

Il disordine informativo e l'odio in rete. Democrazia a rischio

Emiliana De Blasio, Luiss, Roma
Michele Sorice, Luiss, Roma

Information disorder and online hatred. Democracy at risk. *Hate speech is very present in digital communicative ecosystems; although it does not originate in online communication, certainly the speed and pervasiveness of social media and digital platforms have played an important role in the processes of establishing such communicative practices. However, it is impossible to understand hate speech without placing it in the broader framework of information disorder and its relationship with the emergence of the platform society.*

This essay relates online hate speech to the practices of information disorder, which, moreover, are increasingly characterised as systemic and functional to the fragmentation of the public sphere. The essay also illustrates how manipulative dynamics not only feed the information disorder but tend to transform it into a new disinformation order, moreover functional to the information system at the time of neoliberal rationality. This is also why research on hate speech and the fight against online hatred become indispensable elements in the defence of democracy.

Keywords: democracy, information disorder, digital ecosystem, hate speech, neoliberalism, platform society, public sphere.

I discorsi d'odio costituiscono elementi sempre più presenti negli ecosistemi comunicativi digitali e sebbene essi non siano stati “inventati” dalla comunicazione in rete, sicuramente la velocità e la pervasività dei social media e delle piattaforme digitali hanno giocato un ruolo importante nei processi di affermazione di tali pratiche comunicative. Non si può, tuttavia, comprendere l'*hate speech* senza collocarlo nella più ampia cornice del disordine informativo (di cui costituisce una delle componenti) e senza individuare le connessioni fra odio e costruzione del nemico.

1. La menzogna e la costruzione del nemico

Le retoriche di costruzione del nemico – nelle società fortemente mediatizzate della modernità, spesso sovrapposte a quelle dell'antipolitica – sono per lo più gridate, emotive, altamente spettacolari nella loro capacità di convocare passioni, prima che elementi di progettualità razionale. Insomma, l'ideale per l'arena mediatica, sia per quella televisiva che proprio nelle dinamiche di spettacolarizzazione ha trovato uno dei suoi strutturali punti di forza, sia per la

complessa sfera degli ecosistemi comunicativi digitali. Esattamente come nelle retoriche dell'antipolitica, la costruzione del nemico si nutre dello scontro “noi” vs “loro”; adotta cioè un meccanismo di *unificazione* ideologica (De Blasio, Hibberd, Higgins, Sorice 2012) e di delegittimazione dell'avversario (reale o presunto poco importa). Molto spesso, poi, la costruzione delegittimante del nemico si intreccia con la diffusione di *fake news* (espressione non priva di ambiguità), talvolta prive di ogni fondamento¹. Non si tratta di un fenomeno nuovo, in realtà, dal momento che la costruzione del nemico-demoniaco è attestata già nel IV secolo (contro gli ebrei, accusati di essere cospiratori per conto di Satana) e successivamente contro soggetti e comunità considerate non “conformi” alla maggioranza sociale: dalle minoranze “ereticali” a gruppi sociali marginali o giudicati “devianti”². In tutti i casi, l'accusa al “nemico” si basa su informazioni senza alcun fondamento ma capaci di imporsi per la loro pervasività sociale e l'autorevolezza (presunta) delle fonti (spesso legate al potere politico, religioso e/o economico); più recentemente anche grazie all'istantaneità della comunicazione social.

L'insorgenza dei social media ha in effetti favorito e accelerato anche l'emersione di un pubblico diverso da quello rappresentato dai consumatori “razionali”, che sceglievano le fonti in funzione di un'utilità (presunta) o di un'adesione (culturale o ideologica) e si muovevano come corpo sociale in grado di definire l'opinione pubblica. Ammesso che il pubblico “razionale” sia realmente esistito, esso era tuttavia espressione di un'élite culturale (spesso anche economica); è tuttavia vero che il XXI secolo ha visto lo sviluppo di quello che molti studiosi hanno definito pubblico “emozionale” (Higgins 2008; De Blasio e Selva 2019). La dimensione privata del pubblico “emozionale” assume, grazie ai social media, legittimazione pubblica. È in questa sfera, peraltro, che si attivano

¹ Sono moltissimi i casi di *fake news*, assolutamente destituite di fondamento, che hanno ottenuto un grande successo mediatico fino a generare persino atti criminali. Una delle più note è quella definita “pizzagate” che si basava su una fantasiosa accusa a Hillary Clinton di essere a capo di una setta satanica e pedofila internazionale impegnata a favorire un traffico di bambini. Una pizzeria di Washington (la *Comet Ping Pong*) fu “individuata” come la centrale di tali operazioni e fu bersagliata da minacce online e telefoniche, fino all'irruzione di un soggetto armato di un fucile che sparò alcuni colpi all'interno del locale (senza, fortunatamente, colpire nessuno). Successive azioni di “debunking” hanno ovviamente smontato il caso, individuandone anche la genesi e le procedure di sviluppo.

² Una delle prime accuse mosse ai valdesi (e soprattutto alle donne) nell'XI secolo fu quella di praticare la stregoneria.

quei processi di costruzione della partecipazione politica che sfuggono alle regole istituzionali delle democrazie liberali, soprattutto le esperienze che si collocano all'esterno dei processi formalizzati delle pratiche politiche (si pensi a tutte quelle forme di aggregazione e mobilitazione che hanno trovato spazio e legittimazione nei media e nelle reti informali, prima e nei social media poi). La partecipazione pubblica attivata nella cornice rappresentata dai media (in questo caso, vecchi e nuovi) si caratterizza così come “sfera pubblica emozionale”, in cui entrano in gioco fattori fino a qualche anno fa sconosciuti alla comunicazione politica: fra questi possiamo citare l'emergenza della dimensione individuale degli attori politici, spesso connessa a quella che Christopher Lasch (1979) definiva “cultura del narcisismo”. In questa prospettiva possiamo rubricare fenomeni ormai ampiamente studiati come quelli relativi alla personalizzazione della comunicazione politica e alla frammentazione della sfera pubblica³.

L'attenzione al ruolo dei social media ha generato l'insorgenza di diverse espressioni che se da una parte hanno prodotto riflessioni importanti, dall'altra hanno complicato il quadro teorico, anche a causa della loro ambiguità concettuale. È il caso del concetto di “post-verità”⁴ che è stato troppo frettolosamente connesso in via esclusiva all'evoluzione dei social. È invece evidente che non ci sono solo i social media nel radicamento *post-truth politics* (politica della post-verità), a meno che non si voglia pensare che gli ecosistemi informativi dell'epoca “pre-social” fossero dispensatori di verità, cosa che sarebbe essa stessa una post-verità dal momento che i meccanismi di controllo e manipolazione nonché lo scarso pluralismo delle fonti costituivano elementi peculiari dell'informazione politica (e non solo), e non esclusivamente nei regimi totalitari.

³ Hartley (1996, pp. 156-157) descriveva questa tendenza legandola al processo di costruzione di quella dallo studioso definita “sfera pubblica postmoderna”.

⁴ L'espressione post-verità è stata veicolata grazie alla pubblicazione, nel 2004, del libro *The Post-truth Era* di Ralph Keyes, che faceva riferimento all'irrelevanza della verità nello scenario mediale. Secondo *l'Oxford Dictionary*, tuttavia, l'espressione post-verità non sarebbe così recente: essa, infatti, sarebbe stata usata già nel 1992 dal drammaturgo statunitense di origine serba, Steve Tesich, sul magazine *The Nation*. Tesich lo aveva utilizzato a proposito dello scandalo Iran-Contra, facendo riferimento all'esistenza di un “mondo della post-verità”, creato dai media attraverso una dinamica *top-down*; quindi, con un'accezione diversa da quelle usate oggi nel dibattito pubblico. Nel secondo decennio del XXI secolo, infatti, si è sviluppato un altro significato: quello in cui post-verità rappresenterebbe l'impossibilità del discernimento. In pratica, la velocità dei media (diventata di fatto “istantaneità”) renderebbe impossibile distinguere il vero dal falso.

2. *Propaganda e manipolazione*

Un altro importante aspetto di cornice che dobbiamo considerare quando parliamo di discorsi d'odio online, è la complessa relazione fra comunicazione/propaganda e manipolazione. Il tema della manipolazione dell'informazione non è nuovo negli studi sui media. Esso, infatti, era già presente a Paul Lazarsfeld e a Robert Merton e poi, in modi diversi, agli studiosi della Scuola di Francoforte, che infatti studiarono come la dimensione ideologica dei media produca meccanismi di controllo e orientamento delle masse popolari; e ancora negli studi sull'influenza selettiva che si erano concentrati sulle relazioni fra media e pubblico e la capacità dei primi di generare forme di influenza su diversi registri. Il tema della manipolazione è apparso poi in molte altre "scuole" e correnti di ricerca, sebbene in forme diverse e non sempre coerenti fra loro. Dopo una fase di relativo "ottimismo" sulle capacità del pubblico di produrre letture/interpretazioni autonome e critiche dei messaggi mediali, l'attenzione degli studiosi si è spostata sulle nuove forme di influenza e persino sul ritorno del concetto di manipolazione, che ha costituito il territorio di molte ricerche recenti (Morlino, Sorice 2021).

