UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Dipartimento di Scienze Sociali

Dottorato in Scienze Sociali e Statistiche
XXIV ciclo

Tesi di dottorato

Tra computazione e tradizione. Un approccio ibrido alla Content Analysis per lo studio della disinformazione nelle echo chambers di Facebook

Tutor
Chiar.ma Prof.ssa
Enrica Amaturo

Co-tutor
Chiar.ma Prof.ssa
Gabriella Grassia

Coordinatore
Chiar.mo Prof.
Roberto Serpieri

Candidata
Suania Acampa

Anno Accademico 2021-2022

“Lo scandalo, ai nostri giorni,
non consiste nell'attentare ai valori morali,
bensì al principio di realtà.”
Jean Baudrillard, 1991

INTRODUZIONE

Le menzogne sono vecchie quanto il mondo, ma negli ultimi anni le notizie false, la cattiva informazione e le teorie del complotto si sono diffuse con una facilità e una pervasività tali da esser diventate un pericolo per la tenuta democratica. I *social media* hanno radicalmente cambiato il modo in cui comunichiamo, discutiamo con gli altri e accediamo alle informazioni. Con l'avvento delle piattaforme gli editori hanno dovuto affrontare il difficile passaggio al mondo digitale, in cui le procedure di selezione dei contenuti - prima eseguite secondo le logiche gerarchiche della redazione – diventano molto più complesse e sono costrette a rispondere a logiche prima d'ora mai considerate. Questo cambiamento è spesso attribuito al fenomeno della disintermediazione che ha reso i *social media* un palcoscenico globale per chiunque abbia qualcosa da dire. Per alcuni versi questa considerazione non è esatta: la quantità di informazione a cui è possibile accedere oggi è così enorme che la mediazione è indispensabile; a farla però non sono più le redazioni giornalistiche ma i nostri gusti e gli algoritmi di *feed* che ci propongono contenuti adeguati ai nostri gusti. I *social media*, nati per intrattenere, sono ottimizzati per catturare l'attenzione dell'utente che diventa la moneta di scambio del nuovo ecosistema informativo. Per questo motivo, l'esperienza dell'utente sulle piattaforme è finalizzata ad essere permanente e interattiva. Il cortocircuito inizia quando questo nuovo ecosistema diventa al tempo stesso: fonte d'accesso alle informazioni, spazio d'intrattenimento, vetrina di *business* e piazza in cui discutere. Questa pluralità e combinazione di spazi diversi ha prodotto una serie di effetti collaterali come la diffusione di disinformazione, la nascita di spazi iper-partigiani, modelli di *business* basati su *clickbait*, propaganda estremista e complottista. In ognuna delle direzioni in cui si è mossa la comunità scientifica nel tentativo di comprendere e contrastare il fenomeno disinformativo ha incontrato dei limiti. A partire da questi limiti muove i passi questo lavoro di ricerca: il punto di partenza è una riflessione teorica che mira ad inserire la disinformazione in un'ampia cornice fenomenologica che non si limita a indagare il fenomeno a partire dall'avvento delle nuove tecnologie, ma ricostruisce quegli elementi storici, tecnologici e umani che hanno contribuito (e contribuiscono ancora oggi) alla buona riuscita delle strategie disinformative. Gli stessi elementi fanno emergere forti dubbi sulle soluzioni automatiche messe a punto dai colossi digitali e su quelle adottate dalle redazioni giornalistiche internazionali per far fronte al problema. Ponendo particolare attenzione al potere che le piattaforme algoritmiche esercitano sulla selezione e diffusione dei contenuti, il lavoro che qui si presenta pone l'obiettivo di indagare la disinformazione dal punto di vista delle strategie che utilizza, dei temi di cui si nutre e dei processi culturali e comunicativi che innesca tra gli utenti che la consumano, andando oltre un approccio esclusivamente *data driven* che tende a considerare questi ultimi come un soggetto unico dal comportamento indistinto. Gran parte del lavoro è stato dedicato

allo studio del concetto di disinformazione e alla messa a punto di un processo di operativizzazione capace di far comunicare la profondità delle categorie con la scalabilità dei dati, provando così a superare i limiti concettuali presenti in letteratura.

La ricerca si inserisce nell'approccio metodologico dei *Digital Methods* e nella prospettiva epistemologica digitale proposta da Enrica Amaturo e Biagio Aragona (2019). Per rispondere agli obiettivi dell'analisi, lo studio assume una posizione metodologica pluralistica, pragmatica e al tempo stesso critica: in ottica *"follow the medium"* (Rogers, 2013) lo studio si ispira al modo in cui i dati sono naturalmente organizzati dalle infrastrutture tecniche degli ambienti digitali ma tenta di andare oltre le informazioni preconfezionate restituite dalle procedure di collezione dei dati, in modo da evidenziare i limiti, le risorse e le soluzioni adottate nel portare avanti un'indagine su un fenomeno completamente digitale.

Il percorso di ricerca si basa sull'utilizzo di tecniche di analisi tradizionali e computazionali in una prospettiva ibrida capace di trovare soluzioni metodologiche ogni qual volta ci si scontra con problemi connessi ai dati non strutturati, al funzionamento delle tecniche di *machine learning* o alle infrastrutture tecnologiche. Il percorso di metodo proposto combina il rigore sistematico e la consapevolezza contestuale dell'analisi del contenuto tradizionale con l'efficienza delle tecniche computazionali. L'analisi è stata condotta su due livelli, in ognuno dei quali è presente un momento ermeneutico e uno computazionale: il primo livello focalizza l'attenzione sulla trasmissione del messaggio, mentre il secondo livello sulla risposta degli utenti. Questa articolazione nasce dall'esigenza di inserire nella pratica di ricerca una soluzione in grado di far luce su un fenomeno che è già di per sé multilivello: post e commenti. Mentre il momento ermeneutico è stato fondamentale per far emergere aspetti particolari del fenomeno indagato, quello computazionale è stato necessario per estendere quegli aspetti su larga scala. La ricerca è interamente condotta su *user generated data* raccolti grazie all'accesso all'API della piattaforma indagata. Questi dati semi strutturati e non strutturati provenienti dalle piattaforme forniscono l'opportunità di osservare i comportamenti, le relazioni e le pratiche sociali che emergono quando le persone non sanno di essere indagate, ottenendo così un livello di autenticità non riscontrabile nei dati raccolti attraverso intervista (Lupton, 2014). Con il percorso di metodo proposto si può dire di essere gradualmente entrati in un'*echo chambers* e i risultati ottenuti hanno fornito un ulteriore aspetto interpretativo al fenomeno disinformativo. Inquadrandola in termini di spaccatura ideologica degli utenti, la disinformazione diviene elemento di frattura della sfera pubblica digitale; una specie di oggetto irrazionale che circola in uno spazio a lungo ritenuto governato da norme di coinvolgimento razionali (Habermas, 1974). Le posizioni politiche populiste diventano il meccanismo guida degli utenti nel complesso universo disinformativo: gli elementi della

comunicazione populista, prepotentemente emersi durante l'analisi del messaggio disinformativo vengono estremizzati dagli utenti nelle loro pratiche di consumo e interazione.

Nella sua applicazione, ciò che ci restituisce l'approccio proposto è un inedito percorso di metodo modellato esclusivamente per la ricerca digitale e dunque volto ad affrontare le sfide che questa pone, pur basandosi su strumenti analitici ampiamente radicati nelle scienze sociali.

CAPITOLO I - Le condizioni di contesto. I fattori che hanno contribuito alla buona riuscita della disinformazione

Osservare il fenomeno della disinformazione prendendo in considerazione solo la forma e l'entità che esso ha assunto nell'ultimo decennio sarebbe riduttivo oltre che fuorviante. Generalmente, la letteratura sul fenomeno analizza la disinformazione seguendo l'evoluzione mediale della rete e l'avvento delle piattaforme, che senza ombra di dubbio sono stati elementi determinanti nel rendere il fenomeno un problema di rilevanza mondiale, ma questo esclude tutta una serie di fattori che hanno contribuito alla radicazione del fenomeno e all'attuale buona riuscita delle campagne disinformative. Per questo motivo, il lavoro parte dall'inserimento della disinformazione in una cornice fenomenologica più ampia che tiene insieme tre fattori: quello storico, quello tecnologico e quello umano. Il fattore storico è necessario per spiegare come le caratteristiche del postmodernismo abbiano gettato le basi per un distacco del pubblico dall'autorevolezza dell'informazione professionale, con la conseguente migrazione verso canali informativi alternativi, che hanno favorito la nascita di una vera e propria sfera pubblica alternativa in cui trovano terreno fertile le teorie cospirazioniste. Il secondo fattore è quello tecnologico, che ha reso la disinformazione (fenomeno esistente già prima dell'avvento della rete) un vero e proprio rischio per le democrazie occidentali: l'aumento dell'offerta informativa ha frammentato l'*audience*, le logiche delle piattaforme hanno influenzano le tradizionali pratiche editoriali di confezionamento delle notizie e disintermediato la loro produzione e distribuzione. Attribuendo un ruolo sempre più rilevante all'utente *prosumer* e alle sue preferenze, l'elemento tecnologico ha promosso la semplificazione dei contenuti informativi e favorito (attraverso il suo funzionamento) ambienti omofili in cui la disinformazione prolifera con conseguente polarizzazione delle opinioni. L'economia postindustriale, altro elemento distintivo del postmodernismo, ha determinato una massiccia produzione di contenuti informativi tale da generare negli utenti un'*information overload*[1] con conseguenti ripercussioni sull'attenzione, sulla comprensione e sulla verifica dei fatti. Inoltre, la tendenza tutta umana di non uscire dalla propria *comfort zone* di credenze condivise ha generato una sorta di partigianeria nell'esposizione da parte dei soggetti a particolari tipi di informazione e allo stesso tempo ha favorito negli stessi la tendenza inconsapevole a credere acriticamente ad un tipo di contenuto che non discordi da quelle stesse credenze. L'essere umano si manifesta così in tutta la sua meravigliosa natura: imperfetta, multiforme e spesso irrazionale. Per comodità questi fattori saranno trattati separatamente nei paragrafi che seguono ma è chiaro che nessuno prescinde dall'altro. Infine, saranno evidenziate quelle cause che hanno contribuito al

[1] *Information overload* si verifica quando l'informazione a disposizione eccede la capacità dell'utente di elaborarla (Klapp, O. E. 1986)

deterioramento dei confini già fragili dell'autorità giornalistica, rendendo sempre più difficile la distinzione tra giornalismo e ciò che giornalismo non è.

1 La fine delle grandi narrazioni e la legittimazione del singolo. L'elemento storico

Dalla fine degli anni Sessanta del XX secolo inizia a rendersi evidente la traccia di una crisi profonda della modernità, segnata da particolari cambiamenti economici, sociali e culturali. Per definire queste profonde trasformazioni viene formulato il concetto di postmoderno: un movimento ideale sviluppatosi in quegli anni che ci ha fornito indicazioni sulle coordinate reali del tempo complesso in cui viviamo. Questa crisi della modernità ha portato a compimento alcuni dei processi iniziati nell'epoca moderna e altri, invece, li ha radicalizzati a tal punto da metterne in discussione le stesse premesse (Beck, 1997[2]).

Il primo radicale cambiamento riguarda l'organizzazione produttiva. In campo sociologico, il termine postmoderno viene spesso accompagnato da quello complementare di postindustriale (Bell, 1973), impiegato per descrivere le complesse trasformazioni della società a seguito all'avvento delle nuove tecnologie informatiche e telematiche in ogni ambito della vita dell'uomo. La società moderna è giunta al culmine del processo di industrializzazione e per continuare a sopravvivere deve necessariamente concentrare i propri sforzi produttivi verso beni immateriali come servizi e informazioni. In campo economico il termine "informazione" assume un'accezione molto ampia: è informazione tutto ciò che può essere digitalizzato, immagazzinato e distribuito come sequenza di *bit*[3]: in questo tipo di merce ciò che conta è il patrimonio delle informazioni e il valore che esse possono avere sul mercato. La trasformazione del quadro economico del mondo occidentale è stata accompagnata dalla liberalizzazione delle telecomunicazioni e dalla diffusione massiccia di Internet; queste caratteristiche assunte dall'economia hanno così determinato la nascita della cosiddetta società dell'informazione (Castells, 2000). Castells con questo termine non intende sottolineare il semplice ruolo che generalmente ricopre l'informazione nella società perché, come egli stesso scrive (p.21), nel suo senso più ampio l'informazione (intesa come comunicazione di conoscenza) è stata fondamentale in tutte le società, da quella moderna a quella medievale. Con la definizione di "società dell'informazione" Castells vuole indicare l'attributo di una specifica forma di organizzazione sociale in cui la generazione, l'elaborazione e la trasmissione dell'informazione diventano fonti fondamentali di produttività e potere a causa delle nuove condizioni tecnologiche emerse in un particolare periodo

[2] Di preciso Beck parla di una "tarda modernità" che ha portato all'estremo i processi della modernità.
[3] Dai libri alle banche dati, dalle riviste ai film, dalla musica alle quote azionarie, dai blog alle pagine web, ecc.)

storico. Cercando di stabilire la stessa distinzione tra industria e industriale, Castells sottolinea come la società dell'informazione non è solo una società dove c'è informazione ma una società dove le forme sociali e tecnologiche di organizzazione dell'informazione colmano tutte le sfere di attività: a cominciare da quelle dominanti fino a raggiungere gli oggetti e le abitudini della vita quotidiana. Partendo da queste premesse, si delineano tre elementi caratterizzanti della società dell'informazione: la conoscenza come base dell'economia e come fonte principale di crescita e di produttività; la valorizzazione di attività, occupazioni e professioni ad alto contenuto di informazione e conoscenza (dato il moltiplicarsi delle fonti e dei canali informativi) e infine, il trasferimento delle attività economiche dalla produzione di beni alla prestazione di servizi, quindi dalla produzione di risorse materiali a quello di risorse simboliche (Castells, 2008). Già Baudrillard[4] (1976) aveva individuato nella *«rivoluzione strutturale della legge del valore»* (p. 18) il fenomeno che avrebbe interessato la fine della modernità, per cui si sarebbe passato da un tempo (quello dell'economia classica) in cui il valore rinvia comunque al prodotto nella sua materialità, a un tempo in cui l'immagine della merce si svincola dalla merce stessa e non rinvia più a nulla se non a se stessa. Il trionfo delle merci simboliche è ciò che caratterizza l'attuale ecosistema digitale e che ha affiancato al potere economico, politico, coercitivo, il potere simbolico, ossia la *«capacità di intervenire sul corso degli eventi, di influenzare le azioni degli altri e di creare avvenimenti producendo e trasmettendo forme simboliche»* (Thompson, 1995 p. 31).

La crisi della modernità determina un radicale mutamento anche nel clima culturale. L'etichetta di "società postmoderna" richiama la riflessione di J.F. Lyotard (1979) secondo il quale la condizione postmoderna è il risultato della fine delle grandi narrazioni che hanno accompagnato la modernità. L'analisi di Lyotard coinvolge in prima battuta l'ambito della scienza ma è una riflessione che si allarga progressivamente ad altre sfere del sapere. Rispetto al secolo precedente, il Novecento presenta alcune peculiarità: l'affermazione su larga scala del capitalismo in campo economico, il progressivo sfaldamento del comunismo, ultimo elemento ideologico del secondo dopoguerra, la trasformazione della società da un tutto omogeneo a una realtà plurale, frammentata, sempre più individualista. Nel momento in cui Lyotard scrive, le narrazioni che l'occidente ha costruito per rappresentare se stesso, per raccontarsi e quindi per legittimarsi, sono oramai al tramonto. Al centro dell'analisi lyotardiana vi è dunque la questione della legittimazione del sapere e il modificarsi del suo statuto nell'era postindustriale e nella cultura postmoderna: i sistemi filosofici della modernità, così come le ideologie novecentesche, hanno avuto l'obiettivo di spiegare il mondo e di confermare un certo modo di

[4] La riflessione del sociologo francese appare in questo caso profetica, in quanto negli anni Settanta del secolo scorso non era possibile nemmeno immaginare quella che sarebbe stata la portata e il ruolo assunto dal digitale.

interpretare la realtà; ma ognuno di questi grandi racconti del moderno è stato poi confutato dalla storia, per così dire *«invalidato nel suo fondamento»* (ivi, p.38). La razionalità del reale è stata confutata con Auschwitz *«questo crimine, che è reale, non è razionale»* (ibidem), mentre la rivoluzione proletaria è stata confutata da Stalin; la dottrina del liberalismo parlamentare e l'ideale di emancipazione della democrazia sono stati confutati con i movimenti del '68 quando *«il sociale quotidianamente mette in crisi l'istituzione rappresentativa*» (ibidem). Infine, la dottrina del liberalismo economico e la validità dell'economia di mercato in cui *«tutto ciò che è libero gioco della domanda e dell'offerta favorisce l'arricchimento generale»* (ibidem) sono state confutate nelle numerose crisi[5] del sistema capitalistico. Così le grandi narrazioni della modernità perdono di credibilità, mettono in discussione i presupposti di razionalità proclamati dall'Illuminismo e non lasciano spazio a nessun altro tipo di sapere universale. La condizione postmoderna si caratterizza dunque per lo scetticismo e il richiamo alla natura contingente della conoscenza e dei sistemi valoriali, intesi come prodotti di supremazie politiche o culturali. Da questo momento si diffonde una crescente sfiducia nei confronti della scienza e trionfano tendenze autoreferenziali, relativiste e pluraliste. Con il postmodernismo cambia anche la concezione del rischio: da entità calcolabile e gestibile poiché prodotta dall'industrializzazione, ad entità invisibile, immateriale, globale e soprattutto incontrollabile (come i rischi ambientali, economici e tecnologici). Nel suo rapporto con i media, la fine delle grandi narrazioni ha disgregato il confine tra rappresentazione del mondo e realtà del mondo stesso[6], i media hanno trasformato la realtà in *reality* (Ferraris, 2012) rafforzando le percezioni degli individui attraverso l'adozione di pratiche comunicative che hanno giustapposto fatti a opinioni, scienza ad esperienza, fatti a ricostruzione dei fatti. Questi tratti salienti della condizione postmoderna si collegano immediatamente anche alla crisi dell'autorità politica: venendo meno la struttura del partito, la forma di legittimazione diventa il politico stesso, il capo carismatico di cui ci si fida sulla base del processo mediatico che lo figura. A questo punto la domanda cruciale per Lyotard diventa la seguente: dove risiede la legittimità dopo la fine delle metanarrazioni?

Potremmo dire che nella società contemporanea, in cui la dissoluzione delle certezze moderne ha portato alla relativizzazione dei punti di vista, per cui ogni parere vale quanto un altro, il sapere è legittimato non più dall'autorità ma dal singolo. Negli ambienti digitali questo coincide spesso con il proprio gruppo di riferimento con il quale si condivide un particolare sistema valoriale. In mancanza del riconoscimento di un'autorevolezza epistemica, quel che sembra legittimare la conoscenza è quindi

[5] Secondo Lyotard, le crisi del 1911 e del 1929 confutano la dottrina del liberalismo economico mentre quelle degli anni 1974 e 1979 confutano la versione post-keinesiana di essa.

[6] Eclatante fu l'affermazione di Jean Baudrillard riguardo una guerra del Golfo mai accaduta nella realtà e progressivamente ridotta a pura rappresentazione mediatica.

l’esperienza diretta, una sorta di *vox populi* custode delle credenze dei singoli, che frammenta, moltiplica e spettacolarizza le realtà. La spettacolarizzazione e il richiamo alle credenze personali caratterizza un concetto molto caro negli studi sulla disinformazione, quello di post-verità, che tende a configurarsi sempre più come un elemento costitutivo del postmodernismo. Il concetto di post-verità è stato spesso usato (e abusato) in diversi campi del sapere; ma andando oltre le tradizionali definizioni, è utile considerare la post-verità nell’ottica semiotica di Lorusso (2018), ossia come un regime discorsivo in quanto riguarda i modi in cui, attraverso le pratiche discorsive, costruiamo la realtà. Queste pratiche discorsive si caratterizzano per un forte appello all'emotività e alle credenze diffuse che tendono a sostituirsi alle verità consolidate. Il lavoro del filosofo francese Michel Foucault (1971) sembra particolarmente attuale in riferimento al concetto di post-verità perché riguarda principalmente la conoscenza, la verità e il potere. Egli ha sostenuto che il linguaggio, nelle strutture che lo sostengono, aiuta a modellare il modo in cui vediamo le cose; le parole contano perché inquadrano il dibattito e come comprendiamo il mondo. Oggi Foucault sarebbe affascinato dai *social media* e da come l'autore un post possa non esistere alla vista di chi lo legge, ma di quanto il suo messaggio possa assumere una vita autonoma e autoreplicante. Foucault insegna che il discorso opera come una modalità di potere che costituisce e contesta i confini sia delle nostre realtà materiali, sia dei significati che usiamo per plasmarle e dar loro un senso (Schulte, 2013). In quest’ottica, non si può prescindere dal considerare la post-verità nel suo rapporto con i *media digitali*, perché è nei media digitali che oggi si producono e consumano la maggior parte delle nostre pratiche discorsive. Questo ha incredibilmente frammentato la sfera pubblica contemporanea e l’ha resa dipendente dalle infrastrutture tecnologiche che le piattaforme proprietarie controllano (Fuchs, 2014). Per questo i media non possono più esser pensati come semplici spazi di rappresentazione della realtà ma come veri e propri luoghi di formazione del reale: una costruzione a cui partecipiamo attivamente e che negoziamo continuamente con il dispositivo tecnologico attraverso il quale entriamo in contatto con informazioni che orienteranno poi le nostre scelte. Nella società dell’informazione i media digitali hanno portato alle estreme conseguenze le caratteristiche culturali del postmoderno: lo sgretolamento dell’autorità ha elevato il singolo a portavoce di conoscenza che spesso coincide con la propria esperienza dei fatti. In questo modo la realtà in cui si crede diventa plurima e spesso contraddittoria. Dunque, la disinformazione non prolifera (come accadeva prima dell’avvento della rete) per sottrazione o negazione di informazione, ma per moltiplicazione ed eccesso; in questo la rete e i *social network* hanno avuto un ruolo determinante nell’abbattere le barriere classica di fruizione dell’informazione. I media digitali hanno così estremizzato quella frammentazione sociale avviata con la fine della modernità, la quale ha lasciato spazio ad un pluralismo radicale nel cui orizzonte non c’è possibilità di manovra per

legittimazioni totalizzanti del sapere; piuttosto si fanno largo giustificazioni individuali, parziali e differenziate a dar un senso alla complessità del reale.

2 Disintermediazione e reintermediazione algoritmica. L'elemento tecnologico

Alle condizioni di contesto storico, si aggiungono quelle del contesto tecnologico. Il web 2.0 ha guidato il passaggio dal modello informativo classico, in cui le notizie sono mediate da esperti giornalisti e distribuite da testate *mainstream,* ad un modello informativo nuovo, caratterizzato dall'ambiente disintermediato di piattaforme e motori di ricerca e re-intermediato dagli algoritmi che ne regolano il funzionamento. L'evidente cambiamento che le tecnologie proprie del *web 2.0* hanno determinato in termini di forma e contenuto delle notizie è ragionevolmente guidato dal mercato e riflette da un lato l'incapacità dei giornali di dotarsi di un modello di *business* che gli consenta di competere con la gratuità e la velocità della rete, dall'altro il progressivo adattamento della produzione di informazione ai principi organizzativi che guidano l'ecosistema delle piattaforme. L'effetto composto di questi due fattori sta ridisegnando l'intero ecosistema dell'informazione. Le campagne disinformative sfruttano in particolare due elementi costitutivi del *web 2.0:* la graduale entrata degli utenti nel processo produttivo delle notizie e il meccanismo di distribuzione dei contenuti ad opera della *black box* algoritmica. Il modello *top-down*, che prevede l'acquisizione di notizie spostarsi dall'alto verso il basso, ha lasciato il posto a un modello orizzontale in cui gli attori sociali si trasformano in soggetti ibridi: smettono di essere semplici consumatori di notizie e diventano *prosumer* (Ritzer & Jurgenson, 2010), dando vita a ciò che Castells (2017) definisce l'epoca dell'autocomunicazione di massa.

2.1 Il ruolo degli utenti nel processo di creazione delle notizie

Anche se il giornalismo resta vivo come elaborazione simbolica della realtà e finalizzato a legittimare una valida rappresentazione del mondo, oggi i giornalisti non più i soli in grado di attivare la macchina delle notizie: la rete e i dispositivi tecnologici permettono anche ai lettori di produrre contenuti utilizzando strumenti e linguaggi che fino a poco fa erano di esclusiva competenza del mondo dell'informazione. Nonostante gli studi sul giornalismo abbiano tardato a indagarne gli effetti, la questione legata al rapporto tra chi produce informazione di professione e chi dovrebbe solo fruirla è uno degli elementi che confonde i confini dei due ruoli e mette in discussione alcune distinzioni analitiche da sempre date per scontate (Hartley, 2009). L'origine della progressiva imposizione del

lettore nel lavoro giornalistico va ricercata a partire dalle pratiche del *Citizen Journalism,* ossia quell'insieme di attività di condivisione delle informazioni da parte dei cittadini diffuse alla fine degli anni Novanta. Queste pratiche sono figlie della rete e della sua cultura partecipativa, nonché del progresso tecnologico che permette a tutti di avere a portata di mano strumenti per registrare, fotografare e condividere informazioni. L'*input* dei cittadini ha consentito un processo di creazione di notizie fluido e orizzontale: i *blog* sono esplosi rapidamente e i giornalisti cittadini hanno svolto il ruolo chiave di testimoni nel catturare e condividere, in tempo reale, eventi tragici come disastri naturali (Tsunami del 2004 nel sud-est asiatico) o attacchi terroristici (gli attentati di Londra del 2005, gli attentati di Boston del 2013), prima che i giornalisti potessero raggiungere i luoghi interessati, rendendosi così attori chiave del processo informativo. I *blog*, già dopo i primi anni di vita, avevano prodotto nel mondo dell'informazione un cambio di paradigma, avvicinando cittadini e giornalisti, combinando commenti, conversazioni e informazione (Splendore, 2017). Le rubriche "lettere al direttore" dei giornali francesi e inglesi, di cui parlava Walter Benjamin nel 1937[7] (che per primo aveva captato questa graduale perdita di distinzione tra autore e lettore) sono diventati oggi i *social media*, in cui sono migrate molte delle attività tradizionali di *Citizen Journalism*. Con i *social network* si è dunque ridotta al minimo la tradizionale separazione tra chi produce e chi consuma la notizia. Le piattaforme hanno consentito la presa di parola di chiunque fosse intenzionato a prenderla, avviando così una produzione di massa indistinta di informazioni e contemporaneamente dando vita a forme di disintermediazione di quei soggetti che fino a quel momento sono stati i mediatori "naturali" della pratica di informare (Bentivegna e Boccia Artieri, 2019). Questa partecipazione alla creazione delle notizie associa la dimensione del consumo a quella della produzione dei contenuti, rendendo gli utenti dei *prosumers*. Christian Fuchs (2012) è stato tra i primi ad indagare il fenomeno del prosumerismo sulle piattaforme e lo ha fatto in relazione alle teorie di Karl Marx: squarciando il velo di Maya di una rete libera e partecipativa, capace di far emergere una sfera pubblica digitale, Fuchs considera le piattaforme come un sistema di sfruttamento del lavoro e accumulo del capitale. Con l'illusione della gratuità, gli utenti sono sfruttati dalle multinazionali dei *new media* per la continua produzione di contenuti e di dati che tracciano le loro attività, elementi indispensabili per la sopravvivenza stessa delle piattaforme[8]. Mentre nel *mainstream,* ad esser venduta come una merce ai pubblicitari è l'*audience*, con l'emergere degli *user-generated contents,* ad essere venduti come merce sono gli utenti stessi.

[7] Nel 1937, Walter Benjamin, osservando gli effetti della rubrica "Lettere al Direttore" dei giornali francesi e inglesi, evidenziò la perdita della distinzione di ruoli tra autore e lettore e la graduale presa di parola di quest'ultimo.
[8] Il continuo lavoro gratuito dei *prosumer* è considerato da Fuchs come il risultato del trasferimento del lavoro produttivo agli utenti, questi massimizzano il plusvalore affinché i profitti dei *new media* aumentino e il capitale sia cumulabile.

Secondo Bentivegna e Boccia Artieri (2019) gli utenti, durante la propria attività di *prosumer*, non concepiscono come "prodotti" i propri contenuti o le relazioni sociali che sviluppano attraverso le pratiche di condivisione, piuttosto interpretano il proprio ruolo come quello di un individuo che, partecipando all'ambiente mediale, entra in rapporto con la propria comunità di riferimento Queste modalità di creazione collaborativa di contenuti coinvolgono grandi comunità di utenti che agiscono senza un controllo e un coordinamento gerarchico ma operano secondo linee fluide, flessibili, eterarchiche e organizzate *ad hoc* per rispondere agli attuali modelli di produzione (Bentivegna e Boccia Artieri, 2019). Questi modelli di produzione favoriscono la circolazione di contenuti virali indipendentemente dal fatto che contengano informazioni devianti, approssimative o completamente false. Gli utenti *prosumer* sono attori fondamentali nel processo di disintermediazione dell'informazione innescato dalle piattaforme, questo ha dato potere e autonomia all'utente e deteriorato i confini di chi può ritenersi legittimamente produttore di informazione. La disintermediazione però non risiede solo nella produzione dei contenuti informativi ma anche nelle pratiche di consumo: mentre le redazioni tradizionali offrono un pacchetto informativo che fornisce al lettore una copertura accurata ed estesa, che spazia dall'inchiesta all'intrattenimento, con l'avvento delle piattaforme e dei motori di ricerca, gli utenti consumano notizie isolate, accedendo al singolo contenuto di interesse aggirando le tradizionali logiche di distribuzione editoriali. Questa pratica compromette la capacità delle testate di generare ricavi: il prodotto passa dall'essere il "pacchetto informativo" all'essere la singola notizia, consumata in maniera disaggregata e che vive o muore in base alla propria capacità di monetizzare.

2.2 L'informazione *data-driven* e il ruolo degli algoritmi

Ogni disintermediazione prevede inevitabilmente una re-intermediazione, in questo caso operata dagli algoritmi. I social media e i motori di ricerca sono tra i più determinanti fattori di cambiamento della professione giornalistica e i giornalisti hanno assunto in maniera estensiva l'uso dei social media per raccogliere occorrenze da trasformare in *news* (Splendore, 2017). La re-intermediazione algoritmica influenza profondamente il processo informativo in ciascuna delle fasi di produzione (raccolta, selezione, verifica) e distribuzione delle notizie. Partendo dalla produzione, le piattaforme proprietarie che operano sul piano internazionale, tendono indirettamente a fissare degli *standard* globali rispetto ai contenuti condivisi, lo fanno attraverso l'arma del posizionamento dei contenuti nel *feed* degli utenti.

Nella distribuzione dei contenuti, le piattaforme in generale e Facebook[9] in particolare, tengono conto di numerose variabili (ponderate in maniera diversa) che si sommano in un singolo punteggio tale da definire l'ordinamento dei post nel *news feed* della piattaforma, determinando così quali contenuti saranno visibili o meno all'utente che vi accede. Questo unico sistema di punteggio onnicomprensivo viene utilizzato per classificare e ordinare vaste aree di interazione umana nei *social media*, in quasi tutti i paesi del mondo e sicuramente in tutte le lingue.

La stampa *mainstream* invece, a lungo intrappolata tra la carta e gli spazi digitali, ha per molto tempo trasposto le classiche pratiche editoriali nell'ecosistema delle piattaforme, questa strada si è rivelata presto inefficace a rispondere agli interessi e al tipo di attività svolta dagli utenti negli ambienti digitali. La preoccupazione di rispondere alle esigenze dell'*audience* e al posizionamento nei propri contenuti nel *news feed* delle piattaforme ha costretto le redazioni a trasformare le proprie pratiche informative per adattarle al continuo monitoraggio delle preferenze dei propri lettori. Gli utenti, nel consumare informazione, lasciano traccia delle proprie attività che vengono poi capitalizzate dalle piattaforme attraverso servizi di misurazione e analisi dei dati. I servizi di misurazione dell'*audience* forniscono all'industria giornalistica indicazioni molto dettagliate sul tipo di utente, sulle modalità e sui contenuti che maggiormente lo coinvolgono, queste informazioni trasformano le *dashboard* dei social media, in vere e proprie bussole per orientare le redazioni nel confezionamento della notizia (dalla titolazione, alle immagini, alle *keyword*) affinché sia ben ottimizzata sui motori di ricerca e produca il giusto *engagement* nelle piattaforme *social media*. Una buona agenzia stampa però è quella che distribuisce le notizie che dovremmo sapere, non quelle che vogliamo sapere. Come vedremo, non è certo priva di faziosità l'industria editoriale: nella propria *agenda setting* anche la stampa tradizionale seleziona le notizie a cui intende dare più rilievo, ma il rischio di imbattersi in un'informazione completamente aderente alle proprie preferenze personali è drasticamente ridotto nell'universo *mainstream*. Inoltre, ad una testata di una fazione ne corrisponde sempre una dell'altra, questo fornisce al lettore una spiegazione della complessità del reale, certamente modellata sulle gerarchie valoriali dell'editore, ma sicuramente più competitiva nel dibattito pubblico (Lorusso, 2018).

Il sistema di targhettizzazione delle notizie è estremizzato nelle testate native digitali: un classico esempio è quello dell'esperienza BuzzFeed che concentra la propria produzione esclusivamente su contenuti che rispondono alla logica della viralità e dei *topic* in tendenza. Lo fa raccogliendo sistematicamente i dati originati dai propri utenti, combinandoli con le metriche della piattaforma e grazie al supporto di una squadra di *data scientist,* che processa tutto attraverso algoritmi di *machine*

[9] La comprensione del funzionamento del *news feed* di Facebook è approfondita nei capitoli che seguono e partono dall'inchiesta del Wall Street Journal titolata "The Facebook File". L'inchiesta è l'unica risorsa attualmente disponibile per provare a comprendere il funzionamento dell'algoritmo della piattaforma.

learning in grado di mappare le relazioni tra variabili e condivisione (Van Dijck et al., 2018). Essendo che è la cattura dell'attenzione e dell'emozione dell'utente a render virali i contenuti (che si tradurranno poi in guadagno) e che solo alcuni di essi hanno la probabilità di diventare virali, gli editori hanno messo su una vera e propria produzione di massa di contenuti informativi per assicurarsi che almeno una percentuale su tutti generi il giusto *engagement* (ivi). Il contino *user feedback check* porta alle estreme conseguenze quel "*clash of cultures"* di cui parlano Alfred Hermida e Neil Thurman (2008) cioè quel confronto, scontro, negoziazione tra giornalisti e lettori che impone significative trasformazioni nella logica, nelle pratiche e nel concetto stesso di notizia. Questo scontro di culture mostra la vulnerabilità dei giornalisti che, nelle pratiche di interazione con il lettore e di normalizzazione dei loro flussi comunicativi nelle proprie prassi, mettono in discussione la propria autorevolezza e indeboliscono la propria credibilità (Splendore, 2017). Risultato di questa contesa è un continuo rimodellamento dei confini del giornalismo in cui i consumatori acquisiscono credibilità tale da trasformare semplici commenti in un contenuto informativo alternativo al prodotto giornalistico.

Un più alto livello di interazione con il consumatore determina inevitabilmente una maggiore difficoltà nell'ignorare le sue richieste. Dunque, quando si fa riferimento ad un tipo di informazione *data-driven*, non s'intende l'analisi dei dati all'interno degli articoli *mixata* alle diverse modalità di *storytelling*, quanto alla pervasività che i dati hanno nell'orientare le pratiche di selezione e produzione delle notizie da parte dei giornalisti e che riguardano, ad esempio, la scelta di tenere o meno una notizia in primo piano o la scelta di approfondire alcune tematiche a discapito di altre. In generale, sono i dati a definire l'agenda degli editori e questo dimostra quanto la necessità di accrescere il traffico ai *siti web* delle testate, determini l'adozione di logiche che, prima della digitalizzazione, non rientravano affatto nelle prassi giornalistiche. Infine, il fatto che piattaforme come Facebook forniscono questi servizi di *audience analysis* in modo gratuito agli editori[10], evidenzia quanto è importante per le società proprietarie di diventare un nodo centrale nella produzione, circolazione e mercificazione di notizie, regolate rispetto alle attività degli utenti a loro volta disciplinati dall'algoritmo (Van Dijck et al., 2018). In questo modo, la selezione dei contenuti - che ha sempre rappresentato l'essenza della professione giornalistica - poiché frutto di una valutazione rispetto ai valori pubblici sul piano sociale, politico e culturale - è oggi affidata al funzionamento poco chiaro della *black box* algoritmica (ivi).

[10] Nel 2016 Facebook acquista CrowdTangle e lo offre gratuitamente agli editori; questo è un tool che permette alle testate giornalistiche di tracciare il modo in cui i propri contenuti e quelli dei *competitor* si diffondono nelle piattaforme di proprietà del gruppo Zuckerberg.

2.3 La McDonaldizzazione della produzione delle notizie

La frammentazione dell'*audience* e la distribuzione algoritmica delle notizie costringe i giornali a riorganizzare il proprio modello di *business*. La pubblicità ha storicamente sovvenzionato la produzione dei contenuti mediali: Paolo Mancini (2000) scrive che l'industria dell'informazione italiana trova sollievo con l'esplodere della pubblicità come principale modello di *business*, dato dall'avvento della televisione commerciale negli anni Ottanta del secolo scorso. Ciò che differenzia i giornali e le società mediali di quegli anni e che questi connettevano direttamente il lettore, lo spettatore o l'ascoltatore agli inserzionisti pubblicitari, esercitando quindi pieno controllo sul mercato pubblicitario (Van Dijck et al., 2018). Oggi, questo settore ha subito (al contrario) una mediazione da parte dei giganti digitali che si impongono come mercati multilaterali (Alleman et al. 2019), ossia: dal lato dell'utente i servizi offerti sono gratuiti, mentre dal lato degli editori i ricavi sono determinati dalla pubblicità, consumata insieme alle notizie e gestita dal sistema economico delle inserzioni. Questa commistione tra informazione e pubblicità si rileva un altro fattore che ha contribuito a rinegoziare i confini della notizia: questa perdendo i suoi tratti distintivi, rende poco agevole la distinzione tra un contenuto informativo professionale da un contenuto "attira *click*", oltre al fatto che una notizia ufficiale presentata con un continuo rimando pubblicitario mina il prestigio della fonte. Come sottolinea Splendore (2017), prima dell'avvento delle piattaforme la pubblicità, anche nelle imprese editoriali in cui se ne faceva massiccio uso, era sempre ben distinta sia nel prodotto informativo sia nell'organizzazione delle redazioni, in cui la parte *manageriale* era differenziata da quella dei giornalisti. Nel *native advertising* la divisione dei due ambiti è invece sostituita dalla loro mescolanza e la pubblicità viene inserita direttamente nel flusso dell'informazione. Oggi, i *Social Ads* rappresentano una tipologia di pubblicità interattiva basata su inserzioni e veicolata esclusivamente all'interno dei *social network*. I prezzi delle inserzioni pubblicitarie sono monopolizzati dalle piattaforme (con le quali gli editori condividono una percentuale del ricavo) in quanto rappresentano l'unica possibilità per le testate di accedere un'alta profilazione dei propri lettori[11]. Per ottimizzare questo meccanismo di *business* è necessario che le notizie rispondano a quei canoni imposti dalle piattaforme accennati poco prima. Possiamo dunque affermare che questa logica innesca è un vero e proprio processo di McDonaldizzazione della produzione delle *news*, concetto che meglio descrive il drastico cambiamento nella forma del contenuto informativo.

[11] Sebbene siano possibili strade alternative di business che rendono gli editori più indipendenti dalle piattaforme, ad esempio gli abbonamenti online, purtroppo non sempre sono opzioni sostenibili per quelle testate che non godono di reputazioni forti e distinguibili (come il Corriere della Sera; The Wall Street Journal) o semplicemente di lettori disposti a pagare.

Il termine *"McDonaldization"* è stato coniato nel 1993 da Ritzer per indicare il processo mediante il quale i princìpi del *fast food* arrivano a dominare un numero sempre maggiore di settori della società. Il *fast food* è basato su princìpi di efficienza, calcolabilità, prevedibilità e controllo delle macchine; in questo ambiente gli indicatori di valore quali la qualità e la varietà sono sostituiti dalla quantità e dalla standardizzazione. Secondo Ritzer, le aree della vita sociale soggette alla McDonaldizzazione sono in continuo aumento e questo determina la diffusione di tipi di produzione e consumo che, pur essendo estremamente razionali, presentano un gran numero di conseguenze irrazionali. Vale dunque la pena considerare la disinformazione come una delle conseguenze irrazionali generate da questo processo di produzione: gli editori devono adottare strategie efficienti ed efficaci per sopravvivere in rete, la categoria dell'efficienza si posa sull'ottimizzazione del processo di produzione e sulla semplificazione dei prodotti (Ritzer, 1993). Come in un *fast food*, la produzione di un contenuto informativo destinato al consumo *always on* della rete - quindi veloce, costantemente in aggiornamento e disponibile h24 – necessita di essere standardizzata e in linea con il mercato nel quale si pone, quello delle piattaforme, nel quale è indispensabile la condivisione di prodotti semplici e *user-friendly*, ossia generici, coinvolgenti, con lessico semplicistico, lunghezza ridotta e supporto delle immagini. Queste ultime possono essere considerate come il giusto *packaging* per cogliere l'attenzione dell'utente e assicurarsene l'interazione. Continuando ad utilizzare la metafora del *fast food*, è possibile dire che il mercato dell'attenzione ha trasformato il giornalismo professionale in un *McNugget Journalism* (Franklin, 1997), esempio efficace per comprendere quanto siamo disposti a consumare informazione assicurandoci di non imbattere in qualche tipo di complessità: come un pollo (di per sé animale complesso) l'informazione sulle piattaforme diventa un bocconcino piccolo, gustoso e facile da ingerire, avvolto in una panatura sensazionale e confezionato in uno scatolino accattivante. Ciò non significa che le testate d'informazione *mainstream* migrate negli spazi digitali o anche quelle native digitali abbiano smesso di praticare un tipo di giornalismo di qualità, come quello d'approfondimento o investigativo[12], il problema è che in un ecosistema mediale così strutturato, c'è il rischio che questo tipo di giornalismo non raggiunga la quantità di utenti che dovrebbe a causa della pressione che le piattaforme esercitano nella selezione dei contenuti. È proprio questa selezione di contenuti ad animare gran parte del dibattito sulla disinformazione: nella selezione algoritmica, sono inclusi indicatori di interesse personale e globale, questo significa che gli utenti sono esposti a un flusso maggiore di contenuti generati da coloro con i quali interagiscono di più e allo stesso tempo a contenuti che generano *engagement* più alto. Questo ha dato vita ad un sistema chiuso in se stesso in grado di

[12] Un buon esempio italiano è Fanpage.it: nata come testata generalista ha successivamente condotto importanti inchieste come "Bloody Money" su rifiuti e politica in Campania, oppure "Lobby Nera" sulla destra neofascista di Milano e i legami con i partiti locali e nazionali.

trasmettere visioni faziose e personalizzate che rispondono in maniera esaustiva a tutte le richieste di costruzione dei contenuti promosse dalle piattaforme: emotività, semplicità, viralità. Questo funzionamento ha generato una sorta di tensione tra datificazione e autonomia giornalistica: come scrive Van Dijck (2018), la sostenibilità dell'informazione è oggi modellata sulle strategie tecno-commerciali delle piattaforme e nel momento in cui la testata definisce le proprie priorità in base all'indicazione dei dati provenienti da quelle strategie, incorpora automaticamente le prospettive della piattaforma nel proprio ruolo di mediatore della realtà. Dunque, nonostante si dileguino da qualsiasi responsabilità editoriale, le piattaforme favoriscono la diffusione della disinformazione, attraverso la propria infrastruttura, il proprio funzionamento e la pressione che esercitano nel mondo editoriale, rendendola una rilevante arma politica ed economica.

2.4 Il caso italiano

Mentre nel contesto americano il giornalismo ha mantenuto la sua autonomia e legittimato la sua pretesa di raccontare la verità seguendo i principi di oggettività (Splendore, 2017) - grazie ai quali ha saputo render chiari i confini della propria professione e adattarsi ai cambiamenti storici e tecnologici - nel contesto italiano (Mancini parla di modello mediterraneo di giornalismo), la bassa professionalizzazione che lo caratterizza (ivi) ha messo in dubbio la sua autorevolezza che, col tempo, ha prodotto una drastica diminuzione dei livelli di fiducia nei lettori. Nel sistema fragile (Mancini, 2000) dell'informazione italiana, lo scollamento tra pubblici e testate è riconducibile a tre elementi: l'origine letterario-elitaria, il parallelismo politico, l'editoria impura. La nascita del giornalismo italiano tra i salotti letterari del XVII secolo ha determinato un progressivo allontanamento dall'attuale modello di giornalismo dominante, quello appunto di discendenza anglosassone, caratterizzato da pratiche discorsive centrate sul racconto fattuale, neutrale e informativo. L'ambiente letterario italiano ha determinato, già prima dell'avvento dei *social media* e della commistione tra giornalista e lettore, la tendenza a far prevalere lo spazio del commento e dell'interpretazione sulla narrazione cronistica degli eventi. Il parallelismo politico - seconda caratteristica storica dell'informazione italiana - nutrito prima da idee risorgimentali, poi dalla propaganda di regime e infine dalla Resistenza, ha generato un forte sentimento anti-elitario, nutrito dall'idea che i mezzi d'informazione fossero strumenti a servizio del mondo politico e il cui interesse fosse ben lontano dal dar voce al popolo. Un sentimento accentuato dalla forte presenza di un tipo di editoria impura resa necessaria dalla storica debolezza del mercato della stampa italiana, il quale ha avuto bisogno prima dell'intervento del potere pubblico per sopravvivere (già alla fine dell'800), e successivamente dell'aiuto di grossi gruppi industriali e

bancari[13]. L'esistenza di editori impuri (ovvero editori con interessi diversi dall'editoria) ha avuto (e ha ancora) ripercussioni sia sull'operato dei giornalisti, sia sulla percezione del pubblico nei confronti delle notizie diffuse dalle testate *mainstream*, tendenzialmente in linea con le battaglie politiche dei finanziatori. Infine, l'informazione italiana è stata generalmente lenta nell'adottare le nuove tecnologie, introdotte nelle redazioni quando ormai non se ne poteva più fare a meno (Splendore, 2017): le grandi testate nazionali hanno iniziato ad utilizzare la rete per distribuire informazione solo vent'anni fa, riproponendo in digitale i contenuti del cartaceo[14] e arrivando in ritardo ad imporsi nel contesto informativo delle piattaforme con una conseguente acquisizione da parte dei *social network* della funzione fino a quel momento ad essi riservata, quella di prima fonte di formazione di immagini e icone sui principali fatti pubblici. Queste caratteristiche hanno fatto sì che fosse delegittimata la professione nel più ampio contesto sociale (Sorrentino, 2003). La ricerca di fonti alternative assume così i tratti di una risposta a un'offerta valutata in termini negativi dai lettori, questo genera un progressivo venir meno della fiducia nei confronti del sistema informativo e influisce sulla valutazione dell'argomentazione che essi offrono (Bentivegna e Boccia Artieri, 2019).

Indipendentemente dal contesto italiano, la delegittimazione dell'informazione *mainstream* ha promosso fonti alternative e dato vita ad una sorta di "sfera pubblica alternativa". Questa è alimentata da posizioni cospirazioniste che fanno leva sul gioco di sovrapposizione tra finzione e realtà: la separazione tra queste due sfere è così tenue da affermare una loro sostanziale equivalenza che sfocia nel fenomeno della disinformazione. Queste dinamiche sono state facilitate, estremizzate e amplificate (ma non generate) dal supporto tecnologico della rete e delle piattaforme.

3 Siamo esseri imperfetti. L'elemento umano

Lo spazio di confronto, dissenso e dialogo in un sistema mediale così strutturato sono lontane dalle logiche di dibattito pubblico e sfera pubblica così come le intendeva Habermas (1974), piuttosto vengono a generarsi delle sfere ideologiche ben delimitate e abbastanza impermeabili, ognuna delle quali supportata dalla propria verità e consolidata nella propria esperienza: le cosiddette *echo chambers* (Quattrociocchi e Vicini, 2016). Le camere d'eco sono naturalmente prodotte dall'architettura delle piattaforme che, come si è detto, distribuisce informazioni sulla base delle preferenze degli utenti confinandoli in spazi digitali in cui risuona lo stesso rumore di fondo che convalida ricorsivamente lo

[13] È il caso del Corriere della Sera, della Stampa, del Sole 24 Ore, della Repubblica, del Messaggero.
[14] Funzione che Lippmann (1942) riconosce come tradizionalmente ricoperta dai giornali

stesso punto di vista. Ciò che è di preferenza per un gruppo di utenti è adeguato ai loro gusti e personalità ed è quindi una rappresentazione del mondo parziale ma allo stesso tempo assolutizzata dalle loro percezioni. Tutto ciò amplifica un meccanismo cognitivo completamente umano che, in diversa misura, tutti mettiamo in atto involontariamente: il *confirmation bias.* Siamo esseri imperfetti e lontani dalla figura dell'uomo razionale settecentesco, per cui tendiamo a muoverci entro uno spazio delimitato che conferma atteggiamenti e convinzioni preesistenti in ognuno di noi evitando quella che Festinger (1957) chiama "dissonanza cognitiva", ossia un sentimento che si attiva quando entriamo in contatto con informazioni o idee divergenti alle nostre. Ne consegue che gli individui si sottraggono a quei messaggi che appaiono in contraddizione con le proprie opinioni preesistenti e qualora non riuscissero nell'intento, entra in gioco un meccanismo di "percezione selettiva" che lo porta ad una distorsione del significato del messaggio fino a renderlo coerente e integrato al proprio sistema valoriale (Festinger, 1957). Una vera e propria dissociazione mentale tra la realtà e la percezione. Dunque, ciò che da esseri imperfetti riteniamo credibile non sempre corrisponde alla realtà: un fatto è credibile in quanto capace di suscitare adesione (Lorusso, 2018) ma nella logica del filtraggio sulla base delle preferenze, l'adesione può essere data verso qualsiasi tipo di contenuto in diversa misura contraffatto.

