



Università degli Studi di Padova

---

DIPARTIMENTO DI PSICOLOGIA DELLO SVILUPPO E DELLA  
SOCIALIZZAZIONE

*TESI MAGISTRALE IN PsICOLOGIA SOCIALE*

Haters gonna (make you) hate: il Negative  
Campaign come predittore dell'hate speech online  
nelle elezioni europee del 2019 in Italia.

*RELATRICE*  
PROFESSA CATERINA SUITNER  
UNIVERSITÀ DI PADOVA

*Co-RELATRICE*  
PROFESSA LUCIANA CARRARO  
UNIVERSITÀ DI PADOVA

*AUTORE DELLA TESI*  
SALVATORE ROMANO



A CHI LOTTA CONTRO L'OPPRESSIONE.



# Abstract

Questo articolo indaga il Negative Campaign come predittore dell'hate speech online. A partire dal database elaborato da Amnesty International per lo studio "Il barometro dell'odio" realizzato in occasione delle elezioni europee del 2019, in cui sono stati codificati 78mila commenti / re-tweet in base del livello di odio presente nel testo, sono stati codificati ulteriori 10mila post / tweet di politici sulla base del tipo di campagna utilizzata. I risultati mostrano come effettivamente ci sia una relazione significativa tra questi due fenomeni, in particolare quando il target degli attacchi è rivolto a privati cittadini e a categorie di persone che non hanno nessuna rilevanza né in ambito politico né in quello più generalmente pubblico viene generata la maggior parte dei commenti di *hate speech*. I livelli di *negative campaign* vengono attestati su percentuali simili a quelle riscontrare in altri studi condotti in paesi con un alta frammentazione degli schieramenti politici. Chi utilizza di più questo linguaggio risiede all'opposizione dimostrando come questo tipo di comunicazione dipenda più dalla posizione istituzionale (governo o opposizione) che dalla fazione politica (sinistra o destra). Confrontando i dati relativi ai due social network analizzati (Facebook e Twitter) emerge come le campagne negative siano più frequenti su Twitter che su Facebook, in relazione alla lunghezza media dei contenuti che risultano più brevi quando contengono odio.



# Indice

## ABSTRACT

v

1	INTRODUZIONE	3
1.1	Il contesto storico . . . . .	3
1.2	I contenuti della tesi . . . . .	7
2	LA CAMPAGNA POLITICA NEGATIVA	9
2.1	La definizione . . . . .	9
2.2	Fenomeno in aumento? . . . . .	12
2.2.1	Nel mondo . . . . .	12
2.2.2	In Italia . . . . .	15
2.3	Bias di negatività . . . . .	16
2.4	Cosa spinge a usare questo linguaggio nel contesto politico? . . . . .	17
2.4.1	Negatività e medium . . . . .	18
2.4.2	Negatività e contesto elettorale . . . . .	22
2.4.3	Negatività ed efficacia nelle campagne politiche . . . . .	23
3	HATE SPEECH	27
3.1	Libertà di parola o linguaggi d'odio? . . . . .	27
3.2	La definizione . . . . .	29
3.3	Discorsi d'odio e crimini d'odio . . . . .	34
3.4	Discorsi d'odio online . . . . .	37
3.5	Discorsi d'odio e caratteristiche personali . . . . .	38
3.6	Discorsi d'odio e politica . . . . .	44
4	LA PRESENTE RICERCA	45
4.1	Studio esplorativo . . . . .	45
4.2	Le nuove direzioni di sviluppo della ricerca . . . . .	46
4.3	Le ipotesi . . . . .	48

<b>5 METODOLOGIA</b>	<b>51</b>
<b>5.1 Il Barometro dell’Odio . . . . .</b>	<b>51</b>
<b>5.1.1 Amnesty International e il Tavolo per il contrasto ai discorsi d’odio . . . . .</b>	<b>51</b>
<b>5.1.2 La raccolta dati . . . . .</b>	<b>53</b>
<b>5.1.3 La valutazione dei contenuti . . . . .</b>	<b>54</b>
<b>5.1.4 Le elezioni europee del 2019 . . . . .</b>	<b>55</b>
<b>5.1.5 I risultati del Barometro . . . . .</b>	<b>55</b>
<b>5.2 La codifica del <i>Negative Campaign</i> . . . . .</b>	<b>58</b>
<b>5.3 Raccogliere <i>Big Data</i>: le criticità delle API ufficiali . . . . .</b>	<b>60</b>
<b>5.4 Analizzare <i>Big Data</i>: Digital labor e Turco Meccanico . . . . .</b>	<b>62</b>
<b>6 RISULTATI</b>	<b>67</b>
<b>6.1 Il Negative Campaign . . . . .</b>	<b>67</b>
<b>6.1.1 Ip.1: Le percentuali di campagna negativa . . . . .</b>	<b>67</b>
<b>6.1.2 Ip.2: Negative campaign: chi lo usa? . . . . .</b>	<b>69</b>
<b>6.1.3 Ip.3: Il <i>negative campaign</i> in relazione al tipo di social network . . . . .</b>	<b>73</b>
<b>6.1.4 Ip.4: Il <i>negative campaign</i> contribuisce a far diventare virali i contenuti? . . . . .</b>	<b>75</b>
<b>6.2 L’odio nei commenti . . . . .</b>	<b>77</b>
<b>6.2.1 Ip.1: Il <i>negative campaign</i> genera più odio nei commenti? . . . . .</b>	<b>79</b>
<b>6.2.2 Ip.2: quale target genera più odio nei commenti? . . . . .</b>	<b>80</b>
<b>6.2.3 Ip.3: i commenti su Twitter sono più negativi? . . . . .</b>	<b>86</b>
<b>7 CONCLUSIONI</b>	<b>89</b>
<b>7.1 Limiti . . . . .</b>	<b>91</b>
<b>7.2 Futuri sviluppi . . . . .</b>	<b>93</b>
<b>BIBLIOGRAFIA</b>	<b>95</b>
<b>RINGRAZIAMENTI</b>	<b>103</b>

*At the root of all politics is the universal language of conflict.*

Schattschneider

# 1

## Introduzione

### I.1 IL CONTESTO STORICO

A partire dalle elezioni americane del 2008, la politica ha capito di dover fare i conti con i social media per rendere le proprie campagne elettorali realmente efficaci. In quell'anno Barack Obama puntò molto sul suo sito internet my.barackobama.com, sul suo profilo Myspace e sul suo profilo Facebook. All'epoca, il ventiquattrenne Chris Hughes, co-fondatore di quest'ultimo social network, era anche direttore della parte online della campagna del futuro presidente.

Da allora, negli ultimi 12 anni, tutte le campagne elettorali si sono progressivamente spostate online e sui social media. Simbolo ne sono le elezioni del 2016 in cui l'azienda Cambridge Analytica diretta da Steve Bannon ha condotto una profilazione di massa di profili Facebook (indebitamente collezionati) utilizzando una versione semplificata dei Big5.

Questa operazione che ha unito *behavioral science*, *data analytics* con il *micro-targeted advertisements*, ha avuto esiti sicuramente positivi, anche se difficilmente quantificabili ([Rathi, 2019](#)), per la vittoria di Trump. In quelle elezioni infatti hanno fatto la differenza una manciata di voti in alcuni stati chiave.

Più in generale, l'utilizzo di questo tipo di campagne basate sull'utilizzo di informazioni personali, sulla cui raccolta si fondano i social network, è diventata una pratica comune nel marketing e non stupisce che sia utilizzata anche per trasmettere contenuti politici. Questo fenomeno segue il costante aumento di persone connesse in rete in tutto il mondo. Proprio

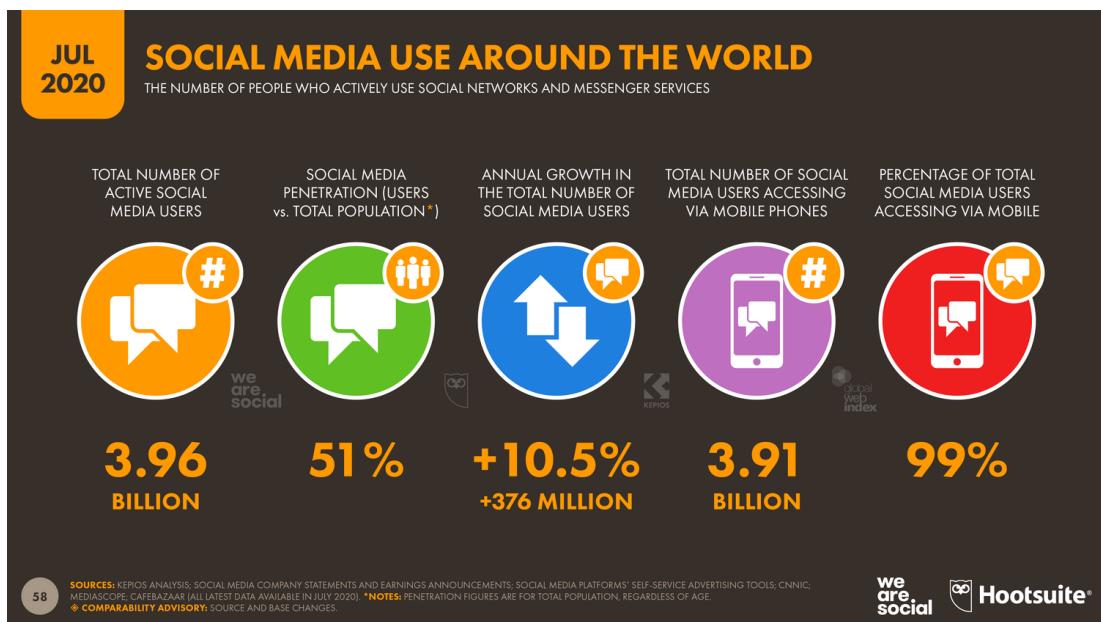
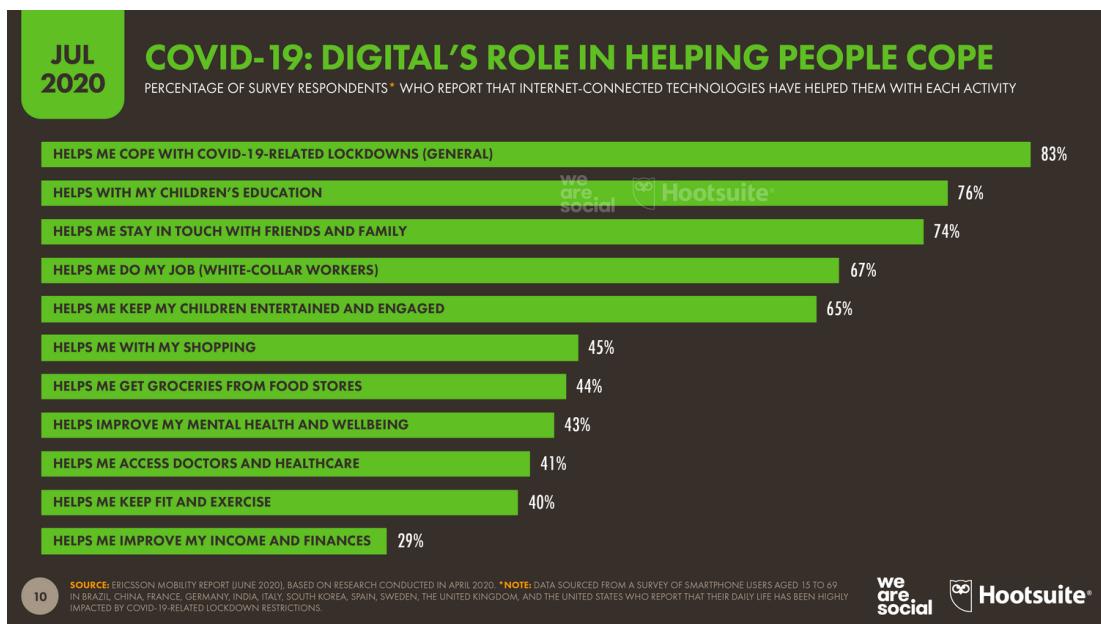


Figura 1.1: L'utilizzo dei social media aggiornato a giugno 2020. Immagine pubblicata all'interno del "Digital 2020: july global statshot"

nel 2020 il numero di persone che utilizzano i social media è arrivato a 3.96 miliardi, pari al 51% della popolazione globale (Kemp, 2020b) [Fig. 1.1], con un aumento di 160 miliardi rispetto ai dati di inizio febbraio (Kemp, 2020a). Il numero di persone che comunicano tramite social è in rapido aumento, solo quest'anno è cresciuto del 10% (376 milioni in più rispetto al 2019), e il trend non accenna a diminuire. In media le persone spendono circa due ore e mezza al giorno sui social e nel 40% dei casi lo fanno per lavoro, dimostrando quanto questi strumenti stiano diventando parte fondamentale delle nostre vite per diverse ragioni, non solo ricreative [Fig. 1.2]. L'utilizzo delle piattaforme commerciali sembra essere diventato ancora più predominante durante la recente pandemia, come dimostra uno studio (Global Web Index, 2020) secondo cui in Italia è cresciuto addirittura del 43% (in linea con l'aumento globale medio).

Più di 2 miliardi e 600 milioni di persone usano almeno una volta al mese Facebook, che si conferma la piattaforma più utilizzata al mondo, ma l'utilizzo di un social non ne esclude altri. Twitter, che conta "solo" 326 milioni di utenti attivi al mese, riscontra una sovrapposizione dei suoi utenti con quelli di Facebook dell'88% dei casi, il che mostra come diverse piattaforme, con diverse regole e stili comunicativi, si completino a vicenda svolgendo ruoli diversi e complementari. Ad esempio, benché Twitter sia uno dei social con meno utenti



**Figura 1.2:** I settori della vita in cui internet è stato più utile durante la pandemia. Immagine pubblicata all'interno del "Digital 2020: July global statshot"

attivi al mese (si colloca al sedicesimo posto), risulta essere il quinto sito più visitato sul totale del traffico globale a causa del suo utilizzo da parte dei politici per la trasmissione di comunicati stampa seguiti da persone anche non attive sulla piattaforma. La comunicazione politica viene affidata spesso a questo social di microblogging, come emerge anche dal Digital News Report 2020 (Newman, Fletcher, Schulz, Andi, & Nielsen, 2020) dell'università di Oxford, in cui il social dei cinguettii risulta essere una fonte di notizie di attualità per l'11% dei soggetti, percentuale che sommata al 46% di intervistati che invece utilizzano Facebook per cercare news ed informarsi, rende evidente l'importanza di questi nuovi canali di comunicazione.

Se da una parte i social network hanno visto un enorme sviluppo in questi anni, andando a definire aspetti importanti della vita come la comunicazione politica e l'informazione giornaliera di una fetta sempre maggiore della popolazione, dall'altra hanno generato alcuni fenomeni negativi come la diffusione di disinformazione e la crescita dei discorsi d'odio che in alcuni casi sfociano addirittura in veri e propri *hate crime* (Chan, Ghose, & Seamans, 2015). Un esempio diventato tristemente celebre è quello del *Pizza Gate*. Sempre durante la campagna elettorale americana del 2016, alcune email sottratte a John Podesta, responsabile della campagna elettorale di Hillary Clinton, vengono pubblicate da WikiLeaks. A partire da questo materiale, alcuni account Twitter di politici conservatori e personaggi dell'*alt-right*

iniziano a diffondere teorie del complotto riguardanti presunti abusi su minori e riti satanici condotti da esponenti democratici in alcuni ristoranti di Washington. Le tragiche conseguenze di questa vicenda ci mostrano come "Fake News Brought Real Guns" (Kang & Goldman, 2016). Infatti, a seguito della campagna di disinformazione online, si è generata una campagna di odio online sulle pagine delle pizzerie coinvolte che ha portato poi Edgar M. Welch, un giovane ventottenne della Carolina del Nord, a fare fuoco (offline) all'interno di una delle pizzerie al centro del falso scandalo (fortunatamente senza ferire nessun\*).

Nel corso degli ultimi anni la società civile, le grandi corporazioni che comprano spazi pubblicitari sui social media e i social media stessi hanno discusso di come contrastare l'odio online, prendendo contromisure che vanno dalla rimozione di account particolarmente problematici, alla categorizzazione automatizzata (o semi-automatizzata) dei contenuti problematici, fino alla modifica dei termini del servizio che regolano le discussioni all'interno delle varie piattaforme. Già nel 2017, un articolo pubblicato sul Wall Street Journal (Nicas, Marshall, & Vranica, 2017) segnava l'avvio della prima "Adpocalypse", termine che letteralmente significa "apocalisse delle pubblicità" e indica il boicottaggio massiccio da parte di grandi aziende delle piattaforme social, con l'obiettivo di far cambiare le regole alla base dell'organizzazione dei contenuti. Nel primo caso, a seguito della pubblicazione di alcuni *screenshots* sul Twitter del giornalista Jack Nicas, in cui pubblicità di grandi aziende come Coca-Cola venivano suggerite su Youtube in relazione a video inneggianti al razzismo se non addirittura a messaggi di propaganda di terroristi islamici vicini all'Isis, molte importanti aziende decisero di bloccare le sponsorizzazioni fino a quanto Google non fosse stata in grado di evitare questo tipo di eventi moderando meglio l'odio sulle sue piattaforme. Da allora diverse volte episodi di lobbismo contro i social media hanno puntato alla revisione delle regole di *content curation*. L'ultimo attacco in ordine temporale è avvenuto nel giugno 2020, quando l'azienda di telecomunicazioni Verizon ha iniziato a ritirare le sue pubblicità da Facebook (Paul & Hern, 2020), riprendendo un appello della campagna "Stop hate for profit" (Stop Hate For Profits, 2020). La questione si è in qualche modo risolta tramite un accordo interno alla Global Alliance for Responsible Media (GARM), di cui fa parte anche lo stesso Facebook, in cui si individuano delle linee guida per definire in modo comune l'*hate speech* e per aumentare la sicurezza digitale (World Federation Advertisers, 2020).

Un altro evento recente di estrema rilevanza riguarda l'account Twitter di Donald Trump. Lo scorso 29 Maggio, per la prima volta, il social network di microblogging ha moderato un contenuto proprio del presidente degli Stati Uniti (Tacopino, 2020). In seguito alla dichiarazione del *Tycoon* riguardo le proteste scoppiate dopo l'uccisione di Jorge Floyd, in cui

afferma che “Quando cominciano i saccheggi, si inizia a sparare”, il media statunitense ha deciso di sovrapporre al tweet una “public interest notice” in quanto l'affermazione sarebbe una “glorification of violence” non in linea con gli standard della comunità [Fig. 1.3]. Benchè l'affermazione non sia stata rimossa dalla piattaforma per la sua rilevanza pubblica, questo episodio rappresenta un punto cruciale nell'evoluzione della propaganda politica, e della sua relazione con l'odio, nel contesto online.

Se da una parte le multinazionali si ritrovano a dover fare pressioni sui social network affinchè limitino l'odio, le stesse piattaforme diventano moderatrici addirittura della massima carica politica statunitense, proprio per prevenire ulteriori episodi di violenza, dimostrando come queste piattaforme si trovino al centro, nel bene e nel male, del dibattito contemporaneo sulla libertà di parola.

A partire da tali considerazioni sulla relazione tra odio online e politica, in questa tesi si indaga la relazione tra il *negative campaign* e l' *hate speech*. L'obiettivo della ricerca è comprendere come alcuni tipi di campagna politica possano generare violenza online. Il nostro auspicio è che i risultati raggiunti possano fornire strumenti scientifici utili alla creazione di strategie efficaci per la riduzione dei livelli di odio online, anche cominciando con il far comprendere ai politici gli effetti della loro comunicazione.

## 1.2 I CONTENUTI DELLA TESI

Il presente lavoro si apre con una rassegna bibliografica della teoria del *negative campaign* (capitolo ”**La campagna politica negativa**”) e dell’*hate speech online* (capitolo ”**Hate Speech**”). Dopo aver presentato l’origine della presente ricerca (“**La presente ricerca**”) si passa poi ad analizzare nel dettaglio il ”Barometro dell’odio” sviluppato da Amnesty International durante la campagna elettorale per le elezioni europee in Italia del 2019, che fornisce il database di partenza per le ulteriori codifiche effettuate per il presente elaborato (capitolo ”**Metodologia**”). Seguiranno le analisi svolte con i principali risultati (capitolo ”**Risultati**”) e le conclusioni con le riflessioni finali (capitolo ”**Conclusioni**”).



*Negative advertising is the crack cocaine of politics.*

Senator Tom Daschle

# 2

## La campagna politica negativa

Negli ultimi tre decenni le scienze sociali hanno indagato a fondo i messaggi politici negativi. I candidati investono molto nelle strategie delle loro campagne che rappresentano la modalità principale di comunicazione con gli elettori, dagli stikers, volantini, cartelloni elettorali, alle interviste e spots sui media tradizionali, fino alle campagne online tramite contenuti sponsorizzati sulla propria pagina e quella del partito sui vari social networks.

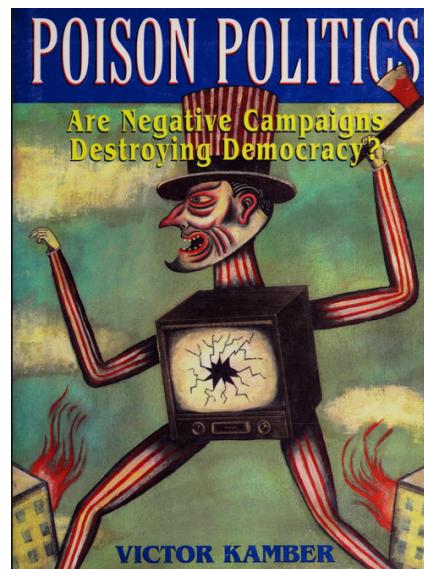
Questi messaggi elettorali possono avere un'accezione prevalentemente positiva, cioè volta a diffondere le proprie proposte e i propri punti di forza, oppure possono avere un'accezione più negativa quando si cerca di far risaltare le debolezze dell'avversario, criticandolo o contrapponendosi al suo programma politico. Su questa distinzione si genera la definizione di “campagna politica negativa”.

### 2.1 LA DEFINIZIONE

Come riportato da Walter ([Walter, 2014](#)), delineare il *negative campaign* può risultare difficile. Diverse sono le definizioni in uso all'interno della comunità accademica. Nel libro di Swint ([Swint, 1998](#)) vengono discusse molte di queste possibili interpretazioni, considerando che, a livello di opinione pubblica, molt\* faticano a convergere su un solo significato , generando, così, fraintendimenti. A riguardo viene citato uno studio di Surlin e Gordon del 1977 ([Surlin & Gordon, 1977](#)) in cui si mostra come la maggior parte degli elettori veda le campagne negative come quelle in cui è contenuto un attacco scorretto nei confronti

di un avversario. In questi casi si sottintende una definizione valutativa che fa riferimento alla illegittimità degli attacchi. Si tratta, però, di una definizione difficilmente quantificabile, che lascia molto spazio all'interpretazione soggettiva. Secondo questo approccio, le campagne negative vengono anche chiamate "dirty" oppure "poison politics". Jamieson (Jamieson, 1993) cita vari esempi in cui gli attacchi risultano palesemente falsi come quello in cui Bush ha attaccato il rivale Dukakis dicendo che era "praticamente contro ogni arma che abbiamo sviluppato". Tale affermazione ha spinto Dukakis a girare uno spot in cui guidava un carro armato M1, per togliere ogni dubbio agli elettori, risultando però ancora più debole, poiché costretto sulla difensiva, e punito in seguito dagli elettori alle urne. Con esempi come questo l'autore vuole far capire come la "dirty politics" possa risultare efficace se utilizzata intelligentemente, nonostante si possa basare anche sulla palese menzogna, strumento che dovrebbe essere responsabilmente evitato da candidati alle elezioni.

Anche Kamber (Kamber, 1997), come viene chiaramente trasmesso dalla copertina del suo libro [Fig. 2.1], si interroga su come la *poison politics* possa avere conseguenze sulla sorti della democrazia. In questo filone di discorso il *negative campaign* risulta interpretato nella sua definizione valutativa, con l'obiettivo di tracciare dei confini che il discorso politico non dovrebbe oltrepassare per evitare che possa avere conseguenze sulla pacifica convivenza. Nel libro di Kamber viene descritta la responsabilità necessaria per candidarsi a una carica pubblica, rimarcando quanto sia importante per tutta la comunità che i rappresentanti politici evitino scorrettezze in campagna elettorale, come l'utilizzo di menzogna, calunnia e attacchi a caratteristiche personali non influenti ai fini della candidatura. L'autore, noto senatore democratico statunitense e consulente per campagne elettorali dei rappresentanti di questo schieramento, quindi molto focalizzato nell'attacco di parte dei "evil Republicans", arriva a descrivere queste tattiche come "anti-politica" per gli effetti nefasti che possono avere sulla fiducia degli elettori nelle istituzioni in generale.



**Figura 2.1:** Copertina del libro di Kamber: "Poison politics: Are negative campaigns destroying democracy?", 1997

La definizione di *negative campaign* largamente più utilizzata in letteratura, soprattutto negli studi quantitativi come il nostro, risulta però essere quella direzionale che vi fa rientrare tutti i tipi di attacchi al di là della valutazione di merito, includendo anche le critiche del tutto legittime, argomentate senza insulti o toni eccessivi e necessarie al processo democratico. In base a questo approccio la campagna positiva risulta per antitesi quella in cui si propone in modo costruttivo la propria posizione e le proprie proposte senza alcun riferimento a quelle degli avversari.

Da alcuni autori (Jamieson & Waldman, 1997) (Jamieson, Waldman, & Devitt, 1998), è stata proposta una definizione direzionale tripartita che tiene conto anche delle campagne comparative come separate da quelle negative e positive. In particolare è stata proposta la divisione in "advocacy" con argomenti in favore del candidato, "attacks" con riferimenti agli avversari e "comparison" che unirebbe entrambi i primi due tipi di campagna. Questa definizione viene applicata allo studio dei singoli spot pubblicitari dei politici statunitensi durante le elezioni del '96. In questa campagna vengono riscontrate differenze molto significative tra i due politici presi in considerazione riguardo in particolare alle campagne comparative, dimostrando come questa categoria non risulti superflua, ma invece importante per discernere al meglio i messaggi elettorali negativi che citano l'avversario senza citare la propria posizione.

Lau e Pomper (Lau & Pomper, 2004) criticano questa ulteriore suddivisione adducendo come spiegazione il fatto che tutte le campagne sono intrinsecamente anche comparative, rendendo questa definizione passibile di interpretazioni soggettive. Viene quindi proposta una valutazione che spazia su un continuum in cui i due opposti sono le campagne positive e quelle negative. Questi termini vengono preferiti rispetto ai precedenti perché più esplicativi.

In tutte le accezioni di definizione, viene presa in considerazione solo la comunicazione relativa a contenuti politici. Analizzando i media tradizionali questa sottolineatura risulta superflua, dato che tutti i discorsi presi in considerazione risultano pienamente politici. Analizzando invece contenuti di social media, capita abbastanza spesso di imbattersi in esternazioni che nulla hanno a che fare con la campagna elettorale. Per questa ragione, come vedremo meglio nel capitolo "**Metodologia**", in questo studio si fa uso di una definizione direzionale con quattro categorie: campagna positiva, neutro, comparativa e negativa.

## Share of Negativity in Presidential Campaigns, 1960-2008

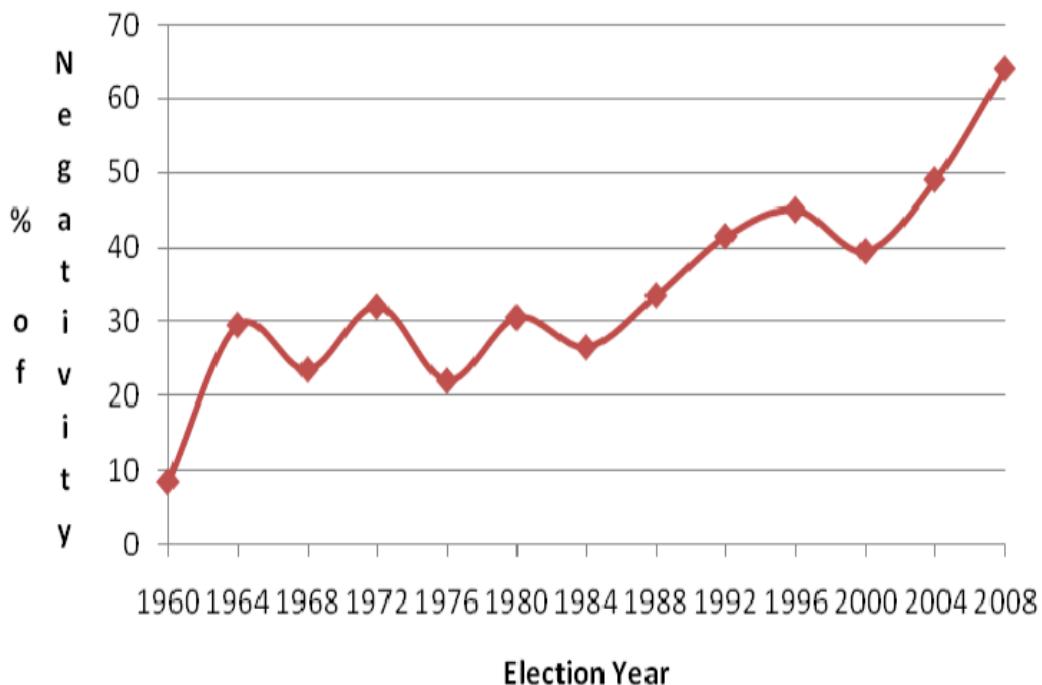


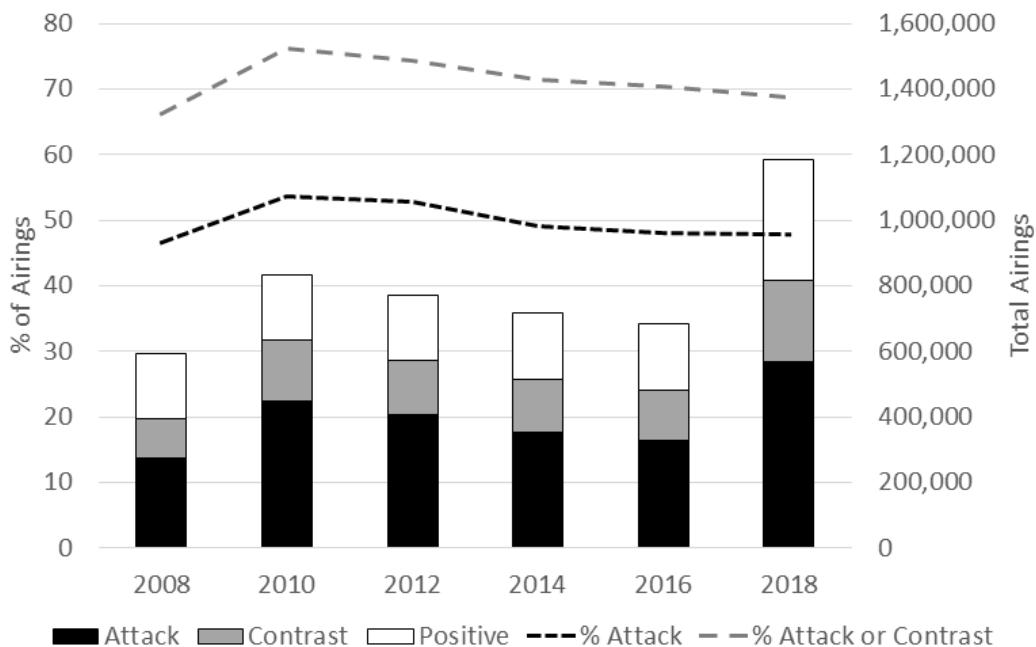
Figura 2.2: Aumento degli spot di campagna politica negativa dal 1960 al 2012. Immagine pubblicata nello studio di Geer (2012).

