

Università degli Studi di Torino
Dipartimento di Culture, Politica e Società
Centro "Luigi Bobbio" per la ricerca sociale pubblica e applicata

FUEL MY FIRE

**Social media, politica
e hate speech in Italia**

**BLUE
PAPERS
CLB-CPS**

A cura di
**Federico Vegetti
Arturo Bertero
Giuliano Bobba
Moreno Mancosu
Antonella Seddone**



Università degli Studi di Torino
Dipartimento di Culture, Politica e Società
Centro "Luigi Bobbio" per la Ricerca sociale, pubblica e applicata

Blue Papers CLB-CPS

FUEL MY FIRE

Social media, politica e hate speech in Italia

Federico Vegetti
Arturo Bertero
Giuliano Bobba
Moreno Mancosu
Antonella Seddone

1 / 2022



Quest'opera è stata rilasciata con licenza Creative Commons Attribuzione 4.0 Internazionale. Per leggere una copia della licenza visita il sito web <http://creativecommons.org/licenses/by/4.0/> o spedisci una lettera a Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

ISBN - 9788875902100

Blue Paper CLB-CPS 1/2022

Centro "Luigi Bobbio" per la ricerca sociale, pubblica e applicata
Dipartimento di Culture, Politica e Società
Università degli studi di Torino
Lungo Dora Siena, 100 - 10153
Torino, TO
www.dcps.unito.it

INDICE

Abstract	iii
Executive Summary	iv
1 HATE SPEECH E SOCIAL MEDIA: UN'AFFINITÀ ELETTIVA?	1
1.1 Social media, politica e cittadini	1
1.2 Il dibattito sui discorsi d'odio online tra hate speech e incivility .	4
1.3 Conseguenze politiche dell'hate speech online	9
2 CASO STUDIO, OPERAZIONALIZZAZIONE E METODO	12
2.1 La diffusione dei social media in Italia e tra i politici	12
2.2 Una definizione operativa di incivility e hate speech	15
2.3 Periodo e fonti analizzate	16
2.4 Metodo e descrizione del dizionario utilizzato	19
3 RISULTATI	23
3.1 Il linguaggio d'odio nei discorsi dei politici sui social media . . .	23
3.2 Incivility ed HS nei commenti degli utenti	26
3.3 L'interazione tra leader e utenti su Facebook: l'odio alimenta odio?	29
Conclusioni	35
Riferimenti bibliografici	37

ABSTRACT

Questo blue paper riporta i risultati di una ricerca volta a offrire una concettualizzazione e quantificazione del linguaggio d'odio tra le pagine Facebook dei principali leader politici in Italia. La prima parte di questo blue paper propone una rassegna teorica della letteratura su *incivility* (utilizzo di linguaggio volgare o che viola norme di educazione condivise) e *hate speech* (utilizzo di insulti o minacce direzionate a individui o gruppi), due fenomeni spesso sovrapposti ma che invece presentano importanti differenze. Nella seconda parte viene descritto un dizionario, o "lessico", basato sulla letteratura sul tema e finalizzato a quantificare *incivility* e *hate speech* all'interno di testi, come i commenti prodotti dagli utenti di piattaforme *social media*. Infine, la terza parte del blue paper presenta i risultati di uno studio empirico che utilizza metodi di analisi testuale quantitativa per osservare la prevalenza di *incivility* e *hate speech* tra i post e i commenti a tali post di 5 leader politici italiani (Giorgia Meloni, Matteo Salvini, Nicola Zingaretti, Silvio Berlusconi e Vito Crimi) osservati da febbraio 2020 a febbraio 2021. Come evidenziato dalle analisi, Meloni, Salvini e Crimi sono i leader che utilizzano con maggior frequenza post caratterizzati da termini catalogabili come *hate speech* o *incivility* (principalmente per riportare commenti di terzi rivolti a loro); allo stesso modo, anche i commenti relativi ai loro post sono quelli in cui è più probabile trovare contenuti analoghi, pubblicati dagli utenti. L'aggressività verbale è la dimensione più prominente dei discorsi d'odio sia per i leader sia per le loro community, mentre il linguaggio volgare (*incivility*) è la seconda dimensione per rilevanza. I post che presentano maggiore *incivility* e *hate speech* sono associati a un numero più elevato di *like*, condivisioni e commenti rispetto a post neutri. Infine, si osserva una corrispondenza, particolarmente significativa per Meloni e Salvini, fra post e commenti in termini di linguaggio contenente *hate speech*. Questo suggerisce che, per quanto riguarda il confronto verbale sui *social media* in Italia, l'odio alimenta odio.

EXECUTIVE SUMMARY

I discorsi d'odio stanno acquisendo centralità nel dibattito pubblico italiano, in particolare a causa della loro crescente diffusione sui social media. Guardando alle conversazioni sui temi politici, questo studio si interroga sulla portata del fenomeno, sulle dimensioni che lo compongono e ragiona sulle implicazioni del ricorso al linguaggio d'odio verificando se il ricorso a questi registri da parte dei leader politici si rintracci anche nei commenti di coloro che interagiscono con quelle pagine.

Benché sempre più spesso nella copertura informativa capiti di incontrare espressioni quali *hate speech*, *shitstorm*, *tempesta d'odio* o altre espressioni ascrivibili alla categoria dei discorsi d'odio, il fenomeno appare estremamente complesso da circoscrivere e quantificare da un punto di vista scientifico. Non che i tentativi non ci siano stati, al contrario. Ciò che ancora manca è un accordo tra gli studiosi su una definizione teorica e sulla sua traduzione empirica. La letteratura, infatti, distingue almeno tra due declinazioni del fenomeno: *l'incivility*, vale a dire il ricorso a contenuti, registri e toni travalicano i confini delle normali conversazioni per virare verso la volgarità, violando le norme di educazione e rispetto senza però avere un target specifico, e *l'hate speech* che si caratterizza, al contrario, proprio per prendere di mira uno o più soggetti che vengono chiamati in causa facendo ricorso a forme di aggressività, quando non di vera e propria discriminazione.

Sulla base di questo dibattito, il *blue paper* propone un'operazionalizzazione originale dei discorsi d'odio che comprende la dimensione dell'*incivility* (intesa come ricorso a un linguaggio volgare) e quella dell'*hate speech* (declinata in aggressività verbale, aggressività fisica e discriminazione). Avendo come obiettivo quello di studiare le conversazioni che si sviluppano intorno ai messaggi dei leader politici, è stato privilegiato un approccio linguistico, sviluppando un dizionario *ad hoc* i cui termini colgono le sopra citate dimensioni. Questo dizionario è stato applicato ai post (n=9.413) e relativi commenti (n=19.753.444) pubblicati sulle pagine Facebook dei leader dei principali partiti politici italiani nel periodo febbraio 2020-febbraio 2021, vale a dire Silvio Berlusconi, Vito Crimi, Giorgia Meloni, Matteo Salvini, Nicola Zingaretti. Facebook è stato scelto poiché è il social media più diffuso tra i cittadini. Inoltre, è ampiamente utilizzato dai

politici e i commenti sono una delle modalità standard di interazione offerte dalla piattaforma.

I principali risultati dell'analisi empirica sono i seguenti:

- **I leader politici si caratterizzano per una frequenza di pubblicazione e, di conseguenza, per un volume complessivo di post molto differenziato:** Salvini (4.818) è di gran lunga il più attivo, Meloni (2.063) non raggiunge la metà dei messaggi del leader della Lega, Zingaretti (1.215), Berlusconi (842) e Crimi (475) seguono a livelli ancora più bassi.
- **I leader politici si caratterizzano per un volume di commenti ai loro post molto differenziato:** il numero medio di commenti per ogni messaggio è elevato non solo per Salvini (2.814) ma anche per Meloni (2.655) che segue a stretto giro. La distanza rispetto agli altri tre leader, invece, è molto ampia dal momento che il loro livello medio si attesta intorno a 300 commenti a post.
- **La percentuale di parole ascrivibili al concetto di *hate speech* nei post dei leader politici raggiunge circa il 10% nella pagina di Meloni, l'8% in quella di Salvini, il 5,5% in quella di Crimi,** mentre per Zingaretti (1,9%) e Berlusconi (0,7%) rappresenta una quota del tutto trascurabile dei loro messaggi. Tuttavia, nella stragrande maggioranza dei casi, si tratta di parole utilizzate per riportare commenti di terzi rivolti ai leader stessi.
- La percentuale di parole ascrivibili al concetto di *hate speech* nei commenti vede gli stessi tre leader nelle prime posizioni, ma con un ordine differente: **la *community* di Crimi fa ricorso a parole d'odio nel 17,2% dei commenti, mentre quella di Meloni (14,7%) e Salvini (13,5%) seguono poco sotto.** Gli stessi registri sono presenti nelle conversazioni sulle pagine di Zingaretti (6,2%) e Berlusconi (3,9%) ma con volumi decisamente più bassi. In questo caso, si tratta principalmente di parole d'odio rivolte al leader, ad altri commentatori o a persone terze.
- Guardando la strutturazione di *incivility* e *hate speech* sia nei post, sia nei commenti emergono alcuni tratti comuni: **l'aggressività verbale è la dimensione più prominente dei discorsi d'odio sia per i leader sia per le loro *community*. Il linguaggio volgare (*incivility*) è la seconda dimensione per rilevanza: lo si ritrova in particolar modo nei commenti di Crimi (7%), Meloni (3,3%) e Salvini (2,9%).** Le dimensioni che conten-

gono riferimenti più offensivi o discriminatori risultano invece marginali sia nei post sia nei commenti.

- **La presenza di *incivility* e *hate speech* nei post di Salvini e Meloni - vale a dire quei leader che fanno maggiore ricorso a questo registro - è in media correlata ad un livello più elevato di *like*, condivisioni e commenti se comparato con post neutri.** Questo *engagement* facilita inoltre la visibilità e la circolazione di questo genere di messaggi all'interno della *community* del politico.
- La corrispondenza fra post e commenti in termini di linguaggio contenente *hate speech* - in altre parole la probabilità che l'odio alimenti l'odio - è particolarmente significativa per Meloni e Salvini. **Quando i due leader pubblicano post che presentano elementi di discriminazione, aggressività verbale o fisica, circa un commento su cinque a sua volta riproduce lo stesso contenuto fungendo da cassa di risonanza per l'*hate speech*.**

L'insieme di questi dati conferma la rilevanza dell'*incivility* e dell'*hate speech* all'interno del dibattito pubblico italiano e sottolinea la necessità di una riflessione sulla qualità del dibattito politico. Se non ci sono dubbi sul fatto che le diverse forme e declinazioni dei discorsi d'odio debbano essere considerate come una vera e propria degenerazione di un dialogo civile e democratico, la sua diffusione rappresenta un fenomeno endemico che va monitorato e contrastato. Occorre intervenire prima che un processo di normalizzazione produca un'ulteriore polarizzazione della contesa politica, esasperando i toni del dibattito e incidendo negativamente sulla convivenza pacifica dei gruppi politici e sociali. In questo contesto, le istituzioni politiche e il giornalismo possono e devono giocare un ruolo chiave nel gestire e disinnescare questi serbatoi d'odio, ma ci riusciranno soltanto se consapevoli del problema e se supportati da evidenze empiriche che restituiscano la rilevanza del fenomeno tenendo conto delle sue diverse manifestazioni e modalità. Questo *blue paper* intende fornire un contributo iniziale, ponendosi come obiettivo quello di offrire a cadenza annuale un report sull'andamento dell'*incivility* e dell'*hate speech* nelle conversazioni sui social media in Italia.

HATE SPEECH E SOCIAL MEDIA: UN’AFFINITÀ ELETTIVA?

1.1 SOCIAL MEDIA, POLITICA E CITTADINI

Internet e i social media hanno profondamente influenzato l’arena e gli attori della comunicazione politica. Da un lato, i social hanno concesso a politici e cittadini di interagire in maniera più immediata e diretta rispetto all’età dell’oro del mezzo televisivo. Dall’altro lato, i social sono ora centrali nel definire i criteri di notiziabilità mediatica, forzando i media tradizionali ad una maggior interazione con il loro pubblico. Chadwick (2013) ha definito i sistemi mediatici contemporanei come *ibridi*, ossia caratterizzati da un equilibrio tra le vecchie logiche di trasmissione e ricezione e le nuove logiche di diffusione, ricircolo e ridefinizione dei contenuti. Se le forme classiche di partecipazione politica non sono state favorite dall’online, forme meno impegnative di interazione con contenuti politici, come i *like*, le condivisioni e i commenti su Facebook, Twitter o Instagram, sono diventate sempre più comuni e stanno contribuendo a ristrutturare il modo in cui la comunicazione e l’informazione politica circola e viene consumata. L’interazione dei cittadini può avere implicazioni nella fase di definizione dei contenuti e delle forme di comunicazione, influenzando per esempio la selezione degli argomenti e i *frame* con cui gli stessi contenuti sono narrati. D’altra parte, sui social media gli utenti sono ormai considerati come parte integrante del processo di distribuzione dei contenuti e delle notizie politiche (Krumsvik 2018; Hille e Bakker 2013).

È stato suggerito che il pubblico online rivesta spesso un ruolo attivo attraverso la sua attività di selezione, interpretazione e produzione (Picone 2015). Queste attività dovrebbero essere considerate come un «continuum su cui un utente può essere attivo in varia misura» (Picone 2015, 128). Hermida si riferisce al pubblico digitale come a "destinatari attivi" (Hermida 2011) che scelgono, cliccano e condividono contenuti politici o notizie, diventando una sorta di "secondary gatekeeper" (Singer 2014).

Il comportamento dei cittadini sui social media si articola lungo diversi tipi

di reazioni e diversi livelli di coinvolgimento. Leggere e condividere contenuti politici, in altre parole, può implicare azioni diverse con esiti diversi. Capire le motivazioni esistenti dietro un clic è fondamentale, ma piuttosto difficile da misurare: Kormelink e Meijer (2018) suggeriscono che le metriche dovrebbero sempre essere interpretate criticamente, perché ci sono almeno tre tipi di considerazioni – cognitive, affettive e pragmatiche – che potenzialmente influenzano i clic.