Gli studi classici sulla manipolazione facevano riferimento a tre presupposti: a) la dimensione intenzionale della manipolazione; b) la dimensione occulta della manipolazione, che fa riferimento all'esistenza di un pubblico inconsapevole e quindi più facilmente manipolabile; c) la dimensione del controllo, connessa alla capacità dell'emittente di orientare, filtrare e dirigere le informazioni (soprattutto nei casi di regimi totalitari o di soggetti svincolati da relazioni sociali di difesa, sebbene tale dimensione è facilmente operazionalizzabile anche per i pubblici "consapevoli" dei regimi democratici). I tre presupposti individuano forme di manipolazione in relazione all'emittente, al contenuto, al pubblico e al contesto (Gili 2001). La dimensione manipolatoria dell'emittente si fonda sulla costruzione di credibilità e di reputazione da parte dello stesso emittente, funzionale alla creazione di fiducia da parte del pubblico. Nella comunicazione politica, per esempio, una strategia manipolatoria "dal punto di vista dell'emittente" è quella che tende a gettare discredito sull'avversario: l'emittente, cioè, si pone come credibile e/o autorevole a fronte di un avversario delegittimato. Le tecniche di

negative campaigning, per esempio, si muovono in questa prospettiva e non è un caso che esse molto spesso si affianchino a forme di “incivility” (altro concetto utile ma ambiguo) e di vero e proprio *hate speech*. Nella sua vecchia ma ancora oggi attualissima analisi sulla manipolazione, Guido Gili (2001) afferma che, dal punto di vista del contenuto le informazioni possono subire quattro tipi di distorsioni, che costituiscono il risultato di due coppie di variabili oppostive: a) distorsione palese *vs* nascosta; b) distorsione intenzionale *vs* volontaria. In questa classificazione, Gili individua: 1) la distorsione intenzionale e palese (come nel caso della “partigianeria” giornalistica) che ha il pregio di essere dichiarata e quindi di lasciare ai destinatari strumenti di difesa e/o di rifiuto; 2) la distorsione intenzionale e nascosta, che avviene nei casi di fonti informative dichiaratamente neutrali ma che lavorano per un “committente” o hanno conflitti di interesse con qualcuna delle parti di cui forniscono informazioni; 3) la distorsione involontaria e palese, risultato delle *routines* produttive giornalistiche e dai meccanismi di *newsmaking* (ovvero le pratiche di selezione, gerarchizzazione, tematizzazione e incorniciamento dei fatti idonei a diventare notizie); 4) la distorsione involontaria e nascosta, che è per lo più l’adozione di una prospettiva ideologica nei meccanismi informativi. Nel caso del pubblico, invece, è il suo stesso coinvolgimento a determinare meccanismi manipolatori. Rientrano in questa categoria sia le tecniche di *emozionalizzazione* dei pubblici (che abbiamo citato sopra) sia quelle legate allo sviluppo di legami omofilici (come nel caso delle cosiddette “echo chambers”) sia quelle che fanno leva sulle capacità di trascinarsi (*bandwagon effect*) da parte di informazioni che riflettono (o rifletterebero) l’opinione della maggioranza. Il contesto, infine, gioca un ruolo fondamentale perché determina le dinamiche di riconoscibilità, l’accettabilità dell’informazione e i meccanismi di pluralismo: è evidente, infatti, che un ambiente con scarse fonti informative (e dunque meno pluralista) sia più predisposto a favorire attività manipolatorie.

Dal punto di vista teorico è anche utile notare la differenza fra persuasione e manipolazione. La prima è un’attività ricorrente in cui un attore (sociale, politico, etc.) tende a influenzare in maniera legittima (attraverso l’argomentazione) le scelte di un altro attore sociale (Stoppino 2001). La seconda, invece, è un’attività

che forza l'opinione in modo surrettizio, usando strumenti nascosti, per lo più ricorrendo a informazioni false (*fake news*) o a processi di negativizzazione dell'avversario o di altri soggetti, anche attraverso la sua demonizzazione e l'attacco con linguaggi d'odio.

I processi manipolatori negli ecosistemi comunicativi si realizzano per lo più attraverso la creazione di influenza sul contesto sociale di riferimento e/o sull'agenda. In questa prospettiva, i meccanismi più frequenti sono quelli della polarizzazione e della saturazione comunicativa (Morlino, Sorice 2021). La prima, come abbiamo visto qualche riga sopra, determina aggregazione verso poli ideologici e radicalizzazione tra posizioni diverse (spesso iper-frammentate); la seconda tende proprio a ridefinire l'agenda e a imporre un'ideologia: in tale meccanismo, il pubblico dei media tende a considerare una visione del mondo caratterizzata e parziale (ideologica) come una sorta di modello unitario di interpretazione possibile della realtà (Sorice 2020a).

I meccanismi manipolatori si sviluppano, in forma deliberata, attraverso la propaganda, altro concetto tornato prepotentemente alla ribalta dell'attenzione della ricerca a causa della crescente centralità delle piattaforme; non è un caso, peraltro, che la "propaganda" politica contemporanea usi, al tempo stesso, retoriche di negativizzazione dell'avversario e diffusione di *fake news*. Nel corso degli ultimi anni, molte ricerche hanno ripreso il vecchio modello della propaganda di Herman e Chomsky (1988)⁵. Una rielaborazione molto convincente di quel modello applicato agli ecosistemi digitali è quella proposta da Christian Fuchs (2018). Il punto di partenza del nuovo modello è la considerazione che l'attenzione online è diventata una *commodity*, al punto che la misurazione della *followership* è essa stessa un indicatore di successo sociale e di legittimazione politica. E, per inciso, molto spesso, anche i discorsi d'odio contribuiscono alla crescita della *followership*. Vanno poi considerate altre variabili di supporto: il

⁵ Edward S. Herman e Noam Chomsky pubblicarono *Manufacturing Consent. The Political Economy of the Mass Media*, nel 1988 (l'edizione italiana fu pubblicata un decennio dopo). Herman e Chomsky sostenevano che i media svolgono un'azione propagandistica, sia adottando un atteggiamento conformista sia operando una sorta di obbedienza acritica alle proprietà da cui dipendono (in quest'ultimo caso, spesso senza alcuna sollecitazione e quindi adottando una specie di auto-censura preventiva o di allineamento ideologico). Il modello prevedeva l'esistenza di cinque "filtri" capaci di determinare forme di influenza mediale: 1) la dimensione economica, la struttura proprietaria e l'orientamento al profitto dei media; 2) il ruolo della pubblicità; 3) il meccanismo di "fornitura" delle notizie (e quindi il ruolo delle fonti); 4) i meccanismi di *nudging* e talvolta di *lobbying* mediale; 5) l'ideologia espressa dai grandi media, spesso "anticomunista" o comunque conservatrice.

ruolo dei gruppi di potere (spesso piccoli) nel determinare dinamiche di attenzione e visibilità negli ecosistemi comunicativi; la funzione di “engagement” svolta da *bot* o da gruppi di soggetti che si prestano a comportamenti coordinati e inautentici, l'esistenza di un sottobosco di notizie false o parzialmente vere (De Blasio 2021).