L'esposizione selettiva può essere riconducibile anche ad un intervento consapevole degli individui. Bentivegna e Boccia Artieri (2019) individuano queste pratiche in tre tipi di interventi: il primo è la riduzione del costo dell'elaborazione delle informazioni, ossia quell'insieme di scelte per cui gli utenti si espongono consapevolmente ad informazioni coerenti con le proprie visioni del mondo perché questo richiede meno sforzo cognitivo. Il secondo è la consultazione di fonti alternative, selezionate in base ai propri pregiudizi, alle quali si attribuisce un giudizio di qualità a fronte di una valutazione negativa espressa nei confronti dei media *mainstream*. Il terzo intervento è la ricerca, su base volontaria, di una consonanza con l'*audience*: questo indica il desiderio di dar vita ad un'omofilia informativa costruita grazie alla scelta consapevole di un'informazione faziosa che si rivolge a lettori affini e con le stesse opinioni. Questo meccanismo umano ha reso insufficienti le strategie di contrasto al fenomeno della disinformazione, come il *debunking* e il *fact checking*, proprio perché gli individui non sono interessati a sentire versioni che vadano in contrasto con la propria visione del mondo, soprattutto se costretti a questo esercizio (Zollo et al., 2017)). Alcuni studi (ivi) hanno empiricamente rilevato quanto la somministrazione, all'interno di un gruppo cospirazionista, di informazioni volte a smontare le teorie sostenute dal gruppo, non ha prodotto nessun risultato di correzione ma, al contrario, ha generato l'effetto controproducente di ulteriore rinforzo delle stesse teorie.

Le piattaforme digitali hanno tradotto il meccanismo umano dell'esposizione selettiva in quello delle *filter bubbles* (Parisier, 2011), che regolano la visibilità e la circolazione dell'informazione influendo

sul consolidamento di credenze e sentimenti condivisi. La piattaforma filtra contenuti che appariranno certamente credibili perché non dissonanti dalle convinzioni dell'utente e da quelle della camera d'eco a cui appartiene; questi contenuti riceveranno a loro volta l'adesione attraverso *like, share* e commenti, attivando così un meccanismo di rafforzamento in cui la credibilità dei contenuti è rinforzata dall'*engagement* della propria rete. Le *filter bubbles* contribuiscono a restituirci una realtà a misura dell'utente e delle sue preferenze, una sorta di versione "addomesticata" della realtà (Lorusso, 2018) a cui si crede acriticamente proprio perché rispecchia la visione del mondo condivisa, con una conseguente difficoltà di distinzione del vero dal falso. La facoltà di distinguere i fatti dai contenuti manipolati è inibita da un altro fenomeno figlio della società dell'informazione: l'*information overload*. Le informazioni oggi disponibili sono diventate talmente abbondanti che risulta difficile reperire quelle utili nel momento necessario. L'enorme quantità di informazioni, molte delle quali irrilevanti, imprecise o completamente false, ha generato nel lettore un sovraccarico cognitivo che gli crea disordine e confusione tale da impossibilitarlo non solo nel prendere una decisione riguardo uno specifico tema, ma anche nella scelta delle fonti a cui affidarsi (Renjith, 2017).

Per quanto le pratiche non mediate dei *social network* ci siano apparse come il trionfo della libertà d'espressione, come esseri imperfetti abbiamo necessariamente bisogno di una mediazione. Seguendo la lezione di Peirce, ogni azione, discorso o pratica, ogni processo di cui gestiamo il senso si basa su una mediazione interpretativa. Abbiamo bisogno di una mediazione perché siamo immersi in un mondo che ci precede, ci condiziona e ci influenza, così l'unico modo per elaborare e condividere il senso è ri-mediarlo attraverso un vero e proprio schema interpretativo necessario a collegare giudizio e realtà (Lorusso, 2018). Ci sono sempre fatti interpretabili e i media sono luoghi in cui si modellano i fatti che poi vengono definiti da interpretazioni. Ogni mezzo di comunicazione è un dispositivo, inteso nel senso foucaultiano del termine, ossia un insieme di elementi tecnologici e discorsivi attraverso i quali i soggetti negoziano i propri personali stati di realtà, attraverso comportamenti, credenze e interpretazioni. Nelle *eco-chambers*, il senso comune si costruisce così su tante verità condivise e personalizzate a tal punto da non essere confutabili perché protette dal rischio di dissenso, queste verità si naturalizzano e stabilizzano creando un falso senso di veridicità anche per le notizie più assurde.

Conclusioni

I fattori individuati e descritti fin qui provano a dare una spiegazione non all'esistenza della disinformazione, quanto alla sua buona riuscita. Cercando di tirare le somme di quanto detto, il postmodernismo è la lente d'ingrandimento giusta per comprendere quanto facilmente la disinformazione riesca ad attecchire nei suoi fruitori. In generale potremmo dire che la cultura postmoderna ha preparato il terreno, minando valori e convinzioni fino a quel momento consolidati nel mondo occidentale: autorità, scienza e conoscenza oggettiva. Il rifiuto dell'autorità e la diffusione di sentimenti anti-elitari hanno allontanato il pubblico dalle fonti autoritarie portatrici di conoscenza, considerate rappresentati di un sistema lontano dagli interessi collettivi e per questo non più accettato. Un sentimento alimentato dalla trasformazione dell'informazione nella principale merce della nuova economia postindustriale. Anche in questo caso, l'informazione (nel suo senso più ampio di condivisione di conoscenza) è stata merce di scambio fin dall'invenzione della stampa, ma è con l'avvento delle nuove tecnologie a diventare una fonte fondamentale di produttività e potere. Negare la possibilità di una verità oggettiva ha aperto la strada ad affermazioni di verità in competizione e alla diffusione di fatti alternativi. Con l'avvento del digitale e delle piattaforme questo sentimento antiautoritario si è radicato nel momento in cui gli utenti hanno avuto l'opportunità di condividerlo: si è moltiplicato grazie alla possibilità di diffondere una verità personale, di assolutizzata a versione alternativa della verità e nella possibilità di riunire in suo nome chiunque avesse la stessa visione del mondo. A diffondere e radicalizzare questi atteggiamenti è stato il funzionamento stesso delle piattaforme che, come descritto, rendono gli ambienti di comunicazione omofili e inadeguati a produrre differenze rispetto al proprio modo di pensare. Per questo, nell'analisi del fenomeno della disinformazione, l'ambiente digitale non va inteso solo come tecnologia innovativa che ha disintermediato le pratiche di fruizione dell'informazione, ma anche come un processo profondamente trasformativo che agisce sulle prassi e sul modo di pensare degli individui. Se da un lato l'economia postindustriale e le tecnologie digitali sono state il trionfo della libertà di espressione, dall'altro hanno contribuito al deteriorato dei confini giornalistici. La stampa tradizionale lavora con prassi e parametri che la stessa disinformazione utilizza. Come si è visto, il mondo editoriale è stato costretto e rimodellare tutte le logiche su cui si è sempre fondata la professione giornalistica per sopravvivere nell'ecosistema mediale delle piattaforme; questo ha generato un cortocircuito per cui l'unico *business* che riesce e far competere le testate con la gratuità della rete, ossia la pubblicità, è diventato facile preda di chi ha prodotto contenuti contraffatti con lo scopo di monetizzare e allo stesso tempo influenzare l'opinione pubblica. La buona riuscita del fenomeno disinformativo risiede, in ultima istanza, nell'uomo. L'elemento umano è quello più impermeabile alle operazioni di intervento, ma se

solo si riuscissero a mettere in dubbio i contenuti che circolano sulle piattaforme nello stesso modo in cui si mette in dubbio l'informazione diffusa dalle autorità, probabilmente la disinformazione, nonostante il megafono della rete, non riuscirebbe a produrre gli stessi preoccupanti effetti.

CAPITOLO II - La ricerca sulla disinformazione. Lo stato dell'arte

Nell'ultimo decennio, il vocabolario politico, sociologico, economico e informatico si è arricchito con l'introduzione del termine *fake news*. Nonostante la presenza di notizie false sia riscontrabile ben prima degli ultimi dieci anni, nel 2016 l'espressione ha acquisito grande rilevanza perché portata all'attenzione da due significativi eventi politici: il processo elettorale statunitense, che ha portato Donald Trump alla Casa Bianca e la Brexit, il tumultuoso processo di uscita dall'Unione Europea da parte del Regno Unito. Durante questi due processi politici si è registrata una forte ondata di voci e informazioni false utilizzate nel tentativo (in entrambi i casi riuscito) di ottenere vantaggi e raggiungere obiettivi. L'Italia non è passata indenne al fenomeno: il *referendum* del 2016 è stato caratterizzato da feroce una campagna disinformativa tale che la notizia più condivisa sui *social media* in relazione al voto è stata proprio una *fake news*[15]. L'ubiquità del funzionamento dei motori di ricerca e dei *social media* hanno fatto da cassa di risonanza al fenomeno. Nonostante le piattaforme proprietarie siano state chiamate ad impegnarsi nel mettere a punto meccanismi utili nella lotta alle notizie false, il problema non è affatto ridimensionato, anche a causa della testimoniata incapacità di queste ultime a farvi fronte[16]. Il fenomeno disinformativo solleva interrogativi di vario ordine e fortemente interdisciplinari su cui la comunità scientifica si è spesso scontrata. Una questione su cui tutti concordano è che il termine *fake news* è ormai inadeguato a descrivere i complessi fenomeni di inquinamento dell'informazione ed è spesso abusato da chi lo utilizza indistintamente per indicare che "qualcosa non va" nella sfera pubblica digitale.

Il capitolo propone una panoramica degli aspetti esaminati dalla ricerca accademica in relazione al fenomeno della disinformazione e le tecniche di analisi adottate. Di certo non fornisce una mappatura esaustiva della produzione scientifica sul fenomeno, ma mira a fornire un punto di partenza per chiarire in quali direzioni la comunità accademica si è effettivamente mossa nel tentativo di comprenderlo e contrastarlo.

Gli studi si sono finora concentrati su cinque macroaree: definizione e categorizzazione della disinformazione; profilazione degli *account* che diffondono contenuti controversi; analisi delle

[15] L'analisi è di Pagellapolitica che ha considerato, tra il 1° ottobre e il 30 novembre 2016, i post in italiano con più alto coinvolgimento su Facebook e più condivisioni sui social media Twitter, Linkedin e Google+ all'interno dei quali era contenuta la parola "referendum". I risultati hanno dimostrato che la notizia con l'*engagement* più alto era una bufala e riguardava il presunto ritrovamento, nel paese inesistente di "Rignano sul Membro", di cinquecentomila schede elettorali con il "sì" già segnato. Nei due mesi preelettorali il *link* di questa notizia ha suscitato oltre 233 mila reazioni, più di ogni altro contenuto preso in considerazione.

[16] L'inchiesta "Facebook Papers" ha evidenziato (tra le altre cose) quanto la piattaforma sia stata incapace di arginare il fenomeno della disinformazione e dell'incitamento all'odio al di fuori degli Stati Uniti e della lingua inglese.

caratteristiche del messaggio; indagini sulle dinamiche sociali in ambienti disintermediati e controllo delle misure di contrasto alla disinformazione.

1. Definire la disinformazione

Considerato quanto le caratteristiche della disinformazione siano molteplici e tra loro combinate, definirla resta un compito piuttosto arduo. La letteratura considera le *fake news* come una forma di falsità intenta a ingannare le persone imitando l'aspetto e il linguaggio delle notizie vere. Questa definizione consente di identificare due domini della disinformazione: l'intenzione dell'autore, che si riferisce al grado in cui il creatore di notizie false intende fuorviare il lettore e il livello di *facticity*[17] (Tandoc et al., 2018), che si riferisce al grado in cui le notizie false si basano su fatti realmente accaduti. Entrambi i fattori sono utilizzati in riferimento a varie forme di contenuto che vanno dalla propaganda alla pubblicità ingannevole, dalle notizie provenienti da siti di informazione alternativa a contenuti con titoli fuorvianti, fino ad arrivare alla satira. Claire Wardle (2017), utilizza due termini distinti per riferirsi a pratiche e contenuti che sono in contrasto con le informazioni verificate da fonti ufficiali, questi sono "disinformazione" e "misinformazione". Gran parte del discorso sulle *fake news* confonde queste due nozioni: si tratta di disinformazione, quando qualcuno crea o condivide deliberatamente contenuti falsi o fuorvianti con l'intento di causar danni; si tratta di misinformazione, quando qualcuno condivide inconsapevolmente contenuti falsi con nessun intento doloso. A queste categorie, Claire Wardle aggiunge la "mal-informazione" definita come quell'insieme di informazioni che si basano sulla realtà ma sono utilizzate per causar danni a persone, organizzazioni o paesi. La distinzione della Wardle è puramente indicativa e nonostante le conseguenze sull'ambiente informativo siano le stesse, spesso ci si imbatte in diverse combinazioni di queste tre concettualizzazioni o queste si presentano come parte di una più ampia strategia volta all'inganno. Ciò che emerge subito dal difficile lavoro di classificazione è la grande attenzione verso la dimensione intenzionale. Le intenzioni che motivano la produzione di disinformazione possono esser mosse principalmente da due giustificazioni: quella finanziaria e quella ideologica (Tandoc et al., 2018). La motivazione finanziaria si riferisce alla produzione di contenuti disinformativi con l'intento di convertire i *click* in ricavi pubblicitari, mentre la motivazione ideologica si riferisce alla produzione di disinformazione con l'intenzione di screditare determinati attori, insinuare il dubbio, influenzare l'opinione pubblica o istigare alla violenza.

[17] Il riferimento proviene da Sartre (1943) e dall'articolazione dei concetti: *facticité* (tutte le realtà concrete e immutabili di un individuo) e *transcendance* (il bisogno di interpretare il mondo e fare scelte). La traduzione italiana potrebbe essere "fatticità" ma si preferisce riportare la parola in inglese così come proposta dalla letteratura a cui si fa riferimento.

Riprendendo i lavori della Warlde è possibile mettere a punto una panoramica del modo in cui la disinformazione viene definita e categorizzata in letteratura, avendo chiari i fattori di intenzione e *facticity*.

- *Satira e parodia*

La presenza della satira nella letteratura sulla categorizzazione della disinformazione può forse sorprendere. Tuttavia, in un'epoca in cui le persone ricevono sempre più informazioni tramite *news feed* e dedicano sempre meno tempo alla lettura attenta, spesso la satira è scambiata per informazione. La satira, per sua natura, è utilizzata in contenuti che si servono dell'umorismo e dell'esagerazione per presentare al pubblico i fatti più rilevanti. Questi contenuti sono in genere incentrati sull'attualità e anche se utilizzano lo stile delle *news*, sono prodotti con una motivazione dichiaratamente umoristica e trasparente. La satira è una parte sempre più rilevante dell'ecosistema delle piattaforme[18] alcune ricerche hanno evidenziato che gli individui che consumano contenuti satirici sono informati sui fatti tanto quanto gli individui che consumano notizie attraverso altri mezzi di informazione (Kohut et. al., 2007). Il basso livello di *facticity* della satira fa riferimento solo al formato, mentre il contenuto principale si basa su avvenimenti accaduti. La parodia condivide molte caratteristiche con la satira poiché entrambe si affidano all'umorismo per attirare il pubblico. La parodia differisce però per l'utilizzo di informazioni non fattuali: invece di fornire commenti diretti sull'attualità attraverso l'umorismo, ridicolizza i problemi e li mette in evidenza attraverso notizie inventate (Tandoc et. al., 2018). La parodia gioca sulla vaga plausibilità della notizia e ha successo sul pubblico perché si caratterizza per un sofisticato equilibrio tra ciò che può essere possibile e ciò che è assurdo, ciò che il lettore potrebbe credere o voler credere. Il problema sorge nei casi in cui la parodia è così sottile da esser scambiata per una notizia, distorcendo completamente il senso degli avvenimenti. Nonostante ciò, non può esser considerato un disturbo informativo al pari di un contenuto manipolato perché nei contenuti satirici e parodici esiste tra l'autore e il lettore una sorta di accordo implicito sull'intento umoristico veicolato. Questi contenuti restano comunque fondamentali per lo studio della disinformazione perché permettono di misurare fino a che punto il *confirmation bias* sia determinante nella scelta delle informazioni.

[18] Un esempio è il dilagare di studi sui *memi:* immagini, GIF o video dai contenuti satirici con un alto grado di viralità. Un *meme* anche se fa riferimento a fatti realmente accaduti, è privo di un reale contenuto informativo.

- *Fabbricazione di notizie*

L'assenza di quell'accordo implicito tra autore e lettore è una delle caratteristiche su cui si basa la fabbricazione di notizie o comunemente chiamate *fake news*. La fabbricazione di notizie si riferisce, in letteratura, ad articoli che non hanno basi fattuali ma sono prodotti e condivisi con lo stesso stile delle notizie provenienti da fonti ufficiali. Il produttore del contenuto, con l'intento di ingannare, fornisce obiettività e spiegazioni dei fatti in maniera così convincente che il lettore riscontra difficoltà nel distinguerlo da una notizia proveniente da fonti ufficiali. Una notizia falsa di successo è un articolo che attinge a parzialità preesistenti, intrecciate in una narrazione fittizia (spesso contenente pregiudizi) che il lettore accetta come legittima. È importante sottolineare che il successo della fabbricazione delle notizie dipende molto dalla tensione sociale preesistente (Tandoc et. al., 2018): se c'è tensione sociale, se ci sono serie differenze politiche, sociali, razziali o culturali, le persone saranno più vulnerabili a credere a informazioni fittizie che supportano queste differenze. Le due dimensioni rilevanti nella letteratura sulle *fake news* sono il movente finanziario e quello della manipolazione dell'opinione pubblica. Guardando al primo, possiamo considerarlo come uno degli effetti collaterali del *business* di cui si sono dotati gli editori nel passaggio all'ecosistema digitale: più la storia è sensazionale e credibile, più suscita l'interesse dell'utente, maggiore sarà il reddito per il produttore proveniente dalle logiche del *clickbait*. Queste logiche sono supportate dall'esistenza dei *bot* che forniscono all'utente l'illusione che il contenuto fabbricato sia ampiamente diffuso: i siti di notizie false spesso si affidano ad un ecosistema di propaganda in tempo reale composto da una rete di *bot* che condividono continuamente la stessa serie di notizie false. Questo sistema fornisce al lettore la sensazione che molti altri stiano leggendo, condividendo ed eventualmente apprezzando la notizia, aggiungendo un ulteriore fattore di legittimità e rinforzo, che non proviene dalla fonte da cui è prodotta la notizia, ma dalla propria comunità in rete.

- *Contenuti manipolati*

Mentre le categorie precedenti si riferivano esclusivamente a contenuti basati sul testo, in questa categoria sono inseriti anche i contenuti visuali. Le immagini e i video sono veicoli particolarmente potenti per diffondere informazioni fuorvianti poiché catturano più facilmente l'attenzione dell'utente ed è più complesso controllarli. L'alterazione delle immagini è una pratica sempre più comune e sofisticata, facilitata da numerose *app* disponibili in qualsiasi *smartphone*. La manipolazione di notizie o immagini si basa su fenomeni reali ai quali però vengono aggiunti elementi che non hanno alcuna base fattuale e che fanno leva su un pregiudizio di parte verso il fenomeno trattato. Attualmente le piattaforme non sono dotate di sistemi per il riconoscimento di immagini manipolate, tanto meno esiste

un modo per far rispettare un qualsiasi codice etico a garanzia del fatto che le immagini pubblicate non diffondano informazioni false o ingannevoli.

- *Propaganda*

Come la satira, anche la propaganda è sorprendentemente rientrata nella letteratura sui disturbi informativi. La propaganda si basa sui fatti ma include pregiudizi che promuovono una particolare prospettiva. Ciò che differenzia la disinformazione dalla propaganda è che l'obiettivo di quest'ultima è quello di persuadere piuttosto che informare, è quindi più apertamente manipolativa che disinformativa. La disinformazione però serve spesso gli interessi della propaganda. È ciò che accade soprattutto nel contesto mediale russo, in cui esiste una perfetta commistione dei due concetti. I canali di notizie ufficiali (come Channel One), trasmessi sia a livello locale nella Federazione Russa che a livello internazionale, non aderiscono allo stesso codice giornalistico delle agenzie stampa occidentali: le notizie diffuse da queste emittenti sono di fatto false e costruite per influenzare la percezione pubblica sulle azioni politiche russe[19] (Khaldarova e Pantti 2016). Questo tipo di disinformazione promossa dalle istituzioni può essere definita come una narrazione strategica: uno strumento in mano agli attori politici per orientare le posizioni dei cittadini su questioni delicate e modellarne così le percezioni e le azioni.

Le dimensioni fin qui descritte circoscrivono il modo in cui possiamo utilizzare il concetto di *"fake news"* nel discorso contemporaneo. Sulla base di questa revisione della letteratura è stato costruito lo strumento di rilevazione utilizzato in questo lavoro di ricerca che prova a superare alcuni limiti emersi dalla produzione scientifica di riferimento e al tempo stesso prova ad applicare i concetti esaminati ad un set di dati particolarmente ampio.

[19] Nel momento in cui si scrive, la più recente e intensa campagna di disinformazione digitale messa in atto dalla Federazione Russa riguarda l'invasione dell'Ucraina. La campagna disinformativa ha l'obiettivo di preparare l'opinione pubblica alla necessità dell'intervento e per incoraggiare il sostegno interno russo all'azione militare. Queste narrazioni diffuse sui media incolpano anche l'Occidente per l'*escalation* della tensione e fanno leva sulla presenza di gravi problemi umanitari in Ucraina che l'intervento russo potrebbe risolvere, promuovendo così il sentimento di patriottismo e il sostegno all'azione. Nel momento in cui si scrive, sui media russi si registra già un aumento del 200% della media giornaliera di diffusione di tali narrazioni. Fonte: Restuccia, A. & McBride, C. (2022), *White House Says Russia Planning 'False Flag' Operation as Pretext for Invading Ukraine*, Wall Street Journal, disponibile al *link*: https://www.wsj.com/articles/white-house-says-russia-is-planning-false-flag-operation-as-pretext-for-invading-ukraine-11642182308

2. Profilazione dei produttori

Mentre il precedente filone di studi tenta di spiegare la tipologia e gli elementi che definiscono la disinformazione, un secondo filone di ricerca è interessato all'analisi dei produttori di contenuti disinformativi. Le ricerche che si pongono quest'obiettivo si focalizzano sulla profilazione degli utenti per rilevare *account* dannosi. Utilizzando funzionalità basate sul tempo, gli studi descrivono il comportamento di pubblicazione degli *account* creatori di contenuti falsi e cercano di rilevare caratteristiche comuni e costanti. Attraverso l'utilizzo di serie storiche si analizza la frequenza di pubblicazione (data dall'intervallo di pubblicazione tra due post) in relazione alla frequenza di risposta e commento degli *account* su determinati contenuti (Ferrara et. al., 2016). L'idea di base è che il comportamento di pubblicazione di *account* dannosi è guidato da una strategia ben definita per arrivare al maggior numero di utenti possibili, soprattutto se supportati da una rete di *bot*. Con questo tipo di analisi, gli *account bot* vengono individuati facilmente: essendo guidati da timer o programmi automatici tendono ad assumere comportamenti temporali precisi e standardizzati rispetto a quelli assunti da utenti umani. Attraverso l'individuazione di costanti nel comportamento di pubblicazione è così possibile rilevare attività di pubblicazione sospette e identificare la produzione contenuti controversi direttamente alla fonte di produzione.

3. L'analisi del messaggio

Un filone di studi molto proficuo si concentra sullo studio delle caratteristiche del messaggio disinformativo. Gli studi condotti negli ultimi anni in questa direzione hanno adottato (in diversa misura) tecniche di *Content Analysis*, con l'obiettivo di studiare il tipo di comunicazione veicolata dai messaggi disinformativi e approcci *Text Mining*, con l'obiettivo di identificare particolarità linguistiche del testo da utilizzare nell'addestramento di modelli di apprendimento automatico e tentare così di costruire rilevatori accurati di *fake news*. L'utilizzo dell'apprendimento automatico combina dunque lo studio del messaggio con la messa a punto di misure di contrasto alla disinformazione. L'analisi del contenuto di articoli pubblicati da siti attenzionati come fonti di notizie false mira ad indagare il diverso grado di *facticity* in relazione a diverse questioni d'interesse pubblico (politica, vaccini, immigrazione) facili prede di campagne disinformative. In queste ricerche, la rilevazione del falso e la classificazione del tipo di disinformazione avvengono manualmente: il ricercatore deve necessariamente possedere una vasta conoscenza dell'argomento nonché delle pratiche di *fact-checking* che adopererà per l'identificazione del falso. La classificazione manuale è sicuramente

efficiente ma utilizzata da sola presenta inevitabilmente problemi di scalabilità nella gestione di grandi volumi di dati tipici di un fenomeno prevalentemente digitale. Nonostante ciò, l'utilizzo della *Content Analysis* ha prodotto una corposa letteratura empirica soprattutto in campo politico con particolare riferimento alle elezioni americane del 2016. Questo approccio è stato utile a comprendere quanto i contenuti disinformativi utilizzano diversi livelli di sensazionalismo, tecniche di *clickbait* e pregiudizi per ottenere maggior *engagement*. È proprio la presenza di pregiudizi ad essere il più forte predittore di *engagement* di un contenuto fasullo pubblicato sui social media (Mourão & Robertson, 2019).

Dal punto di vista delle caratteristiche linguistiche, la letteratura è ricca di sperimentazioni circa l'applicazione del *Natural Language Processing* e del *Machine Learning* nella rilevazione di notizie false. Questo approccio si basa sull'idea che, sebbene gli autori di notizie false cerchino di imitare lo stile di scrittura di un giornalista professionista per ingannare gli utenti, ci sono ancora alcune differenze che possono essere utilizzate per discriminare i contenuti falsi da quelli veri: le notizie false, nel loro intento di raggiungere più utenti possibili, adottano delle particolari caratteristiche linguistiche che le differenziano dalle notizie provenienti da fonti attendibili. Individuate queste caratteristiche è possibile ricostruire i diversi stili di scrittura adottati nella stesura di notizie false.

In questi studi, le caratteristiche linguistiche vengono differenziate tra funzionalità a livello di parola e funzionalità a livello di frase.

A livello di parola, le analisi basate su *bag-of-word*, *n-gram*, *term frequency* (TF), *term frequency-inverted document frequency* (TF-IDF) sono quelle più comunemente utilizzate per l'elaborazione del linguaggio naturale, alle quali si aggiunge l'analisi dei *token*. L'estrazione di queste caratteristiche su un *corpus* di notizie false viene utilizzata per allenare gli algoritmi di classificazione nel lavoro di identificazione con la stessa logica di rilevazione delle mail *spam/no spam.* Le caratteristiche linguistiche possono includere esclamazioni multiple, punti interrogativi, parole in grassetto o in maiuscolo. Anche la presenza di *stop words* è stata utilizzata per il rilevamento di notizie false (Horne & Adali, 2017): questa tiene conto di un particolare utilizzo di negazioni, sostantivi, pronomi e avverbi, ma anche della presenza o meno di parolacce. Questi elementi sono spesso analizzati in combinazione ad elementi tipici della scrittura *social media* come menzioni degli utenti, *hashtag* ed *emoticon*.

Altre caratteristiche linguistiche a livello di parola riguardano la presenza di parole emotive nel contenuto delle notizie; questo può aiutare a determinare la polarizzazione del testo e quindi classificarlo come falso o non falso. Ahmed, Traore e Saad (2017) propongono un modello di rilevamento di notizie false che utilizza proprio l'analisi di *n-gram* unita a tecniche di estrazione delle

caratteristiche per testare le *performance* di sei diversi algoritmi di classificazione[20]. Gli esperimenti studiando l'impatto della dimensione ***n*** di *n-gram:* ogni valore *n è* stato testato in combinazione con un diverso numero di caratteristiche estratte con la funzione TF-IDF che, combinata al classificatore *linear support-vector machine* (LSVM) ha raggiunto una precisione di rilevazione di contenuti falsi pari al 92%. Il lavoro di Horne e Adali (2017) invece, confuta quel presupposto di fondo che considera le notizie false scritte per sembrare notizie vere. I ricercatori hanno estratto caratteristiche stilistiche e sintattiche di articoli falsi provenienti da diversi *set* di dati. I risultati hanno dimostrato che la struttura generale del titolo e l'uso di nomi propri sono molto significativi nel differenziare i contenuti falsi da quelli reali e che le notizie false appaiono (nello stile linguistico) più simili alla satira che alle notizie attendibili. Le caratteristiche linguistiche sono state utilizzate per costruire un modello di apprendimento automatico con una precisione di circa il 71%.

A livello di frase, le funzionalità si riferiscono a tutti quegli attributi che forniscono informazioni sulla complessità della frase, come la lunghezza media, la frequenza della punteggiatura, la quantità di parole funzionali e quella delle *stopwords*, nonché informazioni sulla polarità media (positiva, neutra o negativa). Più nello specifico, l'analisi della complessità può utilizzare il calcolo della profondità dell'albero sintattico di ogni frase del testo per estrarre caratteristiche utili al rilevamento di particolari contenuti. Un esempio di questo tipo di analisi è il lavoro condotto da Conroy, Rubin e Chen (2015), che uniscono l'approccio *n-gram* con l'analisi delle strutture sintattiche profonde. Le frasi vengono trasformate in un insieme di regole di scrittura utile a descrivere la struttura della sintassi di un testo e quindi rilevare le notizie ingannevoli con un'accuratezza che oscilla tra l'85% e il 91%. Hanno inoltre analizzato il grado di veridicità di un contenuto attraverso la compatibilità tra l'esperienza dell'utente e il contenuto stesso. L'intuizione è che uno scrittore con l'intento di ingannare non abbia conoscenza del fenomeno che tratta e questo lo induce a contraddizioni o omissioni di fatti presenti invece in notizie attendibili sugli stessi *topic*. Allo stesso modo, i titoli delle notizie false sono spesso esagerati per attirare l'attenzione dei lettori ma non coerenti (o addirittura in conflitto) con il contenuto testuale delle notizie. La coerenza tra titolo e corpo della notizia o tra le diverse parti della stessa notizia potrebbe dunque essere utile a differenziare un contenuto vero da uno fasullo.

Nonostante gli sforzi, approcci di questo genere hanno comunque i loro limiti. Questi sono dovuti alla scarsa disponibilità di *corpora* per la modellazione predittiva e alla quasi totale mancanza di set di dati

[20] K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Linear Support Vector Machine (LSVM), Decision tree (DT) e Stochastic Gradient Descent (SGD).

onnicomprensivi di informazioni accessorie[21] necessarie ad allenare gli algoritmi di *machine learning* nel classificare una notizia falsa con elevata precisione.
Infine, sia l'approccio *Content Analysis* che quello *Machine Learning* sono applicati sulla componente testuale e questo lascia scoperta una grossa quantità di contenuti disinformativi che invece fanno uso delle immagini nel tentativo di ingannare gli utenti.

4. L'analisi delle dinamiche sociali

Un ulteriore filone di studi si inserisce pienamente nel campo delle scienze sociali computazionali (Lazer et al., 2009) con l'obiettivo di comprendere le dinamiche sociali che emergono tra gli utenti che consumano informazione negli ambienti disintermediati delle piattaforme. Le ricerche empiriche sulle dinamiche sociali in rete fanno uso della *Social Network Analysis* e si caratterizzano per un approccio esclusivamente *data-driven* il cui scopo è quello di sfruttare la grande mole di dati provenienti dai *social media* per mettere a punto modelli matematici e strumenti computazionali precisi a tal punto da comprendere, anticipare e ipoteticamente controllare massicci fenomeni sociali in rete (Quattrociocchi, 2021). Questo approccio vanta di aver empiricamente rilevato l'esistenza delle *echo chambers* e dimostrato il ruolo del *bias* di conferma nella selezione dell'informazione. Il filone di studi sul comportamento *online* ha mostrato come informazioni qualitativamente diverse (provenienti da fonti di informazione ufficiali e fonti di informazione alternativa) presentano caratteristiche molto simili in termini di diffusione, numero di utenti che vi interagiscono e costanza nell'interazione. Informazioni provenienti da fonti ufficiali e alternative si diffondono allo stesso modo e non mostrano sostanziali differenze in termini di fruizione degli utenti (Mocanu et. al., 2015), ciò che è rilevante è che questi contenuti sono fruiti in maniera mutualmente esclusiva. Analizzando il comportamento di circa un milione di utenti italiani su Facebook che hanno interagito con pagine complottiste e scientifiche è stato mostrato quanto all'aumentare dell'interazione degli utenti con una specifica narrativa, aumenti linearmente la probabilità di avere una rete sociale digitale composta solo da utenti con lo stesso profilo e che condividono la stessa narrazione (Bessi et. al., 2015). Questo filone di studi ha respinto la buona riuscita di una delle azioni di contrasto alle *fake news* messe in atto dalle grandi redazioni giornalistiche internazionali: quella del *fact checking* e *debunking.* L'analisi del comportamento degli utenti che consumano fonti di informazione complottista - divisi tra quelli sottoposti a *debunking* e quelli che non lo sono stati - ha dimostrato quanto i primi hanno una probabilità più elevata di continuare a interagire con informazioni complottiste rispetto ai secondi. Questo dimostra che provare a convincere

[21] Come tempi di condivisione, *link*, utilizzo di diverse lingue e altri metadati.

del contrario un sostenitore delle teorie del complotto non fa che produrre un effetto di rinforzo della sua convinzione che si manifesta in una maggiore interazione con le fonti di informazione piegate sulla propria visione del mondo (Bessi et. al., 2014).

È indubbio che lo studio delle dinamiche sociali sulle piattaforme algoritmiche abbia contribuito in maniera rilevante alla comprensione del comportamento degli utenti che fruiscono di informazione alternativa, ma l'approccio esclusivamente *data-driven* sembra aver ridotto la complessità del fenomeno disinformativo ad una mera dicotomia vero/falso, scientifico/complottista e considerato individui diversi e comportamenti sociali complessi come unico soggetto/utente dal comportamento indifferenziato. Ciò che ci si tenta di fare in questo studio è proprio indagare la possibilità di distinguere l'atteggiamento degli utenti sulla base di qualche fattore emergente e provare a ricostruire le narrative e le visioni del mondo che questi costruiscono durante le loro pratiche di interazione con i contenuti disinformativi diffusi in queste bolle.

Conclusioni

Dal lavoro della comunità scientifica sul fenomeno disinformativo emerge l'incapacità di fondo delle soluzioni algoritmiche pensate dalle piattaforme proprietarie per arginare la formazione, diffusione e rinforzo di opinioni provenienti dal consumo di contenuti disinformativi. Creare uno strumento unico, automatico ed efficace per il rilevamento di notizie false è molto complesso: setacciare informazioni provenienti da milioni di messaggi di natura eterogenea e dinamica si scontra inevitabilmente con una serie di limiti che riguardano principalmente il tempo, l'accesso ai dati, il *bias* della lingua e la complessità di un ambiente e di un fenomeno in continua trasformazione. I dati *online* sono sensibili al tempo e di conseguenza anche agli argomenti ed eventi di tendenza; inoltre, considerata la disponibilità ormai limitata di dati provenienti dalle piattaforme, risulta complesso analizzare le caratteristiche dei messaggi disinformativi con l'obiettivo di creare modelli di apprendimento supervisionato che operano in piena autonomia. La maggior parte delle risorse *open access* esistenti si concentrano principalmente su un solo tipo di notizie (spesso politiche) in lingua inglese, questo rende il *training* e la verifica di soluzioni basate sul *machine learning* piuttosto limitata. Sebbene vi sia un evidente successo di analisi che utilizzando vari approcci di apprendimento automatico nel rilevamento del falso, tuttavia, la continua evoluzione delle strategie disinformative rappresentano una continua sfida alla categorizzazione del fenomeno che non può mai definirsi conclusa. A questi limiti si aggiunge una lacunosa ricerca empirica sulla disinformazione veicolata attraverso le immagini: la maggior parte dei lavori si basa sulla componente testuale escludendo un tipo di contenuto in continua crescita sulle piattaforme *social* e che non è possibile analizzare utilizzando le soluzioni fin qui proposte.

La letteratura sulle *fake news* è in continua in espansione. Fornendo uno sguardo alle soluzioni maggiormente adottate dalla comunità scientifica, questo capitolo ha provato a sintetizzare l'attuale conversazione sul fenomeno specialmente alla luce dell'avanzamento tecnologico. Non è sufficiente combattere le notizie false con quelle verificate, occorre una comprensione più sfumata e sempre aggiornata dei processi - individuali e sociali, culturali e comunicativi - che creano e sostengono la disinformazione.

CAPITOLO III - *Content Analysis* tra tradizione e computazione.

Corbetta nel 1992 scriveva che *«il maggior produttore di materiale documentario sulla società è probabilmente costituito dal sistema dei mezzi di comunicazione di massa»* (p. 454) e gli ultimi vent'anni gli hanno dato ragione.

L'abbondanza di dati digitali è diventata una caratteristica distintiva della complessa società contemporanea e in particolare della comunicazione che oggi si esprime soprattutto attraverso il *web* e le piattaforme sociali. Recensioni, *tweet*, *like*, collegamenti, condivisioni, *post*, *tag*, *ecc*. sono solo una parte di miliardi di tracce digitali che quotidianamente lasciamo in rete e attraverso le quali è possibile ricostruire con precisione i nostri gusti, le nostre opinioni e i nostri atteggiamenti. Una delle peculiarità della comunicazione mediata dai dispositivi digitali è la trasposizione del linguaggio parlato nello scritto che ha generato una produzione senza precedenti di dati testuali che si caratterizzano per essere una continua e sempre crescente risorsa di informazione e opinioni. I *big corpora* provenienti dalla rete rendono l'analisi testuale una delle tecniche più rilevanti per la comprensione dei fenomeni sociali *inside* e *outside platforms*.

È possibile prevedere sondaggi di opinione locale utilizzando grandi volumi di dati provenienti dalle piattaforme (Beauchamp, 2017)? È possibile comprendere la censura in Cina basandosi su milioni di post dei *social network* (King, Pan e Roberts, 2013)? E ancora: è possibile prevedere i conflitti armati utilizzando il testo dei giornali e riducendo grandi quantità di articoli ad argomenti interpretabili (Mueller e Rauh, 2018)?

È possibile perché gli scienziati sociali, da decenni interessati all'analisi della comunicazione testuale, negli ultimi anni hanno sfruttato la proficua base empirica proveniente dagli spazi digitali per creare nuove opportunità di studio sulla natura interattiva dell'informazione e della comunicazione mediata dalle piattaforme nonché sull'impatto che questa mediazione genera nelle pratiche del quotidiano.

Di pari passo con l'esplosione dei dati testuali c'è stata un'impressionante evoluzione degli strumenti di analisi: l'abbondanza e varietà di *software* e *tools* spingono le scienze sociali in uno scenario in cui *«la ricerca mediata dal web [...] sta già trasformando il modo in cui i ricercatori praticano metodi di ricerca tradizionali trasposti sul web»* (Amaturo e Punziano, 2016, pp. 35,36).

La metafora del computer come macroscopio (Bennato, 2015) - strumento utile ad osservare i processi sociali che si estendono sia nel tempo che nello spazio - è perfetta per comprendere l'importanza del supporto tecnico alla comprensione della complessa società contemporanea non escludendo il ruolo del ricercatore che, attraverso di esso, interpreta la realtà e produce nuova conoscenza. Difronte all'ampiezza e all'eterogeneità dei dati che oggi è possibile acquisire e di fronte alla complessità degli oggetti di indagine e dei fenomeni osservabili, il ricercatore necessita di costruire percorsi di studio e

di analisi tali da combinare tradizione e computazione, avvalendosi di strumenti che non siano esclusivi né dell'uno né dell'altro approccio. Come accade per ogni grande cambiamento, il rapido processo di quantificazione, dovuto all'esplosione dei *big data* e all'inevitabile sviluppo di strumenti in grado di gestirne il peso computazionale, ha diviso la comunità scientifica tra critici e ottimisti. C'è chi ha considerato quest'abbondanza una rivoluzione nel modo di indagare la società (Lazer et al., 2009) e chi invece l'ha ritenuta una possibile minaccia per la classica sociologia empirica nonché un pericoloso passaggio da un approccio paradigmatico *theory driven* ad uno *data driven* (Amaturo e Aragona, 2019). L'idea portata avanti dai critici è che i *big data* e le nuove tecnologie ci permettono di tracciare qualsiasi tipo di comportamento umano con fedeltà e precisione tali da rendere obsoleto il metodo scientifico tradizionalmente inteso. Secondo quest'idea la ricerca guidata dai dati permetterebbe di avanzare conoscenza anche senza teorie di riferimento. La preoccupazione riguardo il favorire di una scienza volta esclusivamente a tirar fuori modelli dai dati è fuorviante in quanto confonde l'identificazione di regolarità all'interno dei dati con l'interpretazione e la scoperta dei meccanismi dai quali quelle regolarità derivano (ivi). Questa interpretazione risulta problematica senza la dotazione di un orientamento paradigmatico o la guida di una teoria di riferimento. In un'ottica più ampia invece, una coorte di sociologi digitali - con conoscenza profonda e contestualizzata del proprio oggetto di indagine nonché dei limiti e delle possibilità offerte dal digitale - opera elaborando percorsi di ricerca volti ad adottare una postura critica sul ruolo che la tecnologia digitale può avere nella ricerca scientifica, ma allo stesso tempo creativa, combinando le diverse possibilità offerte proprio dalla tecnologia (ivi).

Le stesse scienze sociali computazionali, regine dell'approccio *data-driven*, possono essere inserite in un filone paradigmatico ben preciso che ne definisce le prospettive e ne orienta l'indagine, quello dell'elaborazione delle informazioni. Come descritto da Bennato (2015) il paradigma considera le informazioni come gli elementi alla base dello scambio sociale e orienta lo studio dei sistemi sociali e dei loro processi a partire proprio dallo scambio informativo e dall'attivazione di canali di comunicazione. Su queste basi, l'informatica non può che fornire validi strumenti di supporto all'analisi della complessità sociale proprio in qualità di scienza che si occupa dell'ordinamento, del trattamento e della trasmissione delle informazioni per mezzo dell'elaborazione elettronica (ivi). Il citato supporto dell'informatica sottolinea la seconda preoccupazione di quanti considerano l'avvento dei *big-data* una minaccia alle pratiche empiriche delle scienze sociali, ossia l'idea che le tecniche computazioni possano in qualche modo sostituire quelle tradizionali. Questa idea è figlia di una dicotomia analogico-digitale che non ha più ragione di esistere: lo stesso Rogers (2013) ha distinto le tecniche di rilevazione digitalizzate, ossia tecniche "analogiche" trasposte negli spazi digitali (come le

web survey), da quelle native digitali, ossia nate direttamente nelle infrastrutture della rete (come le interrogazioni delle API o le tecniche di *scraping*).

A guardar bene, le tecniche computazionali si basano su approcci ampiamente riconosciuti e consolidati nelle scienze sociali, ciò che cambia è la possibilità di elaborazione di grosse quantità di dati e quindi la possibilità di condurre studi su larga scala in grado di analizzare i fenomeni in modo longitudinale. È proprio il superamento dei limiti spaziali e temporali uno degli elementi che entusiasma quella parte di comunità scientifica che considera rivoluzionario il contributo del digitale alla ricerca sociale: il ricercatore ha la possibilità di avere un accesso (relativamente) immediato a *set* di dati che possono essere analizzati in maniera diacronica, nonché la possibilità di osservare i comportamenti, le relazioni e le pratiche sociali che emergono quando le persone non sanno di essere indagate, fornendo così un livello di veridicità alla ricerca non riscontrabile nei dati raccolti attraverso intervista (Lupton, 2014). Questa robusta base empirica necessita, al tempo stesso, di metodi computazionali per assemblare, ordinare, filtrare e processare gli oggetti digitali, e solide teorie per interpretare tutta la ricchezza di informazioni in essi contenuta.

Seguendo l'approccio epistemologico digitale proposto da Enrica Amaturo e Biagio Aragona (2019), questo studio assume una posizione metodologica pluralistica, pragmatica e al tempo stesso critica. Il percorso di ricerca messo a punto utilizza tecniche tradizionali e computazionali in una prospettiva ibrida ed escogita creativamente soluzioni metodologiche ogni qual volta si scontra con i limiti connessi ai dati digitali, al funzionamento delle tecniche di *machine learning* o alle componenti tecnologiche. Nel percorso di metodo il ruolo conferito alla tecnologia è chiarito già in fase di disegno della ricerca: le tecniche computazionali vengono utilizzate per scoprire evidenze empiriche successivamente sottoposte a controllo del ricercatore, mentre la teoria è impiegata per dare profondità all'interpretazione dei dati e direzionare, di volta in volta, lo svolgimento della scoperta. Il metodo utilizzato è quello abduttivo proposto da Peirce che mira ad inserire risultati inaspettati in un quadro interpretativo generale (Amaturo e Aragona, 2019).

1 Un approccio ibrido alla *Content Analysis*

Molti studiosi hanno discusso sul ruolo dell'analisi del contenuto tradizionale nel nuovo ambiente di informazione digitale (Amaturo e Punziano, 2013; McMillan, 2000; Weare & Lin, 2000). Chiunque consultasse un classico manuale di riferimento dell'analisi del contenuto (da Krippendorf a Losito, da Rositi ad Amaturo) si renderebbe conto dell'attualità e della valenza che questi testi hanno tutt'oggi nell'orientare un ricercatore che voglia approcciarsi a questo tipo di analisi, anche individuando nuovi ambiti applicativi o nuove forme di comunicazione tipiche dello sviluppo dei social media (ad esempio video *stories*, *meme*, *ecc*.). L'analisi del contenuto sembra sposare bene l'approccio epistemologico adottato in questo studio: essendo una tecnica versatile è possibile ottenere risultati interessanti utilizzando tecniche e approcci differenti in combinazione tra loro; quindi, nella stessa esperienza di ricerca si presta bene ad approfondimenti qualitativi e altri suscettibili di quantificazione, anche sofisticata, come la modellizzazione di argomenti o il *machine learning* (Faggiano, 2016). Mentre per qualcuno solo la codifica umana può offrire la profondità necessaria per classificare validamente i significati testuali, per altri gli algoritmi sono in grado di eseguire una crescente varietà di compiti con precisione ad essi comparabile. In realtà, uno dei campi in cui il fattore umano conserva il suo vantaggio è proprio quello dello studio della comunicazione *online*, che spazia dall'identificazione di modelli semantici mesoscopici (valori, identità, simboli) a pratiche pragmatiche (minacce, *hate speech*, *body shaming*) in un discorso naturale altamente variabile (Baden, Kligler-Vilenchik e Yarchi, 2020). Oltre alla versatilità sopra descritta, si riconosce all'analisi del contenuto anche una vigorosa varietà che riguarda i tipi di testo sottoposti ad analisi: questi possono spaziare dagli articoli di giornale alle dichiarazioni dei politici, dalle recensioni ai *tweet* e commenti lasciati dagli utenti. Ai *big corpora* provenienti dalle piattaforme *social media* e codificati in linguaggio naturale, l'analisi del contenuto può offrire sfumature e *insight* che la sola elaborazione computazionale non restituirebbe.

L'analisi del contenuto, nella sua definizione classica, è una tecnica utile a:

«scomporre, in modo sistematico, l'unità comunicativa […] in elementi più semplici utilizzando cioè criteri espliciti e standardizzati da applicare all'intera unità in oggetto; successivamente gli elementi individuati sono classificati in un insieme di categorie e dunque trasformati in variabili categoriali o ordinali che è possibile sottoporre a trattamenti statistici di vario tipo» (Amaturo e Punziano, 2013, 24).

Nonostante la disputa sulle numerose definizioni[22], la maggior parte degli studiosi concorda sul fatto che l'analisi del contenuto debba essere oggettiva e sistematica: ogni passaggio deve seguire

[22] Per un approfondimento si rimanda al volume Amaturo E, Punziano G. (2013), *Content Analysis. Tra comunicazione e politica* (a cura di), Milano, Ledizioni.

costantemente e in modo esplicito, regole e procedure standardizzate. Tutte queste considerazioni sollevano inevitabilmente una serie di domande: la quantità di dati a nostra disposizione e i potenti strumenti per analizzarli ci forniscono maggiori e migliori informazioni sui fenomeni indagati? È possibile sfruttare al meglio il rigore sistematico e la consapevolezza contestuale dell'analisi del contenuto tradizionale servendosi allo stesso tempo dell'efficienza dei metodi computazionali?