Le piattaforme dei social media, come Facebook, Twitter o Instagram, permettono agli utenti di interagire con i contenuti prodotti e distribuiti dagli attori politici tradizionali. Facebook, in particolare, offre tre tipi principali di interazione che permettono un feedback da parte degli utenti: il like, la condivisione e il commento.¹ Il *like* di Facebook rappresenta il più basso grado di impegno, in quanto espressione di gradimento (o antipatia) verso qualcosa o qualcuno non richiede molto tempo e sforzo. Questa funzionalità estremamente semplice e senza costi è stata spesso indicata come responsabile della riduzione della partecipazione politica, dal momento che favorirebbe processi di “*clicktivism*” (Karpf 2010) o “*slacktivism*” (Morozov 2011). Altri autori hanno invece considerato i *like* come “una versione ‘leggera’ della partecipazione” (Hille e Bakker 2013) che di solito suggerisce una generica forma di «sostegno verso il contenuto a cui si riferisce» (Larsson 2018, 46), ma che non sempre coincide con un’effettiva approvazione, e può essere usata anche per esprimere stupore o addirittura antipatia. La condivisione è il livello successivo di coinvolgimento. Permette agli utenti di ridistribuire contenuti originariamente postati da altri, compresi i politici e le testate giornalistiche. Dato che la condivisione implica dichiarare pubblicamente le proprie posizioni e incoraggiare gli altri a leggere il contenuto, può essere socialmente più costoso, e per questo motivo è meno diffuso del *like*. Commentare, infine, comporta l’impegno maggiore dei tre tipi di azioni, poiché scrivere un contenuto con le proprie parole richiede più tempo e impegno e rappresenta esplicitamente o implicitamente la volontà di trasmettere un’opinione personale. Non sorprende che questo tipo di interazione sia il meno utilizzato, rispetto ai due precedenti. Commentare implica anche che il post originale è stato in qualche modo abbastanza coinvolgente da provocare una forte reazione. Non solo, il commento può potenzialmente influenzare il modo in cui gli altri utenti di Facebook percepiranno quel contenuto.

¹ Dal 2016, oltre ai *like*, gli utenti possono anche esprimere le loro reazioni attraverso cinque nuove emozioni: amore, *haha*, wow, triste e arrabbiato. Ad aprile 2020, in occasione della pandemia COVID-19, a queste si è aggiunta anche l’opzione *hug*

Nell'ambiente online le relazioni fra gli individui sono caratterizzate da una diluizione dello spazio-tempo. Chiunque in possesso di un accesso internet può intervenire, offrire la propria opinione, contribuire alla conoscenza e alla creazione di una cultura condivisa, una "*cyberculture*" (Sacristán 2013). L'impatto a livello micro, meso e macro è difficile da stimare, ma è chiaro a tutti che internet è la rivoluzione tecnologica che più di tutte negli ultimi decenni ha ridefinito la vita quotidiana degli individui. La comunicazione, intesa come produzione e condivisione di informazione, è il cuore della funzionalità di internet, ma presenta anche alcuni aspetti critici e potenzialmente dannosi (Assimakopoulos et al. 2017).

L'interazione online prevede un isolamento dalle convenzioni sociali, e questo indurrebbe un meccanismo di disinibizione rendendo gli individui più inclini a esprimere disaccordo rispetto alle opinioni altrui (Santana 2014). Suler (2004) sottolinea che questo meccanismo può avere esiti positivi nel momento in cui contribuisce all'espressione di opinioni personali altrimenti taciute, ma può produrre effetti perversi nel momento in cui l'ambiente online è percepito come separato da una dimensione reale offline. I social media, dunque, diventerebbero un'arena dove i codici comportamentali classici possono essere sovvertiti, o semplicemente disattesi e le relazioni sociali seguono dinamiche differenti. Questa percezione di dissociazione fra online e offline è favorita dalla possibilità di utilizzare l'anonimato nelle interazioni sul web o sui social media. Gli utenti utilizzano la dimensione online per esprimere i loro punti di vista e le loro opinioni e l'ambiente stesso offre la possibilità di farlo giocando sull'anonimato, proiettando sulle loro interazioni online una erronea e falsata idea di impunità.

La natura globale, immediata e partecipativa di internet è dunque contemporaneamente il suo punto di forza ma anche la sua debolezza. L'ambiente online in altre parole sarebbe lo spazio in cui l'espressione e la diffusione di idee o credenze intolleranti trovano più facile e massiccia diffusione (Assimakopoulos et al. 2017).

La ricerca scientifica ha cercato di comprendere il fenomeno sotto punti di vista differenti, ma mettendo al centro il web e i social media in particolare come ambienti dove i discorsi d'odio trovano una risonanza e una circolazione più estesa (Andrisani, 2014; Ziccardi, 2016; Hopp e Vargo, 2017). Nel prossimo paragrafo questa connessione tra discorsi d'odio e social media viene affrontata all'interno del dibattito scientifico su *hate speech* e *incivility*.

1.2 IL DIBATTITO SUI DISCORSI D'ODIO ONLINE TRA HATE SPEECH E INCIVILITY

Discutere di *hate speech* (HS) non è cosa semplice. In genere con questo termine si indicano quei contenuti comunicativi che mirano a biasimare e attaccare specifici gruppi sociali (Jacobs e Potter 2000; Walker 1994). Eppure, quando si tratta di tracciare il perimetro del fenomeno scopriamo che non disponiamo di una definizione univoca e condivisa e che le declinazioni di linguaggio d'odio sono piuttosto eterogenee per quello che è stato definito un "*moving concept*" (Bentivegna e Rega 2020). Questo complica non poco i tentativi di identificare, comprendere e dunque intervenire sul fenomeno. Da un lato, la libertà d'espressione è essa stessa oggetto di interpretazioni che risentono delle specificità culturali e normative nazionali rendendo quindi arduo convergere non solo su una definizione condivisa, ma anche sulle stesse misure di contenimento del fenomeno (cfr. Alkiviadou 2019; Assimakopoulos et al 2017). Dall'altro lato, invece, il web e i social media, in particolare, si sono configurati come uno spazio di diffusione dei discorsi d'odio che ha introdotto dinamiche e modalità comunicative e relazionali del tutto inedite (cfr. Rossini 2021; Poletto et al. 2020).

L'ambiente digitale si configura infatti come uno spazio di circolazione del tutto nuovo per questo tipo di contenuti (Theocharis et al. 2020). I meccanismi di viralità che caratterizzano il web contribuiscono ad enfatizzare e per certi versi a rafforzare il fenomeno (Alkiviadou 2019). Peraltro, Sydnor (2018) osserva che i soggetti che partecipano a dibattiti ad alto tasso di *incivility* più che risultare offesi segnalano un maggiore coinvolgimento e per questo sarebbero più inclini a replicare quel registro. Brown (2018) nel riflettere sulle specificità che distinguono i discorsi d'odio nella dimensione online e offline, osserva come l'ambiente digitale e in particolare i social media rappresentino uno spazio in cui facilmente si innescano meccanismi di de-individualizzazione e deresponsabilizzazione. I classici codici sociali a cui gli individui si affidano generalmente nella quotidianità delle loro interazioni sociali appaiono meno stringenti e vincolanti. Le ragioni sono diverse. Ad esempio, secondo alcuni studiosi nella dimensione online mancano quegli elementi visuali che contribuiscono a definire i limiti delle interazioni (Oz et al. 2017): l'invisibilità – fatta salva la crescente sovrapposizione di elementi visuali e testuali nello spazio web – agevola l'idea di una sospensione delle norme sociali (Brown 2018). La letteratura sottolinea inoltre come l'anonimato – più percepito che reale – che contraddistingue le piattaforme digitali rafforzi meccanismi comportamentali più estremi, impulsivi e meno accettabili da un punto di vista sociale (Rowe

2015; Halpern e Gibbs 2013; Santana 2014). In altre parole, gli studi sembrano concorde nel sostenere che il web, i social media e in generale le piattaforme digitali rendono più facile (e accettabile) travalicare i canoni classici della comunicazione interpersonale e dunque le norme dell'interazione sociale. Il risultato è che con sempre maggiore frequenza le interazioni sono contraddistinte da elementi di natura offensiva o addirittura violenta. Ciò che è ancor più importante è che le implicazioni, non si limitano alla dimensione virtuale. Gli studi infatti osservano che i discorsi d'odio non riflettono semplicemente opinioni già diffuse e presenti nella società, semmai radicano stereotipi e consolidano sentimenti negativi contro specifici gruppi (Martin et al. 2013). Analogamente, Bunarp e Williams (2014) sottolineano che i discorsi d'odio possono facilmente tradursi in crimini veri e propri con conseguenze talvolta molto severe per le vittime.

Ricostruire il dibattito su HS e linguaggi d'odio da un punto di vista giuridico non è sicuramente fra gli obiettivi di questo blue paper. Tuttavia, può essere utile richiamare alcuni snodi normativi rilevanti che hanno segnato un cambiamento nell'approccio generale al tema. Da questo punto di vista, il primo riconoscimento ufficiale a proposito dei discorsi d'odio è rintracciabile a metà degli anni '60, con il Patto internazionale relativo ai diritti civili e politici, adottato a New York nel 1966 (Assimakopoulos et al. 2017)². Il testo bandisce ogni forma di discorso d'odio, e più precisamente vieta appelli all'odio nazionale, razziale o religioso, che incitano la discriminazione, l'ostilità, la violenza e la guerra. Sarà poi il *Council of Europe* ad aggiornare questa definizione con la Raccomandazione n. 20 (30 ottobre 1997). In quella sede, sono identificati come linguaggio d'odio tutte quelle espressioni che diffondono, incitano, promuovono o giustificano odio razziale, xenofobia, antisemitismo, o altre forme di disprezzo basate su intolleranza come ad esempio: nazionalismo aggressivo, etnocentrismo, discriminazione e ostilità contro le minoranze, i migranti o le persone di origine straniera. Per quanto questa definizione allarghi lo spettro delle declinazioni possibili di HS, in verità le categorie trascurate sono ancora numerose. Più specificamente, manca qualsiasi riferimento ai discorsi d'odio indirizzati alla comunità LGBTQ+ e alle persone disabili, che pure sono un target frequente (Assimakopoulos et al. 2017). Nel 2003 il *Council of Europe* torna sul tema³ per dare più precisa definizione al concetto, soprattutto alla luce della sua

² Il provvedimento sarà reso esecutivo in Italia nel 1977 con la Legge n. 881.

³ *Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems* (COE – European Treaty Series n 189 – 28/1/2003). Hyperlink: <https://rm.coe.int/168008160f>

diffusione nello spazio web. Tuttavia, ancora una volta i discorsi d'odio sono circoscritti alle dimensioni di razzismo, xenofobia e nazionalismo trascurando di citare elementi relativi a omofobia, bifobia e transfobia, disabilità o genere. Analogamente, l'Unione Europea nel 2008, pur citando esplicitamente il tema dell'HS, persiste nel focalizzare l'attenzione su contenuti d'odio a tematica razzista o xenofobica, tralasciando ancora una volta tematiche relative alla comunità LGBTQ+ e alla disabilità.⁴ Alkiviadou (2019) sottolinea che questa ambiguità nella definizione giuridica lascia spazio non solo ad ampia discrezionalità nell'identificazione (e quindi nella risposta) del fenomeno, ma soprattutto restituisce l'idea di una sorta di gerarchia dell'odio. Legalmente la protezione dai discorsi d'odio è riconosciuta solo per alcune vittime, mentre altre categorie e gruppi restano in uno spazio d'ombra dove diventa più difficile contrastare l'HS.

In questa sede, ci preme sottolineare che, nonostante a livello internazionale esistano delle linee guida ufficiali, quando si entra nel dettaglio delle normative nazionali il tema si fa ancora più complesso e si registra una grande eterogeneità. Infatti, il fenomeno dell'HS e la sua definizione risentono di specificità culturali e normative nazionali che inevitabilmente impattano anche sulla efficacia delle policy a contrasto del fenomeno.

In Italia la legislazione più rilevante in merito è la L. 205/1993 che definisce crimine la propagazione di idee basate su idee di superiorità razziale o odio etnico, o che istighino a commettere atti di discriminazione razziale, etnica o religiosa. La legge, inoltre, punisce anche chi istighi o commetta atti di violenza o provocazione di violenza sulla base di motivazioni razziali, etniche o religiose. Il caso italiano è interessante da questo punto di vista: secondo la normativa vigente è fatto divieto ai media di trasmettere contenuti che presentino incitamento all'odio o comportamenti intolleranti sulla base di razza, genere, religione o nazionalità. Ma – come è facile intuire – la normativa italiana non tiene conto dell'evoluzione tecnologica e, quando si tratta di gestire fenomeni di HS in ambienti online, la questione diventa più complessa e l'elaborazione di strategie di contrasto più difficile. Soprattutto, in Italia non esiste ancora una definizione normativa specificamente focalizzata sull'HS al punto che in questa legislatura è stata istituita un'apposita Commissione parlamentare – la cosiddetta Commissione Segre – al fine di fornire strumenti legislativi di indirizzo e controllo sui fenomeni di intolleranza, razzismo, antisemitismo e

⁴ *Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law.* Hyperlink: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32008F0913>

istigazione all'odio e alla violenza nei confronti di persone o gruppi sociali sulla base di alcune caratteristiche quali l'etnia, la religione, la provenienza, l'orientamento sessuale, l'identità di genere o di altre particolari condizioni fisiche o psichiche.

L'evoluzione del fenomeno e la sua diffusione nello spazio web richiede quindi un ulteriore aggiornamento definitorio in grado di coglierne le specificità. A questo proposito è interessante notare come anche le compagnie di social network offrano una definizione più larga dei contenuti di odio e prevedano esplicitamente alcune sanzioni per gli utenti che violano questi principi (es. la rimozione dei contenuti o anche il *ban* dell'utente). YouTube, ad esempio, dichiara di incoraggiare la libertà di parola e asserisce di difendere la possibilità di ognuno di esprimere punti di vista impopolari, ma stabilisce anche dei chiari limiti in tal senso. Più precisamente non sono ammessi attacchi o insulti a individui o gruppi sulla base di questioni di razza o etniche, religiose, disabilità, genere, età, status di veterano (rilevante negli Stati Uniti), orientamento sessuale o identità di genere. In maniera molto simile anche Facebook prevede la rimozione diretta di contenuti che attacchino direttamente le persone sulla base della loro razza, etnia, origini nazionali, fede religiosa, orientamento sessuale, identità di genere, disabilità o malattie. Twitter invece su questo tema adotta una posizione differente. Non fa esplicito riferimento a contenuti d'odio, né elenca modalità o declinazioni possibili: non proibisce, ma semmai avverte i suoi utenti che potrebbero essere esposti a contenuti ritenuti offensivi, pericolosi, inaccurati o anche inappropriati. Soprattutto, evitando di monitorare o controllare il contenuto postato, declina di fatto ogni responsabilità.