Fuchs riprende, come dicevamo, il modello di Herman e Chomsky e lo applica alla comunicazione negli ecosistemi digitali; restano i cinque filtri originari del modello primitivo (si veda la nota 5) ma mutano in maniera significativa le dimensioni connesse con ciascuno di essi. Il modello di Fuchs può essere graficamente rappresentato nel modo seguente (figura 1):

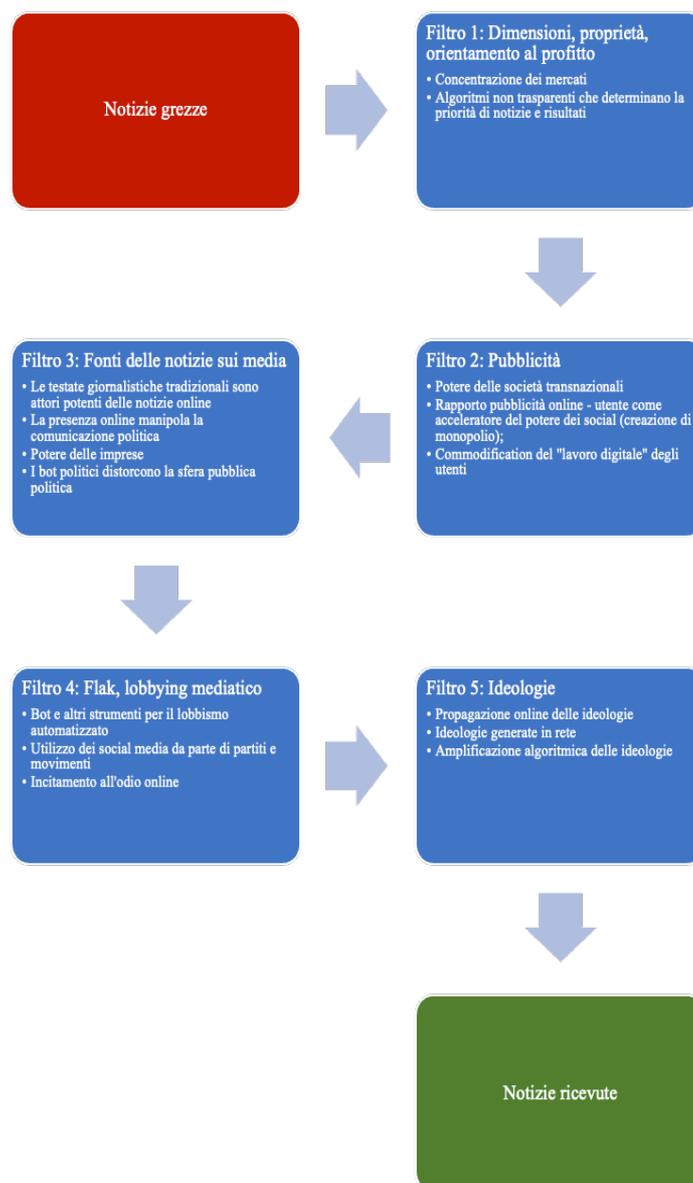


Fig. 1 Il modello della propaganda nella comunicazione online. Rielaborazione con modifiche da Fuchs (2018, p. 86)

La figura 1 evidenzia alcuni aspetti che vanno sottolineati. In particolare: a) nonostante la loro perdita di significatività sociale, restano centrali le testate giornalistiche come strumenti di attivazione di meccanismi di manipolazione; b) i bot, e più in generale le forme algoritmiche e/o di intelligenza artificiale svolgono un ruolo crescente nei meccanismi di distorsione della sfera pubblica; c) la

manipolazione del contesto appare evidente, soprattutto attraverso il ricorso alle “ideologie” generate online dagli utenti (o, più spesso, gruppi organizzati); d) viene individuata e definita in maniera chiara una connessione fra attività di lobbying mediatico e sviluppo dell’odio online: in particolare, dal modello appaiono palesi le interconnessioni fra *hate speech*, bullismo politico e meccanismi di attivazione di troll e specifici algoritmi; e) a proposito di algoritmi, risulta notevole la loro capacità di determinare bolle, produrre gerarchie e determinare l’agenda pubblica, senza che gli utenti possano controllarne il funzionamento e gli esiti; f) il modello evidenzia in maniera incontrovertibile l’importanza del sistema economico, inteso sia come centralità della pubblicità sia come “sfruttamento” di fatto del lavoro digitale degli utenti, assolutamente coerente, peraltro, con i processi di “piattaformizzazione” sociale (van Dijck, de Waal, Poell 2018)⁶. Va ancora notato che l’incitamento all’odio online è promosso dal filtro 4 (il lobbying mediatico) ma è reso possibile e/o velocizzato sia dal filtro delle ideologie sia dalle dimensioni economiche che sono alla base della scarsa trasparenza degli algoritmi che determinano la scala delle priorità in rete. Dovremmo a questo proposito fare una riflessione – come studiosi/studiosi ma ancora di più come cittadine e cittadini – sull’intima connessione fra logiche della propaganda, meccanismi economici, insorgenza e affermazione dell’*hate speech*. La sua presenza, infatti, non può essere derubricata all’azione asistemica di piccoli gruppi di soggetti, magari disagiati e marginali. Sebbene questo fenomeno esista, tuttavia i discorsi d’odio non riguardano solo le esternazioni di disagio di alcuni soggetti più o meno adusi a forme di “inciviltà”, ma costituiscono invece un elemento strutturale della comunicazione negli ecosistemi digitali. I social media, peraltro, sembrano evidenziare la loro capacità di porsi sia come acceleratori dei processi di frammentazione, polarizzazione e radicalizzazione sia come strumenti di anestetizzazione del conflitto sociale, come notato da Wilson e Swyngedouw (2015) nella loro analisi sui processi di depoliticizzazione. I meccanismi di incitamento all’odio si collocano

⁶ La rivisitazione del modello della propaganda evidenzia anche l’esistenza di un processo di frammentazione dell’opinione pubblica, un fenomeno che è strettamente connesso con la più generale frammentazione della sfera pubblica contemporanea. È a questo livello che – proprio in considerazione della crescente centralità delle piattaforme digitali di comunicazione – molti ricercatori hanno parlato di “post-sfera pubblica” (Davis 2019; Schlesinger 2020; Sorice 2020b).

coerentemente dentro questo quadro complicato in cui convivono tendenze sociali centrifughe e non sempre di facile decifrazione.

3. *Il disordine informativo e l'hate speech*

Nel corso degli ultimi anni, esperti, autorità politiche e associazioni di cittadinanza hanno rivolto la loro attenzione sia al tema delle *fake news* sia all'affermazione del fenomeno dell'odio online. Pur trattandosi di fenomeni diversi, è evidente l'esistenza di una relazione: se, infatti, non necessariamente una notizia falsa produce discorsi d'odio, molti discorsi d'odio si basano su informazioni del tutto infondate o sostanzialmente false. Possiamo quindi affermare che, in molti casi, le *fake news* costituiscono il substrato dell'*hate speech* nonché un elemento di accelerazione e diffusione pervasiva dei discorsi d'odio online. “*Fake news* è diventata una categoria postmoderna che in realtà identifica un fenomeno antico che coinvolge eventi basati su calcoli politici” (Ceccarini 2019, p. 9). In effetti, la nozione di *fake news* rappresenta una trasformazione dei vecchi meccanismi di costruzione del consenso, realizzati a partire da strategie politiche di tipo manipolatorio. È per queste ragioni che, accanto al concetto di *fake news*, sono tornati al centro dell'attenzione dell'opinione pubblica, dello stesso *media system* e degli studiosi tutte le nozioni e i concetti connessi alla vasta cornice della disinformazione. Bisogna, a questo proposito, ricordare che il concetto di *disordine informativo* si sostanzia di tre sub-dimensioni. La prima è quello di *disinformazione* propriamente detta che individua il processo col quale un emittente cerca deliberatamente di confondere il pubblico e creare danno (di solito reputazionale) agli avversari politici (o anche economici o di gruppo) attraverso la creazione e diffusione di informazioni false o distorte⁷. La seconda dimensione è quella della *misinformazione*, che è un fenomeno più sfuggente: fa infatti riferimento a informazioni false o inesatte, create e diffuse senza una specifica intenzionalità comunicativa. Alcuni meccanismi del *newsmaking* determinano quasi strutturalmente forme potenziali di *misinformazione*, causati dalla necessità della velocità di diffusione

⁷ Non bisogna confondere la disinformazione con l'informazione di parte (*partisan bias*) che è per lo più evidente e dichiarata mentre la disinformazione spesso si traveste da “informazione” veritiera e “oggettiva”.

dell'informazione con il conseguente sviluppo di errori, spesso dovuti all'assenza di controllo delle fonti (o a controlli poco accurati). In molti casi, le informazioni inaccurate (o false) diventano un pericoloso volano di disordine informativo a causa dell'effetto di moltiplicazione derivante dalla diffusione di tali notizie attraverso i social media. La terza dimensione, infine, è quella nota come *mala-informazione* o *mal-informazione*. Questo termine riguarda lo svelamento truffaldino (e di solito illegale) di informazioni vere ma riservate e private. L'hackeraggio di mail private (come quelle di Hillary Clinton nel 2016 o di Emmanuel Macron nel 2017) sono esempi di mala-informazione, dal momento che le mail non erano false ma erano comunque private e la loro diffusione un evidente reato. Proprio in questa dimensione categoriale si collocano anche altri due sottogruppi di mala-informazione: l'*hate speech* contro gli oppositori politici (Wardle, Derakhsham, 2017) e l'organizzazione di campagne politiche di *negativizzazione* (molto presenti nella comunicazione politica come abbiamo detto sopra) che, con espressione giornalistica vengono definite *mud machine* (macchina del fango). L'insieme di queste tre dimensioni costituisce il fenomeno che viene usualmente definito disordine informativo e che è schematicamente rappresentato dalla figura 2.

