si manifesta quando il commento originale presentava un presupposto errato o una generalizzazione implicita: in questi casi la contronarrazione fornisce chiarificazioni fattuali, evidenzia la complessità dei fenomeni sociali coinvolti o sostituisce l'accusa generalizzata con un riferimento a dinamiche più articolate. Una seconda strategia ricorrente consiste nella ricontestualizzazione empatica: di fronte a giudizi assoluti privi di basi verificabili, gli annotatori propongono alternative interpretative che inseriscono i comportamenti citati in un quadro sociale, economico o storico più ampio. In altri casi, infine, la controretorica non interveniva sul contenuto fattuale, ma sulla cornice valoriale, trasformando il commento in un'occasione per riaffermare principi di equità, diritti e convivenza civile.

6. Discussione

L'insieme delle analisi presentate mette in luce come le configurazioni dell'odio online rivolto alle comunità romanès si articolino lungo assi discorsivi coerenti con i processi storici di etero-costruzione, stigmatizzazione e marginalizzazione già evidenziati nella letteratura. La persistenza di categorie come "zingari", "nomadi" e, più in generale, di una rappresentazione dei rom come corpo estraneo rispetto alla cittadinanza nazionale, trova infatti corrispondenza tanto nelle forme di razzializzazione di lungo periodo quanto nelle narrazioni più contemporanee che legano la presenza delle comunità romanès a temi di devianza, assistenzialismo e conflitto socioeconomico.

I risultati del topic modeling confermano la sovrapposizione tra livelli strutturali, identitari e narrativi del discorso ostile: da un lato emergono con-

ISBN Open Access: 9788835191971

trapposizioni dicotomiche che ricalcano l'opposizione storica "noi/loro", dall'altro compaiono giudizi essenzializzanti che attribuiscono ai rom tratti culturali presunti e immutabili, insieme a una vasta gamma di aneddoti che vengono utilizzati come prove di un pregiudizio generalizzato. Questa stratificazione del discorso riproduce dinamiche sociali note, nelle quali elementi economici, simbolici e morali si intrecciano nel consolidare una rappresentazione omogeneizzante.

Il processo di annotazione ha inoltre mostrato come tali strutture discorsive non si limitino a contenuti esplicitamente offensivi, ma includano forme di tossicità radicate in presupposti impliciti, generalizzazioni e inferenze non esplicitate. La presenza di commenti per i quali risulta impossibile formulare una controNarrazione o operare una detossificazione segnala un livello di ostilità talmente elevato da annullare qualsiasi spazio di negoziazione semantica: si tratta di enunciati privi di argomentazione, ridotti a insulti etnici o a costruzioni linguistiche che non consentono l'individuazione di un contenuto dialogico. Questi casi funzionano come indicatori della forma più rigida e impermeabile del discorso d'odio.

Allo stesso tempo, l'analisi qualitativa degli interventi degli annotatori evidenzia come, laddove sia presente un minimo margine interpretativo, esistano diversi modi di disinnescare le premesse tossiche o di proporre letture alternative. Le strategie emerse, dalla correzione dei presupposti fattuali alla ricontestualizzazione socio-storica, fino al richiamo a cornici valoriali comuni, mostrano che non tutti i contenuti ostili operano sullo stesso piano e che alcuni di essi, pur problematici, mantengono una struttura semantica che permette un intervento trasformativo. Questa osservazione è particolarmente rilevante in quanto riflette la pluralità delle forme discorsive che caratterizzano l'odio online e suggerisce che i processi di contrasto non possono essere univoci ma necessitano di strategie diversificate.

Il controllo sulla distribuzione dei commenti tra annotatori e la verifica dell'assenza di strutture assortative nella rete semantica confermano inoltre la robustezza del processo di campionamento e l'assenza di bias imputabili alla fase di attribuzione dei testi.

Nel loro complesso, i risultati fin qui discussi offrono un quadro delle forme di ostilità rivolte alle comunità romanès nello spazio pubblico digitale, evidenziando continuità storiche, specificità contemporanee e margini variabili di trasformazione discorsiva.

ISBN Open Access: 9788835191971

Riferimenti bibliografici

Albarello, F. and Rubini, M. (2011). Outgroup projection: Il caso degli stereotipi negativi attribuiti a Rom e Rumeni, *Psicologia Sociale*, 6(3): 355–365

Andrews, M. (2004), Counter-narratives and the power to oppose. In Bamberg M. and Andrews M., eds., *Considering counter-narratives: Narrating, resisting, making sense* (pp. 1–6). John Benjamins, Amsterdam.

Bamberg, M. (2008), Considering counter narratives. In *Considering counter-narratives* (pp. 351-371), John Benjamins Publishing Company, Amsterdam.

Bamberg, M. and Andrews, M., eds. (2004), *Considering counter-narratives. Narrating, resisting, making sense*. John Benjamins, Amsterdam.

Bezzecchi, G. (2004), *Il Porrajmos dimenticato. Le persecuzioni di Rom e Sinti in Europa*, Edizione Opera Nomadi, Milano.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3, 993-1022.

Bravi, L. and Bassoli, M. (2013). *Il Porrajmos in Italia: La persecuzione di rom e sinti durante il fascismo*, Emil di Odoya, Bologna.

Breitfeller, L., Ahn, E., Jurgens, D. and Tsvetkov, Y. (2019), Finding microaggressions in the wild: A case for locating elusive phenomena in social media posts. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 1664-1674).

Buturoiu, D.R. and Corbu, N. (2020), Exposure to Hate Speech in the Digital Age. Effects on Stereotypes About Roma People, *Journal of Media Research*, 13(2).

Campigotto, A., Aresu, M., Bianchetti, P. and Piasere, L. (2020), *Questo genere di uomini. Testi su egiziani cingari zingari zingani nell'Italia moderna (1422-1812)*, CISU, Roma.

Corradi, L. (2018). *Il femminismo delle zingare: intersezionalità, alleanze, attivismo di genere e queer*, Mimesis, Milano.

Correa, T. and Jeong, S.H. (2011), Race and online content creation: Why minorities are actively participating in the Web, Information, *Communication & Society*, 14(5): 638-659.

De Bar, G. (1998), *Strada, patria sinta. Cento anni di storia nel racconto di un saltimbanco*, Fatatrac, Firenze.

De Ragna, A. (2023), *Vite in cammino Storia di una famiglia rom di Milano*, Upre, Roma.

del Gobbo, E., Cucco, A. and Fontanella, L. (2025). *Identification of misogynistic accounts on Twitter through graph convolutional networks*. In Giordano G., La Rocca M., Niglio M., Restaino M. and Vichi M., eds., *Statistical models and learning methods for complex data. CLADAG 2023. Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, Cham

Dos Santos, C., Melnyk, I. and Padhi, I. (2018), Fighting offensive language on social media with unsupervised text style transfer. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (volume 2: short papers) (pp. 189-194).

ISBN Open Access: 9788835191971

Elia, A., Fantozzi, P. (2017), *Discriminazioni in una regione del Mezzogiorno. I risultati di una ricerca in Calabria*. Rubbettino, Soveria Mannelli.
Faloppa, F. (2012), *Razzisti a parole (per tacer dei fatti)*, Laterza, Roma-Bari.
Fings, K. (2018). *Sinti e Rom: Storia di una minoranza*, il Mulino, Bologna.
Fondazione Domani Italia, a cura di (2014), *99 domande Romanipè 2.0. 99 domande sulla popolazione romanì*, Futura, Roma.
Giuffrè, M., a cura di (2014), *Uguali, diversi, normali. Stereotipi, rappresentazioni e contronarrative del mondo rom in Italia, Spagna e Romania*, Castelvecchi, Roma.
Guarnieri, N. (2003), *Cultura romanì. Molti giudicano, pochi conoscono. Spazio di confronto di conoscenza*, Editoria Roman, Roma.
Guarnieri, N. (2000), *Il laboratorio del sentimento. Dal pregiudizio alla tolleranza, dalla mediazione culturale alla conoscenza*, Editoria Romani, Roma.
Logacheva, V., Dementieva, D., Ustyantsev, S., Moskovskiy, D., Dale, D., Krotova, I. and Panchenko, A. (2022), Paradetox: Detoxification with parallel data. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 6804-6818).
Lundholt, M.W., Maagaard, C.A. and Piekut, A. (2018), Counternarratives, *The International Encyclopedia of Strategic Communication* (pp. 1-11).
Manzo, F. (2022), *Gli effetti dell'esclusione. 20 anni dal trasferimento dei Rom da Gergeri a San Vito Alto*, Coessenza, Torino.
Manzo, F. and Cosentino, M. (in corso di stampa), *Calabria romaní. Condizione e sviluppo dei rom in Calabria*, Reportage Edizioni, Calabria.
McKenzie-Mohr, S. and Lafrance, M.N. (2017), Narrative resistance in social work research and practice: Counter-storying in the pursuit of social justice, *Qualitative Social Work*, 16(2): 189-205
Meneghini, A.M. (2017). Stereotipi e paure degli italiani nei confronti degli zingari: una rassegna degli studi psicosociali condotti in Italia. *Psicologia sociale*, 1: 3-32.
Minh Tran, Y., Zhang, Y. and Soleymani, M. (2020), Towards a friendly online community: An unsupervised style transfer framework for profanity redaction. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 2107-2114).
Miškolci, J., Kováčová, L. and Rigová, E. (2020), Countering hate speech on Facebook: The case of the Roma minority in Slovakia, *Social Science Computer Review*, 38(2): 128-146.
Morelli, B., Soravia, G. (1998), *I pativ mengr. La lingua e le tradizioni dei rom abruzzesi*, Centro studi zingari, Roma.
Nakamura, L. (2007), *Digitizing race: Visual cultures of the Internet*, Vol. 23, University of Minnesota Press, USA.
Nelson, H. L. (2001), *Damaged identities, narrative repair*. Cornell University Press, USA.
Pasta, S. (2018), Razzismi 2.0. *Analisi socioeducativa dell'odio online*, Morcelliana, Brescia.
Pavlovic, D. (2025), *Irriducibili. Alterità nell'anima zingara*, Upre, Roma.

ISBN Open Access: 9788835191971

Pérez-Almendros, C., Espinosa-Anke, L. and Schockaert, S. (2020), Don't patronize me! an annotated dataset with patronizing and condescending language towards vulnerable communities, *arXiv preprint arXiv*, 2011.08320
Piasere, L. (2006), *Buoni da ridere, gli zingari. Saggi di antropologia storico-letteraria*, CISU, Roma.
Piasere, L. (2009), *I rom d'Europa. Una storia moderna*, Laterza, Roma-Bari.
Piasere, L. (2012), *Scenari dell'antiziganismo*. *Tra Europa e Italia, tra antropologia e politica*, SEID Editori, Firenze.
Piasere, L. (2015), *L'antiziganismo*, Quodlibet, Macerata.
Piasere, L. and Pontrandolfo, S., a cura di (2002), Italia romaní, vol. III: *I Rom di antico insediamento dell'Italia centro-meridionale*, CISU, Roma.
Piasere, L. and Pontrandolfo, S., a cura di (2016), Italia Romaní vol. VI: *Le migrazioni dei rom romeni in Italia*, CISU, Roma.
Piasere, L. and Saletti Salza, C., a cura di (2004), Italia romaní, vol. IV: *La diaspora rom dalla ex Iugoslavia*, CISU, Roma.
Pontrandolfo, S. (2013), *Rom dell'Italia meridionale*, CISU, Roma.
Pontrandolfo, S. and Rizzin, E. (2024). La produzione dell'antiziganismo nei discorsi dei politici dell'Italia contemporanea, *Antropologia Pubblica*, 6(1): 85–108.
Pontrandolfo, S. and Rizzin, E., (2021), Discorsi pubblici su rom e sinti in Italia. Un esempio di dangerous speech?. In De Vita A., a cura di, *Fragilità contemporanee. Fenomenologie della violenza e della vulnerabilità* (pp. 23-55), Mimesis, Milano.
Rizzin, E. and Bravi, L. (2024), *Lacio Drom. Storia delle "classi speciali per zingari". Rom e sinti a scuola 1965-1982*, Anicia, Roma.
Rizzin, E. (2020), *Attraversare Auschwitz. Storie di rom e sinti: identità, memorie, antiziganismo*, Gangemi, Roma.
Rizzin, E. and Bravi, L. (2024), *Lacio drom. storia delle "classi speciali per zingari": Rom e sinti a scuola, 1965-1982*, Anicia, Roma.
Sabiescu, A. (2005), Narratives and counter-narratives in the representation of The Other. The case of the Romani ethnic minority. In Bidwell N. J., and Winschiers-Theophilus H. (2015). *At the Intersection of Indigenous and Traditional Knowledge and Technology Design*, Informing Science Press, Santa Rosa, California.
Somers, M.R. (1994), The narrative constitution of identity: A relational and network approach, *Theory and Society*, 605-649.
Spinelli, S. (2016), *Rom, questi sconosciuti: storia, lingua, arte e cultura e tutto ciò che non sapete di un popolo millenario*, Mimesis, Milano.
Spinelli, S. (2018), *Una comunità da conoscere: storia, lingua e cultura dei Rom italiani di antico insediamento*, D'Abruzzo edizioni Menabò, Ortona (CH).
Spinelli, G. (2022), *Rom e Sinti dieci cose che dovresti sapere*, People, Busto Arsizio.
Topidi, K. and Metcalfe, J. (2024), Digital (mis)-representations: Understanding ethno-cultural minority identity formation online, *Digital Society*, 3(3): 45.
Trevisan, P. (2024). *La persecuzione dei rom e dei sinti nell'Italia fascista: Storia, etnografia e memorie*, Viella, Roma.