### 2.2 FENOMENO IN AUMENTO?

#### 2.2.1 NEL MONDO

Risulta difficile tracciare l'aumento o la diminuzione degli spot di campagna negativa nell'arco del tempo poiché cambiano i canali comunicativi e i contesti nazionali risultano difficilmente comparabili. Geer propone un'analisi retrospettiva dal 1960 al 2012 (Geer, 2012) negli Stati Uniti che ha il pregio di utilizzare una metodologia omogenea su tutto il materiale analizzato. Come si vede nel grafico riportato in questo studio [Fig. 2.2], il trend risulta in costante aumento.

Sempre sulle elezioni statunitensi, questa volta dal 2002 al 2006, prendendo in considera-

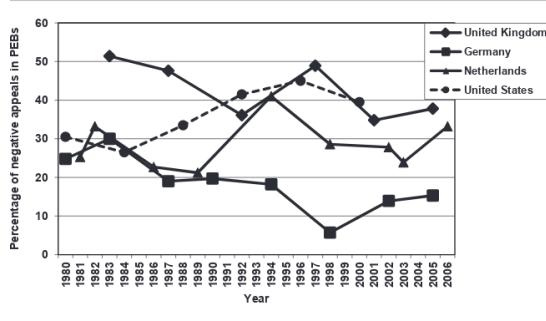


**Figura 2.3:** Volume e proporzione della negatività nelle elezioni federali americane tra il 2008 e il 2018. Kantar Media-CMAG, con analisi di Wesleyan Media Project.

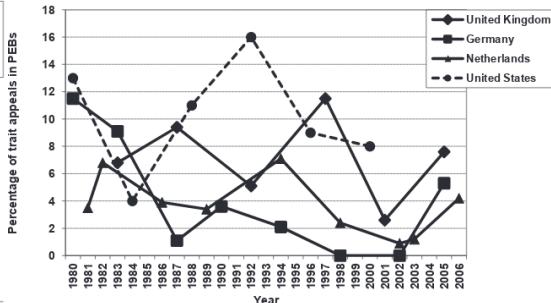
zione gli spot TV e quelli sui siti online nati in quel periodo, vengono riscontrati aumenti progressivi per ogni biennio passando dal 38% al 45% e poi addirittura al 57% nell'online. Percentuali simili sono riscontrate anche in televisione, seppur con un campione meno significativo a causa di candidati minori che non utilizzano questo mezzo perché troppo costoso (Druckman, Kifer, & Parkin, 2010).

Risultati paragonabili, con percentuali di *negative campaign* addirittura intorno al 70%, vengono riportati anche in studi più recenti (Wesleyan Media Project, 2018). Come si può vedere dal grafico presente nella ricerca del Wesleyan Media Project [Fig. 2.3], i livelli rimangono pressoché invariati negli anni presi in considerazione a partire dal 2008, quello che risulta in crescita è invece il numero di pubblicità totali realizzati da chi si è candidat\*. Il volume delle campagne risulta maggiore del 61% nel 2018 rispetto alle elezioni di *midterm* di due anni prima e alla media del decennio precedente, facendo sempre riferimento allo stesso periodo di tutti gli anni presi in considerazione (4 Settembre – 25 Ottobre).

In controtendenza è invece uno studio del 2014 di Walter (Walter, 2014) in cui si compara la negatività registrata nelle campagne elettorali negli Stati Uniti dallo studio sopracitato di Geer con quella di 23 elezioni parlamentari di Germania, UK, e Olanda, avvenute tra il 1980



**Figura 2.4:** Quantità di campagna negativa durante le elezioni parlamentari dal 1980 al 2016, sono analizzati gli spot ufficiali dei partiti in televisione. Immagine pubblicata 2016. Immagine pubblicata nello studio di Walter (2013).

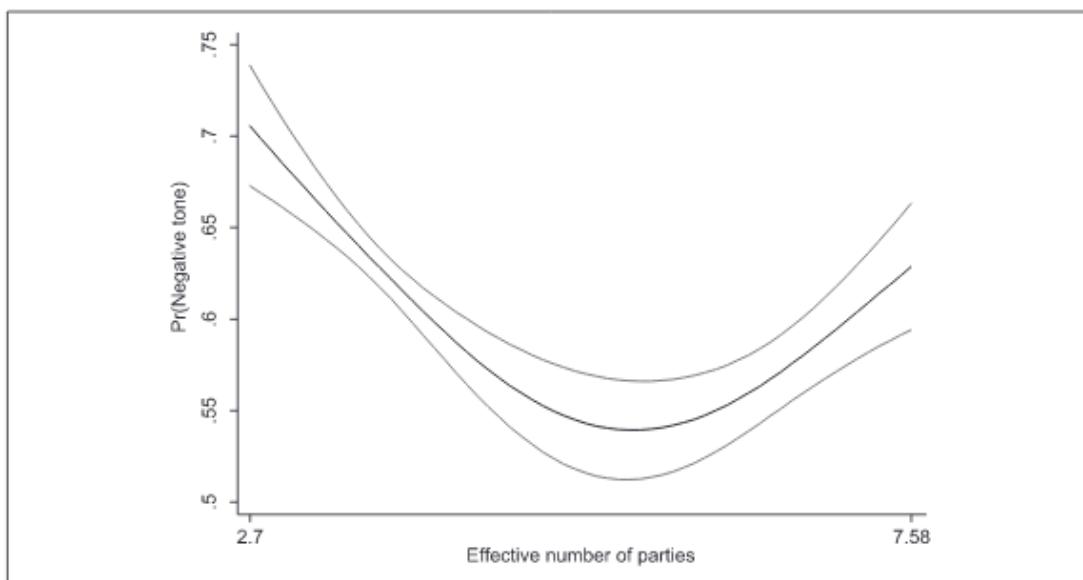


**Figura 2.5:** Quantità di campagna negativa rivolta a singoli candidati rivali nelle elezioni parlamentari dal 1980 al 2006. Immagine pubblicata nello studio di Walter (2013).

e il 2006. Da questo articolo risulta che la negatività media non sarebbe aumentata nelle campagne europee prese in considerazione [Fig. 2.4]. I livelli medi registrati sono quindi i seguenti: 43% per l'UK (il più alto), il 29% per l'Olanda e il 18% per la Germania. La media nello studio di Geer riguardo l'USA era invece del 36%. L'interpretazione dello studio quindi verte principalmente sui diversi sistemi elettorali degli stati anglofoni rispetto agli altri due stati europei: il bipolarismo tipico di questi paesi accentuerrebbe l'utilizzo di campagna negativa, cosa che invece non succederebbe in Germania e Olanda, paesi con un arco parlamentare tradizionalmente più frammentato. Come vedremo nei prossimi capitoli, anche dal nostro studio emerge una dimensione del fenomeno in linea con quella rilevata da Walter per i paesi non anglofoni e non basati sul bipolarismo.

Altrettanto interessante per le analisi che verranno affrontate in questo elaborato, in particolare riguardo il target degli attacchi, un secondo grafico [Fig. 2.5] sempre della stessa ricerca di Walter (Walter, 2014). Qui vengono illustrate le percentuali degli attacchi rivolti a un singolo candidato, arrivando sempre alla conclusione che il processo di *personalization*, quanto meno negli attacchi, non ha subito significativi incrementi negli ultimi anni.

Anche in un altro studio (Papp & Patkós, 2018) si afferma che la frammentazione dello scenario politico sia in grado di determinare il livello di negatività utilizzato dei politici in gara. Viene specificato come però, l'aumento non sia costante e lineare, ma avvenga secondo una curva a "U", in cui livelli alti di frammentazione, presupponendo la necessità di creare alleanze, riporti a una situazione simile a quella del bipolarismo, cioè a frammentazione minima. Questo studio ha analizzato gli articoli di prima pagina e il 5% di tutti gli articoli relativi alle elezioni presenti nei principali quotidiani di nove paesi europei, nel mese precedente al-



**Figura 2.6:** Relazione tra il tono delle campagne politiche e il livello di frammentazione dell'arco politico. Immagine pubblicata nello studio di Papp (2018).

le elezioni, dal 2005 al 2014. Non viene utilizzata direttamente la definizione direzionale di campagna negativa, ma viene effettuata una *content analysis* in base al tono dei politici in riferimento ad avversari tramite una scala nominale che comprende toni negativi, positivi o neutri. Anche in questo caso non viene riscontrato un aumento nel corso degli anni presi in considerazione, ma solo variazioni tra un'elezione e l'altra o tra diversi paesi (danesi e tedesche le campagne meno negative, mentre Ungheria Inghilterra e Olanda si collocano all'estremo opposto). La media generale di negatività riscontrata si attesta al 46.8%, 33.8% quella positiva e il 19.4% quella neutra. In uno dei grafici di questo studio [Fig. 2.6] possiamo vedere la relazione tra frammentazione e tono negativo. Benché non sia direttamente paragonabile con la presente ricerca, risulta interessante vedere come a una frammentazione intorno agli otto partiti in competizione (come nel caso delle elezioni italiane del 2019), la negatività si attesta intorno al 63%.

### 2.2.2 IN ITALIA

Per quanto riguarda il panorama italiano è interessante lo studio di Ceron (Ceron & Adda, 2016), in cui vengono categorizzati a mano sulla base della definizione direzionale di *negative campaign* più di 15 mila tweet di politici in occasione delle elezioni nazionali del 2013. Qui le percentuali di campagna negativa rivolta a un avversario risultano decisamente inferiori a

quelle d'oltre-oceano, attestandosi in media al 16.8%. Vengono riscontrati risultati molto più elevati nei partiti non in carica in quel momento, come anche riscontrato nel nostro studio e in generale nella letteratura. Questo punto verrà trattato in modo più approfondito in un paragrafo successivo.

Un'altra ricerca che indaga proprio le elezioni europee del 2019 in Italia è quella di Seddone ([Seddone, Bobba, & Roncarolo, 2019](#)). Viene analizzata la negatività presente in questa campagna elettorale prendendo in considerazione i media tradizionali (cinque giornali e cinque televisioni) e la loro tendenza a utilizzare un tono negativo soprattutto nelle discussioni sull'Europa. Dai risultati si evince che i giornali sono più propensi, rispetto alle televisioni, a dare risalto a messaggi negativi: in media il 39.5% dei messaggi sui giornali è stato codificato come negativo, rispetto al 32.5% delle TV. In particolare però, Tg5 e Il Giornale, entrambi di proprietà di Silvio Berlusconi, hanno fatto registrare i valori più alti in assoluto (rispettivamente 47.5% e 56.8%), lasciando ipotizzare che l'elettorato di centrodestra, uscito con risultati inferiori alle aspettative nelle elezioni del 2018, sia stato particolarmente stimolato ad andare al voto tramite attacchi agli schieramenti sia più a destra che a quelli più a sinistra. Risulta interessante inoltre che i toni più negativi siano stati riscontrati su tematiche di politica interna piuttosto che verso l'Europa. Quest'ultimo tema risulta però differentemente coper-to da giornali e TV: sugli schermi le tematiche europee appaiono con toni decisamente più negativi.

### 2.3 BIAS DI NEGATIVITÀ

Trasversale a tutte le affermazioni e posizioni sostenute negli studi sul *negative campaign*, risulta la teoria del *negative bias*: l'effetto negatività fa in modo che i messaggi negativi risultino più persuasivi rispetto a messaggi positivi. Questo effetto è dimostrato anche in uno studio sull'attenzione, in cui vengono fatti eseguire dei "color-naming task", per indagare le valutazioni automatiche ([Pratto & John, 1991](#)). In tutti e tre gli esperimenti condotti in questo studio, gli stimoli negativi ricevono in media l'85% in più di attenzione e l'effetto viene riscontrato in tutti i partecipanti. E' possibile rintracciare le cause originarie di questo comportamento in una prospettiva evoluzionistica: gli stimoli negativi vengono sovraesposti per difenderci dai pericoli di tutti i giorni.

Rozin e Royzman ([Rozin & Royzman, 2001](#)), nella loro rassegna, evidenziano come ci siano diversi effetti collegati a questo fenomeno e che non sia possibile ridurli a una teoria onnicomprensiva. Spinte a livello adattativo, evolutivo e meccanicistico contribuiscono a

creare il *negativity bias*. In generale possiamo dire che “Bad is stronger than good” ([Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001](#)). Le emozioni negative, come i feedback negativi, hanno maggiore impatto rispetto ai corrispettivi positivi poiché le informazioni a riguardo vengono elaborate in modo più approfondito.

Per quello che ci interessa in questo studio, è un risultato consolidato che gli elettori prestino maggiore attenzione ai messaggi negativi rispetto ai positivi e che la percezione di paura generata dalla negatività stimoli l’interesse nelle campagne politiche e riesca a mobilitare gli elettori più radicali.

#### **2.4 COSA SPINGE A USARE QUESTO LINGUAGGIO NEL CONTESTO POLITICO?**

I motivi che spingono a utilizzare questo tipo di linguaggio, in particolare nel contesto politico, possono essere raggruppati in tre filoni principali: l’efficacia riscontrate da messaggi negativi su media tradizionali e, più recentemente sui social media; la maggiore efficacia nel mobilitare l’elettorato da parte di questo tipo di campagna; i fattori contestuali al tipo di sistema elettorale e la generale polarizzazione del dibattito. Nei prossimi sottoparagrafi affronteremo queste tre linee di spiegazione in modo separato, nonostante siano fattori che procedono sempre di pari passo.

Un autore che indaga tutte queste possibili cause è Geer ([Geer, 2012](#)). Nel suo articolo discute tre ipotesi: la prima è che questo tipo di campagna risulti più efficace nel creare consenso attorno a un\* candidat\*. Questa spiegazione non risulta pienamente dimostrata ed è tuttora molto dibattuta, tuttavia l’autore cita questa convinzione comune che, anche se non sempre supportata da dati oggettivi, risulta trainare le strategie comunicative di molte campagne politiche.

La seconda indicazione emersa dallo studio è una forte correlazione ( $r = .88$ ) tra la polarizzazione del dibattito e il livello di negatività espresso durante la campagna. A questo riguardo viene specificato che, anche in base a studi precedenti dello stesso autore ([Geer, 2008](#)), risulta più probabile che sia la crescente polarizzazione a causare l’aumento della negatività e non il contrario, come afferma invece Ansolabehere in uno studio precedente ([Ansolabehere & Iyengar, 1995](#)). Emerge, quindi, che la negatività generale di una campagna politica può derivare dal contesto elettorale. In seguito vedremo anche la relazione con un’altro aspetto del contesto politico: la frammentazione degli schieramenti.

La terza ipotesi riguarda i media tradizionali e la loro tendenza a sovraesporre questo tipo di contenuto poiché risulterebbe più coinvolgente per gli/le utenti. Un celebre esempio a

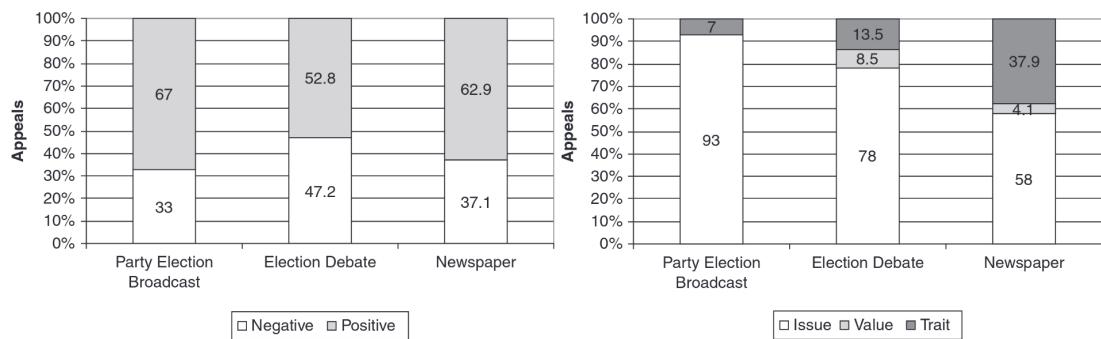
riguardo è la campagna elettorale del 2004 in cui la Swift Boat Veterans for Truth, un' associazione di veterani americani, ha diffuso un video ([youtube.com/watch?v=V4Zk9YmED48](https://www.youtube.com/watch?v=V4Zk9YmED48)) in cui si attaccava direttamente il candidato alla presidenza John Kerry, riportando diverse testimonianze in cui venivano smentite alcune sue dichiarazioni riguardo le azioni eroiche svolte durante la leva militare in Vietnam. La cosa interessante di questo episodio è che lo spot pubblicitario è stato trasmesso solo sulle televisioni di Iowa, Wisconsin e Ohio, raggiungendo circa un milione di elettori. A causa però della grande diffusione che ne è stata fatta da altri media, alla fine della campagna elettorale quasi l'80% dell'elettorato dichiarava di conoscere l'episodio. Un aspetto che non viene trattato in questo studio è invece la funzione svolta dai social media che verrà invece affrontata in questa rassegna.

#### 2.4.I NEGATIVITÀ E MEDIUM

La relazione tra il tono della campagna e i media viene discussa in vari articoli, in particolare vengono indagate l'esposizione di questi messaggi nei media tradizionali, e la maggiore predisposizione dei contenuti negativi a diventare virali sui social media.

#### I MEDIA TRADIZIONALI

Da almeno trent'anni a questa parte le istituzioni subiscono un processo di continua evoluzione, dettata anche dal progressivo aumento della loro dipendenza dai mass media. Ancora prima dell'avvento dei social media, Mezzoleni e Schulz si interrogano su questo processo di *mediatisation* della politica ([Mazzoleni & Schulz, 1999](#)) e su come potrebbe diventare un cambiamento positivo o negativo per le democrazie di tutto il mondo. Dopo aver analizzato i trend nella fiducia nelle istituzioni, comparando diversi paesi, e dopo aver descritto come la professionalizzazione dei *campaign strategist* sia ormai (nel '99) un fatto consolidato e in qualche modo collegato a un aumento della negatività del discorso politico, non arrivano a una conclusione univoca. Propongono, quindi, due teorie opposte. Da una parte le istituzioni starebbero andando verso una “media-driven democracy”, dall'altra sarebbe semplicemente in atto una “third age” della comunicazione politica, senza eccessive cattive ripercussioni sulla democrazia in sé. Viene però negata l'esistenza di prove a favore della nascita di un “party of the media”, il fatto cioè che le democrazie sarebbero in balia di partiti fondati da chi detiene il potere mediatico, sostenendo piuttosto che il consenso sarebbe sempre più organizzato attraverso i media e che quindi risulta fondamentale utilizzarli al meglio per vincere le elezioni, ma questo è possibile anche per chi non detiene il controllo dei media stessi. In quegli anni



**Figura 2.7:** Quantità di campagna negativa registrata durante le elezioni olandesi del 2010, sono analizzati gli interventi dei partiti in tre diversi canali di comunicazione. Immagine pubblicata nello studio di Walter (2010).

**Figura 2.8:** Tipo di target di campagna negativa registrata durante le elezioni olandesi del 2010, sono analizzati gli interventi dei partiti in tre diversi canali di comunicazione. Immagine pubblicata nello studio di Walter (2010)

la vittoria di Fernando Collor de Mello in Brasile, di Tony Blair in Inghilterra e dello stesso Berlusconi in Italia, aveva infatti posto molti interrogativi sull’uso che questi politici avevano fatto di televisioni private e di strategie comunicative spregiudicate.

Come viene sottolineato in un articolo relativo alle elezioni olandesi del 2010 ([Walter & Vliegenthart, 2010](#)), il medium sul quale viene diffuso un messaggio è molto rilevante per poterne analizzare il livello di negatività. Non tutti i canali di diffusione di informazioni sono ugualmente predisposti ad accogliere questo tipo di messaggi. Nello studio di Walter, prendendo in considerazione le comunicazioni degli stessi partiti, i dibattiti elettorali andati in onda sulle televisioni pubbliche e quelle commerciali e i principali quotidiani, vengono codificati tutti gli attacchi positivi e negativi con i rispettivi target. Dalle analisi effettuate risulta che le differenze tra i tre canali di comunicazione nell’esposizione delle campagne negative sono significative, dimostrandosi più elevate nei dibattiti televisivi rispetto a giornali e canali dei politici stessi [Fig. 2.7]. Per quanto riguarda il target utilizzato, l’analisi evidenzia come più ci si allontana dal controllo diretto dei candidati sui messaggi, più questi riportano poco gli attacchi alle argomentazioni degli avversari (“issue”) per spostarsi su attacchi verso il candidato avversario (“value” e “trait”) [Fig. 2.8].

In uno studio del 2008 in Scandinavia ([K. M. Hansen & Pedersen, 2008](#)), viene riscontrato un fattore che potrebbe portare i politici ad aumentare sia la negatività delle loro campagne sia la sovraesposizione da parte dei media tradizionali di questo tipo di messaggio. Analizzando i principali quotidiani si riscontra un 22% di articoli con citazioni dirette ad attacchi mossi da politici in corsa per le elezioni. Risulta dallo studio che però i politici abbiano solo un ’8% di questo tipo di attacchi nelle loro campagne, dimostrando con evidenza

che i media tradizionali preferiscono parlare di questo tipo di critiche rispetto ad altre meno sensazionalistiche.

Anche uno studio sulle elezioni austriache del 2013 ([Haselmayer, Meyer, & Wagner, 2019](#)) riporta risultati simili, dopo aver analizzato 1496 esternazioni dei partiti e 6512 notizie apparse sui giornali riguardo la campagna elettorale. Nella ricerca viene eseguita una codifica prima automatica e poi anche manuale atta a identificare quali delle esternazioni dei politici abbiano generato anche articoli sulla stampa. Ogni contenuto prodotto dai politici è stato anche codificato in base alla definizione direzionale di campagna negativa. I risultati rilevano come i messaggi negativi vengano maggiormente ripresi dalla stampa (17% dei casi) se confrontati con quelli positivi (14%). Viene, quindi, confermata l'ipotesi secondo cui la negatività dei messaggi politici sarebbe un'efficace arma per accaparrarsi l'attenzione della stampa.

#### DAI MEDIA TRADIZIONALI AI SOCIAL NETWORK

Se giornalisti e redattori di media tradizionali quali radio, stampa e TV devono sottostare a un codice deontologico necessario per entrare nell'albo di categoria che in qualche modo limita la possibilità di creare distorsioni nel riportare le affermazioni dei politici, il meccanismo di selezione dei contenuti mostrati agli/alle utenti dei social networks è gestito da un algoritmo proprietario che difficilmente può essere indicato come responsabile di distorsioni. Questi algoritmi sono stati definiti spesso come delle "black box" ([Pasquale, 2015](#)) dagli studiosi di *algorithmic accountability* poichè, essendo alla base del valore commerciale delle piattaforme che li utilizzano, rimangono *trade secrets* scarsamente accessibili agli utilizzatori e quindi anche agli studiosi. Alla base del concetto di *black box* vi è l'idea che a partire da un input, un sistema complesso ed opaco (*black appunto*) restituisca un output visibile, facendo in modo che questo tipo di sistemi risultino, almeno in parte, oscuri non solo a chi ha accesso solo al risultato finale (un\* utente che si ritrova uno specifico post in bacheca), ma anche a chi conosce gli input utilizzati dall'algoritmo (i programmati della piattaforma). Alcuni studiosi si stanno interrogando su come ripensare gli algoritmi dei social networks per renderli più comprensibili e in grado di rispondere direttamente alle necessità di chi li utilizza ([Reviglio & Agosti, 2020](#)), ma la discussione rimane al momento ancora aperta.

Le scelte editoriali dei media tradizionali vanno quindi ricondotte al loro scopo informativo, ai bias individuali dei giornalisti e alla volontà degli editori di aumentare gli acquisti da parte dei loro clienti (l'obiettivo è un comportamento consci, come l'acquisto di un determinato giornale). Nel caso dei social networks, invece, la selezione dei contenuti va ricondotta ad algoritmi pensati per massimizzare il tempo trascorso sulla piattaforma dagli/delle utenti

(*engagement*, non informatività) per poter aumentare il numero di pubblicità che è possibile somministrare e quindi vendere. Questi software vengono implementati a partire dalle abitudini degli/delle utenti, in particolare basandosi sul tempo trascorso su ogni tipologia di contenuto (agendo cioè ad un livello pre-conscio, tramite ad esempio l'analisi dei millisecondi con cui si scorre un post sulla propria bacheca) (Han, 2017). Nonostante i due tipi di media abbiano modelli di *business* diversi e valutino la riuscita della selezione di un contenuto in base a metodi diversi di stima del gradimento, dagli studi emerge, però, che in entrambi i casi i messaggi negativi sono preferiti.

## I SOCIAL MEDIA

Risultati simili a quelli rilevati sui media tradizionali sono infatti riscontrati in analisi svolte sui social network.

Ad esempio, in un articolo relativo alle elezioni ungheresi del 2014 (Bene, 2017), emerge come anche su Facebook la negatività abbia ripercussioni sulla viralità dei post condivisi dai politici. Analizzando 7048 post di 183 diversi candidati e dopo aver categorizzato ogni contenuto in base alla sua struttura (quantità di testo, immagine, tipo di immagine postata...) viene codificato manualmente il tono del messaggio, basandosi sulla presenza di "critiche, attacchi o espressioni di vergogna". Il risultato è molto interessante: solo il tono negativo dei messaggi è in grado di aumentare il numero di interazioni (likes, condivisioni e commenti). Come affermano gli autori, i contenuti negativi sono in grado di portare gli/le utenti ad esprimere le loro opinioni sul social in modo più consistente. A differenza dei media tradizionali, il social rende molto più rilevanti le reazioni degli elettori alle campagne politiche, poiché proprio sulla capacità di generare interazioni determina la visibilità dei contenuti stessi. In questo modo i commenti e i *likes* (sia pro che contro il politico che li riceve) diventano direttamente parte della strategia comunicativa dei partiti, che non possono più prescindere da questi elementi se vogliono diffondere al meglio i propri contenuti.

In generale diversi studi (Berger & Milkman, 2010) (Stieglitz & Dang-Xuan, 2013) hanno trovato evidenze riguardo la diffusione e propagazione di messaggi basati sull'emotività, che risultano essere molto efficaci nei contesti dei social media. Analizzando 165 mila tweet di comunicazione politica, Stieglitz e Dang-Xuan evidenziano come i tweet con forte emotività vengano condivisi molto più spesso rispetto a quelli neutrali e anche in modo più veloce, senza però riscontrare particolari differenze tra contenuti positivi e negativi. Hansen e colleghi (L. K. Hansen, Arvidsson, Nielsen, Colleoni, & Etter, 2011), sempre su Twitter, trovano evidenze di una maggiore viralità dei contenuti negativi, ma con un campione diverso rispetto

allo studio precedente, composto da notizie di attualità. Questa potrebbe essere la differenza in grado di spiegare i due diversi risultati.

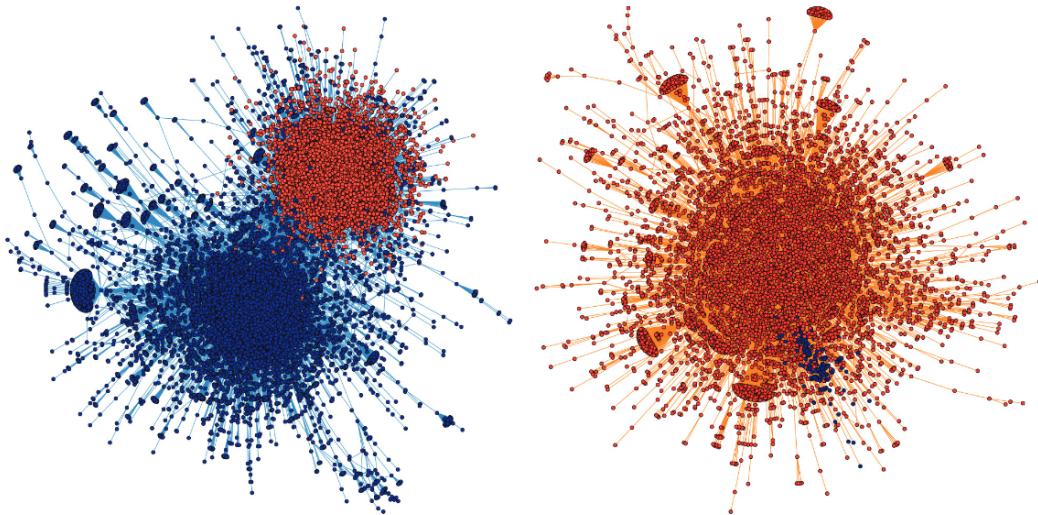
Compiendo una *sentiment analysis* di più di 100 mila tweet contenenti messaggi rivolti a partiti o personaggi politici, uno studio svolto durante le elezioni tedesche del 2009 (Tumasjan, Sprenger, Sandner, & Welpe, 2011) dimostra come Twitter sia un specchio fedele della realtà politica offline. Viene infatti mostrata una correlazione molto forte tra il numero di menzioni dei vari partiti ed i risultati delle elezioni. Questo conferma come, da almeno un decennio, i social networks siano parte integrante della vita politica e non si possano vedere solo come luoghi virtuali senza ripercussioni sulla vita reale. Attraverso l'utilizzo del software LIWC, gli autori analizzano anche il *sentiment* dei contenuti analizzati, giungendo alla conclusione che sulla piattaforma può essere riscontrato, con buona approssimazione, il sentimento degli elettori in tempo reale.

Indagando nello specifico la piattaforma Twitter, Conover e colleghi (Conover et al., 2011) tramite clusterizzazione degli/delle utenti e network analysis di 355 milioni di tweet pubblicati durante le elezioni americane di *midterm*, mostrano come i due tipi di interazioni (retweet e menzioni) possano produrre reazioni molto diverse [Fig. 2.9] e diversamente conflittuali. Chi usa il retweet prevalentemente discute con persone con cui è già d'accordo, le menzioni invece vengono utilizzate anche per discutere con persone lontane dal proprio punto di vista, aumentando la possibilità di scontri anche accesi.

#### 2.4.2 NEGATIVITÀ E CONTESTO ELETTORALE

Analizzando le elezioni danesi del 2005 Hansen e Pedersen (K. M. Hansen & Pedersen, 2008) confermano uno tra i principali punti fermi nella teoria del *negative campaign*: i partiti d'opposizione sono i più propensi ad una campagna negativa. Spiegano questa relazione evidenziando come i partiti in carica abbiano molta più facilità ad ottenere attenzione dei media. Chi è invece all'opposizione ha bisogno di farsi notare tramite campagne negative in grado di sfruttare la sovraesposizione di questo tipo di messaggi nei media tradizionali.

Anche Druckman e Parkin (Druckman et al., 2010) mostrano evidenze a riguardo. Comparando 700 candidati al congresso statunitense nelle loro pubblicità in televisione e sui loro siti web durante tre diverse tornate elettorali consecutive (2002, 2004, 2006), dimostrano come la negatività sia praticamente uguale sui diversi media e come sia collegata invece a variabili quali la posizione istituzionale ricoperta al momento: chi è all'opposizione usa quasi 4 volte più campagne negative. Questo risultato, inoltre, viene avvalorato dal fatto che la negatività aumenta in base al livello di competizione per il posto in palio e i candidati arrivano



**Figura 2.9:** Network dei re-tweet (sinistra) e delle menzioni (destra) su twitter, i due colori rappresentano due cluster di utenti con opinioni politiche avverse. Mentre in un caso la struttura è molto più polarizzata, nel secondo network lo è meno. Immagine pubblicata nello studio di Conover e colleghi (2011).

a mostrare negatività ugualmente molto elevate quando il posto vacante è conteso da due politici entrambi non in carica.

Anche il già citato Walter ([Walter, 2014](#)), nella sua comparazione delle campagne di diversi paesi europei, adducendo la tesi secondo cui le cause della negatività vadano riscontrate prevalentemente nel tipo di sistema elettorale, testa (e conferma) l’ipotesi della diversa propensione al *negative campaign* tra politici in carica e non.

Un’altra tesi presente in letteratura è quella secondo cui sarebbero lo schieramento politico e relativa polarizzazione (destra o sinistra) a determinare i livelli di negatività. Curini e Martelli ([Curini & Martelli, 2010](#)), hanno analizzato l’Italia dal dopoguerra alla fine della seconda repubblica (tra il 1946 e il 1994), in un periodo in cui era presente un bipolarismo chiamato “imperfetto” (tra Democrazia Cristiana e Partito Comunista). Considerando il tema della corruzione, mostrano come gli attacchi su questo argomento aumentino in modo inversamente proporzionale alla distanza ideologica che separa i partiti dalla controparte al governo.