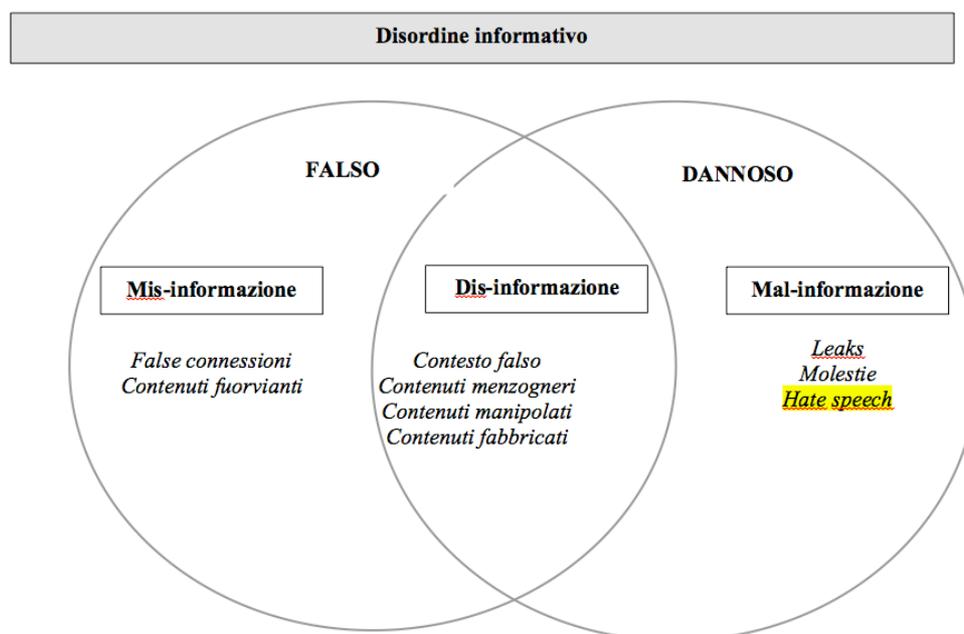


Fig. 2 Disordine informativo. Fonte: rielaborazione da Wardle & Derakshan (2017)

La figura 2 evidenzia le continuità e le connessioni fra i diversi tipi di disordine informativo⁸. Come si può vedere, l'*hate speech* si colloca come sottocategoria della malinformazione, ma è evidente che esso è supportato da meccanismi di disinformazione e finanche dai contenuti fuorvianti che spesso si trovano in rete a proposito di determinati soggetti e/o gruppi sociali. Proprio per la sua trasversalità, l'*hate speech* viene talvolta considerato come una dimensione a sé stante del disordine informativo. Noi preferiamo continuare a considerare i discorsi d'odio come una sottocategoria della malinformazione, sia per accentuare il suo legame con la disinformazione sia per notarne in maniera ancora più decisa la pericolosità. È evidente, infatti, che l'*hate speech* non costituisca una variabile impazzita del sistema comunicativo ma – purtroppo – un elemento strutturale dei meccanismi di disordine informativo, a loro volta stabilmente inseriti nelle logiche comunicative degli ecosistemi digitali.

Una definizione “istituzionale” di *hate speech* è quella proposta nel 2015 dalla Commissione contro il razzismo e l'intolleranza del Consiglio d'Europa. L'*hate speech*, in quel caso, veniva definito in questo modo:

si intende per discorso dell'odio il fatto di fomentare, promuovere o incoraggiare, sotto qualsiasi forma, la denigrazione, l'odio o la diffamazione nei confronti di una persona o di un gruppo, nonché il fatto di sottoporre a soprusi, insulti, stereotipi negativi, stigmatizzazione o minacce una persona o un gruppo e la giustificazione di tutte queste forme o espressioni di odio testé citate, sulla base della “razza”, del colore della pelle, dell'ascendenza, dell'origine nazionale o etnica, dell'età, dell'handicap, della lingua, della religione o delle convinzioni, del sesso, del genere, dell'identità di genere, dell'orientamento sessuale e di altre caratteristiche o stato personale⁹

⁸ Un esempio di misinformazione è rappresentato da titoli fuorvianti di un articolo che facilmente possono favorire la costruzione di un “contesto falso”, in cui vengono attivate vere e proprie attività di manipolazione informativa. I casi più evidenti di disinformazione sono rappresentati dai contenuti fabbricati allo scopo di orientare l'opinione pubblica, spesso a opera di professionisti della disinformazione e persino da gruppi organizzati e con notevole capacità di investimento economico. La mala-informazione, infine, può a sua volta appoggiarsi a dinamiche di disinformazione o usare l'inganno involontario provocato dalla misinformazione. In questo caso, vengono anche usati *troll* oppure gruppi organizzati di *haters* (odiatori) in quelli che vengono definiti *hate groups* o ancora strumenti “automatici” come i *bot*.

⁹ Il testo completo della Commissione (Raccomandazione di politica generale n. 15) è reperibile qui: <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech-ital/16808b5b04>. Si noti che questa definizione è stata usata anche dalla Commissione parlamentare italiana sull'intolleranza, la xenofobia, il razzismo e i fenomeni di odio, attiva per 14 mesi dal 2016.

La definizione di *hate speech* è comunque inevitabilmente imprecisa, dal momento che riguarda un fenomeno in continua evoluzione, non nuovo nel panorama delle società umane ma fortemente trasformatosi all'interno della comunicazione online. Un'analisi utile è quella prodotta dalla commissione "Jo Cox" istituita presso la Camera dei Deputati nel 2016 e attiva per 14 mesi. La commissione (sull'intolleranza, la xenofobia, il razzismo e i fenomeni di odio) individuò – sulla base della documentazione raccolta – l'esistenza di una "piramide dell'odio". Significativamente, la Commissione colloca alla base della piramide gli stereotipi negativi, le rappresentazioni false o fuorvianti e il linguaggio ostile banalizzato, oltre ovviamente gli insulti. Le forme di stereotipizzazione (che sono una semplificazione spesso fuorviante) e le informazioni false (cioè gli elementi di disordine informativo) sono quindi alla base della piramide dell'odio, ne costituiscono la struttura portante e in qualche modo rendono possibile l'odio come discorso sociale. Sono proprio queste forme di disordine informativo che rendono possibili discriminazioni che vengono poi perpetrate negli ambiti più disparati e che, a loro volta, funzionano da detonatore per l'insorgenza del linguaggio d'odio; questo, come abbiamo visto, si rivolge principalmente a soggetti (o gruppi) identificati per caratteristiche che diventano elemento di "negativizzazione", dal sesso all'etnia, dal colore della pelle all'orientamento sessuale, dalle scelte religiose all'appartenenza a determinati gruppi sociali e così via. L'ultimo stadio di questa ipotetica piramide è costituito dai "crimini d'odio" che si configurano come veri e propri atti di violenza fisica contro i soggetti individuati come "nemici" e appunto stigmatizzati e sottoposti ad attacco negli ecosistemi comunicativi digitali.

Una delle caratteristiche strutturali dell'*hate speech* è che esso si basa su porzioni di discorso – frasi brevi, epiteti, slogan – spesso in forma non dialogica e che possiamo considerare come "discorso" solo facendo riferimento al contesto (gruppi online, profili social) in cui questi si ricollocano e assumono un senso sociale. L'*hate speech*, in altri termini, costituisce una narrazione sociale, radicata su precondizioni (gli stereotipi) e capace di influenzare l'opinione pubblica. La narrazione dell'odio può assumere molte e differenziate forme: può avere come

scopo quello di colpire un gruppo sociale (solitamente minoritario e/o marginale) oppure di ridicolizzare soggetti in situazioni specifiche (per motivi sanitari o perché residenti in realtà liminali) o ancora di discriminare soggetti e/o gruppi specifici (migranti, rifugiati, etc.). Un tipo particolare di narrazione dell'odio è quella che colpisce, in maniera trasversale, le donne¹⁰: l'ultimo caso – solo in ordine di tempo, rispetto al momento in cui scriviamo – è quello che ha visto come vittima la campionessa di nuoto sincronizzato Linda Cerruti, attaccata con commenti discriminatori di tipo sessista per una foto pubblicata sul proprio profilo Instagram, in cui esibiva le medaglie (otto, per la cronaca) vinte ai campionati europei dell'agosto 2022. I commenti volgari e sessisti di cui l'atleta è stata vittima (e che hanno provocato la sua risposta e quella della Federazione Italiana Nuoto) costituiscono una specifica (e purtroppo non rara) modalità di narrazione d'odio contro una donna¹¹.