In origine, l'analisi del contenuto si caratterizzava per due elementi: l'oggetto specifico della conoscenza, limitato ad un singolo elemento della comunicazione (che cosa è comunicato) e l'ambito metodologico in cui si muoveva, profondamente influenzato dall'approccio *survey* (Amaturo e Punziano, 2013). A partire dagli anni Sessanta, lo sviluppo della linguistica e della semiotica ha richiamato l'attenzione sulla complessità degli atti comunicativi rendendo insufficiente un'analisi del contenuto che si limitasse esclusivamente allo studio manifesto della comunicazione[23]. Con la possibilità di stabile inferenza, l'analisi del contenuto si impone dalla fine degli anni Cinquanta come tecnica in grado di far emergere i significati latenti della comunicazione, caratterizzata dalla preferenza per l'analisi quantitativa e intensificata dallo sviluppo dell'informatica e dal supporto del computer per il trattamento di ampi *corpora* testuali. L'amore folle tra analisi del contenuto e computer (Rositi, 1989) ha reso questo settore di studi bersaglio di numerose critiche (soprattutto di stampo semiotico) circa il timore che l'aspirazione di presunta oggettività, ottenibile attraverso l'approccio *computer-aided,* non tenesse conto di tutti i delicati problemi di analisi per i quali è necessario l'intervento del ricercatore (Amaturo e Punziano, 2013). Queste preoccupazioni sono ancora vive nei recenti approcci computazionali all'analisi del testo, i cui limiti risiedono nell'incapacità di classificare in maniera automatica dei testi secondo categorie teoricamente fondate (Boumans & Trilling, 2016) e nella difficile interpretazione del linguaggio umano in tutta la sua ricchezza, complessità e sofisticatezza. D'altro canto, le tradizionali forme di analisi del contenuto non sono state messe a punto per gestire gli enormi set di dati testuali con cui entriamo in contatto quando vogliamo condurre una ricerca su *web* e piattaforme. Un discreto filone di studi internazionali ha così iniziato a testare quanto possa essere fruttuoso combinare gli approcci computazionali e tradizionali di analisi del contenuto all'interno della stessa ricerca utilizzando schemi di codifica derivati da categorie guidate dalla teoria. In questa prospettiva, i metodi computazionali sono utilizzati per migliorare e non soppiantare tecniche consolidate, fornendo uno strumento in grado di affrontare l'analisi di ampi *corpora* non rinunciando alle sfumature contestuali. Un approccio ibrido può conservare i punti di forza della tradizionale analisi del contenuto, ossia rigore sistematico e consapevolezza testuale, massimizzandone le capacità su larga

[23] Berelson (1952, p.18) definisce l'analisi del contenuto come «una tecnica di ricerca per la descrizione obiettiva, sistematica e quantitativa del contenuto manifesto della comunicazione».

scala attraverso le tecniche computazionali. Inoltre, potrebbe risultare ulteriormente produttivo se combinato con strumenti di elaborazione del linguaggio naturale, particolarmente sensibili al contesto linguistico e più adatti a gestire istruzioni complesse.

> *«Il potere di calcolo garantisce qualità, precisione e scala nella registrazione di elementi specifici della piattaforma, mentre le pratiche collaudate dell'analisi del contenuto assicurano la valutazione della categorizzazione tematica»* (Sjøvaag et al., 2012, p.93).

È in questa prospettiva che si muove questo lavoro di ricerca: la tradizionale analisi del contenuto è stata combinata con tecniche di *text mining* e *machine learning* implementate su un set di dati ampio e non strutturato a diversi livelli di complessità computazionale. L'idea è quella di uno scambio fertile e armonioso tra tecniche tradizionali, attualizzate nell'ambito del digitale e nuove strategie di analisi, provenienti da più ambiti disciplinari.

2 Fare ricerca digitale *Api-based*

Come viaggia e muta la disinformazione nell'ambiente disintermediato delle piattaforme *social media*? Quali strategie utilizza? Di quali temi si nutre? Come si adatta ai cambiamenti della piattaforma stessa? Quali processi culturali e comunicativi emergono tra gli utenti che consumano contenuti disinformativi?

Per rispondere a queste domande, la ricerca si inserisce nella prospettiva metodologica dei Digital Methods con un approccio che si ispira al modo in cui i dati digitali sono naturalmente organizzati dalle infrastrutture tecniche degli ambienti digitali (Rogers, 2013). Lo studio è stato oggetto di un'ampia riflessione metodologica, il cui obiettivo è stato quello di ponderare l'affidabilità della definizione operativa con l'efficienza della raccolta dei dati digitali e del processo di analisi.

La ricerca digitale per molti anni si è basata sulla raccolta e l'analisi di dati semi strutturati e non strutturati provenienti dalle piattaforme e raccolti attraverso l'interrogazione delle API (*Application Programming Interface*): un approccio alle scienze sociali computazionali e alla sociologia digitale basato sull'estrazione di *record* di dati resi disponibili dalle piattaforme attraverso le loro interfacce di programmazione (Venturini & Rogers, 2019). Attraverso queste interfacce, la raccolta di dati è stata resa (relativamente) semplice grazie al lavoro congiunto di ricercatori e sviluppatori che hanno collaborato affinché l'esperienza di chi ne facesse uso fosse più intuitiva possibile proprio perché consapevoli delle potenzialità offerte da questa procedura di collezione dati nell'esplorazione e conoscenza dei fenomeni sociali in rete. Per molto tempo, questo tipo di ricerca ha permesso di raccogliere informazioni dettagliate su vaste popolazioni digitali, superando efficacemente la distinzione tra metodi qualitativi e quantitativi.

A causa dello scandalo Cambridge Analytica e di una serie di restrizioni già precedentemente avviate, oggi l'accesso libero alle API è completamente chiuso e regolamentato esclusivamente dalle società proprietarie delle piattaforme che consentono l'accesso soprattutto a quelle realtà con le quali è più facile monetizzare (Bruns, 2019) come ad esempio grandi società di *marketing* e pubblicità. A soffrire maggiormente di questa chiusura è stato il mondo accademico. Oggi, non esiste alcun modo gratuito per estrarre contenuti dalle piattaforme senza violarne i TOS *(Terms of Service)*; possibilità alternative, come la *partnership* di Facebook con Social Science One o CrowdTangle, rappresentano sostituti insufficienti alle API in quanto incapaci di garantire una fornitura di dati completamente indipendente dagli interessi della piattaforma. L'attuale cimitero di API e di *tools* interrotti - molti dei quali ideati, programmati e messi a disposizione della comunità scientifica proprio dal gruppo della Digital Methods Initative di Richard Rogers – dimostra quanto questi strumenti abbiano rappresentato una preziosa risorsa nella cassetta degli attrezzi del ricercatore digitale. L'attuale chiusura ha però

evidenziato quanto sia pericoloso affidarsi completamente ed esclusivamente ad un unico strumento di raccolta dati.

A mio avviso, l'attuale situazione di immobilità generata dalla chiusura delle API sta producendo tre conseguenze: la prima è la migrazione verso "dati facili" (da raccogliere) quindi siti *web* o piattaforme meno restrittive sull'accesso all'API e alla possibilità di *data scraping.* Questo atteggiamento nutre l'idea errata che tutte le piattaforme siano utili per indagare indistintamente un particolare fenomeno digitale: ogni piattaforma ha la sua particolarità strutturale e interazionale, nonché una differente pervasività nella quotidianità delle persone, che si differenzia per fascia di età e cambia da paese in paese[24]. Ci sono piattaforme più permissive sulla raccolta dati che però, in Italia, non possono essere considerate popolari (in termini di utilizzo) al pari di Facebook, Instagram o WhatsApp e dunque non sempre utili nell'indagare particolari fenomeni sociali *online*.
La seconda conseguenza è un ritorno a forme di raccolta dati diretta; questo costringe i ricercatori a lavorare su *set* di dati di piccole dimensioni e ad osservare le dinamiche sulle piattaforme attraverso le stesse interfacce degli attori che studiano. La terza conseguenza è l'incoraggiamento alla violazione deliberata dei TOS: il crescente interesse della comunità accademica a questa pratica risponde all'idea che, in alcune circostanze, i benefici per la società derivanti dalla violazione dei termini di servizio superi i danni alla piattaforma (Rogers, 2018; Venturini & Rogers, 2019), ignorando volontariamente numerose e delicate questioni etiche a cui il ricercatore sociale è chiamato a rispondere.
L'alternativa del *web scraping* sulle piattaforme *social* è problematico non solo dal punto di vista legale ma anche da un punto di vista pratico: considerando il funzionamento di piattaforme come Facebook, che modella i contenuti da sottoporre agli utenti attraverso sofisticati algoritmi, lo *scraping* raccoglierebbe ciò che l'algoritmo ha modellato sugli interessi del ricercatore/utente che accede alla piattaforma in quel dato momento piuttosto che una selezione di contenuti neutra come quella fornita dall'API. Inoltre, c'è da considerare che i rigidi controlli avviati dalle società proprietarie non permettono quasi mai la buona riuscita di questa pratica. Per affrontare l'era "post APIcalittica", come l'ha definita Axel Bruns (2019), la Digital Methods Initative di Richard Roger (edizione Winter School 2020[25]) ha sollevato la necessità di battere strade alternative per la ricerca digitale, che non si pieghino all'ottimizzazione di informazioni contenute nei dati messi a disposizione da *partnership* come Social Science One o CrowdTangle, né all'acquisto di dati da servizi terzi che potrebbero presentarsi troppo

[24] In Italia, ad esempio, l'utilizzo di Twitter è completamente diverso da quello negli Stati Uniti o in altri paesi europei. Questa consapevolezza è indispensabile per scegliere la piattaforma (o le piattaforme) che meglio si prestano ad indagare in fenomeno d'interesse.
[25] Alla quale chi scrive ha personalmente preso parte.

ancorati a bisogni di mercato e quindi non sempre adatti alle esigenze di ricerca. La discussione sul libero accesso alle API non deve essere considerata come una dipendenza da *big social data* per le ricerche sul digitale (esistevano dinamiche collettive *online* anche prima dei *social network*) ma è indiscutibile quanto queste piattaforme siano diventate fondamentali per comprendere molte delle sfide urgenti della società globale: dalle campagne di disinformazione, alla polarizzazione del dibattito pubblico, al monitoraggio della diffusione delle pandemie.

Secondo Rogers (2018), quando si discute di alternative si deve partire dall'osservazione fatta da Tim Berners-Lee (co-inventore del *web*) che denuncia quanto il *web* aperto sia in declino a causa dell'aumento della sorveglianza e della crescita del potere delle piattaforme. Sono state avanzate una serie di proposte per cambiare il panorama della ricerca digitale Api-*based*, ma la produzione accademica basata sull'utilizzo di dati provenienti da piattaforme proprietarie di Facebook è enorme rispetto a quella che utilizza le alternative.

È necessario dunque spingere il ragionamento sulle conseguenze di questa chiusura proprio per il mondo accademico: la ricerca sistematica sui *social media* e i loro effetti è ormai insostenibile se non portata avanti attraverso l'utilizzo di servizi che la stessa piattaforma mette a disposizione dei ricercatori.

Questa sorta di autoritarismo della conoscenza sta trasformando quello che era già un *social network* opaco in una vera e propria scatola nera che fugge dalle proprie responsabilità nei confronti dei legislatori e del pubblico. Al tempo stesso, questi ultimi hanno beneficiato del lavoro portato avanti dagli accademici nel conoscere e comprendere le dinamiche che prendono vita nella piattaforma e gli effetti che si determinano al di fuori di essa.

In ogni caso, qualunque sia la tecnica di raccolta che si intende adottare, ciò che riusciamo a collezionare rappresenta solo una parte dell'immensa quantità di dati presenti sulle piattaforme.

Infatti, una delle critiche mosse alla ricerca digitale sui *social media* è che le piattaforme non sono strumenti scientifici per raccogliere dati sulle tendenze della società in quanto la loro attività di archivio dati ha lo scopo di segmentare il pubblico per vendere pubblicità (Rogers, 2018). La raccolta dati dalle piattaforme è effettivamente uno dei punti più controversi da affrontare nelle ricerche sui *social media*: questi dati non sempre sono completi e stabili nel tempo: alcuni elementi svaniscono altri appaiono. Questo introduce il concetto di "complessità interattiva" (ivi) poiché alcuni oggetti raccolti in un preciso momento (ad esempio i *like*) possono essere influenzati da nuovi elementi che vengono introdotti successivamente (ad esempio le *emoticon*). Dunque, se vogliamo esaminare i *like* nel tempo per determinare sentimenti o preferenze, i cali o gli aumenti di questa metrica possono dipendere tanto da modifiche apportate all'algoritmo della piattaforma quanto al reale cambiamento di opinione degli

utenti. Secondo Richard Rogers (2018), i buchi nei dati si creano per una serie di motivi, il più comune dei quali è l'impostazione di restrizioni nazionali: le autorità nazionali potrebbero chiedere alle piattaforme di eliminare o limitare la circolazione di particolari contenuti, ad esempio contenuti politici con posizioni estremiste. Così per rispettare le richieste del governo, la piattaforma non rende più fruibili quei particolari contenuti in quel paese, ma ciò non toglie che potrebbero esser visti altrove. La raccolta di dati di posizione politiche estremiste di quel paese sarebbe teoricamente eseguita meglio al di fuori del paese stesso al fine evitare vuoti nei dati dovuti a restrizioni governative.
Un altro elemento che può restituire *set* di dati incompleti è legato alle impostazioni sulla privacy: le pagine Facebook, ad esempio, possono aver impostato limiti di paese ed età; dunque, a seconda di dove si raccolgono i dati o di chi li raccoglie, alcune di queste pagine potrebbero essere escluse senza che il ricercatore ne sia a conoscenza. A queste incertezze si aggiunge che le piattaforme *"doesn't represent all people"* (Boyd & Crawford, 2012) ma un sottoinsieme molto particolare: non tutti gli individui sono raggiungibili in rete, c'è chi non ha materialmente la possibilità di connettersi alla rete e c'è chi si configura esclusivamente come un consumatore passivo di contenuti senza partecipare attivamente alle pratiche prosumeristiche del *web* 2.0. È necessario infine tener presente che la natura mutevole delle piattaforme rende difficile generalizzare i risultati di un campione di dati ad una popolazione che cambia rapidamente.
Gli elementi della piattaforma sono dunque instabili e lo sono ancor di più le metriche integrate ai dati che ci restituiscono i servizi di raccolta post-APIera. Un esempio è CrowdTangle, lo strumento di monitoraggio dei *social media* di proprietà di Facebook. Questo strumento, nato per misurare l'impatto dei post sugli utenti, restituisce[26] un valore di *engagement rate* (o tasso d'interazione): questo valore segue le logiche commerciali della piattaforma e fa riferimento al concetto più generale di *customer engagement* quindi in alcun modo utile a rilevare il coinvolgimento effettivo degli utenti con un particolare tipo di contenuto.

Per tutti questi motivi, fare ricerca digitale significa prima di tutto cambiare l'orientamento della ricerca: non solo sulle piattaforme ma anche con le piattaforme; considerando questi elementi non solo come fonte di dati ma anche come fonte di metodi e tecniche (Rogers, 2013). I dati digitali ci permettono di condurre analisi di tipo post demografico, questo implica necessariamente una diversa identificazione dell'unità di analisi che non potrà essere l'individuo nel suo essere singolo ma l'individuo come parte integrante di aggregati sociali non riconducibili a particolari categorie socio-demografiche (Airoldi, 2017) ma considerato sulla base delle proprie attività che si costituiscono

[26] Nella versione utilizzata al momento in cui si scrive.

dall'interazione con altri individui e con il *digital device*. Questo tipo di orientamento consente di cogliere la cultura condivisa, le opinioni e la percezione dei fenomeni che emergono dall'esposizione degli utenti agli ambienti digitali (in generale) e algoritmici (in particolare). Oggetto di studio diventano anche le stesse tecniche adottate dal ricercatore per analizzare questo tipo di dati: considerato che gli oggetti di studio del digitale dipendono tanto dalle pratiche sociali trasposte in rete quanto da quelle tecniche utilizzare per indagarli, rappresentare i propri oggetti di studio attraverso tecniche digitali vuol dire anche avere le tecniche digitali come oggetti di studio, in una continua dialettica tra oggetti e metodo (Amaturo e Aragona, 2019).
In questa prospettiva, l'assunto *"follow the medium"* (Rogers, 2013) non significa solo aggiungere un nuovo strumento tecnico alla cassetta degli attrezzi metodologica ma è un invito a far proprie le logiche che il *web* e le piattaforme applicano a sé stesse per raccogliere, ordinare e analizzare i dati. Il ricercatore deve conoscere a fondo gli ambienti digitali, il loro funzionamento, nonché i vantaggi e i limiti ad essi connessi.
Le sfide che pongono questi dati sono numerose ma sarebbe un grave errore escluderli. L'incontro tra tecniche tradizionali e tecniche computazionali è necessario proprio per superare i limiti di una con i punti di forza dell'altra e per far emergere quelle dimensioni profonde che l'utilizzo esclusivo di una sola prospettiva analitica lascerebbe latenti. In questo modo è possibile fornire un prezioso arricchimento alla comprensione dell'agire sociale in rete senza rinunciare all'esplorazione longitudinale.

Nel caso specifico di questo studio, la raccolta di dati tramite API è stata fondamentale per studiare non solo come la disinformazione viaggia nel sistema mediatico delle piattaforme, ma anche per indagare nel profondo come si radicalizzano le posizioni degli utenti immersi nelle proprie *echo chambers* disinformative. In questo studio, le restrizioni imposte dalla piattaforma si sono trasformate in un'opportunità per costruire nuove strade metodologiche e per estrarre conoscenza dai dati.

3 Fasi e tecniche di analisi

Per circoscrivere il campo digitale d'indagine si è scelto di concentrare l'analisi sulla piattaforma Facebook perché in Italia si classifica come il *social network* più utilizzato[27]: circa il 75% (51% lo utilizza per accedere alle *news*) contro il 22% di Twitter (11% lo utilizza per accedere alle *news*)[28]. Oltre al diffuso utilizzo, la strutturazione degli spazi digitali della piattaforma (come gruppi e *fan page*) favorisce più di altre la possibilità per gli utenti di riunirsi intorno a narrative condivise, a promuoverle e a discutere con persone che la pensano allo stesso modo. Infine, la scelta è stata sostenuta da precedenti studi empirici che hanno dimostrato la presenza limitata del fenomeno indagato su piattaforme come Twitter (Cinelli et al., 2020).

La sfida immediatamente emersa nel dedicare l'analisi alla piattaforma Facebook è stata quella relativa alla collezione di una consistente quantità di dati malgrado i limiti tecnici che Facebook stava gradualmente imponendo. L'API della piattaforma ha fornito la possibilità di selezionare le pagine d'interesse e scaricarne i dati in un arco temporale definito. Alla data in cui si scrive, questa possibilità di raccolta libera non è più possibile perché definitivamente chiusa o fortemente ridimensionata su quasi tutte le piattaforme.

Per questa ricerca, questa tecnica di raccolta dati (probabilmente tra le ultime a farne uso) è stata preferita ad altre per due motivi: non costringe il ricercatore ad osservare le dinamiche sociali sulla piattaforma attraverso la stessa interfaccia degli attori che sta studiano e fornisce una selezione neutra di contenuti non influenzata dalla personalizzazione algoritmica che è alla base del funzionamento stesso della piattaforma.

Per la costruzione della base empirica sono state individuate dieci pagine disinformative[29], selezionate dalle *blacklists* di siti specializzati in *fact checking* e *debunking.* Questo tipo di selezione è stata preferita alla *query* perché anche questa pratica risente della pressione dell'algoritmo che spinge verso una scelta di pagine modellata sulle preferenze del ricercatore/utente che sta effettuando l'interrogazione[30].

[27] La televisione risulta essere ancora la principale fonte di notizie, ma è importante tener conto del fatto che l'Italia è un paese con una diffusa percentuale di anziani che fanno affidamento a fonti di notizie tradizionali. D'altro canto, c'è una fascia di persone, i giovani e i nativi digitali, che utilizzano i nuovi media in modo esclusivo.

[28] Reuters Institute Digital News Report 2017, Oxford University

[29] Considerata la natura delle pagine e l'alto rischio di censura a cui sono soggette è stata creata una lista di sostituzione nella quale sono state individuate altre dieci pagine che rispondevano agli stessi criteri. Queste sarebbero state utilizzate nel momento in cui qualcuna delle pagine individuate fosse stata eliminata dalla piattaforma.

[30] Interessante è stato notare come, una volta seguite le pagine d'interesse, la piattaforma Facebook ha iniziato a proporre numerose altre pagine dai contenuti simili.

I criteri alla base della scelta delle pagine sono:

a) Alla data di raccolta dei dati le pagine dovevano avere un numero di seguaci superiore o uguale a 10.000 che rappresenta la soglia minima di *followers* per essere considerato "influente" su un *social media*. Il superamento di questa soglia permette alla pagina di emergere dalla massa di pagine presenti sulla piattaforma[31].

b) Alla data di raccolta dei dati le pagine dovevano essere già attive nel 2016.

In una società *"data intensive"* (Amaturo e Aragona, 2016) l'obsolescenza dei dati è un fattore di rischio per le ricerche sul digitale e con il digitale. Per limitare questo effetto si è optato per la scelta di un arco temporale molto ampio con l'obiettivo di individuare delle costanti del fenomeno che fossero empiricamente rilevanti: l'intervallo temporale della ricerca va da gennaio 2016, passato alla storia come anno della disinformazione (in cui la parola post-verità diventa neologismo dell'anno secondo gli esperti di Oxford Dictionaries[32]) fino ad aprile 2020 (data dopo la quale è stato impossibile collezionare dati).

Questa diacronia è stata possibile grazie proprio al libero accesso all'API (sfruttato fino all'ultimo giorno disponibile) che ha consentito di raggiungere una popolazione di circa 18.767 post e oltre un milione e mezzo di commenti (precisamente 1.549.872). Questi numeri, soprattutto per quanto riguarda i commenti degli utenti, sono oggi impensabili per un semplice ricercatore che voglia condurre una ricerca come questa in completa autonomia.

Il campione è stato costruito sul 10% della popolazione ed estratto attraverso un tipo di campionamento stratificato proporzionale.

Per risolvere il problema della poca numerosità di alcune pagine si è optato per un'aggregazione sulla base del numero di seguaci e degli anni considerati nell'analisi.

Il numero di seguaci è stato aggregato creando tre strati:

- pagine da 10.000 a 100.000 *followers*;
- pagine da 101.000 a 500.000 *followers*;
- pagine con oltre 500.000 *followers*.

Gli anni sono stati aggregati considerando il 2019 come spartiacque tra l'era pre-API e quella post-API, determinante nella raccolta dei dati.

I due strati creati sono:

- anni 2016-2018: pre-API;

[31] Sui social media sembra correre una vera e propria "sindrome dei 10k" poiché da questa soglia di follower in poi, l'utente entra in una vera e propria "élite digitale" nella quale gode di alcuni servizi forniti dalle piattaforme esclusivamente a chi raggiunge questo traguardo.

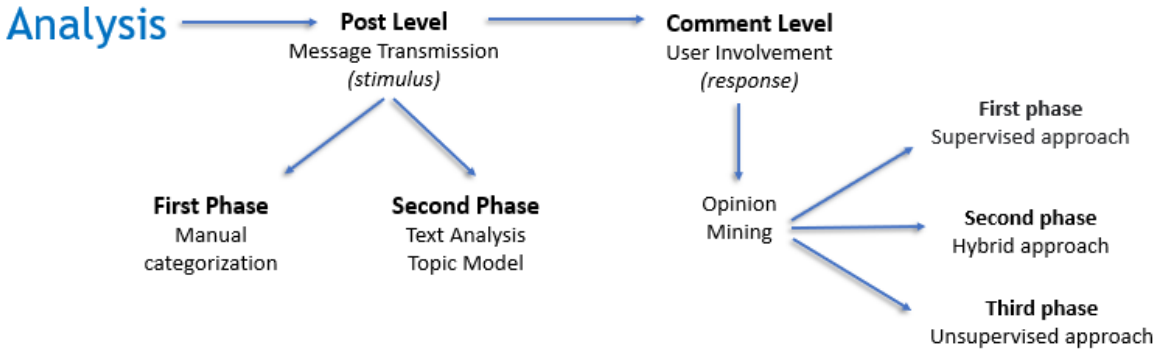
[32] Consultabile al link: https://languages.oup.com/word-of-the-year/2016/

- anni 2019-2020: post-API.

Il campione estratto è composto da 1.877 post con i rispettivi 112.000 commenti.

L'analisi è stata condotta su due livelli, in ognuno dei quali è presente un momento ermeneutico e uno computazionale. Questa articolazione nasce dall'esigenza di ritrovare nella pratica di ricerca, delle soluzioni capaci di far luce su una realtà che è già di per sé multilivello: post e commento. Mentre l'approccio ermeneutico è stato utile a far emergere aspetti particolari del fenomeno indagato, l'approccio computazionale ha esteso l'analisi di quegli aspetti su larga scala.
Per questo scopo è stato messo a punto un percorso di metodo adeguatamente integrato e replicabile che risponde efficacemente alle domande di ricerca poste.

Figura 1 - Fasi di analisi[33]



Il primo livello di analisi focalizzerà l'attenzione sulla trasmissione del messaggio (stimolo) e si dividerà in due momenti:

- *Ermeneutico. Categorizzazione ed etichettatura manuale*

La scelta di etichettare manualmente i dati è dettata dalla consapevolezza della diversità di elementi presenti nel campione, sia dal punto di vista tipologico sia dal punto di vista contenutistico. La volontà è quella di preservare sfere di significato che altrimenti sarebbero andate perse.

- *Computazionale. Text mining e analisi fattoriale.*

Le tecniche di *Text Mining* sono state utilizzate per analizzare il messaggio disinformativo, per estrarre e analizzare gli *hashtag* e per identificare temi latenti nel corpus attraverso il *Topic Modeling*. Queste tecniche sono state integrate con analisi fattoriali e di classificazione volte a far emergere le dimensioni

[33] Se non differentemente indicato, i grafici contenuti in questa ricerca sono stati tutti creati su elaborazioni dati personalmente effettuate, la lingua inglese è stata preferita all'italiano per conservare i significati originali di categorie, elementi e tecniche così come sono riportate nella letteratura di riferimento.

latenti del fenomeno disinformativo e a delineare le strategie messe in atto raggruppando gli elementi utilizzati nella produzione di contenuti controversi.
Il secondo livello di analisi pone invece l'attenzione sul coinvolgimento degli utenti come risposta al messaggio stimolo. Questo livello è stato analizzato con diverse tecniche di *Opinion Mining* che hanno il compito di individuare, estrarre e classificare le opinioni con riferimento alle questioni emerse nel precedente livello di analisi. Anche questo livello si divide in due momenti, quello ermeneutico che caratterizza la costruzione dei *set* di allenamento per le tecniche di *Machine Learning* e un momento computazionale in cui l'algoritmo scelto opererà in piena autonomia.

Seguendo l'insegnamento di Enrica Amaturo (1989), si proverà di seguito ad illustrare ogni soluzione tecnica adottata non partendo dall'algoritmo ma dai problemi a cui risponde: comprendendo il funzionamento generale non sarà difficile poi seguire i passaggi matematici (molto ridotti) e i risultati prodotti.

3.1 Primo livello. Trasmissione del messaggio

Con il passaggio al digitale e lo sviluppo del *web* il nostro rapporto con i dati testuali e si è modificato, trasformando anche le pratiche e i modelli tradizionalmente utilizzati nelle analisi. Tutti i campi che prendono il testo come loro oggetto di studio hanno dovuto far fronte a questa nuova disponibilità di dati grazie alla quale gli approcci computazionali e statistici ricoprono (oggi più che mai) un ruolo rilevante nel garantire il collegamento tra elemento umano e testuale (Lebart et al., 2019). Tra i metodi disponibili per analizzare un insieme di dati testuali, la statistica esplorativa multidimensionale ha dato origine a numerose implementazioni e applicazioni in diversi ambiti di ricerca, soprattutto nell'area nota oggi in letteratura come *Text Mining* (Bolasco, 2005).
L'idea alla base del *Text Mining* è proprio quella di far leva sulla potenza del *computing* per minimizzare la gestione manuale di ampi *corpora*, trasformando il testo libero e non strutturato in dati strutturati e normalizzati con l'obiettivo di estrarre quante più informazioni possibili utilizzando sistemi automatizzati. Il campo dell'analisi statistica dei dati testuali è un campo metodologico molto attivo che ha trovato il giusto equilibrio grazie al dialogo interdisciplinare tra statistici, linguisti, informatici e ricercatori testuali delle scienze sociali e politiche.
Gli strumenti informatici sono stati a lungo utilizzati per elaborare grandi *corpora* testuali: così sono apparse le *Digital Humanities* negli anni '60 e l'uso della statistica per caratterizzare lo stile di autori classici della letteratura e attribuire loro opere anonime. Dagli anni '60 agli anni '90, molto prima che

il *Text Mining* diventasse popolare, la Francia ha assistito a un periodo eccezionalmente attivo nel campo dell'automazione dell'analisi del testo: la scuola francese di *Analyse Des Données* è stata un attore fondamentale in questo sviluppo grazie al lavoro di Jean-Paul Benzécri e colleghi, la cui influenza è ancora viva nella pratica del *Text Mining*, non a caso gli algoritmi e *software* a supporto di questa pratica portano avanti la stessa filosofia.

Riassumendo molto, l'ambizione teorica di Benzécri era quella di aprire le porte a una nuova linguistica in un'epoca dominata dalla linguistica generativa. Contrario alla tesi idealistica di Chomsky in quale, negli anni '60, riteneva che solo una modellazione astratta potesse rivelare le strutture linguistiche, Benzécri ha proposto un metodo induttivo di analisi dei dati linguistici. In questo senso, era abbastanza vicino agli obiettivi di Harris, che mirava a costruire le leggi della grammatica da un *corpus* di affermazioni con un approccio distributivo. I metodi sviluppati da Benzécri erano dal suo punto di vista più efficienti per una comprensione approfondita della lingua rispetto ai lavori di linguistica statistica realizzati da Guiraud o Muller, troppo incentrati esclusivamente sul vocabolario (Benzécri, 1981). Benzécri propone un metodo utile a rispondere ai problemi fondamentali che interessano i linguisti; questo consiste in un'astrazione quantitativa: partendo da tabelle di dati, attraverso il calcolo si costruiscono grandezze che possono misurare nuove entità situate a un livello di astrazione superiore a quello dei dati inizialmente raccolti (Benzécri, 1981). Il passaggio dai dati alle entità astratte è per lui possibile grazie al supporto del *computer*: «*I nuovi mezzi di calcolo ci permettono di affrontare descrizioni complesse di un gran numero di individui e quindi posizionarli su mappe piane o spaziali, in immagini affidabili e accessibili alle intuizioni dalla nebula dei dati iniziali*» (Benzécri, 1968, p. 21).

Il dispositivo informatico è però ausiliario per la sintesi: non fornisce né nomi né significati agli elementi estratti, spetta al ricercatore fornire loro delle interpretazioni. L'ambizione filosofica di Benzécri era di riassegnare valore all'approccio induttivo e quindi di opporsi all'idealismo che supponeva l'esistenza di un modello e del controllo della sua rilevanza attraverso l'osservazione; infatti, l'analisi delle corrispondenze è stata inizialmente proposta come un metodo induttivo per l'analisi dei dati linguistici (Benzécri, 1982). Infine, un elemento che contraddistingue il lavoro di Benzécri è l'organizzazione del suo lavoro in libri collettivi che propongono teoria, esempi di applicazioni provenienti da diversi campi del sapere (come scienze naturali e umane) e programmi da utilizzare nei *personal computer*. Questa organizzazione è un elemento che spiega l'importante diffusione dei suoi metodi, dovuta all'esplicitazione e condivisione delle procedure statistiche e di analisi (una sorta di *open source* esistito prima del tempo). Alla fine degli anni '80, le procedure di analisi sono state incluse nei principali pacchetti *software* statistici, in particolare SPSS, SPAD_T, e oggi implementati in R.

Nel corso del tempo, gli studi quantitativi sulla lingua hanno cambiato progressivamente il loro obiettivo, spostandolo da una logica di tipo linguistico (sviluppata fino agli anni ‘60) ad una di tipo lessicale (intorno agli anni ’70 del secolo scorso), per approdare negli anni Ottanta e Novanta ad analisi di tipo testuale o lessico-testuale (Bolasco, 2005). Recentemente si è provato che l’analisi dei dati testuali migliora con l’apporto di meta informazioni di carattere linguistico (dizionari elettronici, lessici di frequenza, grammatiche locali) e con interventi sul testo (normalizzazione, lemmatizzazione e lessicalizzazione) dunque attraverso un’analisi integrata statistico-linguistica di tipo lessico-testuale (ivi).

L'esame statistico dei testi si è evoluto dal conteggio delle parole allo studio delle associazioni tra le parole: l'informatica si è rivelata molto utile per calcolare le co-occorrenze aprendo così la strada alla ricerca di segmenti ripetuti (Lebart & Salem, 1988) piccole frasi o elementi di linguaggio che riempiono i discorsi politici, i messaggi dei mezzi di comunicazione di massa e di quelli pubblicitari. Con i progressi della visualizzazione grafica del computer, le mappe e le nuvole di parole hanno fornito un supporto visivo di queste concordanze.

I metodi di analisi fattoriale applicati ai testi sono stati utili a sintetizzare l'estrema rigidità della ricerca di segmenti ripetuti e l'eccessiva focalizzazione di mappe cognitive centrate su una sola parola. Questi metodi consentono di studiare sistematicamente le associazioni lessicali e quindi di identificare le affinità tra i termini frequentemente associati. Nell’ADT la disposizione delle parole in relazione tra loro sugli assi fattoriali fanno emergere universi lessicali che rivelano temi latenti del testo in relazioni ad altre variabili osservate. Si ottiene così una mappa cognitiva generalizzata che visualizza non solo le parole nel loro ambiente ma anche posizionando i diversi universi in relazione tra loro. Spetta all'analista farne una lettura semiotica collegando significanti (le parole), referenti (gli universi lessicali) e significati (le idee e le conoscenze del ricercatore) (Lebart et al., 2019).

È possibile evidenziare degli universi lessicali anche attraverso la partizione di unità di significato (risposte, frasi o sequenze di parole, ecc.*)* simili dal punto di vista delle parole che le compongono. Reinert (1983) seguendo Benzécri, ha proposto un metodo per creare questo tipo di partizione operando una classificazione gerarchica discendente da diverse analisi fattoriali utilizzate per definire progressivamente classi omogenee (Lebart et al., 2019). Questo approccio rientra nei metodi di classificazione tematica in cui il ricercatore attribuisce a ciascuna classe, un concetto che meglio riassume gli universi lessicali che la caratterizzano.

A partire dagli anni 2000 si diffonde la necessità di analizzare in maniera automatica il contenuto del *web* per sintetizzare l’informazione contenuta nei dati non strutturati provenienti dalla rete. A queste esigenze risponde il *Text Mining* (TM) e il *Natural Language Processing* (NLP), l’area del *Machine Learning* (ML) dedicata al senso della parola scritta che permette di estrarre automaticamente il

significato dei testi identificando temi o argomenti *(Topic Modeling)* e individuando e classificando le opinioni in essi contenuti *(Opinion Mining)*. Per *Machine Learning* (ML) s'intende l'abilità del computer di apprendere autonomamente grazie ad algoritmi che, in modo esperienziale, migliorano le proprie prestazioni man mano che gli si forniscono esempi da cui imparare.
Le opportunità di analisi sono dunque molteplici; la scelta del tipo di tecnica da utilizzare dipende molto dal *corpus* di cui si dispone. Questi metodi ben si adattano a corpora sufficientemente ampi la cui lettura rappresenta un vero ostacolo. La molteplicità di tecniche e approcci potrebbe però indurre nell'errore di moltiplicare le analisi perdendosi nella complessità dei dati e perdendo il quadro generale della ricerca.

Per una buona analisi automatica del testo, Bolasco (2005) propone di individuare diversi *step* e inserirli in una filiera immaginaria, evitando di "cristallizzare" le procedure possibili ma fissandone i passi fondamentali.
Seguendo la proposta di Bolasco, per l'analisi del messaggio disinformativo è stata immaginata una filiera ideale che prevede i seguenti step:

a) **Lexical pre-processing**: consiste nella codifica, pulizia e normalizzazione automatica del testo. Il testo proveniente dal *web* e dai *social media* necessita di esser sottoposto a ricodifica; se questa non viene eseguita o viene eseguita in maniera errata, ci si trova a lavorare ad un *corpus* dalle parole incomprensibili.
   Si riporta come esempio il *tweet* sotto:

Figura 2 - Tweet (fonte: Quanteda[34])

[34] Lezione con Ken Benoit, Essex Summer School 2020.

Che in Windows diventa:

“â ï¸â ï¸â ï¸RESISTER FOLLOW BACK PARTY â ï¸â ï¸â ï¸ Itâ€™s time for WE HOLD THESE TRUTHS WEDNESDAY! Our founders fought against tyranny. We fight against #Moronism. ðŸ‘‰LETâ€™S PARTY, FOLKS!ðŸ‘‰”.

Questo accade perché i sistemi precedenti utilizzavano *code page* a 8 *bit* che avevano solo 256 caratteri, oggi la maggior parte dei sistemi utilizza invece Unicode: un sistema di codifica che assegna un numero univoco ad ogni carattere usato per la scrittura di testi. Essendo che una pagina *html* utilizza più codifiche e non è possibile trasformare parti diverse di un documento con codifiche diverse, il testo importato in R è stato convertito in Unicode UTF-8, capace di supportare molte lingue e ospitare pagine e moduli in qualsiasi combinazione di tali lingue indipendentemente dalla piattaforma informatica e dal programma utilizzato. Codificato il testo, sono stati uniformati gli spazi, le maiuscole, sono stati eliminati segni di punteggiatura e le *stop words*. È stato stilato un elenco di parole tipiche del linguaggio dei *social media* e sono state eliminate quelle forme che fanno rimando alle *emoticon (*😊, 😕, 😅, *ecc.),* ad alcune onomatopee *(ohooo, naaaaah, woow, hahaha)* e agli acronimi *(https, html, http).*

b) **Information extraction**[35]: in questo *step* è stata effettuata un’analisi sugli *hashtag*, ed è stato portato avanti un lavoro sul vocabolario che si è concluso nella visualizzazione di una rete di *bigrams* dalle relazioni simmetriche in cui è stato possibile analizzare le relazioni tra le parole che compaiono insieme più frequentemente.

c) **Machine Learning**: l’obiettivo di questa fase è stato quello di individuare le dimensioni semantiche latenti del *corpus* in oggetto attraverso un *Latent Topic Modeling* (LTM). Nell'apprendimento automatico e nell'elaborazione del linguaggio naturale il LTM è un

[35] In questo studio le pratiche di *Text Mining* sono state eseguite attraverso due *library* di R, scelte perché implementate durante i corsi seguiti all’Università di Essex. La prima *library* è Quanteda: un pacchetto di analisi quantitativa del testo *open-source* messo a punto per le scienze sociali. Lo strumento è stato sviluppato da Kenneth Benoit e le sue capacità superano quelle fornite da altri pacchetti R (come tm) e persino del pacchetto Python Gensim; è completamente multilingue e funziona bene anche con testi in lingua cinese. È un pacchetto autonomo ma può essere utilizzato anche come *framework* per altre *library*. Sebbene l'utilizzo di Quanteda richieda conoscenze di programmazione, la sua API è progettata per consentire un'analisi potente ed efficiente con un minimo di passaggi, abbassando così le barriere all'apprendimento automatico, dell'utilizzo del NLP e dell'analisi quantitativa del testo. La seconda *library* è Tidyverse una raccolta di pacchetti R *open source* per la gestione dei dati introdotta da Hadley Wickham e dal suo gruppo che condividono una filosofia di progettazione, una grammatica e una struttura di dati ordinata. Il *tidyerse* per la sua semplicità di programmazione è diventato il quinto pacchetto più scaricato di R e il linguaggio di programmazione più utilizzato.

modello statistico che permette di identificare una serie di temi latenti all'interno di una collezione di testi.

3.2 Latent Dirichlet Allocation Topic Model

Chi utilizza il *Text Mining* si trova spesso a lavorare con ampie raccolte di documenti, come post, *blog* o fascicoli di notizie che necessitano di essere divisi in gruppi tematici per poterli analizzare separatamente e compararli. La modellazione degli argomenti è un metodo per la classificazione non supervisionata di documenti che estrae gruppi di elementi dai corpora anche quando l'analista non sa cosa aspettarsi.

La Latent Dirichlet Allocation (da ora LDA) è una tecnica di modellazione di argomenti molto utilizzata per estrarre temi da ampi corpora.

Il termine latente ci fa comprendere subito l'esistenza di qualcosa che non è immediatamente visibile. Senza immergerci nella matematica alla base del modello, possiamo immaginare il processo come guidato da due principi chiave: ogni documento è un insieme di argomenti, ogni argomento è un insieme di parole (Blei et al., 2003). Nello specifico, i documenti sono considerati come la densità di probabilità (o distribuzione) degli argomenti e gli argomenti sono la densità di probabilità (o distribuzione) delle parole. LDA stima entrambi contemporaneamente: trova la combinazione di parole associata a ciascun argomento determinando anche la combinazione di argomenti che descrive ciascun documento.

È importante sottolineare che, piuttosto che essere separati in gruppi discreti, nei modelli tematici probabilistici, le parole possono essere condivise tra gli argomenti ai quali appartengono con diversa probabilità, ciò consente ai documenti di "sovrapporsi" l'un l'altro in termini di contenuto in modo da rispecchiare l'uso tipico del linguaggio naturale.

Riprendendo i lavori di David Blei è possibile definire l'allocazione latente di Dirichlet come un modello probabilistico gerarchico utilizzato per scomporre una raccolta di documenti nei suoi argomenti salienti, dove un "argomento" per LDA è una distribuzione di probabilità su un vocabolario (Blei et al., 2010).

LDA e i suoi parenti sono chiamati modelli topici probabilistici.

Come funziona il processo?

In primo luogo, LDA applica i due principi citati al *corpus* che si intende analizzare.

Supponiamo di avere cinque documenti che compongono un *corpus*, questo *corpus* sarà rappresentato in una matrice di documenti per termini (DTM).

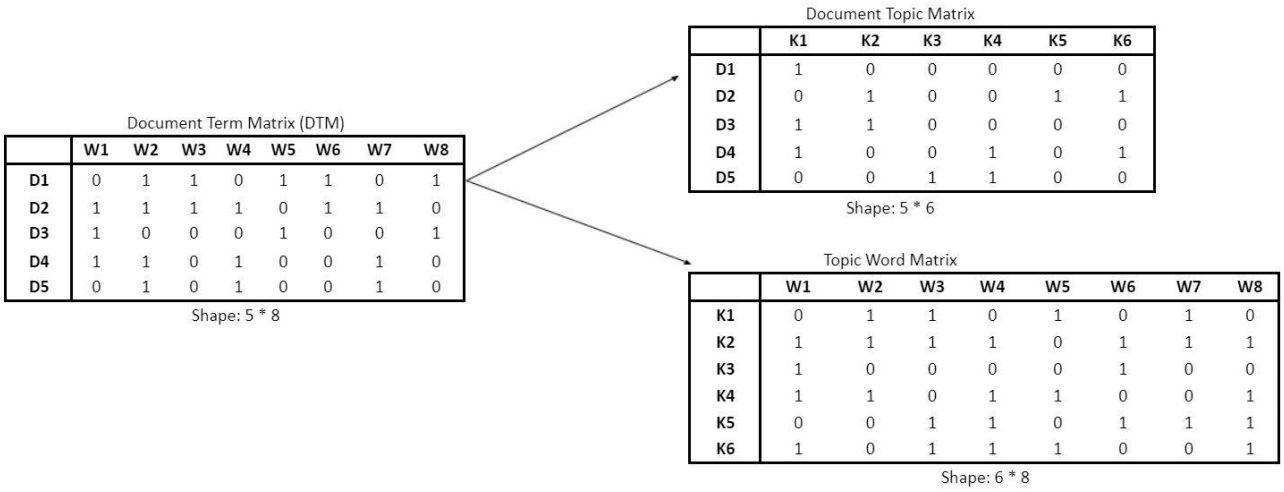
Figura 3 - Esempio di matrice documenti per termini

Document Word Matrix

	W1	W2	W3	W4	W5	W6	W7	W8
D1	0	1	1	0	1	1	0	1
D2	1	1	1	1	0	1	1	0
D3	1	0	0	0	1	0	0	1
D4	1	1	0	1	0	0	1	0
D5	0	1	0	1	0	0	1	0

LDA converte questa matrice in altre due matrici: la matrice "documento per *topic*" e quella "*topic* per parola".

Figura 4 - Esempio di conversione matrici in un modello LDA[36].



	W1	W2	W3	W4	W5	W6	W7	W8
D1	0	1	1	0	1	1	0	1
D2	1	1	1	1	0	1	1	0
D3	1	0	0	0	1	0	0	1
D4	1	1	0	1	0	0	1	0
D5	0	1	0	1	0	0	1	0

	K1	K2	K3	K4	K5	K6
D1	1	0	0	0	0	0
D2	0	1	0	0	1	1
D3	1	1	0	0	0	0
D4	1	0	0	1	0	1
D5	0	0	1	1	0	0

	W1	W2	W3	W4	W5	W6	W7	W8
K1	0	1	1	0	1	0	1	0
K2	1	1	1	1	0	1	1	1
K3	1	0	0	0	0	1	0	0
K4	1	1	0	1	1	0	0	1
K5	0	0	1	1	0	1	1	1
K6	1	0	1	1	1	0	0	1

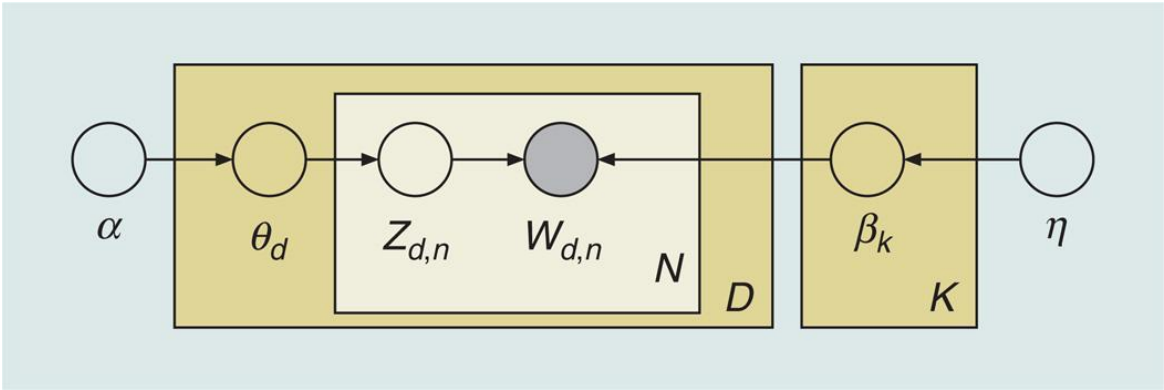
La matrice "documento per *topic*" contiene già i possibili argomenti (rappresentati da K) che i documenti possono contenere, mentre la matrice "*topic* per parola" contiene i termini che tali argomenti possono contenere. La scelta di un numero fisso di argomenti è fatta a priori e presuppone che ogni documento rifletta la combinazione perfetta di tali argomenti. Quando una raccolta di

[36] Immagini raccolte dal sito *analyticsvidhya.com*

documenti viene analizzata in base a questi presupposti, gli algoritmi di inferenza probabilistica rivelano una struttura tematica incorporata (visibile nella tabella 2) (Blei et. al., 2010). Con questa struttura, LDA fornisce un modo per riassumere, esplorare e cercare rapidamente raccolte di documenti di grandi dimensioni.

Secondo Blei (2010) le ipotesi probabilistiche generative di LDA presuppongono che una raccolta di documenti sia graficamente espressa come segue[37].

Figura 5 - Rappresentazione grafica di LDA (di Blei et al. 2010)



Come descritto nel lavoro di Blei (2010) il modello grafico nella figura 5 rivela la struttura nidificata a più livelli delle ipotesi LDA, composto a sua volta da una gerarchia di modelli misti. Ogni documento si configura attraverso un modello a miscela finita in cui le proporzioni della miscela (cioè le proporzioni degli argomenti) sono disegnate in modo univoco per ogni documento ma i componenti della miscela (cioè gli argomenti) sono condivisi nella raccolta. Nelle statistiche, questo è noto come modello di appartenenza mista (ivi).