ISBN Open Access: 9788835191971

Villano, P., Fontanella, L., Fontanella, S. and Di Donato, M. (2017), Stereotyping Roma people in Italy: IRT models for ambivalent prejudice measurement, *International Journal of Intercultural Relations*, 57: 30-41.
Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N. and Kumar, R. (2019), Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval), *arXiv preprint arXiv*, 1903.08983.

Sitografia

Agenzia per i Diritti Fondamentali dell'Unione Europea (FRA) Discrimination and Hate Speech Against Roma (2023). Disponibile su: fra.europa.eu
European Commission Roma in the EU: The Case for a European Strategy (2011). Disponibile su: ec.europa.eu
European Roma Rights Centre (ERRC) Roma Rights Defenders (2021). Disponibile su: errc.org
Hancock, I. We Are the Romani People (2002). Available on various academic platforms.
Human Rights Watch Hate Speech and Its Impact on Society (2022). Disponibile su: hrw.org
Osservatorio Nazionale sull'Antiziganismo (ONA) Relazione Annuale (2023). Disponibile su: osservatorioantiziganismo.it
Oxford Internet Institute Online Hate Speech: A Critical Analysis (2020). Disponibile su: oii.ox.ac.uk
Pew Research Center the Impact of Hate Speech on Marginalized Communities (2021). Disponibile su: pewresearch.org
Rivista Lacio Drom (1965-1999), consultabile al seguente link: https://digital.sturzo.it/biblioteca-sturzo-numeri?periodico_id_s=IT-STURZO-BIB012-000001&type=periodico
UNESCO Combating Antigypsyism (2020). Disponibile su: unesco.org

ISBN Open Access: 9788835191971

Automazione e agency: i sistemi di counter-speech tra efficienza algoritmica e resistenza comunitaria

di *Mara Maretti*[*], *Clara Salvatori*[***]

1. Introduzione

L'hate speech online rappresenta una delle questioni più dibattute per le società democratiche contemporanee, situandosi al crocevia di valori potenzialmente in tensione: da un lato la tutela della dignità delle persone e dei gruppi, dall'altro la salvaguardia della libertà di espressione come fondamento del pluralismo democratico (Brown e Sinclair, 2019; Lepoutre, 2017). Il fenomeno solleva questioni fondamentali sulla natura della sfera pubblica digitale e sulle responsabilità delle piattaforme nella moderazione del discorso (Gillespie, 2018; Matamoros-Fernández, 2017).

È opportuno riconoscere che la definizione stessa di hate speech rimane oggetto di dibattito accademico e giuridico. Non esiste, infatti, una definizione universalmente accettata, principalmente a causa delle determinazioni vaghe e soggettive riguardo al fatto che un discorso sia "offensivo" o veicoli "odio" (Strossen, 2018). Sellars (2016), in una rassegna esaustiva delle diverse definizioni proposte in ambito accademico, giuridico e dalle piattaforme online, identifica alcuni tratti ricorrenti: il prendere di mira un gruppo, o un individuo in quanto membro di un gruppo; la presenza di un contenuto che esprime odio, causa un danno o incita ad azioni nocive e non ha alcuna finalità che lo riscatti; l'intenzione di arrecare danno; la natura pubblica del discorso; un contesto che rende possibile una risposta violenta. Tuttavia, come lo stesso Sellars sottolinea, questi tratti non costituiscono una definizione unitaria, ma criteri che, compresenti in misura maggiore o minore, possono contribuire ad accrescere la certezza che il discorso in questione meriti di essere identificato come discorso d'odio. Questa indeterminatezza definitoria non è un limite marginale, ma attraversa l'intero campo di studi e, come

[*] Dipartimento di Scienze Giuridiche e Sociali, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, mara.maretti@unich.it

[**] Dipartimento di Tecnologie Innovative in Medicina & Odontoiatria, Università degli Studi "G. d'Annunzio" di Chieti-Pescara, clara.salvatori@unich.it

ISBN Open Access: 9788835191971

si vedrà, si riflette nelle scelte progettuali dei sistemi automatizzati di rilevamento e contrasto.

Come evidenzia l'analisi bibliometrica di Tontodimamma *et al.* (2021), il campo di studi ha subito una trasformazione significativa nell'ultimo trentennio, con un'accelerazione della produzione scientifica e una crescente complessità degli approcci metodologici e teorici. Crucialmente, l'hate speech online non può essere trattato come semplice trasposizione digitale di forme preesistenti di discorso d'odio: si tratta di un fenomeno qualitativamente distinto, caratterizzato da velocità di diffusione, permanenza, anonimato e capacità di aggregazione senza precedenti (Gagliardone *et al.*, 2015).

In questo contesto, l'automazione del counter-speech emerge come una delle risposte tecnologiche più promettenti e al contempo problematiche (Munger, 2017; Wright *et al.*, 2017). La promessa è quella della scalabilità: sistemi capaci di identificare e rispondere a milioni di contenuti d'odio con una copertura impossibile per gli interventi umani. Ma questa promessa tecnologica solleva interrogativi che eccedono la dimensione puramente tecnica: i sistemi automatizzati possono davvero contestare l'autorità simbolica del discorso d'odio? Quale spazio rimane per l'agency delle comunità marginalizzate quando la risposta all'odio viene delegata agli algoritmi?

È precisamente la tensione tra efficienza algoritmica e agency comunitaria che questo capitolo si propone di indagare. L'analisi muove dal presupposto che i sistemi di counter-speech automatizzato costituiscano un fenomeno intrinsecamente sociotecnico, che richiede attenzione simultanea alle dimensioni strutturali, culturali e performative del discorso d'odio (Brown, 2017; Pohjonen e Udupa, 2017).

Il framework teorico adottato si colloca all'intersezione tra Science and Technology Studies (STS) e teoria della governamentalità algoritmica (Rouvroy, 2013; Rouvroy e Berns, 2013; Introna, 2016). Questa prospettiva consente di analizzare i sistemi automatizzati di counter-speech come tecnologie di governo: dispositivi che non si limitano a "rispondere" a contenuti preesistenti, ma che producono attivamente categorie, soggettività e relazioni di potere (Bucher, 2018; Gorwa *et al.*, 2020; Yeung, 2017). Il concetto di coproduzione (Jasanoff, 2004) guida l'analisi: i sistemi di counter-speech non sono strumenti neutrali applicati a un problema sociale preesistente, ma partecipano alla definizione stessa di cosa costituisca hate speech e di quali risposte siano considerate appropriate.

La riflessione si fonda su uno studio di caso multiplo che esamina quattro architetture sociotecniche paradigmatiche: la classificazione automatica di Perspective API (2017), la generazione knowledge-grounded di CONAN (2019), l'intelligenza artificiale generativa dei Large Language Models (2022+), e il sistema collaborativo human-in-the-loop CounterQuill (2024).

ISBN Open Access: 9788835191971

Questa selezione non segue una logica puramente tecnologica, ma mira a mostrare come differenti architetture distribuiscano in modo diverso l'agency tra algoritmi, esperti e comunità.

L'argomentazione si sviluppa attraverso sei paragrafi. Dapprima è presentato il framework teorico, elaborando una concezione performativa del counter-speech a partire dalla teoria degli atti linguistici (Austin, 1962), dalla riflessione butleriana sulla performatività (Butler, 1997) e dall'analisi bourdieusiana del potere simbolico (Bourdieu, 1991). Successivamente si propone un'analisi delle architetture delle principali piattaforme social, mostrando come fattori strutturali, più che le risorse investite nella moderazione, determinino la percezione e la diffusione dell'hate speech. In seguito,viene presentato uno studio di caso multiplo finalizzato a comparare i quattro sistemi di counter hate speech automatico in base a tre dimensioni interpretative: epistemologico-ontologica, politica ed etica. I risultati sono discussi nel quinto e nel sesto paragrafo, elaborando il concetto di "ortoprassi algoritmica" e interrogando i limiti del counter-speech come intervento sintomatico, presentando infine la prospettiva di un approccio community-in-the-loop.

2. Il counter-speech come pratica performativa: fondamenti teorici

Il counter-speech rappresenta una strategia comunicativa finalizzata al contrasto dell'hate speech o della disinformazione attraverso la presentazione di narrative alternative, evitando la censura del discorso offensivo e rispondendo invece con empatia e sfida costruttiva alle narrative d'odio (Benesch *et al.*, 2016). Le radici concettuali di questo approccio risalgono alla tradizione giuridica americana sulla libertà di espressione, e in particolare alla celebre formulazione del giudice Louis D. Brandeis nel caso Whitney v. California (1927): "Se c'è tempo per esporre attraverso la discussione la falsità e gli errori, per evitare il male attraverso i processi di educazione, il rimedio da applicare è più discorso, non il silenzio forzato" (Dangerous Speech Project, 2023)[1].

Questa formulazione, conosciuta come la "dottrina del counter-speech", ha profondamente influenzato non solo l'approccio costituzionale americano alla libertà di espressione, ma anche le politiche contemporanee delle piattaforme digitali. Brandeis sviluppò questa dottrina senza alcuna evidenza empirica della sua efficacia; si trattava piuttosto di un'affermazione di principio

[1] https://www.dangerousspeech.org/counterspeech

ISBN Open Access: 9788835191971

basata sulla fiducia illuministica nel potere della ragione e del dibattito pubblico (Benesch *et al.*, 2016). Solo recentemente, con l'avvento dei social media e la proliferazione dell'hate speech online, la ricerca si è focalizzata sul testare empiricamente se e quando il counter-speech possa effettivamente contrastare il discorso d'odio, trasformando un ideale normativo in oggetto di indagine scientifica (Hangartner *et al.*, 2021).

Prima di procedere all'analisi delle pratiche di counter-speech, è necessario esplicitare il quadro teorico-linguistico entro cui tale fenomeno può essere adeguatamente compreso. Il fondamento epistemologico della nostra analisi risiede nella concezione performativa del linguaggio, che segna una discontinuità radicale rispetto al paradigma rappresentazionalista classico.

Secondo l'approccio rappresentazionalista, dominante nella tradizione filosofica occidentale da Aristotele al positivismo logico, il linguaggio funzionerebbe primariamente come sistema di designazione referenziale: le espressioni linguistiche rappresenterebbero entità extralinguistiche, e il valore di verità degli enunciati dipenderebbe dalla loro corrispondenza con stati di cose nel mondo, cioè con la sua oggettivazione. In questa prospettiva, il linguaggio è concepito come medium trasparente, strumento neutro di rappresentazione di una realtà ontologicamente indipendente e preesistente all'atto enunciativo.