Nel 2016 la Commissione Europea riconoscendo internet e i social media come un ambiente peculiare per lo sviluppo e la diffusione dell'HS chiama in causa direttamente le compagnie di social network e web. Il 31 maggio 2016, la Commissione Europea concorda con Facebook, Microsoft, Twitter e YouTube il *EU Code of conduct on countering illegal hate speech online*.⁵ Si tratta di un approccio al fenomeno del tutto inedito. Infatti, seppure le società IT abbiano in larga misura disatteso le indicazioni o comunque preso sottogamba le predisposizioni della Commissione Europea, la scelta di coinvolgere le piattaforme IT e social media ha rinnovato l'attenzione sul fenomeno e indotto molti paesi ad adottare

⁵ Hyperlink: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

normative nazionali in materia.

Se ci spostiamo sul fronte della ricerca scientifica, il ragionamento non è meno complesso. Infatti, anche in questi ambiti, ci si scontra con l'assenza di una definizione condivisa in grado di coprire tutte le possibili articolazioni dei discorsi d'odio in termini di contenuti e modalità. Il che implica, sul versante empirico, una grande varietà di operazionalizzazioni del concetto e la conseguente moltiplicazione di misure e misurazioni.

La letteratura in genere fa riferimento al concetto di *incivility* per identificare quelle dinamiche comunicative dove contenuti, registri e toni travalicano i confini delle normali conversazioni per virare su modalità di offesa, violando le norme di educazione e rispetto (Rossini 2021). Si tratta di contenuti irrispettosi mirati a silenziare o anche criticare visioni alternative (Jamieson 1997). Più precisamente, l'online *incivility* si caratterizza per l'utilizzo di stereotipi negativi, nomignoli dispregiativi, disprezzo e maleducazione (Chen e Lu 2017; Coe et al 2014). Rossini (2020) sottolinea che da un punto di vista empirico - e soprattutto quando si analizza la comunicazione online - l'*incivility* coincide generalmente con contenuti comunicativi che includono linguaggio volgare (Coe et al. 2014), l'uso di lettere maiuscole a simulare l'urlo, la calunnia, il linguaggio peggiorativo, screditante, infamante, estremo o emozionale, l'exasperazione dell'argomentazione (Chen 2017; Chen e Lu 2017; Coe et al. 2014; Papacharissi 2004; Gervais 2015; Sobieraj e Berry 2011).

Di fatto, il termine *incivility* copre una serie di contenuti e comportamenti piuttosto ampi in termini di offesa e tossicità. Molte ricerche concettualizzano l'*incivility* come un continuum dove l'utilizzo di linguaggio volgare - spesso non direzionato a un preciso target - rappresenta il polo più blando (Coe et al. 2014; Muddiman 2017), mentre contenuti comunicativi di natura aggressiva e discriminatoria (razzismo, sessismo, xenofobia e omofobia o attacco diretto a soggetti o gruppi minoritari) si configurano come il polo più estremo ed esasperato di *incivility* (Chen 2017). Rossini (2021) mette in guardia rispetto a questa definizione ampia e inclusiva di *incivility* perché non consentirebbe di cogliere le distinzioni tra il semplice linguaggio volgare (ma non necessariamente offensivo) e i contenuti intolleranti e discriminatori (Papacharissi 2004). Queste definizioni sembrano suggerire che il fenomeno dell'*incivility* sia circoscritto al ricorso ad un linguaggio volgare, talvolta offensivo che non sfocia tuttavia nella violenza o nell'odio esplicito: mancherebbe qui un target, o comunque il discorso non consisterebbe in attacchi interpersonali (Su et al. 2018). In questa

prospettiva, l'HS sarebbe quindi una forma *sui generis* ed estrema di *incivility*. Ai generici elementi offensivi o volgari si accompagnano contenuti espliciti di attacco, abuso, intimidazione e denigrazione direzionati verso singoli o gruppi (Walker 1994). L'HS è quindi quel tipo di linguaggio utilizzato per esprimere odio verso un target che nei casi più estremi ricorre anche a minacce, aggressività ed esplicita promozione della violenza (Davidson et al. 2017). Fortuna et al. (2020) hanno definito l'HS come un linguaggio che attacca o sminuisce, che incita alla violenza o all'odio contro gruppi di individui sulla base di specifiche caratteristiche come l'aspetto fisico, la religione, la discendenza, l'appartenenza nazionale o etnica, l'orientamento sessuale, l'identità di genere o altro. L'HS può inoltre essere espresso attraverso registri linguistici differenti, per cui accanto al linguaggio offensivo o aggressivo si rintracciano anche forme canzonatorie e di presa in giro. Anche in questo caso, quindi, ciò che distingue l'HS da forme generali di *incivility* è la presenza di un target (Fortuna et al. 2020).

Tabella 1: Sistematizzazione teorica della composizione dei discorsi d'odio

Variabile	Tipo di messaggio	Target	Indicatore di
<i>Incivility</i>	Non direzionato	-	Incivility
Aggressività verbale	Direzionato	Individuo generico	Hate speech
Aggressività fisica	Direzionato	Individuo generico	Hate speech
Discriminazione	Direzionato	Gruppo sociale/categoria	Hate speech

La Tabella 1 sintetizza la distinzione fra *incivility* e HS facendo riferimento alle quattro macrocategorie di linguaggio che ricorrono nelle numerose definizioni offerte in letteratura. I due concetti sono legati e per certi versi contigui. Tuttavia, mentre l'*incivility* si caratterizza principalmente per il ricorso a un linguaggio volgare privo di un target preciso, la specificità dell'HS è quella di esprimere aggressività verbale o fisica e di promuovere contenuti discriminatori, direzionati verso target individuali o collettivi in ragione di stereotipi negativi legati a razza, nazionalità, etnia, genere, identità di genere, orientamento sessuale, aspetto fisico e disabilità.

1.3 CONSEGUENZE POLITICHE DELL'HATE SPEECH ONLINE

La letteratura si è molto focalizzata sul fenomeno dell'HS e dell'*incivility* in ambito politico. Come sottolineano Theocharis et al (2020), la politica si articola

e struttura intorno a contrapposizioni e la discussione si caratterizza spesso per i toni accesi, tale dinamica stimola spesso l'HS. Recentemente, si è osservata una decisa crescita del fenomeno: se l'*incivility* era già un tratto che connotava il dibattito politico sui media (Mutz e Reeves 2005), anche il confronto fra attori politici può virare talvolta su registri equivoci con ampio uso di linguaggio volgare o offensivo (Mutz 2015). La diffusione dei social media ha orientato il dibattito accademico su questi temi. Le ricerche più recenti hanno evidenziato che piattaforme quali Twitter e Facebook, consentendo un'interazione quasi diretta fra pubblico e politica, hanno tra le altre cose creato delle condizioni favorevoli allo sviluppo di dinamiche di *incivility* o anche di HS (Theocharis et al 2020).

L'*online incivility*, specialmente se declinata a livello politico, ha precise e rilevanti implicazioni sul funzionamento delle democrazie contemporanee soprattutto perché, dati i meccanismi di viralità che connotano le piattaforme di social media, si ingenera facilmente una sorta di normalizzazione del fenomeno, vale a dire un progressivo adattamento – e contestuale assuefazione – a registri e toni incivili (Coe et al. 2014). Le conseguenze a livello sistemico sono rilevanti. Alcune ricerche hanno evidenziato come l'*incivility* fra politici sia spesso un fattore che favorisce sentimenti di insoddisfazione verso le istituzioni politiche e rafforza pregiudizi negativi nei confronti delle élite (Brooks e Geer 2007). Mutz e Reeves (2005) giungono a risultati analoghi per quanto riguarda la fiducia nei media, mentre Muddiman (2017) registra una associazione fra contenuti di *incivility* e disinformazione.

Le implicazioni si estendono tuttavia anche in termini di fiducia interpersonale: l'*online incivility* fra cittadini all'interno di forum, chat o blog non solo impatta sulla fiducia politica a livello individuale, ma determina anche un incremento di aggressività che si traduce in polarizzazione delle opinioni (Anderson et al. 2014). Soprattutto comportamenti di replicazione del linguaggio possono determinare una normalizzazione del fenomeno (Gervais 2015). Inoltre, studi sperimentali hanno segnalato che contenuti di HS contribuiscono ad accrescere i livelli d'ansia e producono un effetto di de-mobilitazione e questo effetto si riscontra specialmente fra individui appartenenti a minoranze socioculturali (Munger 2016).

Una parte della letteratura, infine, sostiene che alcuni aspetti dell'*incivility* possano anche generare effetti positivi o comunque non necessariamente negativi (Chen et al. 2019). In alcuni contesti, come rilevato da Faulkner (2009) il linguaggio volgare può essere un elemento che crea coesione e condivisio-

ne. In altre ricerche invece emerge come l'*incivility* sia un elemento che attira l'attenzione e accresce l'interesse per la politica, come pure l'inclinazione alla partecipazione (Mutz e Reeves 2005; Brooks e Geer 2007; Chen 2017). Altri studi sottolineano come i contenuti di *incivility* presentano un livello elevato di coinvolgimento che spinge più facilmente altri utenti a partecipare e commentare (Borah 2014).

CASO STUDIO, OPERAZIONALIZZAZIONE E METODO

2.1 LA DIFFUSIONE DEI SOCIAL MEDIA IN ITALIA E TRA I POLITICI

Secondo i dati più recenti (Digital 2021), in Italia si registrano circa cinquanta milioni di utenti connessi ad Internet. Ciò significa che più dell'ottanta per cento della popolazione naviga in rete. Gli *smartphone* attivi nel nostro paese sono più di settantasette milioni, e quarantuno milioni di italiani, pari a circa il 70% della popolazione, dichiarano di avere un account su un social network.

I social media sono indubbiamente il mezzo di comunicazione che contraddistingue l'epoca attuale. Le piattaforme di social media, come Facebook, Twitter o Instagram (per citare le più diffuse) consentono agli utenti di interagire con i contenuti pubblicati online attraverso flussi comunicativi individuali o collettivi, mescolando la logica *one-to-one* con la logica *one-to-many* e *many-to-many*. Gli utenti, in altre parole, possono usare i social media per inviare messaggi privati ad altri utenti, oppure conversare all'interno di gruppi, pagine pubbliche o sulle proprie bacheche non solo con i loro contatti, ma anche con persone che non conoscono. Quest'ultimo elemento è particolarmente rilevante per quanto riguarda *incivility* ed HS perché, come già ricordato, questa assenza di conoscenza reciproca – mi rivolgo a qualcuno che non conosco e che non mi conosce di fronte a un pubblico che non mi conosce e che non conosco – favorisce la percezione di mancanza di sanzioni sociali, presenti invece nella vita offline.

Dal momento che questa ricerca si concentra sulla parola scritta – verificando la presenza di HS nei commenti – la scelta è caduta sul social media in cui più frequentemente avvengono questo tipo di interazioni tra gli utenti, vale a dire Facebook. Mentre altri social media sono incentrati sulla condivisione di immagini (Instagram, Snapchat, etc.) o video (YouTube, TikTok, etc.), oppure si caratterizzano per un pubblico più di nicchia (Twitter, Reddit, LinkedIn), Facebook pur offrendo la possibilità di condividere contenuti multimediali prevede che l'interazione tra utenti avvenga principalmente attraverso conversazioni testuali che si sviluppano nei commenti ai post. Facebook è una piattaforma con un'attitudine generalista che lo ha portato a essere il social media più diffuso al mondo con una media di 2 miliardi e 274 milioni di utenti attivi al mese (Digital

2021). Anche in Italia, si conferma tra i social media più utilizzati, secondo solo a Youtube (Tabella 2). Tuttavia, mentre gli utenti di Youtube normalmente non creano contenuti ma accedono per vedere contenuti, Facebook è di gran lunga la prima piattaforma per quanto riguarda gli utenti che attivamente contribuiscono al flusso comunicativo attraverso la pubblicazione di post, reazioni e commenti. Al momento, Facebook conta trentasette milioni di utenti attivi in Italia (Statista 2020).

All'interno di questo ambiente digitale, la politica non si è potuta esimere dall'essere presente. Partiti, leader e attori politici vari hanno ormai account ufficiali sui principali social media, diventate di fatto i principali canali di comunicazione con il proprio elettorato, i giornalisti e talvolta anche il pubblico generalista. Ovviamente non tutti – partiti o leader che siano – hanno avuto pari fortuna sui social media e non sempre il risultato elettorale o le intenzioni di voto segnalate dai sondaggi vanno di pari passo con il successo sui social media in termini di *follower* e di interazioni con gli utenti (Tabella 3). I dati italiani confermano tendenze radicate anche nel resto d'Europa (Engesser, Fawzi e Larsson 2017): sui social media i leader hanno più seguito dei partiti, mentre i leader dei cosiddetti partiti populistici (Lega, M5S, FdI) ottengono performance migliori rispetto agli altri leader (Tabella 4).

Tabella 2: Percentuale di utenti che dichiarano di aver usato una delle seguenti piattaforme nell'ultimo mese (età: 16-64). Fonte: Digital 2021

	%
Youtube	85,3
Facebook	80,4
Instagram	67,0
Twitter	32,8
Linkedin	31,2
Pinterest	28,0
TikTok	23,9

Caratteristica italiana è invece quella di avere Facebook non solo come social media più diffuso tra la popolazione, ma anche come luogo privilegiato della comunicazione politica. Tutti i leader, con la parziale eccezione di Zingaretti, hanno infatti un seguito maggiore su questo social media piuttosto che su

Twitter o Instagram (Tabella 3, pagina seguente). Facebook è la piattaforma più frequentata, ed è dunque anche quella in cui si registra il maggior volume di interazioni tra leader e cittadini.

Tabella 3: *I principali partiti e leader politici italiani nei sondaggi e sui social media*

	Elezioni Europee 2019	Intenzioni di voto (giugno '21*)	Numero di <i>follower</i>		
			Facebook	Instagram	Twitter
Giorgia Meloni (Fdi)	6,4%	20%	1.812.494	936.000	1.000.000
Matteo Salvini (Lega)	34,3%	21%	4.514.241	2.300.000	1.400.000
Nicola Zingaretti (PD)	22,7%	19%	371.114	121.000	571.485
Silvio Berlusconi (FI)	8,8%	7%	1.091.525	352.000	161.650
Vito Crimi (M5S)	17,1%	16%	178.771	4.449	84.224

*Fonte: Politico.eu. <https://www.politico.eu/europe-poll-of-polls/italy/>

Tuttavia, queste interazioni – che si tratti di cliccare una semplice reazione, condividere o commentare – non sono distribuite in maniera proporzionale alla forza elettorale o al ruolo giocato nel dibattito pubblico. La Tabella 4 mostra una netta spaccatura tra Salvini e Meloni – capaci di generare nei tredici mesi analizzati decine di milioni di reazioni, commenti e condivisioni – e gli altri leader che ottengono performance anche 25/30 volte inferiori.