Il riquadro di color ocra più grande si riferisce a tutti i documenti del *corpus* (rappresentato da *D*), mentre la casella gialla si riferisce al numero di parole in un documento (rappresentato da N). Dentro questo riquadro giallo ci sono diverse parole, una di queste parole è *W*, nel cerchio di colore grigio. Secondo LDA, ogni parola è associata a un argomento latente che è rappresentato da Z. L'assegnazione di Z a una parola presente in questi documenti fornisce una distribuzione di parole argomento presente nel *corpus* che è rappresentato da *theta* (θ). Il modello LDA ha due parametri che

[37] *D* rappresenta il totale dei documenti nel corpus; *N* rappresenta il numero di parole nel documento; *W* rappresenta la parola oggetto in un documento; *Z* rappresenta l'argomento latente assegnato a una parola; *Theta* (θ) rappresenta la distribuzione degli argomenti; *Alpha (a)* e *Beta (β)* sono i parametri di controllo del modello.

controllano le distribuzioni: *alpha* (α) controlla la distribuzione dei *topic* per documento e *beta* (β) controlla la distribuzione delle parole per *topic*. Un argomento β è una distribuzione su un vocabolario fisso di V termini (Blei et al., 2010).

L'obiettivo finale di LDA è trovare la rappresentazione più ottimale delle parole per ogni *topic*.
LDA è un processo iterativo: nella prima iterazione, assegna casualmente i *topic* a ciascuna parola nel documento poi torna indietro al livello del documento per identificare quali *topic* avrebbero generato i documenti e quali parole avrebbero generato i *topic*. Per ottimizzare i risultati ottenuti, LDA itera su tutti i documenti e su tutte le parole. A questo punto LDA presuppone che tutti i *topic* assegnati sono corretti tranne la parola osservata così, delle assegnazioni parole-*topic* già corrette, il processo regola l'attribuzione di argomenti della parola osservata con una nuova assegnazione.
LDA eseguirà un'iterazione su ogni documento "*D*" e per ogni parola "*W*" calcolando due probabilità:

- Probabilità_1: proporzione di parole nel documento (*D*) già assegnate al *topic* (*K*).
- Probabilità_2: proporzione di assegnazioni al *topic* (*K*) su tutti i documenti che derivano dalla parola *W*.

Su queste probabilità LDA stima una nuova probabilità data dal prodotto di p1 e p2 e attraverso questa probabilità identifica il nuovo argomento *K* che è quello più rilevante per la parola *W*.
Per la scelta del nuovo argomento *K* vengono eseguite un gran numero di iterazioni fino a quando non si ottiene uno stato stazionario.

Tutto questo può essere riassunto con i passaggi matematici che seguono (Blei et al., 2010):
Disegna $\boldsymbol{K}$ argomenti da una distribuzione simmetrica di Dirichlet:

$$\beta_{k} \sim \text{Dir}_{V}(\eta), \boldsymbol{k} \in \{1, \ldots, \boldsymbol{K}\}.$$

Per ogni documento $\boldsymbol{d}$, traccia le proporzioni dell'argomento da un Dirichlet simmetrico:

$$\theta_{d} \sim \text{Dir}_{K}(\alpha), \boldsymbol{d} \in \{1, \ldots, \boldsymbol{D}\}.$$

Per ogni parola $\boldsymbol{n}$ in ogni documento $\boldsymbol{d}$, disegna un'assegnazione di argomento dalle proporzioni dell'argomento:

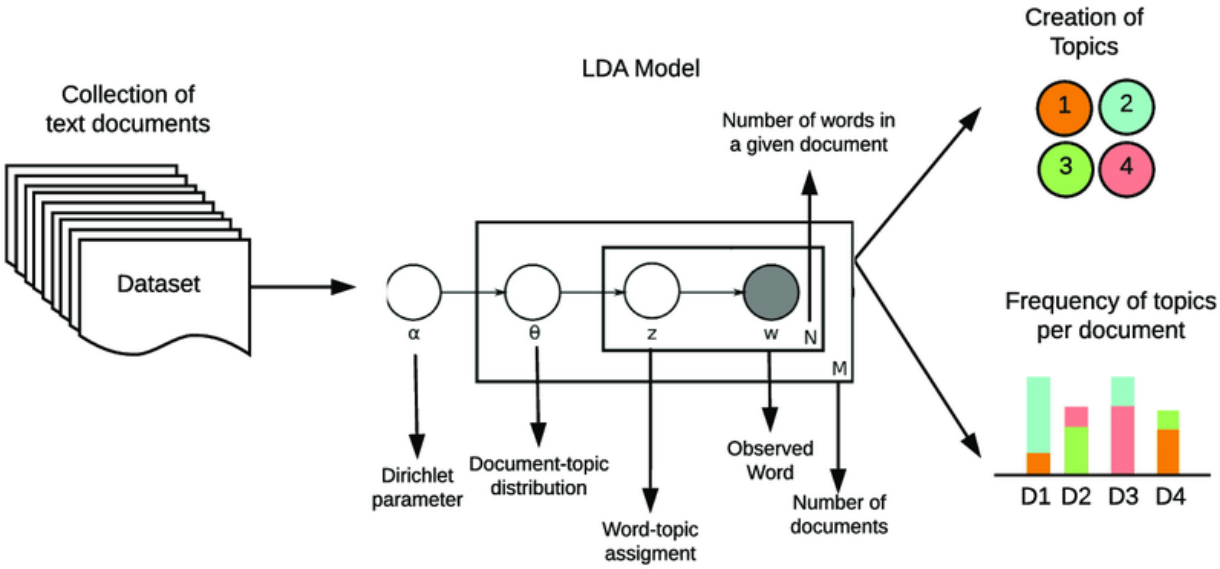
$$\boldsymbol{z}_{d,n} \mid \theta_{d} \sim \text{Mult}(\theta_{d}).$$

Disegna la parola dall'argomento corrispondente

$$\boldsymbol{w}_{d,n} \mid \boldsymbol{z}_{d,n}, {}_{1:K} \sim \text{Mult}(\beta_{z\,d,n}).$$

Per concludere, l'allocazione latente di Dirichlet svolge due compiti: trova gli argomenti del corpus e, allo stesso tempo, assegna questi argomenti al documento presente all'interno dello stesso corpus.
Il diagramma seguente può essere utile per chiarire bene tutti i passaggi sopra descritti.

Figura 6 - Schema dell'algoritmo LDA (di Buenaño-Fernandez et. al, 2020)



Recentemente, l'applicazione di modelli tematici è stata implementata anche a dati non testuali in interessanti ricerche sull'analisi delle immagini[38].

[38] Per un approfondimento: Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic Topic Models: A focus on graphical model design and applications to document and image analysis. *IEEE signal processing magazine, 27(6), 55–65.*

3.3 Analisi fattoriale e di classificazione

L'analisi delle corrispondenze multiple (da ora ACM) è una branca dell'analisi fattoriale, iniziata con Fisher nel 1940; è un tipo di analisi utile ad analizzare grandi matrici di dati in cui vi siano in prevalenza variabili categoriali. L'obiettivo principale dell'ACM consiste nell'analisi delle relazioni esistenti tra un insieme di variabili attraverso l'identificazione di uno spazio ottimale di dimensione ridotta capace di sintetizzare l'informazione contenuta nei dati originari attraverso l'estrazione di variabili latenti (o fattori) attraverso i quali descrivere fenomeni o concetti non direttamente osservabili nella realtà ma dati dalle relazioni intercorrenti tra essi e le modalità delle variabili analizzate (Gherghi e Lauro, 2004). La metodologia statistica ha registrato un forte ritardo nei confronti delle problematiche derivanti dall'osservazione di fenomeni relativi al campo delle scienze sociali: l'analisi multidimensionale di variabili categoriali ha avuto un'evoluzione più lenta di quanto non sia avvenuto parallelamente nel campo delle variabili numeriche perché per molti anni l'interesse verso le variabili qualitative si è risolto tentando di adattare metodi nativi per variabili continue, con esiti che risultavano già in partenza fortemente penalizzati (Gherghi e Lauro, 2004). Non c'è da stupirsi se proprio da studiosi del campo delle scienze sociali (Goodman, Haberman, Fisher) sono arrivati i primi contributi indirizzati all'individuazione di tecniche multivariate per l'analisi simultanea di più variabili. Oggi l'approccio multivariato può essere impostato seguendo due orientamenti: quelli di tipo confermativo, che ha caratterizzato a partire dagli anni '70 la scuola anglosassone e quello di tipo descrittivo, proposto dalla scuola francese di Benzécri che formalizza il metodo mettendo a punto una serie di strumenti analitici in grado di gestire vasti insiemi di dati offrendo un'impostazione volta ad evidenziare le proprietà algebriche e geometriche e fornendo rappresentazioni grafiche di tipo dimensionale capaci di sintetizzare al meglio la molteplicità di relazioni esistenti tra oggetti, proprietà e le loro modalità (Amaturo, 1989). In questo approccio - battezzato prima Analisi delle Corrispondenze che è poi diventato il filone francese di *Analyse des Données* - non si ipotizza nessun modello o probabilità per i dati osservati[39], ma l'associazione tra variabili viene studiata approssimando la matrice in una di rango ridotto contenente le stesse informazioni e capace di definire le relazioni attraverso la definizione di operatori di proiezione associati ai piani fattoriali (ivi). L'elemento di originalità introdotto dalla scuola francese, a cui si deve la notorietà, non risiede soltanto nella possibilità di applicare tecniche fattoriali anche a variabili categoriali ma soprattutto all'attenzione per gli aspetti grafici dei risultati che ne migliorano e facilitano l'interpretazione. Sui fattori emersi dall'ACM sono stati proiettati dei

[39] L'approccio della scuola anglosassone faceva proprio l'utilizzo di modelli log-lineari introdotti in quegli anni da Goodman.

cluster con lo scopo di individuare sia le caratteristiche delle diverse strategie disinformative (nel primo livello di analisi) sia gli elementi che differenziano gli utenti che interagiscono con quelle strategie (nel secondo livello di analisi). L'associazione dell'analisi della corrispondenza con tecniche di *clustering* consente una comprensione più profonda dei dati e un'interpretazione più chiara.
La *Cluster Analysis* comprende un insieme di tecniche di analisi multivariata dei dati volte alla selezione e raggruppamento di elementi omogenei in un insieme di dati. Le tecniche di *clustering* si basano su misure relative alla somiglianza tra due oggetti allo scopo di massimizzarla nel caso in cui gli oggetti appartengano allo stesso gruppo e minimizzarla nel caso in cui gli oggetti appartengano a gruppi diversi. La tecnica di *clustering* adottata in questo studio è di tipo gerarchico *top-down* con un tipo di clusterizzazione non esclusiva, in cui un elemento può appartenere a più *cluster* con gradi di appartenenza diversi.

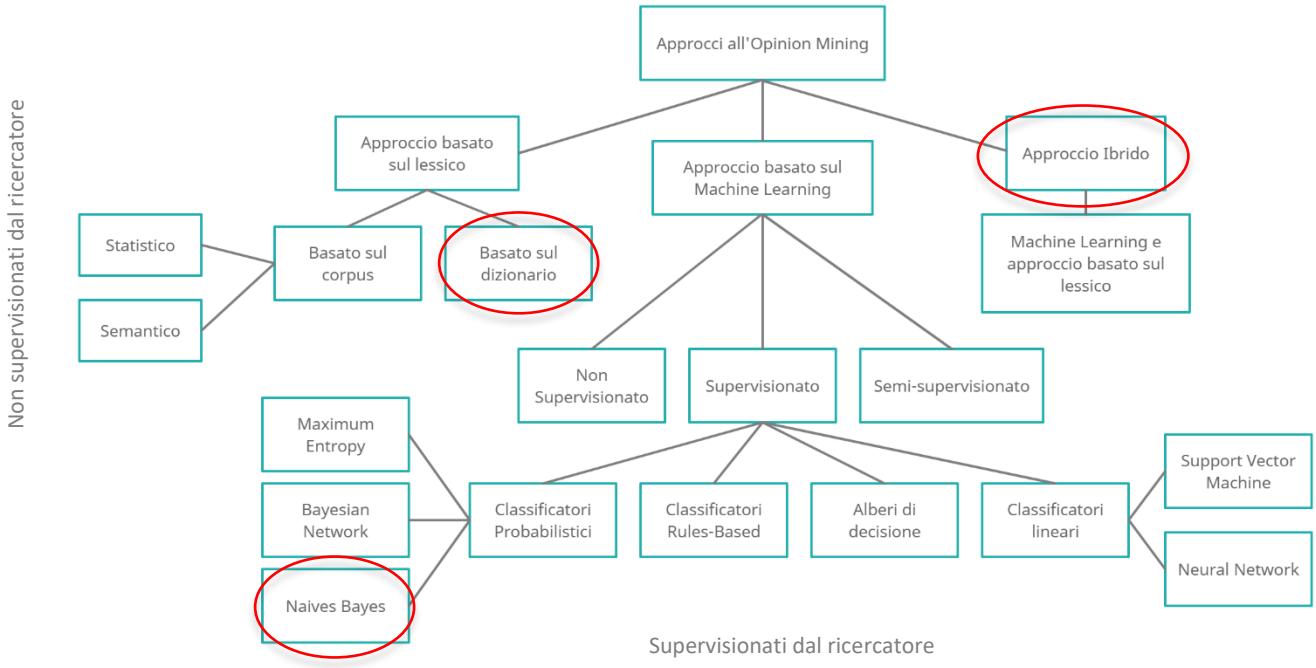
3.4 Secondo livello. *Opinion Mining*

Quando un individuo legge un testo usa la propria comprensione dell'intento emotivo delle parole per dedurre se la comunicazione è positiva, negativa o caratterizzata da qualche tipo di sfumatura emozionale. È possibile utilizzare diverse tecniche di *Opinion Mining* per avvicinarci a rilevare il contenuto emotivo del testo in modo più o meno automatico. Come già detto, il campo dell'apprendimento automatico che fornisce ad un *computer* la capacità di comprendere, analizzare, manipolare il linguaggio umano è il *Natural Language Processing* del quale un ramo molto proficuo è l'*Opinion Mining*. Questo fornisce strumenti adatti ad analizzare l'opinione pubblica rilevando automaticamente le informazioni contenute in un *corpus* testuale e determinandone la polarità nei confronti di un tema, un prodotto, un personaggio pubblico.
L'*Opinion Mining* si suddivide in due macro-approcci: approccio basato sul lessico *(lexicon-based approach)* e approccio basato sull'apprendimento automatico *(machine learning approach)*. Mentre il primo approccio non prevede la supervisione del ricercatore, il secondo approccio si differenzia sulla base dell'intervento o meno del ricercatore nell'implementazione dell'algoritmo. Mentre gli approcci basati sul lessico sfruttano appositi dizionari per la classificazione di frasi o documenti, gli approcci basati sull'apprendimento automatico utilizzano tecniche di intelligenza artificiale che, a partire da un *training set*, permettono di generalizzare la classificazione a qualsiasi altro contenuto testuale. Ognuno di questi approcci possiede una serie di tecniche che vengono scelte dal ricercatore sulla base alla quantità di dati e al tipo di analisi che si intende effettuare. Nella figura 7 sono stati riassunti gli

approcci e gli algoritmi più utilizzati nella classificazione nell'*opinion mining* e nella *sentiment analysis*.

Si passerà brevemente alla rassegna di ciò che differenzia di diversi approcci per poi approfondire il funzionamento di quelli scelti e implementati in questo studio.

Figura 7 - Differenti approcci all'analisi delle opinioni. Schema dell'autore



- *Approccio basato sul machine learning.*

Gli approcci basati sul *machine learning* si affidano ad algoritmi di intelligenza artificiale che, dato un *set* di dati etichettati (*training*), hanno l'obiettivo di prevedere l'attributo classe su un *set* di dati non etichettati. Durante la fase di addestramento, gli algoritmi imparano attraverso un particolare *set* di dati *input* grazie al quale saranno autonomamente in grado di classificare oggetti che non hanno mai visto prima. L'apprendimento può essere supervisionato *(supervised learning)* in cui gli algoritmi acquisiscono conoscenza per la classificazione a partire da un *training set* manualmente pre-etichettato dal ricercatore, o non supervisionato *(unsupervised learning)* in cui gli algoritmi acquisiscono esperienza per la classificazione dall'estrazione di caratteristiche comuni da un *set* di testi non etichettati.

Esistono diversi tipi di classificatori che vantano una matematica via via sempre più complessa: i classificatori probabilistici come il Naive Bayes e il Maximum Entropy; i classificatori basati su alberi

di decisione e quelli basati su regole; i classificatori lineari come il Support Vector Machine o le Reti Neurali.

Il problema con molte di queste tecniche non è solo la complessità concettuale e tecnica (molte *library* disponibili nascondono questa complessità dietro strati di astrazione) ma anche il loro costo computazionale, che cresce in maniera direttamente proporzionale alla crescita delle *feature* da analizzare. Il costo computazionale richiesto per alcune operazioni di *machine learning* e *data mining* è stato effettivamente un ostacolo incontrato nel secondo livello di analisi, ma si approfondirà la questione nei capitoli successivi.

Per concludere, l'approccio *machine learning* all'*opinion mining* ha il vantaggio di non dipendere dalla disponibilità dei dizionari, ma l'accuratezza dei metodi di classificazione dipende fortemente da una corretta etichettatura dei testi utilizzati per il *training* e da un'attenta selezione delle *feature* considerate dall'algoritmo.

- *Approccio basato sul lessico, non supervisionato***.**

L'approccio basato sul lessico rientra tra le tecniche di classificazione non supervisionate. La classificazione avviene calcolando l'orientamento semantico di frasi e documenti a partire dalle parole. I termini e le espressioni verbali che trasmettono un'opinione ben definita sono raccolti ed etichettati in un lessico di opinioni (*opinion lexicon*) necessario alla classificazione senza nessun intervento umano.

Questo tipo di classificazione può essere basata su *corpora* (*corpus-based approach*) o basata su dizionari *(dictionary-based approach)*. In entrambi i casi, parole semanticamente rilevanti vengono individuate all'interno della frase o dei documenti da classificare, ma con delle differenze:

- Nel primo approccio *(corpus-based)* la classificazione avviene attraverso la ricerca di parole che vengono confrontate con un corpus molto ampio già classificato. Sfruttando la correlazione e la co-occorenza fra parole, l'approccio assegna una determinata polarità (es. positiva, negativa o neutra) a parole che con quella polarità hanno un buon grado di associazione. La classificazione avviene confrontando la co-occorrenza tra parole fino a quel momento sconosciute e un insieme di parole già selezionate perché significative nel rilevare un particolare tipo di polarizzazione. Questo approccio fa uso di appositi metodi statistici e semantici: i primi eseguono ricerche per trovare occorrenze di termini che esprimono opinioni, determinandone la polarità a posteriori tramite co-occorrenze con altri termini presenti in un insieme di documenti opportunamente classificati; i secondi assegnano valori di polarità alle parole affidandosi all'idea che parole vicine possano avere la stessa polarità per principio semantico.

- Nel secondo approccio *(dictionary-based)* la classificazione avviene confrontando le parole dei documenti con quelle inserite nei dizionari esistenti, etichettandole sulla base di un punteggio di orientamento. Il punteggio complessivo di un documento dipenderà dal confronto tra la frequenza delle parole emotive presenti con le parole già pesate del dizionario di riferimento[40].

Mentre il problema principale dell'approccio basato sul corpus è la difficoltà di costruzione di *corpora* abbastanza grandi da coprire tutte le parole e le loro combinazioni, il problema principale dell'approccio basato sul dizionario è che fatica a distinguere l'eventuale differenza d'opinione di una parola quando è inserita in un contesto di contraddizione (es. "non buono"). Generalmente, per qualsiasi approccio si intende adottare *l'output* richiede sempre un'attenta convalida, soprattutto per i metodi di classificazione supervisionati nei quali bisogna assicurarsi che la classificazione dell'algoritmo replichi la codifica del *training*.

Nel caso di questa ricerca è stato utilizzato un approccio *machine learning* supervisionato basato sul classificatore probabilistico Naives Bayes; un approccio basato sul lessico che fa uso del dizionario LIWIC (*Linguistic Inquiry and Word Count*) capace di etichettare i testi sulla base delle parole associate a categorie psicologicamente ed emotivamente significative e infine è stata adottata una terza strada, messa a punto per far fronte all'incapacità del classificatore scelto di etichettare le classi con riferimento a particolari categorie.

3.5 Il dizionario LIWC

L'analisi del linguaggio naturale apre una finestra sull'esplorazione di pensieri, sentimenti e personalità degli oratori. Il dizionario LIWC, *Linguistic Inquiry and Word Count*, (pronunciato "*luke*") è una risorsa lessicale sviluppata dallo psicologo sociale James Pennebaker e dal suo gruppo di ricerca presso l'Università del Texas (Pennebaker et al., 2001). Le sue informazioni lessicali sono memorizzate in un dizionario costituito da un numero di categorie e da un numero di parole o termini assegnati a una o più di queste categorie. Le categorie - che hanno un significato psicologico, emozionale o legato a processi cognitivi e preoccupazioni per la vita - sono organizzate in gerarchie: ad esempio la categoria dei pronomi, contiene la categoria dei pronomi personali, che a sua volta contiene la categoria dei pronomi personali per la prima persona singolare e così via.

[40] Ogni dizionario utilizza una diversa scala di valutazione del *sentiment* delle parole.

Questo dizionario può essere utilizzato in ambienti di programmazione o attraverso un *tool* che dispone di un'interfaccia intuitiva attraverso la quale è possibile elaborare una raccolta di testi e ricavare le frequenze relative delle parole appartenenti alle categorie scelte. La distribuzione di tali categorie nel *corpus* fornisce indicazioni sullo stato psicologico del suo autore o può riflettere la sua condizione personale. Il dizionario LIWC è stato pubblicato in più versioni (in particolare Pennebaker et al., 2001, 2007, 2015) e tradotto in molte lingue. I contenuti e il numero di categorie sono aumentati nel corso degli anni: dalle circa 2.319 parole della versione del 2001 alle circa 6.549 nella versione del 2015, migliorando così gli *output* del programma. LIWC è stato progettato per funzionare con più dizionari consentendo ai ricercatori un ampio grado di libertà; c'è infatti la possibilità di inserire propri dati nella lingua d'interesse. La funzione di LIWC conta le occorrenze delle parole nei testi in base alle parole contenute nel suo dizionario ma essendo basata sul principio BoW (modello della borsa di parole) non tiene conto del contesto né esegue un processo di disambiguazione del senso delle parole. Per gli *standard* della linguistica computazionale, il programma è molto semplice e per questo motivo è diventato in poco tempo uno strumento di ricerca ampiamente utilizzato (vedi Tausczik e Pennebaker, 2010, per esempi) anche al di fuori del suo campo originale che è quello della psicologia sociale.

In questa ricerca, per indagare le dimensioni psicologiche sottese al *corpus* degli utenti sono state selezionate cinque categorie considerate rilevanti dai risultati ottenuti dalla prima fase di analisi, dalla fase ermeneutica di costruzione del *set* di *training* e dalla letteratura di riferimento sul fenomeno disinformativo. Queste categorie sono: ottimismo, ansia, rabbia, tristezza e certezza. Il punteggio assegnato da LIWC ad ogni dimensione selezionata è stato categorizzato *post hoc* in: "alto", "medio", "basso", "assente".

Infine, è stata sfruttata la categoria "*bad words"* per analizzare il grado di formalità del linguaggio in base all'utilizzo delle parolacce.

3.6 L’ingenuo classificatore di Bayes

Il classificatore Naive Bayes è storicamente uno degli algoritmi più utilizzati nella classificazione automatica del testo: dalla *spam detection*, alla *sentiment analysis*, alla classificazione per argomento in più categorie. L'ingenuo classificatore di Bayes sta recentemente vivendo una rinascita come fulcro della ricerca sull'apprendimento automatico soprattutto nel campo del linguaggio naturale, ma è stato a lungo indispensabile nell'*information retrieval* con il quale condivide non solo un ampio riconoscimento in letteratura, risalente agli anni ’60 (Lewis, 1998; McCallum & Nigam, 1998), ma anche il principio stesso di ordinamento delle informazioni; questo è basato sul potenzialmente autonomo processo classificatorio che mette in atto una forma di osservazione continua e adattativa dalla quale derivano i successivi processi decisionali (Rieder, 2020). Sebbene negli ultimi decenni siano state sviluppate diverse, nuove e più complesse tecniche, il classificatore di Bayes è ancora uno dei «*più efficienti ed efficaci algoritmi di apprendimento automatico e data mining*» (Zhang, 2004, p. 1) e per questo è rimasto popolare nel corso degli anni.

La classificazione Bayesiana è una tecnica statistica con la quale si determina la probabilità di un elemento di appartenere a una certa classe dati alcuni attributi dell’elemento stesso. Nel caso di questa ricerca, gli elementi sono i commenti, le classi sono il sentimento e l’orientamento politico espresso dagli utenti e gli attributi sono le parole; ragion per cui se una parola compare molto spesso in un documento assegnato a una certa categoria ma raramente ad altri diventa un forte indizio di quella categoria. Il classificatore si basa sul Teorema di Bayes (dal nome di Thomas Bayes 1702-1761) che, in statistica e teoria delle probabilità, calcola la probabilità di un evento sulla base alla conoscenza preliminare delle condizioni che potrebbero essere correlate all’evento stesso. Si tratta dunque di comprendere la probabilità condizionata. Il teorema consente di aggiornare un’ipotesi ogni volta che viene introdotta una nuova evidenza, fornendo così un mezzo per ragionare in uno spazio di incertezza in cui abbiamo alcune conoscenze preliminari che possono essere utilizzate per valutare un caso particolare.

L’equazione che rappresenta il teorema è la seguente:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Dove:

P = simbolo per indicare la probabilità.

P(A|B) = La probabilità P dell'evento A (ipotesi) di verificatasi dato che B (evidenza) si è verificata.

Essa viene definita anche come probabilità posteriore.

P(B|A) = La probabilità che l'evento B (evidenza) si verifichi dato che A (ipotesi) si è verificata.

P(B) = La probabilità che l'evento B (ipotesi) si verifichi.

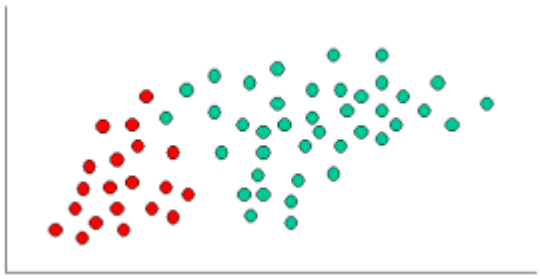
P(A) = La probabilità che l'evento A (evidenza) si verifichi.

La peculiarità principale del classificatore è quella di trattare le caratteristiche come statisticamente indipendenti l'uno dall'altro e per questo motivo è chiamato *Naive*, ossia ingenuo. Molti studiosi considerano questa un'assunzione irrealistica (Rish, 2001), tuttavia le performance di classificazione bayesiana sono sorprendentemente accurate (Zhang, 2004) anche confrontandolo con tecniche molto più sofisticate nell'applicazione a grandi *set* di dati (Domingos & Pazzani, 1997).

Per comprendere meglio il classificatore può essere utile visualizzarne il funzionamento con un esempio tipicamente proposto nei corsi di *machine learning*[41].

Nella figura sotto sono presenti 60 oggetti, di cui 20 rossi e 40 verdi. L'obiettivo è decidere a quale etichetta di classe dovranno appartenere i nuovi oggetti in arrivo tenendo in considerazione gli oggetti già presenti.

Figura 8 - Schema di oggetti rossi e verdi[42]



Poiché ci sono il doppio di oggetti verdi è ragionevole pensare che un nuovo caso (che non è stato ancora osservato) abbia il doppio delle probabilità di avere l'appartenenza "verde" piuttosto che "rossa". Nell'analisi bayesiana questa convinzione è nota come probabilità precedente, ossia basata sull'esperienza precedente (in questo caso di oggetti verdi e rossi) che è usata per prevedere i risultati

[41] I corsi a cui si fa riferimento sono quelli seguiti durante l'Essex Summer School in Social Science Data Analysis dell'Università di Essex, anno 2020.

[42] Fonte: towardsdatascience.com

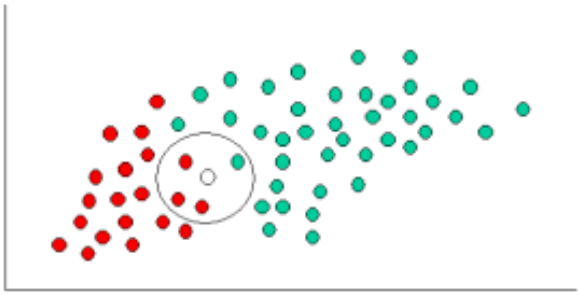
prima che accadano realmente. In questo caso, dunque, le probabilità precedenti per l'appartenenza alla classe sono:

a) Probabilità preliminare per verde: 40 / 60

b) Probabilità preliminare per rosso: 20 / 60

Formulata la probabilità precedente, è ora necessario classificare un nuovo oggetto X in arrivo (bianco nell'immagine che segue).

Figura 9 - Schema oggetti



È ragionevole ipotizzare che la prossimità degli oggetti verdi e rossi a X determini la probabilità di X di appartenere a una classe piuttosto che ad un'altra. Per misurare questa probabilità disegniamo un cerchio attorno a X che comprende un numero di punti indipendentemente dalle loro etichette di classe. A questo punto, sulla base del numero di punti nel cerchio appartenenti a ciascuna etichetta di classe, calcoliamo la verosimiglianza (in inglese solitamente si utilizza la parola *likelihood*):

- verosimiglianza di X per verde = numero di verdi in prossimità X / il totale dei casi di verde.
- verosimiglianza di X per rosso = numero di rossi in prossimità di X / il totale dei casi rossi.

Considerando il raggruppamento della figura sopra è chiaro che la verosimiglianza di X su verde è minore della verosimiglianza di X su rosso, poiché il cerchio comprende un oggetto verde su 40 e 3 oggetti rossi su 20. Sebbene le probabilità precedenti indichino che X possa appartenere al verde perché ci sono il doppio dei casi rispetto al rosso, la verosimiglianza indica al contrario che l'appartenenza alla classe di X è rosso, dato che ci sono più oggetti rossi nelle vicinanze di X. Nell'analisi bayesiana la decisione finale viene prodotta combinando entrambe le informazioni (probabilità precedente e verosimiglianza) per formare la probabilità posteriore sulla base del teorema di Bayes.

- Probabilità posteriore di X di appartenere a verde = probabilità precedente*la verosimiglianza di X per verde. Ossia: $= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60}$

- Probabilità posteriore di X di appartenere a rosso = probabilità precedente*la verosimiglianza di X per rosso. Ossia: $= \frac{2}{6} \times \frac{3}{20} = \frac{1}{20}$

Dunque, X sarà classificato come rosso poiché la sua appartenenza alla classe raggiunge la maggiore probabilità posteriore.

Tornando al caso di questa ricerca, sono ancora molte le domande aperte sull'applicazione dell'ingenuo Bayes ai dati testuali come ad esempio: quanto è rilevante la lunghezza del documento? Qual è la strategia ottimale per selezionare i dati di allenamento? Ricordiamo che l'efficacia della classificazione dipende dal *set* con cui alleniamo il classificatore.

Non esistendo delle regole specifiche per la costruzione di un *training set* per il classificatore bayesiano, la soluzione ritenuta metodologicamente più coerente in questa ricerca è stata quella di estrarre dal campione una percentuale tale di commenti da assicurare al classificatore un *set* di *training* che contenesse tutte le caratteristiche del campione distribuite in maniera rappresentativa per i diversi scenari possibili.

Il campione estratto dalla popolazione raggiunta è composto da 1.878 post e da altrettanti 112.089 commenti legati ai post di appartenenza attraverso una variabile di ancoraggio fornita dall'API, ossia un codice identificativo composto da tre serie numeriche: la prima identificativa della pagina Facebook, la seconda identificativa del post e la terza identifica del commento stesso, rispettando in questo modo l'anonimato di chi lo ha prodotto. Tenendo come guida il codice identificativo è stato costruito un nuovo livello della matrice originaria in cui, ad ogni commento/caso sono state attribuite tutte le variabili del post a cui il commento risponde. Grazie a questa struttura è stato possibile estrarre un *training set* di 5.604 commenti (pari al 5% dei commenti del campione) in maniera casuale stratificata e proporzionale alle variabili *keyword* e *spectrum*.

I commenti estratti per la costruzione del *training set* sono stati poi etichettati secondo il *sentiment* espresso (positivo, negativo, neutro) e secondo l'orientamento politico espresso (orientato prevalentemente a destra, prevalentemente a sinistra, prevalentemente al centro, orientamento politico rifiutato o non espresso). Terminata la fase di formazione, non è più previsto l'intervento umano.

La capacità di classificazione dell'algoritmo è stata verificata attraverso la matrice di confusione che ci informa sul funzionamento del classificatore rispetto alle diverse classi (Provost & Kohavi, 1998). La matrice di confusione altro non è che una tabella di contingenza delle classi previste e quelle effettive di un classificatore. Le informazioni contenute nella matrice di confusione ci aiutano ad interpretare i diversi aspetti della qualità del classificatore e comprendere le performance del modello

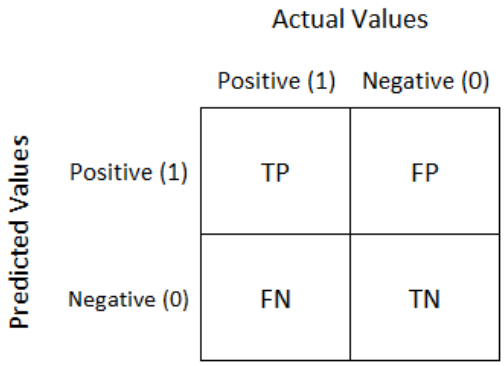
predittivo in modo da determinare quanto questo sia accurato ed efficace. Possiamo avere due tipi di matrice di confusione: quelle che ci forniscono informazioni sulle performance di un modello predittivo a due classi (Si/No, Vero/Falso, Spam/No Spam), quelle ci forniscono informazioni sulle performance di un modello predittivo formato da più di due classi. L'apprendimento automatico restituisce un flusso costante di misure che comunicano gli stati interni del classificatore (come i livelli di precisione o i tassi di errore) queste rendono l'operazione di classificazione osservabile e le scelte prese dall'algoritmo più comprensibili. Queste misure sono utili anche ad analizzare quali caratteristiche dei dati possono influenzare negativamente o positivamente la performance dell'ingenuo Bayes (Rish, 2001) ed essere così grado di definire quali condizioni possono essere sufficienti ad ottimizzarlo (Domingos & Pazzani, 1997).

È la matrice di confusione a restituirci la rappresentazione dell'accuratezza della classificazione.

Attraverso la matrice di confusione si calcolerà prima di tutto l'*accuracy* - ossia la percentuale delle classificazioni corrette data dalla somma degli elementi della diagonale sul numero di elementi totali - e successivamente una serie di altri indicatori sintetici che contengono informazioni riguardo la capacità previsiva del classificatore.

In una matrice a due classi, in cui le colonne rappresentano la classe reale e le righe la classe predetta, sono possibili quattro tipi di classificazione (Hay, 1988).

Figura 10 - Esempio di matrice di confusione[43]



TP: *true positive*, ovvero il numero di osservazioni correttamente previste positive.

TN: *true negative*, ovvero il numero di osservazioni correttamente previste negative.

FP: *false positive*, ovvero il numero di osservazioni erroneamente previste positive (errore di I tipo).

FN: *false negative*, ovvero il numero di osservazioni erroneamente previste negative (errore di II tipo).

[43] Fonte: towardsdatascience.com

Nel caso di questa ricerca, la matrice di confusione è di tipo multi-classe. In una generalizzazione multi-classe, per calcolare il numero di osservazioni corrette e il numero di quelle errate, l'algoritmo procede a calcolare il valore singolo di TP, TN, FP e FN specifico per ogni classe.
Gli indicatori sintetici che è possibile calcolare partendo dalla matrice di confusione sono:

- *Accuracy*, data dal rapporto tra i dati previsti correttamente e il totale delle previsioni. Questa cii informa sulla capacità del modello di prevedere correttamente: la migliore accuratezza è 1, mentre la peggiore è 0 e si calcola con la formula:

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP}$$

- *Sensitivity (*detta anche *Recall* o *True Positive*), data dal rapporto tra TP e il totale delle unità previste positive. Questa misura quanti dati sono stimati correttamente positivi sul totale delle osservazioni effettivamente positive rispondendo alla domanda: "per tutte le istanze che erano effettivamente positive, quale percentuale è stata classificata correttamente?". L'indicatore risponde alla formula:

$$Recall = \frac{TP}{TP + FN}$$

- *Specificity*, data dal rapporto tra TN e il totale delle unità previste negative. Questa misura quante unità previste sono realmente negative sul totale dei valori osservati negativi rispondendo alla domanda *"quando il valore effettivo è negativo, quanto spesso la previsione è corretta?"*. La migliore specificità è 1, mentre la peggiore è 0 e si calcola con la formula:

$$SP = \frac{TN}{TN + FP} = \frac{TN}{N}$$

Altre metriche per la valutazione di un classificatore Naive Bayes (Kuhn, M. 2009.; Kuhn M., et. al, 2020) sono:

- *No Information Rate (NIR):* è il tasso di errore quando l'*input* e l'*output* sono indipendenti. Questo indica la percentuale di precisione ottenibile prevedendo sempre l'etichetta della classe maggioritaria. Per dimostrare che il modello creato è significativo, il valore NIR deve essere inferiore al valore dell'Accuracy.

- *Error Rate:* è il tasso di errore (ERR) e viene calcolato come il numero di tutti i pronostici errati diviso per il numero totale del set di dati. Il miglior tasso di errore è 0, mentre il peggiore è 1.

$$ERR = \frac{FP + FN}{TN + FP + FN + TP}$$

- *Precision*: la precisione è l'abilità di un classificatore di non etichettare un'istanza positiva che è in realtà negativa. Per ogni classe è definito come il rapporto tra veri positivi e la somma di veri e falsi positivi rispondendo alla domanda: *"per tutte le istanze classificate come positive, quale percentuale era corretta?".*

$$Precision = \frac{TP}{TP + FP}$$

- *False positive rate:* Il tasso di falsi positivi (FPR) viene calcolato come il numero di previsioni positive errate diviso per il numero totale di negativi. Il miglior tasso di falsi positivi è 0 mentre il peggiore è 1.

$$FPR = \frac{FP}{TN + FP} = 1 - SP$$

- *F-score:* è una media armonica ponderata delle metriche *Precision* e *Recall* in modo tale che il punteggio migliore sia 1 e il peggiore sia 0. Come regola generale, la media ponderata di F1 dovrebbe essere utilizzata per confrontare i modelli di un classificatore, non la precisione globale.

$$F - score = \frac{2\ x\ Recall\ x\ Precision}{Recall + Precision}$$

Sulla base di questi indicatori, il classificatore applicato ai dati della ricerca restituisce un'accuratezza del 82% nel prevedere la variabile *sentiment,* con una probabilità del 95% che la vera precisione per questo modello si trovi nell'intervallo di confidenza tra 0.80 e 0.83 e una precisione di oltre l'80% sulle tre classi, che fa riferimento all'abilità del classificatore di non etichettare un'istanza con un'etichetta quando in realtà appartiene ad un'altra. Si evince una certa difficoltà di previsione del neutro e del positivo, questo potrebbe essere dovuto al *training set* che (estratto casualmente) è formato solo dal 3% della classe neutra e dal 20% della classe positiva contro il 77% della classe negativa.

Figura 11 - Statistiche di accurata classificazione Naive Bayes per variabile "*sentiment*"

```
Overall Statistics

               Accuracy : 0.82
                 95% CI : (0.801, 0.8379)
    No Information Rate : 0.7391
    P-Value [Acc > NIR] : 1.274e-15

                  Kappa : 0.4479

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: Negativo Class: Neutro Class: Positivo
Sensitivity                   0.9779       0.42105         0.36290
Specificity                   0.3862       0.99573         0.97993
Pos Pred Value                0.8186       0.82051         0.83333
Neg Pred Value                0.8607       0.97378         0.84759
Precision                     0.8186       0.82051         0.83333
Recall                        0.9779       0.42105         0.36290
F1                            0.8912       0.55652         0.50562
```

L'accuratezza del classificatore diminuisce drasticamente nella previsione dell'orientamento politico (53% con un NIR del 72% che ci informa sul cattivo funzionamento del modello). Nonostante la precisione del 84,8% sull'orientamento di destra, ha difficoltà a prevedere gli altri orientamenti (34,4% non espresso; 10,5 rifiutano collocazione; 0,66% sinistra).

Figura 12 - Statistiche di accurata classificazione Naive Bayes per variabile "*orientamento politico*"

```
Overall Statistics

               Accuracy : 0.5393
                 95% CI : (0.5145, 0.5639)
    No Information Rate : 0.7219
    P-Value [Acc > NIR] : 1

                  Kappa : 0.2005

 Mcnemar's Test P-Value : <2e-16

Statistics by Class:

                     Class: Destra Class: Non Espresso Class: Rif. Collocazione Class: Sinistra
Sensitivity                 0.5285              0.6723                 0.333333        0.121622
Specificity                 0.7556              0.6384                 0.967844        0.916993
Pos Pred Value              0.8488              0.3449                 0.105263        0.066176
Neg Pred Value              0.3817              0.8731                 0.992243        0.955722
Precision                   0.8488              0.3449                 0.105263        0.066176
Recall                      0.5285              0.6723                 0.333333        0.121622
F1                          0.6514              0.4559                 0.160000        0.085714
```

I risultati del classificatore bayesiano nel prevedere l'orientamento politico si sono rilevati non sufficienti. La difficoltà può esser dovuta alla composizione del *training set*[44] ma anche al particolare scenario politico che ha caratterizzato l'Italia negli ultimi anni, avvicinando posizioni politiche prima d'ora impensabili. La presenza di due posizioni politiche in egual misura trattate all'interno dello stesso commento (pensiamo a Lega-5Stelle o PD-5Stelle) potrebbe aver creato difficoltà nel processo di disambiguazione.

Dunque, in che direzione muoversi affinché al potenziamento delle tecniche si accompagni un miglioramento della qualità dell'analisi (Amaturo, 1989)?

Per rispondere a questa domanda è stata messa a punto una strada ibrida con un'inedita combinazione tra approccio supervisionato al *machine learning* e approccio basato sul dizionario. Questa strada fa propria la convergenza, tesa a superare i limiti di un approccio con i vantaggi dell'altro e viceversa (Amaturo e Punziano, 2016).

3.7 Una strada ibrida all'*Opinion Mining*

Secondo Marradi (1996) l'essenza del lavoro di un ricercatore sociale è intrinseca nella scelta tra le tecniche già individuate e sviluppate da altri ricercatori e nella possibilità di concepirne di nuove. Riconoscendo i limiti specifici di ciascuna famiglia di algoritmi, diversi ricercatori hanno proposto strategie ibride, ovvero strategie che uniscono i punti di forza di differenti approcci, affiancando o incorporando elementi di una strategia all'interno di un'altra. In questo studio, seguendo il classico funzionamento dell'approccio all'*opinion mining* basato sul lessico è stato creato un dizionario dell'orientamento politico degli utenti. Come descritto nei paragrafi precedenti, l'approccio non supervisionato all'*opinion mining* utilizza liste composte di *"opinion word"* di cui calcola le occorrenze nei testi per attribuirne la polarità. Allo stesso modo, partendo dal *training set* etichettato, è stato costruito un vocabolario dell'orientamento politico espresso attraverso il quale sono state estratte le caratteristiche associate ad ogni categoria definendo un vettore di proprietà che le rappresenta (dette *feature*[45]). Dato un testo, l'estrazione delle *feature* è quel processo di estrapolazione delle sue proprietà salienti che rappresentano appunto le caratteristiche fondamentali del testo.

Messo a punto il vocabolario, sono stati classificati i commenti dell'intero campione. Nel caso specifico di questa ricerca, l'approccio messo a punto prevede un *check* umano post classificazione

[44] 4% Rifiutano Orientamento; 8% Sinistra; 44% Destra; 44% Non Espresso.

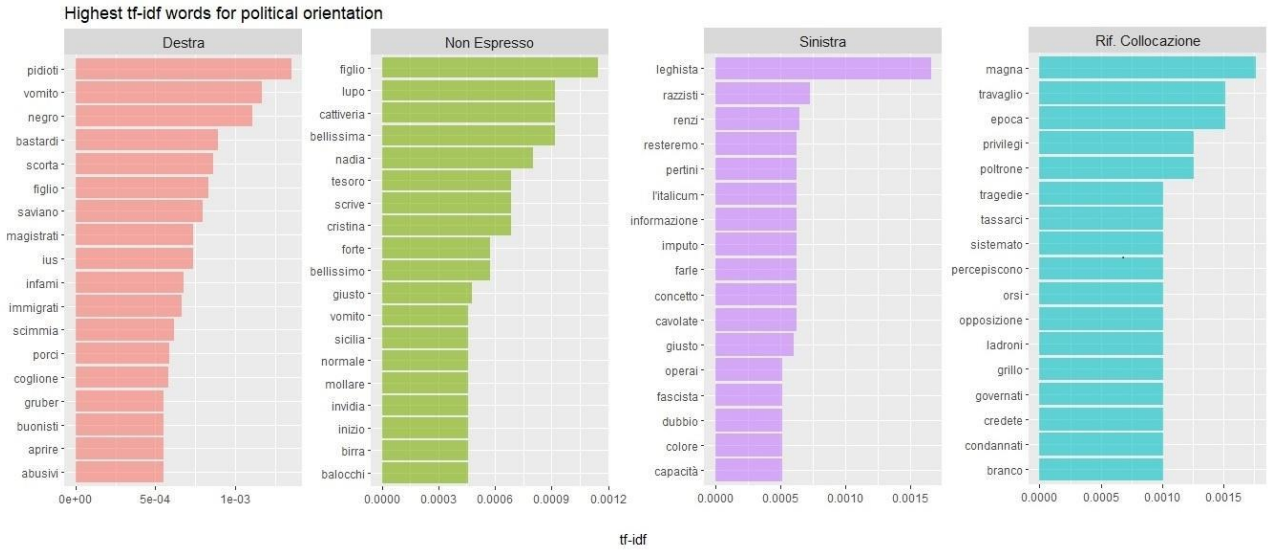
[45] L'estrazione delle *feature* di un testo è il processo di estrapolazione delle sue caratteristiche salienti, le più utilizzate sono: parole, parti del discorso, opinion words, negazioni.

date le particolarità che hanno caratterizzato la politica italiana negli ultimi anni, ma non è escluso che questo possa essere evitato se l'approccio viene applicato a testi di diversa natura.

Considerato che l'estrazione delle *feature* non è così immediata - perché se da un lato devono discriminare e descrivere il più possibile il testo originale, dall'altro devono anche ridurre l'ampia dimensione dei dati di origine ed evitare ridondanze - la funzione ritenuta più adatta a far emergere in maniera evidente le caratteristiche lessicali dei diversi orientamenti politici al fine di creare il vocabolario è la Term Frequency-Inverse Document Frequency (da ora TF-IDF). Questa funzione è ampiamente utilizzata nel campo dell'*information retrieval* per misurare l'importanza di una parola per un documento in rapporto ad una raccolta di documenti. L'idea di TF-IDF è trovare le parole rilevanti per il contenuto di ogni documento diminuendo il peso per le parole di uso comune e aumentando il peso per le parole non comuni presenti in una raccolta di documenti (Silge & Robinson, 2017). Semplificando al massimo possiamo dire che la funzione non estrae le parole più frequenti ma quelle più caratteristiche per ogni categoria; in questo modo se quelle parole dovessero apparire, ad esempio in una *query* di un motore di ricerca, il documento apparirebbe nei risultati perché ritenuto d'interesse per l'utente che ha effettuato la ricerca[46]. Negli studi di linguistica computazionale è frequentemente evidenziato quanto una classica misura di frequenza ponga troppa enfasi sui termini ad alta frequenza e quanto, invece, le misure di specificità assegnino troppo peso ai termini a bassa frequenza (Manning & Schutze, 1999). La difficoltà con la selezione o la ponderazione dei termini sta nello stabilire un buon equilibrio tra popolarità e specificità (Aizawa, 2003). La statistica TF-IDF sembra il giusto compromesso, in quanto non misura quanto spesso un termine appare in un singolo documento ma aiuta a capire l'importanza di un termine calcolando quanto spesso appare in confronto ad altri documenti simili.
Prima di procedere all'analisi, il testo è stato pulito da un elenco di parole ritenute non rilevati così da ricevere una migliore e più significativa trama di analisi.

[46] L'esempio è solo a titolo esplicativo. I motori di ricerca sfruttano funzioni molto avanzate dell'*information retrieval* per quantificare il livello di pertinenza di un documento rispetto alla query dell'utente e non si basano solo sulla funzione base TF-IDF.

Grafico 1 - Parole con valore TF-IDF più alto in base all'orientamento politico



Il grafico mostra con evidenza quanto il lessico utilizzato dagli utenti sia chiaro e si presti a poche ambiguità. Allora perché è necessario un *check* umano post classificazione?

Come ipotizzato riguardo il malfunzionamento del classificatore bayesiano, i commenti restituiscono tutta la particolarità dello scenario politico italiano degli ultimi anni, esprimendo nello stesso momento e in egual misura posizioni politiche diverse. La tabella sotto mostra il corretto funzionamento del modello che assegna lo stesso punteggio di *feature* su più orientamenti politici presenti nello stesso commento. Per risolvere l'impasse è necessario dunque l'intervento umano.