Austin (1962) rivede questa concezione neutra del linguaggio elaborando la teoria degli atti linguistici (speech acts) distinguendo tra enunciati constativi, che descrivono stati di cose e sono valutabili in termini di verità/falsità, ed enunciati performativi, che non descrivono ma compiono azioni nel momento stesso della loro enunciazione. Per comprendere meglio, le affermazioni: "Vi dichiaro marito e moglie", "Battezzo questa nave Queen Elizabeth", "Scommetto dieci euro che domani pioverà" non significano rappresentare una realtà preesistente, ma renderla reale in virtù della sua stessa enunciazione e non indipendentemente da essa, anche in considerazione delle aspettative che producono nell'interlocutore.

La dimensione performativa non è limitata ad una classe circoscritta di enunciati, ma costituisce una proprietà generale del linguaggio. Ogni atto linguistico possiede simultaneamente tre dimensioni: un atto locutorio (la produzione di suoni dotati di significato), un atto illocutorio (affermare, promettere, ordinare, minacciare) e un atto perlocutorio (gli effetti prodotti mediante il dire sul destinatario: persuadere, spaventare, convincere). La forza illocutoria di un enunciato, ciò che esso "compie" nel contesto comunicativo, non è riducibile al suo contenuto proposizionale. Su questo tracciato è Judith Butler (1997) che con la sua riflessione radicalizza la teoria della performatività applicandola al dominio dell'identità sociale e, specificamente, al discorso d'odio. Per Butler, il linguaggio non si limita a rappresentare identità

ISBN Open Access: 9788835191971

precostituite, ma partecipa attivamente alla loro costituzione. Gli atti linguistici operano attraverso la citazionalità: ogni enunciazione ripete, cita, riarticola convenzioni sedimentate, norme sociali cristallizzate in forme linguistiche. L'insulto razzista o sessista non è semplicemente l'espressione di un pregiudizio individuale preesistente, ma un atto che, citando e riattivando una storia di esclusione, posiziona il destinatario in una struttura sociale gerarchica, gli assegna un posto nel mondo, lo costituisce come soggetto sovraordinato o subordinato. In questa prospettiva, l'hate speech non "riflette" l'odio ma lo performa (rende reale un posizionamento sociale): produce effetti materiali sui corpi e sulle vite dei destinatari, contribuisce a riprodurre e naturalizzare le condizioni della loro marginalizzazione. Parallelamente, e questo è il punto teorico cruciale per la nostra analisi, anche il counter-speech non si configura come mera "risposta" rappresentazionale, come confutazione di affermazioni false, ma come pratica performativa che interviene sul terreno stesso della costituzione dei soggetti e delle relazioni sociali, contestando l'autorità simbolica del discorso dominante e aprendo spazi per risignificazioni alternative. Questa prospettiva si intreccia con l'analisi bourdieusiana del potere simbolico. Per Bourdieu (1991), il linguaggio non è mai un medium neutro di comunicazione, ma un campo di forze in cui si esercitano rapporti di dominio. Il potere simbolico è precisamente quella forma di potere che si esercita attraverso la comunicazione, imponendo significati come legittimi e dissimulando i rapporti di forza che ne sono il fondamento. Il linguaggio opera come strumento di dominazione nella misura in cui riproduce e naturalizza gerarchie sociali: le categorie attraverso cui nominiamo il mondo sociale non sono innocenti descrizioni, ma strumenti di classificazione che distribuiscono riconoscimento e misconoscimento, dignità e stigma. La violenza simbolica, concetto centrale nell'apparato teorico bourdieusiano, designa quella forma di violenza che si esercita con la complicità di chi la subisce, poiché viene misconosciuta in quanto tale e percepita come naturale o legittima.

L'hate speech si distingue dalla violenza simbolica propriamente detta per il suo carattere esplicito: mentre la violenza simbolica opera attraverso il misconoscimento e viene subita senza essere riconosciuta come violenza, percepita come ordine naturale delle cose, il discorso d'odio manifesta apertamente l'ostilità e l'intento di ferire. Tuttavia, le due forme di violenza non sono discontinue ma si collocano lungo un continuum: l'hate speech può essere letto come il momento in cui la struttura di dominio, che normalmente opera in modo sotterraneo (attraverso stereotipi naturalizzati, rappresentazioni egemoniche, pratiche discriminatorie routinizzate), emerge in superficie, rendendo visibile ciò che altrimenti resterebbe invisibile. Come hanno evidenziato i *Critical Race Theorists* a partire dal volume *Words That Wound*

ISBN Open Access: 9788835191971

(Matsuda *et al.*, 1993; si veda anche Delgado, 2004), le "parole che feriscono" producono danni reali e documentabili: effetti psicologici, esclusione sociale, perpetuazione di stereotipi che informano pratiche discriminatorie concrete. In questa prospettiva, l'hate speech non è semplicemente l'espressione di pregiudizi individuali, ma si inscrive in una struttura più ampia di violenza simbolica che pervade il tessuto sociale.

In tale cornice il counter-speech non rappresenterebbe semplicemente una "risposta" all'odio, ma una pratica performativa in grado di contrastare l'autorità simbolica del discorso dominante. È necessario superare una concezione meramente reattiva del fenomeno per coglierne la dimensione costitutiva: il counter-speech partecipa alla ridefinizione dei confini del dicibile e dell'accettabile nella sfera pubblica. Questa prospettiva implica un ripensamento della metafora spaziale del "marketplace of ideas" che ha tradizionalmente informato le difese liberali del counter-speech. L'idea classica, articolata nel pensiero di John Stuart Mill (XIX secolo), è che il rimedio al discorso d'odio sia sviluppare più narrazioni: nel libero mercato delle idee, le posizioni migliori prevarrebbero attraverso il confronto razionale. Attraverso questa prospettiva la libertà di espressione sarebbe il miglior antidoto alle idee false o dannose: se lasciamo che tutte le opinioni circolino liberamente, attraverso il dibattito razionale le idee migliori (più vere, più giuste) finiranno per prevalere. Applicato all'hate speech significa non censurare il discorso d'odio, ma rispondergli con più voci. Questa prospettiva postula che: 1) tutti abbiano uguale accesso alla parola; 2) il dibattito sia effettivamente sempre razionale o quantomeno ragionevole; 3) le persone cambino idea di fronte ad argomenti migliori. Ma sappiamo che non funziona sempre in questo modo. Chi subisce hate speech spesso non ha le stesse risorse (tempo, visibilità, legittimità sociale) per "rispondere" in modo efficace e competente. E chi odia raramente cambia idea perché qualcuno lo "confuta" logicamente.

Gelber (2002), in *Speaking Back*, propone un framework, alternativo alla prospettiva illuministica di Mill, fondato sulla teoria delle capabilities di Nussbaum: il counter-speech non è semplicemente "più parole" da aggiungere al mercato delle idee, ma una pratica che richiede supporto istituzionale, materiale ed educativo per permettere ai gruppi marginalizzati di rispondere all'hate speech. L'obiettivo non è solo contraddire i messaggi d'odio, ma contrastare i loro effetti di silenziamento, disempowerment e marginalizzazione, ovvero ripristinare le condizioni che permettono ai gruppi colpiti di partecipare effettivamente al discorso pubblico, permettendo loro di definire attivamente la loro identità, diventando soggetti politici.

ISBN Open Access: 9788835191971

3. Ecologie dell'odio online: piattaforme, algoritmi e dinamiche di diffusione

Non vi è dubbio che l'emergere dei social media ha amplificato sia la diffusione dell'hate speech che la necessità di strategie di contrasto efficaci.

La crescente consapevolezza tra i policy maker nazionali e internazionali, le organizzazioni della società civile e le aziende di social media indica che l'hate speech online impatta negativamente le comunità colpite, in particolare donne, adolescenti, persone LGBTQIA e gruppi etnici minoritari.

Le differenze algoritmiche e demografiche delle piattaforme definiscono una variabilità rilevante della distribuzione dell'hate speech tra i vari social.

Prima di analizzare i sistemi automatizzati di counter-speech, è necessario interrogare le condizioni strutturali entro cui essi operano. Se l'hate speech fosse un fenomeno essenzialmente testuale, ossia messaggi isolati da contestare con altri messaggi, l'automazione della risposta apparirebbe come soluzione naturale. Ma cosa accade se la diffusione dell'odio dipende primariamente da fattori architetturali, algoritmici e sociodemografici su cui il counter-speech, automatico o meno, non può intervenire direttamente?

Il sondaggio globale condotto da Ipsos per UNESCO nel settembre 2023, coinvolgendo 8.000 partecipanti in 16 paesi, rivela che i cittadini percepiscono Facebook come la piattaforma dove l'hate speech è più prevalente (58% degli intervistati), seguita da TikTok (30%), X/Twitter (18%) e Instagram (15%) (Ipsos-UNESCO, 2023)[2].

Nel contesto statunitense, i dati sono ancora più allarmanti. L'Anti-Defamation League (tab. 1), nel suo sesto rapporto annuale "Online Hate and Harassment: The American Experience 2024", basato su un sondaggio rappresentativo di 2.479 adulti americani condotto da YouGov tra gennaio e febbraio 2024, documenta che più della metà degli americani (56%) riporta di aver subito hate speech o harassment online nel corso della propria vita, la percentuale più alta registrata dal 2020. Il dato più preoccupante riguarda l'harassment severo, che include minacce fisiche, stalking, harassment sessuale, doxing e swatting, che ha colpito il 22% degli americani negli ultimi 12 mesi, in aumento rispetto al 18% del 2023. Facebook rimane la piattaforma dove si verifica la maggior parte dell'harassment, con il 61% delle vittime che riportano di aver subito almeno parte dell'harassment su questa piattaforma, seguita da Instagram (39%) e Twitter (28%). Particolarmente significativo è l'aumento dell'harassment su piattaforme di messaggistica:

[2] *Survey on the Impact of Online Disinformation and Hate Speech*. Il report è disponibile su: https://www.ipsos.com/sites/default/files/ct/news/documents/2023-11/unesco-ipsos-online-disinformation-hate-speech.pdf

ISBN Open Access: 9788835191971

WhatsApp è passato dal 14% al 25% e Telegram dal 7% al 13% (ADL, 2024).

Tab. 1 – Condotte offensive online per piattaforma (ADL 2019-2024)

Platform	2019	2020	2021	2022	2023	2024
Facebook	56%	77%	75%	68%	54%	61%
Instagram	16%	17%	24%	26%	27%	39%
Twitter/X	19%	27%	24%	23%	27%	27%
YouTube	17%	18%	21%	20%	—	—
TikTok	—	—	—	14%	19%	—
WhatsApp	13%	6%	11%	—	14%	25%
Reddit	11%	8%	9%	5%	15%	—
Telegram	—	—	—	—	7%	13%
Snapchat	10%	8%	15%	—	—	—
Discord	7%	4%	7%	—	—	—
Twitch	8%	4%	6%	—	—	—

Note. Le percentuali rappresentano la quota di vittime che riportano di aver subito harassment su ciascuna piattaforma (risposta multipla). Dati tratti dall'ADL Annual Online Hate and Harassment Survey (2019-2024). Il simbolo "—" indica dati non riportati per quell'anno. I valori del 2022 si riferiscono alla metrica "lifetime". I report completi sono disponibili su: https://www.adl.org/resources/reports?topics=6096

Le ragioni di questa proliferazione differenziata sono molteplici e strutturali; è difficile da cogliere in un'argomentazione basata su dati spesso non comparabili. Detto ciò, possiamo provare a ipotizzare una spiegazione della percezione degli utenti (tab. 1) in base alle caratteristiche dei differenti social senza pretesa di esaustività.

Tab. 2 – Correlazioni tra caratteristiche delle piattaforme e percezione di hate speech

Platform	HS perc.	Architettura	Gruppi	Età utenti	Contenuto	Moderazione
Facebook	58%	Social graph	Sì	35-65	Testo/misto	~15.000
TikTok	30%	Algoritmico	No	16-30	Video brevi	>40.000
X/Twitter	18%	Interest graph	No	25-50	Testo breve	~1.849
Instagram	15%	Social+interest	No	18-35	Visual	(Meta)

Note. HS perc. = Hate speech percepito (% utenti che dichiarano di aver visto contenuti d'odio). Social graph = connessioni basate su relazioni tra utenti; Interest graph = connessioni basate su interessi/follow; Algoritmico = contenuti selezionati prevalentemente da algoritmo (For You Page). Fonte HS: UNESCO & Ipsos (2023), N=16.000.