#### 2.4.3 NEGATIVITÀ ED EFFICACIA NELLE CAMPAGNE POLITICHE

Sebbene risulti abbastanza chiaro come le campagne negative siano più efficaci nella diffusione del messaggio, la letteratura non indica che questi attacchi siano sempre più efficaci dal

punto di vista elettorale, come si evince dalla rassegna di Lau e colleghi ([Lau, Sigelman, & Rovner, 2007](#)).

Sembra infatti che tali campagne siano in grado di mobilizzare elettori radicali in favore del soggetto che viene attaccato ([Ansolabehere & Iyengar, 1995](#)) ottenendo, così, risultati opposti a quelli desiderati. Inoltre, non sembra essere il modo migliore per contrapporre la propria immagine a quella di un rivale, a causa delle possibili ripercussioni negative nell'opinione dei propri elettori per l'uso di un linguaggio non consono a una figura istituzionale e alla discussione civile.

Ci sono, però, anche evidenze di segno opposto secondo le quali, in un sistema multipartitico, campagne negative rivolte ad avversari vicini alle proprie posizioni politiche portano gli indifferenti con posizioni simili a entrambi i partiti a supportare chi attacca rispetto a chi è attaccato ([Curini & Martelli, 2010](#)). È stato successivamente specificato ([Ceron & Adda, 2016](#)) che però, l'effetto positivo su chi attacca viene accresciuto dall'essere allo stesso tempo sotto attacco: quando un partito è sotto attacco l'effetto “backlash” diminuisce poiché gli elettori difficilmente incolpano chi risponde ad altri attacchi per difendersi.

Uno studio condotto in Italia ha sondato gli effetti delle campagne negative sugli elettori a livello implicito ed esplicito ([Carraro, Gawronski, & Castelli, 2010](#)). Dopo aver presentato affermazioni negative e positive verso avversari utilizzate da politici in corsa nelle ultime elezioni precedenti allo studio, sono state sottoposte ai partecipanti dell'esperimento valutazioni implicite ed esplicite dei candidati. I risultati mostrano come le valutazioni esplicite diminuiscono il gradimento dei candidati che attaccano ma non di quelli che subiscono gli attacchi, mentre a livello implicito il gradimento diminuisce per entrambi i candidati presentati.

Anche uno studio condotto sulle elezioni statunitensi ([Fridkin & Kenney, 2011](#)) del 2006 giunge a conclusioni simili, prendendo in considerazione 30 mila soggetti che hanno effettuato una valutazione dei candidati prima e dopo le elezioni e mettendoli in relazione con gli spot elettorali andati in onda in TV e la copertura mediatica che ne è stata fatta sui giornali. La particolarità di questo studio è che viene analizzato, oltre alla negatività dei messaggi, anche il livello di linguaggio incivile contenuto degli spot e la tematica più o meno rilevante per l'opinione pubblica. I risultati mostrano come linguaggi negativi e incivili influenzano effettivamente l'elettorato. In particolare messaggi su temi rilevanti con toni incivili abbassano le valutazioni per chi è in carica, mentre chi è all'opposizione non subisce questo effetto “backlash”.

Dall'altra parte, nello stesso studio ([Fridkin & Kenney, 2011](#)), viene anche sottolineato come gli attacchi negativi risultano però più presenti nella memoria degli elettori rispetto a

quelli positivi. Risultato confermato in maniera diversa anche da un altro studio ([Perloff & Kinsey, 1992](#)), in cui vengono confrontate le opinioni sulle campagne negative di consulenti e giornalisti politici, mostrando come entrambi pensino che possano avere importanti effetti sul comportamento dei votanti, indicando l'appello alle emozioni come una delle tecniche più utile per raggiungere gli elettori.

In generale, comunque, è importante considerare che gli attacchi negativi possono essere rivolti sia ad aspetti personali del candidato rivale che ai suoi valori e al suo programma. Nel secondo caso il “backfire” è molto minore che nel primo caso ([Carraro et al., 2010](#)).



*"Le parole sono atti, dei quali è necessario fronteggiare le conseguenze"*

Gianrico Carofiglio

# 3

## *Hate Speech*

### 3.1 LIBERTÀ DI PAROLA O LINGUAGGI D'ODIO?

La riflessione sul discorso d'odio viene spesso ricondotta alla *vexata quaestio* dei limiti della libertà di espressione. In particolare, negli Stati Uniti questo dibattito risulta molto spostato verso la garanzia del diritto di parola, sancito dal primo emendamento della costituzione che vieta esplicitamente al congresso di limitare la libera espressione di individui e stampa.

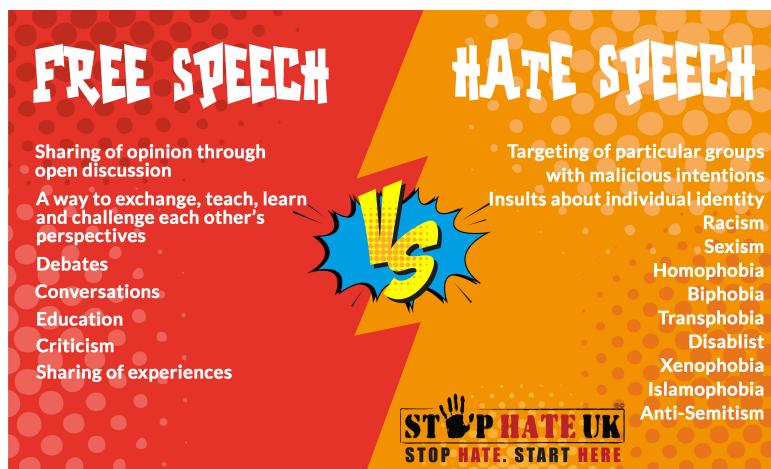
Proprio una ricercatrice statunitense, Katia Campbell, in un suo discorso del 2019 ([K. Campbell, 2019](#)), spiega molto bene le origini di questo dibattito, in particolare nel mondo anglosassone. L'idea secondo cui la libertà di parola possa essere incondizionata si basa principalmente sulla teoria filosofica del "marketplace of ideas" che ha avuto un suo grande fautore nell'economista dell'ottocento John Stuart Mill ([Mill, 1859](#)). A partire dalla corrente di pensiero liberista viene affermato come lo stato debba limitare il meno possibile gli individui (e il mercato), riconoscendo l'importanza del confronto dialettico, a partire dalla necessità di conoscere l'opinione opposta per poter essere realmente convinti della propria. Per evitare quindi la "tirannia della maggioranza" non si può proibire nessuna idea, neanche quelle ritenute scorrette o false dalla società stessa. Andando ancora più indietro nel tempo è possibile riscontrare nei sofisti greci le origini profonde di questo punto di vista. Per questi filosofi la natura della verità era infatti relativa e solo nello scontro tra argomenti diversi poteva nascere una vera democrazia. Per questo motivo i sofisti si sono dedicati all'insegnamento della retorica e della persuasione politica, ritenendoli strumenti indispensabili alla vita di ogni cittadin\*,

ancora più degli specifici valori morali.

In contrasto con questa teoria scettica, che arriva addirittura a domandarsi se la Conoscenza in quanto tale possa esistere, Campbell afferma come sia necessario porre dei limiti al linguaggio su basi etiche, affinché si possano salvaguardare tutte le componenti della popolazione che non hanno i mezzi per combattere ad armi pari all'interno dell'arena della società (K. G. Campbell, 2004). Viene sottolineato come il linguaggio sia alla base della costruzione culturale, rimarcando la natura costituente della comunicazione. Utilizzando i fondamenti della "Critical Race Theory", Campbell afferma come sia necessario contestualizzare alcune delle norme presenti anche nella costituzione, affinché le leggi non diventino uno strumento per silenziare e marginalizzare le minoranze presenti nel paese. E', inoltre, importante creare un dibattito sano all'interno della società, che si basi anche sullo scontro di opinioni, ma che escluda categoricamente menzogne o insulti. L'*hate speech* viene descritto come dannoso su più livelli (individuale, comunitario e della società nel complesso) e una minaccia diretta all'equità necessaria a permettere la partecipazione di tutt\* alla vita democratica.

Anche Popper prende una posizione molto netta nel dibattito sui limiti alla libertà di pensiero (Popper, 1945). Con il suo "Paradosso della tolleranza", dimostra come sia necessario essere intolleranti con gli intolleranti, affinché si possa avere una società effettivamente tollerante. Nel libro "La società aperta e i suoi nemici" afferma che non è possibile assimilare la libertà di parola a discorsi discriminatori che vorrebbero mettere a tacere idee alternative. Un mondo in cui gli intolleranti possono promulgare i loro discorsi di odio in modo indisturbato rischia di diventare velocemente una società intollerante in cui, conseguentemente, la stessa libertà di parola risulta in pericolo. A partire da questo ragionamento, però, non si dice favorevole a limitare a priori la possibilità di espressione di tutte le persone definite come intolleranti. Piuttosto sostiene che sia necessario mettere in campo meccanismi per combattere questo tipo di ragionamenti con argomentazioni razionali e, soprattutto, ritiene che sia indispensabile la massima vigilanza e partecipazione da parte dell'opinione pubblica sul tema. Queste strategie possono risultare molto più efficaci nel contrastare l'intolleranza della semplice censura generalizza. In casi estremi comunque, rivendica il diritto di poter sopprimere queste idee dannose, anche con la forza.

Un articolo in grado di sintetizzare e confrontare efficacemente tutte le posizioni emerse nel dibattito sui limiti alla libertà di pensiero è quello di Massaro (Massaro., 1990). L'autore divide infatti le possibili interpretazioni dell'argomento in tre filoni principali: il primo tende a massimizzare le libertà individuali, proponendo di non punire l'*hate speech* in nessun caso; al contrario il secondo punto di vista prevede l'utilizzo di tecniche di repressione del discorso



**Figura 3.1:** Rappresentazione schematica delle differenze tra libertà di parola e discorsi d'odio. Immagine pubblicata sul sito di Stop Hate UK.

d'odio per favorire la salvaguardia delle categorie maggiormente a rischio arrivando a ritenerne punibile questo tipo di discorso in tutti i casi, non solo quando rivolto a settori della società e a categorie di persone definibili come vulnerabili; la terza opzione risulta una via di mezzo tra le prime due, prevedendo precise e progressive normative in grado di sanzionare solo gli attacchi basati su "razza, genere, religione, origini etniche, orientamento sessuale e altre caratteristiche protette". Il ricercatore si schiera in favore della terza via, in quanto in grado di recepire i tratti positivi delle prime due e raggiungere una sintesi di compromesso. Questa è infatti la posizione condivisa, come vedremo, dalla maggior parte delle associazioni che si occupano del tema e dalle stesse piattaforme social.

### 3.2 LA DEFINIZIONE

Anche la definizione di *hate speech*, come quella di *negative campaign*, risulta non univoca e differente a seconda dei contesti. Lo studio di Amnesty International, sulla base del quale è stata costruita la presente ricerca, utilizza la definizione che viene data dalla Commissione Europea contro il razzismo e l'intolleranza nel 2015 ([European Commission against Racism and Intolerance, 2015](#)):

"[...] si intende per discorso dell'odio il fatto di fomentare, promuovere o incoraggiare, sotto qualsiasi forma, la denigrazione, l'odio o la diffamazione nei confronti di una persona o di un gruppo, nonché il fatto di sottoporre a soprusi, insulti, stereotipi negativi, stigmatizzazione o minacce una persona o un gruppo

e la giustificazione di tutte queste forme o espressioni di odio testé citate, sulla base della "razza", del colore della pelle, dell'ascendenza, dell'origine nazionale o etnica, dell'età, dell'handicap, della lingua, della religione o delle convinzioni, del sesso, del genere, dell'identità di genere, dell'orientamento sessuale e di altre caratteristiche o stato personale."

Sempre nello stesso documento viene anche sottolineato che:

"[...] il discorso dell'odio può assumere la forma di una pubblica negazione, banalizzazione, giustificazione o legittimazione dei crimini di genocidio, dei crimini contro l'umanità o dei crimini di guerra accertati dai tribunali, come pure di un'apologia delle persone condannate per avere commesso tali crimini"

Nell'accezione proposta dalla Commissione Europea, quindi, l'odio non è riducibile ai semplici insulti generici, ha la particolarità di rivolgersi a una precisa minoranza già vittima di discriminazioni.

Uno studio pubblicato dall'UNESCO fornisce un quadro chiaro del contesto all'interno del quale si colloca la definizione di *hate speech*, anche riguardo la sua versione online ([Gagliardone, Gal, Alves, & Martinez, 2015](#)). Viene spiegato come quanto riportato nella Dichiarazione dei Diritti Umani (UDHR) all'interno della "International Covenant on Civil and Political Rights" (ICCPR) del 1966 ([UDHR, 1966](#)) sia il vero fondamento della lotta all'odio, nonostante le trasformazioni della società e delle sue forme comunicative avvenute negli ultimi 60 anni. In questo importante testo sui diritti politici e civili, all'articolo 19, viene sancita la libertà di parola, sottolineando però come sia necessario rispettare i diritti e la reputazione degli/delle altr\*.

### **Article 19**

- ”1. Everyone shall have the right to hold opinions without interference.
2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.
3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary: (a) For respect of the rights or reputations of others; (b) For the protection of national security or of public order (ordre public), or of public health or morals.”

Nell'articolo 20, i limiti alla libertà di parola, già delineati nell'articolo 19, vengono ulteriormente specificati con l'affermazione che è necessario perseguire le discriminazioni razziali e religiose.

#### **Article 20**

- ”1. Any propaganda for war shall be prohibited by law.
- 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.”

I/le ricercat\* che hanno partecipato allo studio UNESCO evidenziano che, benché esistano moltissime sfumature su base nazionale delle possibili definizioni di *hate speech*, è necessario che tutte si conformino a questi principi generali, allargandoli il più possibile e non riducendone la capacità concreta di contrasto del fenomeno.

Viene inoltre sottolineato come sia importante anche per le piattaforme digitali raggiungere, oltre a una comune definizione di linguaggio d'odio, anche un accordo sulle iniziative da intraprendere per limitarne la diffusione.

Un passo in questo senso è stato fatto anche grazie alla campagna ”Stop Hate for Profit” ([Stop Hate For Profits, 2020](#)), già citata nel capitolo ”[Introduzione](#)” della presente ricerca. Parallelamente a questa iniziativa, anche parte degli inserzionisti delle piattaforme social ([World Federation Advertisers, 2020](#)) hanno deciso di affrontare l'argomento riuscendo a convergere su un'unica definizione di *hate speech* fondamentalmente uguale a quella utilizzata nel presente studio. Nonostante ciò, rimangono ancora alcuni dubbi su quali iniziative concrete verranno prese dalle piattaforme nel breve periodo ([Fischer, 2020](#)).

Allo stato attuale delle cose, analizzando i termini del servizio delle due piattaforme prese in considerazione in questo studio, possiamo trovare le seguenti informazioni.

#### TWITTER

Twitter non menziona direttamente l'*hate speech*, ma avverte i suoi utenti che ”potrebbero essere esposti a contenuti che potrebbe essere offensivi”, continua poi specificando che ”in nessuna circostanza Twitter sarà responsabile in nessun modo di nessun contenuto” né dei relativi ”danni per perdite di ogni tipo” che ne potrebbero conseguire. Nelle regole sulla pubblicazione di tweets viene aggiunto però che non è possibile ”pubblicare o postare dirette, specifiche minacce o violenze contro altr\*” ([Twitter, 2019](#)) e viene specificato come verrà

applicata "tolleranza zero" a messaggi "violenti o discriminatori" sia nei tweet pubblici che nei *Direct Messages*, pena "l'immediata e permanente sospensione dell'account".

## FACEBOOK

Facebook, invece, cita direttamente l'*hate speech* definendolo come "gli attacchi diretti a persone sulla base di razza, etnicità, origini nazionali, credo religioso, orientamento sessuale, sesso, *gender identity*, serie disabilità o malattie" e dichiarando la facoltà e la volontà di rimuoverlo dalla sua piattaforma (Facebook, 2019). Aggiunge poi che però "permetterà humour, satira sociale e commenti relativi a questi temi" ed invita gli/le utent\* della piattaforma a utilizzare vere informazioni anagrafiche nel momento della registrazione, ritenendo che questo "aiuterebbe le persone ad essere più responsabili" e allo stesso tempo renderebbe "le persone sempre più consapevoli del tipo di *audience* a cui si rivolgono" quando utilizzano la piattaforma. Viene quindi espressa una posizione molto precisa sull'anonimato come fattore in grado di aumentare il numero di contenuti d'odio, idea confermata anche dalla letteratura che affronteremo nei prossimi paragrafi.

## DANGEROUS SPEECH

Una ulteriore riflessione sul concetto di *hate speech* viene invece proposta da Susan Benesch (Benesch, 2012). Secondo l'autrice la definizione dei discorsi d'odio, anche in questo caso intesi come "discorsi che denigrano le persone sulla base dell'appartenenza a gruppi come quelli etnici o religiosi", risulta troppo ampia e non adatta a prevenire atrocità di massa per due ragioni: l'*hate speech* è presente anche in nazioni in cui non vi è pericolo di genocidi; questo tipo di discorso non aumenta il rischio di crimini violenti di massa, anche se può arrecare seri danni emotivi e psicologici. L'obiettivo che si pone la ricercatrice keniota è quello di contestualizzare questa definizione per renderla più efficace in specifici contesti in cui i crimini d'odio sono molto frequenti ed estesi.

Viene quindi proposto il "*dangerous speech*" inteso come discorso d'odio con una ragionevole probabilità di catalizzare o amplificare violenza di un gruppo verso un altro. La specificità di questa particolare accezione di *hate speech* è quella di essere definita in base alle circostanze nelle quali viene propagandato [Fig. 3.2]. Vengono infatti identificati alcuni fattori che, se presenti contemporaneamente, sono in grado di rendere più "pericoloso" un discorso d'odio: uno *speaker* con un'altra influenza sul pubblico; un *audience* con una paura latente che può essere coltivata da chi comunica; un discorso con un chiaro appello alla violenza; un contesto



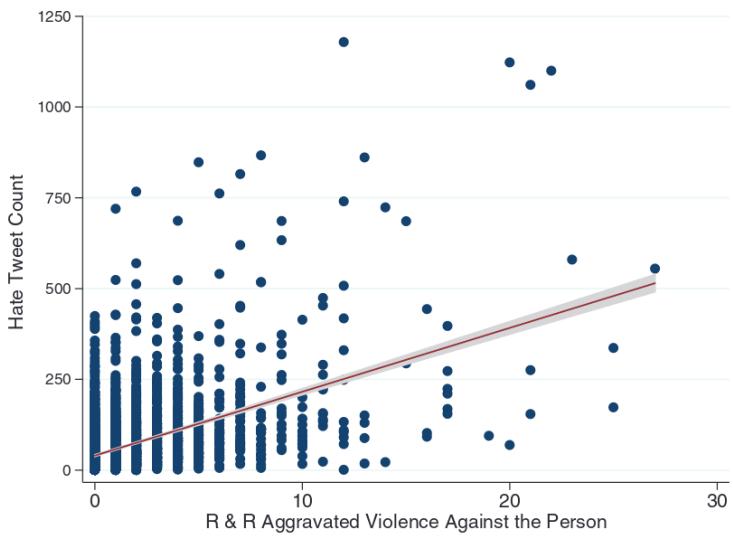
Figura 3.2: Rappresentazione della definizione di Dangerous Speech realizzata da Peter Durand in seguito a una conferenza di Susan Benesch

storico e sociale propizio alla violenza; la propagazione del discorso attraverso mezzi di comunicazione molto influenti, se non addirittura con il monopolio delle forti d'informazione disponibili.

### 3.3 DISCORSI D'ODIO E CRIMINI D'ODIO

Il primo caso conclamato di effetti estremamente negativi nella vita offline a seguito di campagne di odio diffuse attraverso canali di comunicazione digitali è quello delle elezioni in Kenya del 2007 (Osborn, 2008). In seguito a quella tornata elettorale si generò un'onda di violenza che portò all'uccisione di più di mille persone, altre 600 mila furono sfollate. In quell'occasione le *Information Communication Technologies* hanno svolto per la prima volta un ruolo fondamentale nel contesto elettorale. Come suggerito da Osborn infatti, tramite SMS, e-mail e siti web, sono state diffuse una serie di notizie false e incitazioni all'odio che hanno determinato un contesto sociale estremamente instabile e pericoloso. Attraverso un grande numero di interviste e osservazioni partecipate, l'autore è riuscito a ricostruire l'origine e la forma di diffusione di questi contenuti, ottenendo una significativa mappa degli effetti che ne sono susseguiti, nonostante questo fosse uno tra i primi studi del suo genere.

Durante le successive elezioni Keniane del 2013, è stato condotto uno studio molto più ampio e sistematico dei canali di comunicazione digitale e delle conseguenze che ne sono scaturite. Utilizzando la definizione già citata di *dangerous speech* elaborata da Susan Benesch (Benesch, 2012), sono stati raccolti per sei mesi tutti gli episodi di *hate speech* tramite interviste, monitoraggio di siti online e una raccolta di messaggi ricevuti da singoli individui, sia in inglese che nei dialetti locali. Successivamente sono stati codificati i contenuti selezionati in base a due livelli di odio: *hate speech* o *dangerous speech*. Benché all'epoca solo il 35% della popolazione locale avesse accesso ad internet e solo metà di queste persone avesse un profilo Facebook, nonostante quindi la parte di popolazione presente su questo social network fosse quella più giovane e urbanizzata, quindi la meno incline all'odio razziale, è stato riscontrato che ben un contenuto d'odio su quattro era assimilabile al *dangerous speech* in quanto conteneva incitazioni dirette all'uccisione di membri della fazione rivale. Interessante notare che molti degli utenti protagonisti di questi attacchi, a poca distanza dalle esternazioni avevano cambiato nome del profilo, dando un'ulteriore conferma della teoria (che vedremo tra poco) dell'anonimato come disinibitore di atteggiamenti aggressivi online. Altro dato, particolarmente interessante per alla presente ricerca, riguarda i canali su cui l'odio si è propagato durante il periodo di monitoraggio: solo il 3% del totale è stato riscontrato su Twitter, mentre

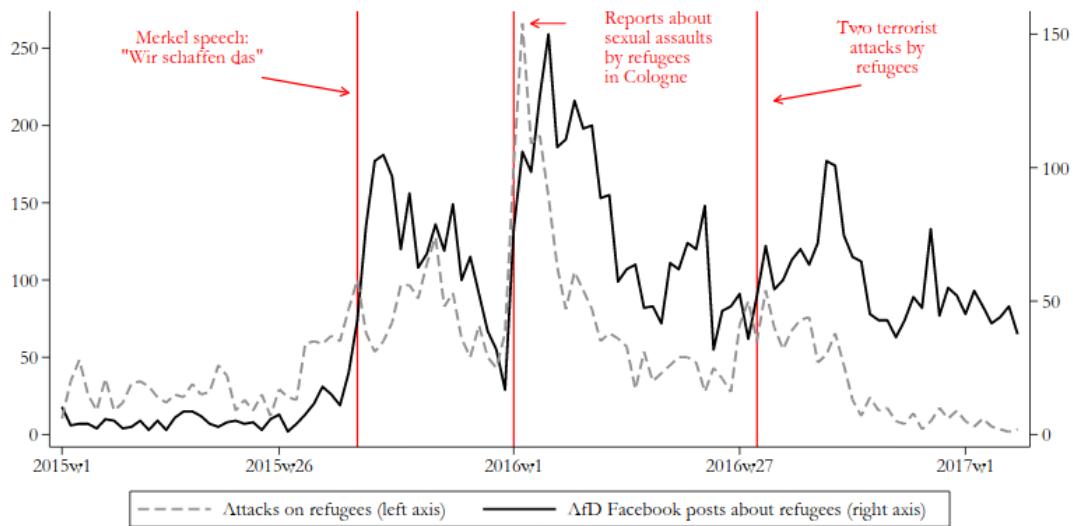


**Figura 3.3:** La correlazione tra il numero di tweet di odio e il numero di crimini d'odio. Immagine pubblicata nello studio di Williams (2019).

oltre il 90% ha avuto luogo su Facebook. Secondo gli/le ricercatori\*, questo potrebbe essere dovuto alle caratteristiche della piattaforma di Mark Zuckemberg nella quale è possibile discutere anche in gruppi e pagine specifiche e non solo in trend pubblici come su Twitter, dando la possibilità di comportarsi in modo diverso in base all'ambiente con il quale si interagisce. Quest'ultimo dato è però legato anche all'utilizzo particolare dei social media da parte della popolazione Keniota, rimane quindi un dato difficilmente generalizzabile ad altre nazioni e periodi storici.

La relazione con l'*hate speech* online è stata indagata anche in nazioni in cui l'*hate crime* non si configura come un fenomeno di massa. Come dimostrato da uno studio condotto nel 2019 a Londra tramite l'utilizzo di tecniche di *Computational Criminology* (Williams, Burnap, Javed, Liu, & Ozalp, 2019), è possibile riscontrare un'associazione tra i discorsi d'odio riguardo razza e religione registrati su Twitter e crimini d'odio aggravati da discriminazioni dello stesso tipo avvenuti in un determinato luogo. Per questa ricerca sono stati confrontati, in un periodo di otto mesi, i dati della polizia anticrimine e quelli dei social di microblogging, trovando una relazione spaziale e temporale tra i due tipi di fenomeno. Come possiamo osservare nella figura 3.3, la correlazione tra le due serie di dati risulta significativa (l'intervallo di confidenza al 95% è rappresentato in grigio).

Anche in uno studio analogo svolto in Germania, in cui sono state analizzate le discriminazioni rivolte ai rifugiati, le conclusioni sono le stesse: il sentimento anti-rifugiati rilevato



**Figura 3.4:** Il numero di post anti-rifugiati sulla pagina Facebook dell'Afd e il numero di crimini anti-rifugiati in Germania, nel corso del tempo. Immagine pubblicata nello studio di Muller e Schwarz (2019).

su Facebook predice il numero di crimini d'odio legati a questa categoria di persone (Müller & Schwarz, 2019). Nello studio viene preso in considerazione il partito “Alternative für Deutschland” (Alternativa per la Germania, AfD), fazione di estrema destra con posizioni fortemente anti-rifugiati diventata il terzo partito nelle elezioni federali del 2017. Qui viene analizzata la loro pagina Facebook poiché in grado di far registrare un numero di interazioni molto alto e nettamente superiore agli altri due principali partiti tedeschi (300 mila followers e 175 mila posts nel 2017). La particolarità di questa pagina è quella di lasciar pubblicare a qualsiasi utente direttamente sulla propria bacheca, non moderando in nessun modo i messaggi di odio che vi vengono generati. L'esito di questa strategia è una percentuale di post relativi alla questione rifugiati molto maggiore rispetto ai media tradizionali e con molti messaggi identificati come *hate speech*. I ricercatori hanno incrociato i dati provenienti da questa pagina con quelli di varie associazioni ed istituzioni che monitorano i crimini d'odio, giungendo a un database complessivo che ha coperto 4 466 municipalità tedesche in un periodo di più di due anni, tra il 2015 e il 2017. Tramite le API di Facebook sono stati estratti 176 153 posts, 290 854 commenti, 510 268 likes, e 93 806 user ID individuali dalla pagina dell' AfD. Le analisi confermano l'esistenza di una significativa correlazione tra i due fenomeni, soprattutto in presenza di avvenimenti particolarmente salienti per l'opinione pubblica, come possiamo vedere nel grafico riportato [Fig. 3.4].

### 3.4 DISCORSI D'ODIO ONLINE

Se è stata accertata una relazione tra odio online e crimini di odio offline, diventa rilevante una riflessione sulle specificità dell'*hate speech* generato sul web rispetto a quello presente nei media tradizionali. Di questo si occupa uno studio di Brown ([Brown, 2018](#)). Lo studioso inizia riprendendo le caratteristiche proprie dell'odio online, definito anche come "cyberhate", già presenti in letteratura: la facilità di accesso allo strumento, la dimensione dell'*audience* facilmente raggiungibile e la possibilità di restare anonimi (tratteremo quest'ultimo aspetto in modo approfondito più avanti). Per sintetizzare e specificare meglio accessibilità e potenza della comunicazione digitale, Brown propone un quarto aspetto che chiama "istantaneità", facendo riferimento alla estrema velocità con la quale è possibile pubblicare contenuti su internet, rendendo più spontaneo e più sconsiderato il linguaggio di chi comunica. Tramite l'esempio dell'ipotetica diffusione di un messaggio antisemita, confrontato il tempo richiesto per raggiungere la stessa *audience* con un tradizionale volantinaggio rispetto alla pubblicazione di un contenuto sui social media, appare chiaro quanto i tempi vengano ridotti da ore (se non giorni) a solo alcuni *istanti*, appunto. Viene sottolineato inoltre che non ci sono particolari differenze di sostanza negli attacchi d'odio tra online e offline, quello che risulta sicuramente diverso è il modo in cui questo linguaggio viene moderato (o meglio non moderato) su piattaforme digitali rispetto a media tradizionali e vita offline. Sulle principali piattaforme digitali non vengono infatti imposte le stesse restrizioni adottate dalle *media companies* tradizionali come, ad esempio, le linee editoriali o i codici di condotta. Inoltre, il carattere di "istantaneità" dei messaggi online permette la pubblicazione di messaggi scritti istintivamente, senza alcuna riflessione e prima che si possano attivare freni inibitori. Tutto questo favorisce la diffusione dell'odio.

### STRUMENTI DI CROWDSOURCING PER COMBATTERE L'ODIO

Se da una parte lo sviluppo dei social media ha dato luogo a un nuovo tipo di modalità di propagazione dell'odio, ha anche dato la possibilità di ripensare gli strumenti con i quali è possibile combatterlo. Esistono almeno due associazioni che nel mondo si occupano di creare mobilitazioni tra gli utenti di internet tramite *crowdsourcing* di contributi indirizzati a catalogare e studiare i contenuti di odio online. Il loro obiettivo pratico è quello di costruire strumenti innovativi in grado di tracciare l'*hate speech* sui social network contribuendo a migliorare le policy delle varie piattaforme e a contrastare i fenomeni d'odio in generale. Una di queste è HateBase ([hatebase.org](#)) e si occupa di censire e localizzare i discorsi d'odio su

internet per costruire una mappa degli attacchi e della loro propagazione. Un fondamentale contributo di questa associazione è la creazione (e l'aggiornamento costante) di un dizionario con le specifiche parole d'odio usate in tutto il mondo, a seconda delle particolarità che il lessico assume nelle varie nazioni e lingue. La realizzazione di questo strumento è stata possibile grazie alla collaborazione di migliaia di persone che hanno aiutato a verificare esempi di discorsi d'odio, confermandone la natura in una specifica comunità.

The Online Hate Prevention Institute ([ohpi.org.au](http://ohpi.org.au)), invece, combatte l'odio con un applicazione web che consente a tutti gli utenti di internet di segnalare l'*hate speech* pubblicato sui vari social network. La possibilità di tenere traccia delle segnalazioni, che poi vengono inoltrate alle stesse piattaforme, permette di registrare i tempi di reazione di queste ultime e la loro celerità nel rimuovere i contenuti incriminati.