Tabella 4: *Interazioni totali dei principali leader politici italiani su Facebook (febbraio 2020-febbraio 2021). Fonte: Crowdtangle.com*

	Totale interazioni	Totale reazioni	Totale commenti	Totale condivisioni
Giorgia Meloni	64.7 M	44.9 M	9.7 M	10.1 M
Matteo Salvini	160.1 M	111.5 M	25.5 M	23.1 M
Nicola Zingaretti	4.0 M	3.0 M	708.7 K	334.2 K
Silvio Berlusconi	3.1 M	2.4 M	461.2 K	202.3 K
Vito Crimi	1.3 M	741.5 K	332.0 K	207.3 K

2.2 UNA DEFINIZIONE OPERATIVA DI INCIVILITY E HATE SPEECH

Come accennato in precedenza, il termine "*hate speech*" ha avuto grande diffusione negli ultimi anni. Nel linguaggio comune, lo si utilizza per indicare il generico ricorso a un linguaggio volgare, offensivo, violento o discriminante. Concentrando la nostra analisi sui social media e sugli ambienti digitali, abbiamo adottato una definizione ampia e articolata che identifica quattro principali dimensioni dei discorsi d'odio che muovono da un livello meno intenso e non direzionato (*incivility*) a un livello più intenso e direzionato (HS, cfr. Tabella 5).

1. La prima dimensione riguarda l'utilizzo di *linguaggio volgare*. Pur non essendo una forma di istigazione d'odio in senso stretto, il ricorso ad un linguaggio volgare è stato considerato come un primo indicatore di una interazione che avviene al di fuori o ai confini di quelli che sono considerati i requisiti base di una conversazione civile. Non a caso nella letteratura accademica, accanto all'espressione "*hate speech*", ha fatto la sua comparsa anche il termine "*incivility*", che riguarda un ricorso generico a parolacce ed espressioni volgari che però non sono rivolte specificatamente contro un altro soggetto.
2. La seconda dimensione si riferisce all'*aggressività verbale*. In questo caso parolacce e insulti sono direzionati, ossia utilizzati con il fine di offendere un altro soggetto verbalmente. Questo tipo di linguaggio può rivolgersi ad almeno due diversi target. Da un lato, l'offesa può essere indirizzata un altro soggetto passivo, con cui non può esserci interazione. Dall'altro lato, invece, questo tipo di linguaggio può essere rivolto verso uno o più utenti attivi che reagiscono e interagiscono tra loro, generando una spirale di insulti reciproci.
3. La terza dimensione implica (l'evocazione di) *aggressività fisica*. Si tratta, in buona sostanza, di una versione più violenta della dimensione precedente che implica minacce direzionate. Anche in questo caso, possono essere genericamente rivolte verso un soggetto passivo oppure verso un utente che può prendere parte alla conversazione. Questo tipo di aggressività si distingue dalla precedente per l'intenzione di minacciare un altro soggetto facendo riferimento alla sua incolumità.
4. La quarta dimensione, infine, coglie in senso stretto l'essenza dell'HS poiché identifica quelle espressioni di *discriminazione* che sono volte ad offendere un altro soggetto sulla base delle sue caratteristiche quali etnia, religione, genere, orientamento sessuale, condizioni fisiche o psichiche o semplicemente per la sua diversità rispetto a modelli dominanti,

considerati normali.

Tabella 5: *La composizione di incivility e hate speech: definizioni ed esempi*

Definizione delle dimensioni di hate speech	Esempi
<i>Incivility</i>	
Uso generico di parolacce ed espressioni volgari non rivolte a un altro soggetto	«Cazzo vi ho votato per 12 anni non me lo perdonerò mai» «Ma che due coglioni con questa storia»
Aggressività Verbale	
Uso di parolacce e insulti volti a offendere un altro soggetto verbalmente	«Ma vaffanculo va, traditore» «Siete dei vigliacchi» «Vergognatevi!»
Aggressività Fisica	
Uso di parolacce e insulti volti a sopraffare un altro soggetto o minacciarlo fisicamente	«Darei fuoco a tutti!» «Dovete morire» «Ci vogliono le ruspe!»
Discriminazione	
Uso di parolacce e insulti volti ad offendere un altro soggetto a causa della sua appartenenza a una categoria sociale subordinata	«Ritardati del cazzo» «Buffone zingaro» «Ciccione flaccido dovresti cambiare dieta» «Ma chi è sta zinnona?»

2.3 PERIODO E FONTI ANALIZZATE

Le analisi proposte in questo report si focalizzano su un periodo di 13 mesi. I dati raccolti, infatti, coprono un arco di tempo che va dal 1° febbraio 2020 al 28 febbraio 2021. Si tratta di un frame temporale piuttosto particolare dove l'emergenza sanitaria legata alla pandemia da COVID-19 ha segnato l'agenda politica italiana e orientato dunque l'attenzione dei leader politici.

La ricerca ha preso in esame tutti i post Facebook pubblicati sugli account ufficiali dei principali leader politici italiani. In totale sono stati analizzati 9.413 post. Per ogni contenuto sono stati poi presi in considerazione tutti i commenti ricevuti. In totale si tratta di 19.753.444 di commenti. Post e commenti sono stati quindi analizzati tramite un questionario sviluppato ad hoc al fine di

individuare l'eventuale presenza di incivility e HS.

La Tabella 6 mostra la distribuzione di post e commenti tra i partiti e leader analizzati. I dati mostrano uno scenario piuttosto differenziato. Matteo Salvini emerge come il leader di gran lunga più attivo su Facebook con un totale di 4.818 post pubblicati, vale a dire una media di circa 14 post al giorno. Anche la leader di Fratelli d'Italia, Giorgia Meloni, presenta un'intensa attività social ma si ferma a 2.063 post (in media poco meno di 6 al giorno). Gli altri leader, invece, si distinguono per volumi meno rilevanti. Luca Zingaretti nel periodo osservato pubblica un totale di 1.215 post, Silvio Berlusconi, invece, ne pubblica 842 e chiude Vito Crimi con 475 post pubblicati lungo i 13 mesi di rilevazione. In altre parole, se il leader PD pubblica circa 4 post al giorno e per Berlusconi siamo a circa 2 post, con il leader incaricato del Movimento 5 Stelle l'attività su Facebook si riduce a una pubblicazione quotidiana. Chiaramente queste cifre si riflettono anche sul fronte delle interazioni e dunque dei commenti. In media i post pubblicati da Salvini registrano circa 2.813 commenti, mentre quelli di Meloni si attestano a circa 2.655 commenti. Questo dato – confermato anche dalle mediane – è interessante. A fronte di una significativa differenza nel volume di post pubblicati e nell'numero di *follower*, infatti, i due leader ottengono un numero di commenti piuttosto simile. L'attività social di Meloni, in altre parole, sembra avere un potenziale superiore rispetto a quello del leader leghista. Osserviamo in questo frangente che è proprio Meloni a registrare il numero massimo di commenti a un post (20.343).

Tabella 6: Volume di post e commenti nelle pagine dei principali leader politici italiani (febbraio 2020 – febbraio 2021)

	Post totali	Media post giornalieri	Media commenti per post	Mediana commenti per post	Min commenti per post	Max commenti per post
Giorgia Meloni	2.063	6	2.655	1725	29	20.343
Matteo Salvini	4.818	14	2.814	1.716,5	26	20.088
Nicola Zingaretti	1.215	4	263	127	5	17.938
Silvio Berlusconi	842	2	293	145	13	18.961
Vito Crimi	475	1	324	117	6	16.799

Gli altri tre leader inclusi nell'analisi giocano di fatto su un livello diverso. In media i post di Zingaretti ottengono 263 commenti, mentre i contenuti pub-

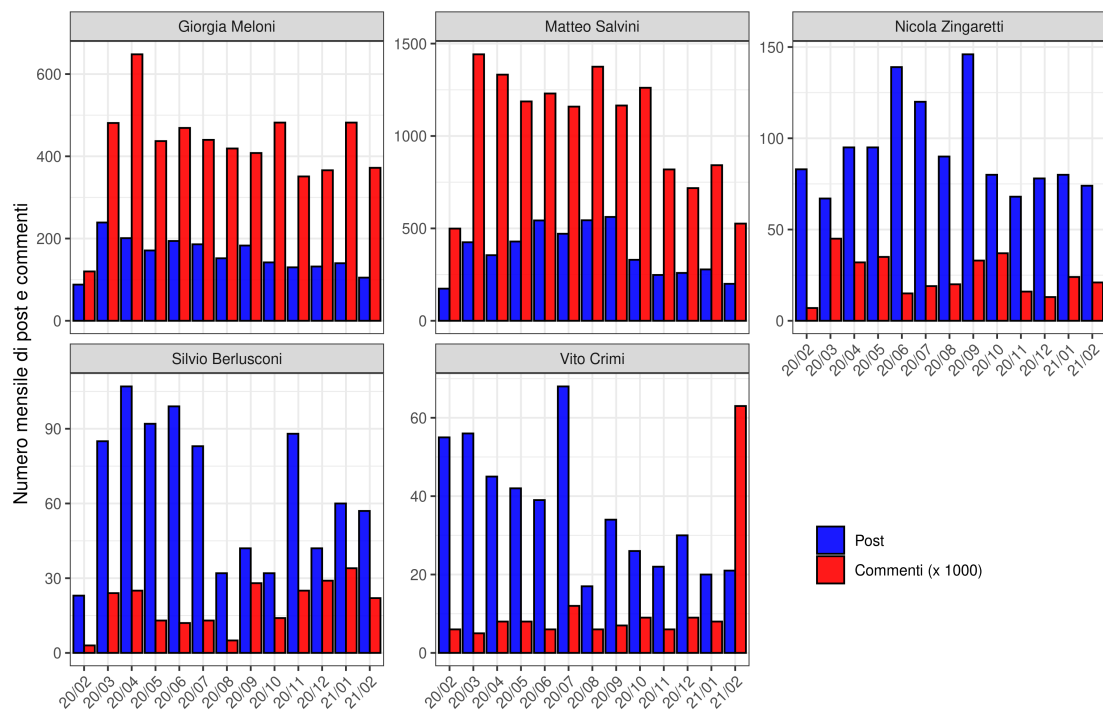
blicati su Facebook da Berlusconi raggiungono la quota di circa 293 commenti. Crimi – che pure avevamo visto in coda per volumi di post pubblicati – registra però un numero medio superiore di commenti (324 circa). Anche in questo caso, a fronte di un minore investimento sulla comunicazione social registriamo una risposta differente da parte degli utenti che interagiscono con questi contenuti.

La Figura 1 presenta la distribuzione mensile di post e commenti nelle pagine dei leader politici inclusi nella nostra analisi. In prima battuta notiamo come nel caso di Giorgia Meloni la pubblicazione di contenuti su Facebook segua un trend piuttosto stabile. Certamente nelle prime fasi della nostra rilevazione che coincide esattamente con i primi mesi dell'emergenza COVID-19 e dunque con le misure di lockdown più rigide registriamo un numero maggiore di post Facebook pubblicati e in quel frangente anche un picco nei commenti. Tuttavia, non si evidenziano differenze rilevanti in termini di volumi di contenuti pubblicati lungo il periodo analizzato. Se invece spostiamo la nostra attenzione sul caso del leader leghista lo scenario muta. Salvini, infatti, mantiene un livello sempre molto alto di attività su Facebook ma con alcune marcate differenze. Durante tutta la prima fase dell'emergenza COVID-19 - come già si ricordava per Meloni - in concomitanza con la fase più acuta di pressione sulle strutture sanitarie e quindi le misure più drastiche di contenimento, Matteo Salvini risulta meno solerte nella pubblicazione di contenuti. Semmai, riscopre un certo attivismo a ridosso delle prime riaperture (giugno/luglio 2020), quando si intensifica il dibattito in seno al Consiglio Europeo a proposito del Recovery Fund, e soprattutto quando si acuisce il conflitto con il Governo Conte a proposito delle misure emergenziali. Di seguito, con l'inizio della seconda ondata di contagi il leader leghista sembra invece meno attivo e il numero di post, infatti, si riduce notevolmente. Questo pattern si consolida e conferma fino alla fine della nostra rilevazione che coincide peraltro con la crisi del Governo Conte e l'insediamento di Mario Draghi alla guida di un nuovo esecutivo che vede la partecipazione di tutti i partiti politici con la sola eccezione di Fratelli d'Italia. Il dato dei commenti segnala una tendenza piuttosto simile.

Per quel che attiene invece il caso di Nicola Zingaretti, il volume dei post pubblicati – che sappiamo essere molto inferiore rispetto ai due leader appena discussi – non fa registrare particolari differenze con la sola eccezione di alcuni picchi collocati proprio nelle fasi di mitigazione dei contagi nel giugno 2020 fino all'inizio della seconda ondata alla fine dell'estate. Silvio Berlusconi, dal canto suo, pur nel quadro di numeri limitati, mostra una più intensa attività di pubblicazione su Facebook nelle fasi iniziali della nostra rilevazione e in

particolare, dunque, nei primi mesi di emergenza pandemica, si osserva poi una significativa riduzione del numero di post pubblicati per poi registrare un nuovo incremento alla fine del 2020. La crisi del Governo Conte, nel suo caso, funge da *boost* per l'attività su Facebook. Vito Crimi, infine, mostra anche lui una maggiore attività nelle prime fasi della crisi pandemica, per poi ridurre drasticamente a partire dall'estate 2020 il suo investimento su Facebook.

Figura 1: Distribuzione mensile di post e commenti nelle pagine dei principali leader politici italiani (febbraio 2020 – febbraio 2021)



2.4 METODO E DESCRIZIONE DEL DIZIONARIO UTILIZZATO

Per analizzare il contenuto dei commenti ai post su Facebook, abbiamo utilizzato uno degli strumenti più semplici e al contempo più efficaci messi a nostra disposizione dall'analisi quantitativa dei testi: la classificazione delle parole presenti in diverse categorie predefinite di *incivility* e HS. Questa procedura consiste in due stadi: (1) la creazione di un dizionario contenente una lista di termini che assumiamo essere propri dell'HS o del linguaggio volgare, e (2) la ricerca all'interno del corpus dei commenti di tali termini.