Studiare le modalità di narrazione dei discorsi d'odio (nonché dell'incitamento all'odio) costituisce un aspetto fondamentale per comprendere l'insieme delle relazioni che connettono l'odio online con le altre forme di disordine informativo, nonché per individuare le caratteristiche di funzionamento e le logiche di ingaggio delle diverse forme di *hate speech*.

4. Il contrasto all'odio online

Anche se consideriamo l'*hate speech* come una variabile specifica del disordine informativo, le misure di contrasto a esso non possono limitarsi alle forme tradizionali di *fact-checking* e *debunking*. Com'è noto, il contrasto al disordine informativo viene realizzato in vari modi. Il primo è quello che viene definito *fact-checking*, ovvero il controllo sull'attendibilità e veridicità dei fatti (informazioni) riportate e delle fonti che li hanno originati. Il secondo è quello che viene definito *debunking* (letteralmente “demistificazione”), cioè l'adozione di

¹⁰ La ricorrenza di “attacchi” alle donne è rilevata anche da studi che non riguardano direttamente l'*hate speech*. In una ricerca sulla copertura mediatica della regolarizzazione dei lavoratori e delle lavoratrici migranti, appariva con evidente drammaticità come le donne costituissero un obiettivo specifico di *incivility* o comunque di strategie di delegittimazione (De Blasio, Rega, Valente 2022).

¹¹ Uno degli elementi rilevati nelle ricerche sull'*hate speech* è la dimensione intersezionale del fenomeno, che produce un effetto di rinforzo. Le donne, in particolare, costituiscono un obiettivo specifico: le coppie migrante/donna, avversario politico/donna, nemico/donna sono molto presenti nel novero delle vittime dei discorsi d'odio. In molti casi, infatti, nelle narrazioni dell'odio, alla delegittimazione dell'avversario/nemico si unisce una tendenza sessista e misogina.

metodi scientifici per mettere in dubbio e confutare affermazioni prive di fondamento scientifico. Si tratta di misure importanti ma che scontano una serie di problemi, soprattutto quando parliamo della diffusione di notizie false che possono generare odio, e dove la velocità della smentita (e/o della risposta) raramente può intercettare la velocità di diffusione dell'input informativo. Sono necessarie, quindi, altre misure: da un'effettiva presa di responsabilità da parte delle istituzioni e delle agenzie formative fino al contrasto alle forme di concentrazione e allo strapotere del capitalismo digitale che rappresenta comunque un elemento da non sottovalutare. In questa direzione ci sono state iniziative importanti, come il progetto *Controlodio*, promosso congiuntamente dalle Università di Torino e di Bari. In questo caso, è stato sviluppato un algoritmo che consente di monitorare i discorsi di incitamento all'odio (in italiano) provenienti da Twitter. A febbraio 2019, i ricercatori del progetto *Controlodio* avevano calcolato che il 14% dei tweet in lingua italiana sono classificabili come *hate speech*; in tutti i casi, tuttavia, la maggiore "potenza" dei discorsi d'odio deriva dalle modalità attraverso cui i personaggi pubblici (soprattutto i leader politici) riprendono le notizie e le "incorniciano". I meccanismi di *engagement* sono, a questo riguardo, molto importanti. Secondo diversi studiosi, nelle reti sociali, i contenuti che utilizziamo e a cui siamo esposti rappresenta una porzione di realtà, selezionata da algoritmi che analizzano le nostre attività e preferenze generando di fatto un fenomeno noto come *filter bubbles* (Pariser 2011). In questa prospettiva, la massimizzazione del tempo di permanenza sulle piattaforme di comunicazione social (funzionale a legittimi interessi economici degli "owners" delle piattaforme stesse) è prodotta dall'incremento dell'*engagement*, ovvero dalla moltiplicazione delle interazioni (anche pubblicitarie, attraverso per esempio, i click su video e contenuti di *advertising*). Secondo Derakhshan e Wardle nel loro ormai notissimo rapporto per lo *Shorenstein Center on Media, Politics and Public Policy*, la sinergia fra social media e televisione rappresenta un "ricettacolo per politici sensazionalistici che traggono beneficio da popolazioni polarizzate attraverso divisioni politiche, economiche, religiose, razziali o etniche". In tale quadro, un ruolo importante è rivestito anche dalle *dark ads*, o per essere più precisi dal legame fra queste e le

fake news (o menzogne appositamente costruite a scopo manipolatorio). I social media avrebbero anche favorito una crescente diffusione di notizie false – come abbiamo visto – capaci di influenzare e orientare il contesto decisionale esistente. Alcuni studi, infatti, hanno evidenziato l’incapacità di segmenti significativi della popolazione nell’identificare la veridicità o meno delle *fake news* (Allcott, Gentzkow 2016). Proprio allo scopo di individuare come funzionano i meccanismi di *engagement*, si è sviluppata la ricerca MINE (*Mapping Italian News Election*) dell’Università di Urbino. In quel caso, le “notizie” erano tratte da tre fonti: Google News, il Global Database of Society (<https://www.gdeltproject.org>) e Twitter. Le tendenze politiche sono poi state studiate usando il *Media Partisanship Attention Scores* (Faris et al 2017), un indice sviluppato da Yochai Benkler, Robert Faris, Hal Roberts e Ethan Zuckerman per il loro studio sul panorama mediale durante le elezioni presidenziali negli Stati Uniti. In questo modo, è stata verificata l’esistenza di elementi di “moltiplicazione” delle informazioni provenienti dai leader politici o da altri soggetti che a essi si ispirano; i meccanismi di *engagement*, in altre parole, sono stati studiati misurandone l’ampiezza e il potenziale di distribuzione (cfr. Ceccarini et al. 2019; Ceccarini et al. 2020).

Sicuramente più efficace – ma su tempi inevitabilmente lunghi – è l’adozione di prospettive di *media literacy*, ovvero l’acquisizione di competenze medialità a largo spettro, capaci di fornire strumenti di verifica delle informazioni e della loro fondatezza, nonché di dare agli utenti della rete una sorta di “educazione alla comunicazione responsabile”. La *media literacy* (più correttamente *digital media literacy* nell’ambito della comunicazione digitale) è appunto l’acquisizione di competenze specifiche, la capacità di accedere, analizzare, valutare e creare messaggi in una varietà di contesti (Livingstone 2004) o almeno la capacità di assumere la cultura del dubbio (boyd 2017). Secondo diversi studiosi (Bulger e Davison 2018), la *media literacy* è uno dei metodi più efficaci per contrastare il disordine informativo. Secondo l’Unesco, la *digital media literacy* (sempre congiunta alla *information literacy*) si basa su un set di conoscenze che possono essere riassunte come nella figura 3.

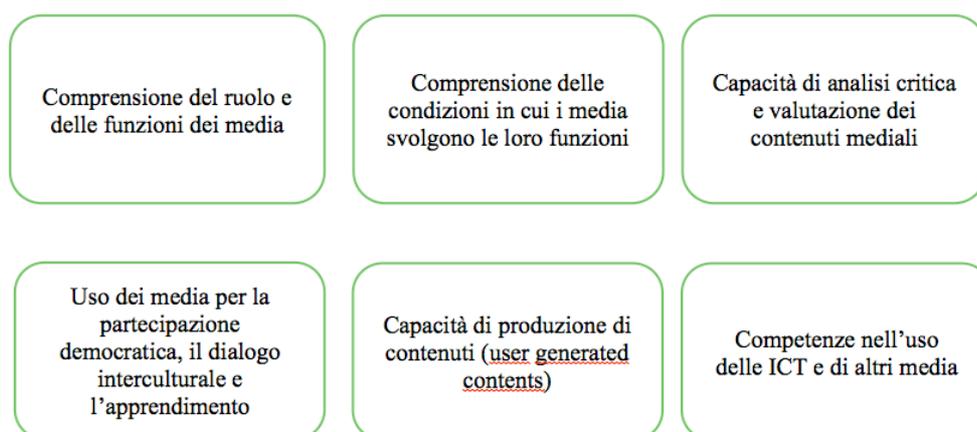


Fig. 3 Le componenti della *media literacy*. Fonte Unesco – *Media and Information Literacy Curriculum for Teachers*

La *media literacy* si è sviluppata a partire dal paradigma della *pedagogia critica*, che si fonda sull'idea che sia necessario educare i bambini/ragazzi e le bambine/ragazze all'acquisizione di competenze per saper riconoscere le dinamiche di potere, di esclusione e di conflitto sociale (Mclaren 1997). In alcuni paesi sono stati fatti esperimenti di introduzione di programmi di *media literacy* nelle scuole con ottimi risultati in termini di acquisizione di capacità di uso critico della comunicazione digitale¹².