Il ricercatore è qui chiamato a prendere delle decisioni spesso di non facile risoluzione.

Tabella 1 - esempi di etichettatura dell'orientamento politico con approccio ibrido

ID	Text	Right	Left	Not Expressed	Refuse
text29288	Prova a domandarti invece se gli italiani voteranno ancora PD dopo le schifezze della famiglia Boschi su Banca Etruria o gli intrallazzi del padre di Renzi con gli outlet o il caso Consip o il Matteo che va dalla Gruber a dire che il debito pubblico è calato ma viene smentito meno di 24 ore dopo dalla Banca d' Italia. Perché non fai questo sondaggio con gli Italiani? Sono certo che riceverai solo un vaffanculo e w il movimento 5stelle!	1	1	0	1
text11445	AVANTI Movimento 5 Stelle - LEGA!!! LA MAGGIORANZA DEGLI ITALIANI SOSTIENE QUESTO GOVERNO PERCHÉ CONVINTA CHE È L'UNICA SOLUZIONE PER IL PAESE!!!	1	0	0	1
text80139	Ho votato 5Stelle, ma se rompono per la TAV voto Salvini. Vogliamo un'Italia al passo coi tempi e senza immigrati. Quindi di Maio e Di Battista non rompete i coglioni. Si va avanti così!	1	0	0	1

Le statistiche di funzionamento del modello messo a punto ci dimostra che i risultati migliori sono stati ottenuti nel momento in cui non ci si è affidati totalmente all'automazione ma quando il ricercatore ha assunto un ruolo rilevante nel processo decisionale.

Figura 13 - Accuratezza di classificazione del modello ibrido per la variabile *"orientamento politico"*

```
Overall Statistics

               Accuracy : 0.781
                 95% CI : (0.77, 0.7918)
    No Information Rate : 0.4863
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.652

 Mcnemar's Test P-Value : < 2.2e-16

 Statistics by Class:

                      Class: DESTRA Class: NO ESPRESSO Class: RIFIUTANO Class: SINISTRA
 Sensitivity                 0.9026             0.7728          0.34944          0.83801
 Specificity                 0.7850             0.8864          0.99585          0.98483
 Pos Pred Value              0.6791             0.8656          0.89952          0.83262
 Neg Pred Value              0.9411             0.8048          0.93513          0.98540
 Precision                   0.6791             0.8656          0.89952          0.83262
 Recall                      0.9026             0.7728          0.34944          0.83801
 F1                          0.7750             0.8166          0.50335          0.83531
```

4 Umano, troppo umano. Il *bias* dell'approccio supervisionato

Nel chiudere lo spazio dedicato alla trattazione delle tecniche riporto quanto Marradi nel 1989 scriveva a proposito della diffusione in Italia delle tecniche di analisi delle corrispondenze multiple: «*ciò che più mi piace di questa tecnica sono i suoi limiti, quelli che fanno inorridire alcuni statistici*» (p.1).

Nell'utilizzo di tecniche di *machine learning* si conoscono l'*input* e l'*output* e spesso vengono utilizzate senza essere messe in discussione, di fatto reificandosi e diventando oggetti veri e propri. Diventa invece fondamentale porre attenzione al processo che è stato messo in atto per raggiungere la rappresentazione dell'oggetto indagato e allo scopo che si intende perseguire.

Anche se la proliferazione della lingua scritta ha inevitabilmente reso essenziale il supporto computazionale, è necessario confutare due idee di fondo che accompagnano l'utilizzo della tecnica: l'idea che si possano ottenere buoni risultati con qualsiasi tipo di testo grazie solo alla sofisticata elaborazione dell'algoritmo e l'idea che il procedimento, al suo completamento, ci restituisca dei risultati oggettivi, neutri e perfettamente replicabili.

Riguardo la prima idea: la complessità del linguaggio umano è una delle difficoltà più grosse da affrontare nel trattamento automatico dei testi. A questa complessità va aggiunta la particolarità del tipo di linguaggio derivante dalle piattaforme *social media* che risente non solo dei classici elementi linguistici come sarcasmo o ironia, ma anche del funzionamento della piattaforma stessa: ad esempio, la pratica di scrittura su Twitter è diversa da quella di Facebook perché diverso è lo spazio dedicato al testo (Twitter ha un massimo di battute) e diverse sono anche le logiche di diffusione dei contenuti. Inoltre, il linguaggio dei *social* si caratterizza particolarmente per dialettalismi ed errori ortografici propri della trasposizione della lingua parlata nello scritto.

Riguardo la seconda idea: essendo la fase di addestramento molto delicata, in quanto da essa dipende la corretta classificazione dell'algoritmo, si è ritenuto necessario provare a comprendere l'eventuale presenza di soggettività nel procedimento di classificazione dei testi e l'entità di tale soggettività. Per questo scopo sono stati estratti casualmente cento commenti dal *training set* e sottoposti all'etichettatura di due tester esterni alla ricerca, ma con pari competenze.

Grafico 3 - Confronto tra etichettature su variabile *sentiment*

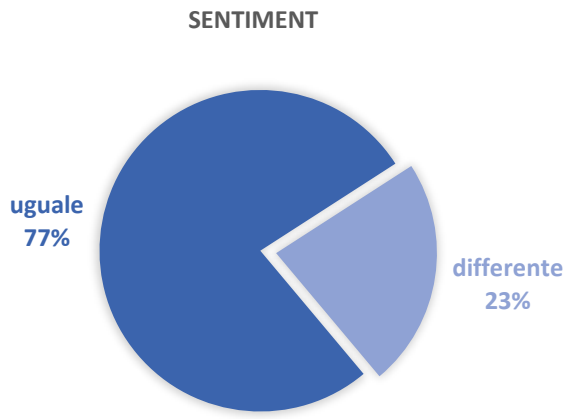
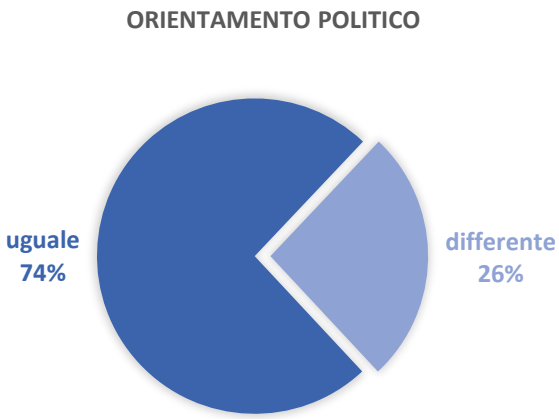


Grafico 2 - Confronto tra etichettature su variabile orientamento politico



Il confronto tra le procedure fa emergere con chiarezza di quanta soggettività si caratterizza questa fase: 26% sull'orientamento politico e 23% sul *sentiment*. Questa differenza di valutazione apre a diversi interrogativi e respinge il mito dell'oggettività e dell'automazione di queste tecniche, mettendone in crisi l'impiego del tutto acritico in diversi campi e con diversi tipi di dati.

Il processo di apprendimento degli algoritmi supervisionati per gestire dati testuali è decisamente antropocentrico poiché l'azione umana resta necessaria per interpretare il significato del testo. Il linguaggio verbale può avere molteplici valenze soprattutto se ad un messaggio è associata un'immagine o un'*emoticon*, come nel caso dei *social network*. L'interpretazione del testo guida la classificazione e le scelte nel *set* di allenamento; queste scelte riguardano principalmente cosa tenere in considerazione e come elaborarlo con un impatto sui dati finali, sul modello scelto e sugli obiettivi perseguiti. Queste scelte sono culturalmente radicate e posso influenzare pesantemente i risultati. Dunque, la possibilità nel *machine learning* di "mostrare" alla macchina come funzionare non significa solo mettere a punto un modello analitico ma più in generale un modo di guardare e agire sul mondo. È per questo motivo che Knorr Cetina (1999, p. 5) considera le tecniche di apprendimento automatico come "macchine del sapere".

Un modello analitico funzionante è l'esito di un complesso processo negoziale tra uomo e macchina che può avere ampio respiro ma anche conseguenze impreviste. Le conseguenze delle decisioni prese dal ricercatore sono difficili da stimare in anticipo a causa del numero elevato di elementi, proprietà e iterazioni. Per mettere a punto una "ricetta decisionale" (Rieder, 2020), la riflessione sullo scopo dell'applicazione della tecnica è certamente parte della pratica stessa: un'interpretazione "corretta" è possibile solo attraverso un'etichettatura manuale o *feedback* continuo, intervenendo dunque attivamente negli spazi che si cercano di rappresentare. La classificazione non è fissa e inalterabile ma

è frutto di una serie di schematismi variabili, ciascuno costruito per uno scopo specifico. Lo scopo guida il processo di creazione del modello ed è espresso attraverso la definizione di classi (*spam-no spam*, positivo-negativo, vero-falso, *ecc*). In questa fase, l'analista sta mettendo a punto una base di conoscenza da utilizzare nel processo decisionale dell'algoritmo, questa non è una formula matematica neutra ma una valutazione personale che andrà a nutrire un modello statistico adattativo.
Un algoritmo incorpora il "pregiudizio" di chi sviluppa il modello che successivamente traduce in codice. La conoscenza prodotta nei contesti in cui gli algoritmi vengono applicati è dunque orientata piuttosto che neutra e imparziale (Aragona e Felaco, 2018). Quello che sembra essere un processo automatizzato, è in realtà il risultato di una serie di scelte fatte da un determinato punto di vista. Gli algoritmi, quindi, mediano la conoscenza e contribuiscono ontologicamente alla produzione della realtà (Aragona e Felaco, 2020). Scelte diverse porteranno a diversi risultati.

La comprensione delle specificità di questi processi diventa necessaria se allarghiamo l'ambito oltre le singole esperienze di ricerca e cerchiamo di capire come la soggettività di questi processi si ripercuote nelle pratiche di utilizzo delle piattaforme. Un esempio sono gli strumenti di segnalazione dei contenuti da parte degli utenti: questi servono chiaramente come *input* di addestramento dei classificatori per identificare contenuti illegali, dannosi o generalmente indesiderati. Ma come vengono decisi questi valori *input*? Come vengono operativizzati e come trovano espressione in termini tecnici?
Non sono certo una novità i casi di discriminazione dovuti alla censura o alla minor visibilità di immagini ritraenti corpi femminili come donne grasse, nere o in generale lontane da un determinato tipo di fisicità (caso Celeste Barber, 2020; caso Nyome Nicholas-Williams, 2020; caso Sara Newmann, 2014). Queste perplessità necessitano senz'altro di più specifici approfondimenti, ma è certo che il peso della soggettività in fase di addestramento degli algoritmi di classificazione e previsione è un argomento che trova ancora poco spazio di discussione nella comunità scientifica che ne fa largo uso.
Per concludere, possiamo considerare qualsiasi modello di apprendimento automatico una risposta pratica a decisioni umane assunte in ogni fase di creazione del processo. L'utilizzo del classificatore di Bayes fornisce sicuramente la possibilità di portare avanti analisi complesse e interessanti, ma proietta inevitabilmente lo scopo e le opinioni di chi sta portando avanti l'analisi che, con il proprio bagaglio di interessi e aspettative, va in qualche modo ad indirizzare il disegno dell'algoritmo.
Questa considerazione si traduce in una preoccupazione rispetto al modo in cui il *machine learning* è spesso utilizzato e discusso nei *media studies*: ossia con un'enfasi posta sui modelli come sinonimo di oggettività. Gli approcci computazionali hanno dei vantaggi che non possono essere certo ignorati, ma il modello non è né oggettivo né neutro e l'analista deve essere consapevole che i risultati ottenuti dipendono in maniera costitutiva, anche se non esclusiva (Marradi, 1989), da una serie di decisioni a

lui stesso rimandate. Una consapevolezza necessaria affinché si possa procedere all'analisi con la responsabilità dei risultati attribuibili a sé stesso e non al tipo di algoritmo scelto.

È importante infine tener presente che non esiste un modello di *machine learning* che funzioni meglio di un altro: nell'apprendimento automatico è celebre il teorema *"No Free Lunch"* di David Wolpert e William Macready (1997), secondo il quale un modello può rivelarsi ottimo per un problema e pessimo per un altro. È quindi necessario testare più modelli o più volte lo stesso modello affinché si possa trovare la strada che funziona meglio per il problema che si sta indagando, questo presuppone una conoscenza approfondita dei propri dati e del contesto in cui sono stati raccolti.

Qualunque sia l'approccio, la figura dell'analista o ricercatore - seppur fortemente sostenuta dagli strumenti computazionali - resta centrale, attiva e vigile. Egli, come nei tradizionali tipi di indagine è decisore, controllore, osservatore attento e garante della qualità dei risultati sia nel mettere a punto e nel testare gli algoritmi, sia nel fissare parametri e nel costruire modelli.

CAPITOLO IV - Strategie e messaggio. Lo spettro disinformativo

Come ampiamente descritto nei capitoli precedenti, il termine *"fake news"* non è sufficiente a descrivere la vasta gamma di contenuti disinformativi che ogni giorno è prodotta e diffusa in rete. La disinformazione ha confini molto labili e in questa particolare fase di analisi ci si muove lungo un *continuum* che va dal completamente falso al completamente vero e che ha nel mezzo una numerosa gamma di sfumature che non possono essere ignorate. Il processo di operativizzazione della disinformazione realizzato in questo studio nasce da due esigenze: quella di superare il criterio intenzionale sul quale si basa la classificazione dei contenuti disinformativi presente in letteratura e quella di adattare una categorizzazione profonda del fenomeno all'analisi quantitativa di dati non strutturati provenienti dalle piattaforme. Così, al lavoro di classificazione di Claire Wardle[47] (Wardle & Derakhshan, 2017; Wardle, 2018) sono state integrate le strategie di costruzione dei contenuti disinformativi sviluppate da MediaLab (Università di Lisbon[48]a); e le categorie sono state modellate in riferimento alla letteratura sulle caratteristiche dell'ambiente disinformativo italiano. Questo lavoro di rielaborazione della disinformazione è stato necessario per superare il limite presente nella categorizzazione della Wardle - fortemente associata alla valutazione dell'intenzione e quindi molto difficile da rilevare empiricamente e da valutare oggettivamente attraverso il tipo di dati e le tecniche utilizzate in questa ricerca – e il limite dell'approccio messo a punto dall'Università di Lisbona, sicuramente più pratico del precedente ma adatto a piccoli *set* di dati in quanto volto a studiare non solo i contenuti disinformativi ma anche tutta la filiera di verifica dei fatti[49].

Infine, a differenza del lavoro di categorizzazione presente in letteratura che (come descritto nel capitolo 2) differenzia i tipi di disinformazione sui fattori di intenzione e *facticity*, in questa ricerca la distinzione è effettuata in base al fattore di rischio disinformativo.

Date queste considerazioni è stato costruito uno spettro dei contenuti disinformativi, uno strumento grazie al quale è stato possibile ottenere un indice di rischio che accompagnerà l'interpretazione dei dati durante tutta la ricerca.

Lo spettro continuo è lo strumento che meglio si addice a rappresentare il *continuum* che caratterizza la disinformazione ed è necessario per non cadere nell'errore, spesso commesso da approcci

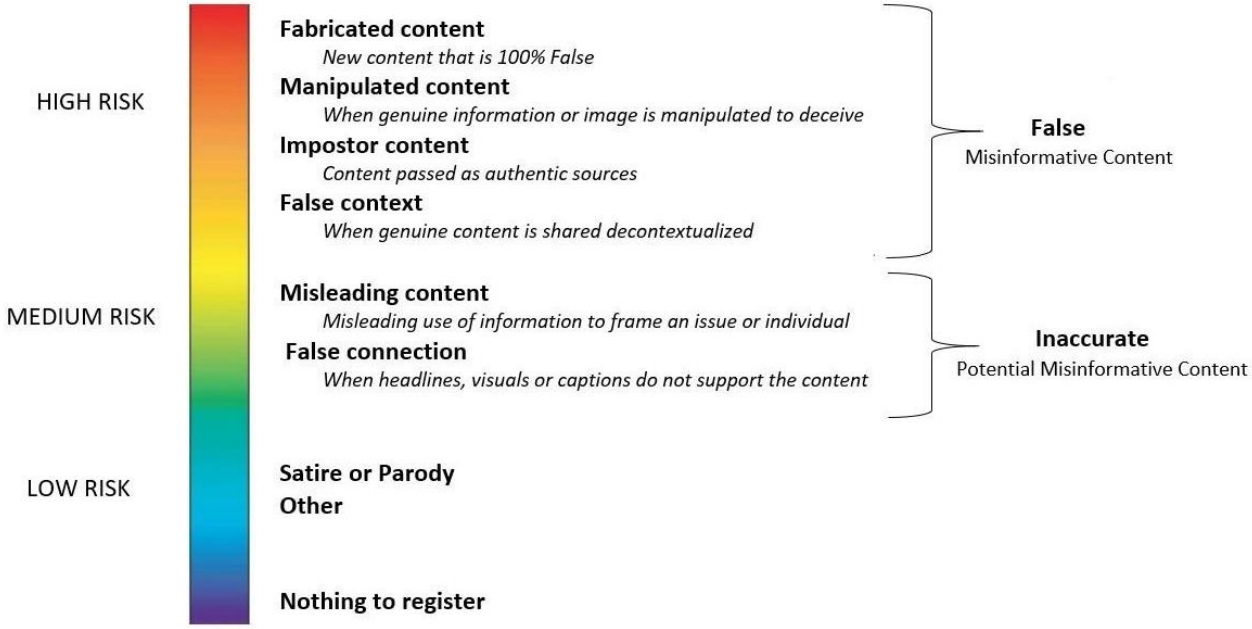
[47] Claire Wardle parte dalla suddivisione di tre elementi: 1. i diversi tipi di contenuto che vengono creati e condivisi; 2. le motivazioni di chi crea questo contenuto; 3, i modi in cui questo contenuto viene diffuso. In questo studio si affronterà solo il primo elemento perché il tipo di dati scelti ci impedisce di rilevare gli intenti che si celano dietro i produttori di notizie.

[48] Personalmente testate prima di questa analisi durante lo studio della disinformazione nei gruppi Facebook pro-Brexit in un disegno Digital Methods messo a punto durante l'esperienza di ricerca nel 2020 alla Digital Methods Initiative di Richard Rogers, Università di Amsterdam.

[49] Questo approccio prevede la compilazione di una dettagliata scheda di analisi per orgni contenuto analizzato attraverso la quale è possibile ricostruire il processo di *fact-checking.*

esclusivamente *data-driven*, di considerare i contenuti diffusi sulle piattaforme come *fake news* indistinte.

Figura 14 - Spettro dei contenuti disinformativi. Elaborazione dell'autore.



Sulla base di questo strumento, è esposto ad alto rischio disinformativo un utente che entra in contatto con:

- *Fabricated Content*: contenuti creati *ad hoc* e completamente falsi (non hanno alcuna relazione con la realtà).
- *Manipulated Content*: quando informazioni e immagini vere sono modificate *ad hoc* per ingannare.
- *Impostor Content*: quando un contenuto falso è presentato come fonte autentica (es. indicazione di fonti attendibili per avvalorare la notizia falsa).
- *False Context:* quando un contenuto vero è condiviso decontestualizzato rendendo l'informazione generale falsa (un esempio frequente è l'utilizzo di foto di celebrità di origine africana ritratte in ambienti facoltosi fatte passare per immigrati in centri di accoglienza).

Il rischio disinformativo diminuisce se l'utente entra in contatto con:

- *Misleading content*: uso ingannevole di un'informazione vera per inquadrare un problema o un individuo (un esempio frequente è l'utilizzo di notizie su attentati *jihadisti* in riferimento alla presenza islamica in Italia gridando all'"invasione islamica").
- *False connection*: quando il titolo, il testo o l'immagine utilizzata non fa riferimento al contesto. Questo è il *Framing Effect*[50] (Goffman, 1986) che fa leva sulla differente modalità di esposizione del messaggio per influenzare la percezione dello stesso. Il *frame* lo si può intendere come una cornice che contorna un determinato elemento e che cerca di incoraggiare una certa interpretazione e scoraggiarne delle altre. Dunque, anche se in presenza di due elementi veri, messi in connessione tra loro in un'unica cornice interpretativa, rendono la notizia generale falsa o inaccurata.

Il rischio disinformativo è basso se l'utente entra in contatto con "*Satire or Parody"*.
Come già discusso, la satira e la parodia non possono esser considerati come rischi disinformativi al pari di un contenuto fabbricato o manipolato, ma presentate in un certo modo possono rovesciare completamente il significato di un messaggio[51].
Infine, la modalità *"Nothing to Register"* è stata utilizzata per indicare tutti quei contenuti sui quali non si è registrato alcun rischio e che quindi possono essere considerati attendibili.

Ogni post è stato analizzato singolarmente in base al contenuto, al contesto e agli indizi visivi. Consapevole dell'esistenza di posizioni molto diverse riguardanti la validità e la coerenza dei metodi di verifica dei fatti (Amazeen, 2015; Marietta, Barker & Bowser, 2015), il *debunking* è stato effettuato consultando i siti governativi, i quotidiani nazionali e il servizio ANSA nonché i servizi *fact-checking* stretti in collaborazione con Facebook e infine i siti *web* firmatari del codice dell'International Fact-Checking Network di Poynter[52], che stila una serie di impegni a cui le organizzazioni si attengono per promuovere il controllo dei fatti.

I post del campione sono stati verificati, etichettati e classificati uno per uno.

[50] Il primo ad analizzare le teorie di framing nel campo della comunicazione fu Gregory Bateson (1954), ma fu il sociologo Erving Goffman (1986) a ritenere che esistano dei veri e propri schemi di interpretazione o *framework* che gli individui utilizzano per contestualizzare la realtà che li circonda. Questi li aiutano a comprendere, definire e giudicare l'accadimento degli eventi. Società e cultura sono i principali influenzatori dei *frame* utilizzati dalle persone.
[51] Un esempio sono gli studi di Kendall e Wolf (1946) nello studio dei pregiudizi razziali e religiosi: utilizzando vignette satiriche che ridicolizzavano il pregiudizio razziale e religioso i ricercatori scoprirono che gli individui che condividevano quei pregiudizi avevano una lettura distorta del materiale mostrato che rafforzava il pregiudizio invece di minarlo.
[52] Consultabile al link: www.poynter.org/ifcn/

Il grosso del lavoro di questa ricerca è stato dunque dedicato alla costruzione di categorie che riuscissero a restituire quella profondità contestuale necessaria a rispondere agli obiettivi della ricerca. Il lavoro di categorizzazione è stato effettuato non solo sull'oggetto della ricerca ma anche sugli oggetti digitali confezionati dalla piattaforma. Come sostengono Enrica Amaturo e Biagio Aragona (2019) i dati digitali che vengono raccolti attraverso pratiche automatizzate (come le interfacce di programmazione) richiedono maggiore attenzione alle procedure di preparazione all'analisi che prevedono il modo in cui i dati vengono selezionati, ridotti e organizzati per le analisi. Per questo motivo uno degli sforzi è stato proprio quello di andare oltre le informazioni preconfezionate dall'API aggirando così la politica della piattaforma che struttura gli oggetti digitali. Il lavoro di pre-analisi è stato particolarmente lungo ma indispensabile per ottenere la profondità che la ricerca intende perseguire.

Il primo obiettivo è stato quello di differenziare nel dettaglio tutti i possibili tipi di contenuto che possono essere utilizzati a supporto dei contenuti disinformativi.

Grafico 5 - Tipo di post derivante dall'API (% n= 1.877)

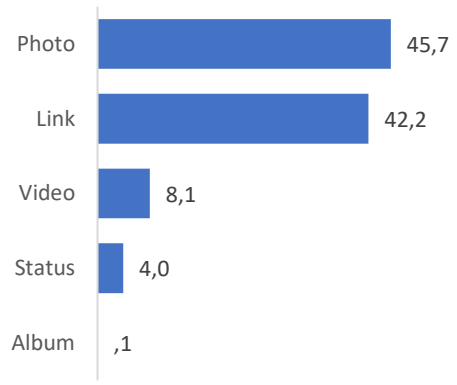
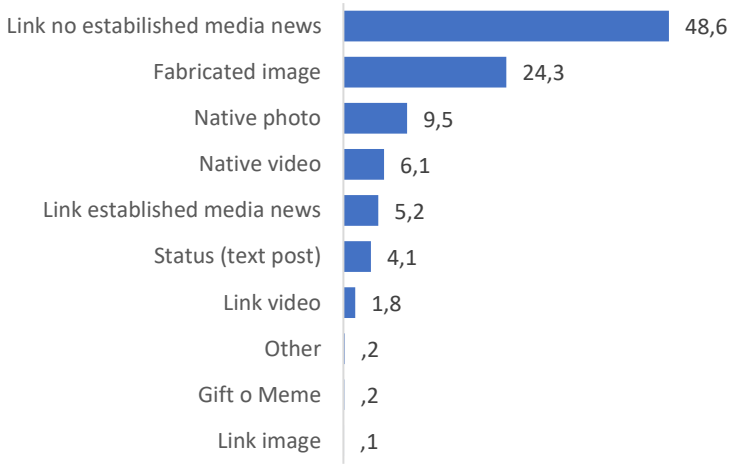


Grafico 4 - Tipo di post rielaborato (% n= 1.877)

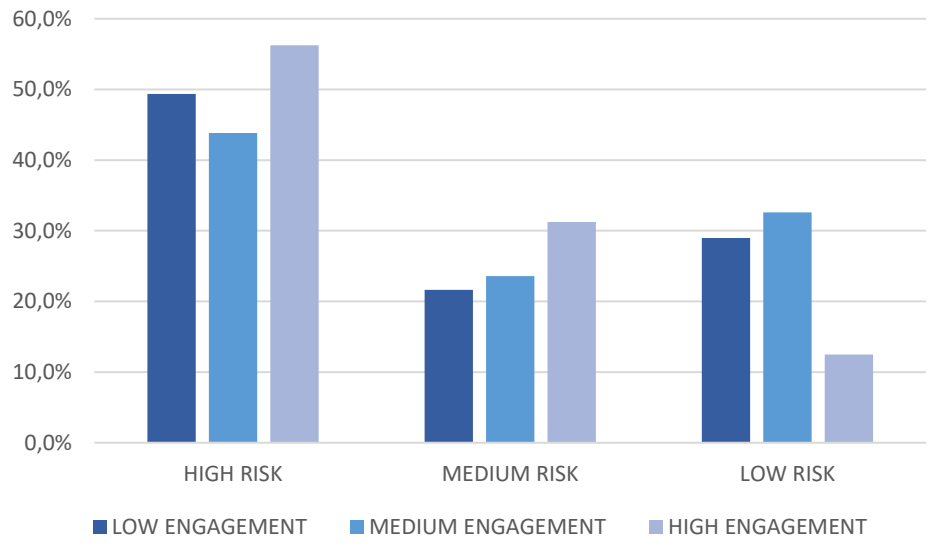


Come si legge nei grafici, la modalità *link* dell'API è stata scomposta in "*link* ad un'immagine", "*link* ad un video", "*link* ad un articolo di una testata giornalistica attendibile", "*link* ad un articolo di una testata giornalistica non attendibile". Anche la modalità foto è stata scomposta in: "*native photo*" e "immagini fabbricate"; questa categoria, a differenza della precedente, comprende immagini sulle quali sono stati aggiunti elementi non originariamente parte di esse. Seguendo lo stesso ragionamento la modalità video è stata scomposta in "*native video*" ossia video caricati direttamente sulla piattaforma e "*link video*" ossia video caricati altrove e riportati attraverso il collegamento *link* in piattaforma.

Inoltre, una particolare attenzione è stata data al coinvolgimento emotivo dell'utente, variabile rilevante soprattutto nella seconda fase di analisi. Considerando che la piattaforma definisce il funzionamento degli oggetti digitali e dunque possiede una sua politica (Hochman & Manovich, 2013), il valore di *engagement rate* restituito dall'API è stato ricostruito in una nuova variabile *engagement*. Questa è data dalla somma del numero di *like*, *share*, commenti e *reaction*, eludendo così le logiche commerciali della piattaforma fondate sulla vendita di visibilità dei contenuti. Ricostruendo questa variabile è stato possibile misurare il coinvolgimento considerando un approccio alla raccomandazione partecipativo e incentrato sull'utente piuttosto che sulla piattaforma. Possiamo a questo punto definire l'*engagement* come il risultato di scambi relazionali ripetuti tra l'utente e il contenuto della piattaforma: il grado di questa interazione può essere considerato un coinvolgimento dato dalla particolare in sintonia tra contenuto e l'utente (dimensione cognitiva), o un coinvolgimento dato da particolari emozioni suscitate nell'utente in contatto con il contenuto (dimensione emozionale).

Considerando il valore medio e quello massimo, l'*engagement* è stato classificato in tre modalità: *low* (da 0 a 5.000 interazioni), *medium* (da 5001 a 20.000 interazioni), *high* (oltre le 20.000 interazioni). Questi valori sono stati analizzati in riferimento al rischio disinformativo.

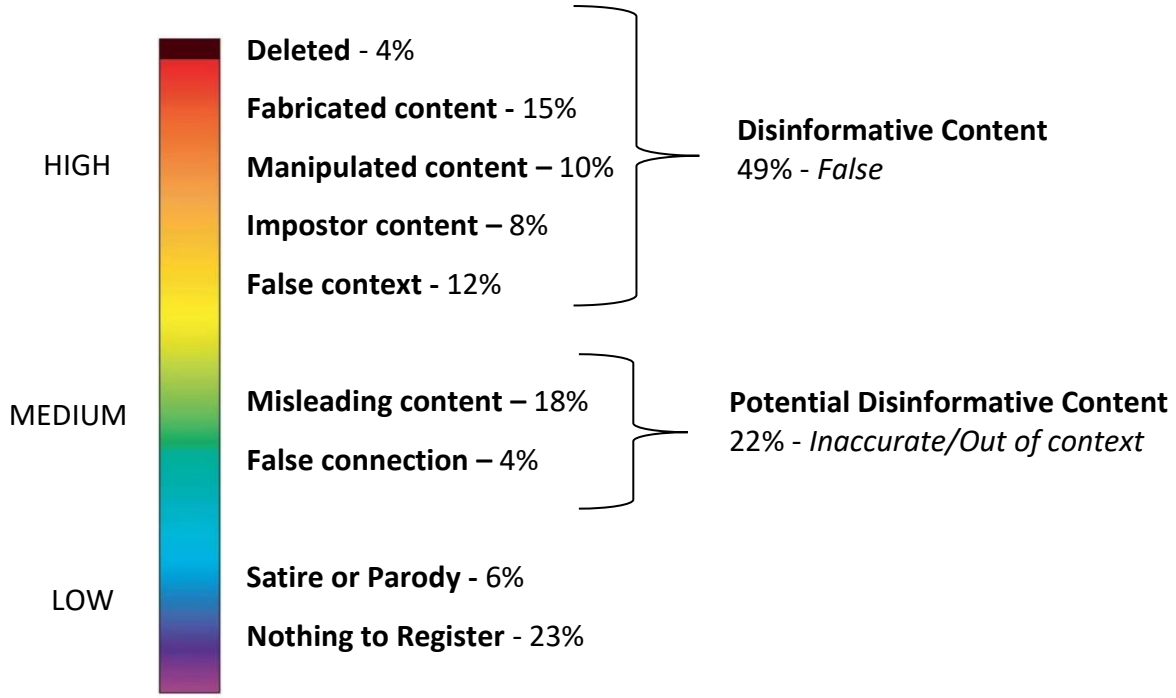
Grafico 6 - Rischio disinformativo per engagement (% n= 1.877)



Le percentuali registrate fanno vacillare l'idea, ampiamente diffusa, che i contenuti disinformativi siano più virali di altri perché rispondono alla logica del *clickbait*; su questa convinzione sono state costruite anche le campagne di sensibilizzazione al riconoscimento di un contenuto falso o fuorviante. Considerando il ruolo della piattaforma, è ragionevole pensare che la diffusione sia determinata dal

suo funzionamento che influisce anche sull'adozione di un particolare tipo di comportamento da parte degli utenti nelle loro pratiche di fruizione dei contenuti.
Il campione estratto è composto per il 49% da contenuti ad alto rischio disinformativo e per il 22% da contenuti potenzialmente disinformativi, dal 23% di contenuti sui quali non si è registrato nessun rischio e dunque non utili agli obiettivi della ricerca e un 6% di ironia e satira.

Figura 15 - Distribuzione dei contenuti sullo spettro disinformativi (% n= 1.877)



La disinformazione non è limitata nel tempo ed è probabile che si intensifichi prima o durante significativi processi decisionali democratici. Approfondendo l'analisi della produzione di disinformazione durante l'arco temporale considerato nell'analisi è infatti emerso che picchi più alti di rischio si registrano durante particolari momenti storico-politici del paese (grafico 7): nei primi mesi del 2016, segnati dalla crisi dei migranti; durante i mesi estivi del 2018, caratterizzati dalle elezioni politiche e dalla creazione del primo governo Conte (M5S+Lega); nei mesi invernali del 2019, durante i quali abbiamo assistito alla crisi di governo e alla nascita del Conte-bis; nei primi mesi del 2020, tristemente noti per il dilagare della pandemia da Coronavirus.
A supporto di questa interpretazione è stato confrontato l'andamento della disinformazione con quello delle *keyword* manualmente prodotte durante la fase di etichettatura. Il grafico 8 mostra i picchi più alti delle *keyword* "immigrazione", "politica", "covid", proprio in corrispondenza dell'alto e medio

rischio disinformativo. Si è ritenuto necessario etichettare ogni post con una *keyword* perché dalla prima esplorazione sui tipi di post restituiti dall'API (45,7% foto, 42,1% *link*, 8% video, 4% *status*) si temeva che l'esclusivo utilizzo delle tecniche di *topic modeling* fosse inadatto visto che solo il 4% dei contenuti è esclusivamente testuale. Durante la fase ermeneutica è poi emerso che quasi tutte le foto, i video e i *link* sono accompagnati da una cerniera introduttiva di testo; questo però più che fornire indicazioni sul tipo di argomento trattato nel *post* fornisce indicazioni sulla posizione di chi produce il contenuto in relazione al tema trattato.

Grafico 7 - Andamento del rischio disinformativo nel tempo (% n= 1.877)

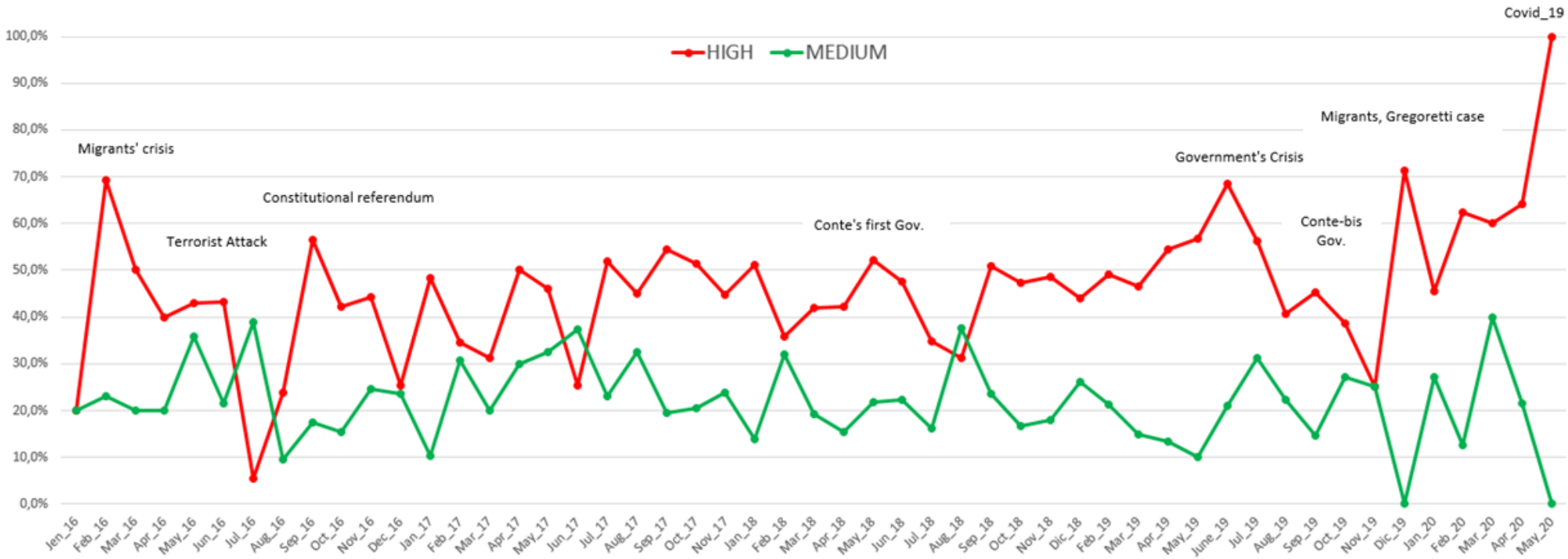
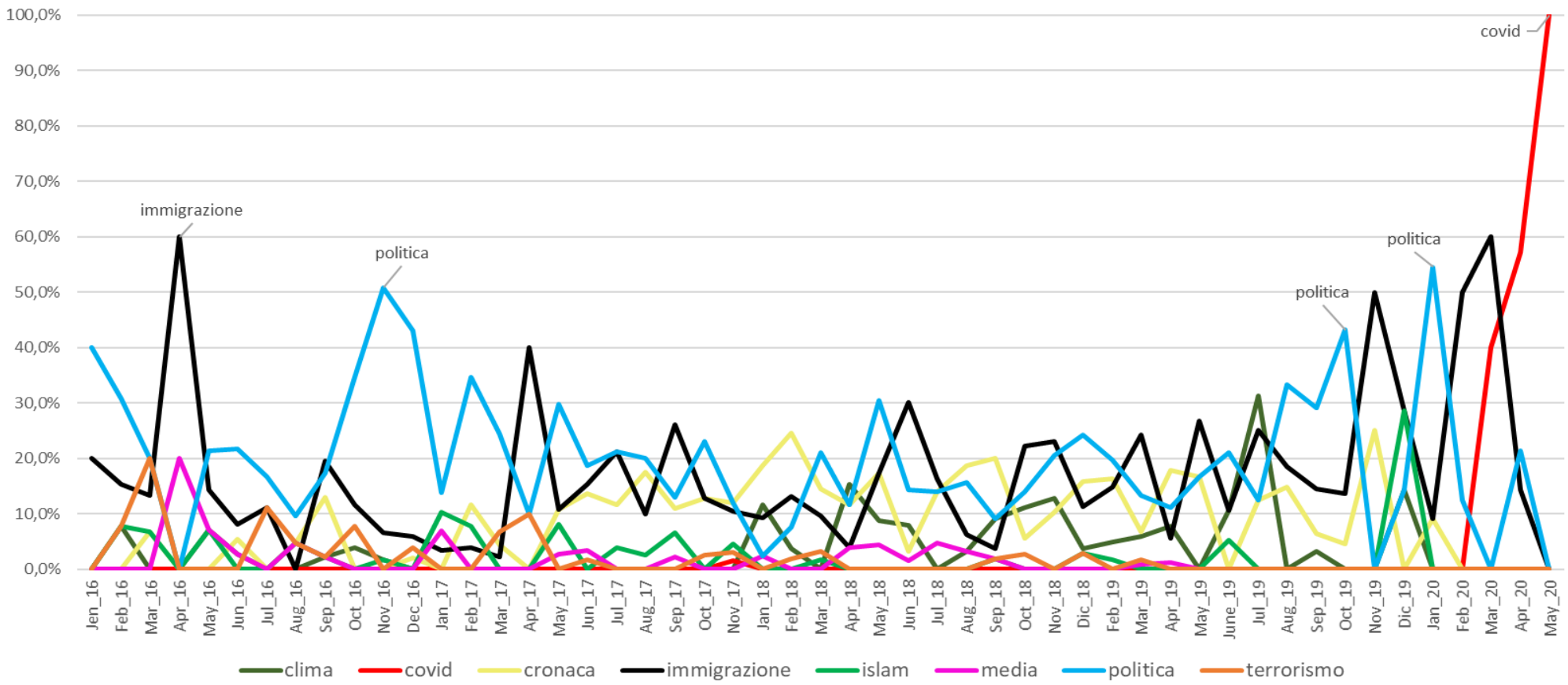


Grafico 8 - Andamento delle keyword nel tempo (% n= 1.877)



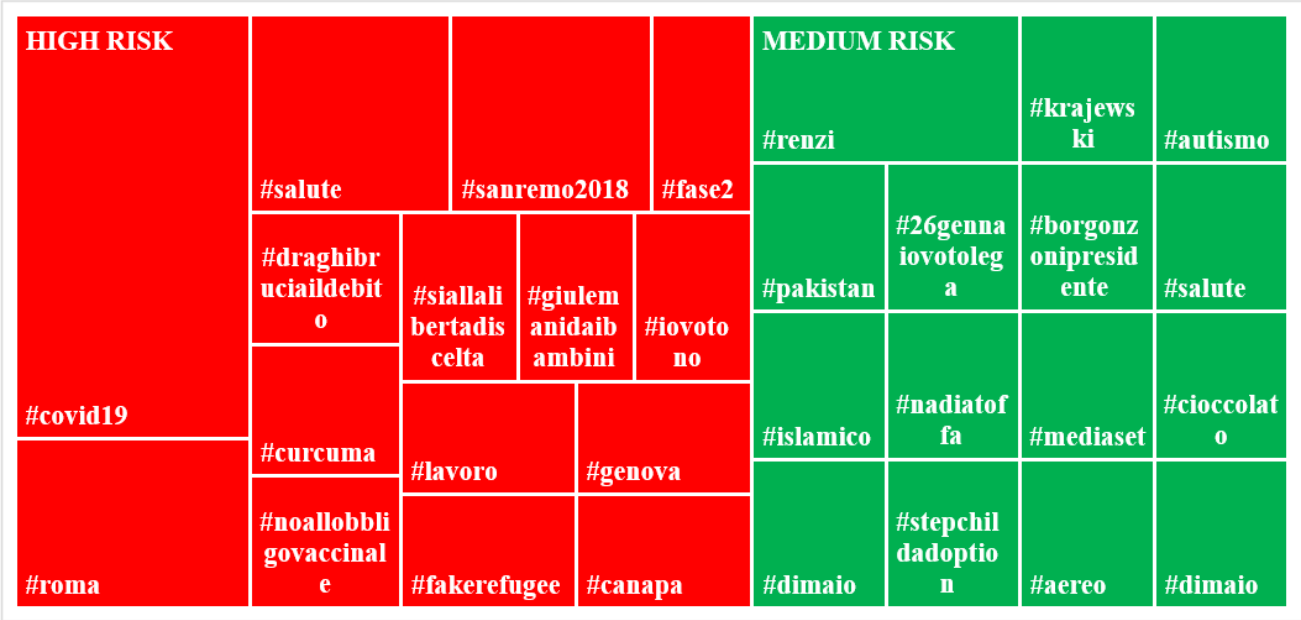
1 Quando la disinformazione incontra il populismo

L'analisi del messaggio disinformativo parte dall'esplorazione degli *hashtag*.

L'hashtag è un aggregatore tematico e quindi la sua funzione è quella di facilitare gli utenti nella ricerca di contenuti su un tema d'interesse. Tra le recenti funzionalità delle piattaforme di *social media* c'è la possibilità di seguire non solo i profili degli utenti ma anche di essere sempre aggiornati su specifici *hashtag*. Seguendo un *hashtag*, l'utente nel proprio *news feed* vedrà comparire tutti i contenuti dedicati a quel tema specifico a prescindere da chi li abbia postati. Questi contenuti vengono organizzati in una selezione di post disposti per rilevanza e attualità.

L'esperienza ci insegna che nel modo in cui formuliamo una richiesta di informazioni ad un algoritmo ci sono già gli orientamenti della risposta; per questo motivo è stato utile analizzare gli *hashtag* in base al rischio disinformativo che veicolano poiché questi funzionano sui *social media* nello stesso modo di una *query*. Nel grafico 9 sono rappresentati gli *hashtag* più utilizzati per far circolare contenuti a medio e alto rischio disinformativo: la grandezza della cella è proporzionale alla frequenza dell'*hashtag* in quel dato gruppo.

Grafico 9 - Hashtag più frequenti per rischio disinformativo.



Veicolano contenuti ad alto rischio disinformativo i più generici e frequenti *#Covid19, #salute, #roma #fase2, #canapa* e i più specifici *#noallobligovaccianle, #sìallalibertàdiscelta, #giùlemanidaibambini, #fakerefugee, #draghibruciaildebito*. Questo risultato non solo ci orienta in modo più dettagliato delle *keyword* sul tema trattato, ma ci informa anche sulla posizione espressa dai produttori di contenuti nel dibattito su tema stesso. Gli *hashtag* menzionati sono infatti espressione di una chiara posizione nel dibattito sul tema del vaccino, su quello dell'immigrazione e su quello dell'economia.
Assegnare un *tag* (o parola chiave) ad un contenuto testuale è certamente un sistema utile ma anche piuttosto semplicistico, in quanto manca di sfumature circa l'argomento trattato.

Un'espressione del chiaro orientamento di chi ha prodotto questi contenuti è emersa soprattutto attraverso la modellazione degli argomenti. LDA ha estratto diciotto *topic* latenti[53] dalla matrice termini per documenti[54]. Le sfere di significato a cui rimandano le parole associate ad ogni argomento hanno risentito positivamente dell'affondo ermeneutico effettuato durante la fase di etichettatura. Questo approccio congiunto ha fatto sì che le parole associate ad ogni *topic* rimandassero più facilmente al tema di riferimento anche quando non era immediatamente comprensibile.
Inoltre, appare chiaro quanto le *keyword*, da sole, non sarebbero state utili ad approfondire l'orientamento degli utenti sul tema trattato, come invece è molto chiaro dalle parole che compongono ogni *topic*.

[53] La scelta di *k* è stata arbitraria. Esistono vari metodi per valutare il modello con un numero ottimale di argomenti (Blei e Lafferty, 2009) ma allo stesso tempo non esiste una risposta "giusta" al numero di argomenti appropriati per un dato corpus (Grimmer e Stewart, 2013), così partendo dal numero di *keyword* già in mio possesso (64 totali, 18 delle quali con una frequenza ≥ 5) è stato eseguito il modello un numero definito di volte (K=10, K= 15, K=18, K= 20) scegliendo, infine, quello più performante (k=18). La funzione ha restituito un oggetto contenente i dettagli completi del modello: per ogni *topic* sono riportate le parole ad esso assegnate con la rispettiva probabilità misurata (β).
[54] La matrice documenti per termini (dtm) del *corpus* testuale contiene 1.620 documenti e 10.941 termini distinti con il 100% di *sparsity*. Il dato di *sparsity* è indicativo di una grande dispersione testuale, cioè nonostante la precedente diminuzione della dimensione del vocabolario effettuata in fase di *pre-processing*, ci sono ancora termini che appaiono raramente. Lo sparse è stato dunque ridotto di 0.992.

Tabella 2 - *Topic* generati attraverso LDA. Distribuzioni multinomiali composte dai 10 termini più probabili per ogni *topic*.