La creazione del dizionario è ovviamente lo stadio più laborioso, dalla cui qualità dipende però l'esito dell'intera operazione. Con il termine "dizionario" si intende una collezione di parole o brevi locuzioni associate a diverse categorie. L'esempio più classico di utilizzo di un dizionario nell'analisi quantitativa di testi è la *sentiment analysis*, la tecnica attraverso la quale il livello di positività o negatività espresso da un testo viene ricavato dalla quantità di parole classificate come "positive" o "negative" presenti in esso. Nel nostro caso, per produrre un dizionario di HS abbiamo creato liste di parole partendo dalle diverse categorie di linguaggio d'odio così come definite dalla letteratura (Tabella 1) e da noi operazionalizzate (Tabella 5).⁶ La redazione della lista di parole è partita dal dizionario *Hurtlex* (Bassignana et al. 2018), una delle rare risorse per analisi di HS in lingua italiana. Da questo dizionario abbiamo eliminato alcune parole a nostro avviso cadute in disuso (ad es. "gagliofo", "marrano"), e aggiunto alcuni sinonimi delle parole rimaste. In una seconda fase, abbiamo utilizzato i dati messi a disposizione per *HaSpeeDe* (*Hate Speech Detection*), una competizione organizzata nel contesto della campagna *EVALITA 2018* organizzata dall'Associazione Italiana di Linguistica Computazionale ([qui e qui](#) per ulteriori informazioni). I dati *HaSpeeDe* consistono in una serie di post Facebook e tweet che sono stati classificati manualmente come contenenti HS o meno da due team di linguisti computazionali italiani (Del Vigna et al. 2017; Poletto et al. 2017). Usando tecniche di *machine learning*, abbiamo identificato le parole con il più alto valore discriminante (parole la cui presenza predice con maggiore successo la probabilità che un post o un tweet sia classificato come contenente HS), e utilizzato queste per aggiornare il dizionario. In una terza fase, abbiamo espanso il dizionario includendo termini da noi valutati come propri di linguaggio d'odio nelle diverse categorie. La divisione in categorie e sottocategorie risultante, con alcuni esempi per ognuna di esse, è riportata nella Tabella 7.

⁶ L'approccio basato sul dizionario è anche utilizzato nella ricerca di Rega e Marchetti (2021). Gli autori analizzano le reazioni dei cittadini ai post incivili dei leader su Facebook durante le elezioni generali italiane del 2018 e operazionalizzano l'inciviltà politica come un concetto multidimensionale composto da derisione/sarcasmo, travisamento esasperato, linguaggio insultante, diffamazione e discorsi intolleranti.

Tabella 7: *Categorie del dizionario di hate speech con esempi*

Categorie	Sottocategorie	Esempi	Numero di termini nel dizionario
<i>Incivility</i>		<i>cazzo, figa, stronzat*</i>	132 (20%)
Aggressività verbale			259 (40%)
	Insulto generico	<i>scemo, stronzo</i>	
	Insulto politico	<i>zecca, fascio, sinistroid*</i>	
	Linguaggio sprezzante	<i>maledett*, vergogn*, schifo</i>	
Aggressività fisica		<i>ti_ammazz*, muori, stai_zitt*</i>	97 (15%)
Discriminazione			162 (25%)
	<i>Body shaming</i>	<i>ciccion*, nano, tappo</i>	
	Abilismo	<i>spastic*, storpio, ritardat*</i>	
	Discriminazione etnica o religiosa	<i>negr*, terron*, zingaro</i>	
	Omofobia	<i>finocc*, froci*, ricchion*</i>	
	Sessismo	<i>puttana, cagn*, donnacc*</i>	

La tabella mostra come le quattro categorie generiche di HS si compongono da una o più sottocategorie più specifiche (per esempio la categoria Discriminazione è l'aggregato di diversi tipi specifici di discriminazione verbale, come il *body shaming*, la discriminazione etnica o religiosa, l'omofobia, ecc.). Per ognuna di queste sottocategorie, la terza colonna della tabella mostra alcuni esempi. Notare come il dizionario include parole vere e proprie oppure "temi" (tutti i termini che finiscono con l'asterisco), ovvero le parole private delle desinenze. Il vantaggio di utilizzare i temi, invece delle parole complete, è che in questo modo è possibile identificare tutte le possibili coniugazioni delle parole cercate. Inoltre, il dizionario contiene anche alcune brevi locuzioni, come ad esempio "*devi_morire*". Questo ci permette di identificare un insulto direzionato, con maggiore precisione che includendo semplicemente il termine "morire". In totale, il dizionario contiene 650 termini, di cui 55 si ripetono in più di una categoria (ad esempio il termine "*vaffanculo*" conta sia come *Incivility* che come Insulto Generico). La quarta colonna della tabella mostra il numero di termini in ogni categoria (e la percentuale rispetto al totale dei termini nel dizionario).

Per verificare la capacità del dizionario di riconoscere effettivamente il contenuto di HS dei commenti Facebook e dei tweet, abbiamo analizzato i commenti e i tweet presenti nei dati *HaSpeeDe*, confrontando la classificazione dei testi come "contenenti HS" o meno sulla base del dizionario (includendo anche termini classificati come *incivility*) con la classificazione manuale operata dai team che hanno prodotto i dati per *HaSpeeDe*. I risultati sono alquanto incoraggianti: nel caso dei commenti Facebook, l'*accuratezza* della classificazione dei testi come contenenti HS utilizzando il nostro metodo rispetto alla classificazione manuale è del 66%. Il che significa che in circa due terzi dei commenti il nostro dizionario identifica la presenza di HS tanto accuratamente quanto lo farebbe un essere umano. Notare che la *precisione* della classificazione, ovvero la percentuale di post che effettivamente contengono HS tra quelli classificati come tali dal dizionario, è del 94%. Questo implica che, nel caso dei commenti Facebook, il dizionario minimizza i falsi positivi, o in altre parole, previene che commenti che non contengono HS (stando alla classificazione di *HaSpeeDe*, vedi Del Vigna et al., 2017) vengano invece classificati come contenenti HS. Nel caso dei tweet l'*accuratezza* della predizione è superiore: 71%. Tuttavia, la *precisione* è inferiore: 64%. Questo indica che, per quanto riguarda la classificazione utilizzata da *HaSpeeDe* per i dati Twitter (vedi Poletto et al., 2017), il dizionario ha una maggiore sovrapposizione con la classificazione manuale rispetto ai dati Facebook, tuttavia non fa un lavoro altrettanto buono nel minimizzare i falsi positivi. Questo può essere dovuto a diversi fattori, *in primis* i differenti criteri utilizzati per classificare i commenti o i tweet come contenenti HS, che provengono infatti da studi diversi. In generale, notiamo che una tale sovrapposizione tra una classificazione ottenuta applicando un dizionario (basata quindi su singoli termini, senza considerare il significato delle frasi e altre caratteristiche del contesto linguistico) e il giudizio umano (basato su una comprensione del testo a tutto tondo) rimane un risultato incoraggiante, confermando che il nostro dizionario è un metodo valido, seppur imperfetto, di rilevazione di HS nei commenti degli utenti.

RISULTATI

3.1 IL LINGUAGGIO D'ODIO NEI DISCORSI DEI POLITICI SUI SOCIAL MEDIA

Prima di esaminare in dettaglio le caratteristiche delle conversazioni tra gli utenti, è utile avere un quadro generale dei messaggi pubblicati dai principali leader politici su Facebook. Più nello specifico si tratta di comprendere se anche nella comunicazione dei politici sia possibile riscontrare l'utilizzo di parole d'odio e soprattutto in quale misura. Con questa finalità, il dizionario descritto nel paragrafo precedente è stato applicato al contenuto testuale di tutti post pubblicati dai leader di FI, M5S, FdI, Lega e PD.

La Tabella 8 riporta quindi le percentuali di *incivility* e HS contenute nella comunicazione di ogni leader. I dati sono abbastanza chiari ed emergono due caratteristiche generali. In prima battuta Meloni (10,4%), Salvini (8,4%) e Crimi (7,6%) mostrano un livello cumulato di *incivility* e HS decisamente più alto di quello di Zingaretti (3,3%) e Berlusconi (0,7%). In secondo luogo, osservando la distribuzione tra le diverse categorie che compongono i discorsi d'odio, si registra una prevalenza di contenuti riferiti all'aggressività fisica. Dimensioni come volgarità, aggressività fisica e discriminazioni risultano praticamente assenti. Da un'ispezione dei dati emerge tuttavia che le poche occorrenze certificate nella maggior parte dei casi non dipendono da affermazioni del leader, ma derivano invece da dichiarazioni altrui o da eventi esterni riportati nel messaggio che l'analisi automatizzata del testo non è in grado di distinguere. Questo significa che l'HS o linguaggio volgare utilizzato dai leader è nella stragrande maggioranza dei casi dovuto al fatto che essi riportano commenti di terzi, spesso rivolti proprio a loro. A nostro avviso, questo non neutralizza il potenziale polarizzante del linguaggio d'odio: nonostante i leader tendano a mostrarsi come vittime di HS, il "colpevole" di tale linguaggio viene comunque spesso identificato e riportato, in quanto individuo o in quanto membro di una categoria avversaria (ad esempio un avversario politico o un suo sostenitore).

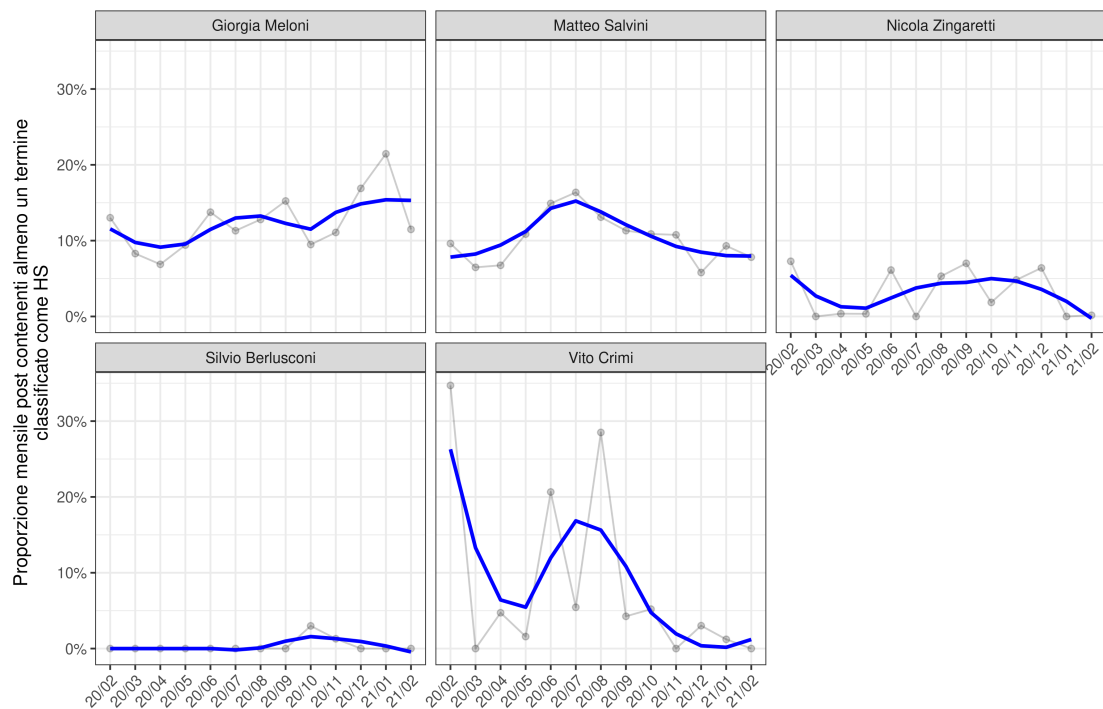
Tabella 8: *Incivility e HS nei post dei leader su Facebook*

	Giorgia Meloni (2.063)	Matteo Salvini (4.818)	Nicola Zingaretti (1.215)	Silvio Berlusconi (842)	Vito Crimi (475)
<i>Incivility</i>	1 (0%)	22 (0,5%)	0 (0%)	3 (0.4%)	1 (0,2%)
<i>HS</i>					
Aggressività verbale	197 (9,5%)	374 (7,8%)	22 (1,8%)	3 (0.4%)	25 (5,3%)
Aggressività fisica	6 (0,3%)	12 (0,5%)	0 (0%)	0 (0%)	2 (0,4%)
Linguaggio discrimi- natorio	4 (0,2%)	13 (0,4%)	1 (0,1%)	3 (0.4%)	0 (0%)
Totale HS	204 (9,9%)	389 (8,1%)	23 (1,9%)	6 (0.7%)	26 (5,5%)

Queste prime evidenze empiriche confermano alcune delle riflessioni già emerse nella discussione della letteratura. I partiti populistici sono generalmente considerati quelli più inclini ad adoperare una comunicazione che cerca (ed esaspera) la contrapposizione – anche aspra – coerentemente alla narrazione manichea della società (Bracciale, Andretta e Martella, 2021). Non stupisce quindi che Meloni, Salvini e Crimi – a capo di partiti considerati come populistici – si distinguano per livelli più elevati di aggressività verbale. La scarsa presenza di linguaggio volgare, aggressività fisica e linguaggio discriminatorio è invece facilmente spiegabile in virtù della carica pubblica e dalla visibilità mediatica dei leader politici. A prescindere dalle retoriche populiste, dall'enfasi su scontri e contrapposizioni, esistono dei confini normalmente invalicabili nella comunicazione di un personaggio pubblico che lo dissuadono dall'esprimersi in maniera troppo esplicita. Capita così che anche un politico come Matteo Salvini – che fa della sua lotta all'immigrazione uno dei propri capisaldi - si guardi bene dal pubblicare messaggi apertamente razzisti o discriminatori. Semmai è frequente il ricorso all'ironia. Ad esempio, nella passata legislatura i migranti – recuperando un termine utilizzato da Laura Boldrini – erano chiamati “risorse” con tono derisorio. In questo modo non solo si esprimeva biasimo sul fenomeno migratorio ma si rinfocolava l'astio nei confronti della ex presidente della Camera dei deputati. Anzi, proprio a questo proposito vale la pena ricordare che fra le modalità spesso utilizzate vi è anche l'utilizzo di una comunicazione che pur non indulgendo su termini e contenuti apertamente di HS è in grado di innescare vere e proprie *shitstorm* contro specifici target. In questi casi, un dizionario – e più in generale una tecnica di analisi testuale quantitativa – non

è in grado di cogliere la complessità del processo comunicativo e finisce con il sottostimare determinati aspetti dell'HS. Questa precisazione serve a ribadire che in larga misura i dati che presentiamo in questo report rappresentano la porzione più esplicita e visibile di un fenomeno che per sua natura è complesso ed eterogeneo, ma soprattutto ambiguo e sfuggente. In altre parole, il volume di HS sui social media è ben più alto di quanto sembri e soprattutto di quanto sia possibile misurarle.