### 3.5 DISCORSI D'ODIO E CARATTERISTICHE PERSONALI

Oltre agli aspetti in grado di aumentare la diffusione dell'odio generati dalle proprietà del mezzo di comunicazione come la già citata istantanietà, sono presenti in letteratura anche studi che considerano le caratteristiche soggettive di chi diffonde l'odio. Forse la più significativa è l'anonimato. In uno studio del 2004 ([Suler, 2004](#)), viene preso in considerazione "l'effetto disinibizione" presente nel mondo online a partire dall'assunto che gli episodi di *self-disclosure* sono più frequenti nelle interazioni nel cyberspazio rispetto a quelle di persona. Se da una parte la comunicazione digitale porta con sé una "disinibizione benigna" in grado di spingere a condividere parti molto personali della propria vita, dall'altra parte può portare anche a una "disinibizione tossica" quando i soggetti si ritrovano ad usare un linguaggio incivile che altrimenti non userebbero o quando si trovano ad esplorare ambienti online legati a criminalità, violenza o pornografia che non frequenterebbero mai nella vita offline. Benché la distinzione tra questi due tipi di disinibizione online possa risultare per alcuni aspetti controversa poichè basata su considerazioni di valore, l'articolo di Suler cerca di individuare le cause alla base del fenomeno indipendentemente dai suoi risvolti positivi o negativi. Vengono individuati sei principali fattori alla base della disinibizione online riassunti nella seguente tabella:

<b>I sei fattori della disinibizione online secondo Surel (2004)</b>	
<b>Nome del fattore</b>	<b>Descrizione del fattore</b>
Anonimato dissociativo	Deriva dall'utilizzo di un profilo falso o animo che rende meno vulnerabili alle conseguenze di un comportamento, dissociando l'identità online da quella offline.
Invisibilità	A causata dall'impossibilità di vedere chi sta frequentando un determinato ambiente digitale come un sito internet, l'invisibilità può generare una propensione a visitare luoghi (digitali) dove altrimenti non si andrebbe se si fosse visti. Eliminare il linguaggio del corpo ed espressioni facciali in grado di far trasparire, per esempio, la disapprovazione o la noia, riduce le inibizioni.
Asincronia	L'asincronia della comunicazione digitale rende possibili interazioni "mordi e fuggi" in cui si risponde a messaggi e subito ci si disinteressa della risposta, per poi leggerla quando si è nello stato emotivo migliore per accettarne le conseguenze.
Introversione solipsistica	Comunicazioni solo testuali possono indurre chi comunica a immaginare la voce e il volto delle persone con cui si sta parlando, trasformando una figura digitale in un personaggio intra-psichico. Questo può portare a pensare discussioni online come se fossero all'interno della propria mente, aumentando la disinibizione.
Immaginazione dissociativa	È possibile che relazioni sviluppate nel cyberspazio diano luogo alla una netta separazione tra le regole e i comportamenti adottati online rispetto a quelli offline, delineando una separazione tra due mondi che in realtà non sono del tutto separati e spingendo ad aumentare la disinibizione dei comportamenti violenti.
Minimizzazione di status e autorità	Negli ambienti digitali molti dei simboli di status e potere presenti nelle relazioni dal vivo vengono meno, diminuendo la paura di disapprovazione e punizioni che ne deriva e facilitando una maggiore disinibizione

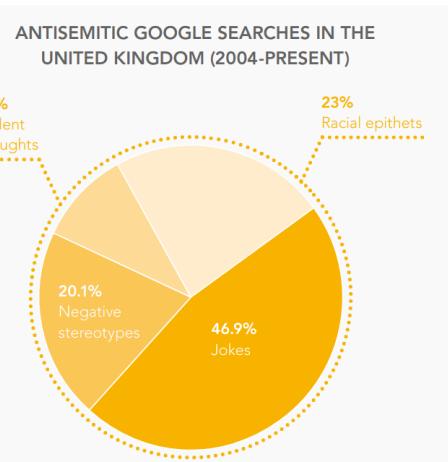
Nello studio viene inoltre sottolineato come anche le differenze individuali possano giocare un ruolo in questo fenomeno. Differenti personalità mettono in campo diversi meccanismi di difesa in grado di regolare la disinibizione. Ad esempio, personalità istrioniche tendono ad essere più aperte e disinibite, mentre persone compulsive possono mostrare tendenze opposte. In ogni caso, la presenza dei sei fattori alla base della disinibizione individuati da Surel favorisce l'esternazione di messaggi di odio online.

Un altro studio più recente ha indagato in laboratorio come le interazioni intergruppo possano incidere sulla percezione dei commenti di odio online (Kim & Wojcieszak, 2018). In questo articolo i commenti online sono stati classificati come contatto intergruppo diretto (quando viene nominato esplicitamente un membro dell'outgroup) o come contatto intergruppo indiretto (quando si legge un commento di un membro dell'outgroup non diretto a sé stessi). In particolare sono state prese in considerazione le reazioni ad articoli di giornali, considerati un luogo di scambio di opinioni, anche opposte, molto utilizzato nell'ultimo decennio. Il modello testato utilizzava le emozioni negative e positive come mediatori degli effetti dei contatti tra gruppi e come output il livello di minaccia ("generale", "simbolica" e "realistica") e la distanza sociale. Sono stati, quindi, selezionati 397 partecipanti dell'esperimento tutt\* contrar\* all'avanzamento dei diritti su due temi, quello dell'omosessualità e quello dell'immigrazione. Sono stati presentati loro alcuni articoli di giornali e i relativi commenti in due situazioni sperimentali: con contatto indiretto o diretto.

Dai risultati emerge che nel caso del tema omosessualità il contatto diretto influisce positivamente sulla riduzione della minaccia generale e simbolica, e sulla distanza sociale, ma non sembra avere effetti sulla percezione di minaccia realistica. Nessuno di questi indicatori è invece significativamente modificato nel caso del tema immigrazione. I test sul contatto indiretto, invece, non hanno registrato nessuna variazione significativa negli indicatori considerati come output. Inoltre, considerando le emozioni positive provate durante la lettura dei commenti come mediatore del tipo di contatto, si registrano aumenti significativi nella riduzione di tutti i tipi di minaccia e di distanza sociale, sia in relazione all'omosessualità che all'immigrazione, ma sempre solo in presenza di contatti diretti. Le emozioni negative diminuiscono alla lettura di commenti riguardanti l'omosessualità, e risultano mediatori nella riduzione di tutti i tipi di minaccia e la distanza sociale, ma non producono effetti significativi quando il tema è l'immigrazione. Nonostante il fatto che a volte le interazioni online siano utilizzati per diffondere odio e quindi aumentare la diffusione di stereotipi, possiamo concludere affermando che i contatti online positivi con membri dell'outgroup, soprattutto se diretti e non estesi, possono anche aiutare a ridurre il pregiudizio.

Anche gli orari in cui l'odio online viene prodotto ci possono fornire degli importanti spunti per capire che tipo di persone mettono in pratica questi comportamenti. Come riportato da un'altro studio di un'associazione inglese ([Community Security Trust \(CTS\) & Antisemitism Policy Trust, 2019](#)), dalle ricerche effettuate su Google è possibile trarre alcune interessanti considerazioni. In questo caso specifico sono state analizzate le ricerche che utilizzano un linguaggio o dei contenuti discriminatori nei confronti della comunità ebraica sul più famoso motore di ricerca. Sono state trovate 170mila ricerche di questo tipo in un anno in Inghilterra, nazione che si posiziona, così, al terzo posto nel mondo per ricerche sull'antisemitismo, dietro solo a Israele e Libano, con percentuali 29 volte superiori, ad esempio a quelle degli Stati Uniti. Come possiamo osservare dal grafico riportato nello studio, il 10% delle ricerche contiene "pensieri violenti", dovuti alla presenza di frasi come "Gli ebrei devono morire" o "Uccidere gli ebrei" [Fig. 3.5]. Interessante come la maggior parte di queste ricerche sia stata effettuata tra le 2 e le 3 del mattino. In questo momento della giornata sono concentrate in generale la stragrande maggioranza delle ricerche che fanno riferimento a comportamenti omicidi, ma anche suicidi, dimostrando come istinti violenti contro sé e contro gli altri abbiano una radice comune, da ricercare nella personalità e negli stati emotivi dei soggetti interessati.

Un altro articolo ha invece indagato le caratteristiche della personalità di chi odia online ([Erjavec & Kovačič, 2012](#)). In questo caso l'ambito di ricerca non ha riguardato i social network, ma i commenti agli articoli giornalistici pubblicati sui tre principali siti web di tre testate giornalistiche slovene. A differenza di altre ricerche, oltre all'analisi del discorso applicata ai contenuti online che compongono il campione di riferimento, sono state eseguite approfondite interviste con chi ha prodotto i contenuti, al fine di indagare le caratteristiche di chi diffonde odio. In seguito a tali interviste è stata fatta una classificazione dei soggetti coinvolti su un continuum che va dalla personalità autoritaria a quella libertaria sulla base di tre dimensioni principali: la relazione con la società; la dimensione cognitiva ed intellettuale; la relazione con gli altri. All'estremo



**Figura 3.5:** Percentuali di tipo di odio registrato nelle ricerche effettuate su Google in Inghilterra. Immagine pubblicata nello studio del CTS (2019).

dell'autoritarismo possiamo trovare, ad esempio, individui con grande rispetto per le autorità, tendenze al pregiudizio e fedeltà incondizionata verso il proprio gruppo di appartenenza. All'altro estremo si collocano persone con vocazione alla libertà ed all'equità, anticonformiste e tolleranti, più indipendenti dal gruppo di riferimento e indirizzate alla realizzazione personale. Sulla base di queste interviste sono stati definiti due diversi profili degli odiatori: da una parte quelli che sono organizzati e fanno riferimento a un particolare gruppo politico e ai suoi interessi, dall'altra quelli che agiscono in autonomia. Gli appartenenti al primo gruppo vengono definiti "i soldati" poiché appartengono a partiti o organizzazioni non governative, utilizzano discorsi militari, rispondendo a ordini dei loro superiori, e agiscono nell'ottica di completare le missioni assegnate difendendo i loro interessi e cercando di "distruggere il nemico". Nel secondo gruppo vengono inseriti i "believers", i "players" e i "watchdogs". I primi seguono la loro ideologia personale e le loro regole nell'ottica di attaccare il nemico, spesso utilizzano pseudonimi con significati precisi che richiamano alla giustizia e la leadership, assumendo il ruolo degli unici a dare la giusta interpretazione delle notizie, giustificando i loro attacchi di odio come l'unica modalità possibile per affermare la verità. Il secondo sottogruppo usa l'*hate speech* come un'arma in un gioco all'interno di una comunità digitale, dimostrando di non avere una particolare visione politica, ma di usare vari punti di vista in modo strumentale. Questo gruppo vede gli attacchi non come veri discorsi d'odio, ma come scherzi. L'ultimo profilo delinea un gruppo di persone che postano commenti di odio per bloccare quella che ritengono essere un'ingiustizia sociale, sono gli unici a rendersi conto di quanto sia problematico utilizzare questo tipo di linguaggio e sono gli unici consapevoli che l'anonimato giochi un ruolo fondamentale in quello che fanno, arrivando addirittura a dire che dovrebbe essere limitato e che in tal caso smetterebbero di diffondere odio.

Un altro studio che analizza le caratteristiche degli odiatori, ma con una metodologia diversa rispetto al precedente, è quelli di Ribeiro e colleghi ([Ribeiro, Calais, Santos, & Almeida, 2018](#)). La ricerca ha preso in considerazione 100 386 utenti di Twitter con almeno 200 tweets pubblicati sul proprio profilo tramite una *random-walk-based crawler* sulla piattaforma. In seguito è stato creato un grafo contenente i retweet tra i vari utenti e sono stati codificati come *hateful users* circa 5 mila individui, in base all'utilizzo di parole d'odio contenute nel dizionario di HateBase. A tutti i contenuti è stato anche assegnato un valore emozionale tramite il software Empath. E' stata inoltre svolta un'ulteriore codifica manuale per aumentare l'accuratezza del sotto-campione selezionato facendo emergere i 500 utenti con più alta incidenza di discorsi d'odio. Infine, è stato registrato quanti utenti sono stati bannati dalla piattaforma a tre mesi di distanza.

Table 3: Occurrence of the edges between hateful  and normal  users, and between suspended  and active  users. Results are normalized w.r.t. to the type of the source node, as in:  $P(\text{source type} \rightarrow \text{dest type} | \text{source type})$ . Notice that the probabilities do not add to 1 in hateful and normal users as we don't present the statistics for non-annotated users.

Node Type	(%)	Node Type	(%)
 $\rightarrow$ 	41.50	 $\rightarrow$ 	13.10
 $\rightarrow$ 	15.90	 $\rightarrow$ 	2.86
 $\rightarrow$ 	7.50	 $\rightarrow$ 	92.50
 $\rightarrow$ 	99.35	 $\rightarrow$ 	0.65

Figura 3.6: Probabilità di retweet tra i vari tipi di profili codificati nello studio. Immagine pubblicata nello studio del Ribeiro e colleghi (2018).

I risultati evidenziano come gli odiatori abbiano un profilo significativamente più recente rispetto agli altri utenti. Analizzando le interazioni generate da questo gruppo di profili emerge che sono anche più attivi degli altri, ricevono più *likes* e più *followers*. Al contrario di quanto ipotizzato inizialmente dai ricercatori però, questi utenti non sembrano utilizzare tecniche di *spamming*, condividono infatti meno url rispetto alla media del campione complessivo. Analizzando le parole utilizzate da questo gruppo al di là di quelle inserite nel vocabolario dei termini d'odio, viene riscontrato che questi utenti utilizzano maggiormente parole collegate ad emozioni positive, negative, legate alla sofferenza e all'amore, ma anche quelle riferibili al tema del lavoro e delle promesse/giuramenti. La parte più interessante della ricerca riguarda forse le interazioni tra gli *hateful users*, gli utenti bannati e tutti gli altri [Fig. 3.6]. Risulta infatti che il 41% dei retweet degli odiatori sono rivolti ad altri odiatori. Considerando che questi utenti sono una percentuale molto ridotta sul totale di quelli presi in considerazione (rappresentano lo 0,68% del campione), la probabilità che interagiscano direttamente tra di loro è addirittura 71 volte più alta rispetto alla probabilità di retwittare altri utenti. Sembra quindi che gli odiatori abbiano un profilo ben specifico e compongano una comunità molto interconnessa.

**Table 1.** Trends in “unjustified incivility or vulgarity”

Target	Percentage of Total
Political Elites	44%
Personal	25%
Out-group	7%
The media	1%
Minorities	1%
Untargeted	22%

**Figura 3.7:** percentuali dei target a cui sono stati indirizzati gli attacchi incivili e volgari. Immagine pubblicata nello studio di Yacubov (2020).

### 3.6 DISCORSI D’ODIO E POLITICA

Risulta rilevante constare come vi siano poche ricerche che mettono in correlazione direttamente l’*hate speech* con il *negative campaing*.

L’unico di cui siamo a conoscenza è uno studio condotto durante le elezioni Statunitensi del 2016 che mette in luce come i commenti d’odio online siano condizionati anche dal tipo di comunicazione politica effettuata dai candidati (Yacubov & Rossini, 2020). Nello studio sono stati analizzati 1501 commenti di odio a partire da un campione originario di circa 26 milioni contenuti pubblicati in risposta a post di Donald Trump e Hillary Clinton. L’odio è stato selezionato tramite l’utilizzo del dizionario sull’*hate speech* reso disponibile da Hatebase. Su questo sottoinsieme di commenti è stata poi effettuata anche una valutazione qualitativa dividendoli tra incivili e intolleranti. I commenti propriamente incivili (700) o con argomentazioni incivili (394) risultano molto più presenti di quelli intolleranti (164). Dall’analisi del target di questi attacchi [Fig. 3.7], emerge come i gruppi politici rappresentino il bersaglio principale, seguiti dai singoli politici, mentre gli attacchi alle minoranze sono l’1% dei commenti analizzati. Questi dati sono in linea con quelli riscontrati nella nostra ricerca, come vedremo più avanti. La relazione tra i discorsi dei politici e i commenti di odio emerge da un’ulteriore analisi qualitativa dei contenuti. Viene riscontrato come gli utenti tendano a riprendere il linguaggio utilizzato dai politici, come ad esempio nel caso della frase “grab her by the pussy”, utilizzata dal presidente in carica durante la campagna e spesso riutilizzata dai suoi sostenitori nei commenti ai suoi post per attaccare altri utenti.

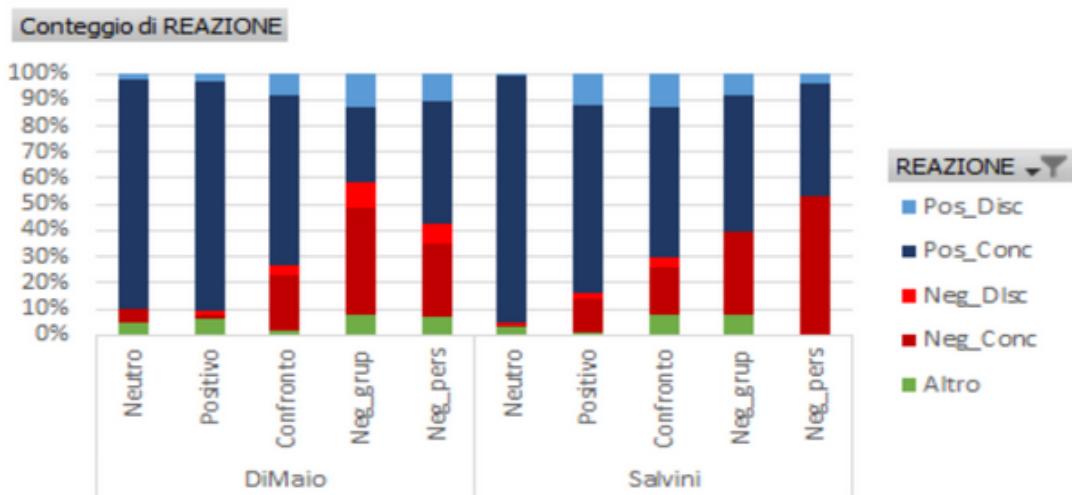
# 4

## La presente ricerca

### 4.1 STUDIO ESPLORATIVO

La presente ricerca nasce da una riflessione avvenuta all'interno del corso di "Psicologia della politica" delle professoressa Carraro e Lenzi all'Università di Padova. Durante il semestre è stato condotto un progetto di gruppo denominato "Hates gonna hate..?" che ha visto la collaborazione di 8 studenti\* (Ilaria Atella, Flavia Calabrese, Enrico della Volpe, Lorenzo Manconi, Giuseppe Mignemi, Salvatore Romano, Marianna Testa, Lavinia Tredici) e una tutor (Dalila Cataldo). Questo progetto, che ha svolto funzione di studio preliminare rispetto alla presente ricerca, si poneva l'obiettivo di esplorare gli effetti delle campagne negative nel contesto online. In quell'occasione sono stati presi in considerazione i due leader politici di Lega e Movimento 5 Stelle in corsa per le elezioni politiche del 2018 (il progetto è stato svolto poco prima delle elezioni europee del maggio 2019). Sia per Di Maio che per Salvini sono quindi stati selezionati 5 post sullo stesso tema (le banche) e con un numero di interazioni paragonabile. I post pubblicati su Facebook sono stati selezionati in base a 5 categorie: un post "neutro", uno di campagna "positiva", uno di campagna "comparativa" e due di campagna "negativa", di cui uno che aveva come target un gruppo e un'altro che aveva come target una singola persona.

Sono stati poi collezionati 100 commenti in modo casuale dai primi 500 pubblicati in risposta ai post considerati, e sono stati ulteriormente codificati nelle categorie di "odio concorde" o "non concorde" con la posizione del politico quando venivano riscontrati insulti, parolacce



**Figura 4.1:** Risultati relativi allo studio preliminare svolto durante il corso di "Psicologia della politica". Vengono messi in relazione il tipo di campagna politica contenuta nei post e i livello di odio presente nei commenti.

o linguaggio scurrite; "positivo concorde" o "non concorde" quando invece il tono era civile, anche se non in accordo con la posizione espressa nel post; "altro" nel caso in cui il contenuto non era interpretabile o usciva dalla discussione iniziata nel post.

Tutti i contenuti sono stati categorizzati da due divers\* valutat\*, con un'omogeneità dei giudizi del 92%, in caso di disaccordo i post e i commenti sono stati rivalutati da tutto il gruppo, arrivando ad ottenere un campione finale di mille contenuti codificati (100 commenti per i 5 post dei 2 politici).

Benché il campione fosse ridotto, si era riscontrata una significativa relazione tra il tipo di campagna politica e il livello di odio nei commenti  $\chi^2$  ( $16, N = 1027$ ) = 204.625,  $p = 0.000$ , evidente anche dal grafico realizzato in quell'occasione [Fig. 4.1], in cui si può vedere come in post negativi si arrivi addirittura a oltre il 50% di commenti di odio.

#### 4.2 LE NUOVE DIREZIONI DI SVILUPPO DELLA RICERCA

Poco dopo questo studio iniziale, è stato pubblicato lo studio di Amnesty International "Barometro dell'odio", come vedremo meglio nel capitolo **Metodologia**. La ricerca presenta la peculiarità di raccogliere contemporaneamente i post e tweet dei politici durante la campagna elettorale, ma anche i relativi commenti, giungendo a un campione di dimensioni considerevoli. Si è deciso quindi di aggiungere al database della ONG anche una codifica relativa al tipo di campagna utilizzata dai politici così da poter indagare, con un campione ben più

esteso, il rapporto tra tipologia di campagna realizzata dai politici e generazione di odio online. Per quanto non sia possibile determinare un principio di causalità diretto tra il tipo di campagna e il tipo di commento di odio che ne scaturisce, poiché la nostra ricerca non si avvale di un modello sperimentale, sicuramente i post dei politici precedono temporalmente le risposte delle/gli utent\* e, quindi, sicuramente le influenzano.

Alla domanda relativa alla relazione tra tipo di campagna e odio nei commenti, ne sono state aggiunte alcune suggerite dalla letteratura del *negative campaign*, e altre che indagano il medium utilizzato, posto che il campione prende in considerazione sia Facebook che Twitter.

Per quanto riguarda l'analisi dei social , alcuni politici utilizzano in modo intercambiabile le due piattaforme analizzate in questo studio. In questi casi viene pensato a priori un contenuto o un messaggio da condividere con la propria base e lo si traduce rispettivamente in un post o un tweet che viene poi pubblicato. Altri politici invece, cambiano il linguaggio e anche i contenuti tra un piattaforma e l'altra. Abbiamo, quindi, ritenuto interessante verificare se il tipo di comunicazione utilizzata dai politici dipenda anche dal medium che decidono di usare se ci sono delle differenze tra l'utilizzo di *negative campaign* tra le due piattaforme. Molti studi hanno dimostrato come i media tradizionali sovraespongano i messaggi di campagna negativa rispetto agli altri tipi di comunicazione (Walter & Vliegenthart, 2010) (K. M. Hansen & Pedersen, 2008) (Haselmayer et al., 2019). Anche in relazione ai social networks, è stato evidenziato come la negatività sia più facilmente apprezzata tramite *likes*, condivisioni e discussioni nei commenti (Conover et al., 2011) (Berger & Milkman, 2010) (Stieglitz & Dang-Xuan, 2013).

La principale differenza tra Facebook e Twitter è sicuramente la lunghezza massima del testo che è possibile condividere: se su Facebook post lunghi e articolati sono più frequenti, su Twitter la lunghezza massima del testo è definita a priori ad un massimo di 280 caratteri. Un'altra differenza è la possibilità di interagire con gli/le utenti: se su Facebook un commento di risposta del politico potrebbe perdersi nella miriade di commenti conseguenti al post, su Twitter, tramite re-tweet e trend, è possibile instaurare vere e proprie discussioni tra politici e utenti comuni. Inoltre, Twitter è comunemente utilizzato per messaggi estemporanei di comunicazione rapida e di commento su fatti che stanno avvenendo in un momento specifico, come avviene spesso nella comunicazione politica, mentre Facebook è più utilizzato per ragionamenti complessi ed estemporanei. Analizzando le statistiche dei post e dei tweet considerati in questo studio, abbiamo cercato indicazioni anche sul diverso uso dei due social network fatto dai politici.

#### 4.3 LE IPOTESI

Le ipotesi di partenza di questa ricerca possono quindi essere suddivise in due gruppi: quelle relative alla teoria del *negative campaign* (paragrafo 5.2), testate sul database dei soli post dei politici, e quelle relative alle conseguenti reazioni di odio (paragrafo 5.3) testate sui commenti in risposta ai post.

Per quanto riguarda la comunicazione politica, abbiamo cercato conferme di alcune teorie riguardo il *negative campaign* presenti nella letteratura e già introdotte nel primo capitolo:

Ip.1) Il *negative campaign* è in aumento rispetto ai dati registrati in altre campagne elettorali? Si vuole verificare l'ipotesi che questo tipo di campagna sia in aumento negli ultimi anni e che sia decisamente più presente rispetto a 20-30 anni fa. Verranno quindi confrontati i dati presenti in letteratura con quelli rilevati in questa ricerca, partendo dal presupposto che risulta difficile fare comparazioni tra studi che prendono in considerazioni definizioni diverse del fenomeno, analizzano mezzi di comunicazione differenti e spesso non riguardano la stessa nazione.

Ip.2) Il *negative campaign* è un fenomeno diffuso prevalentemente all'interno di alcuni schieramenti politici (destra o sinistra), oppure dipende dalla posizione dei vari partiti nei luoghi del potere (governo o opposizione)? Si vuole testare l'ipotesi che questo stile comunicativo derivi dal ruolo istituzionale che gli schieramenti occupano durante la campagna e dalla conseguente strategia elettorale che ne deriva (campagna di conquista o di posizione), più che dall'orientamento politico in sé. Verranno confrontati quindi i livelli di negatività di due schieramenti diversi per orientamento politico con quelli di due schieramenti diversi per posizione istituzionale ricoperta al momento della campagna politica.

Ip.3) Il tipo di campagna politica utilizzato dipende dalla piattaforma che viene utilizzata dai politici per comunicare? In particolare è possibile rintracciare una relazione tra la brevità dei messaggi formulati su Twitter rispetto ai post mediamente più lunghi di Facebook? L'ipotesi è che il tipo di campagna utilizzato derivi dalla strategia politica che quindi dovrebbe risultare indipendente dal medium utilizzato e dal suo stile comunicativo. Verranno confrontati i livelli di negatività del discorso politico suddividendo i post in base al sito di pubblicazione del contenuto (Twitter o Facebook).

Ip.4) I post di *negative campaign* sono in media più virali rispetto a quelli positivi? Si vuole testare l'ipotesi secondo cui questo tipo di campagna negativa sia più efficace e più facile da diffondibile e da far diventare virale, come sostenuto da diversi *campaign stragists* che la prediligono proprio per questo motivo. Verranno confrontati i dati di diffusione dei

contenuti (*likes*, condivisioni e numero di commenti) in base al tipo di campagna utilizzata dai politici.

Per quanto riguarda invece le ipotesi relative ai commenti generati dai vari post:

Ip.1) C'è una relazione tra il tipo di campagna politica proposta e i livelli di odio registrati nei commenti? L'ipotesi è che il livello di insulti e linguaggio scurrile utilizzato dagli/dalle utenti che interagiscono con i contenuti di campagna politica online dipenda anche dal tipo di comunicazione utilizzata dai politici. Essendo i post pubblicati prima dei commenti, questa relazione presupporrebbe un effetto di causalità tra i due fenomeni, fornendo un importante indicazione su come ridurre la produzione dell'odio online. Verranno confrontati i livelli di odio nei commenti (codificati come "hate speech" e "problematici") in base al tipo di campagna politica a cui rispondono (negativa, comparativa, positiva o neutro).

Ip.2) Quali sono i target delle campagne negative e comparative che producono maggiormente odio nei commenti? Oltre al tipo di campagna, anche il target dell'attacco delle campagne negative e comparative può determinare cambiamenti significativi nel tipo di risposta degli/delle utenti. In particolare, l'ipotesi qui presentata è che quando i politici attaccano persone o gruppi non pubblici (privati cittadini o categorie di persone) i livelli di odio siano i più alti in assoluto. Questa ipotesi verrà comparata anche con due ulteriori possibili ipotesi alternative in cui i target vengono raggruppati in modo diverso: a) Target politici (personaggio politico e gruppo politico) contro quelli non politici. In questo caso l'ipotesi è che attaccare personaggi politici genera meno odio rispetto alla discussione sugli altri target poiché si rimarrebbe all'interno della normale dialettica politica. b) Target singolari (privato cittadino, personaggio pubblico e personaggio politico) contro quelli che indicano gruppi di persone. L'ipotesi in questo caso è che i livelli di odio salgono quando si attaccano singoli personaggi e non gruppi.

Ip.3) I livelli di odio sono uguali in entrambe le piattaforme analizzate? La negatività, non solo del tipo di post (Ip.4), aumenta su Twitter? Questa ipotesi si basa sul fatto che la brevità dello stile comunicativo (come per Twitter rispetto a Facebook) sia correlata con toni più accesi, e linguaggio più scurrile. Verranno quindi comparate le statistiche di odio dei due *social network* presi in analisi.



*You cannot kill the truth. You cannot kill justice. You cannot kill what we are fighting for.*

Jean Dominique, Haitian democracy activist.

# 5

## Metodologia

In questo capitolo analizzeremo lo studio di Amnesty International "Il Barometro dell'Odio" che ha fornito il database originario, già in gran parte codificato, che è stato utilizzato per le nostre analisi nel presente lavoro. In seguito vedremo le ulteriori elaborazioni fatte da noi specificamente per questa tesi, discutendone gli aspetti metodologici.

### 5.1 IL BAROMETRO DELL'ODIO

#### 5.1.1 AMNESTY INTERNATIONAL E IL TAVOLO PER IL CONTRASTO AI DISCORSI D'ODIO

Le analisi de "Il Barometro dell'Odio" fanno parte di "Contrasto all'hate speech online", una delle 12 campagne attualmente portate avanti da Amnesty International. Oltre alla ricerca, questa campagna comprende una Task Force "Hate Speech" per la creazione di contro-narrazione sul tema, un videogioco rivolto alla sensibilizzazione dei più giovani chiamato "Hate-Sick" e svariate attività educative per le scuole e per la formazione professionale. Questa articolata iniziativa nasce per rispondere al problema dell'odio online definito, sulla stessa pagina web della campagna, come *diffuso*, poiché coinvolge molti utenti come vittime o colpevoli, *liquido*, perché si propaga ed è difficile da contenere e *pericoloso* perché è causa ed effetto del cambiamento culturale che può portare a discriminazioni anche offline.

---

<https://www.amnesty.it/campagne/contrasto-allhate-speech-online/>

La rete: Action Aid Italia Onlus, Amnesty International Italia, COSPE Onlus; ASGI- Associazione Studi Giuridici sull'Immigrazione, ARCI, Associazione "Carta di Roma", Giulia- Giornaliste Unite Libere Autonome, Lunaria, Pangea Onlus, Vox Diritti, Rete Lenford - Avvocatura per i diritti LGBTI; No Hate Speech Movement Italia, Università Bicocca, Università di Bologna, Università di Firenze, Università di Padova, Reading (UK), Università Statale Milano, Università di Trento, Università di Verona); CNR Palermo, Fondazione Bruno Kessler; OSCAD – Osservatorio per la sicurezza contro gli atti discriminatori, Osservatorio di Pavia, Consiglio Nazionale Forense.

Figura 5.2: Le realtà fondatrici della "Rete nazionale per il contrasto ai discorsi e ai fenomeni d'odio".

Oltre ad aver elaborato una serie di precise richieste ai social media, agli stati europei e alle autorità italiane, questa campagna di Amnesty international ha scritto e creato un vademecum destinato ai singoli cittadini, affinché ognun\* possa fare qualcosa per contrastare questo fenomeno [Fig. 5.1].

Anche grazie a questo progetto, è nata la "Rete nazionale per il contrasto ai discorsi e ai fenomeni d'odio", un importante gruppo di associazioni e università (tra cui quella di Padova) [Fig. 5.2] che coinvolge giornalist\*, avvocat\*, scienziat\* e attivist\* da tutta Italia, con l'obiettivo di ridurre i fenomeni d'odio tramite la ricerca, azioni di *advocacy* e *lobby*, la condivisione di buone pratiche, la promozione di percorsi educativi e, ovviamente, la sensibilizzazione e la mobilitazione della società civile.