Topic 1 - Azione	
term	β
agire	0,0168102
vogliono	0,0116337
avanti	0,011211
attacco	0,0107618
pensate	0,0091408
andare	0,0082407
fare	0,0066479
prima	0,0058169
occhi	0,0058169
parole	0,0056469

Topic 2 - Ambiente e Clima	
term	β
sempre	0,012768829
clima	0,011157063
cielo	0,008898574
popolo	0,008540414
mondo	0,007954815
difendere	0,007280651
ambiente	0,007280651
proprio	0,006872758
pianeta	0,005662729
terra	0,005662729

Topic 3 - Europa	
term	β
europea	0,017125006
guerra	0,00966997
francia	0,008723322
stati	0,007930293
unione	0,007930293
europa	0,006344234
uniti	0,006344234
governo	0,005622348
macron	0,004758176
partiti	0,004758176
potere	0,004758176

Topic 4 - Lobby	
term	β
interessi	0,011018814
fatto	0,010231145
banche	0,00926041
renzi	0,007674939
partito	0,007627299
governo	0,00664999
fare	0,006142662
miliardi	0,005933207
movimento	0,005933207
posizione	0,005933207

Topic 5 - Accoglienza Migranti	
term	β
immigrati	0,0123634
storia	0,0086569
nord	0,007695
richiedenti	0,007695
vediamo	0,0076946
asilo	0,0076898
accoglienza	0,0067331
nazionale	0,0067331
cara	0,0057713
centri	0,0057713
questione	0,0057713
sardegna	0,0057713

Topic 6 - Complottismo	
term	β
dicono	0,06709855
shared	0,059252872
link	0,043450232
nessuno	0,023950581
cose	0,02281049
dirà	0,02143189
nocensura	0,02080154
antipolitica	0,013237344
video	0,012682598
complotto	0,01006515

Topic 7 - Elementi Chimici	
term	β
composition	0,016722408
aluminum	0,014121145
suitable	0,013006317
combustible	0,012263099
combustion	0,011519881
compositions	0,010776663
potassium	0,010776663
sodium	0,010033445
invention	0,008918618
substance	0,008547009

Topic 8 - Infromazione Alternativa	
term	β
rimani	0,032681364
informato	0,032681354
notizia	0,00787928
anni	0,007688676
foto	0,006767929
piace	0,006061898
seguici	0,005455708
italia	0,004783077
pagina	0,004602681
fare	0,004381919

Topic 9 - Antieuropeismo	
term	β
economia	0,0106078
sapere	0,0073858
basta	0,0065651
italia	0,0065651
mondo	0,0065651
senza	0,0055336
europa	0,0049239
italexit	0,0049239
dati	0,0049107
dire	0,0041032
grazie	0,0041032
immigrazione	0,0041032

Topic 10 - Informazione Mainstream	
term	β
informare	0,007612617
prima	0,007205934
presidente	0,005550192
oggi	0,005293863
italiani	0,005258173
Stato	0,005010205
follia	0,00456757
grillo	0,00456757
news	0,00456757
fonte	0,003806308
giornali	0,003806308
mentana	0,003806308

Topic 11 - Virus e Vaccini	
term	β
hiv	0,010801015
virus	0,009082528
montagnier	0,008308473
gallo	0,007477626
nobel	0,006646778
causa	0,005815931
milioni	0,005815929
mondo	0,005091058
falsi	0,004985084
mondiale	0,004985084

Topic 12 - Nazionalismo	
term	β
italiani	0,018327067
migranti	0,017868774
accordo	0,011182598
salvini	0,00763432
polizia	0,007453962
governo	0,006852903
italia	0,006494606
stato	0,005964052
fatto	0,005932213
paese	0,005218546
nazione	0,005218546

ultimi	0,0041032
vaccini	0,0041032

Topic 13 - Oppisizione alla sinistra	
term	β
boldrini	0,0068662
fascismo	0,0060676
piano	0,0060676
buonisti	0,0051977
sinistra	0,00508
batterio	0,004334
famiglia	0,004334
giorni	0,004334
quali	0,004334
sangue	0,004334
toscana	0,004334
italia	0,0118703

Topic 14 - Invito alla condivisione	
term	β
condividi	0,011720799
like	0,01001744
seguire	0,009962679
dicono	0,008281959
seguaci	0,008204559
italiani	0,007924312
opinione	0,007032479
clicca	0,006929238
pagina	0,006573967

Topic 15 - Giustizia	
term	β
reato	0,01944096
essere	0,016680949
art	0,01666368
omissione	0,01481216
condotta	0,014664033
legge	0,013579137
evento	0,01203488
quando	0,010312416
omissione	0,00833184
penale	0,00833184

Topic 16 - Politica Locale	
term	β
words	0,006053107
soldi	0,004904233
incredibile	0,004839375
napoli	0,004291236
città	0,004291236
rabbia	0,004291236
fatto	0,004290911
regione	0,004271773
essere	0,003805104
sicurezza	0,003678202
sindaco	0,003678202
governo	0,003649158

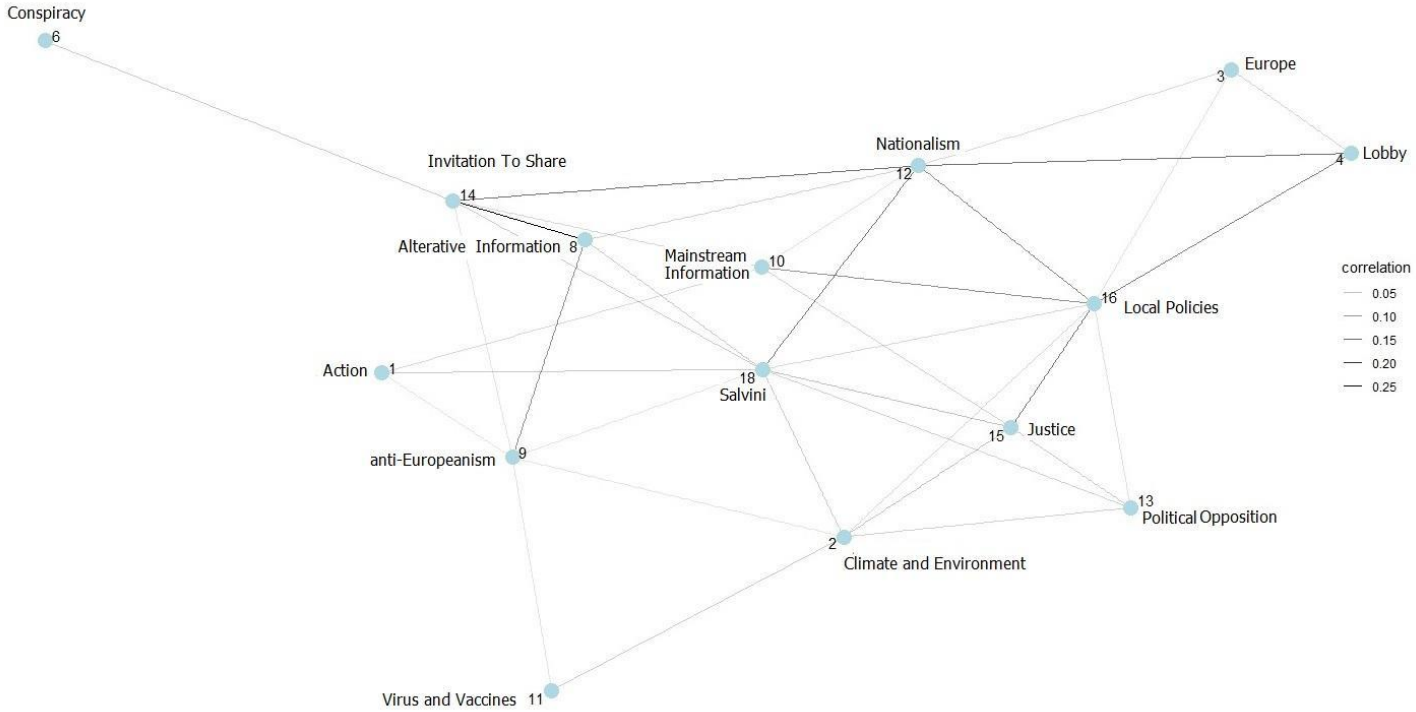
Topic 17 - Cronaca	
term	β
sedicenne	0,0290538
ucciso	0,0253918
arrestato	0,0156832
donna	0,0149364
moglie	0,0119491
uccidere	0,0112023
due	0,0097808
rapina	0,0097086
casa	0,0095611
uccisione	0,008215

Topic 18 - Salvini	
term	β
lega	0,008693837
italia	0,006821569
stato	0,006580054
fare	0,00581668
detto	0,005796935
salvini	0,004691075
persone	0,00436251
quando	0,00436251
parole	0,004261649
avere	0,003909185

Visualizzando i *topic* attraverso un grafo[55] di relazioni simmetriche, in cui ogni nodo rappresenta un *topic* e ogni linea rappresenta l'intensità della correlazione con gli altri, i legami più forti si evidenziano tra il tema che riguarda l'informazione alternativa (*topic* 8), l'invito alla condivisione (*topic* 14) il nazionalismo (*topic* 12), le *lobby* (*topic* 4), l'antieuropeismo (*topic* 9) e Salvini (*topic* 18).

[55] Per ottenere il grafo è stata creata una matrice di correlazioni con l'obiettivo di trovare le correlazioni tra *topic* sulla base delle parole che li caratterizzano e successivamente è stata tracciata una rete della matrice.

Grafico 10 - Rete di correlazione tra topic.



Ciò che emerge da queste correlazioni è quanto la disinformazione utilizzi un tipo di comunicazione che fa leva su narrazioni antagoniste al sistema.
Questa tensione antagonista è figlia del clima di pessimismo, sfiducia e malcontento verso la classe dirigente le cui radici sono proprio da rintracciare nella rottura generata dal postmodernismo che trova spazio nei movimenti populisti, abili a sfruttare proprio questo sentimento di malcontento e distacco tra le *élite* (politiche, finanziarie, dei media) e il popolo.

Il dibattito scientifico sul populismo è molto vivace ed eterogeneo così come la definizione del concetto lungi dall'essere definita[56]. Per sostenere l'interpretazione data è stata adottata la definizione di Jagers e Walgrave (2007) i quali definiscono il populismo come uno specifico stile di comunicazione politica adottata da attori - che possono essere *leader* di movimento, *leader* di partito, rappresentanti di gruppi di interesse, giornalisti, *ecc*. – che fa leva su particolari elementi che la identificano e la differenziano da qualsiasi altro tipo di comunicazione.

[56] Canovan1981; Wievorka 1993; Taguieff 1995, 1998; De Benoist 2000; Taggart 2000; Elchardus 2001; Meny & Surel 2000, 2002; fango 2004; Abs 2004

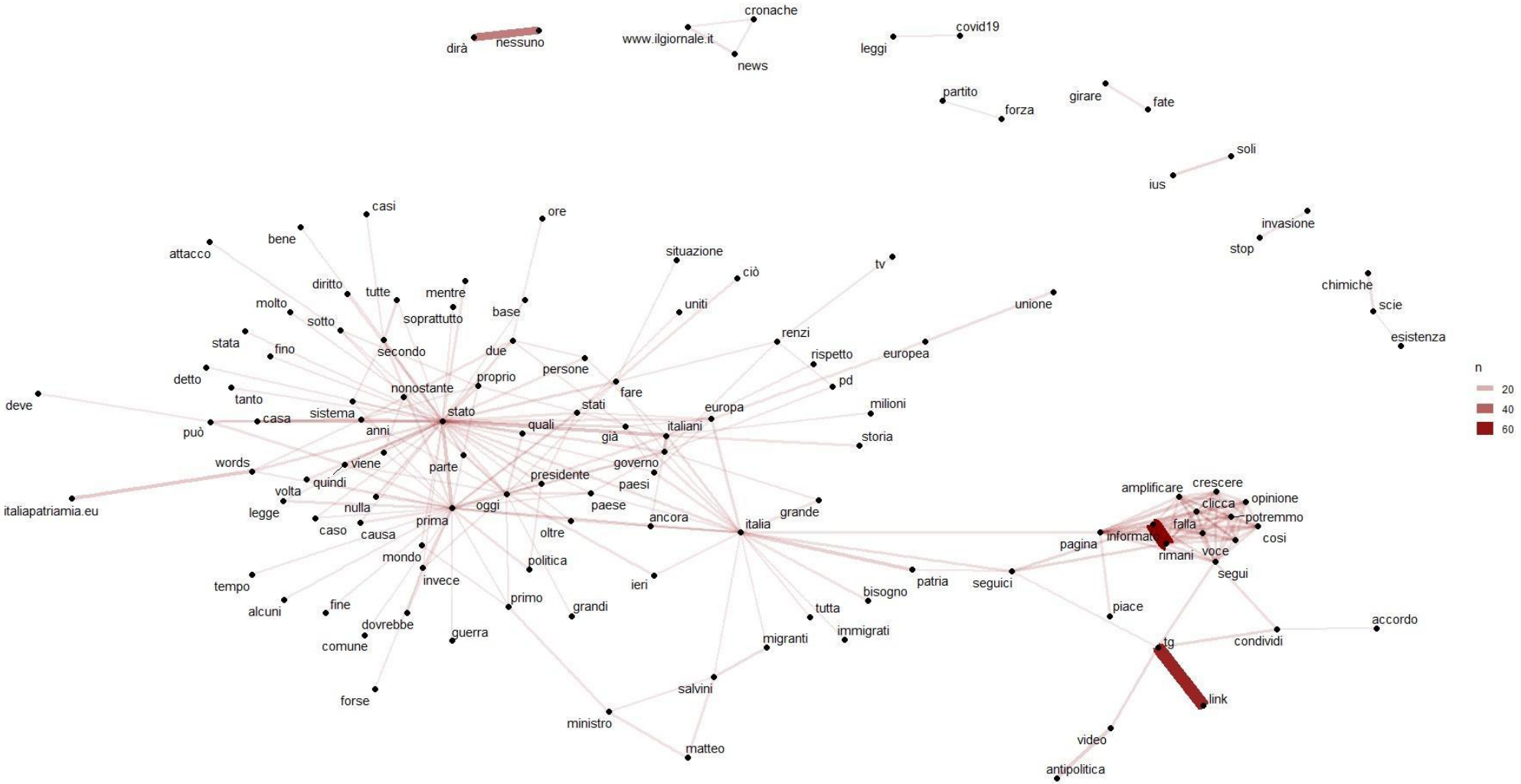
Gli elementi possono essere così riassunti:

- Il populismo si riferisce sempre al popolo e giustifica le sue azioni appellandosi e identificandosi con le persone; considera il popolo come un gruppo monolitico senza differenze interne ad eccezione di alcune categorie molto specifiche che sono soggetto a una strategia di esclusione (*out-group*).
- Il populismo è radicato in sentimenti anti-*élitari* che enfatizza con il rimarcare la distanza e lo straniamento tra il popolo e le *élite* che perseguono solo i propri interessi.
- Il populismo mobilita il supporto attraverso canali comunicativi.

Il tipo di comunicazione veicolata dai contenuti disinformativi sembra toccare tutti i quattro punti evidenziati da Jagers e Walgrave. L'elemento *anti-establishment* emerge non solo dalla relazione tra i *topic*, ma anche dall'insieme di parole a cui rimanda, ad esempio, l'avversione per l'Europa, di cui se ne marca l'ostilità verso i meccanismi politici, istituzionali e sanitari visti come espressione di una *lobby* volta a tenere il popolo assoggettato. In altre occasioni è evidente la costruzione di alcune argomentazioni volte ad affermare l'autorevolezza di certe posizioni, come quelle dei *no-vax* espresse nel *topic* "Virus e Vaccini": qui è rimarcata la dannosità dei vaccini attraverso le dichiarazioni di Montagnier, Nobel per la medicina e scopritore del virus dell'AIDS, recentemente salito sul carro degli antivaccinisti diffondendo l'idea della pericolosità della profilassi e altre posizioni complottiste riguardo la "dittatura sanitaria".

Finora sono state considerate le parole come singole unità e considerate le loro relazioni con i documenti. Tuttavia, molte analisi testuali interessanti si basano sul calcolo e la visualizzazione delle relazioni tra tutte le parole, utili ad approfondire maggiormente lo studio del *corpus* preso in esame. È stata così costruita una rete di parole che ricorrono insieme più spesso. L'intensità della loro connessione, data dall'intensità del colore, è basata sulla frequenza.

Grafico 11 - Rete di frequenza parole (frequenza n > 10).



La rete ci fornisce un prezioso e ulteriore momento interpretativo nell'analisi del messaggio.
La *clique* più ristretta della rete ha un significato ben preciso in quanto caratterizzata da lemmi come "clicca", "voce", "crescere", "amplificare", "rimani" e "informato", che sembrano essere un chiaro riferimento alle azioni di mobilitazione e persuasione messe in atto dalle pagine: invito a crescere in termini di comunità e invito a condividere i propri contenuti. Il nodo ponte "segui", collega la *clique* alle parole "tg", "*link*", "video" e "antipolitica" che rimandano ad un tipo di comunicazione dai toni antagonisti evidente anche dalla presenza della parola "antipolitica". I due nodi che aprono le reti rimandano alle due posizioni politiche: il nodo ponte "Italia" apre la rete verso una posizione nazionalista, dove il tema dell'immigrazione è trattato esclusivamente con posizioni politiche orientate e a destra, mentre il nodo ponte "Europa" apre lo "Stato" alla sinistra del PD e a Renzi.
L'atteggiamento complottista si distacca dalla rete ed è dato da una serie di collegamenti minori che rimandano a temi e posizioni ben precise: l'esistenza delle scie chimiche, l'invasione degli immigrati, l'esclusività delle notizie fornite da fonti alternative ("nessuno-dirà") e l'atteggiamento proselita ("fate-girare") che si differenzia dall'invito a condividere i contenuti data dal collegamento ("d'accordo-condividi"). Infine, il richiamo a fonti *mainstream* (ma politicamente molto orientate) dato dal collegamento tra il lemma "*news*" e il *link* "www.ilgiornale.it" si configura come nicchia separata probabilmente perché inserita in contesti che si differenziano dal resto, al contrario il *link* italiapatriamia.eu[57] (fonte non attendibile) è collegato alla rete dal nodo ponte "*words*".

Le informazioni problematiche si manifestano nella piattaforma sfruttando questioni specifiche e il populismo pare coinvolto nella diffusione della disinformazione più di quanto si credesse. Mentre la satira tenta di minare l'autocrazia, l'orientamento tematico emerso alimenta i pregiudizi e cerca di creare indignazione nella comunità, compromettendo la fiducia degli utenti verso le istituzioni e orientando l'opinione pubblica verso una particolare posizione.
In realtà la scuola di moscovita di *Media and Communication*[58] aveva già posto il problema della disinformazione nella cornice della comunicazione politica con l'assioma *"Just tell the people what they want to hear"* (Ilya Kiriya, 2019[59]). Lo studio del potere della disinformazione in Russia parte dalla riflessione sul ruolo dello Stato in epoca post-sovietica e la guida che esso assume nella rivoluzione dei *mass media*.

[57] La fase di preprocessing del testo ha eliminato l'elemento *https://* tipico del link ma non il link, e questo è stato utile per comprendere il ruolo dato alle diverse fonti nel messaggio veicolato.
[58] HSE University, una delle principali università della Federazione Russa alla quale chi scrive ha preso parte nel 2019 per intraprendere lo studio della disinformazione.
[59] Dagli appunti personali alle lezioni della Summer School "Fake News, Post-Truth and Digital Media: inquiry in relationship between media and politics", National Research University Higher School of Economics of Moscow.

In epoca post-sovietica l'intrusione dello Stato sui media agisce in diversi modi:

- nel campo dei contenuti, per garantire la promozione delle riforme statali;
- nel campo della proprietà dei media, per garantire il controllo dei contenuti;
- nel campo del finanziamento dei media, per risolvere i conflitti tra interessi commerciali e statali.

Negli ultimi due decenni la comunicazione politica in Russia è diventata sempre più cospirativa, un processo in cui i media statali hanno svolto un ruolo fondamentale inasprendo la regolamentazione legislativa per controllare i *social media*. Il controllo di Internet a livello strutturale funziona isolando le *echo chambers* dei discorsi di opposizione politica e creando allo stesso tempo una massiccia ondata di false informazioni e opinioni pro-statali (ivi). Le strategie politiche della produzione di disinformazione in Russia comprendono la creazione *ad hoc* di *account troll* e *bot* per distorcere la comunicazione nei *social media*. Lo Stato, dunque, non influenza solo il *medium* ma anche la forma della notizia, implementata in epoca post-sovietica con un *ethos* specifico (copertura multiforme, applicazione della conoscenza degli esperti, *ecc*.). La manipolazione dei contenuti da parte dello Stato comprende: notizie false, notizie ingannevoli, false perizie, false testimonianze, falsi esperti, falsi dati, falso pluralismo[60], più parti che hanno la stessa posizione o si raccomandano reciprocamente (ivi). Con l'avvento di Internet, le dinamiche del controllo ideologico guidate dallo Stato si sono riversate nella rete. Perché la Russia è *leader* nella produzione di *fake news*[61]? Perché le teorie del complotto - costruite principalmente per diffondere l'idea di una cooperazione occidentale volta a smantellare lo Stato Russo attraverso operazioni segrete - sono state attivamente e direttamente promosse dal Cremlino per un interesse politico interno. Sistematiche e legittime campagne di disinformazione sono portate avanti dalle autorità per i loro stessi interessi, premendo sulla contrapposizione noi/loro e ponendo una minaccia esterna alla nazione con l'obiettivo di generare una massiccia coesione interna.

La diversa evoluzione storica dei media italiani e il diverso ruolo ricoperto dallo Stato sono le motivazioni per le quali l'elemento politico non è stato inserito fin dalla fase di disegno della ricerca, ma che è prepotentemente emerso attraverso l'utilizzo dell'approccio metodologico proposto.

[60] Caratteristiche utilizzate nella costruzione delle categorie dello spettro disinformativo utilizzato in questa ricerca.
[61] Questa è la domanda che ha guidato lo studio della disinformazione durante la permanenza all'HSE University di Mosca.

2 L’anatomia della disinformazione

Ciò che preme a questo punto dell’analisi è far emergere le strategie utilizzate dai produttori di contenuti per veicolare il messaggio cospirativo-populista emerso dall’analisi testuale.

È possibile interpretare le strategie disinformative attraverso quattro dimensioni sintetiche di significato portate all’evidenza dall’analisi fattoriale[62]:

La prima dimensione, che nel grafico 12 è rappresentata sull’asse orizzontale (1° asse), si caratterizza per l’opposizione tra la produzione di contenuti interni ed esterni alla piattaforma. Il fattore è stato dunque denominato *Platform*.

La seconda dimensione, che nel grafico 12 è rappresentata sull’asse verticale (2° asse), prende forma a partire dal *continuum* di veridicità che caratterizza lo spettro disinformativo e si caratterizza per l’opposizione tra contenuti completamente falsi e contenuti ingannevoli. Il fattore è stato dunque denominato *Truth's Continuum.*

La terza dimensione, che nel grafico 13 è rappresentata sull’asse orizzontale (3° asse) si caratterizzata pienamente per l’opposizione tra l’alto e il basso rischio disinformativo. Il fattore è stato dunque denominato *Risk.*

Infine, la quarta dimensione, che nel grafico 13 è rappresentata sull’asse verticale (4° asse), prende forma dall’opposizione tra *keyword e topic* potenzialmente polarizzanti e quelli potenzialmente poco polarizzanti. Il fattore è stato dunque denominato *Theme’ Polarization.*

[62]Per impostare l’ACM si è deciso di usare come variabili attive quelle che forniscono informazioni sul rischio disinformativo, sul tipo di post e sull’argomento trattato (*topic* e *keyword*) e come variabili illustrative quelle che fanno riferimento agli elementi costitutivi del contenuto, all’*engagement* e il numero di *follower* del gruppo dal quale i contenuti sono diffusi.

Grafico 12 - Piano fattoriale derivante da ACM (assi 1-2)

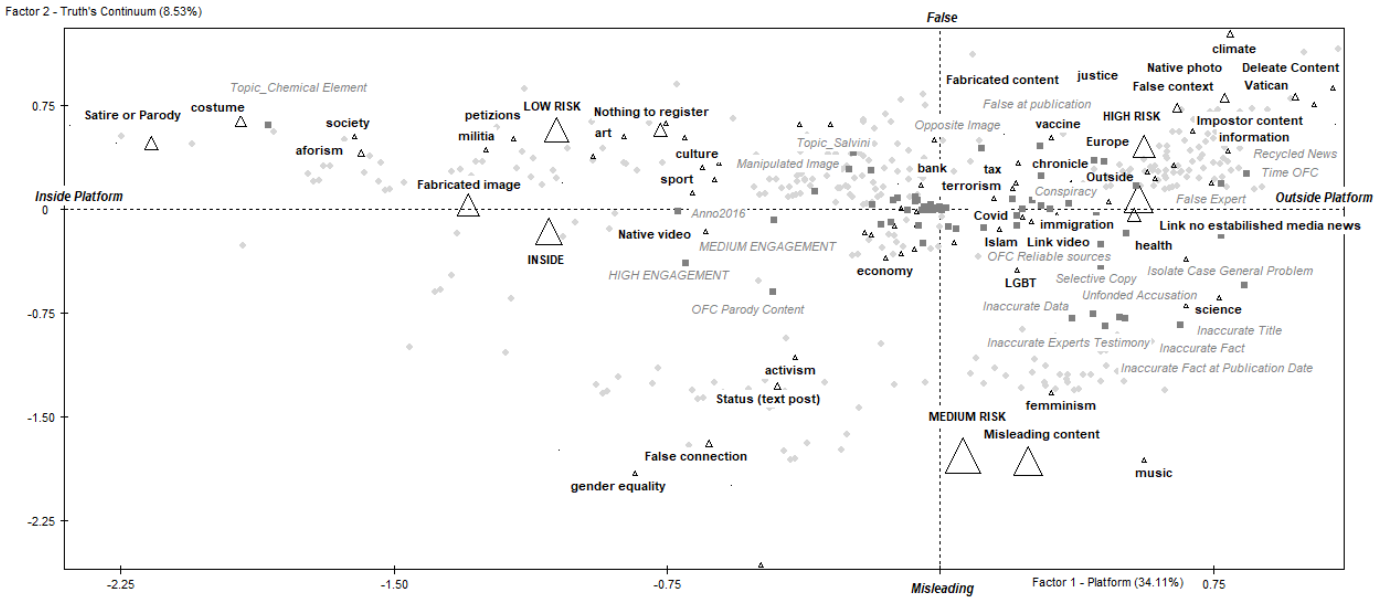
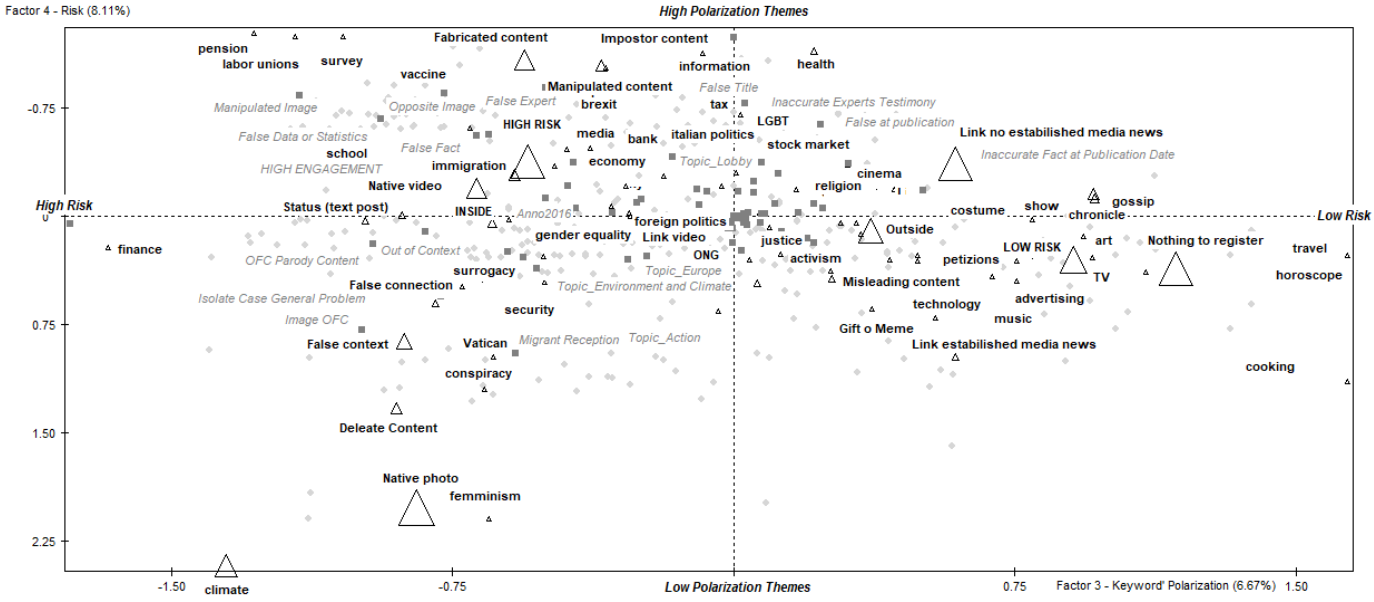


Grafico 13 - Piano fattoriale derivante da ACM (assi 3-4)



Dall'incrocio delle prime due dimensioni (grafico 12) appare chiaro quanto contenuti ad alto rischio disinformativo siano sbilanciati tutti verso l'esterno della piattaforma, mentre quelli prodotti all'interno si caratterizzano per un rischio disinformativo piuttosto basso. Gli spazi di significato prodotti dall'incrocio dell'asse terzo e quarto (figura 18) ci forniscono preziose informazioni sull'associazione tra rischio disinformativo e argomenti polarizzanti, ossia temi che riescono maggiormente a dividere l'opinione pubblica. Più gli argomenti sono divisivi più si prestano alla disinformazione, questi contenuti riportano false statistiche o testimonianze di esperti, immagini

manipolate o opposte, false accuse, contenuti fuori contesto (come luoghi fuori contesto, satira fuori contesto) e teorie del complotto. Temi come clima, femminismo, *gender equality*, giustizia, ONG, pur provenienti da testate informative riconosciute sono condivise utilizzando false connessioni o generalizzando casi isolati, dimostrando che una notizia può diventare potenzialmente disinformativa se condivisa sulle piattaforme in modo tale da essere ingannevole. Un inganno difficilmente riconoscibile se si considera la tipicità del consumo moderno dell'informazione basato sull'abitudine degli utenti a fermarsi al titolo senza leggere l'intero contenuto dell'articolo.
Temi come moda, spettacolo, TV e cucina, provenienti dall'esterno della piattaforma, hanno un basso rischio disinformativo perché appunto poco polarizzanti.

Sulla base dei fattori emersi dall'ACM, attraverso la *cluster analysis* si sono individuate tre tipi di strategie disinformative.

Tabella 3 - Strategie disinformative

Strategy	Post Type	Use of Contents	Contents Construction	Topic	Keyword
PRODUCTION OF FALSE (OUTSIDE) 35,62%	Link NO established media news	Impostor. Manipulated. Fabricated. False context. Deleted content.	False title. False Expert Testimony. Unfounded Accusation. Recycled News. Selective Copying. Out of context of reliable sources.	Conspiracy Theory. Mainstream Information. Political Opposition. Virus and Vaccine. Nationalism. Invite to share	Climate. Health
PRODUCTION OF FALSE (INSIDE) 13,79%	Status. Native video.	Fabricated. False Context. Manipulated. Impostor. Satire or Parody	Manipulated Image. False Data or Statistics. Inaccurate fact. Parody Out of Context	No Topic	Italian Politics
PRODUCTION OF DECEPTION 21,57%	Status. Native Photo.	Misleading; False Connection	Selective Copy, Out of Context	No Topic	Italian Politics
JUNK NEWS (INSIDE) 29%		Satire or Parody		No Topic	Costume. Aphorism.

- La produzione del falso all'esterno della piattaforma è la strategia più utilizzata (35,62%): questa comprende *news* provenienti da fonti d'informazione alternativa e cospirativa che si caratterizzano per contenuti ingannevoli o completamente fabbricati. Le *news* sono costruite riportando falsi titoli, false testimonianze di esperti, accuse infondate o informazioni fuori contesto e i temi di cui trattano risultano essere molto divisivi: troviamo infatti le teorie del complotto, l'informazione *mainstream*, virus e vaccini, il clima e il sentimento nazionalista.

Infine, caratteristica di questo tipo di messaggio è il continuo invito alla condivisione del contenuto promossa dai produttori.

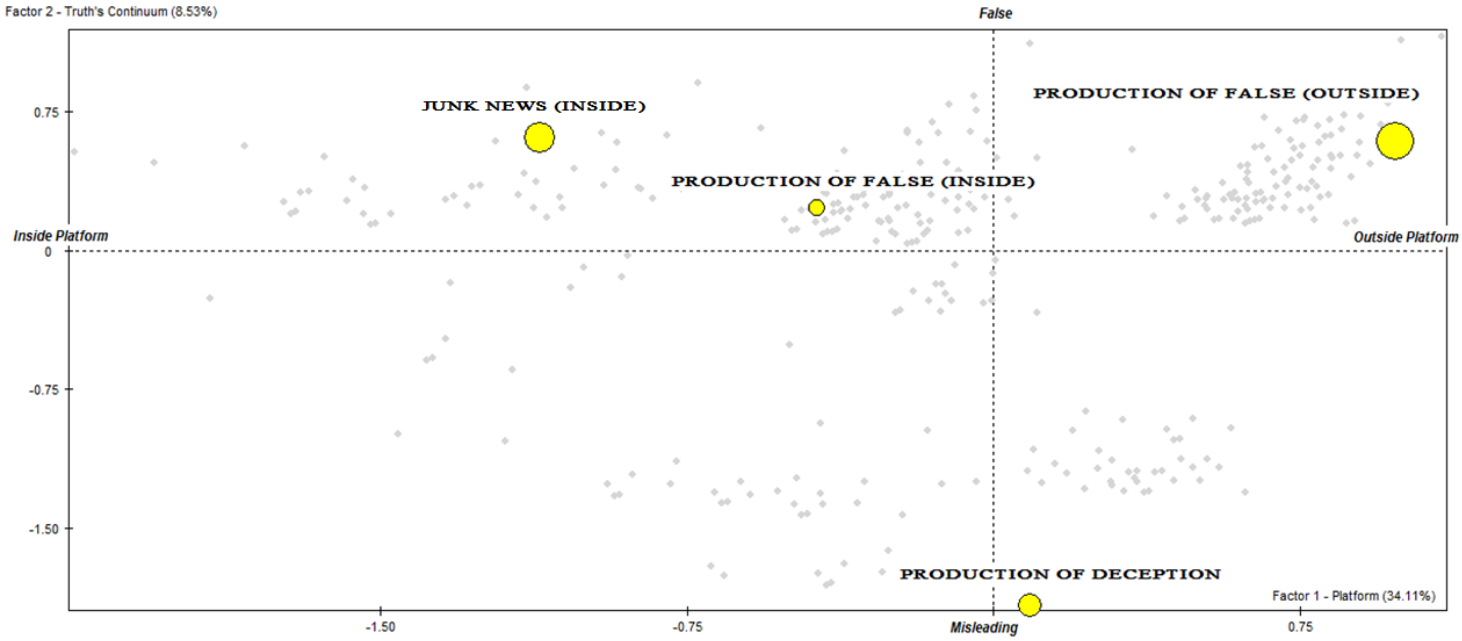
- La produzione del falso all'interno della piattaforma comprende il 13,79% dei contenuti. Sono soprattutto post testuali (*status*) o esclusivamente visivi (*native video*). La caratteristica principale di questa strategia è l'utilizzo della tecnica del "fuori contesto" messa a punto utilizzando immagini manipolate, falsi dati o statistiche fino ad arrivare all'utilizzo di contenuti parodistici e satirici fuorvianti. La manipolazione riguarda l'inquadramento del messaggio, che tratta soprattutto di politica italiana.

- La produzione dell'inganno è la seconda strategia più utilizzata sulla piattaforma *social* (21,57%) e si caratterizza soprattutto per l'utilizzo di copie selettive ossia materiale falso o fuorviante combinato con informazioni vere. La strategia del fuori contesto è utilizzata anche in questo tipo di contenuti e resta in primo piano l'argomento politico.

- Quando i contenuti non contengono informazioni false o disinformative, di cosa si tratta? Come suggerisce Venturini (2019), spostando l'attenzione dalla falsità alla strategia di diffusione, questi contenuti possono essere considerati "notizie spazzatura" (*Junk News*). Nel campione sono il 29% e il fatto che non costituiscano una vera e propria strategia disinformativa non se ne può sminuire l'influenza: le notizie spazzatura sono problematiche non perché false o fuorvianti, ma perché saturano il dibattito pubblico (Venturini, 2019) lasciando poco spazio ad altre discussioni e riducendo così la ricchezza del confronto, impedendo che vengano poste all'attenzione degli utenti storie più rilevanti.

Osservando i temi, quasi tutti i *topic* estratti riguardano la costruzione del falso proveniente dall'esterno della piattaforma; questo perché la condivisione dei *link* provenienti da siti di notizie non attendibili si arricchisce di una cerniera testuale che sottolinea la posizione di chi produce la notizia verso il tema condiviso, invitando la comunità ad uniformarsi. È inoltre interessante notare quanto un contenuto satirico inserito fuori dal contesto originario riesca a trasformarsi facilmente in una strategia disinformativa.

Grafico 14 - clusters proiettati su piano fattoriale derivante da ACM (assi 1-2)



Una categorizzazione così dettagliata sull'oggetto digitale ha permesso di dimostrare quanto la disinformazione più che nascere sulla piattaforma, ha reso la piattaforma un vero e proprio canale di diffusione dei contenuti disinformativi. Guardando le percentuali del campione (grafico 10): il 35% dei contenuti disinformativi proviene dall'interno contro il 65% provenienti dall'esterno, di questi il 72% sono ad alto rischio disinformativo contro il 28% di quelli provenienti dall'interno.

Grafico 15 - Produzione di disinformazione dentro e fuori la piattaforma (% n= 1.877)

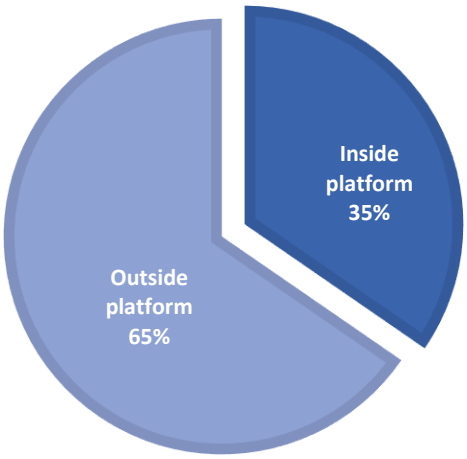
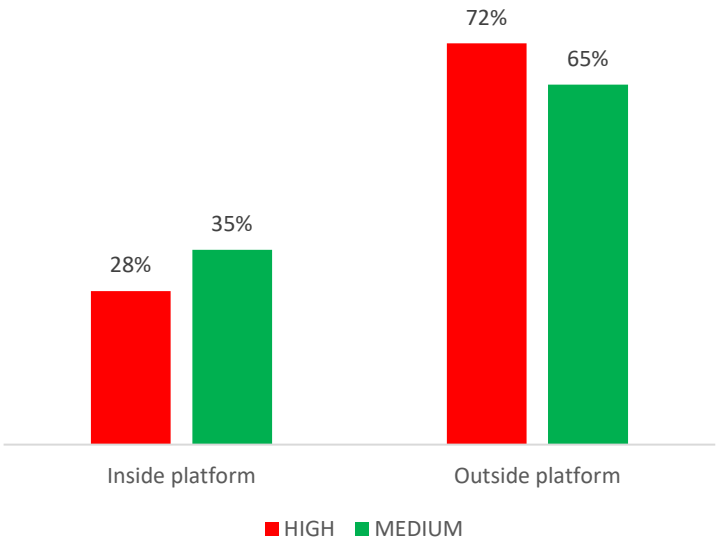
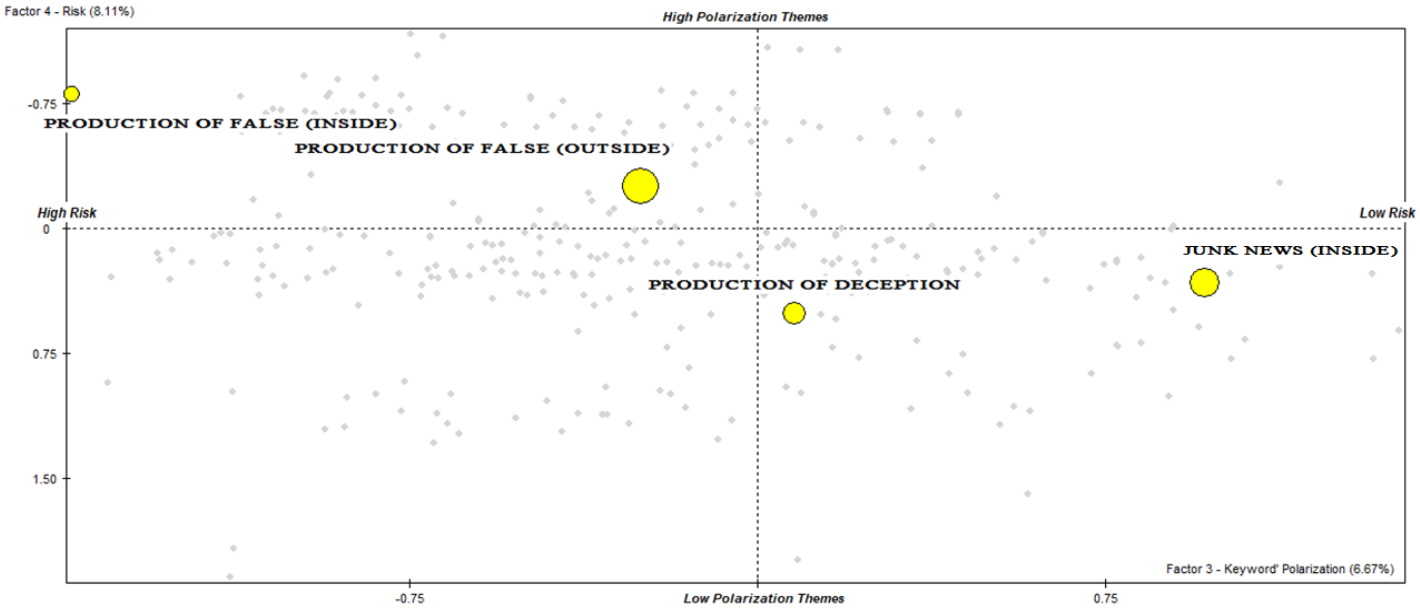


Grafico 16 - Rischio disinfomativo all'interno e all'esterno della piattaforma (% n= 1.877)



Facebook diventa così un vero e proprio *hub* di smistamento degli utenti verso altri spazi digitali, come blog o piattaforme video ricoprendo molte più responsabilità come facilitatore di questo ecosistema che come vero e proprio produttore.

Grafico 17 - Clusters proiettati su piano fattoriale derivante da ACM (assi 3-4)



Muovendosi lungo il confine sottile che divide la verità dalla menzogna, la produzione del falso, che sia interna o esterna alla piattaforma mostra una grande capacità empatica nei confronti della notizia e punta su argomenti particolarmente polarizzanti per l'opinione pubblica attraendo gli utenti nel momento in cui diventano *hot topic* del giorno.

Ciò che non emerge è l'idea, ampiamente diffusa, che il falso sia più virale del vero. Riflettendoci, come potrebbe il valore di verità di un'informazione determinare la sua velocità di fruizione e di diffusione? La viralità della disinformazione va piuttosto ricercata nel funzionamento della piattaforma che orienta le azioni degli utenti. Per questo motivo si è ritenuto necessario comprendere il più possibile il funzionamento della piattaforma Facebook, possibile soltanto attraverso una ricostruzione basata su un'indagine condotta dal Wall Street Journal titolata Facebook Files.

3. Come Facebook favorisce la disinformazione. Un'occhiata dentro la *black box* algoritmica

Tra la fase di estrazione dei dati e quella di analisi, alcuni contenuti sono stati soggetti a censura e dunque eliminati dalla piattaforma.

Dopo le polemiche relative al presunto ruolo dei *social media* nell'incrementare la diffusione di *fake news* in occasione delle elezioni presidenziali americane che hanno visto la vittoria di Trump, tra fine 2016 e inizio 2017, Facebook ha implementato un sistema di monitoraggio della disinformazione che utilizza in maniera combinata le segnalazioni degli utenti e l'attività di verificatori esterni. Questo procedimento prevede che i contenuti che non superano il *fact check*, siano pubblicamente contrassegnati come "contestati dai verificatori". In questo tipo di avvisi ci si è imbattuti anche durante la fase di etichettatura manuale dei post[63].

Figura 16 - Messaggio di segnalazione Facebook su immagine falsa

[63] Nonostante il post non fosse più disponibile per collocarlo lungo il *continuum* dello spettro, l'API mi ha restituito tutte le informazioni di cui avevo bisogno per analizzarli: cornice testuale, metriche di *engagement*, al tipo di post, *ecc*.

Figura 17 - Avviso segnalazione Facebook su video falso

Su questi contenuti è possibile fare *click* per comprendere perché sono stati contestati e qualora qualche utente desideri comunque condividerlo, la piattaforma ne sottolinea con un ulteriore avviso, la poca affidabilità (Jamieson & Solon, 2016).

Questo sistema ha delle criticità; prima tra tutte il fatto che l'apposizione di etichette simili rischia di generare un effetto opposto (come dimostrato dagli studi citati sugli effetti del *debunking*) incentivando gli utenti alla condivisione volontaria delle notizie etichettate come controverse, alimentando così una sorta di lotta contro la volontà di censura "dei potenti" (Levin, 2017).

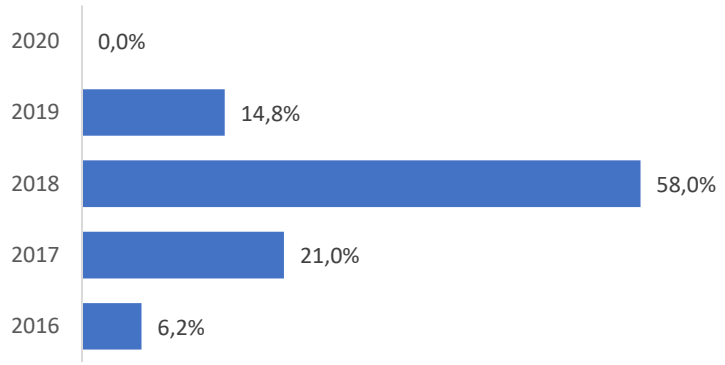
Un'articolata riflessione sugli strumenti utilizzabili per contrastare la disinformazione *on line* si rinviene anche nel documento/manifesto diffuso in rete da Mark Zuckerberg nel febbraio 2017 dal titolo *"Building Global Community*[64]*"*. In questo documento, il CEO di Facebook dichiara di esser consapevole del problema e di aver adottato misure per contrastarlo capaci di tener conto della linea sottile che intercorre tra bufale, satira e opinione. L'approccio adottato è quello utilizzato per combattere lo *spam*, dunque attraverso un massiccio uso di algoritmi. La piattaforma, come lo stesso Zuckerberg dichiara, si è concentrata meno sul divieto della disinformazione quanto sulla diffusione di prospettive e informazioni aggiuntive, inclusi gli avvisi dei *fact checker* sull'accuratezza di un elemento.

La distanza temporale tra la fase di estrazione e di analisi ha permesso di entrare in contatto con i contenuti eliminati dalla piattaforma e comprendere su quali di essi, il sistema messo a punto da Facebook, ha funzionato.

[64] Consultabile al link: www.facebook.com/notes/3707971095882612/

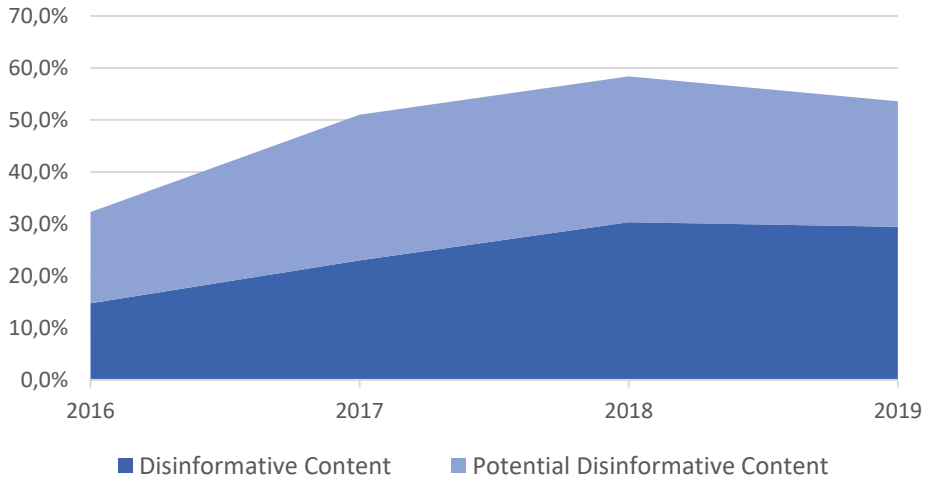
Considerando che su un campione di 1.877 post solo 75 di questi sono stati eliminati alla data di analisi, nelle pagine considerate una percentuale molto alta di censura si registra nel 2018.

Grafico 18 - Contenuti eliminate negli anni (% n=75)



Escludendo il 2020, di cui abbiamo traccia solo dei primi mesi dell'anno (e quindi la censura potrebbe non aver ancora operato al momento dell'analisi) il 2018 è stato anche l'anno in cui si registra il picco più alto di produzione di contenuti controversi.

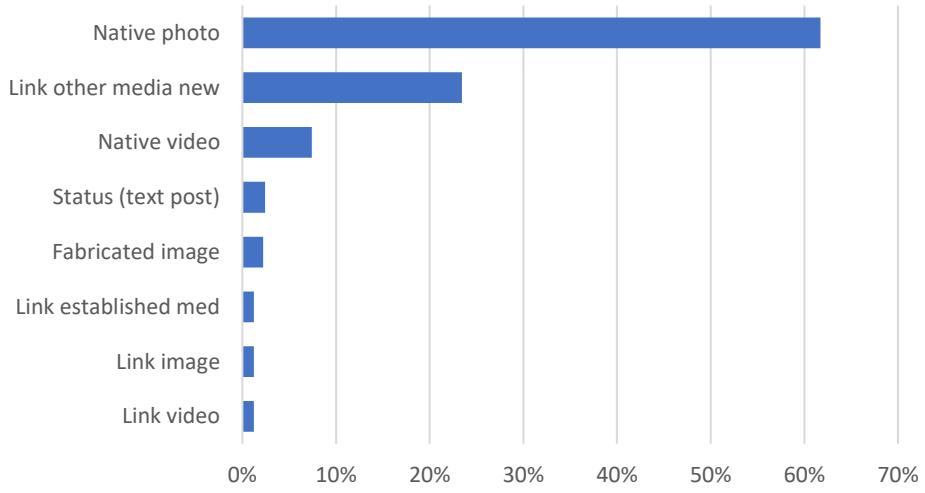
Grafico 19 - Produzione di disinformazione



Queste considerazioni preliminari mettono effettivamente in evidenza quanto le misure di contrasto alla disinformazione italiana adottate dalla piattaforma non abbiano effettivamente avuto successo in queste pagine. La censura ha operato soprattutto su foto native, *link* di articoli provenienti da testate non riconosciute e *native video*, lasciando circolare in maniera indisturbata le immagini fabbricate.