L'attenzione alla *media literacy* sposta la responsabilità del contrasto al disordine informativo sulle agenzie formative; al tempo stesso, un controllo regolatorio troppo articolato sulla produzione delle *fake news* e/o dell'incitamento all'odio rischia – soprattutto nel caso di paesi con democrazie deboli o regimi totalitari – l'incremento di meccanismi di censura¹³. Per questo motivo molti studiosi (Bulger & Davison 2018; Livingstone 2018; Buckingham 2019) hanno proposto l'adozione di misure complesse e ibride, che comprendano sia

¹² La policy d'intervento e le iniziative dell'Unesco sulla media literacy sono disponibili qui: <https://www.unesco.org/en/communication-information/media-information-literacy>

Le linee d'azione e i toolkit d'intervento sono invece reperibili qui: <https://www.unesco.org/en/node/67946>
Infine, i moduli per insegnanti sulla *media and information literacy* sono disponibili qui: <http://unesco.mil-for-teachers.unaoc.org>

¹³ Un tema spesso connesso a quello del disordine informativo è quello della *cybersecurity*, su cui esiste ormai un'ampia letteratura che affronta il tema da diverse prospettive (sicurezza dei dati personali, sicurezza militare, intelligence, diplomazia digitale, etc.). Il collegamento fra *cybersecurity* e i disordini informativi è stato messo in luce in maniera esplicita dalla Commissione Europea fin dal 2019 con un'apposita risoluzione (<https://europa.eu/rapid/press-releaseIP-19-2011en.htm>).

l'incremento delle conoscenze comunicative dei cittadini (attraverso programmi di *media literacy*), sia il ricorso a strumenti di *debunking* e *fact-checking*, in un quadro regolatorio e normativo definito e chiaramente inserito in una prospettiva democratica. È tuttavia evidente che le questioni connesse all'*hate speech*, alla moltiplicazione delle *fake news* e, più in generale, all'arcipelago del disordine informativo non possono essere affrontate con misure parziali né trattate solo all'interno di un quadro repressivo. Sono invece necessarie misure sistemiche, che mettano in relazione forme di contrasto ai fenomeni di disordine informativo con la crescita di una profonda educazione alla comunicazione digitale e che, al tempo stesso, consentano di avere strumenti di difesa efficace per le vittime dell'odio online, senza che questo si traduca in un deficit di libertà d'espressione e di mobilitazione digitale. Un compito non semplice che è diventato il focus, non a caso, di ricerche e interventi attivi.

5. Ricerche e aree d'intervento

Nel corso del 2022, Amnesty International ha pubblicato il suo annuale *Barometro dell'Odio*, dedicato in particolare al tema della cittadinanza e alle leggi relative a essa.¹⁴ Nella pagina di presentazione, l'organizzazione umanitaria scrive che

1 commento su 10 è offensivo, discriminatorio e/o hate speech e quando si parla di riforma della cittadinanza a farla da padrone sono spesso razzismo e xenofobia [...]. Un tema quasi invisibile, così come lo sono i suoi protagonisti. Tra tecnicismi, strumentalizzazioni politiche e una sovrapposizione fuorviante al tema immigrazione, di cittadinanza non solo si parla poco, ma anche male [...]. Dopo le donne, i bersagli preferiti dagli utenti sono le persone con background migratorio, i rifugiati e i migranti, seguiti da chi è impegnato nel mondo della solidarietà¹⁵.

Si tratta di un quadro che purtroppo conferma i risultati di altre ricerche ed evidenzia attraverso dati empirici il legame profondo fra disinformazione, contenuti fuorvianti (come nel caso della misinformazione) e *hate speech*.

¹⁴ Il rapporto "Barometro dell'odio – Senza cittadinanza" è stato realizzato da Amnesty International con il supporto dell'Unione Cristiana Evangelica Battista d'Italia.

Il rapporto completo è disponibile qui:

<https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2022/05/Barometro-dellodio-2022-Senza-cittadinanza-low-.pdf>

¹⁵ Si veda: <https://www.amnesty.it/barometro-dellodio-senza-cittadinanza/>

La ricerca di Amnesty International presenta un quadro molto chiaro, anche grazie alla ricchezza dei materiali raccolti: 6 settimane di monitoraggio, 50 attivisti coinvolti ma soprattutto 6 milioni di tweet e commenti raccolti da pagine e account politici e del sistema dell'informazione, con ben 27.000 contenuti unici catalogati. I social network – ci dicono i ricercatori e le ricercatrici di Amnesty International – costituiscono un “vero e proprio sfogatoio virtuale, con la maggior parte dei commenti degli utenti (56,9%) che esprime opinioni negative, critiche, polemiche e linguaggio d'odio, contro il 43,1% di contenuti dall'accezione neutra o positiva” (Amnesty International 2022, p. 11).

Un aspetto da sottolineare – presente trasversalmente in molte ricerche sull'odio online – riguarda il linguaggio delle narrazioni dell'odio. Esse, infatti, si connotano per essere fortemente polarizzate (noi *vs* loro) e per una totale assenza di argomentazione. In molti casi, infatti, l'*hate speech* funziona in maniera apodittica e sostituisce all'argomentazione la fallacia logica delle pseudo-argomentazioni che, di fatto, rendono impossibile persino la contro-argomentazione (esercizio peraltro difficile e rischioso perché chi lo pratica subisce poi molto spesso lo stesso trattamento di odio di chi intende difendere)¹⁶. D'altra parte, l'assenza di argomentazione è correlata alla dimensione fortemente polarizzante dell'odio online, in cui la presa di posizione e/o la scelta di campo sono sovraordinate rispetto alla logica razionale dell'argomentazione e della prova. A questo livello, peraltro, si attivano meccanismi di *confirmation bias*, in cui la selezione degli argomenti che supportano le proprie opinioni attivano logiche di “tranquillizzazione” (costituiscono cioè la “comfort zone” per alcuni soggetti); al tempo stesso proprio i fenomeni di polarizzazione discorsiva (e politica) accelerano i meccanismi di *confirmation bias* e determinano lo sviluppo di potenziali *echo-chambers*.

Al di là delle dimensioni linguistiche, il tema dell'*hate speech* può, peraltro, essere messo in correlazione anche con la riemersione di una sfera pubblica fortemente emozionale, in cui i social giocano un ruolo importante. La profonda trasformazione della sfera pubblica, in effetti, ha prodotto da un lato la

¹⁶ La fallacia retorica costituisce uno degli esempi più presenti nei discorsi d'odio online. Ma sono presenti anche nella narrazione giornalistica di alcuni temi rilevanti e “divisivi” (come, per esempio, l'immigrazione) e persino nelle loro dinamiche narrative. Cfr. Rega, De Blasio, Valente (2021); De Blasio, Selva, Sorice (2022).

riconsiderazione dell'importanza della teoria della spirale del silenzio e dall'altro una rinnovata attenzione alle numerose peculiarità della sfera pubblica moderna. Un caso specifico è rappresentato proprio dalla sfera pubblica emozionale, il cui studio richiede uno sforzo di analisi dei legami sociali, delle comunità e della solidarietà che sembrano indeboliti dall'egemonia culturale dell'individualismo, spesso rafforzata da algoritmi che guidano l'opinione pubblica (come nel caso delle *filter bubbles*) o facilitano la discorsivizzazione dell'altro come nemico. In questo contesto, l'attenzione degli studiosi e dei politici si sposta dai legami comunitari alle relazioni intime come elemento chiave per definire l'identità personale (Stanyer 2013). Le emozioni acquisiscono un ruolo fondamentale nel processo di integrazione e disintegrazione dei legami sociali, con importanti conseguenze sulla fiducia e sulla coesione sociale (De Blasio, Selva 2019; De Blasio, Selva 2021). Non si può non notare la relazione fra l'emozionalizzazione della sfera pubblica (cornice che facilita – sebbene non determini – lo sviluppo dell'*hate speech*) e la frammentazione sociale. Quest'ultima è un processo orizzontale e verticale a un tempo: da una parte, infatti, essa riguarda la disconnessione tra pubblici e comunità nell'accesso al discorso informativo (le notizie e il commento su di esse), dall'altra parte concerne la “rottura” tra cittadini e intermediari informativi (Bennett, Pfetsch 2018). Il processo di frammentazione diventa così funzionale all'incremento della polarizzazione che è, in parte, dipendente dai processi di individualizzazione del consumo di informazione e in altra parte logica della semplificazione del discorso politico. I processi di frammentazione e le dinamiche di polarizzazione favoriscono la radicalizzazione delle posizioni, legittimate da meccanismi – come abbiamo visto – di *confirmation bias*. L'emozionalizzazione della sfera pubblica, allora, costituisce un esito dei processi di frammentazione sociale e di polarizzazione politica (Iyengar et al. 2019). La drammatizzazione del conflitto – anche attraverso i discorsi d'odio – rappresenta l'altra faccia dei processi di omogeneizzazione, in cui le posizioni critiche (o semplicemente “argomentate”) vengono facilmente marginalizzate.