ISBN Open Access: 9788835191971

La Tabella 2 mette in relazione la percezione di hate speech con alcune caratteristiche strutturali delle principali piattaforme social. La variabile dipendente, come già accennato e riportato in precedenza è la percentuale di utenti che dichiara di aver visto contenuti d'odio (UNESCO e Ipsos nel 2023 su un campione di 16.000 rispondenti in 16 paesi). Si tratta di un dato di percezione soggettiva, non di una misurazione oggettiva della quantità di hate speech presente su ciascuna piattaforma: ciò che viene rilevato è l'esperienza vissuta dagli utenti, influenzata sia dall'effettiva esposizione a contenuti problematici sia dalla sensibilità individuale nel riconoscerli come tali.

Le variabili esplicative incluse nella Tabella sono state selezionate sulla base della letteratura esistente sui platform studies e sulle dinamiche dell'odio online. L'architettura delle connessioni distingue tra piattaforme basate su relazioni sociali preesistenti (social graph, come nel caso di Facebook), piattaforme organizzate attorno a interessi e following (interest graph, come X/Twitter), e piattaforme dove la selezione dei contenuti è prevalentemente delegata all'algoritmo (come TikTok). La presenza di gruppi chiusi o semiprivati ci indica se la piattaforma offre spazi aggregativi difficilmente accessibili alla moderazione esterna. La tipologia di contenuto dominante distingue tra piattaforme prevalentemente testuali, visuali o video. I dati sullo staff di moderazione provengono dai report di trasparenza pubblicati ai sensi del Digital Services Act europeo nel 2024.

Il dato che emerge in modo più evidente è la posizione di Facebook dove il 58% degli utenti dichiara di aver visto contenuti d'odio — una percentuale quasi doppia rispetto a TikTok (30%) e più che tripla rispetto a Instagram (15%). Questa marcata differenza non può essere spiegata dalla sola dimensione della base utenti o dalle risorse investite nella moderazione: Facebook dispone di circa 15.000 moderatori, un numero significativamente superiore ai 1.849 di X/Twitter, eppure la percezione di hate speech su X risulta molto inferiore (18%). Occorre dunque guardare alle caratteristiche strutturali della piattaforma.

Prima però di ipotizzare spiegazioni algoritmiche di diffusione del contenuto d'odio, è utile chiedersi se vi sia qualche possibile bias nella percezione del discorso d'odio nelle varie piattaforme social. Per quanto riguarda la percezione individuale, che in parte giustificherebbe i risultati della survey UNESCO & Ipsos, Facebook è costruito su un social graph che riflette relazioni personali originate offline: amici, familiari, colleghi, compagni di scuola. Questo significa che gli utenti sono esposti a contenuti prodotti da persone che conoscono nella vita reale. Il fenomeno del "context collapse" descritto da Marwick e Lewis (2017) e Boyd (2011), ossia la compresenza in un unico spazio digitale di audiences normalmente separate, amplifica l'impatto emotivo dei contenuti d'odio: leggere un commento razzista da parte di un ex

ISBN Open Access: 9788835191971

compagno di classe o di un parente potrebbe produrre un effetto più disturbante, rispetto allo stesso contenuto proveniente da uno sconosciuto su TikTok. La percezione di hate speech, in altre parole, non dipende solo dalla quantità di contenuti problematici, ma anche dalla relazione con chi li produce. Un altro aspetto, un'altra variabile "confondente" rispetto alla percezione dei contenuti di odio, è la fascia di età prevalente degli utenti delle varie piattaforme. Facebook presenta una popolazione significativamente più anziana rispetto alle altre piattaforme: il suo nucleo centrale si colloca nella fascia 35-65 anni, mentre TikTok è utilizzato prevalentemente da utenti tra i 16 e i 30 anni, e Instagram da giovani adulti tra i 18 e i 35. Questa differenza generazionale potrebbe influire sulla percezione di hate speech attraverso diversi meccanismi. In primo luogo, gli utenti più anziani potrebbero presentare soglie di sensibilità differenti: espressioni che le generazioni più giovani, cresciute negli ambienti digitali, considerano parte del repertorio comunicativo del proprio tempo, incluse forme di ironia aggressiva, sarcasmo e provocazione, possono risultare più offensive per chi ha sviluppato le proprie competenze comunicative in contesti perlopiù offline. In secondo luogo, la *digital literacy*, intesa come capacità di navigare le piattaforme, personalizzare i feed e adottare strategie di difesa attiva (silenziare, bloccare, segnalare utenti problematici), tende a essere meno sviluppata nelle coorti più anziane, con il risultato di una maggiore esposizione passiva a contenuti indesiderati.

Questo quadro interpretativo si inserisce in modo non contraddittorio rispetto alle evidenze empiriche, che mostrano come adolescenti e giovani adulti siano sproporzionatamente più esposti all'hate speech online rispetto alla popolazione generale. In particolare, sebbene circa il 20% dei giovani adulti (18–24 anni) dichiari di aver subìto vittimizzazione diretta, oltre il 70% riferisce di aver assistito a episodi di hate speech negli ultimi tre mesi (Reichelmann *et al.*, 2021). Tale esposizione massiva sembra produrre effetti ambivalenti: da un lato, numerosi studi documentano conseguenze psicologiche negative, tra cui aumento dello stress, sintomi ansiosi e riduzione della fiducia interpersonale (Näsi *et al.*, 2015; Keipi *et al.*, 2017); dall'altro, l'esposizione ripetuta può favorire processi di assuefazione e desensibilizzazione, contribuendo alla normalizzazione dell'hate speech nello spazio digitale (Soral *et al.*, 2018).

L'assuefazione comunicativa potrebbe contribuire a spiegare perché i soggetti più giovani, pur essendo più frequentemente esposti all'odio online, tendano a percepirlo meno come tale o a non riportarlo come hate speech, un'ipotesi che la letteratura suggerisce con crescente convergenza.

Volendo ora passare ad ipotizzare quali possano essere le caratteristiche strutturali delle varie piattaforme che porterebbero a un'effettiva proliferazione dei contenuti d'odio, possiamo partire sempre da Facebook. Qui vi è

ISBN Open Access: 9788835191971

una caratteristica importante da tenere in considerazione, ossia la presenza dei Gruppi, spazi privati che non hanno equivalenti strutturali sulle altre piattaforme analizzate. I gruppi Facebook funzionano come comunità chiuse dove si sviluppano dinamiche di echo chamber e dove il discorso d'odio può normalizzarsi progressivamente, al riparo dalla visibilità pubblica e dalla moderazione automatizzata. L'inchiesta giornalistica "Facebook Files" pubblicata dal Wall Street Journal nel 2021, basata su documenti interni dell'azienda, ha mostrato come l'algoritmo di raccomandazione dei Gruppi abbia sistematicamente indirizzato utenti verso comunità estremiste (Horwitz e Seetharaman, 2021). Questa dinamica è coerente con i risultati delle numerose ricerche condotte sulle filter bubbles (ad esempio Bakshy, Messing e Adamic, 2015), che hanno mostrato come l'esposizione a contenuti ideologicamente omogenei sia più marcata all'interno delle reti sociali strette.

Anche l'architettura algoritmica di YouTube opera non come semplice meccanismo di selezione neutrale, ma come infrastruttura che orienta attivamente i percorsi di consumo verso contenuti sempre più radicali. Ribeiro *et al.* (2020) hanno fornito la prima evidenza empirica su larga scala dell'esistenza di "percorsi di radicalizzazione" algoritmici su Youtube. Analizzando 330.925 video pubblicati su 349 canali e processando oltre 72 milioni di commenti, lo studio ha dimostrato che gli utenti migrano sistematicamente da contenuti moderati verso contenuti progressivamente più estremi, con le raccomandazioni della piattaforma che facilitano questa progressione. Nel caso di Reddit, Massanari (2017) ha dimostrato, attraverso un'etnografia digitale di lungo periodo, come specifici affordances strutturali della piattaforma abbiano fornito terreno fertile per quello che l'autrice definisce *toxic technocultures*. Analizzando i casi emblematici di *#Gamergate* e *The Fappening*, lo studio identifica cinque caratteristiche architetturali che facilitano l'attivismo antifemminista e misogino: il sistema di *karma point* che premia la popolarità sopra la qualità; l'aggregazione di materiale tra subreddit che permette la rapida diffusione virale di contenuti; l'estrema facilità di creazione di account e subreddit che consente la proliferazione di spazi dedicati all'hate speech; la struttura di governance decentralizzata basata su moderatori volontari; le politiche ambigue sui contenuti offensivi che storicamente hanno privilegiato la libertà di espressione sulla protezione dall'odio. Queste scelte di design non sono meramente tecniche ma riguardano le politiche della piattaforma.

Le piattaforme digitali non operano quindi come contenitori neutrali entro cui l'hate speech semplicemente "accade", ma come infrastrutture attivamente configurate da scelte architetturali, algoritmiche e di governance che co-producono le condizioni stesse della sua diffusione.

ISBN Open Access: 9788835191971

4. Automazione e agency nel counter-speech: un'analisi comparativa attraverso il framework STS

Quanto emerso dall'analisi delle architetture di YouTube, Facebook e Reddit trova conferma in una crescente letteratura che documenta come la diffusione dell'hate speech online non sia riducibile ai contenuti testuali, ma dipenda strutturalmente dalle affordance e dalle logiche algoritmiche delle piattaforme. Questa constatazione solleva una domanda fondamentale che costituisce il punto di partenza del presente paragrafo: se i fattori che determinano la percezione e la diffusione dell'odio online sono di natura strutturale, radicati nell'architettura delle piattaforme, nelle dinamiche di gruppo e nelle caratteristiche sociodemografiche degli utenti, quale ruolo può effettivamente svolgere il counter-speech automatizzato? Ci troviamo di fronte a un disallineamento tra il livello dell'intervento (la risposta al messaggio individuale) e il livello delle determinanti del fenomeno (le condizioni strutturali che ne favoriscono la proliferazione)?

Per rispondere a queste domande, adottiamo la prospettiva degli Science and Technology Studies (STS) come framework analitico. Gli STS offrono strumenti concettuali particolarmente adatti ad analizzare come le tecnologie di counter-speech non siano semplicemente strumenti neutrali applicati a un problema sociale preesistente, ma partecipino attivamente alla co-produzione (Jasanoff, 2004) di specifiche comprensioni dell'hate speech, della moderazione e della governance online. Applicato al nostro oggetto di studio, questo principio implica che i sistemi di counter-speech automatico non si limitano a "rispondere" all'hate speech, ma contribuiscono a definire cosa sia l'hate speech, quali risposte siano considerate appropriate e chi abbia l'autorità di produrle.

A differenza degli approcci che trattano la tecnologia come variabile indipendente che produce effetti sociali (determinismo tecnologico) o come mero strumento neutro plasmato da forze sociali esterne (costruzionismo sociale ingenuo), gli STS considerano scienza e tecnologia come non governate esclusivamente da logiche interne di efficienza o verità, ma profondamente embedded in strutture politiche, interessi economici e norme culturali.

Oltre al già citato concetto di co-produzione (co-production), sviluppato da Sheila Jasanoff (2004), che designa il processo attraverso cui la conoscenza scientifica e l'ordine sociale vengono simultaneamente prodotti, un secondo concetto rilevante è quello di immaginari sociotecnici (sociotechnical imaginaries), definiti come «visioni di futuri desiderabili collettivamente condivise, istituzionalmente stabilizzate e pubblicamente performate» (Jasanoff e Kim, 2009, p. 120). Gli immaginari sociotecnici non sono semplici previsioni o utopie: sono visioni normative che orientano lo sviluppo tecno-

ISBN Open Access: 9788835191971

logico, giustificano investimenti e modellano le aspettative pubbliche. Nel caso del counter-speech automatizzato, possiamo identificare un immaginario dominante che configura l'hate speech come problema risolvibile attraverso l'applicazione di tecniche di elaborazione del linguaggio naturale e la risposta appropriata come produzione scalabile di contronarrazioni generate algoritmicamente.