Il Barometro dell'Odio consiste in un monitoraggio costante di diversi social network. A partire dalle elezioni del 2018, periodicamente Amnesty International ha coinvolto centinaia di attivisti e decine di ricercator\* e attivist\* nell'aggregazione e nell'analisi di dati quantitativi e qualitativi che testimoniano la gravità dell'*hate speech* online e ne individuano i bersagli principali.

Per ogni studio svolto vengono pubblicati tre documenti: una pagina sul sito dell'associazione con i grafici e i risultati principali; un articolo a cura dei e delle scienziat\* che partecipano al "Tavolo per il contrasto ai discorsi d'odio", con tutte le analisi svolte e le evidenze che hanno fatto emergere, e una nota metodologica. I documenti prodotti riguardo allo studio sulle elezioni europee del 2019, che ha fornito i dati di partenza per la presente ricerca, verran-



Figura 5.1: Una delle cartoline prodotte da Amnesty International Italia contenuta all'interno del "Kit per contrastare l'hate speech"

<https://www.amnesty.it/aiutaci-a-contrastare-lodio-online/>

no descritti qui di seguito (rispettivamente ([Amnesty International Italia, 2019c](#)), ([Amnesty International Italia, 2019b](#)), ([Amnesty International Italia, 2019a](#))).

### 5.1.2 LA RACCOLTA DATI

Amnesty International Italia ha monitorato per 40 giorni (dal 15 aprile fino alla conclusione della campagna elettorale il 24 maggio 2019) i profili Facebook e Twitter dei candidati al Parlamento europeo più attivi online e quelli dei leader di partito ai quali i candidati facevano riferimento. Sono stati raccolti in un primo momento oltre 4 milioni e 250 mila di contenuti dai *feed* dei profili dei politici, ne sono poi stati selezionati 100 mila per l’analisi rispettando alcune proporzioni che vedremo tra poco [Fig. 5.3]. Non si conoscono, a oggi, rilevazioni – né in Italia né in altri paesi dell’Unione Europea – così estese rispetto alla campagna delle scorse elezioni per il rinnovo del Parlamento europeo. La particolarità di questa raccolta sta nel considerare contemporaneamente sia le campagne politiche (cioè i post prodotti dai candidati), sia le reazioni sviluppate dagli utenti nei commenti. La finestra della raccolta dei commenti e delle statistiche di viralità dei post è stata estesa fino al 31 maggio per comprendere anche le interazioni con i contenuti postati dai candidati a fine campagna. Per la raccolta sono state utilizzate le *Application Programming Interface* (API) ufficiali dei due social network, con l’unica limitazione nel caso di Twitter di non poter avere accesso a *tutti* i contenuti, ma solo ad una selezione casuale di essi. Essendo i commenti analizzati nello studio un campionamento altrettanto casuale di quelli raccolti, questa limitazione non costituisce un problema metodologico. Le API di Facebook sono invece limitate ad un massimo di 24 commenti per post, caso riscontrato solo in pochi contenuti analizzati in questo studio, non riducendo significativamente, quindi, la rappresentatività del campione.

Sono stati presi in considerazione i 40 politici più attivi su Twitter, e i 40 più attivi su Facebook in base alla semplice somma del numero di tweet/post pubblicati con il relativo numero di risposte/commenti ricevuti. Per assicurare un’equa rappresentazione degli schieramenti e del territorio, sono stati valutati i *feed* raccolti dai 4 politici complessivamente più attivi per ogni lista e il *feed* di almeno un rappresentante di ogni circoscrizione per lista. Inoltre, per ogni lista sono stati presi in considerazione almeno una donna e almeno un uomo. A seguito della pubblicazione dei risultati delle elezioni sono stati introdotti nel campione altri candidati eletti e non precedentemente considerati. Il risultato consiste in un campione di 77 politici che, oltre a essere stati monitorati, sono stati anche valutati. Sui 27.009 post/tweet raccolti per questi politici ne sono stati valutati l’80% (21.596). Successivamente sono stati selezionati in modo casuale un numero di commenti proporzionato al totale dei commenti ricevuti per



Figura 5.3: Alcune statistiche riassuntive dei risultati dello studio Il Barometro dell'Odio, pubblicate su [www.amnesty.it/cosa-facciamo/elezioni-europee/](http://www.amnesty.it/cosa-facciamo/elezioni-europee/).

ogni post/tweet, partendo da un minimo di 4 e raggiungendo i 2-3mila commenti per post con 100mila commenti totali.

### 5.1.3 LA VALUTAZIONE DEI CONTENUTI

Circa 180 attivisti di Amnesty International Italia sono stati istruiti su come svolgere la valutazione dei contenuti e hanno realizzato il lavoro tramite il sito shinyapps.io in cui venivano assegnati, in modo casuale, una serie di pacchetti da 50 contenuti da valutare. Nell'applicazione è stato mostrato il contenuto testuale; il nome del politico dal cui *feed* proveniva il contenuto; la distinzione tra tweet/post e risposta/commento; nel caso di risposta/commento, è stato mostrato il tweet/post in replica al quale il contenuto era stato pubblicato.

Per ogni contenuto sono stati valutati: il tema (donne, lgbti, disabilità, migranti rifugiati e persone con background migratorio, rom, minoranze religiose, solidarietà, povertà socio-economica, altro); l'accezione (negativa, positiva); se negativa la tipologia (non problematico, problematico, *hate speech*, ambiguo); se problematico o *hate speech* il target (il politico autore del contenuto, un altro politico, l'autore del commento/risposta precedente, un singolo individuo o un gruppo perché riconducibile a una categoria soggetta a discriminazione, altro); la categoria del target (donne, lgbti, persone con disabilità, migranti rifugiati e persone con background migratorio, rom, musulmani, ebrei, un singolo o un gruppo per lo svolgimento di attività di tipo umanitario e/o solidaristico, persone in condizione di povertà socio-economica, non riconducibile ad alcuna categoria/altro). Per la definizione di *hate speech* si è utilizzata quella contenuta nella Raccomandazione di politica generale n.15 dell'ECRI relativa alla lotta contro il discorso d'odio (adottata l'8 dicembre 2015) ([European Commission](#)

against Racism and Intolerance, 2015), già citata del capitolo ”*Hate Speech*”.

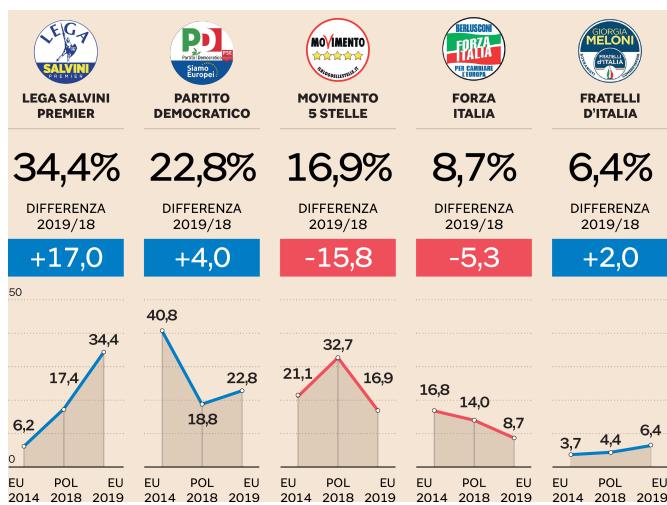
Ogni contenuto è stato valutato da tre valutatori diversi; se le tre valutazioni sono risultate essere coerenti sono state accettate. I contenuti di *hate speech*, quelli ritenuti ambigui da almeno due valutatori e quelli per i quali non c’era allineamento tra i tre valutatori, sono stati rivisti da esperti appartenenti al Tavolo per il Contrasto ai Discorsi d’Odio.

#### 5.1.4 LE ELEZIONI EUROPEE DEL 2019

Prima di affrontare i risultati dello studio, un breve approfondimento sulle elezioni europee e sui risultati dei partiti italiani a questa tornata del 2019. Le elezioni europee assumono una rilevanza secondaria rispetto a quelle nazionali ed hanno caratteristiche proprie come, ad esempio, una minore affluenza complessiva e risultati meno significativi per i partiti in quel momento al governo, dovuti al ”voto di protesta” che in questo tipo di elezioni è più presente (Reif & Schmitt, 1980). Vengono quindi definite ”second-order elections”, anche in funzione del fatto che prevedono budget molto più limitati per la campagna elettorale che in generale dura meno rispetto alle elezioni nazionali. Questo tipo di elezioni concentra le sue tematiche su due filoni principali: quello della politica interna e degli effetti del voto sugli assetti nazionali e la discussione (spesso minoritaria) sulle questioni effettivamente europee. Negli ultimi anni, sempre di più, si è passati a discutere della legittimità del sistema europeo, soprattutto in relazione ai temi della migrazione e a quello dell’*austerity* imposta ad alcuni paesi dopo la crisi del 2008. Anche nella tornata elettorale del 2019 si sono creati due evidenti schieramenti: uno euroscettico e l’altro europeista, dando maggiore rilievo alla discussione su contenuti sovrnazionali. Questo tipo di opinioni contrastanti ha cambiato il posizionamento dei principali partiti nei vari paesi dell’unione e l’equilibrio tra di essi (Meijers, 2017), con effetti importanti anche in Italia, dove almeno uno dei partiti in quel momento al governo ha fatto una campagna dichiaratamente anti-europeista. Se nel 2009 solo il 20% dell’elettorato ha votato partiti antieuropesi, nel 2014 si è arrivati a superare il 50% e nel 2019 è stato addirittura raggiunto il 58%, (Seddone et al., 2019). Nella figura 5.4 possiamo vedere i risultati delle elezioni e le tendenze dei cinque principali partiti.

#### 5.1.5 I RISULTATI DEL BAROMETRO

I due principali risultati di questa ricerca da un punto di vista statistico riguardano l’effetto del tono dei politici e del tema da loro trattato sul tipo di commenti generati. Risulta infatti confermato con significatività statistica che un tono negativo nei post dei politici (post codifi-



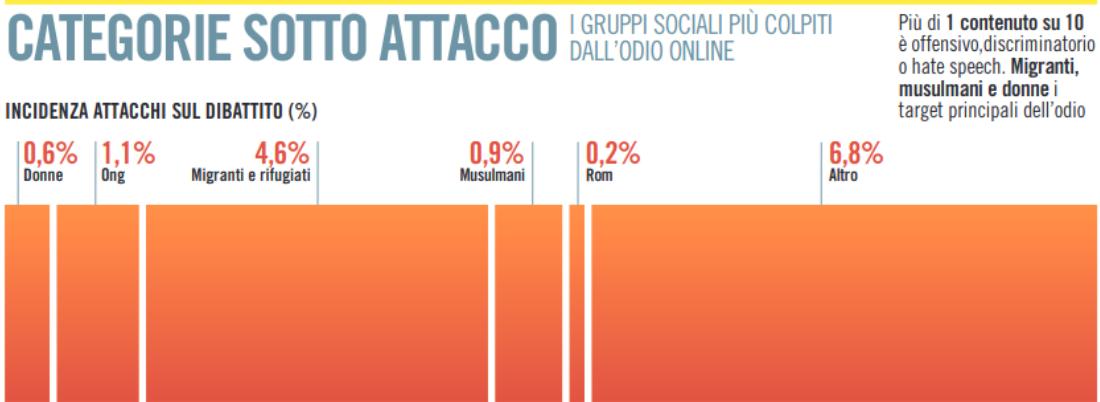
**Figura 5.4:** I risultati delle elezioni europee del 2019 in Italia, con confronti con le precedenti due elezioni. Infografica pubblicata da Weeklymagazine.it

cati come negativi/problematici/*hate speech*) sia in grado di generare più commenti negativi (negativi/problematici/*hate speech*) rispetto a equivalenti post positivi e neutro. L’ipotesi riguardo i temi trattati nei post risulta parzialmente confermata: è stata riscontrata una significatività per i temi ”immigrazione” e ”minoranze religiose”, ma non per quello riguardo il tema ”donne”, come supposto inizialmente [Fig. 5.5].

Vengono inoltre proposte alcune considerazioni più generali riguardo proprio i temi toccati dalla campagna elettorale. Rispetto al precedente Barometro sulle elezioni politiche del 2018, gli attacchi rivolti a comunità migranti non si sono ridimensionate, ma si sono normalizzate passando anche ad attacchi alla solidarità e a chi la promuove. La discussione su questi argomenti risulta ancora più polarizzata, anche a causa dell’intreccio con il dibattito sull’Unione Europea.

Riguardo le categorie utilizzate per individuare i livelli di odio generali, risulta che l’11,5% dei 100 mila contenuti valutati sia risultato essere offensivo e/o discriminatorio o *hate speech*. Limitandoci al solo *hate speech* l’incidenza si attesta invece all’1% del totale dei contenuti. Soffermandosi sui contenuti riguardo l’immigrazione (8% del totale), il 42% di questi risultano discriminatori o *hate speech*, percentuale di poco inferiore solo a quella relativa al tema ”minoranze religiose” che fa riscontrare il 47,5% di messaggi offensivi e molto superiore alla media generale.

Guardando solo ai 21.596 post/tweet dei politici valutati, solo 1 è stato etichettato come *hatespeech* (conteneva riferimenti alla “bonifica sociale e culturale” di un gruppo di per-



**Figura 5.5:** Incidenza delle categorie sotto attacco nei commenti. Immagine pubblicata nel report di Amnesty International sul "Barometro dell'odio: elezioni europee del 2019"

sone rom). Tuttavia ben 415 post/tweet (quasi 2 su 100) sono stati ritenuti offensivi e/o discriminatori e tra questi sono molti quelli al limite dell'hate speech.

Non tutti i/le candidat\* hanno utilizzato i social con uguale intensità ed efficacia: il 51,5% delle interazioni totali è stata fatta registrare da Matteo Salvini. Gli altri esponenti seguono ad ampiissima distanza e solo sei tra loro superano la soglia dell'1%: Luigi Di Maio(14,5%), Giorgia Meloni (8,3%), Silvio Berlusconi (3,6%), Silvia Sardone (2,7%) Carlo Calenda (1,7%), Nicola Zingaretti (1,3%). Questo dato risulta interessante se comparato con quanto emerge dalla sezione “Report della Libreria inserzioni” pubblicata da Facebook in cui è possibile vedere quanto è stato speso dai singoli inserzionisti e, soprattutto, quali contenuti sono stati promossi. Nello studio di Amnesty International viene infatti riportato come Matteo Salvini sia stato il politico a spendere di più su questo social in pubblicità nel periodo considerato dallo studio, arrivando a 128.782 euro per la promozione di 58 post, alcuni dei quali successivamente codificati come problematici. Molti di questi sono inerenti al tema immigrazione: uno di questi pubblicizzava il video di una persona di origine straniera mentre littigava con un controllore su un treno. Sul video sono stati spesi tra i 10 mila e i 50 mila euro, riuscendo a raggiungere 1 milione di utenti e ottenendo 80 mila likes (e reazioni in generale) e più di 22 mila commenti. Questo video rappresenta un esempio chiaro dei temi e della viralità raggiunta dai contenuti del politico, andando a vedere i principali temi trattati e l'incidenza della negatività raggiunta, questo emerge chiaramente [Fig. 5.6]. Inoltre il leader della Lega ha sicuramente investito più di altri candidati sulla comunicazione social, ottenendo risultati nettamente migliori nella diffusione dei suoi contenuti.

	I 4 TEMI PIÙ PRESENTI SUL FEED DEL POLITICO	INCIDENZA SUL TOTALE (%)	POST/TWEET OFFENSIVI E/O DISCRIMINATORI (%)	COMMENTI OFFENSIVI E/O DISCRIMINATORI (%)
<b>MATTEO SALVINI</b>				
Lega	Rifugiati	12,2	41,6	52,6
Interazioni: 2.188.168	Religioni	1,7	63,2	45,3
	Europa	1,7	9,2	1,8
	Un altro politico	8,7	1,1	19,3
<b>LUIGI DI MAIO</b>				
M5S	Solidarietà	2,4	0	2,6
Interazioni: 615.297	Rifugiati	2,4	0	27,9
	Povertà	3,2	0	1
	Un altro politico	13,6	0	5,7

**Figura 5.6:** I temi più citati dai due politici con più interazioni (il tema "altro" non è stato preso in considerazione). I colori utilizzati indicano l'accezione prevalente del tema: verde per i contenuti con accezione negativa inferiore o uguali al 25%, giallo tra il 26% e il 50%, arancione tra il 51% e il 75%, e rosso tra il 75% e il 100%. Immagine pubblicata nel report di Amnesty International sul "Barometro dell'odio: elezioni europee del 2019"

## 5.2 LA CODIFICA DEL *NEGATIVE CAMPAIGN*

A partire dai dati elaborati da Amnesty International, per il presente lavoro abbiamo provveduto alla ulteriore codifica di 10103 post e tweet di politici. Se infatti nel database originario erano presenti circa il doppio dei contenuti prodotti dai politici rispetto a quelli presi in considerazione in questa tesi, per metà di questi post non sono stati raccolti i relativi commenti. Inoltre alcuni commenti (circa 2 mila) non facevano riferimento a nessun post o tweet analizzato. Si arriva così all'utilizzo di 10103 post/tweet e di 78175 commenti, per un totale di 88278 contenuti utilizzati sui 100 mila raccolti nel data base iniziale.

La valutazione dei contenuti è stata possibile grazie alla preziosissima collaborazione di 9 persone con background accademico (Carlotta Baroni, Francesca Carotta, Giorgio Roda, Luigi Romano, Marilina Verna, Paola Biffi, Sara Vergolini, Samuele Maieron, Vanessa Bebi) che, dopo essere state opportunamente

istruite, hanno codificato 2mila post di politici a testa. Ogni post, infatti, è stato codificato due volte da due valutat\* all'oscuro del lavoro dei/delle collegh\*. In caso di non accordo tra i *coders*, il contenuto è poi stato rivisto per garantire omogeneità nell'interpretazione delle categorie anche nei casi più discutibili. L'accordo tra i due valutatori è stato comunque raggiunto nell'87% dei casi. All'inizio del lavoro, dopo un incontro di formazione, è stata fornita



**Figura 5.7:** Un post di campagna neutra, non ha infatti argomento politico.

ad ogni valutatore una scheda con le indicazioni per l'applicazione delle categorie, poi sono stati sottoposti a tutti i/le valutatori gli stessi 500 contenuti, utilizzati come allenamento e come esempio. In seguito c'è stato un confronto collettivo sui dubbi emersi e sulle codifiche errate per garantire la massima omogeneità delle valutazioni. Il lavoro ha richiesto da 7 a 20 giorni di tempo, in base alla velocità di esecuzione delle varie persone che hanno partecipato.

È stato analizzato solo il testo dei messaggi, non si è fatto, quindi, riferimento né alle immagini né ai link presenti nel post e nei tweet poiché la presente analisi voleva concentrarsi sul linguaggio dei politici, e in particolare sulle loro affermazioni, più che sui vari tipi di contenuti multimediali che i social media permettono di condividere.

Le categorie utilizzate per la valutazione sono state fondamentalmente due: la prima prevedeva l'identificazione del tipo di campagna utilizzata nel post. Si poteva scegliere tra "neutro" quando il contenuto risultava impossibile da contestualizzare o quando non si riferiva

a nessun tema politico; "positiva" quando veniva citata una propria proposta politica senza citare avversari politici o quando si promuoveva un proprio appuntamento elettorale; "comparativa" quando veniva citata una propria proposta e contemporaneamente quella rivale, ponendo in contrapposizione le due idee; "negativa" quando invece nessun riferimento alla propria idea veniva citata, ma si citavano posizioni o caratteristiche di altri\*. È stata quindi utilizzata una definizione puramente direzionale, che non comprendeva nessun aspetto valutativo del livello di negatività del contenuto analizzato.



Pina Picierno @pinapic · Apr 29, 2019  
Noi le **#autoblu** avevamo detto che le tagliavamo e le abbiamo tagliate. Loro avevano detto che le tagliavano e invece le aumentano. La differenza è tutta qui.  
[#stellecadenti](#) [#casta5stelle](#)

Figura 5.10: Un tweet di campagna comparativa, viene citata la propria tesi e comparata con quella di, in questo caso, un altro gruppo politico.

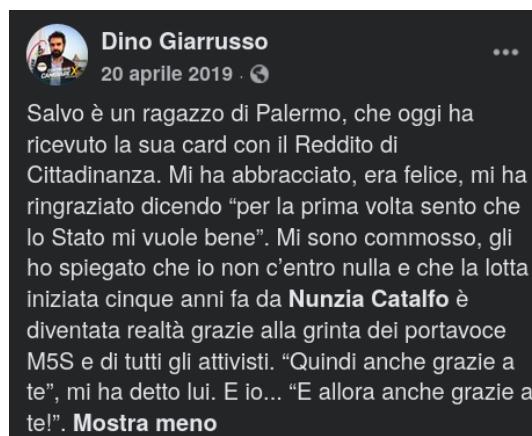


Figura 5.8: Un post di campagna positiva, si rimarca una proposta politica senza attaccare posizioni contrarie.

La seconda categoria utilizzata nelle valutazioni è stata applicata solo ai post classificabili come campagna "comparativa" o "negativa". In questi casi, è stato individuato il target (uno o massimo due) al quale faceva riferimento il contenuto del po-



**Beatrice Brignone** @beabri · 29 Apr 2019

**Eccoli là. Quelli che difendono “le nostre donne”.**

**Altro che uomini. Altro che onore. Altro che italiani veri.**

**Merde. Merde e basta.** [nextquotidiano.it/consigliere-di...](http://nextquotidiano.it/consigliere-di...) via  
@nextquotidiano

**Figura 5.9:** Un tweet di campagna negativa in cui non viene citata la propria proposta politica, ma si attacca solamente, in questo caso, un gruppo politico e un personaggio politico. Si fa riferimento a un consigliere di Casapoud accusato di stupro.

st/tweet. Le tipologie di target definite ex-ante, con cui classificare i contenuti valutati erano: "privato cittadino" quando il post si riferiva a persone non note al pubblico e di cui non si conosceva il nome, come ad esempio, persone fotografate a manifestazioni o comparse in video diventati poi virali; "categoria di persone" quando veniva attaccata una categoria sociale o un gruppo in quanto tale, come per esempio la comunità rom o i "clandestini"; "personaggio pubblico" quando si faceva riferimento a una figura pubblica ma non di ambito politico come per esempio Roberto Saviano o un personaggio della TV; "personaggio politico" per persone impegnate in organizzazioni politiche oppure che ricoprivano, o avevano ricoperto in passato, cariche istituzionali di carattere politico; "gruppo non politico" per gruppi organizzati e riconosciuti in quanto tali come associazioni, ONG, ma anche aziende o associazioni di categoria come i giornalisti; "gruppo politico" per riferimenti a partiti, gruppi politici o schieramenti a tutti i livelli istituzionali. Nelle immagini [Figg. 5.7, 5.8, 5.10, 5.9] possiamo vedere alcuni esempi dei vari tipi di campagna elettorale e dei rispettivi target.

### 5.3 RACCOGLIERE BIG DATA: LE CRITICITÀ DELLE API UFFICIALI

Riguardo alla raccolta dati, una possibile criticità dello studio risiede nell'utilizzo delle API ufficiali, che restituiscono i contenuti delle piattaforme dopo che la *content curation* è stata applicata dalle stesse. Questo significa che i post e tweet rimossi perché non rispettosi delle policy dei social media, non vengono restituiti. È possibile che anche dei messaggi di *hate speech* siano stati rimossi in modo automatizzato o manuale, non comparendo quindi nel database finale. In generale, raccolte dati per studi indipendenti sulle piattaforme social risultano fragili se non basati su dati raccolti indipendentemente dalle piattaforme che si vogliono studiare ed eventualmente sottoporre a critica. A seguito dello scandalo di Cambridge Analytica del 2018 (già citato nell'introduzione del presente lavoro), Facebook ha largamente

ristretto i dati accessibili alle aziende e ai ricercatori, rendendo necessario un ripensamento complessivo dei *computational methods* utilizzati fino a quel momento e portando alcuni studiosi ad interrogarsi sulla necessità di una "Computational Research in the Post-API Age" (Freelon, 2018). Freelon, in particolare, ha riflettuto sulla necessità di utilizzare tecniche di *scraping* per raccogliere questo tipo di dati e sulle possibili implicazioni giuridiche di queste tecniche che consistono fondamentalmente nel "raschiare" (dal verbo inglese *scrape*) in modo automatizzato le pagine HTML dei social, estraendo in modo autonomo le informazioni di cui si ha bisogno dal *document object model* (DOM) con cui sono organizzati i siti. Spesso è possibile però, che questa modalità di raccolta dati sia in violazione dei *terms of service* (TOS) delle piattaforme.

Risulta importante anche rispettare i diritti degli/delle utenti di cui si analizzano i dati, conformandosi agli standard, per l'Europa, del *General Data Protection Regulation* (GDPR). A proporre una riflessione su GDPR e TOS, e a proporre una conseguente soluzione tecnica sono, ad esempio, Mancosu e Vegetti (Mancosu & Vegetti, 2020). Loro mettono in evidenza come le informazioni dei social, benché non più accessibili tramite API, rimangano comunque pubbliche e accessibili sul web. Hanno, quindi, elaborato un *tool* semplice ed efficace per lo *scraping* di profili Facebook.

Bernard Rider, già dal 2013, ha reso disponibile "Twitter Capture and Analysis Toolset" (DMI-TCAT), un software in grado di estrarre praticamente qualsiasi dato disponibile su Twitter tramite API ufficiali, ma registrando più volte al giorno i nuovi contenuti postati, evitando parte della moderazione del social. Nel 2014 ne discutono alcuni usi pratici e possibilità di ricerca (Bruns, Weller, Borra, & Rieder, 2014) poi largamente utilizzati negli anni successivi.

Soluzioni simili per Youtube, Amazon, Pornhub e Facebook sono state proposte dal collettivo Tracking Exposed di cui anch'io faccio parte ([www.tracking.exposed](http://www.tracking.exposed)). Questi *tool*, estensioni browser basate su JavaScript, compiono *scraping* delle pagine social *user-side*, e hanno permesso, tra le altre cose, l'analisi di elezioni nazionali italiane con metodologie innovative di analisi di questi *new media* (Hargreaves et al., 2019) che non prevedono l'utilizzo di API ufficiali. Alcuni sperimentali confronti tra i dati registrati *in vivo* tramite *scraping* e le API ufficiali di Youtube (Agosti et al., 2019), hanno mostrato che il numero di video consigliati da Youtube che si ottengono tramite una richiesta all'API ufficiale differiscono da quelli raccolti utilizzando *tools* di *scraping*, sia con browser privi di cronologia pregressa che (ancora di più), con profili molto personalizzati, come si può vedere nell'immagine estratta dal report di workshop a cui ho partecipato nel 2019 durante la Summer School in Digital



**Figura 5.11:** La differenza tra i video raccomandati da Youtube restituiti dalle API ufficiali e quelli ricevuti da gruppi di utenti con un browser senza cronologia (clean browser) e altri con profili molto utilizzati (personalized).

Methods dell'Università di Amsterdam [Fig. 5.11]. Benchè Youtube sia molto diverso da altre piattaforme social, problemi simili potrebbero sorgere anche analizzando Facebook e Twitter.

Sarebbe interessante, in successivi campionamenti, utilizzare le tecniche e i software sopracitati per verificare se i valori di odio riscontrati con le nostre analisi qui riportate, siano o no minori di quelli reali a causa della metodologia usata nella raccolta dati.

#### 5.4 ANALIZZARE BIG DATA: DIGITAL LABOR E TURCO MECCANICO

Lo sviluppo di nuove tecnologie e la parallela esponenziale crescita delle capacità di calcolo e di immagazzinamento di dati prodotto negli ultimi 20 anni ha dato luogo alla nascita dei *big data*. Questo termine viene riferito a enormi dataset che includono "unstructured data" che richiedono analisi in tempo reale. Da una parte questo tipo di dati pone problemi su quali siano le migliori modalità di organizzazione e gestione dei dati, dall'altra è foriera di nuove opportunità per scoprire varibili nascoste e comprendere fenomeni da un nuovo punto di vista (Chen, Mao, & Liu, 2014).

Anche a livello di organizzazione del lavoro, questa trasformazione ha determinato dei cambiamenti rilevanti. La nascita della "gig-economy" ha portato con sé nuove nuovi status per i lavoratori e nuove organizzazioni del lavoro definite "digital labor" (De Stefano, 2016).

Sono nati le "crowd-sourced/micro-work platforms" e le "work-on-demand platforms". Le prime funzionano con "micro-tasks" che vengono svolti solo online e possono essere ad esempio la classificazione di foto, o la revisione di contenuti online, come ad esempio viene fatto da Amazon Mechanical Turk. Il secondo tipo di piattaforme rende possibile l'incontro tra consumatori e lavoratori che svolgono mansioni fisiche come i guidatori di taxi nell'esempio classico di Uber ([Tan et al., 2020](#)).

Il Turco Meccanico è un marcheggioco costruito da Wolfgang Von Kempelen alla fine del 18° secolo per impressionare la regina Maria Teresa d'Austria. Questo oggetto rappresentava una persona dalla fattezze turche ed era apparentemente in grado di giocare a scacchi in modo magistrale e in modo del tutto automatizzato. All'interno però nascondeva una vera persona, probabilmente di bassa statura, che manovrava la struttura e creava l'illusione di automazione [Fig. [5.12](#)]. Da questa creazione Amazon ha preso il nome per il suo sito di *micro-work*, suggerendo appunto che algoritmi di *machine learning* che richiedono enormi dataset categorizzati per arrivare a prestazioni a volte superiori a quelle umane, in realtà siano basati sul micro-lavoro svolto da migliaia se non milioni di persone sparse per il mondo.

Il presente studio, come quello di Amnesty International, si è confrontato con i *big data*, e ha deciso di impiegare una metodologia di valutazione di questi dati assimilabile ai *micro-tasks* delle *crowd-sourced/micro-work platforms*. Concretamente, sono state utilizzate delle piattaforme digitali basate su *cloud* per distribuire un grande numero di contenuti facenti parte di un unico progetto a vari ricercatori distribuiti in vari luoghi geograficamente anche molto lontani.

La differenza principale tra il presente lavoro di codifica e piattaforme come quella di Amazon risiede nel ruolo del valutatore: non si parla in questo caso di lavoratori e lavoratrici che svolgono "lower-skill tasks", ma di ricercatori e ricercatrici attentamente istruiti\* allo svolgimento di un compito tutt'altro che facile o intuitivo come l'applicazione di categorie come *hate speech* e *negative campaign*. L'altra fondamentale differenza risiede nei diversi obiettivi: nel nostro caso la finalità ultima è una ricerca scientifica e non il profitto, obiettivo tanto di chi ha realizzato la ricerca che di chi ha aiutato con le codifiche. Nel caso della ricerca "Barometro dell'odio" i *coders* si descrivono come attivisti e hanno un obiettivo sia politico che di ricerca, ma la sostanza rimane fondamentalmente diversa da quella di aziende come Amazon.