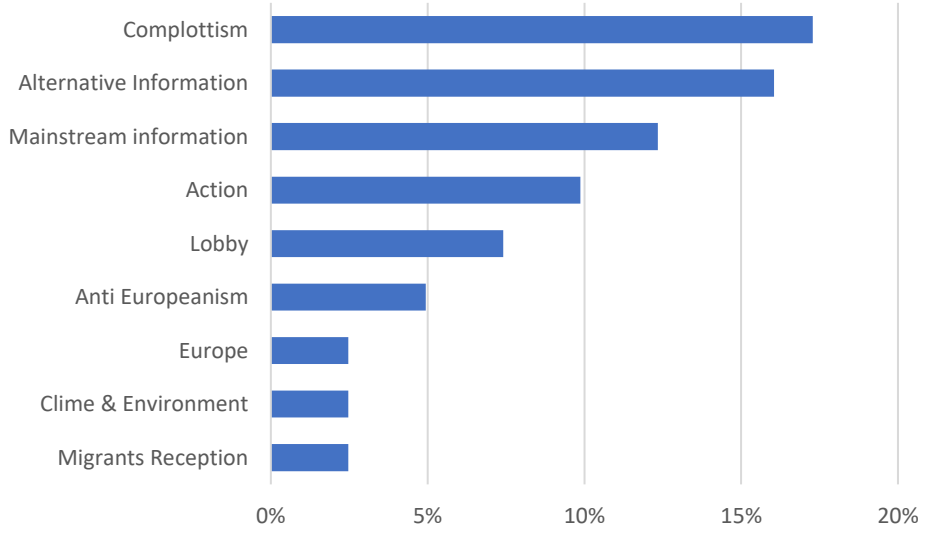
Queste saranno approfondite nel prossimo paragrafo e considerate come un tentativo della disinformazione di sopravvivere al (seppur blando) controllo.

Grafico 20 - Tipo di post contenuti eliminati (n=75)



Gli argomenti trattati dai post presi in esame riguardano in maniera prevalente il complottismo e l'informazione alternativa. Una spiegazione alla diffusione di questo tipo di argomento e alla produzione così massiccia di disinformazione dal 2017 in poi può esser spiegata dando uno sguardo ai cambiamenti apportati dalla società al proprio algoritmo di gestione contenuti.

Grafico 21 - *Topic* dei contenuti eliminati (n=75)



Le azioni di contrasto alla disinformazione da parte della piattaforma sono state fortemente messe in dubbio proprio dall'indagine del Wall Street Journal titolata Facebook Files (successivamente titolata Facebook Papers).

I Facebook Files sono basati su una revisione dei documenti interni della società - divulgati dalla *whistleblower*[65] ed *ex* dipendente Frances Haugen – come rapporti di ricerca, discussioni *online* tra dipendenti e relazioni su risultati di analisi forniti all'alta dirigenza. I documenti mostrano come i ricercatori di Facebook hanno rilevato e segnalato gli effetti negativi della piattaforma e di come il CEO Zuckerberg abbia contraddetto, minimizzato ed evitato di divulgare non solo l'impatto dei propri prodotti sugli utenti ma anche i tentativi fallimentari della direzione nel contenere la disinformazione, l'incitamento all'odio e la violenza sulla piattaforma, a volte per inefficienza tecnica a volte per non danneggiare i profitti che derivano dall'attività stessa degli utenti sulla piattaforma.

Questo gran numero di documenti interni ci offre una visione senza precedenti di come opera il gigante dei *social media* ed è l'unica fonte di informazione di cui si dispone per provare a comprendere come funziona l'algoritmo di selezione del *news feed* della piattaforma.

I documenti sono stati in gran parte scritti da un gruppo di ricercatori interni all'azienda il cui compito era indagare sui problemi della piattaforma e trovare soluzioni. Insieme, questi documenti mostrano quanto Facebook sia consapevole di causar danni (che non si limitano alla sola disinformazione ma anche alla responsabilità di Instagram sulla salute mentale degli adolescenti e alla responsabilità della piattaforma Facebook nel traffico di esseri umani) e di come la dirigenza in molti casi non può o non vuole affrontarli per non danneggiare la crescita aziendale.

L'algoritmo di Facebook è una vera e propria scatola nera: un insieme complesso di equazioni matematiche unite ad un calcolo misterioso che alla fine decide cosa vedremo quando accediamo al *social network*. Uno dei modi possibili per comprendere il funzionamento dell'algoritmo è quello di aprire la scatola nera che li governa e provare a scomporre il processo sul quale quegli algoritmi sono disegnati e usati.

Partendo dall'esame del loro complesso assemblaggio socio-tecnico - inteso come un sistema composto da diversi apparati di natura tecnica e sociale che sono inestricabilmente intrecciati e che vanno a definire la produzione dei dati – è possibile provare a comprendere gli attori coinvolti e provare a costruire le scelte, le negoziazioni e vincoli che vanno a contribuire alla formazione dell'algoritmo stesso (Aragona e Felaco, 2018). Dal punto di vista metodologico, esaminare gli algoritmi, senza considerare il contesto più ampio in cui vengono utilizzati, rischia di "feticizzarli" e quindi di tralasciare alcuni aspetti cruciali che ne determinano il loro funzionamento (Thomas, Nafus,

[65] Il *whistleblower* è un dipendente che segnala alle autorità situazioni sospette all'interno della propria azienda.

Sherman, 2018). La destrutturazione dell'assemblaggio socio-tecnico degli algoritmi può rappresentare una valida strategia per cogliere la loro natura contingente, relazionale e contestuale.

Attraverso l'inchiesta del Wall Street Journal è stato dunque possibile ricostruire e comprendere (seppur in maniera molto superficiale) il funzionamento dell'algoritmo di *news feed* di Facebook, del modo in cui viene ponderato e il ruolo che assume nella produzione di disinformazione.
All'inizio del 2018 Facebook ha dichiarato che stava apportando un grande cambiamento al *social network* con l'ambizione di connettere sempre di più le persone e di rendere la piattaforma un luogo d'interazione più sano. I documenti interni diffusi dal Wall Street Journal rivelano però una storia completamente diversa.
L'algoritmo di Facebook cambia continuamente, ma l'ultimo cambiamento è stato un vero e proprio cambio di paradigma più che un semplice ritocco. Questa modifica è nata dall'esigenza di far fronte ad un anno difficile per la società, il 2017, durante il quale la piattaforma ha vissuto un mutamento angosciante nel comportamento degli utenti: le metriche di coinvolgimento erano diminuite in modo allarmante durante tutto l'anno; sebbene la quantità di tempo che gli utenti trascorrevano su Facebook non stesse diminuendo, coloro che utilizzavano la piattaforma non erano più coinvolti. Questo perché il vecchio algoritmo di *news feed* era pesantemente orientato verso i video e la promozione di contenuti professionali (come *news* di giornali, contenuti d'approfondimento, promozione aziendale, *ecc*). Questo stava trasformando le persone in utenti passivi e distanziati che fruivano dei contenuti senza condividerli, senza produrne di nuovi e in generale senza interagire. A lungo andare questo effetto avrebbe annoiato gli utenti e li avrebbe portati ad abbandonare la piattaforma. Una volta che i *data scientist* hanno rilevato questo problema di coinvolgimento, la società ha lavorato ad una soluzione per convincere gli utenti a pubblicare, commentare e interagire di più durante il loro tempo trascorso su Facebook.
Per questa soluzione era necessario attuare un cambiamento radicale in ciò che gli utenti vedevano nei loro *feed* di notizie; in altre parole, l'algoritmo sarebbe stato modificato in modo che gli utenti potessero vedere più post delle persone a cui sono collegati e meno contenuti editoriali. La modifica dell'algoritmo è stata incentrata su una formula che l'azienda ha chiamato *"Meaningful Social Interactions"* (da ora MSI) (Hagey and Horwitz, 2021). Facebook ha utilizzato il concetto di MSI per creare un sistema di punteggio che prevede che quando i *like*, i commenti e le condivisioni provengono da persone vicine, il punteggio MSI è maggiore. Questo punteggio è dato dalla somma della misurazione dell'interazione di un post (come commenti, *like, share* ed *emoticon*) e la misurazione di quanto sono vicine le persone che stanno interagendo. Rileva quindi sia le interazioni che la vicinanza delle persone che interagiscono. L'obiettivo iniziale era semplicemente quello di

ottenere più punteggio possibile basandosi sull'idea che è più probabile che un utente pubblichi qualcosa se ha maggiori probabilità di ricevere un commento o un *like* da un collegamento a lui vicino.

I documenti esaminati vanno più nel dettaglio e scompongono la formula MSI fornendo uno sguardo raro sul funzionamento interno dell'algoritmo. Nella sua idea la formula è piuttosto semplice: un mi piace vale un punto, una ricondivisione o una reazione vale cinque punti, un commento significativo 30 punti; a questo si aggiunge o sottrae il punteggio in base a quanto queste interazioni siano vicine alle persone che commentano o condividono, quindi, se è un amico, con tante connessioni in comune o un estraneo, con poche connessioni in comune.

Dal punto di vista dell'azienda, la modifica dell'algoritmo ha funzionato: MSI ha neutralizzato il calo dei commenti, aumentato le condivisioni e in generale il coinvolgimento. MSI è stato abile soprattutto a mostrare alle persone quei contenuti che avrebbero suscitato in loro un sentimento tale da indurli a commentare o condividere. Dal punto di vista dei ricercatori interni però, la modifica all'algoritmo ha prodotto parecchie preoccupazioni, legate soprattutto alla tendenza del sistema di *feed* a promuovere contenuti controversi come notizie false, contenuti divisivi, offensivi e razzisti. Uno dei motivi per cui ciò è accaduto è perché MSI è stato costruito attorno a due aspetti: l'ottimizzazione per la vicinanza e per il coinvolgimento. Spesso però il coinvolgimento ha avuto la meglio.

Per comprendere il funzionamento riportiamo un esempio: supponiamo che abbiamo due amici, con il primo siamo buoni amici, parliamo spesso e abbiamo circa 500 amici in comune, ma questo amico pubblica solo noiosi aggiornamenti sul consiglio scolastico o informazioni sulla raccolta dei rifiuti nel quartiere; al contrario, una persona a caso della nostra rete con cui parliamo a malapena e con la quale abbiamo forse altre dieci connessioni in comune, diffonde continuamente contenuti controversi ma che raccolgono molti commenti, come complotti sui vaccini, Covid, immigrati e scie chimiche. Anche se con il primo amico la vicinanza alla rete è forte, molto probabilmente nel nostro *feed* vedremo molte più cose del secondo amico perché ciò che sta pubblicando è molto divisivo e crea un massiccio coinvolgimento.

I documenti spiegano questo effetto: un elemento della formula di MSI prevede quanto qualcosa può diventare virale e quindi spinge attivamente quel contenuto a più utenti. Questo elemento è stato chiamato *"Downstream MSI"* (ivi) e un modo davvero discutibile di filtrare ciò che vediamo poiché il suo compito è prevedere quale contenuto è più probabile che venga ricondiviso o commentato ripetutamente. Data questa previsione, l'algoritmo organizzerà i contenuti da mostrare nel *news feed* di più utenti possibili.

Dal risultato di questo funzionamento viene fuori che ciò che riceve più commenti sono contenuti controversi, cospirativi e altamente divisivi che provocano rabbia politica.

I singoli utenti potrebbero non aver mai notato questi cambiamenti, ma le aziende i cui profitti dipendono dalla visibilità che hanno su Facebook ne hanno subito registrato gli effetti. Un settore che ha prestato particolare attenzione è stato proprio il mondo dell'informazione che ha segnalato quanto questa modifica stava incentivando particolari editori (come Buzzfeed[66]) a creare maggior contenuti dal taglio divisivo (ivi).

Dall'idea che lo scorrimento passivo sulla piattaforma fosse negativo è nato un meccanismo volto teoricamente ad incoraggiare interazioni sociali significative ma che praticamente sta spingendo disinformazione e contenuti tossici tali da portare le persone a litigare e quindi produrre più *engagement*. I documenti mostrano quanto dopo la modifica dell'algoritmo, i contenuti divisivi sono aumentati significativamente: post cospirativi, estremi e disinformativi diventano i contenuti più virali sulla piattaforma. Inoltre, i documenti mostrano che più volte un contenuto è condiviso più è probabile che sia un contenuto discutibile: ad esempio, se un contenuto è stato ricondiviso venti volte di seguito, sarà dieci volte più probabile che contenga nudità, violenza, incitamento all'odio e disinformazione, rispetto ad un altro che non è stato ricondiviso affatto.

I documenti mostrano che i ricercatori hanno proposto soluzioni che potrebbero potenzialmente impedire a così tanti contenuti negativi di diventare virali. Un'idea è stata quella di eliminare proprio il *Downstream MSI*, la parte della nuova formula che fa previsioni su quali contenuti hanno più probabilità di diventare virali e li mostra a più utenti. Secondo i *data scientist*, se Facebook avesse richiamato quella parte della formula avrebbe ridotto drasticamente la diffusione della disinformazione. L'azienda ha accettato di farlo nel 2020 solo per alcuni argomenti delicati come l'informazione sanitaria. Nonostante l'impegno dei ricercatori di Facebook a trovare una soluzione che riducesse i contenuti dannosi e al tempo stesso non impedisse alle persone di esprimere le proprie opinioni, i documenti mostrano che Zuckerberg ha deciso di non adottare nessuna delle soluzioni che gli sono state proposte perché non adatte alle metriche degli utenti, alle metriche di crescita *standard* e quindi in generale al *business* della piattaforma (ivi). Oggi la modifica dell'algoritmo avvenuta nel 2018 rimane in gran parte intatta.

Facebook ha più di tre miliardi di utenti pari a più di un terzo della popolazione mondiale e man mano che la piattaforma diventa sempre più radicata nelle nostre vite, le decisioni che prende e le priorità che si dà riguardano tutti, anche chi non utilizza la piattaforma perché, come mostrano i documenti, ciò che accade sulla piattaforma ha significative ripercussioni nel mondo al di fuori della piattaforma.

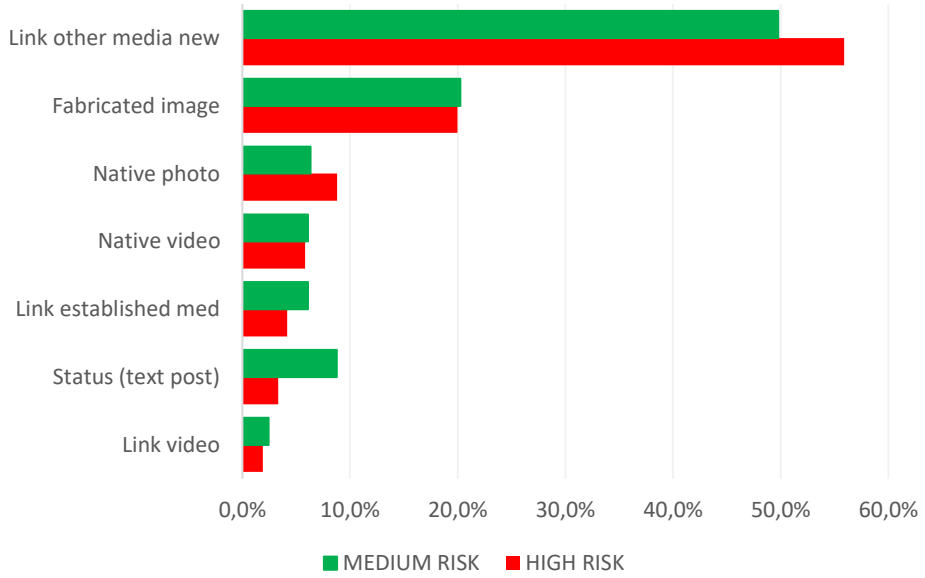
[66] Nell'autunno del 2018, Jonah Peretti, amministratore delegato dell'editore *online* BuzzFeed, ha inviato un'e-mail a un alto funzionario di Facebook sottolineando quanto il cambiamento apportato alla piattaforma stava rendendo virali contenuti divisivi incentivando la redazione a produrne di più. Il contenuto in questione è un post di BuzzFeed intitolato "21 cose che quasi tutti i bianchi sono colpevoli di dire" (disponibile al link: https://www.buzzfeed.com/michellerennex/guys-dont-attack-me-pls) che ha ricevuto 13.000 condivisioni e 16.000 commenti su Facebook di persone che litigavano tra loro sulla razza.

4. Il *bug* della censura: le immagini fabbricate.

I *bug* di controllo della disinformazione da parte della piattaforma sono sapientemente sfruttati dai produttori di contenuti disinformativi. Come descritto nei capitoli precedenti, il tipo di post più condiviso nell'ecosistema disinformativo italiano di Facebook sono le immagini (circa 45%), scomposte per l'analisi in "foto native" e "immagini fabbricate" per differenziare quelle immagini sulle quali sono stati aggiunti elementi non originariamente parte di esse.
Le immagini fabbricate sono il 25% a fronte del 10% delle foto e, dopo i *link* a testate informative non riconosciute, sono i contenuti con più alto rischio disinformativo.

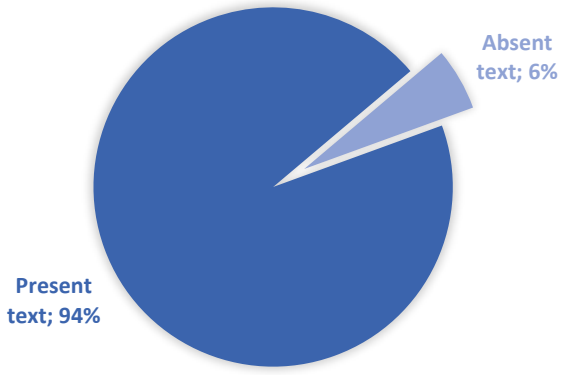
Nonostante ciò, sono state quelle meno soggette ad azioni di censura.

Grafico 22 - Rischio disinformativo per tipo di post (% n= 1.877)



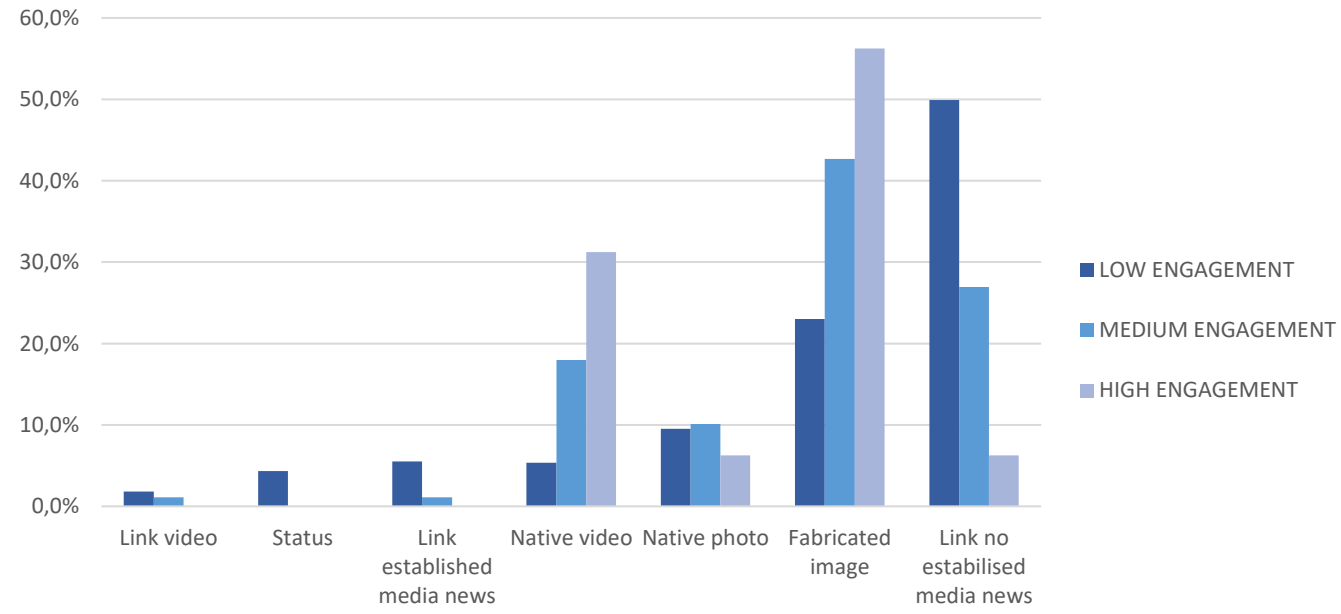
Il 94% delle immagini ad alto e medio rischio disinformativo contengono elementi testuali: un messaggio capace di eludere con più facilità le pratiche di censura degli algoritmi utilizzati dalla piattaforma sull'elemento testuale.

Grafico 23 - Presenza di testo sulle immagini



Il grafico 24 mostra inoltre quanto questo tipo di contenuti abbiano un alto livello di attrazione rispetto agli altri.

Grafico 24 - Tipo di post per engagement (%)



In generale, si registra una maggior viralità dei contenuti visuali rispetto a quelli testuali mentre le *news* provenienti da siti d'informazione riconosciuti registrano un *engagement* molto basso rispetto a quelle provenienti da siti non riconosciuti; come se il riscontro con le fonti ufficiali fosse stato quasi completamente soppiantato dal consenso della rete come parametro di valutazione dell'informazione condivisa. Le direzioni che prendono i contenuti disinformativi dipendono tanto dalla piattaforma quanto da un fattore tutto umano (come emergerà nel prossimo capitolo) che non può essere

tralasciato: la tendenza degli utenti chiudersi dentro narrative condivise, visioni e modi di raccontare le cose che sono aderenti al proprio sistema di credenze, sminuendo o rifiutando ciò che è divergente o dissonante (Festinger, 1957).

Sono dunque lontani i tempi in cui Facebook era popolato quasi esclusivamente da testi che esprimevano gli stati d'animo degli utenti: i contenuti visuali si stanno facendo sempre più largo sulle piattaforme, così tanto che ormai immagini e video la fanno da padroni relegando il testo a una semplice cerniera introduttiva. Per questo motivo si è ritenuto necessario approfondire l'analisi delle strategie disinformative esclusivamente visive attraverso una scheda di analisi del contenuto per immagini, dalla quale emerge chiaramente quanto queste siano un valido supporto alla produzione di disinformazione più difficile da censurare.

Le immagini fabbricate ad alto e medio rischio disinformativo (per un totale di 251 casi) sono state analizzate attraverso una tradizionale scheda di analisi del contenuto per le immagini[67]. Alle informazioni già raccolte nella precedente fase ermeneutica, l'obiettivo è stato quello di combinare informazioni di tipo visivo come il tipo rappresentazione, i personaggi ritratti, l'uso che se ne fa dell'immagine, lo stile e il tono della comunicazione testuale sull'immagine e infine l'obiettivo di tale comunicazione. Inoltre, sulla base delle evidenze emerse dall'analisi sul messaggio veicolato, è stata inserita la variabile della posizione politica espressa. Il testo presente sulle immagini è stato trascritto e sottoposto ad un'analisi delle corrispondenze lessicali[68] che ha permesso di individuare gli argomenti principali in modo semi-automatico assumendo la frequenza delle parole come indicatore della rilevanza di ciascun tema e integrando i temi con le variabili individuate.

Da questo approfondimento emerge che l'espediente del testo è soprattutto politico (57,2%) e la posizione politica maggiormente espressa è prevalentemente orientata verso destra (49,6%).

[67] In appendice

[68] L'Analisi delle Corrispondenze Lessicali (da ora ACL) è una tecnica di analisi dei dati per variabili categoriali elaborata nell'ambito dell'approccio *Analyse des Données* dalla scuola francese di J.P. Benzécri (1973) all'inizio degli anni Settanta. Trattandosi di un procedimento di tipo fattoriale, attraverso l'ACL è possibile individuare dimensioni sottese ai dati, in grado di sintetizzare le molteplici relazioni tra le variabili originarie e le parole presenti nel corpus.

Grafico 26 - Espediente del testo sull'immagine (% n=251)

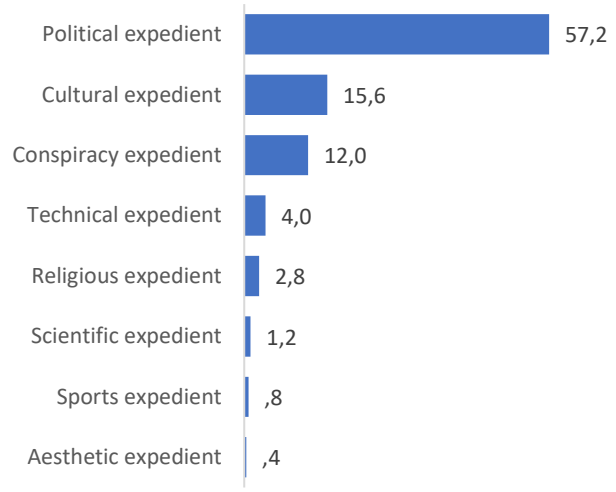
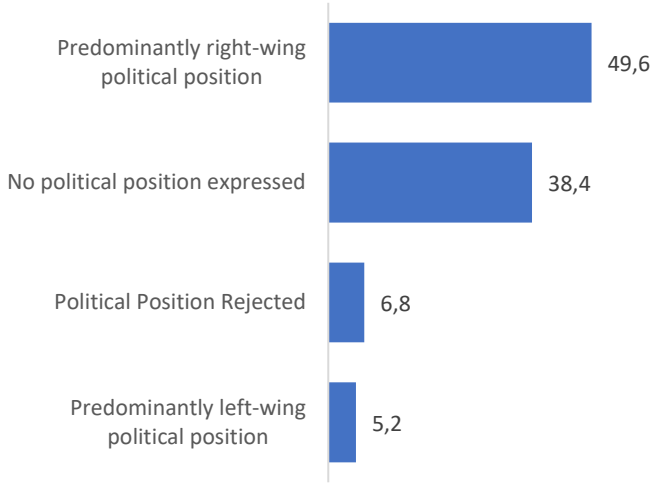


Grafico 25 - Posizione politica espressa dall'immagine (% n=251)



Questa posizione politica si caratterizza per un uso promozionale (60%) e strumentale (57%) dell'immagine e per un tono della comunicazione tendenzialmente aggressivo (70%).

Grafico 27 - Uso dell'immagine (% n=251)

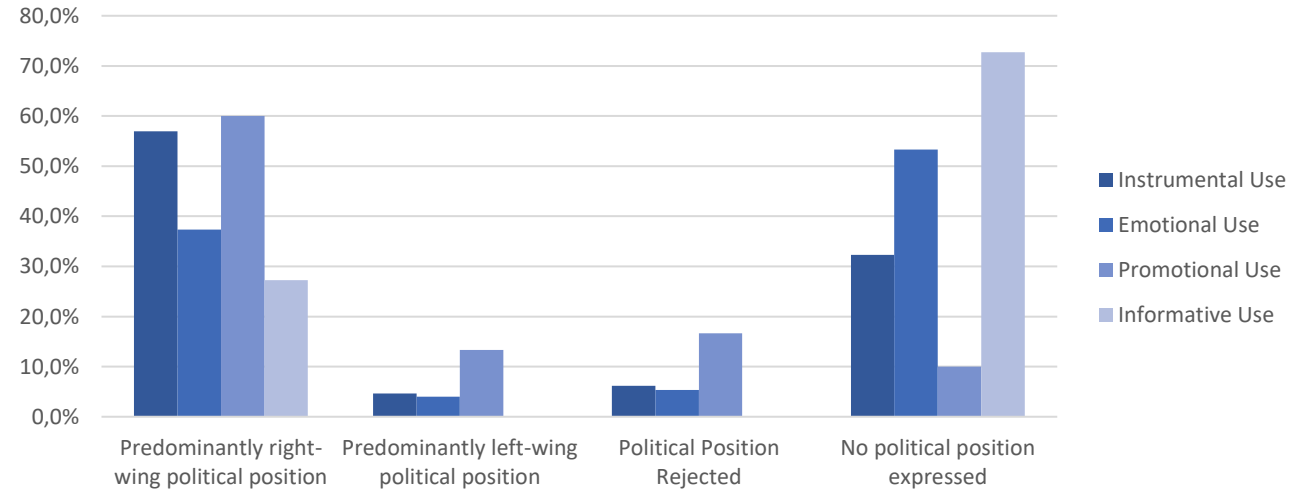
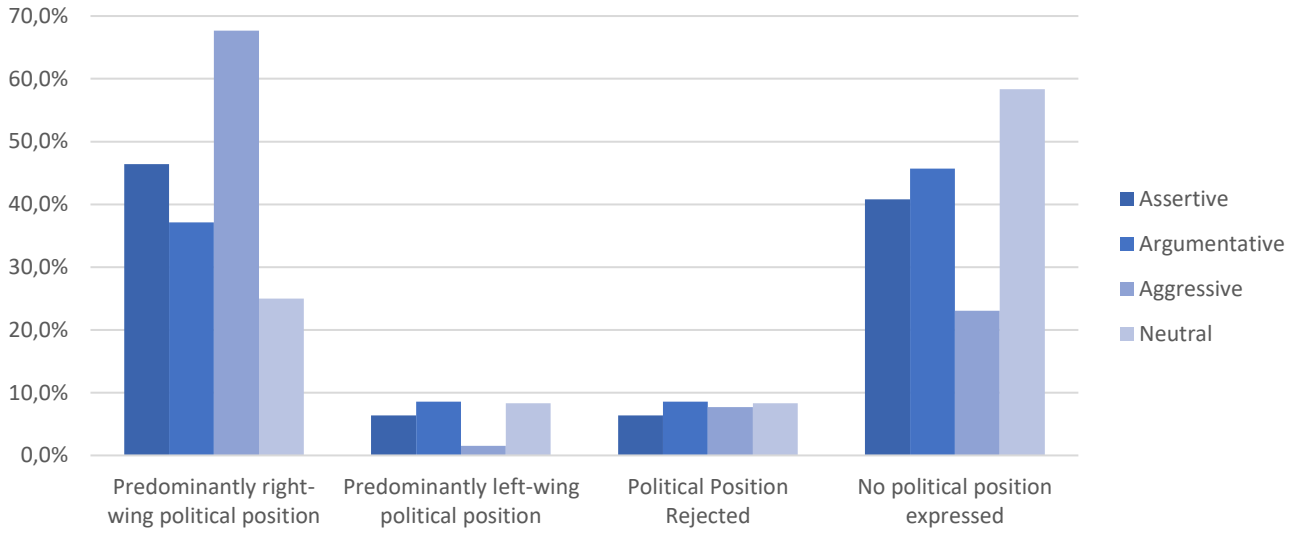


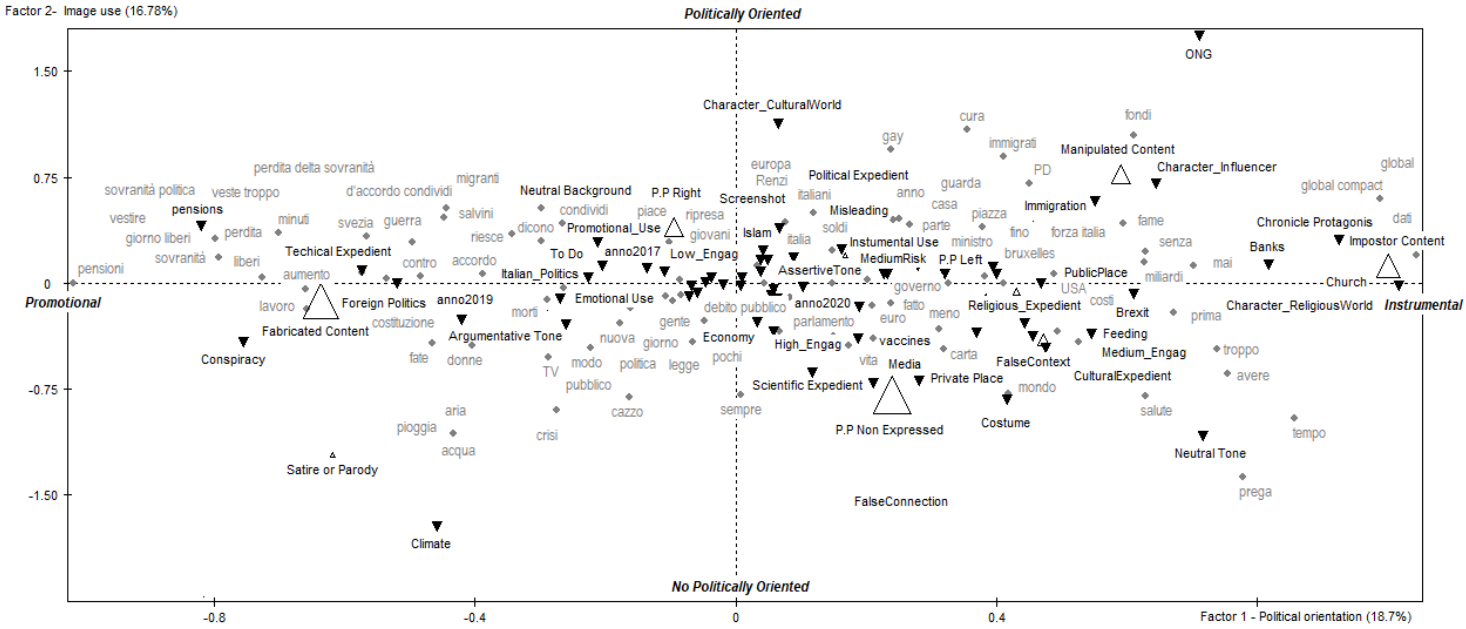
Grafico 28 - Stile e tono della comunicazione visuale (%n=251)



Le strategie disinformative utilizzate dai contenuti visivi possono essere analizzate attraverso due dimensioni sintetiche emerse dall'analisi fattoriale[69]: i contributi più alti nella creazione della prima dimensione, che nel grafico 29 è rappresentata sull'asse orizzontale (1° asse), rimandano a due aspetti strategici della comunicazione disinformativa, caratterizzata dall'opposizione tra promozione e strumentalizzazione. La seconda dimensione, che nel grafico 29 è rappresentata sull'asse verticale (2°asse) prende forma dalla contrapposizione tra una comunicazione politicamente non orientata e una politicamente orientata[70].

[69] Sono attive le variabili: *Disinformation Spectrum* (6 modalità), *Image Use* (3 modalità), *Communication Style and Tone* (4 modalità), *Text's Aim* (3 modalità), *Political Position* (4 modalità). Sono illustrative le variabili: *Representation Type* (7 modalità), *Kind of Characters* (5 modalità), *Number of Characters* (9 modalità), *Text Expedient* (8 modalità).
[70] I primi due assi fattoriali spiegano complessivamente il 33,2% dell'inerzia totale.

Grafico 29 - Piano fattoriale derivante da ACL (assi 1-2)



I quattro spazi di attributo derivanti dall'incrocio di queste due dimensioni si differenziano per toni e obiettivi comunicativi, utilizzo delle immagini e posizione politica espressa. La comunicazione promozionale non orientata politicamente fa un uso delle immagini di tipo emozionale e basa la comunicazione su toni neutrali, volti a "far sapere". L'espediente del testo è soprattutto scientifico[71] e culturale ma di matrice complottista. Infatti, i lemmi presenti in questo spazio rimandano a temi tipici del complotto come quello della salute e dell'ambiente. Nel quadrante opposto emerge una comunicazione politicamente orientata a sinistra e di tipo strumentale. Il tono della comunicazione è assertivo e le immagini ritraggono *influencer*, protagonisti di fatti di cronaca o del mondo culturale in spazi pubblici. Il tema trattato è quello dell'immigrazione, della comunità LGBT e dell'Europa.

Al contrario, la comunicazione politicamente orientata a destra è di tipo promozionale, data dalla presenza di lemmi come "d'accordo", "condividi", le immagini hanno uno sfondo neutro quindi chiaramente volte a trasmettere del testo: il messaggio testuale ricalca una chiara posizione politica sovranista con particolare riferimento al tema della perdita di sovranità legata anche al tema dell'immigrazione e del lavoro in termini di regolamentazione, precariato e retribuzione. Sono infatti presenti lemmi e segmenti come: "giorni liberi", "pensioni", "certificato medico", "retribuzione sufficiente", "stipendio".

[71]La modalità *"sicentific expedient"* e *"technical expedient"* sono state impostate come illustrative dato che il loro peso residuale tendeva a compromettere i risultati.

Nel quadrante opposto, tra una comunicazione politicamente non orientata e un uso strumentale dell'immagine, il messaggio si basa sull'espediente scientifico, il tema climatico ed economico è accompagnato da quello dei vaccini, della salute e dei media in contrasto rispetto alla presenza di personaggi del mondo religioso. Il tono è neutrale e la strategia comunicativa è quella del contenuto ingannevole che utilizza falsi contesti per confondere il lettore: anche se l'informazione è attendibile questa viene condivisa decontestualizzata, rendendo così il messaggio generale falso.
La comunicazione veicolata da queste immagini è ad alto coinvolgimento e tratta temi molto delicati con posizioni negazioniste e pericolosamente radicali e una comunicazione volta a persuadere gli utenti.

Sulla base di queste considerazioni si sono definite due strategie disinformative per i contenuti esclusivamente visivi

Tabella 4 - Strategie disinformative delle immagini

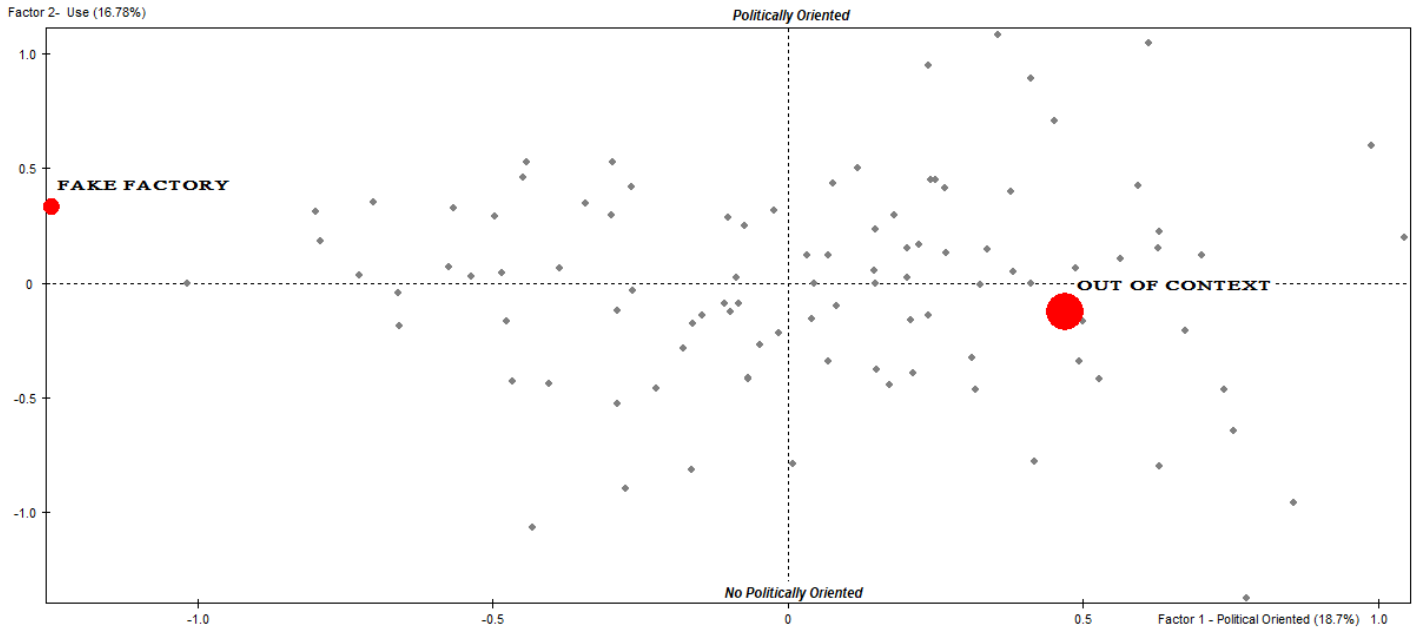
Strategy	Content and Risk	Image Use	Text Use	Visual Elements	Political Orientation	Themes	Bigrams
OUT OF CONTEXT (72,66)	Impostor False Connexion **Medium Risk**	Instrumental	Neutral Tone	Public Place Political Characters	Left; Not expressed	Economy; Climate; Costume	*Italia; banche; bene; governo; parte; dovere; media; gente; soldi; Europa*
FAKE FACTORY (27,34%)	Fabricated **High Risk**	Emotional; Aimed to action	Argumentative Tone	Neutral background	Right; Refuse	Italian politics; Foreign politics	*perdita della sovranità; sovranità politica, d'accordo condividi, migranti, Salvini, guerra, condividi, liberi, perdita, pensioni, aumento, giovani.*

- *Out of context*: più diffusa (72,66%), si caratterizza per un rischio disinformativo medio visto l'utilizzo di contenuti ingannevoli o fuori contesto. L'immagine ha un uso strumentale e utilizza elementi come spazi pubblici e personaggi politici per trattare temi come l'economia e il clima. La posizione politica è orientata a sinistra o non espressa affatto.

- *Fake Factory:* meno diffusa (27,34%), si caratterizza per un rischio disinformativo alto visto l'utilizzo di contenuti completamente fabbricati. L'immagine ha uno sfondo neutro, quindi è presente del testo e fa leva sull'emotività per spingere l'utente all'azione (in questo caso

l'azione di diffusione del contenuto vista la presenza dei *bigrams* "d'accordo-condividi"). La posizione è orientata a destra o rifiutata e il tema trattato è politico-nazionalista.

Grafico 30 - Cluster proiettati su piano fattoriale derivante da ACL



Come emerge, la produzione del falso attraverso i contenuti visuali si posiziona tra una comunicazione politicamente ben orientata e quella promozionale, volta a fidelizzare un pubblico *target*; mentre la strategia basata su contenuti "fuori contesto" quindi tendenzialmente fuorvianti non è influenzata dall'orientamento politico e utilizza un tipo di comunicazione strumentale, volta a conseguire uno scopo tramite un atto comunicativo. Gli inviti all'azione sui *social media* sono noti per essere mezzi efficaci di mobilitazione, sarebbe quindi utile sensibilizzare gli utenti sulle strategie distorsive di questo tipo di contenuto per imparare ad individuarlo e contrastarlo.

Quando il contenuto testuale presente sulle immagini riesce ad eludere con poche difficoltà il controllo algoritmo, le infografiche create dalla piattaforma per i contenuti di tipo *link* volte ad informare l'utente sull'attendibilità della fonte condivisa, non sono più sufficienti.

Per le immagini manca qualsiasi tipo di orientamento ed è dunque molto più difficile per un lettore digitalmente poco alfabetizzato orientarsi nel riconoscimento dell'inganno.

Conclusioni. Non è tutto *fake* ciò che luccica

L'analisi della disinformazione non può che muoversi lungo un continuum di veridicità e finzione: i contenuti analizzati sono estremamente eterogenei, si collocano in posizioni diverse lungo questo *continuum* e si adattano in maniera differente alla piattaforma digitale; pertanto, non possono più essere analizzati in maniera indistinta[72].

Non v'è dubbio che la disinformazione sia sempre esistita, ma quando si è combinata con il potere distributivo dei *social media* è diventata pervasiva. Le regole dell'informazione sono profondamente cambiate perché è cambiato il *medium* su cui sono diffuse le notizie e il paradigma di realtà dentro cui siamo immersi: il racconto disintermediato del reale, fatto di *link*, post, *hashtag*, immagini, trasforma il vero in verosimile, per questo non si può più parlare di vero contro falso, ma di vero e falso insieme.

Adattarsi ai mutamenti della piattaforma sopravvivendo nel tempo alle azioni di censura e sfruttando le possibilità offerte dalla piattaforma stessa, si è rilevata una prerogativa della disinformazione. Questa sorprendente capacità di adattamento ha trasformato Facebook da piattaforma di creazione a megafono di amplificazione: sebbene gran parte dei contenuti provengano dall'esterno della piattaforma è pur vero che senza la vetrina della piattaforma non avrebbero la stessa risonanza proprio perché i meccanismi di funzionamento del *social network*, sono funzionali alla diffusione della disinformazione. La comunicazione veicolata da questi contenuti utilizza le stesse logiche di diffusione dell'*Instant Marketing*; le strategie alla base della manipolazione delle informazioni non sono molto diverse da quelle utilizzate per convincere gli utenti a comprare un paio di scarpe nuove: l'interesse dei produttori è quello di catturare la risorsa più scarsa dell'economia dell'informazione, ossia l'attenzione degli utenti e tenerla il più a lungo possibile. Quest'obiettivo è raggiunto promuovendo sensazionali, controversi e coinvolgenti contenuti manipolati che si servono di argomenti di tendenza particolarmente divisivi e che fanno leva sulla diffidenza, ormai diffusa, verso fonti attendibili o versioni ufficiali dei fatti. Insinuando il dubbio con ragionamenti e prove spesso poco lineari, questi contenuti giocano un ruolo rilevante nel nutrire narrative cospirative che a loro volta guidano il coinvolgimento degli utenti, allineandosi così agli interessi dei proprietari delle piattaforme.

La comunicazione disinformativa basa la sua strategia di diffusione sul sentimento di comunità di quanti si ergono portavoce di verità con il compito di diffondere tutto ciò che non trova spazio nel *mainstream*, e sulla persuasione ad entrare a farne parte di una cerchia esclusiva di utenti che non si

[72] Basta pensare che è solo con l'avvento della pandemia che la piattaforma Twitter ha ampliato le categorie di analisi dei tipi di contenuto per includere informazioni "fuorvianti" (Twitter, 2019).

piegano alle versioni ufficiali dei fatti diffuse dalle grandi lobby. Queste comunità si riuniscono intorno a determinate narrazioni e si comportano come gruppi di opinione dal pensiero unico (Quattrociocchi e Vicini. 2016). L'identificazione automatica e la conseguente eliminazione dei contenuti che violano gli *standard* comunitari, oltre a basarsi sul messaggio, dovrebbero tener in considerazione anche il tipo di post, come è strutturato e il modo in cui è condiviso. L'incrocio di queste caratteristiche potrebbe essere utile ai proprietari della piattaforma per costruire (qualora lo vogliano), un modello di apprendimento automatico che sia in grado di rilevare quei contenuti che si manifestano come potenzialmente disinformativi (e non solo falsi). In realtà sarebbe utile lavorare sulla messa a punto di uno strumento di controllo che non si basi esclusivamente sull'apprendimento automatico.

L'avvento della pandemia, ad esempio, ha spinto soprattutto i motori di ricerca a curare diversamente i risultati delle *query* riguardo questioni di interesse nazionale, il Coronavirus è stato una sorta di "stato di informazione eccezionale" che ha spinto le piattaforme verso il ritorno degli editori (Rogers, 2021). Ciò ha sollevato la questione se il lavoro degli editori possa estendersi, oltre le fonti sanitarie, anche ad altri tipi di informazioni, come ad esempio quelle relative alle elezioni politiche. La cura dei risultati delle *query* è un lavoro laborioso, caduto in declino con la scomparsa generale dell'*editing* umano e l'ascesa degli algoritmi che hanno preso il posto degli editori (ivi). Ci sono sostenitori del recupero editoriale *online* che ritengono che le informazioni trattate dai motori di ricerca riguardo temi delicati non dovrebbero essere completamente affidate al funzionamento degli algoritmi ma restituite al lavoro di redazione, promuovendo così il valore qualitativo dell'aspetto della fonte. Ovviamente la questione della gestione del volume delle fonti la rende una strada difficilmente percorribile. Un'altra soluzione potrebbe essere un maggior appello alla "saggezza della folla" (ivi), attraverso l'adozione di un approccio pubblico attivo, presumendo così che il più significativo strumento di controllo risieda proprio nell'utente capace di segnalare contenuti ingannevoli su varie piattaforme etichettandoli come inappropriati, fuorvianti, cospirativi, falsi *ecc*. La sensibilizzazione alla disinformazione si arricchisce infine con l'educazione al dubbio: approcciando alle dinamiche narrative che emergono con un atteggiamento popperiano che potrebbe diventare lo strumento nelle mani degli utenti/segnalatori per orientarsi nel *mare magnum* dei contenuti diffusi in rete.

Infine, il diritto di avere un libero accesso alle informazioni dovrebbe necessariamente essere accompagnato dalla responsabilità di ciascuno a sottoporsi ad una dieta mediale più variegata ed equilibrata possibile. Lo sforzo collettivo degli utenti dovrebbe dunque essere mirato all'approfondimento e al confronto dialettico, unici atteggiamenti in grado di produrre un pensiero analitico, critico e consapevole.