Qualche ulteriore indicazione sulle differenze nel ricorso a parole d'odio emerge se si guarda l'evoluzione nel tempo (Figura 2). Escludendo il profilo di Silvio Berlusconi che presenta un andamento piuttosto basso e piatto – a confermare che l'ambiente social è in verità estraneo al leader di Forza Italia – si delineano due pattern in particolare. Da un lato, Crimi e Zingaretti presentano un trend piuttosto irregolare, caratterizzato da picchi verso l'alto e verso il basso. Questo indica un ricorso incidentale al dizionario dell'*incivility* o dell'HS, nel caso di Crimi anche nel contesto di un'attività social abbastanza limitata. Discorso opposto vale invece per Meloni e Salvini. Nel loro caso, infatti, il volume medio di HS e *incivility* si conserva costante nel tempo. Questo segnala un ricorso sistematico a questi contenuti, che si configurano dunque come un elemento costitutivo del discorso politico di questi due leader – e questo vale in particolare l'aggressività verbale. È interessante osservare come le due curve raggiungano i livelli più bassi durante il primo lockdown (marzo/aprile 2020), ma tendano poi a divergere: Meloni continua a crescere durante tutto l'anno, mentre Salvini a partire da agosto inizia a decrescere. Una spiegazione di questa divergenza – che non è possibile verificare empiricamente in questa sede – può essere dovuta alla diversa strategia politica dei due leader rispetto al governo e alla gestione pandemica. Infatti, mentre Meloni rivendica saldamente la propria opposizione al governo – mantenendo elevati i toni della sfida – Salvini si dimostra in questa fase più disponibile al compromesso e riduce le occasioni di scontro (non a caso la Lega sarà tra i sostenitori del Governo Draghi nel febbraio 2021, mentre Fratelli d'Italia resterà all'opposizione).

Figura 2: Volume di HS nei post dei leader politici su Facebook (febbraio 2020 - febbraio 2021)

3.2 INCIVILITY ED HS NEI COMMENTI DEGLI UTENTI

Come abbiamo avuto modo di osservare nel paragrafo 2 i profili Facebook dei leader analizzati si caratterizzano per un seguito, un volume di post pubblicati e un livello di interazione piuttosto diversificato. Chiaramente, queste differenze si riflettono anche sul numero totale di commenti, il cui volume varia da un massimo di circa 13.5 milioni per Salvini a un minimo di circa 155 mila per Crimi. Ancora una volta emergono due pattern: da un lato, Salvini e Meloni ottengono, in media, più di 2.500 commenti a post; dall'altro lato, Zingaretti, Berlusconi e Crimi si attestano a livelli quasi 10 volte più bassi, intorno a 300 commenti a post (vedi Tabella 6). Si confermano da questo punto di vista (e ancora una volta) le osservazioni già emerse in letteratura: leader e partiti populistici sono caratterizzati da maggiore attivismo e successo sui social (Engesser et al. 2017; Ernst et al. 2017; 2019).

In questo paragrafo focalizziamo la nostra attenzione sulle interazioni de-

gli utenti e più specificamente ci interessiamo ai commenti. Benché – come segnalato sopra – per alcuni leader la componente di *incivility* e HS sia una caratteristica sistemica della conversazioni online, il loro livello non è semplicemente una funzione del volume dei commenti. La percentuale complessiva di parole d’odio nei commenti evidenzia infatti differenze molto marcate, simili a quelle già riscontrate all’interno dei messaggi dei leader (Tabella 9). In questo frangente, le pagine di Berlusconi e Zingaretti si confermano le meno coinvolte dal fenomeno HS, infatti i contenuti che presentano elementi di *incivility*/HS riguardano, rispettivamente, il 3,9% e 6,2% dei commenti ai loro post. La situazione è decisamente differente se consideriamo invece i commenti ai post di Salvini, Meloni, come pure a quelli di Crimi. I dati, infatti, ci mostrano una presenza di *incivility*/HS all’interno dei commenti di quasi tre volte superiori rispetto a quanto osservato per Berlusconi e Zingaretti. Mentre per il leader ad interim del M5S il numero assoluto di messaggi contenenti parole d’odio è limitato (vale a dire meno di 27 mila commenti in un anno), il dato è decisamente più rilevante per i leader di FdI e Lega. In questi due casi, il numero assoluto di commenti è particolarmente elevato. Per Meloni, infatti, in oltre 800.000 commenti ai suoi post si registra la presenza di *incivility* od HS. Se consideriamo Salvini, invece, raggiungiamo addirittura oltre 1.800.000 commenti con contenuti riconducibili a *incivility*, aggressività fisica o verbale e discriminazione. Questi numeri rappresentano un volume considerevole, vale a dire una massa critica di contenuti riferiti a *incivility* o HS a cui risulta esposta la maggior parte degli utenti che interagisce su quelle pagine. Se poi proviamo a scendere nel dettaglio e considerare più precisamente la composizione di queste conversazioni riscontriamo pattern simili a quanto già emerso nell’analisi della comunicazione dei leader. L’aggressività verbale è la dimensione più presente, seguita dal linguaggio volgare mentre le altre dimensioni risultano più residuali. La community che ruota intorno alla pagina di Crimi è quella che evidenzia le percentuali più elevate di espressioni incivili (7%) e di insulti orientati verso target specifici (16,5%), segue la community di Meloni con 2,9% e 13,6%, infine quella di Salvini con 3,3% e 12,4%. In generale, quindi, le due categorie che nella nostra operazionalizzazione presentano una minore intensità d’odio sono quelle più presenti all’interno delle conversazioni caratterizzati da toni incivili e HS. Se la presenza di discorsi d’odio online è un elemento che non va sottovalutato perché "tossico" per il funzionamento delle democrazie e perché tra i tratti costitutivi del "disordine informativo" (Waisbord 2018, 2020; Wardle e Derakhshan, 2017), il fatto che l’aggressività verbale e la discriminazione – elementi a più alta intensità d’odio – siano utilizzati sporadicamente suggerisce che il livello di "inquinamento" dell’ecosistema politico-comunicativo sui social

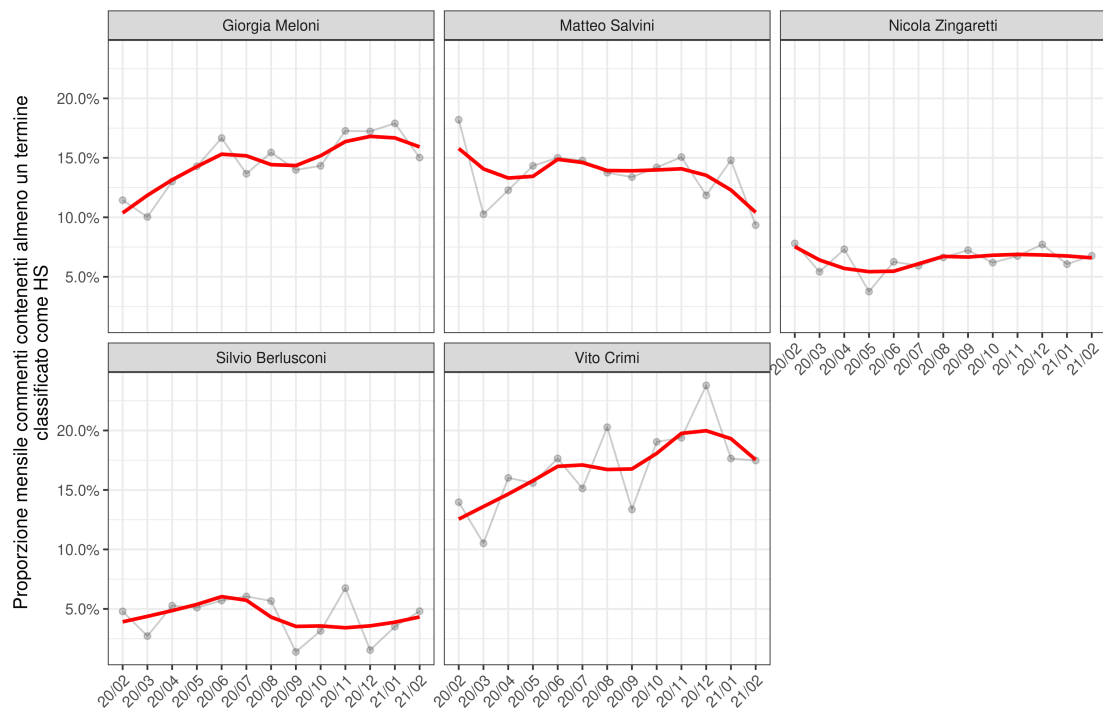
media sia ancora in una fase iniziale.

Tabella 9: *Incivility e HS nei commenti ai post dei leader su Facebook*

	Giorgia Meloni (5.477.286)	Matteo Salvini (13.556.706)	Nicola Zingaretti (319.174)	Silvio Berlusconi (246.268)	Vito Crimi (154.010)
<i>Incivility</i>	161.092 (2,9%)	445.101 (3,3%)	4.154 (1,3%)	1.077 (0,4%)	10.745 (7%)
<i>HS</i>					
Aggressività verbale	742.433 (13,6%)	1.679.464 (12,4%)	17.579 (5,5%)	8.917 (3,6%)	25.410 (16,5%)
Aggressività fisica	23.668 (0,4%)	38.054 (0,3%)	666 (0,2%)	288 (0,1%)	497 (0,3%)
Linguaggio discriminatorio	53.455 (1%)	149.439 (1,1%)	1.921 (0,6%)	601 (0,2%)	1.111 (0,7%)
Totale HS	804.059 (14,7%)	1.834.498 (13,5%)	19.805 (6,2%)	9.691 (3,9%)	26.507 (17,2%)

La Figura 3 consente invece di riflettere sull'andamento nel tempo di *incivility* ed HS all'interno dei commenti ai post pubblicate sulle pagine pubbliche dei cinque leader monitorati. Per quel che riguarda i commenti ai post di Berlusconi e Zingaretti le curve mostrano uno sviluppo sostanzialmente piatto anche in ragione di un basso numero di commenti. La questione è decisamente diversa per gli altri leader. In particolare, le *communities* di Crimi e Meloni dopo una breve fase di *rally round the flag* (Kritzinger et al. 2021), coincisa con il periodo di lockdown a marzo e aprile 2020, fanno registrare una decisa crescita nei livelli di *incivility*/HS all'interno dei commenti ai post dei due leader. Opposto è invece il trend della community di Salvini. In questo caso, infatti, il volume di commenti che contiene elementi di *incivility*/HS decresce in maniera significativa e costante a partire da febbraio 2020. Difficile dire quali siano le spiegazioni di questi andamenti senza potersi avvalere di un'analisi del contenuto di post e commenti in grado di chiarire quali temi fossero affrontati. Ci si può tuttavia interrogare se esista una correlazione tra la presenza di *incivility*/HS nei post dei leader e nei commenti degli utenti. Il prossimo paragrafo prova a rispondere a questa domanda.

Figura 3: Incivility e HS nei commenti ai post dei leader su Facebook



3.3 L'INTERAZIONE TRA LEADER E UTENTI SU FACEBOOK: L'ODIO ALIMENTA ODIO?

In questo paragrafo ci interessa indagare se e in quale misura e la pubblicazione di messaggi contenenti elementi di *incivility*/HS produca degli effetti nelle reazioni ai post. Si tratta di chiarire se contenuti di natura volgare o che enfatizzano discriminazione, aggressività verbale o fisica pubblicati dalle pagine dei leader determinino una maggiore presenza di *incivility*/HS anche nei commenti degli utenti.

Prima ancora di verificare se effettivamente l'odio alimenti a sua volta l'odio, l'impatto di *incivility*/HS può essere osservato anche in termini di engagement, verificando se il linguaggio d'odio nelle sue diverse declinazioni crei maggiori livelli di engagement da parte degli utenti in termini di *like*, commenti o condivisioni.

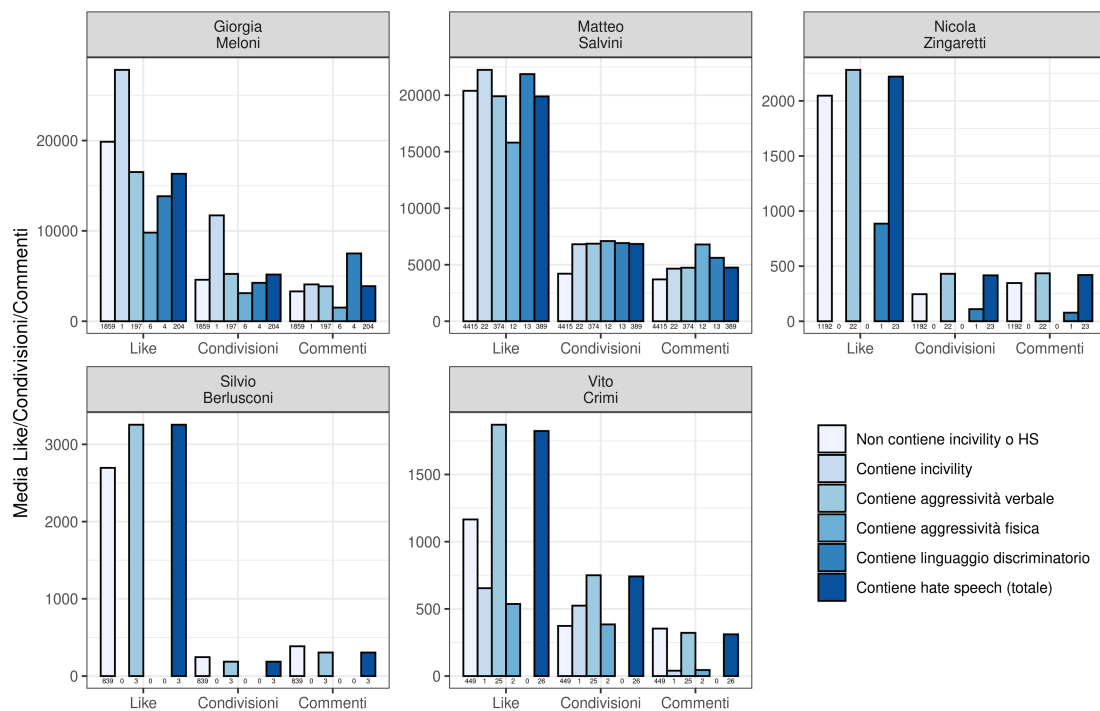
La Figura 4 riporta esattamente questo aspetto, dettagliando per ogni leader il livello medio di engagement riferito ai diversi contenuti di *incivility*/HS pubblicati dai leader politici.