Benché quindi i due processi partano dalla stessa problematica (l'utilizzo di *big data*) e utilizzino la stessa tecnica (*micro-work cloud*), differiscono per tipo di compito svolto (qualificato, che necessita di studio rispetto a compiti semplici) e per obiettivi (ricerca e non profitto). Queste metodologie simili risultano bene descritte solo dal punto di vista dell'evoluzione delle

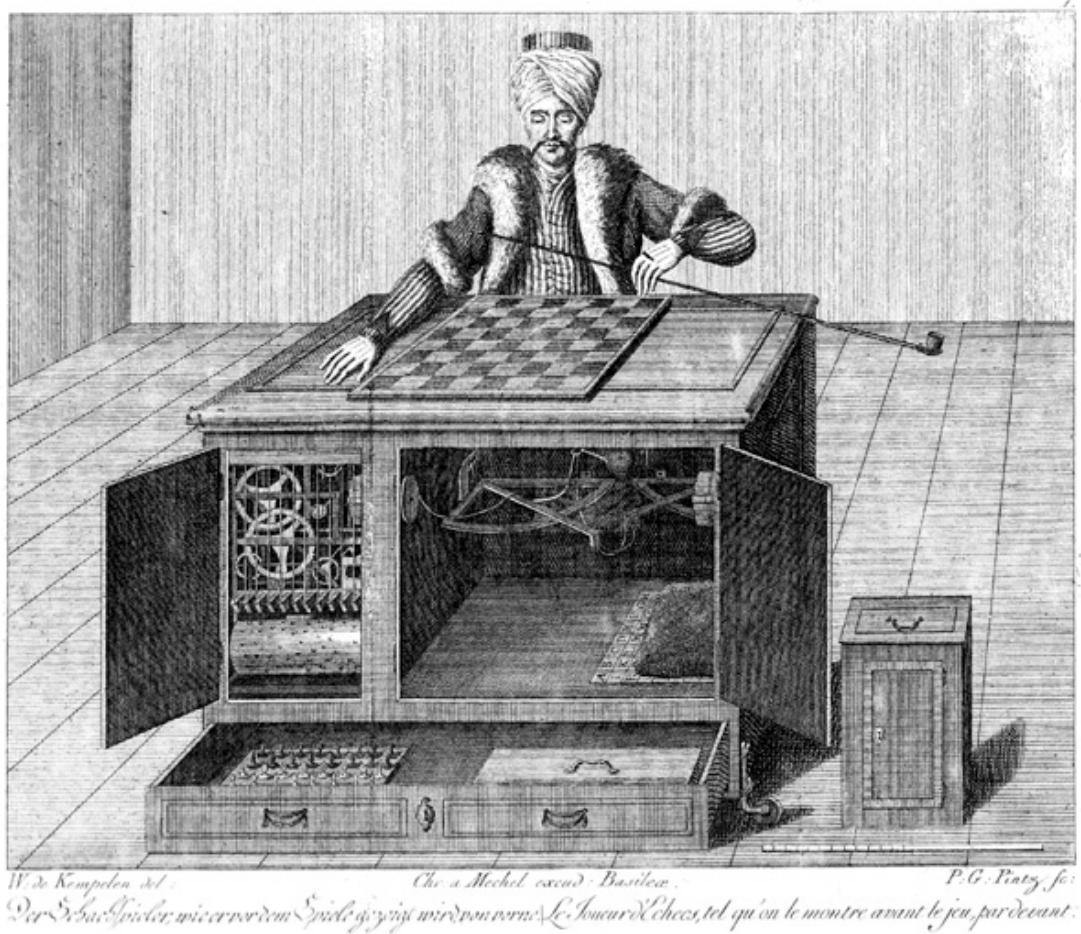


Figura 5.12: Una rappresentazione del Turco Meccanico probabilmente disegnata dal suo creatore Kempelen

forme di lavoro e poco dal punto di vista dell’evoluzione delle tecniche di ricerca, ma risultano fondamentali per poter beneficiare delle più recenti sviluppi tecnologici in un ambito come nell’altro.

La costruzione di *tool* in grado di facilitare questo tipo di lavoro cooperativo all’interno delle università potrebbe risultare una sfida cruciale per il futuro delle ricerche basate sull’analisi qualitativa e manuale di *big data*, la costituzione di reti di collaborazione tra ricercatrici e ricercatori in quest’ottica potrebbe favorire le analisi in molti ambiti del sapere, diventandone un requisito fondamentale, una condizione *sine qua non*.



*The medium is the message. This is merely to say that the personal and social consequences of any medium - that is, of any extension of ourselves - result from the new scale that is introduced into our affairs by each extension of ourselves, or by any new technology.*

Marshall McLuhan

# 6

## Risultati

In questo capitolo verranno descritte le analisi effettuate sui dati codificati e verranno presentati i principali risultati in grado di confermare o specificare meglio la letteratura presa in considerazione.

Essendo le categorie prese in considerazione prevalentemente nominali e basate su una codifica prettamente qualitativa, verrà utilizzato per lo più il test di chi-quadro per testare la significatività delle ipotesi, prediligendo quindi una valutazione qualitativa delle differenze riscontrate nelle percentuali di frequenza rispetto a quelle teoriche.

Tutte le analisi e i grafici sono stati effettuati tramite l'utilizzo del software Python, lo script realizzato è disponibile su Github (<https://github.com/SalvatoreRomano1>), mentre i dati utilizzati, comprendenti le codifiche, sono a disposizione previa richiesta ad Amnesty International - Italia.

### 6.1 IL NEGATIVE CAMPAIGN

In questa sezione verranno presentati i risultati relativi alla codifica di 10103 post di politici sulla base della teoria del *negative campaign*.

#### 6.1.1 IP.1: LE PERCENTUALI DI CAMPAGNA NEGATIVA

Il primo grafico [Fig. 6.1] riguarda il tipo di contenuti creati dai politici e fa riferimento alla prima ipotesi. Come possiamo osservare, i post positivi rimangono di gran lunga i più

Utilizzo dei diversi tipi di campagna politica (DB post)

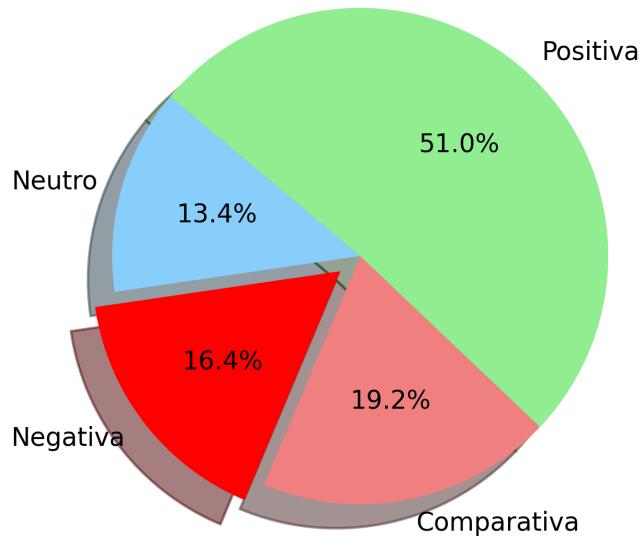


Figura 6.1: L'utilizzo dei diversi tipi di campagna politica sul totale dei post e tweet dei politici.

utilizzati.

Su un totale di 10103 post analizzati, il 51% risultano essere di campagna positiva e il 13% di campagna neutro. I tipi di campagna che comprendono un riferimento a un avversario sono in totale il 35%, rispettivamente il 16% quella negativa e il 19% quella comparativa.

Questi dati rimangono difficilmente confrontabili con quelli di studi precedenti per almeno due motivi: da una parte ogni ricerca analizza solo una parte dei canali informativi all'interno dei quali viene fatta propaganda elettorale e sappiamo che i livelli di odio possono cambiare in base al linguaggio comunicativo dettato dal medium; dall'altra parte l'applicazione della definizione di *negative campaign*, benché nella sua accezione direzionale risulti chiara e sgombra da frantendimenti, è stata declinata con diversi tipi di sottocategorie. È importante aggiungere anche il fatto che i dati relativi ad elezioni nazionali e non europee risultano difficilmente confrontabili a causa delle differenze di coinvolgimento dei diversi tipi di elettorato.

Per quanto poco significativo, il confronto con gli studi relativi alle campagne elettorali degli Stati Uniti sembra indicare, comunque, una minore presenza di campagne negative in Italia. Guardando ad articoli che analizzano questo tipo di percentuali negli ultimi decenni

(Geer, 2012) (Druckman et al., 2010) (Wesleyan Media Project, 2018), possiamo dire che i dati mostrano risultati più vicini a quelli riscontrati in europa (Walter, 2014), in particolare in uno scenario politico non bipartitico (Papp & Patkós, 2018), come registrato in altri studi che prendono in considerazione l'Italia (Ceron & Adda, 2016) (Seddone et al., 2019) citati nel capitolo **"La campagna politica negativa"**.

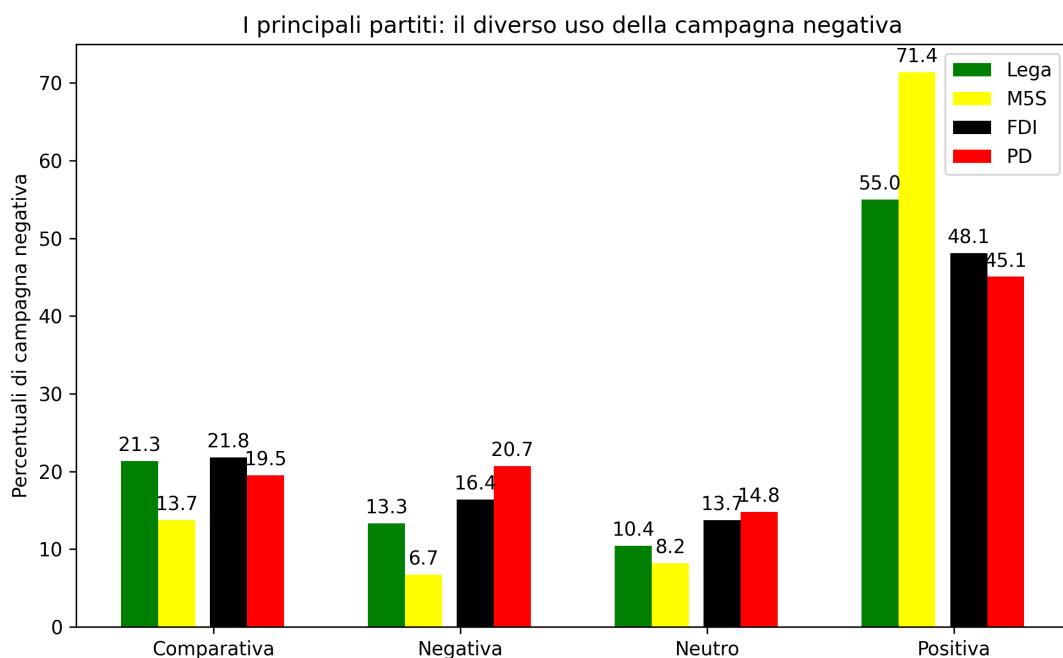
#### 6.1.2 IP.2: NEGATIVE CAMPAIGN: CHI LO USA?

Le percentuali dei vari tipi di campagna politica non risultano omogenee attraverso tutto l'arco elettorale. Dai dati raccolti emergono delle percentuali molto diverse a seconda del partito analizzato. Nel primo grafico di questo paragrafo [Fig. 6.2] possiamo osservare i principali quattro partiti candidati alle elezioni e le loro diverse strategie comunicative. Nell'insieme, essi rappresentano il 75% di tutti i post e tweet raccolti. Risulta che PD e Fratelli d'Italia sono i partiti che usano maggiormente, in percentuale sui post analizzati, rispettivamente la campagna negativa e quella comparativa. Il partito che al contrario fa riscontrare una percentuale largamente maggiore di campagna positiva risulta essere il Movimento 5 Stelle con oltre il 71% di utilizzo di post di questo tipo tra i suoi candidati alle europee.

Nella tabella riportata di seguito [Fig. 6.3] possiamo vedere le percentuali di tutti i partiti a confronto, anche quelli minori non presenti nel grafico precedente. Emerge che anche formazioni minori, che rappresentano poco meno del 5% dei post-tweet analizzati (+Europa e La Sinistra), hanno proposto confronti e attacchi a rivali con una frequenza maggiore rispetto alla media di tutti i partiti considerati contemporaneamente. La Sinistra risulta essere il più propenso a questo tipo di comunicazione, con valori intorno al 30%, sia per la campagna comparativa che per quella negativa.

Sembra interessante, quindi, chiedersi quale sia il motivo alla base di differenze così marcate. A tal fine, sono state prese in considerazione due ipotesi. La prima è che l'utilizzo di campagne negative e comparative sia dovuto prevalentemente alla posizione istituzionale ricoperta dal partito, cioè se esso sia al governo oppure all'opposizione a livello nazionale.

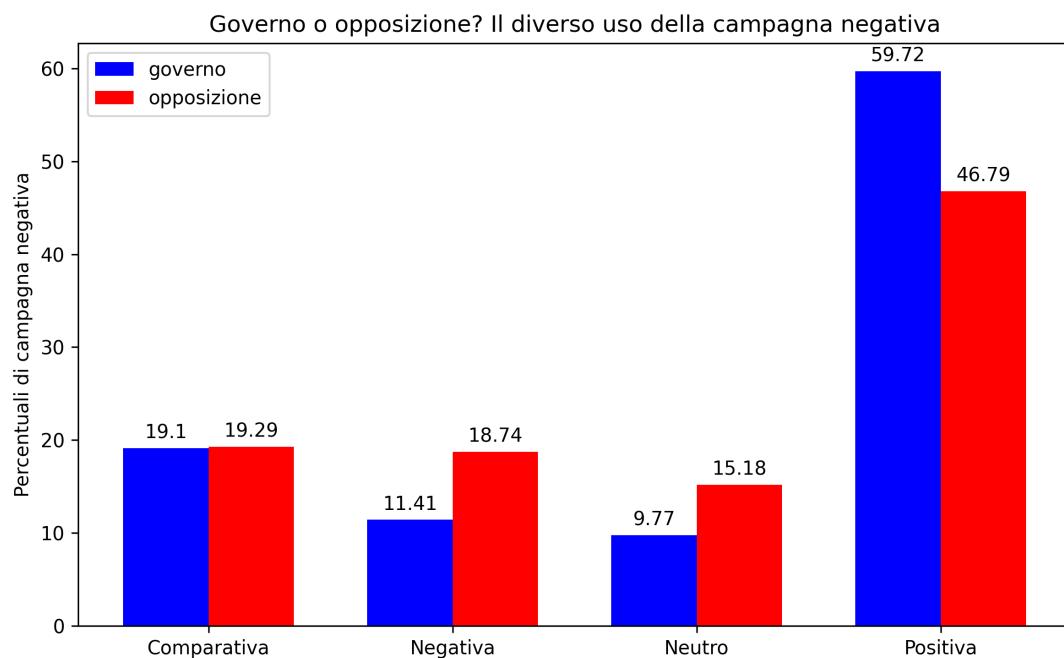
Benché stiamo analizzando la campagna per le elezioni europee, come detto anche nel capitolo **"Metodologia"**, risultano comunque pochissimi riferimenti a ciò che è stato fatto dei rispettivi partiti nella legislazione europea appena conclusa; ci si è piuttosto concentrati sull'idea futura di Europa e sulla politica interna (Seddone et al., 2019). Facciamo quindi riferimento agli schieramenti in carica (Lega e M5S) o meno nel parlamento nazionale al momento della campagna elettorale. Raggruppati i partiti in due gruppi, le differenze nell'utilizzo dei tipi di campagna vengono analizzate tramite test di chi-quadro. Questo test verrà poi



**Figura 6.2:** I quattro principali partiti e il loro utilizzo dei tipi di campagna politica a confronto. Le percentuali di tipo di campagna per ogni partito

p_PARTITO									
p_PARTITO	+ Europa	FDI	FI	La Sinistra	Lega	M5S	PD	Possibile	All
p_campagna									
Comparativa	26.56	21.84	13.10	28.43	21.28	13.69	19.47	12.78	19.43
Negativa	26.76	16.36	13.42	30.88	13.32	6.69	20.67	11.82	16.40
Neutro	12.66	13.66	15.65	7.84	10.41	8.17	14.78	30.99	13.16
Positiva	34.02	48.14	57.83	32.84	54.99	71.44	45.07	44.41	51.01
All	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

**Figura 6.3:** Le percentuali di tipi di campagna utilizzati da tutti i partiti.



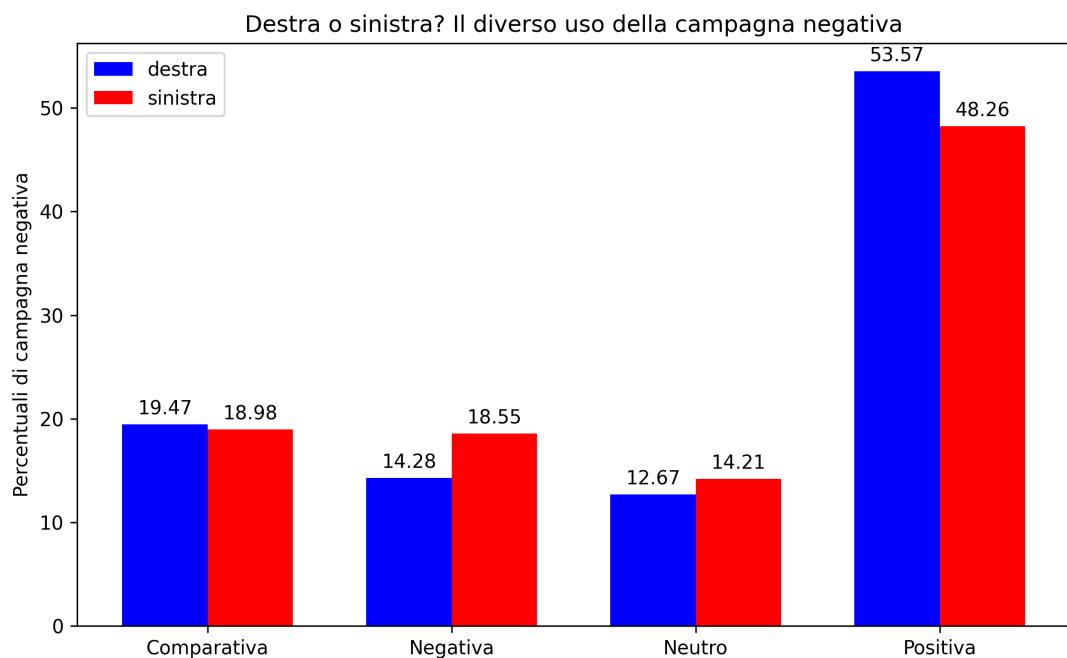
**Figura 6.4:** Le percentuali di tipi di campagna utilizzati dai partiti raggruppati in base alla posizione istituzionale (governo vs opposizione).

confrontato con un secondo test in cui il raggruppamento avverrà in base all'appartenenza del partito a posizioni di sinistra o di destra.

Come vediamo con il test successivo, benché la campagna comparativa risulti utilizzata in egual misura dai due schieramenti divisi in base al loro ruolo istituzionale, nella campagna comparativa e in quella positiva le percentuali risultano opposte tra i due gruppi [Fig. 6.4]. Chi è al governo utilizza molti più messaggi in cui racconta le sue proposte in positivo e parla degli obiettivi raggiunti negli ultimi mesi. Chi è all'opposizione cerca di screditare gli avversari con messaggi negativi, si potrebbe supporre che lo faccia per guadagnare visibilità sui media.

Effettuando un test di chi quadro per verificare la significatività di queste differenze riscontrate, i risultati confermano l'ipotesi:  $\chi^2 (7, N = 10103) = 193.532, p = 0.0000$ . Data l'alta numerosità del campione, interessante il dato del chi-quadro di Pearson che mostra una differenza molto elevata tra le frequenze sottoposte al test.

Sondando la seconda ipotesi, raggruppando cioè i partiti in base allo schieramento ideologico a cui appartengono, otteniamo dei risultati che confermano delle differenze tra destra e sinistra, ma meno marcate rispetto al test precedente [Fig. 6.5]. Interessante il fatto che la campagna comparativa rimanga invariata nelle sue percentuali, dimostrando co-



**Figura 6.5:** Le percentuali di tipi di campagna utilizzati dai partiti raggruppati in base alla posizioni di governo (destra vs sinistra).

me ci sia un'effettiva differenza con la campagna negativa, le cui percentuali invece variano leggermente.

Effettuando un test di chi-quadro anche in questo caso i risultati mostrano un p-value significativo allo stesso modo, ma con un valore del test di Pearson minore:  $\chi^2 (7, N = 10103) = 46.8791$ ,  $p = 0.0000$ . Possiamo quindi affermare che questo tipo di ipotesi risulta meno utile per spiegare cosa spinge a utilizzare una strategia comunicativa rispetto ad un'altra.

#### MOVIMENTO 5 STELLE: SINISTRA O DESTRA?

Per questo test il Movimento 5 Stelle è stato classificato come un partito di sinistra, tuttavia la categorizzazione di questo specifico partito rimane una questione aperta. Alcune posizioni (come quelle sull'ambiente) possono essere sicuramente assimilate a quelle della sinistra classica, sulla questione dell'immigrazione, invece, soprattutto nel periodo di governo con la Lega, le posizioni sono state più tendenti a quelle della destra populista. Il problema alla base di questo dilemma rimane il fatto che i rappresentanti del partito stesso hanno sempre preferito porsi in opposizione a tutta la "casta" rifiutando esplicitamente di essere inseriti in uno dei due schieramenti classici, mirando ad essere votati dagli elettori di entrambi e finendo per

governare prima con la destra e poi con la sinistra. Provando a spostare il Movimento nello schieramento di destra, comunque, le considerazioni generali rimangono invariate poiché il test di chi-quadro, anche in questo caso, rileva minori differenze nell'utilizzo del *negative campaign* rispetto alla teoria della posizione di governo:  $\chi^2 (7, N = 10103) = 70.8986$ ,  $p = 0.0000$ .

In conclusione, l'ipotesi riportata in diversi articoli della letteratura sul genere (K. M. Hansen & Pedersen, 2008) (Druckman et al., 2010) (Walter, 2014) (Curini & Martelli, 2010), secondo cui il *negative campaign* risulta prevalente nei partiti che devono conquistare il potere rispetto a quelli che devono mantenerlo poiché si trovano già al governo, risulta confermata e avvalorata da confronti con altre ipotesi alternative che, seppur in grado di evidenziare alcune differenze, risultano meno efficaci.

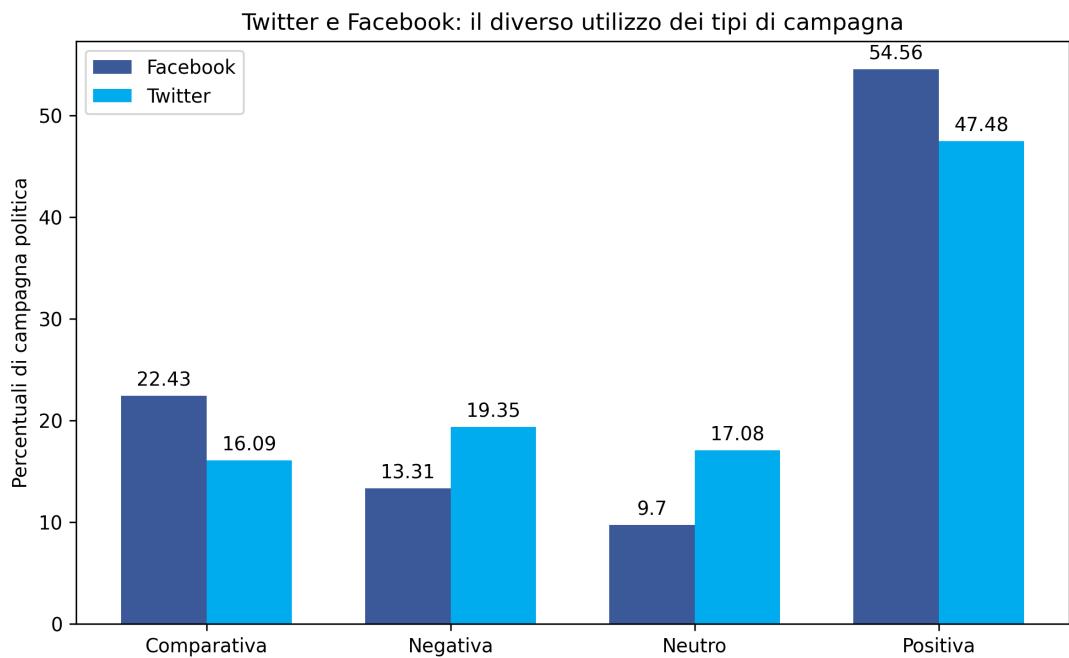
#### 6.1.3 IP.3: IL NEGATIVE CAMPAIGN IN RELAZIONE AL TIPO DI SOCIAL NETWORK

A seguito delle riflessioni contenute nei capitoli precedenti sui due social presi in considerazione, risulta interessante indagare come la comunicazione politica cambi a seconda del medium utilizzato. Comparando i valori generali attraverso tutti i partiti [Fig. 6.6] risulta effettivamente che Facebook è maggiormente utilizzato per campagne comparative e positive che naturalmente necessitano più spazio per essere argomentate. Le campagne negative invece risultano più utilizzate sulla piattaforma dei cinguettii, che sembra quindi prediligere attacchi diretti piuttosto che confronti tra i propri temi e quelli degli avversari.

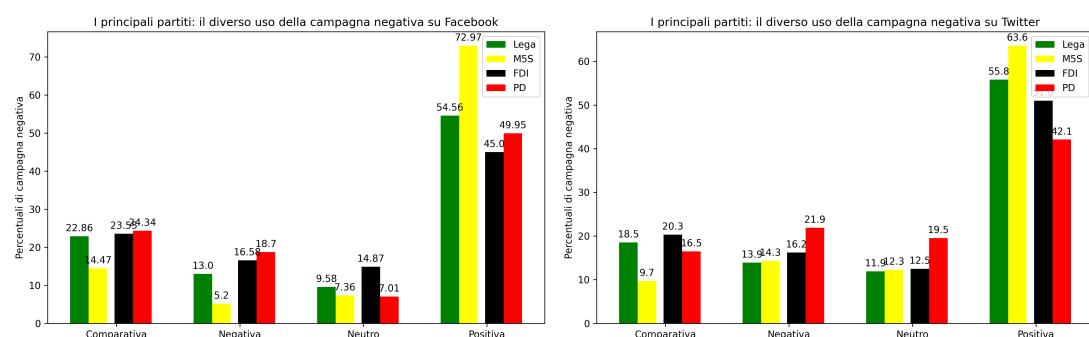
Anche verificando la lunghezza media del testo dei post dei vari tipi di campagna su entrambe le piattaforme risulta che in quelle comparative vengano utilizzati più caratteri rispetto alle negative (Comparativa = 65.0, Negativa = 40.1). E questo fenomeno persiste anche confrontando le lunghezze medie di questi tipi di campagna sui social presi singolarmente: su Facebook quella comparativa ha una media di 86.40 caratteri, la negativa di 53.04 ( $N = 5002$ ); su Twitter invece la comparativa è mediamente di 35.55, la negativa di 31.34 ( $N = 5101$ ).

Analizzando l'utilizzo comparato di Facebook [Fig. 6.7] e Twitter [Fig. 6.8] da parte dei quattro principali partiti (Lega, M5S, FDI, PD) presi singolarmente e poi la differenza tra le percentuali dei diversi tipi di campagna tra le due piattaforme [Fig. 6.9], emergono le stesse indicazioni dell'analisi aggregata.

Tutte le formazioni in campo prediligono Facebook per le comparazioni tra i propri temi e quelli degli avversari. A parte FDI e Lega (che presentano variazioni minime) tutti i partiti fanno riscontrare un aumento percentuale delle campagne negative su Twitter rispetto a Facebook. Questo dato è più evidente nei due partiti minori che hanno utilizzato maggior-



**Figura 6.6:** Le percentuali di tipi di campagna in base alla piattaforma sul quale sono stati postati.



**Figura 6.7:** I principali partiti a confronto: utilizzo dei vari tipi di campagna politica su Facebook.

**Figura 6.8:** I principali partiti a confronto: utilizzo dei vari tipi di campagna politica su Twitter.

p_PARTITO		+ Europa	FDI	FI	La Sinistra	Lega	M5S	PD	Possibile	All
p_campagna										
<b>Comparativa</b>	17.56	3.28	7.15		7.48	4.35	4.73	7.81	17.88	6.25
<b>Negativa</b>	-19.17	0.42	-9.84		-22.31	-0.88	-9.09	-3.16	-2.96	-6.17
<b>Neutro</b>	-1.96	2.32	-12.03		3.65	-2.28	-4.98	-12.48	-24.28	-7.20
<b>Positiva</b>	3.58	-6.03	14.72		11.17	-1.19	9.33	7.83	9.36	7.11
<b>All</b>	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00

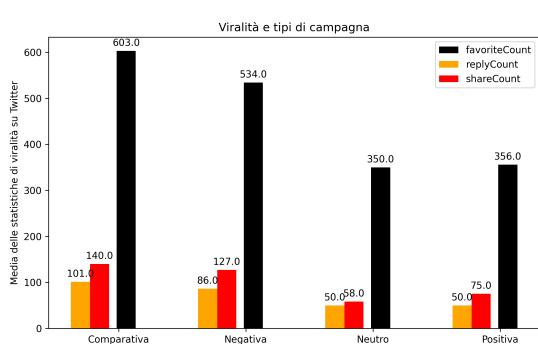
**Figura 6.9:** Differenza tra le percentuali dei vari tipi di campagna politica per i singoli partiti: ai valori riscontrati su Facebook vengono sottratti quelli riscontrati su Twitter. A valori positivi corrisponde quindi una maggiore percentuale di utilizzo di un certo tipo di campagna su Facebook rispetto a Twitter.

mente questo tipo di campagna (+Europa e La Sinistra), in percentuale sul totale di post e tweet pubblicati, arrivando a un aumento di circa il 20% nel passaggio da una piattaforma all'altra.

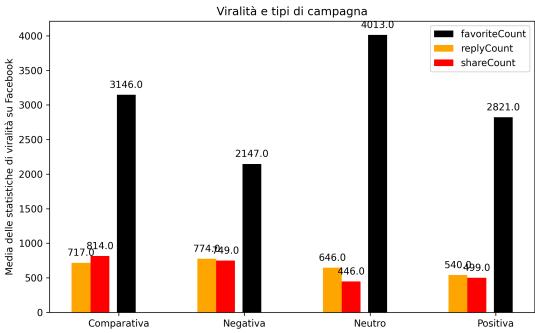
In conclusione: tutti i partiti, a seconda della piattaforma social utilizzata, fanno riscontrare percentuali diverse nelle tipologie di campagna proposta. Questa differenza sembra andare nella direzione di una maggiore negatività per Twitter e un maggiore utilizzo di messaggi comparativi su Facebook. Le variazioni sono probabilmente collegate alla lunghezza media dei due tipi di messaggio, che essendo più lunghi per la campagna comparativa e più brevi per quella negativa, si conciliano meglio con le regole comunicative dei due diversi social. Il medium indirizza quindi le forme di attacco politico.

#### 6.1.4 IP.4: IL NEGATIVE CAMPAIGN CONTRIBUISCE A FAR DIVENTARE VIRALI I CONTENUTI?

Riprendendo la letteratura sul tema, si è deciso di testare la viralità dei vari tipi di campagna politica. Su Twitter, post comparativi e negativi hanno una media di *likes* quasi doppia rispetto agli altri due tipi di campagna che non prevedono di citare avversari politici [Fig. 6.10]. Anche per quanto riguarda le risposte ai tweet e la ricondivisione degli stessi, i valori sono addirittura più che doppi per la campagna comparativa e almeno un 30% più elevati per quella negativa.



**Figura 6.10:** Differenza tra le medie dei principali indicatori di viralità e i tipi di campagna politica su Twitter.



**Figura 6.11:** Differenza tra le medie dei principali indicatori di viralità e i tipi di campagna politica su Facebook.

Su Facebook invece, i risultati sono meno univoci. Senza considerare i post neutri che fanno riscontrare il valore più elevato di likes in media, ma che, appunto, non hanno argomento politico e andrebbero analizzati uno per uno per capire le ragioni di tanto apprezzamento, i post comparativi fanno registrare una media più elevata nel numero di likes, seguiti da quelli positivi ed infine da quelli negativi. [Fig. 6.11]. Le statistiche di ricondivisione e risposte però, anche su questo social, osservano mediamente un 30% percento in più circa nei casi in cui vengono citati gli avversari rispetto a quando non lo sono. A parte la piccola anomalia della media di likes riscontrata nei post negativi su Facebook, che potrebbe in qualche modo spiegare il perché questo tipo di campagna venga utilizzata meno sul social di Zuckerberg rispetto a Twitter, come discusso nell’ipotesi 3, la letteratura e l’ipotesi iniziale qui presa in considerazione risultano confermate: la negatività dei messaggi politici è in grado di aumentare la diffusione dei contenuti.