Non bisogna quindi dimenticare il contesto politico-economico e il frame comunicativo in cui i discorsi dell'odio si manifestano e si affermano. Appare

evidente, per esempio, che le logiche della piattaformaizzazione della sfera pubblica influenzano in maniera decisiva le strategie argomentative degli attori politici (ma anche di quelli non esplicitamente “politici”), soprattutto su temi potenzialmente sensibili o che vengono definiti – talvolta in maniera strumentale – “divisivi”. Un caso specifico è rappresentato dal dibattito sul cosiddetto DDL Zan (il progetto di legge contro l’omotransfobia) che ha evidenziato sia come i rappresentanti politici abbiano portato il tema fuori dalle aule parlamentari, in modo da coinvolgere pubblici più ampi nel quadro della frammentazione sociale, sia come le dinamiche della piattaformaizzazione della sfera pubblica abbiano influenzato in maniera dirimente l’esito del dibattito parlamentare (De Blasio, Selva, Sorice 2022)¹⁷.

Tornando alla ricerca sull’*hate speech*, è utile ricordare la complessità di tali lavori, sia perché insistono su un ecosistema molto ampio (si pensi ai diversi social, ai gruppi su piattaforme di messaggistica nonché alle piattaforme nate proprio per favorire lo sviluppo di culture settarie) sia perché l’universo di discorso è rappresentato da una molteplicità di variabili (dalle notizie ai post non informativi, dai commenti ai re-post, fino all’adesione attraverso “like” o simili che possono a loro volta generare sub-universi di discorso). Non è un caso che in questo tipo di ricerche siano necessari approcci ibridi e multidimensionali (per lo più nella prospettiva *mixed-methods*), capaci di adottare una prospettiva sistemica interpretativa supportata da una coerente analisi quantitativa e, al tempo stesso, capaci di porsi al confine fra discipline e metodi diversi.

¹⁷ La ricerca sulle modalità di discorsivizzazione sociale del dibattito sul DDL Zan si è sviluppata adottando una prospettiva basata su un approccio narrativo. Utilizzando la piattaforma *CrowdTangle* sono stati scaricati i post pubblicati su Facebook durante i momenti di maggiore impatto pubblico della discussione sul DDL; adottando un approccio *mixed-methods* di tipo nidificato sono state poi usate tecniche di analisi dei post di tipo quantitativo. La selezione dei post, in altri termini, rispondeva a criteri qualitativi/interpretativi piuttosto che a logiche di rappresentatività statistica. La ricerca ha preso in considerazione tre aspetti: a) la *polarizzazione politica*, indagata attraverso un’analisi degli asserti valutativi che ne ha misurato l’intensità, ricorrendo a un’accezione della polarizzazione come progressiva divergenza (Lelkes 2016); b) l’*emozionalizzazione*, attraverso un’analisi del contenuto in grado di evidenziare il repertorio emozionale; c) le *strategie argomentative*, attraverso un’analisi del discorso secondo l’approccio dei Critical Discourse Studies (Fairclough 2010; Wodak & Meyer 2015; Wodak 2020). Nonostante la ricerca non avesse come suo obiettivo quello di analizzare i discorsi d’odio, i meccanismi di polarizzazione ed emozionalizzazione trovano spesso supporto in modalità argomentative (come, per esempio, il *détournement* semantico, l’assenza di contro-argomentazioni o diverse forme di “disordine informativo” – dalla strumentalizzazione a forme più o meno esplicite di *minsinfomazione*) che sono le stesse che si trovano alla base delle modalità di sviluppo dell’odio online. Si può ancora notare come l’insistenza dei politici sull’eteronormatività alimenti paure e risposte irrazionali: tali evidenze si trovano sia in ricerche che non trattano l’*hate speech* sia in altre (come quella qui citata di Amnesty International) che invece si concentrano specificamente sui discorsi dell’odio.

6. Contrastare l'odio online. Difendere la democrazia

Il contrasto all'odio online e all'incitamento all'odio (due aspetti diversi di una realtà fortemente interconnessa) non è facile. Nonostante siano stati emanati regolamenti e direttive, nonostante l'intervento delle autorità e nonostante i tentativi di regolazione da parte delle stesse piattaforme digitali, la materia resta di difficile decifrazione. La Commissione Europea entrata in carica nel 2019 (con la presidenza di Ursula von der Leyen) ha inserito subito (nell'agenda politica presentata) l'impegno contro la disinformazione e *l'hate speech*, significativamente connessi. Il disordine informativo e l'odio online vengono definiti pericoli per la democrazia. Non solo, quindi, le ingerenze esterne nei periodi elettorali (da sempre uno dei punti di impegno dell'Unione Europea nell'ambito della comunicazione digitale) vengono definite pericolose per la democrazia ma anche la disinformazione e i discorsi dell'odio.

L'hate speech, in effetti, deriva ed è supportato da fenomeni di disinformazione ma è anche uno strumento di avvelenamento dell'informazione. Le ricerche condotte negli ultimi anni evidenziano alcuni elementi strutturali dei discorsi d'odio online:

- 1 l'assenza di argomentazione (o il continuo spostamento del focus argomentativo);
- 2 la forte presenza di elementi di fallacia argomentativa;
- 3 il ricorso a retoriche polarizzanti, in un quadro di forte frammentazione sociale;
- 4 la presenza di meccanismi di *confirmation bias*, che determinano un ulteriore livello di disordine informativo;
- 5 l'attacco prioritario ad alcuni soggetti (migranti, minoranze, segmenti sociali specifici, volontari) con un deciso incremento della violenza discorsiva nei confronti delle donne;
- 6 l'adozione di meccanismi di semplificazione discorsiva (*dumbing down*) allo scopo di costringere a scelte di campo;
- 7 la saturazione comunicativa che accompagna sempre l'odio verso alcuni soggetti;

- 8 il legame fra disinformazione e *hate speech*;
- 9 la connessione fra processi di emozionalizzazione e incitamento all'odio online;
- 10 la relazione indiretta fra la piattaformaizzazione della sfera pubblica, le trasformazioni dei modelli informativi e l'odio online.

A queste caratteristiche dovremmo aggiungere quelle più generali, riguardanti il quadro generale rappresentato dal sistema dell'informazione. Appare infatti chiaro da diversi studi (De Blasio, Rega, Valente 2022) che la piattaformaizzazione del sistema dell'informazione favorisce una narrazione altamente polarizzante, dove peraltro i social media tendono a rafforzare sia la credibilità delle notizie (“rimbalzandole”) sia la polarizzazione. Va ancora notata la persistenza di una potenziale saldatura tra la logica della propaganda e le nuove routine di produzione giornalistica: una logica che va inquadrata a sua volta nella complessa relazione triadica tra cittadini, sistema informativo e politica (Sorrentino 2009) e che, in alcuni casi, può costituire un territorio fertile per lo sviluppo dei discorsi d'odio. Infine, va nuovamente sottolineato il legame “saturazione/polarizzazione” che accompagna la trasformazione degli ecosistemi comunicativi e costringe a ripensare le teorie e gli approcci della comunicazione politica (Barisione 2020; Morlino, Sorice 2021).

I discorsi dell'odio – al di là delle motivazioni di carattere psicologico che investono i soggetti responsabili – costituiscono un pericolo per la democrazia, dal momento che essa trova alimento solo da uno spazio pubblico aperto e in cui le diverse prospettive possano serenamente confrontarsi. L'odio online – ma, più in generale, le forme di odio che delegittimano specifici segmenti sociali – contribuiscono alla frammentazione della sfera pubblica e, in definitiva, del quadro democratico. Per questo, al di là della legittima tutela delle vittime e della necessità di interventi rapidi di prevenzione e sanzione dei responsabili, è necessaria una presa di coscienza sociale. E, con essa, dell'adozione di strumenti informativi e formativi efficaci, che consentano sia la difesa dall'odio sia la trasformazione del dissenso legittimo in pratiche di confronto rispettoso della diversità e basato sull'argomentazione razionale. Non si tratta di un compito semplice, soprattutto quando persino alcuni attori politici favoriscono o trovano

vantaggio dall'odio online. Per questo motivo, il compito delle istituzioni formative e di ricerca – pur complesso e difficile – resta fondamentale. Combattere l'*hate speech* non significa volere semplicemente un mondo digitale più “civile” o “ben educato”; significa contribuire alla creazione di spazi pubblici capaci di garantire voce a tutte e tutti e, in questo modo, a difendere la democrazia.