Come abbiamo già avuto modo di sottolineare, Silvio Berlusconi e Nicola Zingaretti sono i due leader che meno utilizzano questo tipo di repertorio verbale all'interno della loro produzione comunicativa su Facebook. Non sorprende, dunque, che anche sul fronte dell'engagement rileviamo un livello sensibilmente minore rispetto agli altri leader come ad esempio Salvini o Meloni. In ogni caso il basso volume di contenuti di *incivility*/HS pubblicati sulle pagine di Berlusconi e Zingaretti impedisce di derivare qualsiasi conclusione su questo aspetto. Peraltro, se ci concentriamo sul caso di Berlusconi, ad esempio, notiamo come in media i post che non presentano alcun elemento d'odio ottengono un numero maggiore di *like* rispetto a quelli che invece contengono *incivility*/HS. E se anche consideriamo il caso di Zingaretti la differenza di *like* ottenuti in media dai post che contengono elementi d'odio rispetto ai post neutri è sostanzialmente trascurabile. Il tutto – si ribadisce a scampo di equivoci – in un contesto dove per questi due leader la presenza di parole riconducibili a volgarità, aggressività fisica o verbale o addirittura discriminazione è del tutto marginale, proprio perché ridotta a qualche unità. In ogni caso, anche le loro *communities* sembrano essere poco reattive su questo fronte. Uno scenario piuttosto simile emerge anche considerando i dati relativi a Vito Crimi - capo politico ad interim del M5S. Anche nel suo caso il volume di contenuti d'odio è piuttosto limitato e ancor più contenute sono le reazioni di engagement dei suoi utenti. Tuttavia, pur in un quadro di limitato ricorso a elementi di *incivility*/HS, registriamo che quei pochi post che presentano contenuti di *incivility*/HS – e all'interno di questi sono in particolare quei messaggi che includono elementi di aggressività verbale – ottengono in media un numero superiore di *like* rispetto a quei post che invece sono privi di linguaggio d'odio.

Spostando la nostra attenzione su Matteo Salvini e Giorgia Meloni, invece, notiamo delle differenze rilevanti. In prima battuta, ma lo sapevamo dai paragrafi precedenti, nel loro caso il volume di messaggi che contengono discorsi di natura volgare, discriminatoria, di aggressività fisica o verbale è superiore rispetto agli altri leader. Soprattutto, questo maggiore ricorso a linguaggio d'odio corrisponde a livelli di engagement decisamente più elevati. Se ci concentriamo sul leader leghista, per esempio, vediamo che il numero medio di *like* si attesta attorno a 20.000 con un livello lievemente superiore per quei post che contengono linguaggio volgare, ma senza particolari differenze

rispetto a quei post FB che invece non riportano elementi di *incivility*/HS. Ma se consideriamo invece il dato relativo alle condivisioni ci rendiamo conto che in generale quei messaggi che presentano contenuti riconducibili a volgarità, aggressività verbale o fisica o anche discriminazione registrano un numero medio di condivisioni maggiore rispetto agli altri post. Questo pattern – ma in misura decisamente inferiore – lo riscontriamo anche per quel che riguarda i commenti. Ad ogni modo, il dato è significativo perché implica che contenuti di *incivility*/HS hanno un potenziale di propagazione e pervasività piuttosto rilevante. Un quadro analogo si rintraccia considerando i dati relativi a Giorgia Meloni. In questo caso, però, il volume di messaggi contenenti discorsi d'odio prodotti dalla leader di Fratelli d'Italia è inferiore rispetto a quanto osservato per Salvini. Tuttavia, se ragioniamo in termini di engagement la reazione della community di Meloni è piuttosto simile. Nel suo caso, i messaggi privi di linguaggio d'odio ottengono in media un numero superiore di *like* rispetto a quei post che invece riportano elementi controversi. Mentre per quel che riguarda le condivisioni e i commenti non emergono differenze degne di nota.

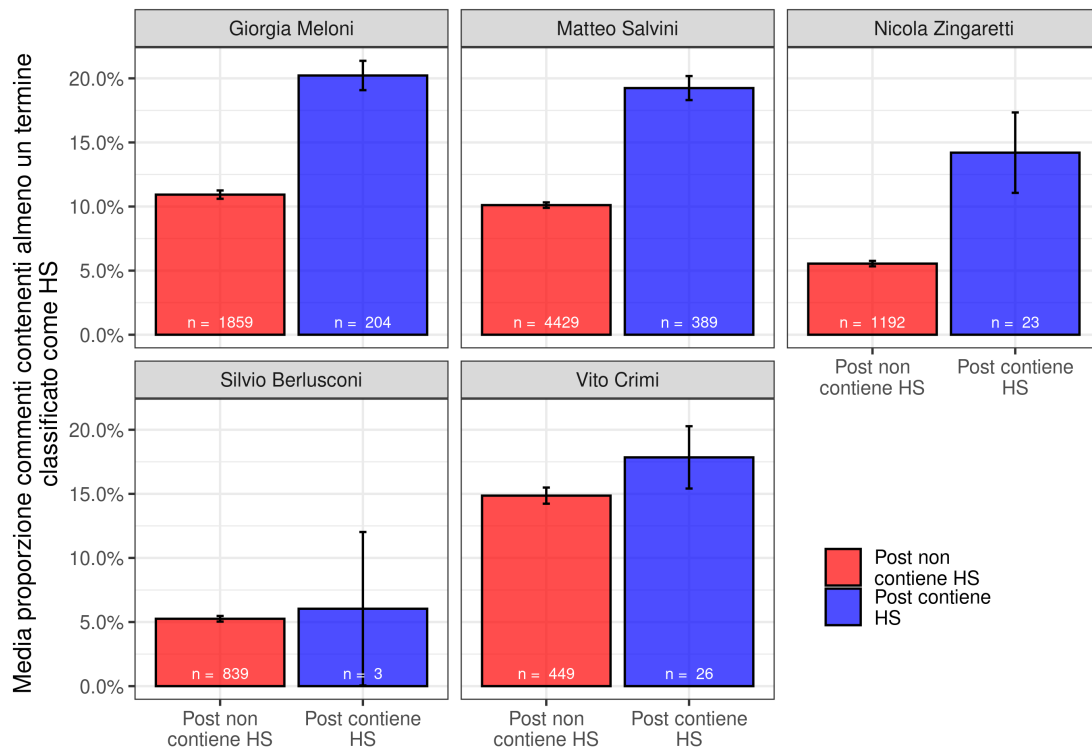
Figura 4: Engagement medio dei post e volume di *incivility*/HS nella comunicazione dei leader politici su Facebook (febbraio 2020 - febbraio 2021)



Come si segnalava sopra, oltre alla dimensione dell'engagement e quindi al potenziale di circolazione e pervasività dei post pubblicati dai leader, l'interazione con le community può essere osservata anche considerato se e in quale misura la presenza di contenuti di *incivility*/HS si rifletta in una maggiore presenza nei commenti di linguaggio discriminatorio, di aggressività fisica o verbale, o volgarità. La Figura 5 restituisce esattamente questo dato consentendo di valutare la corrispondenza fra post e commenti in termini di linguaggio contenente HS. In questa analisi ignoriamo il contenuto di *incivility* in quanto molto marginale nella comunicazione dei leader, se non addirittura assente, come nel caso di Berlusconi e Zingaretti. Il dato generale è piuttosto eloquente. Per Berlusconi, ad esempio, la porzione di commenti che presenta elementi riconducibili a discriminazione, aggressività verbale o fisica non evidenzia differenze particolari legate al contenuto dei post Facebook. Tuttavia, si tratta solo di tre post. Il caso di Nicola Zingaretti è lievemente differente perché effettivamente nel caso dei pochi post contenenti elementi di HS la percentuale di commenti che riflettono questi contenuti passa dal 5% al 10%. In ogni caso, bisogna ricordare che anche qui la quota di tali post è molto limitata (appena 23). I dati relativi a Vito Crimi sono interessanti perché seppure non mostrino grosse differenze fra i commenti ai post che presentano HS e quelli che invece non hanno di questi contenuti, osserviamo che a prescindere dal contenuto dei post pubblicati suoi utenti tendono a commentare ricorrendo a HS (parliamo del 15% dei commenti).

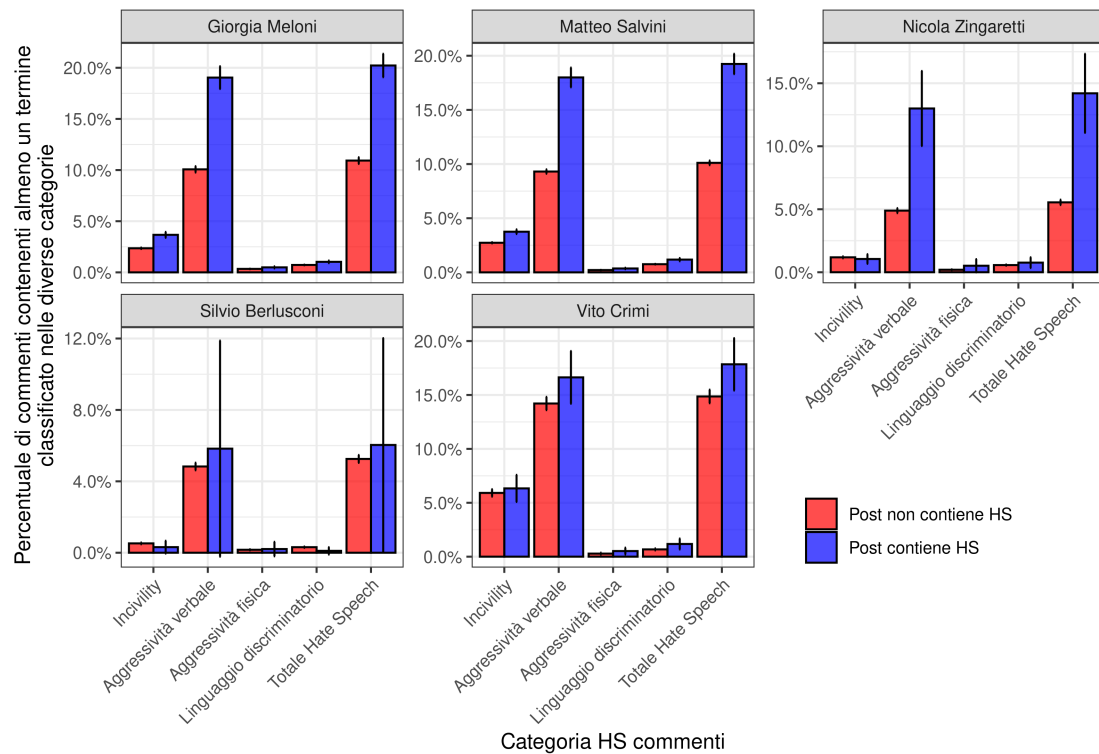
Quando consideriamo invece i dati relativi alla leader di Fratelli d'Italia e al leader leghista lo scenario cambia drasticamente. Non si tratta solo di un significativo incremento di post che riportano contenuti controversi dal punto di vista dei messaggi d'odio. Infatti, per entrambi i leader quando vengono pubblicati messaggi Facebook che presentano discriminazione o aggressività, circa un commento su cinque a sua volta riproduce lo stesso registro linguistico, fungendo dunque da cassa di risonanza per l'HS.

Figura 5: Corrispondenza/correlazione tra post dei leader contenenti HS e presenza di HS nei commenti su Facebook



La Figura 6 va ulteriormente a fondo di questo tema dettagliando la relazione fra HS contenuta nei post pubblicati dai leader politici e le diverse declinazioni di questo linguaggio d'odio presente nei commenti, questa volta includendo anche *incivility*. In questa analisi, tralasciamo il caso di Berlusconi che nella sua produzione Facebook appare decisamente poco incline a frequentare questi registri e linguaggi, e il caso di Crimi che, come abbiamo avuto modo di vedere sopra, fa osservare una porzione piuttosto rilevante di commenti contenenti HS a prescindere dal contenuto effettivo dei suoi post. Questo vale anche per *l'incivility*. Invece, osserviamo come per tutti gli altri tre leader è soprattutto l'aggressività verbale a popolare i commenti ai post. E questo vale anche per Nicola Zingaretti che pure presenta volumi di HS più bassi rispetto agli altri due leader. Nel caso di Meloni e Salvini arriva a sfiorare il 20%. Al contrario, gli altri tipi di *incivility*/HS sembrano essere meno in grado di aizzare analogo linguaggio d'odio nei commenti.

Figura 6: Corrispondenza/correlazione tra post dei leader contenenti HS e tipo di incivility/HS nei commenti su Facebook



Possiamo dunque concludere che "l'odio alimenta odio" nel linguaggio online? In parte, e principalmente per quanto riguarda i leader dei partiti maggiori (Salvini, Meloni, Zingaretti), ovvero coloro che sono più attivi e hanno maggiore seguito su Facebook. Il che non significa che altre *community* siano immuni dal fenomeno: come dimostra il caso di Crimi, anche se il livello medio di *incivility* e HS tra i commenti è molto alto, esso non necessariamente rispecchia in modo significativo il contenuto del post.