Per testare la significatività delle differenze tra i gruppi presi in considerazione, è stato utilizzato il test di Kruskal-Wallis. La numerosità dei post che compongono tutte le categorie relative ai tipi di campagna utilizzata non è omogenea e gli assiomi dell’analisi della varianza non risultano verificati (la distribuzione dei valori non è normale, le rispettive varianze non risultano omogenee). Risulta quindi impossibile confrontare direttamente le medie con test come One-Way-ANOVA. Si ricorre quindi all’utilizzo delle mediane, che sono indici più robusti, e possono essere confrontate anche in casi come questo. I risultati [Fig. 6.12] confermano quanto già osservato nei grafici precedenti: i sei test effettuati (per confrontare i tre indici di viralità per le due piattaforme) restituiscono in tutti i casi p-value significativi e valori della statistica di test in media sopra i 100 punti. L’ipotesi risulta quindi pienamente confermata anche da questo punto di vista.

### Test di Kruskal-Wallis [Facebook]:

variabile analizzata	test statistico	p-value
Mediana dei likes	55.252	6.065e-12
Mediana condivisioni	272.214	1.024e-58
Mediana risposte	207.907	8.250e-45

### Test di Kruskal-Wallis [Twitter]:

variabile analizzata	test statistico	p-value
Mediana likes	132.754	1.378e-28
Mediana condivisioni	272.214	1.024e-58
Mediana risposte	258.403	9.959e-56

Figura 6.12: Risultati del test di Kruskal-Wallis sulla viralità in base al tipo di campagna utilizzato. Tutte le differenze di mediana risultano largamente significative

## 6.2 L'ODIO NEI COMMENTI

In questa sezione cercheremo di indagare la relazione tra l'odio online registrato nei commenti e i post pubblicati dai politici. Verranno considerati quindi tutti i 78175 commenti categorizzati in base ai livelli di odio e i 10103 post analizzati in base al tipo di campagna politica.

Relativamente alla dimensione generale del fenomeno, va segnalato che i livelli generali di odio registrati già nello studio di Amnesty (vedi capitolo "Metodologia"), evidenziano un problema che forse, in questo caso, potrebbe risultare minore delle aspettative dell'opinione pubblica. Se tra i post dei politici solo un contenuto è stato classificato come propriamente *hate speech*, tra i commenti ne sono stati codificati 436 (1%) [Fig. 6.13].

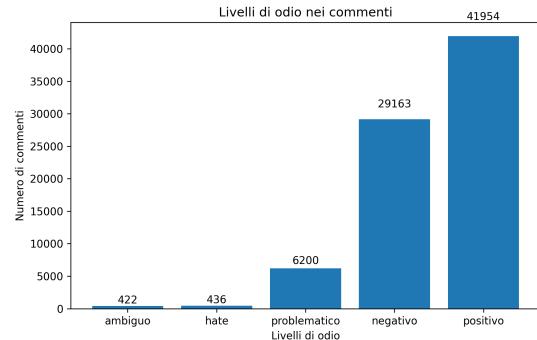
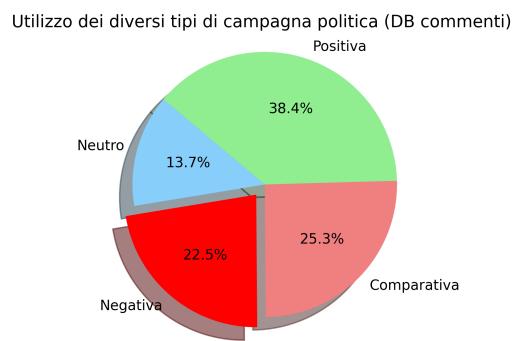


Figura 6.13: Livelli di odio generali riscontrati all'interno del campione

Percentuali più rilevanti sono riscontrate per i messaggi degli utenti che utilizzano linguaggio incivile pur senza sfociare nella discriminazione di minoranze a rischio: 6200 casi equivalenti all'8% del campione. Sicuramente il tipo di contenuti prevalenti è quello positivo 54%, seguito dai commenti negativi, ma civili, che valgono il 34% del totale.

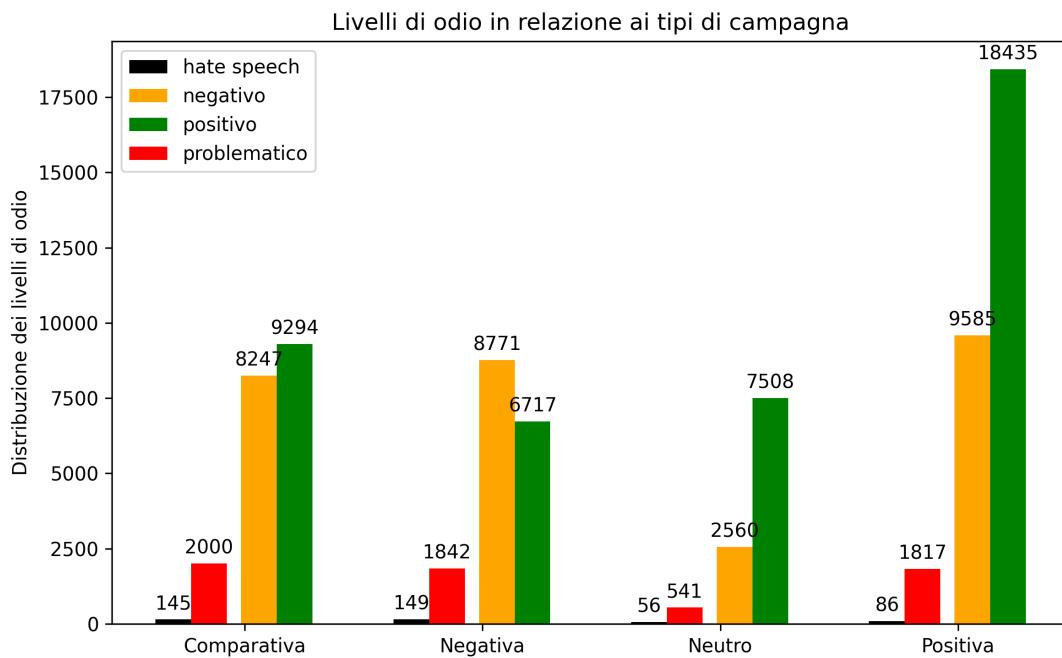
È importante sottolineare che prendendo in considerazione i commenti relativi ai post, le percentuali in cui i vari tipi di campagna sono presenti è differente rispetto al campione costituito dai i soli post [Fig. 6.14]. Il numero di commenti raccolti e poi analizzati per ogni post di politico dipende dalla numerosità dei commenti presenti in relazione a quello specifico contenuto, ed è stata pensata per comporre un campione rappresentativo di tutti i politici in corsa alle elezioni e del loro effettivo seguito sui social.



**Figura 6.14:** Percentuali dei tipi di campagna riscontrati nel database dei commenti

commenti di risposta ai post negativi non hanno la stessa percentuale sul totale dei commenti (22,5%) che hanno i post negativi sul totale dei post (16%). Lo stesso vale per i commenti di post comparativi (25% rispetto al 19,2% nel campione dei soli post) e per quelli positivi (38,4% rispetto al 52%). Le categorie negative e comparative risultano quindi sovradimensionate, come abbiamo avuto modo di spiegare nelle precedenti ipotesi, questi tipi di campagna risultano più virali e quindi anche più commentate. Ciò ci porta ad avere più commenti per post nel campione complessivo che è proporzionato al numero di risposte ad ogni contenuto di politico. Questo non dovrebbe però influire sul tipo di analisi che faremo nei prossimi sottoparagrafi, che partono dall'assunto che i tipi di campagna non sono comunque rappresentati in egual misura e vengono quindi comparati in base alle percentuali di commenti negativi tramite test di chi-quadro.

Il campione non è stato pensato per essere omogeneo nei tipi di campagna politica, che comunque sono stati categorizzati in un momento successivo allo studio de "Il barometro dell'odio". La differenza nella variazione delle percentuali non è in nessun modo spiegabile o legata a questo studio, dipende piuttosto dal campionamento e ha come effetto delle variazioni non tali da invalidare le successive analisi, ma comunque degne di essere spiegate. Per chiarire con un esempio: il numero di



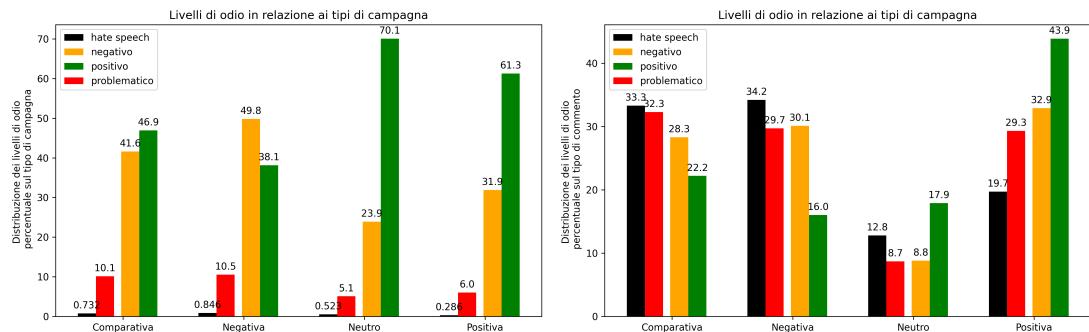
**Figura 6.15:** Valori assoluti dei livelli di odio in relazione con i tipi di campagna politica.

#### 6.2.1 IP.I: IL NEGATIVE CAMPAIGN GENERA PIÙ ODO IN NEI COMMENTI?

Questa domanda risulta essere quella principale all'interno di questo studio poiché permette di sfruttare al meglio tutte le potenzialità del database, utilizzando contemporaneamente post e relativi commenti.

Dai valori assoluti dei contenuti di odio presenti nei commenti [Fig. 6.15] possiamo già fare delle considerazioni. L'*hate speech* compare in uguale misura nelle reazioni a post comparativi e negativi, con valori molto più alti rispetto ai post positivi, nonostante i commenti relativi a questo tipo di post siano più numerosi. I commenti positivi mostrano, invece, un andamento diametralmente opposto. I commenti problematici fanno registrare il valore assoluto più alto in risposta a post comparativi, e risultano uguali in risposta a campagne negative e positive, nonostante la disparità di numero di commenti totale. Interessante che la campagna comparativa faccia registrare più commenti positivi che negativi, mentre nei post negativi avviene il contrario.

Per contestualizzare meglio i valori assoluti, è utile vedere come il numero dei vari tipi di commenti si rapporta con il totale dei commenti di ogni tipo di campagna. Per farlo sono state calcolate le percentuali delle categorie dei commenti in base alle categorie dei post dei



**Figura 6.16:** Livelli di odio nei commenti in relazione ai tipi di campagna politica. Percentuali relative al totale di tipo di campagna. **Figura 6.17:** Livelli di odio nei commenti in relazione ai tipi di campagna politica. Percentuali relative al totale alle categorie dei livelli di odio

politici [Fig. 6.16]. Da qui emerge ancora più chiaramente come i commenti problematici siano più prevalenti quando il politico cita l'avversario nei suoi post: in risposta alla campagna negativa si supera il 10%, poco sotto si collocano i commenti a post comparativi, mentre la percentuale di questo tipo di risposta per i positivi si ferma a quasi la metà. In generale da questo tipo di grafico si comprende meglio come i post negativi generano le maggiori percentuali di tutti i tipi di contenuti classificabili come *hate speech*, problematici, e negativi. Effettuando un test di chi-quadro, le differenze nelle frequenze risultano significative con un valore del test di Person molto alto e un p-value molto basso, avvalorando ulteriormente la conferma delle ipotesi:  $\chi^2 (7, N = 78175) = 4024.679, p = 0.0000$ .

Il ruolo cruciale del *negative campaign* nel generare odio emerge in modo ancora più evidente se si fa riferimento alla ripartizione dell' *hate speech* per tipologia dei post che lo ha generato. [Fig. 6.17]. Risulta infatti che il 67,5% di tutti i commenti di *hate speech* sono stati registrati sotto i post che citano avversari. Senza questo tipo di campagna l'odio sarebbe poco più del 30% del totale. Da questo dato si possono trarre, quindi, indicazioni importanti per identificare delle strategie concrete utili per ridurre i livelli di odio presenti sul web. Evitare o ridurre campagne negative e comparative potrebbe ridurre conseguentemente anche i livelli di odio nei commenti. Indicazioni ancora più precise emergono dall'analisi del tipo di target indicato nelle campagne negative e comparativa.

## 6.2.2 IP.2: QUALE TARGET GENERA PIÙ ODISSEI NEI COMMENTI?

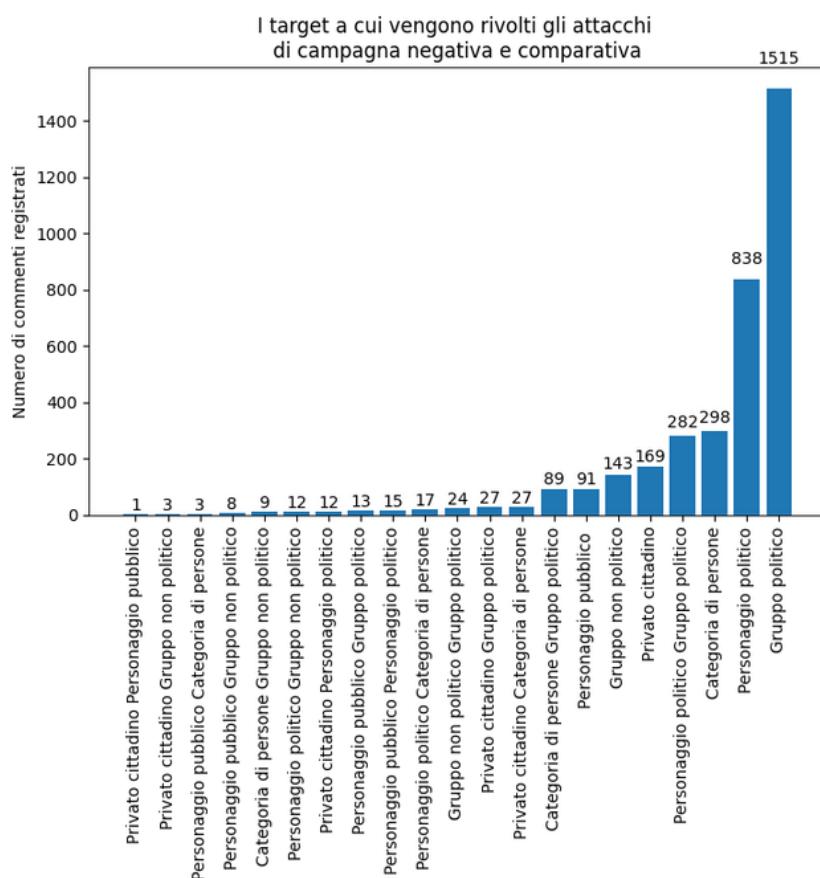
Prendendo in considerazione solo i post di campagna negativa e comparativa, possiamo analizzare i target che vengono maggiormente utilizzati in questo tipo di attacchi verbali e le

conseguenze sui livelli di odio nei commenti. I post/tweet negativi e comparativi sono 3009, cioè il 30% dei post prodotti. I commenti campionati in risposta a questi post sono 37413, il 44.8% di tutti i commenti considerati. A partire da questo nuovo *subset* potremo capire meglio cosa scatena la generazione di odio online.

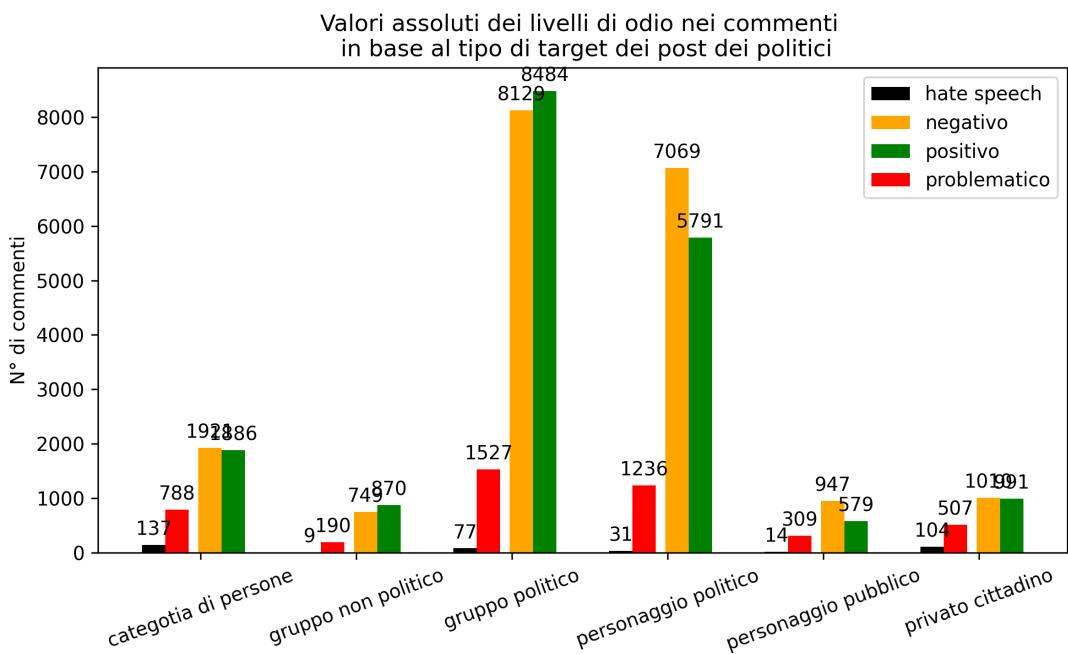
Come si può vedere nel grafico [Fig.6.18], questo tipo di post può avere uno o due target contemporaneamente. Succede spesso, ad esempio, che ci si rivolga a un partito rivale, citando anche un avversario politico specifico, oppure che si metta in relazione un personaggio pubblico a uno schieramento e ai suoi temi. I post/tweet con doppio target sono 565, i commenti che fanno seguito a questi post con doppio target sono 6430. I commenti relativi a post con singolo target risultano quindi nettamente più frequenti: “personaggio politico” e “gruppo politico” insieme valgono il 56% di tutti i commenti relativi a post negativi o comparativi (rispettivamente 23% e 42%) e valgono il 7% quando compaiono insieme raggiungendo il 63% totale, ben al di sopra di tutti gli altri tipi di target. I commenti relativi a post di attacco a “categorie di persone” rappresentano l’8% del totale dei commenti di questo *subset*, le risposte ad altri target singoli sono presenti in percentuali comprese tra il 2 e 4%. Tutti i target doppi hanno generato meno dell’1% dei commenti qui presi in considerazione.

Per ridurre il numero di singole categorie prese in considerazione e per poter vedere più chiaramente la loro relazione con il tipo di commenti generato, ogni commento con target doppio è stato sdoppiato nei due target singoli. Questo metodo introduce una distorsione, aumentando il numero di singoli commenti presi in considerazione da 37413 a 43648 elementi, considerando che 6430 vengono sostanzialmente sdoppiati, ma la distorsione non è tale da inficiare le indicazioni qualitative delle nostre elaborazioni. Inoltre, sempre per questioni di semplicità, nei grafici, non verranno presentati i risultati relativi ai commenti codificati come ambigui, dato che per queste risposte non è stato possibile, in fase di codifica, capire il significato, risultano meno importanti per le considerazioni che andremo a fare. Queste semplificazioni non verranno, però, conservate nel corso delle analisi statistiche necessarie a confermare la significatività dei risultati finali relativi a questa ipotesi.

Dai valori assoluti [Fig.6.19] possiamo vedere come la maggior parte dei commenti di *hate speech* siano stati riscontrati in risposta a target riferiti a “categorie di persone” e “privati cittadini”. Benché “Gruppi politici” e “personaggio politico” facciano rilevare molti commenti positivi e negativi senza uso di linguaggio incivile, quando si fa riferimento a “personaggi politici” il numero di commenti negativi è superiore a quello dei commenti positivi. Stesso discorso vale per i “personaggi pubblici” (es. Roberto Saviano) che generano pochi commenti positivi rispetto ai negativi.



**Figura 6.18:** Percentuali dei tipi di campagna riscontrati nel database dei commenti



**Figura 6.19:** Valori assoluti dei tipi di commenti in base al target dei post comparativi o negativi a cui rispondono

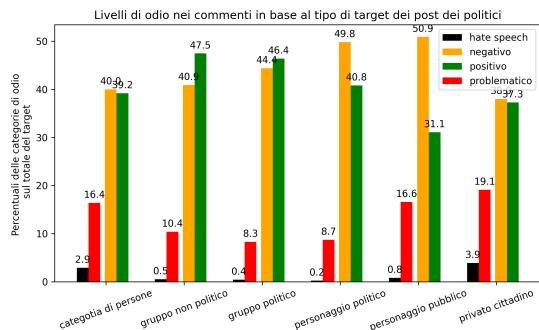
Andando ad analizzare la rilevanza dei vari tipi di commento in percentuale sul totale delle risposte ai target [Fig.6.20], possiamo osservare ancora più chiaramente come i due target non politici e non pubblici (“categoria di persone” e “privato cittadino”) risultino avere rispettivamente circa 6 e 9 volte più commenti di *hate speech* rispetto agli altri target, che hanno invece percentuali simili sotto l’1%. Anche i commenti problematici sono più frequenti in proporzione ai commenti che rispondono a post con questi target, registrando valori tra il 50% e il 100% più alti degli altri. Solamente il citare un personaggio pubblico genera livelli di commenti problematici simili (16,6%). Sempre quest’ultima categoria di target è quella con la maggiore differenza (20%) tra commenti negativi e positivi, con la minore percentuale sul totale di commenti positivi tra tutti i gruppi. Anche quando si fa riferimento a un personaggio politico i commenti sono più negativi che positivi, al di là di problematici ed *hate speech*, ma in questo caso la differenza tra positivi e negativi è solo il 10%.

Dall’analisi, infine, delle percentuali dei vari tipi di commenti sul totale dei contenuti analizzati, e sempre in relazione al target al quale rispondono, emergono indicazioni su come poter ridurre i livelli di odio sul web [Fig.6.21]. Evitando di citare “categorie di persone” e “privati cittadini” nei loro attacchi online (rispettivamente 443 e 239 post, il 6% del totale di tutto quello che è stato postato) i candidati potrebbero ridurre del 55% per cento il totale dei

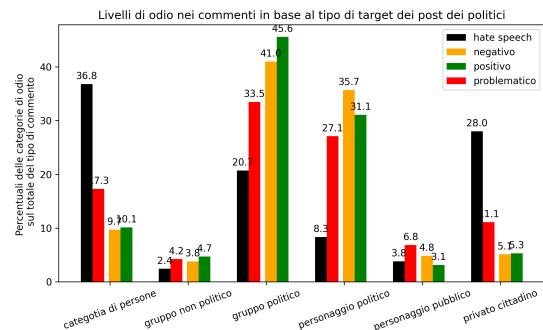
commenti di hate speech registrati (241 commenti di odio verso minoranze su 436 totali). 682 post di politici valgono 241 commenti di *hate speech*? Pensiamo di no.

In sintesi: quando i politici attaccano “categorie di persone” e “privati cittadini”, si registrano i maggiori livelli di *hate speech*, anche perché la stessa definizione di questo specifico tipo di linguaggio comprende attacchi a minoranze specifiche. Stesso discorso però anche per i commenti incivili in generale (problematici) il che indica che il riferimento polemico a gruppi e persone che vivono al di fuori dall’arena del dibattito pubblico, autorizza comportamenti più aggressivi e più incivili, anche quando non degenerano nell’*hate speech* vero e proprio. E’ come se i commentatori percepissero un maggior livello di impunità quando l’obiettivo polemico è un personaggio sconosciuto ai più. Oppure si potrebbe ipotizzare che questo tipo di attacchi in qualche modo esca già dai limiti della campagna elettorale, entrando in quello dell’attacco fine a se stesso. Si potrebbe anche discutere del fatto che attacchi a categorie come i Rom o migranti, oppure a privati cittadini in seguito ad un arresto o alla partecipazione ad una manifestazione, siano fatti che non dovrebbero essere usati all’interno di messaggi pubblici in campagne elettorali di livello nazionale o addirittura sovranazionale come in questo caso perchè non di interesse pubblico, essendo casi specifici e che non dovrebbero spingere a generalizzazioni di nessun tipo. Inoltre, attaccare singoli personaggi (politici o in generale pubblici) rispetto a gruppi (politici o pubblici) genera più commenti negativi che positivi. Anche in questo caso, quindi, si potrebbe dire che i livelli di aggressività siano tanto maggiori quanto maggiore è la vulnerabilità del target attaccato. Privati cittadini e categorie di persone indicate in quanto tali hanno infatti pochissime possibilità di rispondere mediaticamente agli insulti ricevuti, dato che non dispongono di canali social con seguito paragonabile a quello dei politici. Si riscontra inoltre, come il fenomeno dell’odio risulti particolarmente accentuato anche nel caso dei personaggi pubblici, che fanno registrare altri livelli di commenti “problematici”. Anche se è difficile che i commenti in questo caso sfocino in vero e proprio *hate speech* (a causa della definizione che lo delimita), potrebbe risultare dannoso allo svolgimento di una discussione civile indicare persone al di fuori del contesto politico come simboli di un’idea avversa.

Sembra chiaro quindi che uscire dal contesto pubblico e politico, e coinvolgere nella polemica, gruppi e cittadini sconosciuti ai più, sia il fattore in grado di generare più odio. Abbiamo riscontrato però anche una prevalenza di messaggi problematici e negativi generati da target singolari rispetto ai corrispettivi plurali. Per vedere se le due ipotesi sono supportate da significatività statistica è stato eseguito un test di chi-quadro. In questo caso sono stati presi in considerazione tutti i tipi di commenti (anche quelli ambigui) e i target sono stati raggruppa-



**Figura 6.20:** Percentuali dei tipi di commenti sul totale del tipo di target dei post comparativi o negativi a cui rispondono



**Figura 6.21:** Percentuali dei tipi di commenti sul totale del tipo di commento in base al target dei post comparativi o negativi a cui rispondono

	pubnonpub	privato	pubblico_politico	All	singplur	plurale	singolare	All
pubnonpub								
c_rating								
ambiguo	1.72		0.48	0.69	ambiguo		0.62	0.65
hate	3.16		0.32	0.81	hate		0.79	0.69
negativo	39.01		46.25	45.01	negativo		43.09	45.25
positivo	39.80		44.38	43.59	positivo		45.86	43.53
problematico	16.31		8.58	9.90	problematico		9.65	9.89
All	100.00		100.00	100.00	All		100.00	100.00

**Figura 6.22:** Percentuali dei tipi di commenti divisi tra quelli che rispondono a un post in cui il target è privato (privato cittadino, categoria di persone) oppure pubblico (personaggio pubblico e politico, gruppo non politico o politico)

**Figura 6.23:** Percentuali dei tipi di commenti divisi tra quelli che rispondono a un post in cui il target è un singolo (privato cittadino, personaggio pubblico o politico) oppure gruppo (categoria di persone, gruppo non politico o politico)

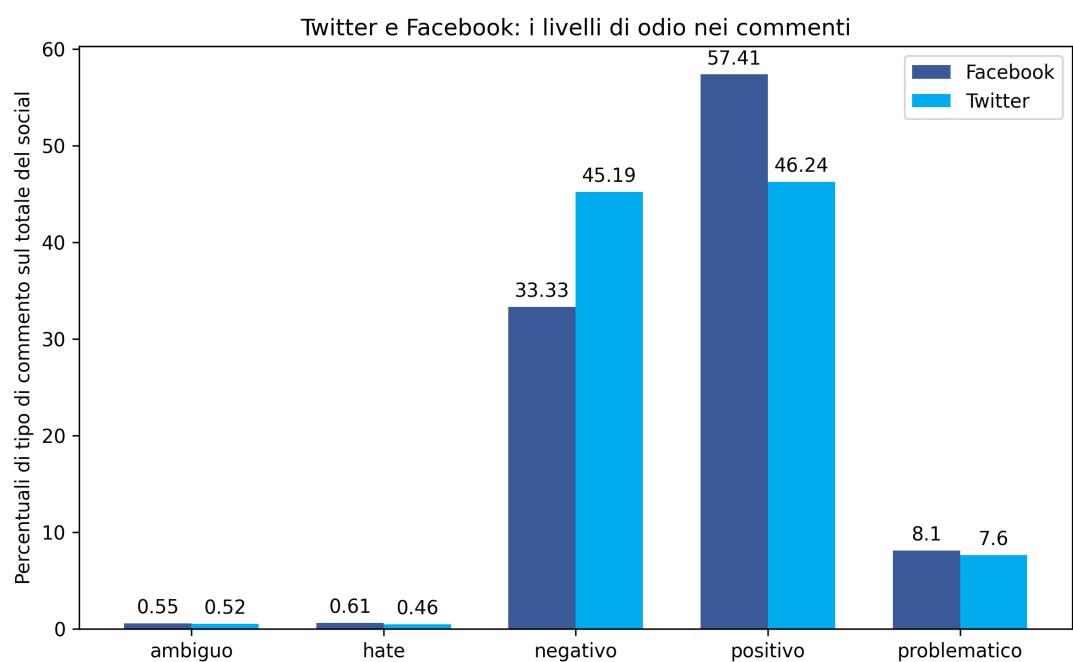
ti in entrambi i casi in due categorie, riducendo il numero di target totali. Sono stati presi in considerazione anche i target doppi in quanto tali, ma sono stati esclusi dai test quei target doppi che non rientravano in nessuna delle due categorie create per l'analisi.

In primo luogo, viene testato se vi è una differenza tra i commenti relativi a post comparativi e negativi con target riferito a personaggi o gruppi di dominio pubblico (“personaggio pubblico” e “personaggio politico”, “gruppo politico e gruppo non politico”, più i target doppi che comprendono due di queste categorie) confrontati con quelli che non lo sono (“privato cittadino”, “categoria di persone e privato cittadino” + “categoria di persone”). Gli altri target doppi sono stati eliminati (es. privato cittadino - gruppo politico)[Fig.6.22]. Il test ha dato risultati molto significativi con un Pearson chi-square molto alto:  $\chi^2 (9, N = 31240) = 879.1844$ ,  $p = 0.0000$ . Possiamo quindi dire che quando si fa riferimento a target non di dominio pubblico si crea un significativo aumento di commenti di *hate speech* e problematici nei commenti.

Allo stesso modo è stata testata la seconda ipotesi. In questo caso i due gruppi consistono nei target singolari da una parte (“privato cittadino”, “personaggio pubblico e personaggio politico”, più relativi target doppi) e target plurali dall’altra (“categoria di persone”, “gruppo politico” o “gruppo non politico”, più relativi target doppi). Anche in questo caso i target doppi che non rientrano nelle due categorie sono stati esclusi (es. privato cittadino + gruppo politico) [Fig.6.23]. I risultati sono significativi anche in questo caso, ma con un effetto minore rispetto al test precedente ( $9, N = 31083$ ) =  $112.3151$ ,  $p = 0.0000$ . Possiamo quindi dire che attaccando target singolari rispetto a gruppi si generano commenti più negativi rispetto a quelli positivi, ma non una percentuale maggiore di linguaggi incivili ed *hate speech*. In generale la prima distinzione testata risulta più esplicativa di quest’ultima.