La prospettiva STS si integra produttivamente con il concetto di governamentalità algoritmica sviluppato da Antoinette Rouvroy e Thomas Berns (2013). Questo concetto designa una forma di governo del mondo sociale basata sull'elaborazione algoritmica di grandi insiemi di dati piuttosto che sulla politica, sul diritto e sulle norme sociali tradizionali. La governamentalità algoritmica opera attraverso tre momenti interconnessi: la datificazione (trasformazione dell'azione sociale in dati quantificati), la profilazione (costruzione di profili comportamentali a partire dai dati) e l'azione anticipatoria (intervento preventivo basato su predizioni algoritmiche).

Rouvroy (2013) descrive questo processo come "comportamentismo dei dati": gli algoritmi non ci interrogano come soggetti pensanti, ma raccolgono le nostre tracce digitali (click, ricerche, tempo di permanenza) per costruire profili predittivi. L'autorità decisionale si sposta così dalle persone agli oggetti tecnici, diventando apparentemente neutrale e difficile da contestare. In questo modo, lo spazio per la riflessione critica e il dialogo viene significativamente ridotto. Applicata ai sistemi di counter-speech automatico, questa prospettiva illumina come tali tecnologie operino una doppia datificazione: dell'hate speech (trasformato in pattern testuali classificabili) e della risposta appropriata (codificata in dataset di contronarrazioni "esemplari"). Il discorso d'odio viene così tradotto da fenomeno sociale complesso, radicato in strutture di potere, dinamiche di gruppo e contesti culturali specifici, in problema tecnico di classificazione e generazione testuale. Questa traduzione non è neutrale: essa co-produce una specifica comprensione dell'hate speech come fenomeno essenzialmente linguistico, decontestualizzato e trattabile algoritmicamente.

Sulla base di questo framework teorico, è possibile proporre l'analisi dei sistemi di counter-speech automatizzato attraverso tre dimensioni interpretative trasversali che guidano l'esame di alcuni significativi casi di studio selezionati per questo capitolo. Ci riferiamo alla dimensione *epistemologico-ontologica* che permette di focalizzare su come il sistema definisce e operazionalizza l'hate speech, quali assunzioni ontologiche sulla natura del discorso d'odio sono incorporate nel design tecnico e come viene costruita la "risposta appropriata" e sulla base di quale autorità epistemica.

La seconda dimensione, quella *politica* ha a che fare con la *performatività, l'agency, il potere e la decisione*. Tale asse interpretativo è utile per

ISBN Open Access: 9788835191971

cercare di comprendere quanto il sistema automatico possa essere performativo (secondo la definizione già fornita in precedenza) e a chi sia attribuita la responsabilità dell'azione. Seguendo Butler (1997), l'hate speech non descrive semplicemente la realtà ma la costituisce performativamente, producendo soggetti subordinati attraverso la replicazione di convenzioni storicamente sedimentate. Quale agency viene attribuita al counter-speech automatico? Può una risposta algoritmica contestare l'autorità simbolica del discorso d'odio, o rischia di riprodurre le strutture di potere che intende sfidare? Quali valori sono incorporati nel sistema? Chi partecipa alla definizione di cosa costituisca una risposta appropriata? Come sono distribuiti i rischi e i benefici dell'automazione tra le diverse comunità coinvolte?

La terza dimensione ossia *etica* (*giustizia, riconoscimento e danno*), risponde alle seguenti domande: quali valori sono incorporati nel sistema? Chi partecipa alla definizione di cosa costituisca una risposta appropriata? Come sono distribuiti i rischi e i benefici dell'automazione tra le diverse comunità coinvolte? Le comunità marginalizzate sviluppano strategie sofisticate di autorappresentazione e contronarrazione che articolano l'identità attraverso temi positivi dell'ingroup piuttosto che attraverso la costruzione dell'alterità? I sistemi automatizzati riconoscono e preservano queste pratiche, o le sostituiscono con risposte standardizzate che eludono la conoscenza situata delle comunità?

La selezione dei quattro casi studio presi in considerazione non segue una logica puramente tecnologica, ma mira a mostrare come differenti architetture sociotecniche distribuiscano in modo diverso l'agency tra algoritmi, esperti e utenti, co-producendo nel processo specifiche comprensioni di cosa costituisca hate speech e di quali risposte siano appropriate. L'ordine in cui sono presentati i casi riflette l'evoluzione cronologica delle tecnologie: dalla classificazione automatica (Perspective API, 2017), alla generazione knowledge-grounded (CONAN, 2019), all'intelligenza artificiale generativa (ChatGPT, 2022), fino ai sistemi collaborativi human-in-the-loop (CounterQuill, 2024).

Nella Tabella 3 è sintetizzata l'analisi comparativa dei quattro casi secondo le dimensioni: tecnologica; epistemico-ontologica; politica; etica.

ISBN Open Access: 9788835191971

Tab. 3 – Analisi comparativa dei sistemi automatici di Counter-Speech: architetture e implicazioni sociotecniche (2017-2024)

Dimensione	Perspective API (Google/Jigsaw, 2017)	CONAN (FBK, 2019)	LLMs (ChatGPT) (OpenAI, 2022+)	CounterQuill (Virginia Tech, 2024)
Funzione primaria	Sistema di moderazione che assegna punteggi di tossicità ai contenuti	Genera controarrative specifiche per tipo di discorso d'odio	Genera risposte di counter-speech senza addestramento specifico	Assistente alla co-scrittura con percorso educativo in tre fasi
Logica operativa	Classificazione automatica basata su probabilità statistiche	Nichesourcing: trasferisce al sistema l'esperienza specializzata delle ONG	Intelligenza artificiale "generalista" con capacità emergenti da addestramento massivo	Collaborazione pedagogica: l'IA come partner di apprendimento, non sostituto
DIMENSIONE TECNOLOGICA: come funziona				
Come è stato costruito	Addestrato su milioni di commenti etichettati da lavoratori retribuiti online (Wikipedia, NYT)	Modello linguistico specializzato addestrato su ~5.000 coppie di hate speech e controarrative create da esperti di ONG	Modello linguistico di grandi dimensioni addestrato su enormi quantità di testi dal web, poi affinato attraverso feedback umano (miliardi di parametri)	Sistema strutturato in 3 moduli (apprendimento, brainstorming, co-scrittura) che guida l'utente
Input/ Output	Input: testo. Output: punteggi da 0 a 1 per diverse forme di tossicità (insulti, minacce, attacchi identitari)	Input: messaggio d'odio + gruppo bersaglio. Output: controarrativa specifica (basata su fatti, empatia, ecc.)	Input: istruzioni + contesto. Output: testo generato liberamente	Input: messaggio d'odio + riflessioni dell'utente. Output: counter-speech co-costruito insieme
Capacità di scala	Altissima: può analizzare milioni di richieste al giorno in pochi millisecondi	Media: richiede adattamento specifico per ogni contesto/lingua, costi elevati per creare dataset	Alta: accessibile via servizio commerciale, nessun addestramento necessario, ma con costi per utilizzo	Bassa: richiede impegno attivo dell'utente (~20 minuti per risposta)

ISBN Open Access: 9788835191971

Dimensione	Perspective API (Google/Jigsaw, 2017)	CONAN (FBK, 2019)	LLMs (ChatGPT) (OpenAI, 2022+)	CounterQuill (Virginia Tech, 2024)
DIMENSIONE EPISTEMOLOGICA: come "conosce" l'hate speech				
Fonte del sapere	Maggioranza statistica da lavoratori online non-esperti retribuiti	Esperienza situata di ONG specializzate nella lotta all'odio (Stop Hate UK, CESIE, ecc.)	Pattern statistici estratti da enormi quantità di testi web, conoscenza implicita	L'esperienza diretta dell'utente + supporto pedagogico del sistema
Definizione di hate speech	Operazionale/comportamentale: "contenuto che può far abbandonare una conversazione"	Categoriale – discorso che attacca gruppi per caratteristiche identitarie	Implicita nei filtri di sicurezza e politiche d'uso predefinite	Educativa/riflessiva: l'utente impara a riconoscere elementi costitutivi (target, deumanizzazione)
Comprensione del contesto	Decontestualizzata: analizza testi isolati, ignora storia conversazionale e contesto culturale	Semi-contestualizzata: target specifici, ma risposte pre-generate	Decontestualizzata: nessuna memoria tra conversazioni, ogni prompt indipendente	Contestualizzata: l'utente porta conoscenza situazionale, riflessione guidata
DIMENSIONE POLITICA o dell'AGENCY: chi decide				
Centro decisionale	Centralizzato: Google/Jigsaw definisce soglie, attributi, dati di addestramento	Mediato: ONG europee definiscono strategie appropriate	Delegato all'algoritmo: il modello decide il contenuto, OpenAI definisce policy	Distribuito: decisione finale sempre umana, AI supporta
Agency comunitaria	Assente: le comunità bersaglio non partecipano alla definizione di tossicità	Mediata: ONG rappresentano (ma non sono) le comunità bersaglio	Assente: nessun coinvolgimento nella generazione delle risposte	Preservata: l'utente mantiene voce e controllo sul messaggio finale
Governance	Aziendale: proprietario, opaco, decisioni unilaterali	Ibrida: ricerca pubblica (dataset), ma implementazione proprietaria	Aziendale: modello proprietario, policy decise unilateralmente	Accademica/Open: sistema di ricerca, principi trasparenti

ISBN Open Access: 9788835191971

Dimensione	Perspective API (Google/Jigsaw, 2017)	CONAN (FBK, 2019)	LLMs (ChatGPT) (OpenAI, 2022+)	CounterQuill (Virginia Tech, 2024)
Relazione con piattaforme	Integrazione diretta: usato da NYT, Reddit per moderazione automatica	Indipendente: strumento di ricerca, non integrato in piattaforme mainstream	Ambivalente: ChatGPT indipendente, ma API usate in vari servizi	Potenziale integrazione come estensione browser progettata per integrazione futura
DIMENSIONE ETICA: rischi e limiti				
Rischio principale	Discriminazione sistemica: sovrasegnalazione di varietà linguistiche minoritarie (AAVE, dialetti)	Paternalismo: esperti decidono la "risposta corretta" per le comunità	Genericità e "allucinazioni" (fatti inventati)	Barriere all'accesso: richiede tempo, motivazione, alfabetizzazione digitale
Bias documentati	Razziale: varietà linguistiche come AAVE classificate più tossiche. Politico: asimmetria destra/sinistra	Culturale: contronarrative riflettono prospettive ONG europee occidentali	Tossicità emergente: modelli più grandi generano risposte più tossiche (+25-44%)	Non ancora documentati su larga scala (studio pilota N=20)
Limiti intrinseci	Non genera risposte, solo classifica. Non distingue ironia, contesto, reclaiming	Dataset limitato, rischio genericità, difficile adattamento cross-culturale	Nessuna comprensione reale, solo pattern statistici, nessuna memoria	Non scalabile per intervento massivo, dipende da motivazione utente
Conseguenze per l'agency	Silenziamento: può censurare voci marginalizzate che usano linguaggio "tossico" per reclaiming	Sostituzione: le risposte "esperte" sostituiscono auto-rappresentazione comunitaria	Alienazione: risposte algoritmiche non riflettono esperienza vissuta delle comunità	Empowerment potenziale: costruisce competenze, rafforza senso di ownership e autoefficacia