Riferimenti bibliografici

- Allcott, H., Gentzkow, M., 2017, *Social Media and Fake News in the 2016 Election*, National Bureau of Economic Research, Cambridge, <http://www.nber.org/papers/w23089>
- Amnesty International Italia, 2022, *Barometro dell'odio. Senza cittadinanza*. Amnesty International sezione italiana, Roma. Disponibile all'indirizzo: <https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2022/05/Barometro-dellodio-2022-Senza-cittadinanza-low-.pdf>
- Barisione, M., 2020, “Contro il comunicazionismo. Per una critica al riduzionismo comunicativo”, in *Comunicazione Politica*, 21(3), pp. 347-370.
- Benkler, Y., Faris, R., Roberts, H., 2018, *Network Propaganda: Manipulation, Disinformation and Radicalization in American Politics*, Oxford University Press, Oxford.
- Bennett, W. L., Livingston, S., 2018, “The disinformation order: Disruptive communication and the decline of democratic institutions”, in *European Journal of Communication*, 33(2), pp. 122–139.
- Bennett, W.L., Pfetsch, B., 2018, “Rethinking Political Communication in a Time of Disrupted Public Spheres”, in *Journal of Communication* 68(2), pp. 243-253.
- boyd, d., 2017, “Did Media Literacy Backfire?” in *Points. Data & Society* (<https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d>)
- Bulger, M., Davison, P., 2018, “The Promises, Challenges, and Futures of Media Literacy”, in *Data and Society*. <https://digital.fundacionceibal.edu.uy/jspui/bitstream/123456789/227/1/DataAndSocietyMediaLiteracy2018.pdf>
- Ceccarini, L., 2019, “Could the Internet be a safety-net for democracy in crisis?”, in *Italian Political Science Review. Rivista italiana di Scienza Politica*, <https://doi.org/10.1017/ipo.2019.24>, pp. 1-17.
- Ceccarini, L., Diamanti, I., 2018, *Tra politica e società. Fondamenti, trasformazioni e prospettive*, Il Mulino, Bologna.

- Colombo, F., 2020, *Ecologia dei media. Manifesto per una comunicazione gentile*, Vita & Pensiero, Milano.
- Davis, A., 2019, *Political Communication. A New Introduction for Crisis Times*, Polity, Cambridge.
- De Blasio, E., 2021, “Contesto politico e *propaganda model*”, in Morlino, L. e Sorice, M., a cura di, *L'illusione della scelta. Come si manipola l'opinione pubblica in Italia*, Luiss University Press, Roma, pp. 247-264.
- De Blasio, E., Hibberd, M., Higgins, M. & Sorice, M., 2011, *La leadership politica: media e costruzione del consenso*, Carocci, Roma.
- De Blasio, E., Rega, R., Valente, M., 2022, “Polarization and Platformization of News in Italian Journalism: The Coverage of Migrant Worker Regularization”, in Palau-Sampio, D., López-García, G., Iannelli, L., eds., *Contemporary Politics, Communication, and the Impact on Democracy*, IGI Global, Heshey, pp. 74-92.
- De Blasio, E., Selva, D., 2019, “Emotions in the Public Sphere: Networked Solidarity, Technology and Social Ties”, in Fox, B., Ed., *Emotions and Loneliness in a Networked Society*, Palgrave Macmillan, London.
- De Blasio, E., Selva, D., 2021, “COVID-19 in Italy: Performing Power and Emotions”, in *Frontiers in Political Science* 3:661378. doi: 10.3389/fpos.2021.661378
- De Blasio, E., Selva, D., Sorice, M., 2022, “Il dibattito sul DDL Zan e la post-sfera pubblica italiana”, in *Mediascapes Journal*, 19/2022, pp. 89-112.
- De Blasio, E. e Sorice, M., 2020, “L'immagine del nemico nei social media”, in Anselmi, M. e Guercio, L., a cura di, *Il nemico*, Mondadori Università, Milano.
- Fairclough, N., 2010, *Critical Discourse Analysis. The Critical Study of Language*, Routledge, London.
- Faris, R. Roberts, H., Etling, B., Bourassa, N., Zuckerman, E. and Benkler, Y., 2017, *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election* (August 2017). Berkman Klein Center Research Publication, 2017-6. Available at SSRN: <https://ssrn.com/abstract=3019414>
- Fuchs, C., 2018, “Propaganda 2.0: Herman and Chomsky's Propaganda Model in the Age of the Internet, Big Data and Social Media”, in Carañana, P. and Klagen, J., eds., *The Propaganda Model Today: Filtering Perception and Awareness*, University of Westminster Press, London.
- Giglietto, F., Iannelli, L., Valeriani, A., Rossi, L., 2019, “Fake news is the invention of a liar: How false information circulates within the hybrid news system”, in *Current Sociology. La Sociologie Contemporaine*, 67(4), pp. 625–642.
- Giglietto, F., Righetti, N., Rossi, L. e Marino, G., 2020, “It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections”, in *Information, Communication and Society*, 23(6), pp. 867-891.
- Gili, G., 2001, *Il problema della manipolazione. Peccato originale dei media?* Franco Angeli, Milano.

- Hartley, J., 1996, *Popular Reality: Journalism, Modernity and Popular Culture*, Arnold, London.
- Herman, E. S., Chomsky, N., 1988, *Manufacturing Consent: The Political Economy of the Mass Media*, Pantheon Books, New York.
- Higgins, M., 2008, *Media and their Publics*, Open University Press, Maidenhead.
- Iyengar, S., et al., 2019, “The Origins and Consequences of Affective Polarization in the United States”, in *Annual Review of Political Science* 22, pp. 129-146.
- Lelkes, Y., 2016, “Mass Polarization: Manifestations and Measurements”, in *Public Opinion Quarterly* 80(S1), pp. 392–410.
- Livingstone, S., 2004, “What is Media Literacy?”, in *Intermedia*, 32 (3), pp. 18-20.
- Mclaren, P., 1997, “Critical Pedagogy”, in *Teaching Education*, 9 (1). doi: <https://doi.org/10.1080/1047621970090101>
- Metzger, M. J., Flanagin, A., 2013, “Credibility and trust of information in online environments: The use of cognitive heuristics”, in *Journal of Pragmatics*, 59, part B, pp. 210-220. doi: <https://doi.org/10.1016/j.pragma.2013.07.012>
- Morlino, L., Sorice, M., a cura di, 2021, *L'illusione della scelta. Come si manipola l'opinione pubblica in Italia*, Luiss University Press, Roma.
- Pariser, E., 2011, *The Filter Bubble: What the Internet is Hiding from You*, Penguin Press, New York.
- Rega, R., De Blasio, E., Valente, M., 2021, “Il giornalismo italiano alla prova delle piattaforme. La copertura della regolarizzazione dei lavoratori immigrati tra polarizzazione e inciviltà”, in *Comunicazione politica*, 3/2021, pp. 385-412.
- Saurwein, F., Spencer-Smith, C., 2020, “Combating Disinformation on Social Media: Multilevel Governance and Distributed Accountability in Europe”, in *Digital Journalism*, 8:6, pp. 820-841, DOI: [10.1080/21670811.2020.1765401](https://doi.org/10.1080/21670811.2020.1765401)
- Schlesinger, P., 2020, “After the post-public sphere?”, in *Media Culture and Society*, 42(7-8), pp. 1545-1563.
- Sorice, M., 2020a, *Sociologia dei media. Un'introduzione critica*, Carocci, Roma.
- Sorice, M., 2020b, “La piattaforma della sfera pubblica”, in *Comunicazione politica*, XXI (3), pp. 371-388.
- Sorrentino, C., 2009, “Il triangolo imperfetto. La negoziazione fra giornalismo, politica e cittadini”, in *Comunicazione Politica*. 1(2009), pp. 121-130.
- Stanyer, J., 2013, *Intimate Politics: Publicity, Privacy and the Personal Lives of Politicians in Media Saturated Democracies*, Polity, Cambridge.
- Stoppino, M., 2001, *Potere e teoria politica*, Giuffrè, Milano.
- Van Dijck, J., Poell, T., & De Waal, M., 2018, *The platform society: Public values in a connective world*, Oxford University Press, Oxford; tr. it. *Platform Society. Valori pubblici e società connessa*, Guerini, Milano, 2019.
- Wardle, C., Derakhshan, H., 2017, *Information Disorder: Toward an interdisciplinary framework for research and policy making*, Council of Europe, Strasbourg (<https://shorensteincenter.org/wp->

Il disordine informativo e l'odio in rete. Democrazia a rischio

[content/uploads/2017/10/PREMS-162317-GBR-2018-Report-de%CC%81sinformation.pdf?x78124](#)).

Wodak, R., Meyer, M., eds, 2015, *Methods of Critical Discourse Studies*, Sage, London.

Wodak, R., 2020, *The Politics of Fear*, Sage, London.