#### 6.2.3 IP.3: I COMMENTI SU TWITTER SONO PIÙ NEGATIVI?

Abbiamo visto come il tipo di canale utilizzato fa cambiare il tipo di comunicazione politica adottata, aumentando la negatività dei messaggi su Twitter rispetto a quelli pubblicati su Facebook. Effetti simili saranno riscontrati anche nella negatività dei commenti? Analizzando il grafico con le percentuali dei livelli di odio nelle risposte a post e tweet [Fig.6.24], vediamo come in questo caso i risultati siano meno netti e l’ipotesi non risulti del tutto confermata. Se da una parte Facebook registra molti più commenti positivi rispetto a quelli con tono negativo, su Twitter le due percentuali sono praticamente uguali. Sembra quindi che il social di microblogging favorisca commenti più negativi rispetto alla piattaforma di Zuckemberg. Ma quando guardiamo al linguaggio incivile e all’*hate speech*, le proporzioni si invertono. Benché la distribuzione risulti significativamente diversa con un test di  $\chi^2$  ( $7, N = 78175$ ) =  $1075.1347$ ,  $p = 0.0000$ , questa differenza non è chiaramente spiegabile a partire da questo campione e servirebbero altri tipi di dati per poter fare considerazioni più solide. Possiamo ipotizzare che gli odiatori siano più propensi ad utilizzare Facebook rispetto a Twitter, mentre gli utenti generici potrebbero essere spinti dal diverso canale comunicativo ad aumentare i loro livelli di negatività, ma le differenze nei tipi di commento che potremmo far risalire agli odiatori sono troppo piccole per poterlo affermare con certezza, e soprattutto non disponiamo di nessuno dato relativo agli/alle utenti dei commenti presi in considerazione, non possiamo quindi paragonare i contenuti nelle due piattaforme per l’\* stess\* utente (come invece abbiamo fatto con i post e tweet dei politici).



**Figura 6.24:** Le percentuali dei vari tipi di commenti sul totale dei due social in considerazione



# 7

## Conclusioni

Il sempre crescente utilizzo della comunicazione digitale in tutto il mondo, e in particolare in ambito politico, ha spinto molt\* ricercat\* a interrogarsi sugli effetti di questo nuovo tipo di comunicazione sulla polarizzazione del dibattito pubblico e sui suoi livelli di negatività. Contemporaneamente, negli ultimi anni ha assunto crescente rilevanza politica e sociale anche il fenomeno dell'odio online, per gli effetti dannosi su chi ne è vittima e per i crimini d'odio che ne sono scaturiti nella vita offline. Poche ricerche hanno, però, messo in relazione diretta l'*hate speech* con le campagne negative. L'ipotesi alla base della nostra ricerca, era, invece, proprio verificare l'esistenza di una correlazione tra il tipo di campagna proposta dai politici nei loro post pubblicati sui social network, e i livelli di odio registrati nei commenti a quegli stessi post. Abbiamo supposto, infatti, che il ricorso agli insulti e ad un linguaggio aggressivo e scurrile da parte degli/delle utenti che interagiscono con i contenuti di una campagna politica online, dipenda anche dal tipo di comunicazione utilizzata dai politici. Per testare questa ipotesi si è utilizzato il database del Barometro dell'Odio realizzato da Amnesty International, un campione di dati senza precedenti in Italia per vastità, ricchezza di tipi di informazioni raccolte, rappresentatività della campagna analizzata e mole di codifica qualitative. Le nostre elaborazioni hanno rilevato una correlazione evidente tra il tipo di messaggio diffuso dai politici e le conseguenti risposte degli/delle utenti sui social media. L'evidenza empirica raccolta sembra, quindi, confermare l'ipotesi principale che abbiamo sottoposto a verifica con la nostra ricerca. Questo risultato è interessante perché dimostra come la responsabilità della presenza di contenuti di odio online non possa essere attribuita solo ai singoli

individui. Anche se si possono riscontrare alcune caratteristiche che accomunano tutti gli odiatori, quando si analizza l'odio all'interno del dibattito politico è necessario considerare il più ampio contesto fornito dalla comunicazione prodotta dai candidati alle elezioni e il linguaggio utilizzato. Nelle nostre rilevazioni, il 67% dell'odio è stato generato da post che attaccano direttamente gli avversari politici. Da questo punto di vista sembrerebbe un fenomeno difficilmente arginabile: le campagne negative (e comparative) hanno sempre fatto parte del dibattito democratico, sarebbe difficile chiedere di utilizzare solo messaggi legati alle proprie posizioni, e forse sarebbe anche sbagliato. Dall'analisi dei target delle campagne di attacco è emerso, però, che il 63% dell'odio generato da attacchi ad avversari è associato ad una piccola percentuale (solo il 7%) di tutti i post pubblicati, ed in particolare a quelli che contengono attacchi a privati cittadini e a categorie di persone che non hanno nessuna rilevanza né in ambito politico né in quello più generalmente pubblico. Questo risultato merita a nostro parere una riflessione. È veramente necessario attaccare privati cittadini per guadagnare consenso elettorale? È appropriato che i politici, il cui ruolo è rappresentare tutt\* i/le cittadin\*, attaccino direttamente alcune categorie di persone del tutto estranee al dibattito politico? I dati emersi da questa ricerca, suggeriscono che evitare questo tipo di messaggi diminuirebbe in modo molto significativo i commenti di odio pubblicati sui social network durante le campagne elettorali e favorirebbe una convivenza civile e democratica se è vero (come emerge dalla letteratura passata in rassegna nel capitolo "*Hate Speech*") che c'è una relazione tra i discorsi d'odio presenti sul web ed i crimini d'odio agiti nella vita reale. Pensiamo, quindi che si possa trarre dalle nostre elaborazioni qualche indicazione utile per mettere a punto una strategia efficace di contrasto all'*hate speech*.

Gli altri risultati emersi dalla nostra ricerca sono allineati, invece, ad indicazioni che sono già consolidate nella letteratura. In particolare trova conferma la tesi che le campagne negative siano utilizzate più dai partiti all'opposizione che da quelli al governo e trova conferma anche la tesi, ampiamente discussa dalla letteratura, che i messaggi negativi abbiano una maggiore viralità ed una maggiore diffusione rispetto a quelli positivi. Nelle nostre elaborazioni questo è vero per entrambi i social network analizzati, cioè sia per Twitter che per Facebook. Questo effetto è dovuto al diverso grado di attenzione che viene data ai singoli post/tweet da parte degli utenti (e che nella letteratura è ricondotto principalmente agli effetti del *negativity bias*), ma è anche dovuto agli algoritmi che gestiscono i flussi d'informazione all'interno dei social. Al momento questi algoritmi sono pensati con l'unico scopo di aumentare il tempo che gli/le utenti trascorrono sulla piattaforma, sulla base di un *business model* volto al profitto e alla vendita di pubblicità, non alla vendita di notizie informative, come per esempio

succedeva nei media tradizionali. Poiché ormai sia Facebook che Twitter svolgono un ruolo fondamentale nell'*agenda setting* anche del nostro paese, bisognerebbe interrogarsi maggiormente su come rendere queste "black box" dei strumenti al servizio della collettività, e non solo "armi di distrazione di massa" (Rosen, 2012).

Dalla nostra ricerca risulta difficile capire, invece, se, in generale, il *negative campaign* sia aumentato, rispetto ad altre elezioni in Italia e nel mondo. Da questo punto di vista è arduo inserire i risultati delle nostre elaborazioni all'interno della letteratura rilevante. Le comparazioni internazionali risultano complesse per la diversità delle metodologie utilizzate e per la specificità dei sistemi elettorali e politici nazionali. Possiamo solo rilevare che i livelli non altissimi di campagna negativa individuati in questo studio (16,4%) riflettano valori simili a quelli registrati in altri paesi con sistemi elettorali che prevedono la competizione tra molti partiti, ed inferiori a quelli di nazioni in cui è presente un bipolarismo perfetto come gli Stati Uniti.

Infine una riflessione sulle piattaforme in generale, e in particolare su quelle prese in considerazione qui: Facebook e Twitter. Abbiamo visto come ci siano delle significative differenze nel tipo di campagna utilizzata sui diversi social network; i politici sembrano capire, e quindi anche sfruttare efficacemente, i due diversi stili comunicativi. Ai messaggi più corti di Twitter corrisponde un utilizzo maggiore di campagne negative, nei lunghi post su Facebook vengono prediletti messaggi positivi che spiegano il proprio punto di vista o che lo comparano in dettaglio con quello altrui. Anche chi commenta su Twitter sembra essere portato ad utilizzare un linguaggio più negativo, mentre su Facebook l'*hate speech* sembra più presente in percentuale, ma servirebbero più dati sui profili degli utenti per poter affermare con la necessaria certezza quest'ultima affermazione.

## 7.1 LIMITI

Benché il database creato da Amnesty International sia uno dei punti di forza di questo studio, presenta un'importante limitazione: non sono presenti i dati relativi ai profili dei commentatori. Questo è dovuto a ragionevoli motivazioni di privacy e agli scopi originari del Barometro dell'Odio. Non è stato quindi possibile capire la relazione tra i produttori di *hate speech*; per assurdo, potrebbero essere stati postati tutti dalla stessa persona.

Un'altra limitazione del campione selezionato è l'utilizzo delle API ufficiali dei social networks. È possibile che molti dei commenti di odio siano stati cancellati dalle stesse piattaforme in modo automatizzato, dai politici che hanno pubblicato i post/tweet o su segnalazione

di altr\* utenti. Tecniche di *scraping* automatizzato costante avrebbero permesso di togliere ogni dubbio sulla effettiva restituzione di tutto ciò che è stato postato.

Inoltre sono stati considerati solo gli account ufficiali dei politici, ma non le pagine dei vari partiti e nessun gruppo Facebook di riferimento per i vari elettorati o gli articoli di giornale relativi alle varie dichiarazioni dei politici, che ormai sono sempre condivise da testate giornalistiche e televisioni anche sui social. Sicuramente in ambienti diversi dalle pagine ufficiali i livelli di odio e i tipi di strategia comunicativa utilizzata da chi crea concretamente le campagne elettorali possono cambiare anche notevolmente da quelli riscontrati nelle nostre rilevazioni.

Inoltre, per quanto riguarda la codifica dei contenuti, è necessario segnalare che la tassonomia da noi utilizzata non è presente in nessuno studio di nostra conoscenza, ma risulta una mediazione tra diverse tassonomie presenti nella letteratura. La definizione di target o "categoria di persone" potrebbe risultare troppo direttamente legata alla definizione di *hate speech*, anche se questo non cambia le conclusioni che si possono trarre dalle analisi: risulta discutibile che questo tipo di attacchi siano identificabili con la tradizionale idea di campagna negativa, si potrebbero anche definire forme *soft* di *hate speech*, ma rimane il fatto che questi contenuti abbiano generato più contenuti d'odio rispetto ad altri tipi di target. Anche la categoria "privato cittadino" risulta poco frequente negli studi sul *negative campaign*. D'altra parte, però, post che attaccano ed espongono mediaticamente una singola persona non nota con intento denigratorio o di accusa (come, ad esempio, l'immigrato X denunciato per l'Y reato), non possono certo essere considerati messaggi di campagna politica positiva o neutro per la loro intrinseca valenza esemplificativa di discorsi più ampi, anche se a volte sotterri.

La valutazione inoltre, avrebbe potuto essere svolta da tre giudici indipendenti, come fatto nello studio di Amnesty International, invece che solo da due. Questo avrebbe aumentato le percentuali di certezza della codifica.

Le analisi infine, non presentano una propria modellizzazione complessiva, in grado di tenere insieme i vari fattori analizzati, come si sarebbe potuto fare realizzando un'analisi bayesiana, da affiancare ai testi di chi-quadro già eseguiti.

Per quanto non sia possibile determinare un principio di causalità diretto tra il tipo di campagna e il tipo di commento di odio che ne scaturisce, poiché la nostra ricerca non si avvale un modello sperimentale, sicuramente i post dei politici precedono temporalmente le risposte degli utenti, influenzando la loro reazione.

Le considerazioni sulla comparazione delle percentuali sul *negative campaign* in relazione ad altri studi risultano complesse e non del tutto compiute. Questi valori variano da ele-

zione ad elezione e da una nazione all'altra. L'utilizzo dei social in particolare aumenta col tempo e le fasce di popolazione che lo utilizzano si modificano rapidamente. Per poter fare considerazioni appropriate servirebbe uno studio longitudinale che prenda in considerazione anche media tradizionali e altre forme di campagna come i cartelloni elettorali riuscendo ad analizzare tutti i messaggi propagandati.

## 7.2 FUTURI SVILUPPI

Sarebbe interessante verificare la correlazione del *sentiment* tra post dei politici e relativi commenti, cercando di individuare quali potrebbero essere le emozioni presenti nei messaggi elettorali che danno maggiormente luogo a discorsi d'odio. Costruendo un network semantico dell'*hate speech* presente nel campione utilizzato, sarebbe inoltre possibile capire quali sono i temi specifici della campagna elettorale per le europee del 2019 che hanno maggiormente generato fenomeni d'odio, e quale sia la relazione tra le parole usate in questo sottoinsieme rispetto al network delle parole raccolte nel complesso.

All'interno del database già collezionato sono presenti moltissime altre informazioni che non sono state utilizzate in questo studio. Non si è indagata, ad esempio, la relazione tra il genere dei politici e i messaggi d'odio, anche se è possibile ipotizzare che questo fattore possa avere un ruolo nelle risposte degli/delle utenti. Non sono stati presi in considerazione i risultati elettorali raggiunti, che invece potrebbe essere usati come indicatore di riuscita delle varie strategie. Si potrebbe studiare anche la geografia del voto: non sono presenti informazioni riguardo all'ubicazione dei/delle commentator\*, ma tutt\* i/le candidat\* si sono presentat\* in una specifica circoscrizione, che potrebbe essere una variabile significativa. La divisione dei/delle candidat\* in macro regioni come Nord-Sud-Centro potrebbe mediare gli effetti elettorali del *negative campaign*.

Amnesty International, inoltre, ha codificato tutti i contenuti in base al “topic” affrontato. Capire quali sono gli argomenti che vengono maggiormente utilizzati nelle campagne negative e comparative sarebbe certamente d'aiuto per comprendere meglio questi fenomeni. Potrebbe darsi che alcuni temi utilizzati dai politici fungano da moderatori insieme al tipo di campagna nella produzione di odio.

Per testare efficacemente le conseguenze dei messaggi politici, sarebbe, infine, interessante svolgere delle analisi sperimentali, che prevedano la somministrazione dei vari tipi campagna politica in laboratorio, riuscendo a controllare molto meglio tutte le variabili coinvolte nel

processo di attivazione dei soggetti. Si potrebbe testare se, anche implicitamente, alcuni tipi di messaggi inducano i soggetti a prediligere risposte d'odio oppure no.

# Bibliografia

- Agosti, C., Beraldo, D., Calzati, S., Charpentier, Z., Choi, S., Del Fabbro, L., ... Vos, J. (2019). *Investigating youtube personalization with ytrex, digital methods summer school 2019.* <https://wiki.digitalmethods.net/Dmi/SummerSchool2019AlgorithmsExposed>.
- Amnesty International Italia. (2019a). *Barometro dell'odioelezioni europee 2019, nota metodologica.* <https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2019/06/28223758/nota-metodologica-barometro-amnesty.pdf>.
- Amnesty International Italia. (2019b). *Barometro dell'odioelezioni europee 2019, rapporto finale.* <https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2019/05/29202706/Amnesty-barometro-odio-2019.pdf>.
- Amnesty International Italia. (2019c). *Barometro dell'odioelezioni europee 2019, report divulgativo.* <https://www.amnesty.it/cosa-facciamo/elezioni-europee/>.
- Anscombe, S., & Iyengar, S. (1995). *Going negative.* (Vol. 95). New York: Free Press.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, 5(4), 323-370. doi: 10.1037/1089-2680.5.4.323
- Bene, M. (2017). Go viral on the facebook! interactions between candidates and followers on facebook during the hungarian general election campaign of 2014. *Information, Communication, Society*, 20(4), 513-529. doi: 10.1080/1369118X.2016.1198411
- Benesch, S. (2012). Dangerous speech: A proposal to prevent group violence. *Voices That Poison: Dangerous Speech Project.* <https://dangerouspeech.org/wp-content/uploads/2018/01/Dangerous-Speech-Guidelines-2013.pdf>.
- Berger, J., & Milkman, K. (2010). Social transmission, emotion, and the virality of online content. *Wharton research paper*, 106, 1-52.

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), 297-326. doi: 10.1177/1468796817709846

Bruns, A., Weller, K., Borra, E., & Rieder, B. (2014). Programmed method: developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*. doi: 2050-3806

Campbell, K. (2019). *The freedom of hate speech; a call for civil dialogue*. [https://www.youtube.com/watch?v=xD6\\_Ma5vutw](https://www.youtube.com/watch?v=xD6_Ma5vutw).

Campbell, K. G. (2004). Freedom of speech, imagination, and political dissent: Culturally centering the free speech principle. *Doctoral dissertation, University of Denver*. <https://search.proquest.com/openview/eede71d1421b4ae1f11265ce0fdfe66c/1?cbl=18750&diss=y&pq-origsite=gscholar>.

Carraro, L., Gawronski, B., & Castelli, L. (2010). Losing on all fronts: The effects of negative versus positive person-based campaigns on implicit and explicit evaluations of political candidates. *British Journal of Social Psychology*, 49(3), 453-470. doi: 10.1348/014466609X468042

Ceron, A., & Adda, G. (2016). E-campaigning on twitter: The effectiveness of distributive promises and negative campaign in the 2013 italian election. *New media and society*, 18(9), 1935-1955. doi: doi:10.1177/1461444815571915

Chan, J., Ghose, A., & Seamans, R. (2015). The internet and racial hate crime: Offline spillovers from online access. *MIS Quarterly*, 40, 381-403. doi: doi:10.2139/ssrn.2335637

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, 19(2), 171-209. doi: 10.1007/s11036-013-0489-0

Community Security Trust (CTS), & Antisemitism Policy Trust. (2019). *Hidden hate: What google searches tell us about antisemitism today*. <https://cst.org.uk/public/data/file/a/b/APT%20Google%20Report%202019.pdf>.

Conover, M. D., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. *Icwsrm*, 133(2011), 89-96. doi: 10.1080/105846099198613

Curini, L., & Martelli, P. (2010). Ideological proximity and valence competition. negative campaigning through allegation of corruption in the italian legislative arena from 1946 to 1994. *Electoral Studies*, 29(4), 636-647. doi: 10.1016/j.electstud.2010.06.004

De Stefano, V. (2016). Introduction: crowdsourcing, the gig-economy and the law. *Comparative Labor Law Policy Journal*, 37(3).

Druckman, J. N., Kifer, M. J., & Parkin, M. (2010). Timeless strategy meets new medium: Going negative on congressional campaign web sites, 2002–2006. *Political Communication*, 27(1), 88-103. doi: 10.1080/10584600903502607

Erjavec, K., & Kovačić, M. P. (2012). “you don’t understand, this is a new war!” analysis of hate speech in news web sites’ comments. *Mass Communication and Society*, 15(6), 899-920. doi: 10.1080/15205436.2011.619679

European Commission against Racism and Intolerance. (2015). *Raccomandazione di politica generale n.15 dell’ecri relativa alla lotta contro il discorso di odio.* <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech-ital/16808b5b04>.

Facebook. (2019). *Facebook community standards.* <https://www.facebook.com/communitystandards>.

Fischer, S. (2020). *Advertising giants agree to evaluate mutual definition of hate speech.* <https://www.axios.com/advertising-hate-speech-tech-companies-facebook-5d2b4ebf-e412-4075-ad1f-77aa2b7a07b3.html>.

Freelon, D. (2018). Computational research in the post-api age. *Political Communication*, 35(4), 665-668. doi: 10.1080/10584609.2018.1477506

Fridkin, K. L., & Kenney, P. (2011). Variability in citizens’ reactions to different types of negative campaigns. *American Journal of Political Science*, 55(2), 307-325. doi: 10.1011/j.1540-5907.2010.00494.x

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech.* UNESCO. [https://www.researchgate.net/publication/284157227\\_Countering\\_Online\\_Hate\\_Speech\\_-\\_UNESCO#pf1f](https://www.researchgate.net/publication/284157227_Countering_Online_Hate_Speech_-_UNESCO#pf1f).

Geer, J. G. (2008). *In defense of negativity: Attack ads in presidential campaigns*. University of Chicago Press.

Geer, J. G. (2012). The news media and the rise of negativity in presidential campaigns. *Political Science and Politics*, 45, 422-427. doi: doi:10.1017/S1049096512000492

Global Web Index. (2020). *Global web index coronavirus research. series 4: Media consumption and sport*. [https://www.globalwebindex.com/hubfs/1.%20Coronavirus%20Research%20PDFs/GWI%20coronavirus%20findings%20April%202020%20-%20Media%20Consumption%20\(Release%204\).pdf](https://www.globalwebindex.com/hubfs/1.%20Coronavirus%20Research%20PDFs/GWI%20coronavirus%20findings%20April%202020%20-%20Media%20Consumption%20(Release%204).pdf).

Han, B. C. (2017). *Psychopolitics: Neoliberalism and new technologies of power*. Verso Books.

Hansen, K. M., & Pedersen, R. T. (2008). Negative campaigning in a multiparty system. *Scandinavian Political Studies*, 31(4), 408-427. doi: 10.1111/j.1467-9477.2008.00213.x

Hansen, L. K., Arvidsson, A., Nielsen, F. B., Colleoni, E., & Etter, M. (2011). Good friends, bad news-affect and virality in twitter. *Future information technology*, 34-43. doi: 10.1007/978-3-642-22309-9\_5

Hargreaves, E., Agosti, C., Menasché, D., Neglia, G., Reiffers-Masson, A., & Altman, E. (2019). Fairness in online social network timelines: Measurements, models and mechanism design. *Performance Evaluation*, 129, 15-39. doi: 10.1016/j.peva.2018.09.009

Haselmayer, M., Meyer, T. M., & Wagner, M. (2019). Fighting for attention: Media coverage of negative campaign messages. *Party Politics*, 25(3), 412-423. doi: 10.1177/1354068817724174

Jamieson, K. H. (1993). *Dirty politics: Deception, distraction, and democracy*. Oxford University Press on Demand.

Jamieson, K. H., & Waldman, P. (1997). Mapping campaign discourse: An introduction. *American Behavioral Scientist*, 40(8), 1133-1138. doi: 10.1177/0002764297040008012

Jamieson, K. H., Waldman, P., & Devitt, J. (1998). Mapping the discourse of the 1996 us presidential general election. *Media, Culture, Society*, 20(2), 323-328. doi: 10.1177/016344398020002009

Kamber, V. (1997). *Poison politics: Are negative campaigns destroying democracy?* Basic Books.

Kang, C., & Goldman, A. (2016). In *washington pizzeria attack, fake news brought real guns*. <https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html>.

Kemp, S. (2020a). *Digital 2020: 3.8 billion people use social media*. <https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media>.

Kemp, S. (2020b). *Digital 2020: July global statshot*. <https://datareportal.com/reports/digital-2020-july-global-statshot>.

Kim, N., & Wojcieszak, M. (2018). Intergroup contact through online comments: Effects of direct and extended contact on outgroup attitudes. *Computers in Human Behavior*, 81, 63-72. doi: 10.1016/j.chb.2017.11.013

Lau, R. R., & Pomper, G. M. (2004). *Negative campaigning: An analysis of us senate elections*. Rowman and Littlefield.

Lau, R. R., Sigelman, L., & Rovner, I. B. (2007). The effects of negative political campaigns: a meta-analytic reassessment. *Journal of Politics*, 69(4), 1176-1209. doi: 10.1215/j.1468-2508.2007.00618.x

Mancosu, M., & Vegetti, F. (2020). What you can scrape and what is right to scrape: A proposal for a tool to collect public facebook data. *Social Media Society*, 6(3). doi: 10.1177/2056305120940703

Massaro., T. M. (1990). Equality and freedom of expression: The hate speechdilemma. *William and Mary Law review*, 32(2), 211-265. doi: 10.1177/1468796817709846

Mazzoleni, G., & Schulz, W. (1999). 'mediatization' of politics: A challenge for democracy? *Political communication*, 16(3), 247-261. doi: 10.1080/105846099198613

Meijers, M. J. (2017). Contagious euroscepticism: The impact of eurosceptic support on mainstream party positions on european integration. *Party Politics*, 23(4), 413-423. doi: 10.1177/1354068815601787

Mill, J. S. (1859). *On liberty*.

Müller, K., & Schwarz, C. (2019). Fanning the flames of hate: Social media and hate crime. *SSRN*. doi: 10.2139/ssrn.3082972

Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2020). Reuters institute digital news report 2020. *Reuters Institute for the Study of Journalism*.

Nicas, J., Marshall, J., & Vranica, S. (2017). *Verizon pull ads from google over hate videos*. <https://www.wsj.com/articles/at-t-pulling-ads-from-google-over-hate-videos-1490206824>.

Osborn, M. (2008). Fuelling the flames: Rumour and politics in kibera. *Journal of Eastern African Studies*, 2(2), 315-327. doi: 10.1080/17531050802094836

Papp, Z., & Patkós, V. (2018). The macro-level driving factors of negative campaigning in europe. *The International Journal of Press/Politics*, 194016121880342, 42-60. doi: 10.1177/19401612188034264

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Paul, K., & Hern, A. (2020). *Verizon pulls ads from facebook over inaction on hate speech*. <https://www.theguardian.com/technology/2020/jun/25/verizon-advertising-facebook-hate-speech-boycott>.

Perloff, R. M., & Kinsey, D. (1992). Political advertising as seen by consultants and journalists. *Journal of Advertising Research*, 32(3), 53-60.

Popper, K. R. (1945). *The open society and its enemies*.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, 61(3), 380-391. doi: 10.1037/0022-3514.61.3.380

Rathi, R. (2019). *Effect of cambridge analytica's facebook ads on the 2016 us presidential election*. <https://towardsdatascience.com/effect-of-cambridge-analyticas-facebook-ads-on-the-2016-us/-presidential-election-dacb5462155d>.

- Reif, K., & Schmitt, H. (1980). Nine second-order national elections—a conceptual framework for the analysis of european election results. *European journal of political research*, 8(1), 3-44. doi: 10.1111/j.1475-6765.1980.tb00737.x
- Reviglio, U., & Agosti, C. (2020). Thinking outside the black-box: The case for “algorithmic sovereignty” in social media. *Social Media+ Society*, 6(2). doi: 10.1177/2056305120915613
- Ribeiro, M. H., Calais, P. H., Santos, Y. A., & Almeida, W., V. A. and Meira Jr. (2018). Characterizing and detecting hateful users on twitter. *arXiv*. 1803.08977.
- Rosen, L. D. (2012). *Weapons of mass distraction: Why we have lost the ability to focus*. <https://www.psychologytoday.com/us/blog/rewired-the-psychology-technology/201212/weapons-mass-distraction>.
- Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5 (4), 296–320. doi: 10.1207/S15327957PSPR0504\_2
- Seddone, A., Bobba, G., & Roncarolo, F. (2019). European versus domestic politics: Media negativity during the 2019 european elections campaign in italy. *Italian Political Science*, 14(2), 1-19. doi: doi:10.1177/1461444815571915
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of management information systems*, 29(4), 217-248. doi: 10.2753/MIS0742-1222290408
- Stop Hate For Profits. (2020). *Statement from stop hate for profit on ad pause success and stophateforprofit campaign*. [stophateforprofit.org](http://stophateforprofit.org).
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology and behavior*, 7(3), 321-326. doi: 10.1089/1094931041291295
- Surlin, S. H., & Gordon, T. F. (1977). How values affect attitudes toward direct reference political advertising. *Journalism Quarterly*, 54(1), 89-98. doi: 10.1177/107769907705400113
- Swint, K. C. (1998). *Political consultants and negative campaigning: The secrets of the pros*. University Press of America. doi: ISBN:0761810080,9780761810087

Tacopino, J. (2020). *Twitter places notice on trump's minneapolis tweet for 'glorifying violence'*. <https://nypost.com/2020/05/29/twitter-places-public-interest-notice-on-trumps-minneapolis/-tweet/>.

Tan, Z. M., Aggarwal, N., Cowls, J., Morley, J., Taddeo, M., & Floridi, L. (2020). The ethical debate about the gig economy: a review and critical analysis. *Available at SSRN*. doi: SSRN-id3669216.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with twitter: How 140 characters reflect the political landscape. *Social science computer review*, 29(4), 402-418. doi: 10.1177/0894439310386557

Twitter. (2019). *The twitter rules*. <https://help.twitter.com/en/rules-and-policies/twitter-rules>.

UDHR. (1966). *International covenant on civil and political rights*. <https://www.ohchr.org/EN/ProfessionalInterest/Pages/CCPR.aspx>.

Walter, A. S. (2014). Negative campaigning in western europe: Similar or different? *Political Studies*, 62, 42-60. doi: 10.1111/1467-9248.12084

Walter, A. S., & Vliegenthart, R. (2010). Negative campaigning across different communication channels: Different ball games? *The International Journal of Press/Politics*, 15(4), 441-461. doi: 10.1177/1940161210374122

Wesleyan Media Project. (2018). "61% increase in volume of negative ads". <http://mediaproject.wesleyan.edu/103018/>.

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2019). Fairness in online social network timelines: Measurements, models and mechanism design. *The British Journal of Criminology*, 60 (1), 93–117. doi: 10.1093/bjc/azz049

World Federation Advertisers. (2020). *Global alliance for responsible media launches to address digital safety*. <https://wfanet.org/knowledge/item/2019/06/18/Global-Alliance-for-Responsible-Media-launches-to-address-digital-safety>.

Yacubov, H., & Rossini, P. (2020). Trends in incivility and intolerance: An analysis of political facebook comments during the 2016 election. *iConference 2020 Proceedings*.

# Ringraziamenti

Il primo grazie va ai miei genitori, per avermi sempre sostenuto oltre ogni ragionevole dubbio sulla mia voglia di studiare. A mia madre: mia editor, consulente, manager e tutti gli altri ruoli che è stata costretta ricoprire per aiutarmi, grazie per ogni sacrificio fatto per me. A mio padre: per il valore che attribuisce alla formazione e per averlo dimostrato investendo nella mia, per avermi riportato di continuo alla realtà dal mondo delle folli passioni in cui spesso mi ritrovo.

Grazie ai professor\* dell'Università di Padova, soprattutto a quell\* che sono stati anche maestr\*: "L'insegnante mediocre racconta. Il bravo insegnante spiega. L'insegnante eccellente dimostra. Il maestro ispira." diceva Socrate. Spero, un giorno, di essere un buon maestro anche io come voi.

Alle mie relatrici di tesi, per la fiducia che avete riposto in me e per la serietà (e pazienza!) con le quali avete seguito il mio progetto. Entrare nella Rete e poter scrivere che anche la mia università ne fa parte è per me un grandissimo onore. Nonostante sia stato difficile arrivare a questo risultato, senza di voi non ce l'avrei mai fatta.

Ai responsabili del Polo, per aver accettato di rendermi parte della discussione e della costruzione degli spazi che accolgono il nostro dipartimento. L'università non è solo le sue lezioni o le sue mura, il buon studente non prende solo appunti e non fa solo esami. "Il Maestro apre la porta, ma tu devi entrare da solo": avremmo aperto quella porta anche da soli, ma discutere con voi per aprirla ed entrarci insieme è stato (inaspettatamente) più divertente.

A Stefania Milan, compagna e professoressa allo stesso tempo come nessun\* altr\*. Grazie per essere il miglior esempio a cui ispirarmi. Criptare quella mail per il tirocinio in PGP e averti incontrato quelle due volte al bar calabrese ha spinto in avanti la mia carriera più di anni di studio. Senza di te, non avrei imparato quel poco che so sul mondo dei computer.

Grazie al team di Tracking Exposed. Grazie a voi, soprattutto grazie a Vecna, posso dire oggi di essere un Ricercatore. Grazie per la fiducia e grazie per avermi considerato dal primo momento come un pari, come uno di voi. Il mondo ha un gran bisogno di rivoluzionari smanettoni come noi, adelante compagneros!

A Marcello Racchini, unico "capitano, mio capitano". Mi scuso per tutte le bravate che ho fatto al liceo, anche se è tardi per scusarsi. Ringrazio per la passione per la scrittura che mi

hai trasmesso, anche se non ne sono degno. Spero che un giorno ci rincontreremo e potremo rincominciare da capo.

A Paola: mia musa, mio cuore, brindiamo a noi! Grazie per avermi sopportato durante questo periodo stressante. Mi hai insegnato cosa vuol dire amare, cosa vuol dire prendersi cura delle persone a cui si vuole bene. Spero che questa laurea possa servirci per costruire un meraviglioso futuro insieme.