ISBN Open Access: 9788835191971

Prima di inoltrarci in una valutazione più sociotecnica dei quattro casi presi in esame risulta utile descriverne sinteticamente le caratteristiche tecniche. Perspective API (2017) opera a monte della risposta: non genera counter-speech ma ne costituisce la condizione di possibilità tecnica, classificando i contenuti attraverso punteggi di "tossicità". Il sistema incarna la logica della *moderazione algoritmica* nella sua forma più elementare: decisioni automatizzate, scalabili, apparentemente neutrali, fondate su pattern statistici estratti da milioni di annotazioni crowdsourced. CONAN (2019) sposta il focus dalla classificazione alla generazione, ma attraverso una mediazione umana qualificata: il *nichesourcing*[3]. Qui non sono crowd (pubblico generico) anonimi a definire cosa sia odio e come rispondervi, ma operatori di ONG specializzate nel contrasto all'hate speech. L'expertise situata viene "congelata" in un dataset che addestra modelli generativi, trasferendo al sistema uno specifico pattern della risposta appropriata. ChatGPT e i Large Language Models (2022+) rappresentano un salto paradigmatico: la generazione di contenuto non richiede più dataset specializzati né fine-tuning[4], ma emerge dalle capacità generaliste di modelli pre-addestrati su contenuti dal web. Il counter-speech diventa un caso particolare di una competenza linguistica universale, accessibile tramite *prompt engineering*[5]. Questa modalità di generazione di contenuti nasconde però una nuova forma di opacità: le "guardrails" etiche sono incorporate nel modello stesso, definite unilateralmente dal produttore. Infine, il CounterQuill (2024) rovescia la direzione del flusso: anziché automatizzare la produzione di risposte, il sistema supporta gli utenti nel produrle autonomamente. L'AI diventa *partner educativo* in un workflow strutturato che preserva l'authorship umana. CounterQuill utilizza lo stesso modello di ChatGPT (GPT-3.5); la differenza cruciale non risiede nella tecnologia sottostante, ma nell'architettura dell'interazione. Questa progressione, dalla classificazione automatica, alla generazione mediata da esperti, alla generazione algoritmica generalista, alla co-produzione collaborativa tra umani e AI, non descrive un'evoluzione lineare verso soluzioni "migliori", ma quattro risposte distinte a una tensione irrisolta: come conciliare la scala del problema (miliardi di contenuti) con la specificità delle risposte (contestualizzazione culturale, sensibilità alle dinamiche di potere,

[3] A differenza del *crowdsourcing*, che raccoglie contributi da una massa indifferenziata di utenti, il *nichesourcing* coinvolge una comunità ristretta di esperti di dominio. Il termine sottolinea il passaggio dalla quantità (la "folla") alla qualità (la "nicchia" competente).

[4] Il fine-tuning, letteralmente "messa a punto" è un processo attraverso cui un modello di AI già addestrato viene specializzato su un compito specifico mediante esposizione a esempi mirati.

[5] Il Prompt engineering è una tecnica che consiste nel formulare istruzioni testuali (dette prompt) in modo da guidare il comportamento di un modello di AI generativa verso l'output desiderato, senza modificarne la struttura sottostante.

ISBN Open Access: 9788835191971

rispetto dell'agency comunitaria). Come mostra la Tabella 3, ciascun paradigma risolve questa tensione privilegiando dimensioni diverse: scalabilità versus contestualizzazione, efficienza versus partecipazione, automazione versus educazione, con implicazioni profondamente differenti sul piano epistemologico, politico ed etico.

4.1. Il prerequisito del rilevamento: Perspective API e la datificazione dell'odio

Perspective API, sviluppata a partire dal 2017 da Jigsaw (società del gruppo Alphabet, holding di Google), focalizzata su tecnologie per la sicurezza online, rappresenta il paradigma dominante in questo ambito. Il sistema non produce risposte all'hate speech, ma ne costituisce la condizione di possibilità tecnica poiché classifica i contenuti assegnando punteggi di "tossicità" su scala 0-1, consentendo l'identificazione automatizzata di contenuti potenzialmente problematici su scala altrimenti ingestibile (considerando la grande quantità di testo prodotto dalle piattaforme). Addestrato sul dataset di commenti del New York Times etichettati da annotatori umani, Perspective API ha processato miliardi di commenti ed è integrato in numerose piattaforme editoriali e social e ad oggi processa circa due miliardi di richieste giornaliere ed è integrato in oltre mille piattaforme, tra cui il Wall Street Journal e Reddit (Jigsaw, 2025)[6]. Questo sistema è rilevante per il nostro oggetto di studio perché le scelte epistemologiche operate a livello di rilevamento si propagano ai sistemi di risposta che su di esso si fondano. Se un classificatore definisce in modo problematico cosa costituisca "hate speech", i sistemi di counter-speech che utilizzano tale classificazione erediteranno e amplificheranno tutti gli eventuali bias. Per quanto riguarda l'asse epistemologico-ontologico, Perspective API opera una specifica traduzione ontologica: la "tossicità" viene definita come un commento scortese, irrispettoso o irragionevole che probabilmente indurrà qualcuno ad abbandonare una discussione (Jigsaw, 2017). Questa definizione costruisce l'hate speech primariamente come problema di *civiltà conversazionale* definito sulla base del rischio che qualcuno abbandoni la discussione, piuttosto che come violenza simbolica (Bourdieu, 1991) o atto performativo di subordinazione (Butler, 1997). La metrica unidimensionale del punteggio riduce qualitativamente la complessità del discorso d'odio, le sue dimensioni strutturali, storiche, contestuali, in un valore numerico. Un insulto razzista e un commento sarcastico possono ricevere punteggi simili, pur avendo nature e conseguenze radical-

[6] https://perspectiveapi.com/

ISBN Open Access: 9788835191971

mente diverse. Alcune ricerche hanno documentato un problema sistematico nei dataset utilizzati per addestrare i classificatori di hate speech. Un esempio è l'indagine Sap *et al.* (2019) che mostra come gli annotatori umani, incaricati di etichettare i contenuti come "offensivi" o "non offensivi", tendevano a giudicare negativamente i tweet scritti in African American English (AAE), una varietà linguistica dell'inglese con caratteristiche proprie, diffusa tra la popolazione afroamericana. Non perché questi tweet fossero effettivamente più offensivi, ma perché il registro linguistico dell'AAE veniva percepito come "aggressivo" o "volgare" da annotatori non familiari con quella varietà. I modelli di machine learning addestrati su questi dati hanno appreso e amplificato tale associazione: i tweet in AAE e quelli di utenti auto-identificati come afroamericani risultano avere fino al doppio delle probabilità di essere classificati come tossici. Perspective API, testato dagli autori, ha mostrato lo stesso pattern. Il caso illustra come l'apparente oggettività dei punteggi algoritmici mascheri giudizi culturalmente situati: il sistema non rileva la tossicità in sé, ma la distanza da norme comunicative dominanti. Questo risultato illustra come il processo di datificazione incorpori e amplifichi le strutture di potere esistenti.

Dal punto di vista dell'asse della rappresentazione delle comunità coinvolte, Perspective API solleva questioni fondamentali. Le comunità marginalizzate spesso impiegano forme linguistiche che possono apparire "tossiche" a sistemi addestrati su norme comunicative dominanti: l'appropriazione ironica di slur, il code-switching, le forme di solidarietà espresse attraverso linguaggio apparentemente aggressivo. Le pratiche di auto-rappresentazione online includono strategie retoriche, provocazione, ironia, satira, che un classificatore automatico di tossicità potrebbe erroneamente segnalare. Il sistema così rischia di disciplinare non solo l'hate speech ma anche le forme di resistenza delle comunità che intende proteggere.

4.2. Costruzione di risorse per l'automazione: CONAN e il nichesourcing

Il progetto CONAN (COunter NArratives through Nichesourcing), sviluppato da Chung *et al.* (2019) presso la Fondazione Bruno Kessler, rappresenta un approccio radicalmente diverso: piuttosto che classificare contenuti, mira a costruire le risorse linguistiche necessarie per addestrare sistemi di generazione automatica di counter-narratives. L'innovazione metodologica consiste nel nichesourcing che abbiamo già citato, ossia la raccolta di dati non da crowd generici ma da esperti di settore, in questo caso, operatori di ONG specializzate nel contrasto all'hate speech.

ISBN Open Access: 9788835191971

Il dataset CONAN originale comprende 4.078 coppie hate speech/counter-narrative in inglese, focalizzate su contenuti islamofobici. Multi-Target CONAN (Fanton *et al.*, 2021) ha esteso questa risorsa a 5.003 coppie coprendo target multipli (musulmani, ebrei, persone LGBT, persone di colore, migranti, donne, persone con disabilità, Rom) in tre lingue (inglese, francese, italiano). L'approccio human-in-the-loop prevede che esperti umani rivedano e correggano le counter-narratives generate automaticamente, creando un ciclo di miglioramento iterativo.

CONAN è un esempio di come i sistemi di counter-speech non si limitino a rispondere all'hate speech ma partecipino attivamente alla definizione di cosa costituisca una risposta appropriata. Il processo di costruzione del dataset incorpora decisioni normative fondamentali: quali strategie retoriche privilegiare? La tipologia CONAN distingue tra "facts" (correzione fattuale), "denouncing" (denuncia morale), "empathy" (appello emotivo), "humor" (risposta ironica), "warning of consequences" (avvertimento sulle conseguenze), "positive" (valorizzazione del gruppo target). Questa tassonomia codifica una specifica strategia del cambiamento, l'assunzione che determinate strategie narrative siano più efficaci di altre nel contrastare l'hate speech, che viene poi "congelata" nel dataset e propagata ai sistemi che su esso vengono addestrati.

Il paradosso dell'expertise: il nichesourcing solleva una tensione fondamentale rispetto all'asse della rappresentazione comunitaria. Da un lato, il coinvolgimento di esperti di ONG garantisce qualità e appropriatezza delle risposte, evitando i problemi del crowdsourcing generico. Dall'altro, gli "esperti" sono tipicamente esterni alle comunità target: professionisti del settore non-profit, spesso appartenenti a gruppi maggioritari. Anche in questo caso CONAN incorpora un'autorità epistemica specifica: quella dell'esperto professionale, piuttosto che quella della comunità situata.

4.3. Generazione automatica: Large Language Models e il campo di forze del counter-speech

L'emergere dei Large Language Models (LLM) come GPT-4 e ChatGPT ha aperto nuove possibilità per la generazione automatica di counter-speech. A differenza dei sistemi basati su retrieval (che selezionano risposte da un database predefinito) o su template (che completano strutture fisse), gli LLM possono generare risposte contestualizzate nel discorso, fluenti e apparentemente "personalizzate" a specifici contenuti d'odio. La ricerca di Saha *et al.* (2024) ha valutato le capacità di ChatGPT (GPT-3.5) nella generazione zero-

ISBN Open Access: 9788835191971

shot[7] di counter-speech, confrontandolo con modelli precedenti (GPT-2, DialoGPT, FlanT5) sui dataset CONAN, Reddit e Gab. I risultati presentano miglioramenti sostanziali di ChatGPT rispetto ai modelli precedenti. Particolarmente rilevante l'aumento della qualità argomentativa (+27%) e della qualità complessiva del counter-speech (+120%) rispetto agli altri sistemi presi in considerazione. Tuttavia, emerge una criticità: la leggibilità diminuisce del 35%: la sofisticazione linguistica dei modelli più avanzati può compromettere l'accessibilità delle risposte per i destinatari.

Un risultato particolarmente problematico riguarda la tossicità emergente: Saha *et al.* (2024) documentano che i modelli di dimensioni maggiori generano counter-speech più tossico, con incrementi tra il 25% e il 44% rispetto ai modelli più piccoli della stessa famiglia. Questo fenomeno controintuitivo, ossia sistemi progettati per contrastare l'odio che producono risposte più tossiche, solleva interrogativi sulla scalabilità dell'approccio LLM e sulla relazione tra capacità generativa e controllo etico. L'introduzione dei modelli linguistici pre-addestrati ha modificato profondamente i sistemi di counter-speech, rendendo possibile la generazione automatica di risposte fluenti e contestualmente plausibili senza il ricorso a dataset specializzati. Studi comparativi mostrano che questi modelli producono contronarrazioni linguisticamente più articolate rispetto ai sistemi precedenti (Tekiroğlu *et al.*, 2022). Tuttavia, le evidenze sull'efficacia sociale di tali risposte sono più controverse.