CONCLUSIONI

Questo studio si interroga sul fenomeno del linguaggio d'odio (*hate speech*) in Italia, ovvero sulla sua portata, sulle dimensioni che lo compongono e su un eventuale ricorso differenziato a questi registri da parte dei leader politici e dei cittadini che interagiscono con loro sui social media. La letteratura scientifica esistente propone un quadro molto complesso e sfaccettato del fenomeno distinguendo grosso modo *incivility* (utilizzo di linguaggio volgare o che viola norme di educazione condivise) da *hate speech* (utilizzo di insulti o minacce direzionate a individui o gruppi). Lo scopo di questa ricerca era di proporre una definizione univoca per entrambi questi tipi di linguaggio, e derivarne un metodo di quantificazione da utilizzare in analisi quantitative dei testi. Questo strumento, attualmente il più ampio e dettagliato dizionario di *incivility* e *hate speech* per la lingua italiana, è stato poi applicato alla quantificazione di linguaggio volgare e di odio tra i post ($n=9.413$) e relativi commenti ($n=19.753.444$) pubblicati sulle pagine Facebook dei leader dei principali partiti politici italiani (Giorgia Meloni, Matteo Salvini, Nicola Zingaretti, Silvio Berlusconi e Vito Crimi) nel periodo febbraio 2020-febbraio 2021.

I risultati delle analisi mostrano alcune tendenze rilevanti. Innanzi tutto, esiste un certo quantitativo di *hate speech* nel linguaggio utilizzato dai politici nei loro post, particolarmente significativo per Meloni (10% dei post), Salvini (8%), e in misura minore Crimi (5,5%), e trascurabile per Berlusconi (0,7%) e Zingaretti (1,9%). È importante, tuttavia, sottolineare che nella stragrande maggioranza dei casi tali parole vengono utilizzate per riportare commenti di terzi rivolti ai leader stessi. Per quanto riguarda i commenti ai post, l'utilizzo di linguaggio d'odio vede gli stessi tre leader nelle prime posizioni, ma con un ordine differente: la community di Crimi fa ricorso a parole d'odio nel 17,2% dei commenti, mentre quella di Meloni (14,7%) e Salvini (13,5%) seguono poco sotto. Gli stessi registri sono presenti nelle conversazioni sulle pagine di Zingaretti (6,2%) e Berlusconi (3,9%) ma con volumi decisamente più bassi. In questo caso, si tratta principalmente di parole d'odio rivolte al leader, ad altri commentatori o a persone terze. La presenza di *incivility* e *hate speech* nei post di Salvini e Meloni è inoltre correlata ad un livello medio più elevato di *like*, condivisioni e commenti rispetto a post che non contengono questo linguaggio. Questo engagement facilita la visibilità e la circolazione di questo genere di

messaggi tra coloro che seguono il politico. Infine, i risultati mostrano che c'è una corrispondenza fra post e commenti in termini di linguaggio contenente *hate speech*, che è particolarmente significativa per Meloni e Salvini. Quando i due leader pubblicano post che presentano elementi di discriminazione, aggressività verbale o fisica, circa un commento su cinque a sua volta riproduce lo stesso contenuto fungendo da cassa di risonanza per l'*hate speech*. Questo risultato è particolarmente importante in quanto suggerisce che, per quanto riguarda il confronto verbale sui *social media* in Italia, l'odio alimenta odio.

L'insieme di questi dati conferma la rilevanza dell'*incivility* e, soprattutto, dell'*hate speech* all'interno del dibattito pubblico e politico in Italia. Questo report intende dunque anche sottolineare la necessità di una seria riflessione sulla qualità del discorso pubblico e politico. Le diverse forme e declinazioni dei discorsi d'odio sono considerabili come vere e proprie degenerazioni del dialogo civile e democratico. La diffusione dell'HS sui social è preoccupante anche perché vi è la possibilità che esso riverberi nei contesti offline, impattando catastroficamente sulla socializzazione politica della società italiana. L'odio espresso online è un fenomeno endemico che va monitorato e contrastato, intervenendo in maniera efficace al fine di evitare un'ulteriore polarizzazione della contesa politica. In questo contesto, le istituzioni politiche e il giornalismo possono e devono necessariamente giocare un ruolo chiave nel gestire e disinnescare questi serbatoi d'odio, ma ci riusciranno soltanto se consapevoli del problema e solamente se opportunamente supportati da evidenze empiriche adeguate. Questo blue paper intende fornire un contributo preliminare ma coerente al problema, ponendosi come obiettivo quello di offrire a cadenza annuale un report dell'andamento dell'*incivility* e dell'*hate speech* nelle conversazioni sui *social media* in Italia.

RIFERIMENTI BIBLIOGRAFICI

- Alkiviadou, N. (2019). Hate speech on social media networks: towards a regulatory framework?. *Information & Communications Technology Law*, 28(1), 19-35.
- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., e Ladwig, P. (2014). The “nasty effect”: Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, 19, 373–387.
- Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., e Xenos, M. A. (2018). Toxic talk: How online incivility can undermine perceptions of media. *International Journal of Public Opinion Research*, 30(1), 156–168.
- Andrisani, P. (2014). Il perverso intreccio tra odio virtuale e odio «virale», in Lunaria (a cura di) *Cronache di ordinario razzismo. Terzo libro bianco sul razzismo in Italia*. Roma.
- Assimakopoulos, S., Baider, F. H., e Millar, S. (2017). *Online hate speech in the European Union: a discourse-analytic perspective*. London: Springer.
- Bassignana, E., Basile, V., e Patti, V. (2018). Hurltlex: A Multilingual Lexicon of Words to Hurt. *Proceedings of the Fifth Italian Conference on Computational Linguistics, CLiC-It 2018*. 5th Italian Conference on Computational Linguistics, CLiC-it 2018, Turin, Italy
- Bentivegna, S. e Rega, R. (2020) I discorsi d’odio online in una prospettiva comunicativa: un’agenda per la ricerca. *Mediascapes journal*, 16: 151-171.
- Borah, P. (2014). Interaction of incivility and news frames in the political blogosphere: consequences and psychological mechanisms. In *Handbook of Research on Political Activism in the Information Age* (pp. 407-424). IGI Global.
- Bracciale, R., Andretta, M. e Martella, A. (2021) Does populism go viral? How Italian leaders engage citizens through social media. *Information, Communication & Society*, <https://doi.org/10.1080/1369118X.2021.1874472>

- Brooks, D. J. e Geer, J. G. (2007) Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science*, 51(1), 1-16.
- Brown, A. (2018). What is so special about online (as compared to offline) hate speech?. *Ethnicities*, 18(3), 297-326.
- Burnap, P. e Williams, M. L. (2014) *Hate speech, machine classification and statistical modelling of information flows on Twitter: interpretation and communication for policy decision making*. Presented at: Internet, Policy & Politics, Oxford, UK, 26 September 2014.
- Chadwick, A. (2013) *The Hybrid Media System: Politics and Power*. Oxford: Oxford University Press.
- Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. London: Palgrave Macmillan.
- Chen, G. M., e Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, 61(1), 108–125.
- Coe, K., Kenski, K., e Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679.
- Davidson, T., Warmesley, D., Macy, M., e Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 11, No. 1).
- Del Vigna, F., Cimino, A., Dell'Orletta, F., Petrocchi, M., e Tesconi, M. (2017). Hate me, hate me not: Hate speech detection on Facebook. *Proceedings of the First Italian Conference on Cybersecurity (ITASEC17)*. First Italian Conference on Cybersecurity (ITASEC17), Venice, Italy.
- Engesser, S., Ernst, N., Esser, F. e Büchel, F. (2017) Populism and social media: How politicians spread a fragmented ideology. *Information, Communication & Society*, 20(8): 1109-1126.
- Engesser, S., Fawzi, N. e Larsson, A. O. (2017) Populist online communication. *Information, Communication & Society*, 20(9): 1279-1292.
- Ernst, N., Engesser, S.; Esser, F. [2017] Switzerland: Favourable conditions for

- growing populism. In T. Aalberg, F. Esser, C. Reinemann, J. Strömbäck e C.H. de Vreese (a cura di) *Populist political communication in Europe*. New York: Routledge (151-164).
- Ernst, N., Blassnig, S., Engesser, S., Büchel, F. e Esser, F. [2019] Populists Prefer Social Media Over Talk Shows: An Analysis of Populist Messages and Stylistic Elements Across Six Countries. *Social Media + Society*. <https://doi.org/10.1177/2056305118823358>
- Faulker, M.E. (2009) Anti-Lesbian, Gay, Bisexual, and Transgendered Victimization in Canada and the United States: A Comparative Study. In *Hate Crime: The Victims of Hate Crime* (a cura di B. Perry), Vol 3, Praeger, Westport.
- Fortuna, P., Soler, J., e Wanner, L. (2020). Toxic, hateful, offensive or abusive? what are we really classifying? an empirical analysis of hate speech datasets. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 6786-6794).
- Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, 12(2), 167–185.
- Halpern, D. e Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, 29(3), 1159–1168.
- Hameleers, M., van der Meer, T., e Vliegthart, R. (2021). Civilized truths, hateful lies? Incivility and hate speech in false information–evidence from fact-checked statements in the US. *Information, Communication & Society*, 1-18.
- Hille, S. e Bakker, P. (2013) I Like News. Searching for the ‘Holy Grail’ of Social Media: The Use of Facebook by Dutch News Media and Their Audiences. *European Journal of Communication* 28(6): 663–680.
- Hopp, T. e Vargo, C. J. (2017). Does negative campaign advertising stimulate uncivil communication on social media? Measuring audience response using big data. *Computers in human behavior*, 68, 368-377.
- Jacobs, J. B. e Potter, K. (2000). *Hate crimes: Criminal Law and Identity Politics*. Oxford: Oxford University Press.

- Jamieson, K. (1997). *Civility in the House of Representatives*. APPC report 10. Retrieved May 26, 2021 <https://cdn.annenbergpublicpolicycenter.org/Downloads/Civility/0ld%20reports/Civility%20report%20March%201998.pdf>
- Karpf, D. (2010) Online Political Mobilization from the Advocacy Group's Perspective: Looking Beyond Clicktivism. *Policy & Internet* 2(4): 7-41.
- Kormelink, T. G. e Meijer, I. C. (2018) What Clicks Actually Mean: Exploring Digital News User Practices. *Journalism* 19(5): 668-683.
- Kritzinger, S., Foucault, M., Lachat, R., Partheymüller, J., Plescia, C. e Brouard, S. (2021) 'Rally round the flag': the COVID-19 crisis and trust in the national government. *West European Politics*, 44(5-6): 1205-1231.
- Krumsvik, A. H. (2018) Redefining User Involvement in Digital News Media. *Journalism Practice* 12(1): 19-31.
- Larsson, A. O. (2018) I Shared the News Today, Oh Boy. *Journalism Studies* 19(1): 43-61.
- Martin, R. C., Coyier, K. R., VanSistine, L. R. e Schroeder, K. L. (2013). Anger on the Internet: the Perceived Value of Rant-Sites. *Cyberpsychology, behavior and social networking*, 16(2):119- 22.
- Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., e Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media+ Society*, 5(3), <https://doi.org/10.1177%2F2056305119862641>
- Morozov, E. (2011) *The Net Delusion: The Dark Side of Internet Freedom*. New York: PublicAffairs.
- Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication*, 11, 21.
- Munger, K. (2016). "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior*, 39: 629- 49.
- Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton, NJ: Princeton University Press.
- Mutz, D. C., e Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, 1-15.

- Oz, M., Zheng, P., e Chen, G. M. (2017). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society*, 20(9), 3400–3419. <https://doi.org/10.1177/1461444817749516>
- Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New media & society*, 6(2), 259-283.
- Hermida, A. (2011) Fluid Spaces, Fluid Journalism: The Role of the Active Recipient. In *Participatory Journalism: Guarding Open Gates at Online Newspapers*, edited by J. B. Singer, A. Hermida, D. Domingo, A. Heinonen, S. Paulussen, T. Quandt, Z. Reich e M. Vujnovic, 177–191. New York: Wiley-Blackwell.
- Picone, I. (2015) Grasping the Digital News User. *Digital Journalism* 4(1): 125–141.
- Poletto, F., Basile, V., Sanguinetti, M., Bosco, C., e Patti, V. (2020). Resources and benchmark corpora for hate speech detection: a systematic review. *Language Resources and Evaluation*, 1-47.
- Rega, R. e Marchetti, R. (2021) The strategic use of incivility in contemporary politics. The case of the 2018 Italian general election on Facebook. *The Communication Review*, <https://doi.org/10.1080/10714421.2021.1938464>
- Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., e Wojatzki, M. (2017). Measuring the reliability of hate speech annotations: The case of the european refugee crisis. *arXiv preprint arXiv:1701.08118*.
- Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research. Advance online publication*. <https://doi.org/10.1177/0093650220921314>
- Rossini, P. (2021). More Than Just Shouting? Distinguishing Interpersonal-Directed and Elite-Directed Incivility in Online Political Talk. *Social Media+ Society*, 7(2), <https://doi.org/10.1177%2F20563051211008827>
- Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, communication & society*, 18(2), 121-138.
- Sacristán, A. (a cura di). 2013. *Sociedad del conocimiento, tecnología y educación*. Madrid: Morata.

- Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, 8(1), 18–33.
- Singer, J. (2014) User-generated Visibility: Secondary Gatekeeping in a Shared Media Space. *New Media & Society* 16(1): 55–73.
- Sobieraj, S., e Berry, J. M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, 28(1), 19-41.
- Su, H.-P., Huang, C.-J., Hao-Tsung Chang, H.-T e Lin, C.-J. (2017) Rephrasing Profanity in Chinese Text. In *Proceedings of the Workshop Workshop on Abusive Language Online (ALW)*. Vancouver, Canada.
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior*, 7(3), 321-326.
- Sydnor, E. (2018). Platforms for incivility: Examining perceptions across different media formats. *Political Communication*, 35(1), 97–116.
- Theocharis, Y., Barberá, P., Fazekas, Z., e Popa, S. A. (2020). The dynamics of political incivility on twitter. *Sage Open*, 10(2), [https://doi.org/10.1177%2F2158244020919447](https://doi.org/10.1177/2F2158244020919447).
- Waisbord, S. (2018) Why Populism is Troubling for Democratic Communication. *Communication, Culture and Critique*, 11(1): 21-34.
- Waisbord, S. (2020) Mob Censorship: Online Harassment of US Journalists in Times of Digital Hate and Populism. *Digital Journalism*, 8:8, 1030-1046.
- Walker, S. (1994). *Hate speech: The history of an American controversy*. University of Nebraska Press.
- Wardle, C e Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe Report. Disponibile online <https://rm.coe.int/information-disorder-toward-an%2Dinterdisciplinary-framework-for-researc/168076277c>
- Ziccardi G. (2016). *L'odio online. Violenza verbale e ossessioni in rete*. Milano: Raffaello Cortina.