Un esperimento sul campo condotto su Twitter/X da Bär *et al.* (2024) mostra che il counter-speech generato automaticamente risulta efficace solo in forme limitate: risposte non personalizzate e basate sull'avvertimento delle conseguenze riducono la probabilità di reiterazione dell'hate speech. Al contrario, le risposte personalizzate ed empatiche generate da modelli linguistici di grandi dimensioni si sono dimostrate inefficaci o addirittura controproducenti, producendo in alcuni casi un aumento della tossicità dei messaggi successivi. Questi risultati mettono in discussione l'idea che la personalizzazione algoritmica e la simulazione dell'empatia costituiscano, di per sé, strategie efficaci di contrasto all'odio online e sembrano confermare le preoccupazioni sollevate dalla teoria della governamentalità algoritmica (Rouvroy, 2013): l'automazione del counter-speech non solo rischia di essere inefficace, ma può produrre effetti perversi precisamente quando tenta di simulare la personalizzazione e l'empatia umana.

Dal punto di vista della performatività, questi risultati sollevano questioni fondamentali. Se il counter-speech opera, come suggerisce Gelber (2002),

[7] Zero-shot significa che il modello opera senza training supervisionato ad hoc sul compito specifico. Non ha visto, in fase di addestramento, esempi espliciti del tipo "questo è counter-speech / questo non lo è".

ISBN Open Access: 9788835191971

contestando l'autorità simbolica del parlante d'odio, questa autorità può essere efficacemente contestata da un'entità non-umana? La forza illocutoria di un atto linguistico dipende, seguendo Austin (1962) e Butler (1997), dalle condizioni di felicità che includono l'identità e la posizione del parlante. Un algoritmo può performare l'atto di contestare l'odio, o la sua non-umanità lo priva della forza performativa necessaria? La risposta sarebbe intuitivamente no, ma da una prospettiva STS, tuttavia, la questione non è riducibile alla dicotomia umano/non-umano: la performatività del counter-speech emerge come proprietà relazionale, co-prodotta dall'architettura della piattaforma, dalle norme di governance e dagli immaginari sociotecnici che attribuiscono, o negano, legittimità all'intervento algoritmico nell'ambito di un frame culturale, politico ed etico.

4.4. Collaborazione umano-macchina: CounterQuill e la preservazione dell'agency

CounterQuill, sviluppato da Ding *et al.* (2024), rappresenta un paradigma alternativo che tenta di risolvere le tensioni identificate nelle logiche precedenti. Piuttosto che sostituire l'agency umana con l'automazione, CounterQuill si configura come sistema collaborativo che supporta e educa gli utenti nella scrittura di counter-speech empatico, preservando la loro voce personale e sviluppando le loro competenze.

Il sistema guida l'utente attraverso un processo in tre fasi: una fase iniziale di apprendimento per comprendere come funziona l'hate speech e quali strategie di risposta siano possibili; una fase di riflessione e ideazione sul caso specifico; e una fase finale di co-scrittura, in cui la risposta viene affinata preservando la voce personale dell'utente.

A differenza di ChatGPT, che genera risposte complete che l'utente può solo accettare o rifiutare, CounterQuill guida l'utente attraverso un processo strutturato che mantiene l'authorship umana. È significativo che entrambi i sistemi utilizzino lo stesso modello sottostante (GPT-3.5): la differenza cruciale non risiede nella tecnologia di base, ma nell'architettura dell'interazione e nella distribuzione dell'agency tra umano e macchina.

Uno studio condotto da Ding *et al.* (2024), con 20 partecipanti, finalizzato a confrontare CounterQuill con ChatGPT, ha prodotto risultati significativi per la nostra analisi. Gli utenti di CounterQuill hanno riportato un senso di ownership significativamente maggiore sul counter-speech co-autorizzato, percependo il sistema come "partner di scrittura" piuttosto che come strumento di automazione. Crucialmente, gli utenti erano più disposti a pubblicare online le risposte co-scritte con CounterQuill rispetto a quelle generate

ISBN Open Access: 9788835191971

da ChatGPT. Questo risultato suggerisce che la preservazione dell'agency umana non è solo un valore normativo ma una condizione di efficacia pratica: le persone sono più propense ad agire quando sentono che le azioni sono genuinamente loro.

CounterQuill esemplifica un diverso immaginario sociotecnico: non l'automazione che sostituisce l'azione umana, ma la tecnologia come dispositivo di supporto che potenzia le capacità umane preservandone l'agency. Questo modello risuona con l'approccio delle capabilities di Gelber (2002): piuttosto che produrre counter-speech "per" le persone, il sistema costruisce le capabilities necessarie perché le persone producano counter-speech autonomamente. Dal punto di vista della governamentalità algoritmica, CounterQuill rappresenta un tentativo di re-introdurre il momento della riflessività che i sistemi puramente automatici tendono a bypassare.

Tuttavia, anche questo modello presenta limiti dal punto di vista dell'asse della rappresentazione comunitaria. CounterQuill rimane uno strumento progettato da ricercatori esterni alle comunità target, che incorpora una specifica teoria del "buon" counter-speech. Le strategie che insegna, empatia, factual correction, reframing, potrebbero non corrispondere alle pratiche di resistenza che le comunità sviluppano autonomamente. L'"educazione" implicita nel sistema veicola norme comunicative che potrebbero disciplinare le forme di espressione delle comunità marginalizzate, anche nel tentativo di supportarle.

4.5. Il counter-speech come intervento sintomatico

Dopo l'analisi presentata possiamo ora tentare di rispondere in modo più "solido" alla domanda posta all'inizio di questo capitolo. I sistemi di counter-speech automatizzato intervengono al livello del sintomo, le singole manifestazioni discorsive dell'odio, piuttosto che delle cause strutturali che determinano la sua diffusione online. L'analisi del capitolo precedente ha mostrato che la percezione dell'hate speech dipende dall'architettura delle piattaforme, dalla presenza di gruppi chiusi, dalle caratteristiche demografiche delle basi utenti: fattori su cui il counter-speech, automatico o meno, non può intervenire direttamente.

Questo non significa che il counter-speech sia inutile. La ricerca di Bär *et al.* (2024) mostra che forme specifiche di counter-speech, non-contestualizzato, basato su warning, possono effettivamente ridurre la produzione di contenuti odiosi a livello individuale. Tuttavia, l'esperimento documenta anche che l'efficacia non aumenta, e può diminuire, con la "sofisticazione" tecnologica: il counter-speech generato da LLM, contestualizzato e personaliz-

ISBN Open Access: 9788835191971

zato, performa peggio delle risposte più semplici. Questo risultato suggerisce che l'immaginario della scalabilità attraverso l'automazione intelligente potrebbe essere fondamentalmente mal orientato.

Dal punto di vista della co-produzione STS, i sistemi di counter-speech automatizzato co-producono una specifica comprensione dell'hate speech: come problema essenzialmente linguistico, trattabile attraverso interventi testuali, decontestualizzato dalle strutture di potere che lo generano. Questa comprensione non è falsa poiché l'hate speech è anche un fenomeno linguistico, ma è parziale: ignora le dimensioni strutturali, relazionali e materiali che determinano la sua diffusione e il suo impatto. L'immaginario sociotecnico che anima questi sviluppi configura una soluzione tecnica per un problema che è, in ultima istanza, politico.

5. Verso un approccio community-in-the-loop

L'analisi condotta suggerisce la necessità di ripensare il ruolo dell'automazione nel contrasto all'hate speech. Il paradigma emergente del human-in-the-loop, esemplificato nella nostra analisi da Multi-Target CONAN e CounterQuill, rappresenta un primo passo importante, ma non sufficiente. Il problema non è solo mantenere l'umano "nel loop" dell'automazione, ma interrogarsi su quale umano e con quale autorità. I sistemi esistenti includono tipicamente ricercatori, sviluppatori e, nel migliore dei casi, professionisti di ONG: raramente le comunità marginalizzate che sono il target primario dell'hate speech e che sviluppano le strategie più sofisticate e contestualmente appropriate di resistenza.

Un modello alternativo è offerto dalle strategie di auto-rappresentazione documentate per le comunità target online (Stroud & Cox, 2018). Queste strategie di resistenza, come l'articolazione dell'identità attraverso temi positivi dell'in-group piuttosto che attraverso la costruzione dell'Altro, emergono dalle pratiche situate delle comunità stesse, non da esperti esterni, e rappresentano una forma di counter-narrazione che i sistemi automatizzati, focalizzati sulla "risposta" puntuale all'hate speech singolo, faticano a catturare e riprodurre.

Un autentico approccio community-in-the-loop richiederebbe di coinvolgere le comunità target non come validatori di contenuti prodotti altrove, ma come co-progettisti dei sistemi stessi: nella definizione di cosa costituisca hate speech nel loro contesto specifico, nella identificazione delle strategie di risposta culturalmente appropriate, nella valutazione dell'efficacia degli interventi. Questo approccio implicherebbe un ripensamento radicale dell'immaginario sociotecnico dominante: dalla scalabilità attraverso l'automazione alla costruzione di capacità attraverso la partecipazione; dalla risposta

ISBN Open Access: 9788835191971

standardizzata al supporto per l'autorappresentazione; dalla sostituzione dell'agency umana al suo potenziamento.

La sfida per la ricerca futura è sviluppare metodologie e tecnologie che supportino questo orientamento. CounterQuill indica una direzione promettente, il sistema come scaffolding educativo piuttosto che come sostituto, ma rimane progettato dall'esterno. Sistemi genuinamente community-centered richiederebbero processi di co-design che pongano le comunità al centro fin dalla fase di progettazione, incorporando le loro epistemologie, pratiche e valori. In termini STS, ciò significa passare dalla co-produzione di soluzioni tecniche alla co-produzione di problematizzazioni: non dare per scontato cosa sia l'hate speech e quale sia la risposta appropriata, ma fare di queste stesse definizioni l'oggetto di un processo partecipativo.

Questa transizione non elimina le tensioni identificate, considerando che ogni intervento incorpora visioni normative, ogni tecnologia distribuisce potere, ma le rende oggetto di negoziazione esplicita piuttosto che di decisioni tecniche opache. Come suggerisce il framework della co-produzione, i modi di conoscere il mondo sono inseparabili dai modi di organizzarlo: sistemi di counter-speech che incorporano diverse epistemologie e includono diverse voci parteciperebbero alla produzione di un ordine sociale diverso. Non una soluzione definitiva, ma l'apertura di uno spazio di possibilità che l'automazione, lasciata a sé stessa, tende sistematicamente a chiudere.

Riferimenti bibliografici

ADL (2021), *Online hate and harassment: The American Experience 2021*. Anti-Defamation League. Disponibile al sito: https://www.adl.org/resources/report/online-hate-and-harassment-american-experience-2021.

ADL (2024), *Online Hate and Harassment: The American Experience 2024* (Sixth Annual Report). Anti-Defamation League. Disponibile al sito: www.adl.org/resources/report/online-hate-and-harassment-american-experience-2024.

Austin, J.L. (1962), *How to Do Things with Words*, Oxford University Press, Oxford.

Bakshy, E., Messing, S. and Adamic, L. A. (2015), Exposure to ideologically diverse news and opinion on Facebook, *Science*, 348, 6241: 1130-1132. DOI: 10.1126/science.aaa1160.

Bar, D., Maarouf, A. and Feuerriegel, S. (2024), *Generative AI may backfire for counterspeech*. arXiv preprint. Disponibile al sito: https://arxiv.org/abs/2411.14986.

Benesch, S. (2014), *Countering dangerous speech: New ideas for genocide prevention*. Harvard Kennedy School.

Benesch, S., Ruths, D., Dillon, K.P., Saleem, H.M. and Wright, L. (2016), *Counterspeech on Twitter: A field study*. Dangerous Speech Project.

ISBN Open Access: 9788835191971

Binns, R. (2017), Algorithmic accountability and public reason, *Philosophy & Technology*, 31, 4: 543-556. DOI: 10.1007/s13347-017-0263-5.
Bonacchi, C. and Krzyzanska, M. (2021), Heritage-based tribalism in Big Data ecologies: Deploying origin myths for antagonistic othering, *Big Data & Society*, 8, 1: 1-16. DOI: 10.1177/20539517211003310.
Bourdieu, P. (1991), *Language and Symbolic Power*, Harvard University Press, Cambridge.
Boyd, D. (2011), Social network sites as networked publics. In Papacharissi Z., ed., *A Networked Self* (pp. 39-58), Routledge, New York.
Brown, A. (2017), What is hate speech? Part 1: The myth of hate, *Law and Philosophy*, 36, 4: 419-468. DOI: 10.1007/s10982-017-9297-1.
Brown, A. and Sinclair, A. (2019), *The Politics of Hate Speech Laws*, Routledge, New York.
Bucher, T. (2018), *If... Then: Algorithmic Power and Politics*, Oxford University Press, Oxford.
Buerger, C. (2019), *Combating hate speech through counterspeech*. Berkman Klein Center, Harvard University. Disponibile al sito: https://cyber.harvard.edu/story/2019-08/combating-hate-speech-through-counterspeech.
Butler, J. (1997), *Excitable Speech: A Politics of the Performative*. Routledge, New York.
Chung, Y.L., Kuzmenko, E., Tekiroğlu, S.S. and Guerini, M. (2019), CONAN – COunter NArratives through Nichesourcing: A multilingual dataset of responses to fight online hate speech, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2819-2829). Disponibile al sito: https://aclanthology.org/P19-1271/.
Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. and Starnini, M. (2021), Dynamics of online hate and misinformation, *Scientific Reports*, 11: 22083. DOI: 10.1038/s41598-021-01487-w.
Dangerous Speech Project (2023), *What is counterspeech?*. Disponibile al sito: https://www.dangerousspeech.org/counterspeech.
Daniels, J. (2013), Race and racism in Internet studies: A review and critique, *New Media & Society*, 15, 5: 695-719. DOI: 10.1177/1461444812462849.
Delgado, R. (2004), *Understanding Words That Wound*, Westview Press, New York.
Ding, Q., Ding, D., Wang, Y., Guan, C. and Ding, B. (2024), Unraveling the landscape of large language models: a systematic review and future perspectives, *Journal of Electronic Business & Digital Economics*, *3*, 1: 3-19. DOI: 10.1108/JEBDE-08-2023-0015.
Eubanks, V. (2018), *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York.
Fanton, M., Bonaldi, H., Tekiroğlu, S.S. and Guerini, M. (2021), Human-in-the-loop for data collection: a multi-target counter narrative dataset to fight online hate speech, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (pp. 3226-3240). Disponibile al sito: https://aclanthology.org/2021.acl-long.250/.

ISBN Open Access: 9788835191971

Gagliardone, I., Gal, D., Alves, T. and Martinez, G. (2015), *Countering online hate speech*. UNESCO Publishing, Paris.
Gelber, K. (2002), *Speaking Back: The Free Speech Versus Hate Speech Debate*, John Benjamins Publishing, Amsterdam.
Gillespie, T. (2018), *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, Yale University Press, New Haven.
Gillespie, T. (2020), Content moderation, AI, and the question of scale, *Big Data & Society*, 7, 2: 1-12. DOI: 10.1177/20539517209432.
Gorwa, R., Binns, R. and Katzenbach, C. (2020), Algorithmic content moderation: Technical and political challenges in the automation of platform governance, *Big Data & Society*, 7, 1: 1-15. DOI: 10.1177/2053951719897945.
Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B.B., Derksen, L., Hall, A., Jochum, M., Murias Munoz, M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F. and Donnay, K. (2021), Empathy-based counterspeech can reduce racist hate speech in a social media field experiment, *Proceedings of the National Academy of Sciences*, 118, 50: e2116310118. DOI: 10.1073/pnas.2116310118.
Horwitz, J. and Seetharaman, D. (2021), The Facebook Files: How Facebook's Data-Driven Culture Is Undermining Its Own Products, *The Wall Street Journal*, 12 settembre. Disponibile al sito: https://www.wsj.com/articles/the-facebook-files-11631713039.
Introna, L. (2016), Algorithms, governance, and governmentality, *Science, Technology, & Human Values*, 41, 1: 17-49. DOI: 10.1177/0162243915587360.
Ipsos-UNESCO (2023), *Survey on the impact of online disinformation and hate speech*. Disponibile al sito: https://www.ipsos.com/sites/default/files/ct/news/documents/2023-11/unesco-ipsos-online-disinformation-hate-speech.pdf.
Jasanoff, S. (2004), States of knowledge: The co-production of science and social order. In Jasanoff S., ed., *States of Knowledge: The Co-Production of Science and Social Order* (pp. 1-32). Routledge, New York.
Jasanoff, S. and Kim, S. H. (2009), Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea, *Minerva*, 47, 2: 119-141. DOI: 10.1007/s11024-009-9124-4.
Keipi, T., Näsi, M., Oksanen, A. & Räsänen, P. (2016), *Online Hate and Harmful Content: Cross-National Perspectives*, Routledge, London.
Lepoutre, M. (2017), Hate speech in public discourse: A pessimistic defense of counterspeech, *Social Theory and Practice*, 43, 4: 851-883. DOI: 10.5840/soctheorpract201711125.
Marwick, A. and Lewis, R. (2017), *Media manipulation and disinformation online*. Data & Society Research Institute.
Massanari, A. (2017), #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures, *New Media & Society*, 19, 3: 329-346. DOI: 10.1177/1461444815608807.

ISBN Open Access: 9788835191971

Matamoros-Fernández, A. (2017), Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube, *Information, Communication & Society*, 20, 6: 930-946. DOI: 10.1080/1369118X.2017.1293130.

Matamoros-Fernández, A. and Farkas, J. (2021), Racism, hate speech, and social media: A systematic review and critique, *Television & New Media*, 22, 2: 205-224. DOI: 10.1177/1527476420982230.

Matsuda, M., Lawrence, C. R., Delgado, R. and Crenshaw, K. W. (1993), *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*. Boulder: Westview Press.

Milan, S. and Treré, E. (2019), Big data from the South(s): Beyond data universalism, *Television & New Media*, 20, 4: 319-335. DOI: 10.1177/1527476419837739.

Mill, J. S. (1859), *On Liberty*, John W. Parker and Son, London.

Munger, K. (2017), Tweetment effects on the tweeted: Experimentally reducing racist harassment, *Political Behavior*, 39, 3: 629-649. DOI: 10.1007/s11109-016-9373-5.

Munn, L. (2020), Angry by design: toxic communication and technical architectures, *Humanities and Social Sciences Communications*, 7, 53: 1-11. DOI: 10.1057/s41599-020-00550-7.

Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E. and Oksanen, A. (2015), Exposure to online hate material and social trust among Finnish youth, *Information Technology & People*, 28, 3: 607-628. DOI: 10.1108/ITP-09-2014-0198

O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Phillips, W. (2015), *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. Cambridge: MIT Press.

Pohjonen, M. and Udupa, S. (2017), Extreme speech online: An anthropological critique of hate speech debates, *International Journal of Communication*, 11: 1173-1191.

Ribeiro, M.H., Ottoni, R., West, R., Almeida, V.A.F., Meira, W. Jr. (2020). Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT** 2020), Barcelona, Spain, January 27-30 (pp. 131-141), ACM, New York.

Reichelmann, A., Zick, A., Glaser, M., Küpper, B. (2021), Hate speech and hate crime victimization among young people: Prevalence, risk factors, and consequences, *Journal of Interpersonal Violence*, 36: 23–24). DOI: 10.1177/0886260520909199.

Rouvroy, A. (2013), The end(s) of critique: Data-behaviourism versus due process. In Hildebrandt M. and de Vries K., eds., *Privacy, Due Process and the Computational Turn* (pp. 143-167). Routledge, New York.

Rouvroy, A. (2013), The end(s) of critique: Data behaviourism versus due process. In Hildebrandt M. and de Vries K., eds., *Privacy Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology* (pp. 143-167). Taylor & Francis, London.

ISBN Open Access: 9788835191971

Rouvroy, A. and Berns, T. (2013), Gouvernementalité algorithmique et perspectives d'émancipation, *Réseaux*, 177, 1: 163-196. DOI: 10.3917/res.177.0163 .
Saha, P., Agrawal, A., Jana, A., Biemann, C. and Mukherjee, A. (2024), On zero-shot counterspeech generation by LLMs, *Proceedings of LREC-COLING 2024* (pp. 12443-12454).
Sap, M., Card, D., Gabriel, S., Choi, Y. and Smith, N.A. (2019), The risk of racial bias in hate speech detection, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1168-1678). Disponibile al sito: https://aclanthology.org/P19-1163/.
Sellars, A. (2016), *Defining Hate Speech*. Berkman Klein Center Research Publication No. 2016-20, Harvard University. Disponibile al sito: https://cyber.harvard.edu/publications/2016/DefiningHateSpeech.
Soral, W., *et al.* (2018), Exposure to hate speech increases prejudice through desensitization, *Aggressive Behavior,* 44(2): 136-146. DOI: 10.1002/ab.21713.
Strossen, N. (2018), *HATE: Why We Should Resist It with Free Speech, Not Censorship*, Oxford University Press, New York.
Stroud, S. R. and Cox, W. (2018), The varieties of feminist counterspeech in the misogynistic online world. In Segrave M. and Vitis L., eds., *Gender, Technology and Violence* (pp. 210-225). Routledge, New York.
Tekiroğlu, S., Bonaldi, H., Fanton, M. and Guerini, M. (2022), Using Pre-Trained Language Models for Producing Counter Narratives Against Hate Speech: a Comparative Study, *arXiv e-prints*, *arXiv-2204*.
Tontodimamma, A., Nissi, E., Sarra, A. and Fontanella, L. (2021), Thirty years of research into hate speech: Topics of interest and their evolution, *Scientometrics*, 126, 1: 157-179. DOI: 10.1007/s11192-020-03737-6.
Wright, L., Ruths, D., Dillon, K. P., Saleem, H. M. and Benesch, S. (2017), Vectors for counterspeech on Twitter, *Proceedings of the First Workshop on Abusive Language Online*: 57-62. Disponibile al sito: https://aclanthology.org/W17-3009/.
Yeung, K. (2017), 'Hypernudge': Big Data as a mode of regulation by design, *Information, Communication & Society*, 20, 1: 118-136. DOI: 10.1080/1369118X.2016.1186713.

ISBN Open Access: 9788835191971

ISBN Open Access: 9788835191971

ISBN Open Access: 9788835191971

COMPUTATIONAL SOCIAL SCIENCE

I social media hanno trasformato radicalmente la comunicazione pubblica, favorendo nuove forme di partecipazione democratica ma, al contempo, amplificando la diffusione del razzismo e della xenofobia con un'intensità senza precedenti. Il volume analizza il fenomeno dell'odio online da una prospettiva autenticamente interdisciplinare, che integra sociologia, linguistica computazionale, statistica e metodologia della ricerca sociale. Il lavoro si sviluppa, attraverso 13 contributi, lungo tre direttrici complementari: l'analisi delle grammatiche del razzismo contemporaneo e delle sue radici storiche; lo sviluppo di strumenti per il rilevamento automatico del linguaggio d'odio, calibrati sulle specificità del contesto italiano; una serie di studi empirici che esplorano la rappresentazione mediatica della migrazione, dell'antisemitismo e dell'antiziganismo negli spazi digitali. Un filo rosso attraversa l'intera opera: l'attenzione alle strategie di contrasto, dai sistemi automatizzati di counter-speech alle pratiche comunitarie di resistenza simbolica. Configurandosi al tempo stesso come sintesi dello stato dell'arte e come strumento operativo, il volume si rivolge a ricercatori, decisori politici, educatori e professionisti dell'informazione, offrendo risorse teoriche e metodologiche per comprendere, monitorare e contrastare un fenomeno che mette in discussione i fondamenti stessi della convivenza democratica.

Giuseppe Giordano è professore associato di Statistica sociale presso l'Università di Salerno.

Mara Maretti è professoressa ordinaria di Sociologia presso l'Università "G. d'Annunzio" di Chieti-Pescara.

Michelangelo Misuraca è professore associato di Statistica sociale presso l'Università di Salerno.

Giuseppina Damiana Costanzo è professoressa ordinaria di Statistica presso l'Università della Calabria.

FrancoAngeli
La passione per le conoscenze