CAPITOLO V - Dentro un *echo-chamber*

Sebbene la definizione non sia ancora ben definita concettualmente, l'immagine della camera di risonanza ha proprio lo scopo di trasmettere l'idea di quanto le persone che utilizzano le piattaforme siano esposte in gran parte (o esclusivamente) ad un tipo di contenuto pro-attitudinale. Le preoccupazioni circa la *black box* algoritmica alla base dell'infrastruttura delle piattaforme, riguarda soprattutto il campo della fruizione delle informazioni in quanto riveste un ruolo fondamentale nella costruzione dell'opinione pubblica. Concetti come governabilità algoritmica (Rouvroy e Berns, 2013) o responsabilità algoritmica (Diakopoulos, 2015) sottoscrivono la premessa centrale che gli algoritmi hanno sia una dimensione politica, sia una dimensione culturale in quanto servono come mezzo per sapere cosa c'è da sapere (Rieder, 2020). I sistemi di gestione dei *feed* nei *social media* non sono solo mezzi per produrre e filtrare conoscenza ma anche per mettere in atto preferenze di valore, partecipando alla definizione stessa di quella conoscenza.

Finora ci siamo focalizzati sulle *echo chamber* guardandole in maniera trasversale, e ciò che è emerso è che le strategie disinformative si differenziano in base alla produzione dei contenuti interna o esterna alla piattaforma, che provano a sopravvivere alla censura attraverso l'utilizzo delle immagini e che veicolano un messaggio politicamente ben definito.

Sull'interno delle *echo cambers* sappiamo dalla letteratura che le comunità di utenti si aggregano attorno ad argomenti specifici e che tendono a confrontarsi per trovare possibili spiegazioni a fenomeni per loro rilevanti. Gli algoritmi di *feed* creano le condizioni per la formazione e il mantenimento di reti omogenee, giocando un ruolo considerevole nella diffusione della disinformazione e nella diffusione di un tipo di messaggio in stile populista (Engesser et al., 2017).

Ciò che ci sfugge è se questi utenti, una volta dentro l'*echo chamber,* assorbono indistintamente tutti i contenuti proposti condividendo la stessa narrazione o se esiste un qualche elemento che li distingue. Ci si è chiesti se quell'orientamento politico, emerso con prepotenza già al livello del messaggio, orienti in qualche modo anche l'interazione degli utenti con i contenuti diffusi in queste camere di risonanza e se possa essere determinante nella radicalizzazione delle loro opinioni. Sebbene il comportamento radicalizzato degli utenti nelle camere d'eco sia ben documentato da un proficuo filone di ricerca *data-driven*, c'è relativamente poca ricerca empirica che tenti di approfondire come e in che direzione si radicalizzano queste opinioni. È ciò che si è tentato di fare in questa parte della ricerca i cui risultati sono descritti nei paragrafi che seguono.

1 Come si differenziano gli utenti nelle camere d'eco disinformative

Ponendo il *focus* all'interno della *echo chamber* disinformativa, si nota che la disinformazione è ragionevolmente associata al sostegno populista. A questo proposito possiamo far riferimento alle *echo chambers* come camere di somiglianza politica sia nell'informazione che nella discussione.

Grafico 32 - Polarizzazione del Sentiment (% n = 110.663)

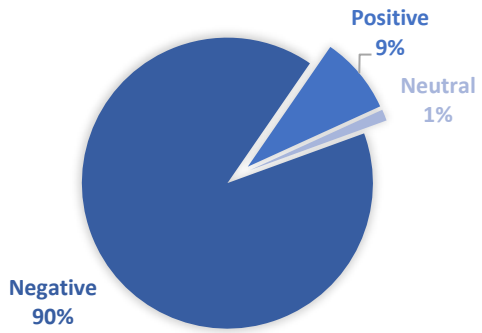
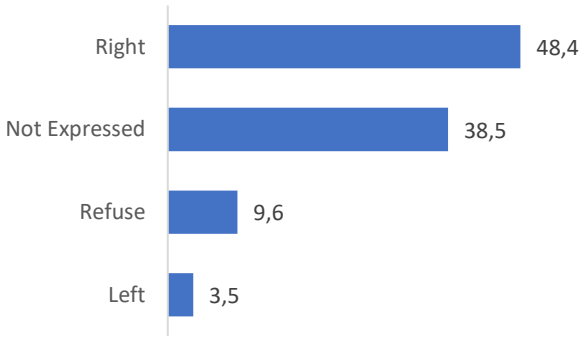


Grafico 31 - Orientamento politico (% n = 110.663)



La polarizzazione del *sentiment* non è solo una componente importante nello studio degli effetti della disinformazione ma sembra anche legata al sostegno dell'orientamento politico populista.
Il *sentiment* degli utenti nelle camere d'eco disinformative analizzate è polarizzato negativamente, una polarizzazione che sembra riguardare soprattutto l'orientamento politico di destra in quale risulta essere anche quello più frequentemente esposto al rischio disinformativo.

Grafico 34 - Polarizzazione del *sentiment* per orientamento politico (% n=110.663)

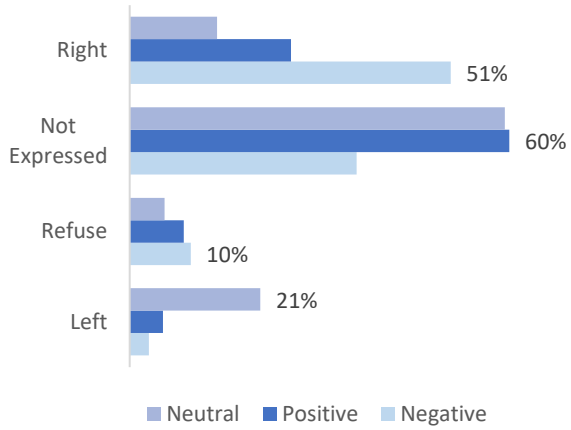
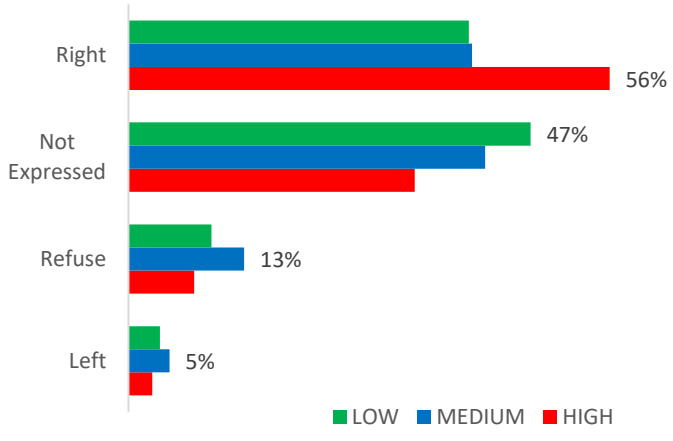


Grafico 33 - Rischio disinformativo per orientamento politico (% n=110.663)



Nella sua forma più elementare, le emozioni sono rilevanti perché influenzano le capacità cognitive degli individui e quindi i modi in cui interpretano, ricordano e reagiscono alla realtà sociale (Neuman, 2007). I contenuti disinformativi sono particolarmente inclini all'impiego di un linguaggio emozionale che in questa ricerca è stato analizzato attraverso l'uso del vocabolario LIWC. Tra le tante categorie emotive di LIWC sono state selezionate la rabbia, l'ansia, l'ottimismo e il consenso. Questa scelta è tutt'altro che arbitraria: nel 2016 Facebook ha offerto ai suoi utenti cinque nuovi modi per interagire con i post presenti sulla piattaforma, aggiungendo agli storici *"like"* e *"share"* le *reaction "love," "haha," "wow," "sad"* e "*angry*".

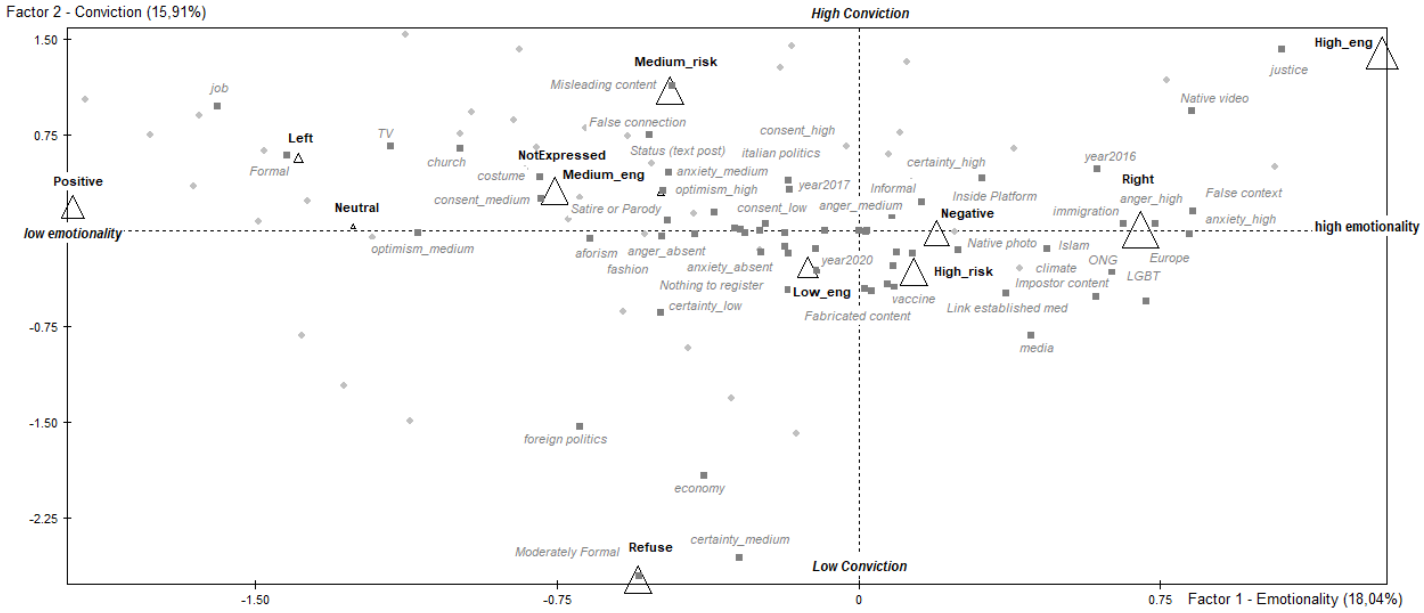
Come descritto, nella proposta di *feed*, Facebook tiene conto di numerose variabili, ponderate in maniera diversa, che si sommano in un singolo punteggio utile a determinare quali tipi di post saranno visibili o meno all'utente che vi accede. Questo unico sistema di punteggio onnicomprensivo viene utilizzato per classificare e ordinare vaste aree di interazione umana nella piattaforma; in quasi tutti i paesi del mondo e in tutte le lingue. A partire dal 2017 l'algoritmo di *ranking* di Facebook ha tarato l'*emoticon* "*angry*" con un valore superiore a tutte le altre perché i contenuti che generano rabbia coinvolgono gli utenti in maniera maggiore rispetto agli altri e li spingono ad una continua interazione (Hagey and Horwitz, 2021). Questa ponderazione ha generato una serie di effetti collaterali in quanto l'algoritmo ha ottimizzato la piattaforma alla circolazione di contenuti controversi che vanno dalle *fake news*, alle voci politicamente conservatrici, fasciste, separatiste e *xenofobe*, a fenomeni di *spamming*, *hate speech*. L'inchiesta Facebook Papers ha confermato che nel 2019 i post che hanno scatenato reazioni di rabbia avevano una probabilità più alta di includere disinformazione, tossicità e notizie di bassa qualità rispetto a tutti gli altri contenuti (Merrill & Oremus, 2020). Rabbia e ansia sono anche le due emozioni più frequentemente associate al populismo e allo stesso modo l'ansia è anche associata a maggiori episodi di pensiero cospiratorio e atteggiamenti *anti-élitari* (Grzesiak-Feldman, 2013).

Diversi studi sulle dinamiche del pensiero complottista (Sunstein, 2009; 2014) suggeriscono che questa narrazione ha l'obiettivo di convogliare la paura degli individui verso ciò che non si conosce che si traduce in sentimenti di ansia e paranoia verso eventi, personaggi o storie specifiche.

A bilanciare i sentimenti negativi nell'analisi sono state inserite le categorie di ottimismo e consenso.

Attraverso questo percorso di metodo possiamo dire di essere entrati gradualmente all'interno di queste tribù dal pensiero unico per comprendere cosa accade. L'analisi ha portato all'evidenza due dimensioni sintetiche di significato che si caratterizzano per l'opposizione tra alta e bassa emotività primo fattore (asse orizzontale) e alta e bassa convinzione (secondo fattore, asse verticale)[73].

Grafico 35 - piano fattoriale derivante da ACM (assi 1-2)



Lungo il primo asse, l'alta emotività contenuta nei messaggi degli utenti corrisponde all'interazione con contenuti ad alto rischio disinformativo che polarizzano negativamente il *sentiment* e generano un coinvolgimento degli utenti molto alto *(high engagement)* come è alta la convinzione espressa verso le opinioni espresse. Questa comunicazione negativa, polarizzata e ferma nelle proprie posizioni caratterizza la posizione politica di destra. Nel lato opposto del piano fattoriale, invece, la bassa emotività espressa nei commenti degli utenti corrisponde all'interazione con contenuti a minor rischio disinformativo che generano un coinvolgimento medio-basso di coloro che esprimono posizioni politiche orientate principalmente a sinistra o che non hanno alcuna posizione politica. Questi utenti esprimono sentimenti positivi o neutri. Lungo il secondo asse, invece, la bassa convinzione e la bassa emotività caratterizzano gli utenti che rifiutano posizioni politiche: stiamo parlando del Movimento Cinque Stelle, che sebbene sia un movimento populista pare, in questo contesto, meno radicalizzato del populismo di destra. Da questa proiezione, appare chiaro che i

[73] I primi due assi fattoriali spiegano complessivamente il 34% dell'inerzia totale. Sono attive le variabili: *sentiment* (tre modalità) orientamento politico (quattro modalità), engagement (tre modalità), rischio disinformativo (tre modalità).

sostenitori del populismo di destra hanno un rischio più elevato di consumare contenuti disinformativi. L'alta convinzione con cui questi utenti esprimono le proprie opinioni rende difficile un eventuale confronto con posizioni diverse, la loro interazione rende virali questi contenuti tenendo vivo un dibattito molto accesso su temi molto delicati come immigrazione, islam, diritti LGBT, media, clima, *ecc*.

Dalla classificazione degli utenti, appare chiaro questi più che grande tribù dal pensiero unico e appaiono piuttosto come una serie di piccole tribù che esprimono convinzione ed emotività differenti, discutono su temi differenti con linguaggi differenti e si espongono in maniera distinta al rischio disinformativo.

Tabella 5 - Cluster di utenti nelle echo chambers disinformative

	Engagment	Language	Convinction	Emotionally	Sentiment	Disinformative Risk	Themes
NO POLITICAL ORIENTATION (78%)	Low	Infomal	High	Absent	Neutral	Medium	Gossip; Chronicle
RIGHT-WING POLITICAL ORIENTATION (12%)	High	Infomal	High	High (anger, anxiety, no optimism)	Negative	High	Immigration; Italian Politics; Climate
POLITICAL ORIENTATION REFUSED (5%)	Low	Moderately Formal	Medium	Low	Neutral	Medium	Economy; Italian Politics
LEFT-WING POLITICAL ORIENTATION (5%)	Medium	Formal	High	Low (optimism high, anxiety medium)	Positive	Low	Job; Italian Politics; Foreign Policy; Environment

L'ansia e la rabbia appaiono essere caratterizzanti dell'utente con posizioni politiche di destra.

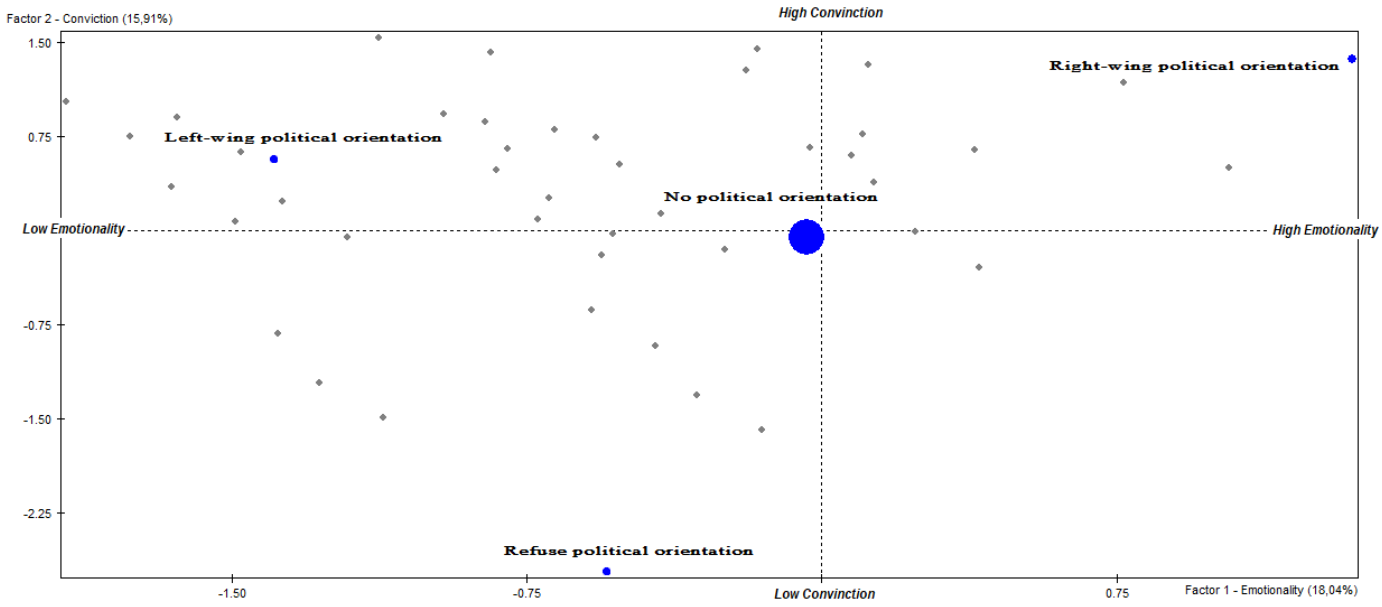
Albertson e Gadarian (2016), dimostrano che innescati dall'ansia, gli individui sono più propensi a cercare informazioni e particolarmente propensi a consumare notizie negative e non attendibili. Allo stesso modo, l'ansia è associata a maggiori episodi di pensiero cospiratorio e atteggiamenti anti-elitari e xenofobi (Grzesiak-Feldman, 2013). Questo sentimento guida l'atteggiamento politico della destra populista ed è associato a temi quali, immigrazione, clima e politica nazionale. All'alto rischio disinformativo già evidenziato, si aggiunge l'utilizzo di un linguaggio informale[74] coerente con il *sentiment* negativo e l'alta emotività espressa nei commenti. L'alto livello di *engagement* pare confermare il funzionamento dell'algoritmo ricostruito nel capitolo precedente che facilita la

[74] Le modalità "formale", "informale" e "mediamente formale" della variabile "linguaggio" sono state costruite sui punteggi della categoria *"bad words*" del vocabolario LIWC che calcola la frequenza delle parolacce contenute nei discorsi, pertanto l'informalità del linguaggio va intesa, in questo caso, come frequente uso di parolacce.

diffusione di contenuti che suscitano forti emozioni, questi atteggiamenti rendono i contenuti virali e rafforzano la coesione dell'*in-group* data dall'alto convincimento verso le posizioni espresse.
Il coinvolgimento diminuisce per gli utenti con posizioni politiche orientate a sinistra (5%), così come diminuisce il rischio disinformativo a cui si espongono e l'emotività della comunicazione, tendenzialmente positiva e ottimista espressa con linguaggio formale. Resta alta la convinzione verso le posizioni espresse sui temi quali il lavoro, la politica e l'ambiente. Il basso coinvolgimento e la bassa convinzione caratterizzano i Cinque Stelle (5%) che, come si vedrà nel prossimo paragrafo, dimostrano di essere poco coesi al proprio interno dal punto di vista delle opinioni espresse.

Nel grafico 36 appare chiaro quanto siano le posizioni politiche - con diverso grado di convinzione ed emotività - a polarizzare e differenziare la risposta ai contenuti disinformativi degli utenti nelle *echo chambers*, diventando così meccanismo guida nel complesso universo disinformativo.

Grafico 36 - Cluster proiettati su piano fattoriale derivante da ACM (assi 1-2)



L'importanza della dimensione "convinzione" rimanda ad una delle teorie attraverso le quali si cerca di rispondere alla domanda: perché si crede ad una notizia falsa? Persone con una fortissima convinzione di base, quando vengono messe di fronte ad un'evidenza che contraddice questa convinzione spesso la rifiutano, questo crea un sentimento di disagio chiamato "dissonanza cognitiva" (Festinger, 1975) che, più o meno inconsciamente, li porterà ad ignorare o addirittura negare tutto ciò che non si adatta a quella convinzione, dirigendosi così verso ambienti che promuovono narrazioni o situazioni che confermano le proprie posizioni, rassicurandole.

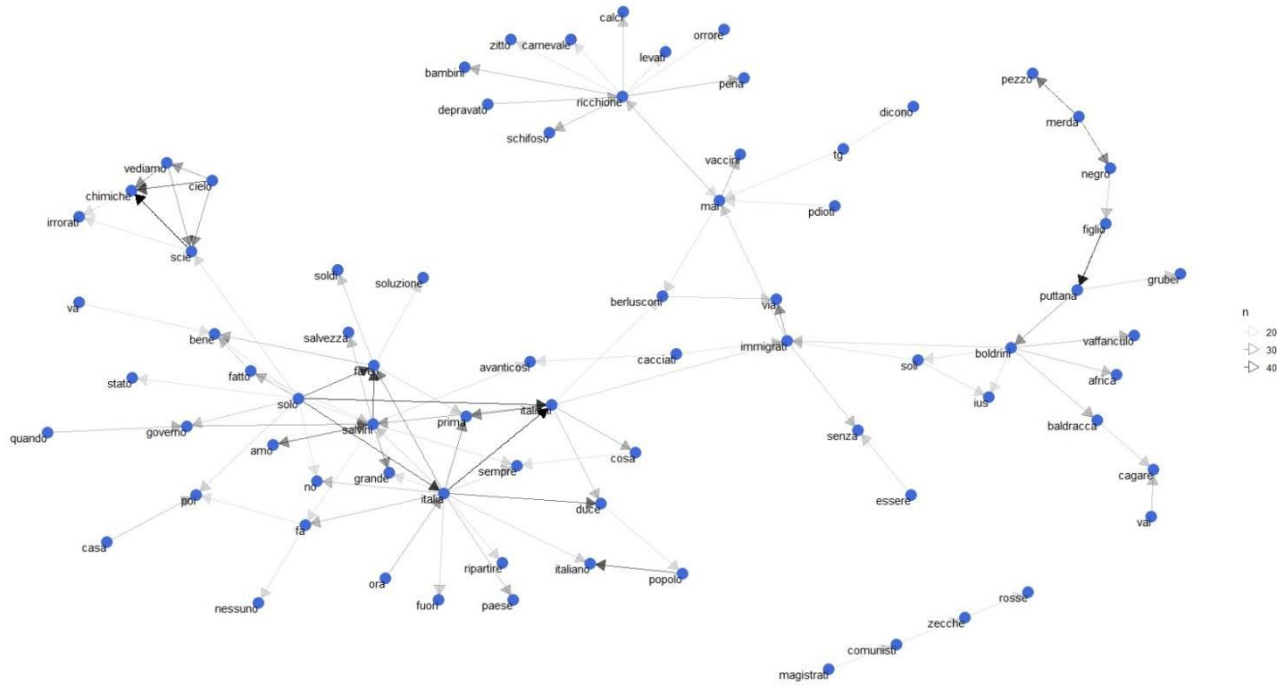
Non siamo gli esseri perfetti come immaginavamo nel ‘700: in maggior o minor misura, la costruzione che abbiamo della realtà è basata su concezioni approssimative e credenze radicate. A questo si aggiunge la complessità di questioni urgenti del nostro tempo, come l’emergenza ambientale, il multiculturalismo, le crisi economiche, la tecnologia può che possono portare le persone, indipendentemente dal livello di istruzione, ad essere incapaci di determinare cause semplici di circostanze complesse e a scegliere così di credere a quelle spiegazioni che non li allontanano dalla propria zona di *comfort*.

2 Le narrative degli utenti polarizzati

La discussione e l'elaborazione di narrazioni in un ambiente disinformativo così segregato politicamente provoca la polarizzazione del gruppo e influenza negativamente le emozioni dell'utente. Pertanto, comprendere su quali temi le posizioni degli utenti si polarizzano sono state ricostruite le narrative dei diversi orientamenti politici attraverso la visualizzazione di una catena di Markov: un modello comune nel *Text Mining* attraverso il quale si costruisce una vera a propria rete di discussione semanticamente orientata, in cui ogni parola dipende solo dalla parola precedente. Per rendere interpretabile la visualizzazione, si è scelto di mostrare solo le connessioni parola-parola più comuni impostando una soglia di frequenza diversa per ogni gruppo di utenti, ma si potrebbe immaginare di proiettare un enorme rete che rappresenti tutte le connessioni possibili che si verificano nel *corpus*.

Ancora una volta, l’interpretazione delle narrative ha beneficiato del momento ermeneutico privilegiato durante la fase di costruzione del *training set.*

Grafico 37 - Rete di Markov utenti con posizioni politiche di destra (frequenza ≥ 20)



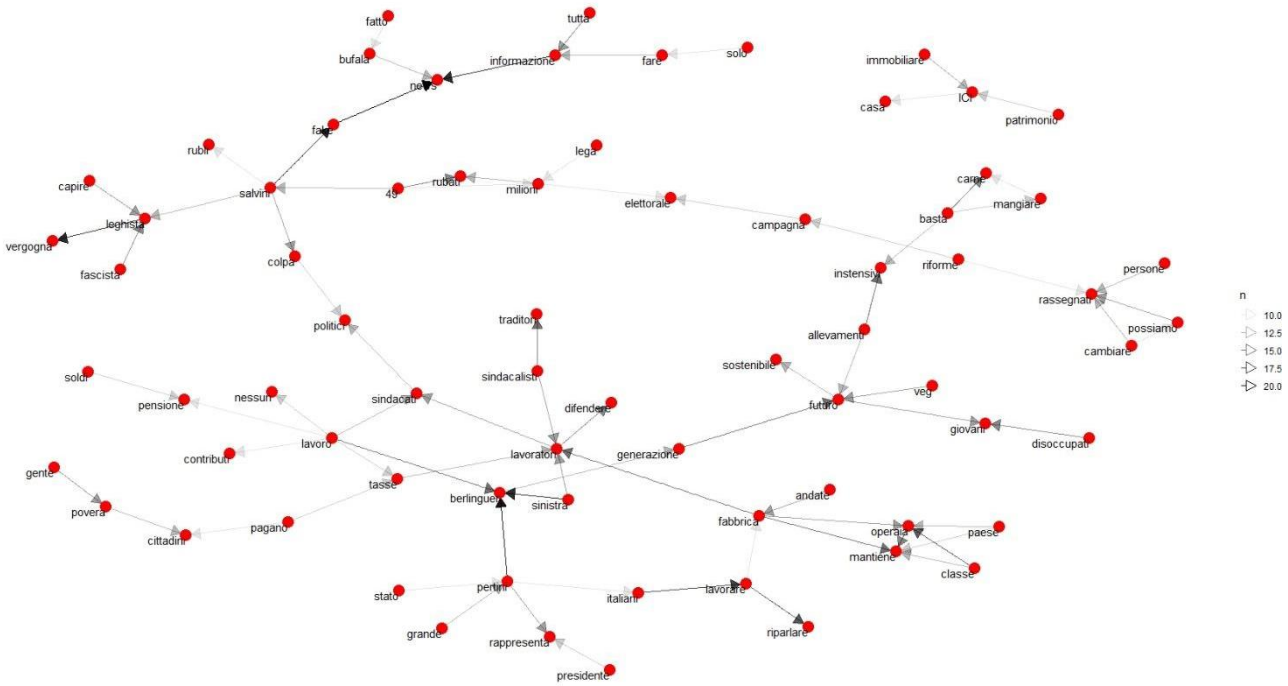
La narrativa portata avanti dagli utenti di destra nelle *echo chambers* disinformative può essere sintetizzata in sei elementi che si contraddistinguono per una forte radicalizzazione e con frequenti richiami all'odio: nazionalismo, complottismo, omofobia, razzismo, misoginia, opposizione politica. Nella parte centrale della rete l'anima nazionalista è supportata dalla figura di Salvini (dal cui nodo si aprono le parole "salvezza", "soluzione", "amo", "grande") ed estremizzata dalla deriva fascista data dalla presenza della parola "Duce"; la stessa anima nazionalista filo-salviniana che si distacca dalla vecchia politica di destra testimoniata dal susseguirsi di parole "Italia", "Berlusconi", "via", "mai".

L'elemento complottista è ubicato in diverse parti della rete: a sinistra si posiziona la narrativa legata alle scie chimiche che si apre dal nodo "solo" e si caratterizzata dalle parole "cielo", "vediamo", "irrorati", "scie", "chimiche". A destra della rete invece, la narrativa complottista si apre dal nodo "mai" e verte sul tema dei vaccini ("mai", "vaccini") e sull'informazione *mainstream* ("mai" "tg", "dicono").

Questo risultato è particolarmente interessante perché porta in evidenza il cosiddetto "paradosso del complottista" (Quattrociocchi, 2021): quelli più attenti alla manipolazione (attuata nelle loro credenze dai mezzi di comunicazione di massa) sono i più propensi a interagire con fonti di informazioni intenzionalmente false e quindi potenzialmente anche i più propensi ad esser soggetti a manipolazione.

L’estremismo delle posizioni di destra è testimoniato dall’elemento omofobo sulla parte alta della rete aperto sempre dal nodo “mai” e caratterizzato dall’universo semantico che sfocia in violenza e *hate speech* (“ricchione”, “schifoso”, “depravato”, “bambini”, “orrore”, “carnevale”, “calci”, “levati”). Violenza e odio si riversano anche verso gli immigrati, rimarcando un atteggiamento innegabilmente razzista (“negro”, “pezzo”, “merda”, “figlio”, “puttana”) e verso le donne. La misoginia che caratterizza il populismo di destra – il cui universo semantico è ben definito e lascia pochi dubbi - prende di mira due differenti figure femminili: Laura Boldrini, legata alla politica tradizionale e di opposizione (atteggiamento *anti-establishment*) e Lilli Gruber legata al mondo dell’informazione *mainstream* (scetticismo verso i media). Infine, la *clique* in basso alla rete esprime un’opposizione politica anti-elitaria anch’essa polarizzata negativamente e data dalla presenza della parola magistrati legata a “comunisti”, “zecche”, “rosse”.

Grafico 38 - Rete di Markov utenti con posizioni politiche di sinistra (frequenza ≥ 10)



La narrativa portata avanti dagli utenti di sinistra nelle *echo chambers* disinformative, contrariamente alle precedenti, appare moderata, lontana dalle logiche populiste, poco radicalizzata e piuttosto positiva. Si caratterizza per cinque elementi: il richiamo alla sinistra storica, la condizione degli operai, il distacco dai sindacati, lo sguardo al futuro sostenibile, la consapevolezza della disinformazione.

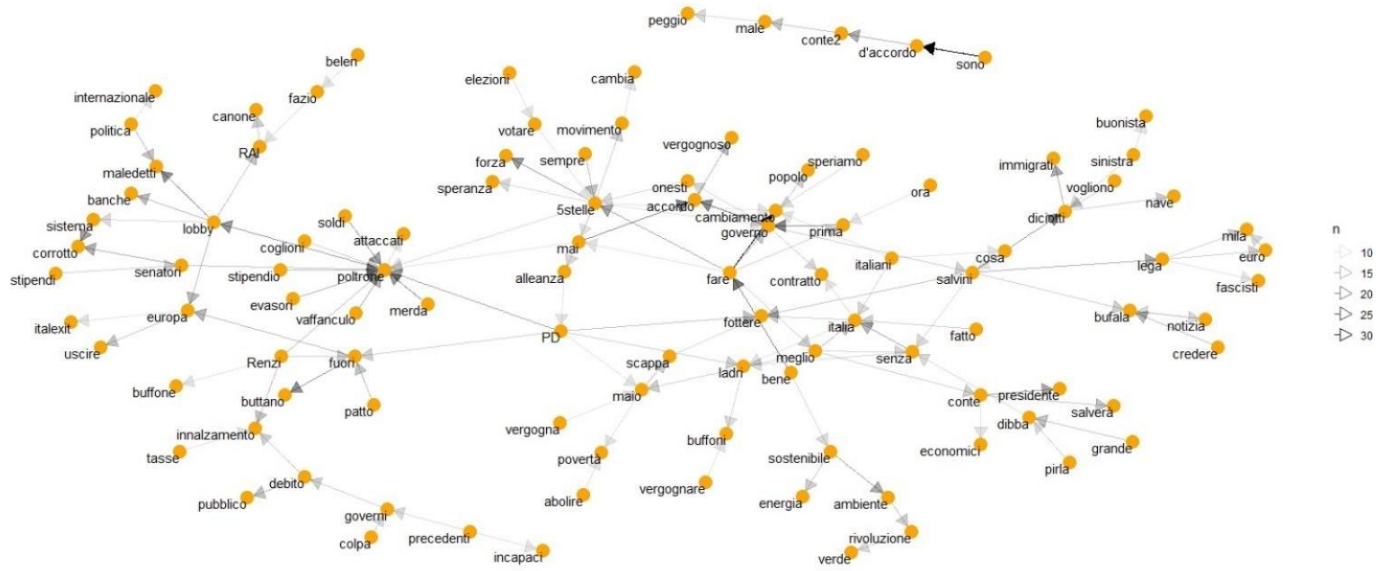
Il forte richiamo alla sinistra storica italiana è visibile nella parte centrale della rete e ben rappresentato dal gruppo semantico *"*Berlinguer", "Pertini", "grande", "Presidente" a loro volta legati dal lemma "lavoro" alle sue tradizionali battaglie: *"*tasse", "pensioni", "povertà", "classe operaia" sulla quale ricadrebbe il peso del paese e la condizione nelle "fabbriche". Gli utenti si distaccano completamente dalla storica istituzione del sindacato e dal ruolo del sindacalista, considerato traditore dei lavoratori.

Lo sguardo al futuro è invece molto forte e visibile in alto a destra della rete, in cui vengono richiamati i temi della sostenibilità ambientale legata agli allevamenti intensivi e al consumo della carne, ma anche alla preoccupazione per la dilagante disoccupazione giovanile.

Dalla parola "riforma" si apre un gruppo semantico ("persone", "possiamo", "rassegnati") che rimanda ad una visione positiva verso il cambiamento, possibile solo collettivamente superando il sentimento di rassegnazione.

L'opposizione politica è soprattutto verso Salvini in particolare e la Lega in generale, al quale si condannano le posizioni fasciste, evidenti nella rete precedente, e i quarantanove milioni di euro rubati. La poca radicalizzazione la si può notare anche in questa occasione: l'universo semantico di trattazione dell'opposizione politica non sfocia mai in un linguaggio d'odio come accade invece nella posizione politica precedente. Infine, la cosa più interessante che emerge da questa narrativa è la consapevolezza della disinformazione che si apre a partire dal lemma "Salvini", quasi ad attribuirgliene le colpe; la presenza delle parole ad alta frequenza quali "*fake*", "*news*", "bufala", "informazione", "fatto" rimanda alla pratica di commentare i contenuti disinformativi segnalandoli come bufale. Una consapevolezza che nella rete precedente mancava e che si vedrà in quella successiva, anche se in maniera meno rilevante.

Grafico 39 - Rete di Markov utenti che rifiutano collocazione politica (frequenza ≥ 10)



La narrativa di chi rifiuta una collocazione politica appare più polarizzata della precedente ma meno polarizzata degli utenti di destra. I temi trattati sono molti e riguardano posizioni *anti-establishment*, *anti-élite* e pro-ambientalismo, gli utenti sembrano spaccarsi in due posizioni: l'opposizione verso il PD e quella contro la Lega che sfocia nel rifiuto per qualsiasi patto di governo e il favore al presidente Conte, custode di questo patto.

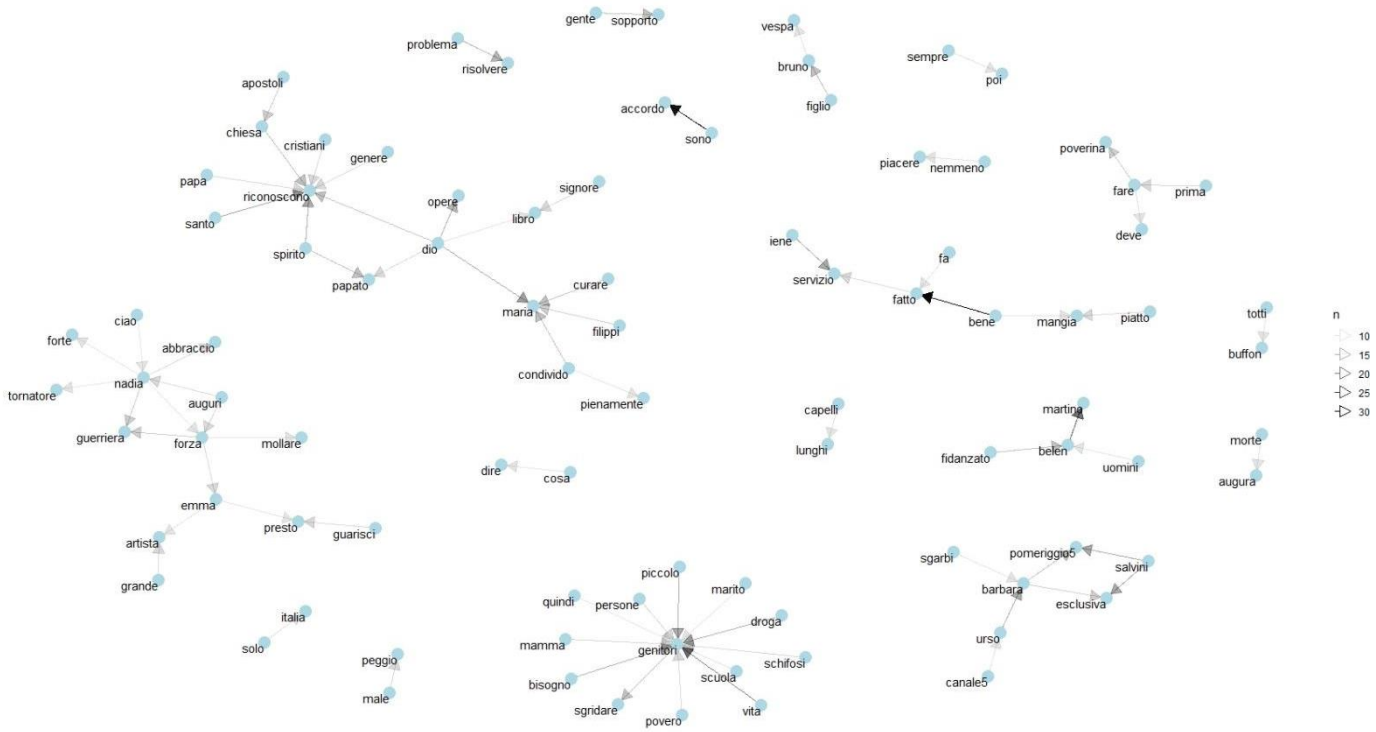
Nella parte destra della rete, un linguaggio polarizzato negativamente rimarca il sentimento *anti-establishment* che, con la parola "poltrone", apre a un insieme semantico che rimanda al tema degli stipendi dei senatori e dell'evasione fiscale. Dal lemma "poltrone" si apre anche il sentimento antieuropeista "uscire", "*italexit*" e *anti-elitario* dato dall'insieme di parole "*lobby*", "banche", "sistema-corrotto", "politica-internazionale". Nelle lobby è inserita anche la RAI, la cui riforma della *governance* è uno dei capisaldi del movimento. L'opposizione politica si concentra sulla colpevolizzazione dei governi precedenti circa il debito pubblico e l'incapacità di migliorare le condizioni del paese, un attacco piuttosto duro è soprattutto al PD di Renzi. Le due anime del movimento emergono chiaramente: quella a favore, espressione di un insieme semantico molto positivo "presidente-salverà" e quella contraria, testimoniata dalle parole "contratto-vergogna", *"*accordo-vergognoso*"* e dalla *clique* in alto in cui si nota un pessimismo verso il governo Conte2. Le stesse due anime sono evidenti anche nei confronti di Di Battista, in basso nella rete, dal quale si aprono le due parole opposte "pirla" e "grande". L'attenzione verso l'universo ambientalista ha una semantica piuttosto positiva, considerando i lemmi "sostenibile", "energia", "ambiente", "rivoluzione-verde".

La narrativa nei confronti di Salvini (a destra della rete) si dirama in tre direzioni con *sentiment* diversi: sulla vicenda della nave Diciotti, le posizioni degli utenti sembrano piuttosto concordi con la linea della Lega considerando l'utilizzo di lemmi molto vicini alla comunicazione salviniana, come "sinistra-buonista", "sinistra-vogliono-immigrati"; il disaccordo con Salvini è espresso invece dall'accusa di fascismo e con il rimarcare della questione dei quarantanove milioni di euro rubati dal partito.

La cosa interessante è la presenza, ancora una volta, del tema delle notizie bufala collegato a Salvini[75]: la capacità dimostrata dalla comunicazione politica populista della Lega è legata ad una coalescenza di caratteristiche, tra cui l'amplificazione di notizie semi-veritiere, che vengono virilizzate a tal punto da diventare cultura condivisa, rafforzata sia dalla fonte considerata carismaticamente affidabile, sia dal numero di condivisioni che la rendono in quel momento difficilmente contestabile. Una vera e propria spirale del silenzio (Noelle-Neumann, 2002) osservata nella sua forma *online* che enfatizza e rinforza le opinioni e i sentimenti prevalenti, silenziando quelle minoritarie e dissenzienti.
Dunque, la connessione tra disinformazione e populismo non risiede unicamente nelle loro somiglianze comunicative ma è supportata anche in alcuni *avatar* dell'ambiente politico. *Leader* populisti di destra come Matteo Salvini, Donald Trump o Jair Bolsonaro sono attori centrali nella costruzione discorsiva della disinformazione (Morris, 2020[76]): incolpando i media *mainstream* nelle loro comunicazioni, mettono in dubbio l'evidenza empirica e la conoscenza degli esperti e danno priorità alla verità e al buon senso delle persone messe al centro della scena dell'onestà e della realtà. Queste pratiche si propagano nelle piattaforme *social media* e agiscono nella costruzione dell'opinione pubblica e nella radicalizzazione delle posizioni degli utenti.

[76] Matteo Salvini si è recentemente guadagnato una menzione particolare della BBC per essere, insieme a Bolsonaro e Trump, tra i politici che diffondono più fake news. Il servizio è stato realizzato dal giornalista Chris Morris per la trasmissione Reality Check, specializzata nel verificare le notizie che girano in rete. La puntata è disponibile al link: https://www.bbc.com/news/av/52299689

Grafico 40 - Rete di Markov utenti che non esprimono posizioni politiche (frequenza ≥ 10)



Dalla rete è chiaro che gli utenti che commentano contenuti senza esprimere posizioni politiche hanno opinioni non radicalizzate espresse con un *sentiment* piuttosto neutro. I temi di discussione sono legati alla Chiesa, alla famiglia, ai personaggi dello spettacolo, come "Nadia" Toffa, "Emma" Marrone e della televisione "Belen", "Barbara D'Urso".

Alla luce di quanto è emerso è possibile fornire alla disinformazione un altro aspetto interpretativo, inquadrandola in termini di spaccatura ideologica.

È chiaro che queste narrazioni non si materializzano al contatto con contenuti disinformativi, piuttosto sono questi ultimi a sfruttare le tensioni preesistenti nella società radicalizzando le convinzioni e rendendo la comunicazione a tratti feroce e virale attraverso il richiamo ad emozioni forti come paura, disgusto o sorpresa. Contrariamente alle aspettative, la scomparsa del ruolo di *gatekeeper* svolto dalle redazioni e dai media *mainstream* non ha favorito un'aperta arena di comunicazione in senso habermasiano, piuttosto ha aperto uno spazio per far prosperare la disinformazione e favorire segregazione e polarizzazione (Törnberg, 2018; Bergmann, 2018). La disinformazione polarizza la realtà: i contenuti divisivi attirano utenti sia favorevoli che contrari, i quali vengono catapultati in una spirale comunicativa che porta visibilità alla fonte ma nessun confronto: il dialogo è debole, si danno giudizi e si disegnano schieramenti secondo logiche quasi plebiscitarie. La difesa delle proprie

posizioni e della propria credibilità è così pronunciata da sfociare in una modalità difensiva persecutoria, rafforzata dallo screditamento aggressivo dell'altro e dall'innalzare barriere per difendersi dal dissenso.

Vittimizzazione e persecuzione diventano ruoli interscambiabili nelle dinamiche comunicative degli utenti immersi in *echo chambers* disinformative, una pratica comunicativa che caratterizza anche i discorsi populisti. Infatti, dai risultati dell'analisi appare chiaro quanto le posizioni politiche orientino gli utenti nel rapportarsi ai contenuti disinformativi; la disinformazione attecchisce lì dove le posizioni diventano intransigenti e si radicalizzano, generando uno spazio dal confronto ridotto in cui qualsiasi tipo di contenuto viene inghiottito senza molto pensiero critico. Ci sono dunque ragioni per credere che il problema della disinformazione riguarda in modo particolare gli attori populisti. La diffusione della disinformazione è stata alimentata dalla proliferazione dei *social media* e favorita dalla circolarità tra mittente e destinatario della comunicazione, resi intercambiabili dall'ambiente disintermediato delle piattaforme. Allo stesso modo, i social media sono stati un valido strumento di propaganda per i partiti populisti nel rivolgere il loro appello direttamente alle persone.

Alla luce di quanto emerso, i contenuti disinformativi e il populismo mediatico alimentano lo stesso tipo di argomenti e talvolta si sovrappongono. Il primo elemento comune è la distinzione manichea tra noi e loro, nucleo interno del populismo e allo stesso tempo contesto in cui le persone si appropriano di contenuti, in varia misura, contraffatti. Le argomentazioni populiste riflettono i valori, le opinioni e a volte le frustrazioni degli utenti, contribuendo allo sviluppo e al consolidamento di percezioni errate sui fatti politici e sugli *out-group* che possono essere indistintamente immigrati, minoranze etniche, minoranze sessuali o talvolta minoranze di estremamente ricchi e potenti, come lobby di banche e case farmaceutiche. I messaggi populisti hanno dunque un forte impatto sugli stereotipi e sull'attribuzione di colpe agli *out-group*. Allo stesso modo, attribuire colpe e diffondere stereotipi sono caratteristiche emerse anche nel messaggio disinformativo, queste generano un effetto mobilitante sugli utenti con opinioni simili, che sfogano sulla piattaforma sentimenti anti-*establishment,* anti-immigrati, anti-media. Come è emerso, più che al valore di verità del contenuto, gli utenti (soprattutto di destra) trattano i contenuti disinformativi come fatti credibili e condivisibili e la convinzione che esprimono innesca dibattiti accesi, che a volte diventano incendiari e offensivi.

Questo ci collega al secondo elemento in comune: disinformazione e populismo si incontrano sul livello di *appeal* emotivo. La disinformazione, così come il populismo, si viralizza e persuade gli utenti grazie all'alto livello di emotività su cui fa leva. Il ruolo delle emozioni è rilevante in entrambi i fenomeni: la comunicazione populista costruisce messaggi che innescano negli utenti risposte emotive molto potenti come rabbia o odio, allo stesso modo il magnetismo degli elementi emotivi è ciò che rende gli utenti più vulnerabili ai contenuti disinformativi. La natura emozionale e

provocatoria di molti contenuti, diffusi e co-creati nelle camere dell'eco disinformative, rende le persone esposte all'indignazione e favorisce il consolidamento di una percezione e di un'identità *in-group* che porta a una crescente frammentazione e radicalizzazione. Come è emerso, l'emotività influisce anche sulla viralità della disinformazione nelle camere d'eco: visioni del mondo simili e l'alta convinzione verso le proprie posizioni sono elementi sufficienti per viralizzare un contenuto che si diffonde grazie al supporto di ambienti omofili.

Come già detto, probabilmente la connessione tra disinformazione e populismo non risiede unicamente nelle loro somiglianze comunicative ma si riscontra anche nei *leader* politici. Per questo motivo non è da escludere un approfondimento che si concentri sugli effetti, spesso combinati o simili, della disinformazione e della comunicazione dei *leader* populisti.

Conclusioni e lavoro futuro

Arrivati a questo punto è utile riflettere sulle implicazioni dei risultati emersi nel corso di questo lavoro. Queste riguardano sia aspetti teorici che metodologici.

Nonostante la rete resti uno strumento formidabile, in un contesto in cui ognuno diventa un *medium* in grado di amplificare messaggi attraverso le piattaforme, la consapevolezza dello strumento che si utilizza è necessaria per orientarsi in ambienti algoritmici che ci propongono contenuti di nostro interesse, rafforzando le nostre convinzioni e non favorendo il confronto con gli altri. Queste camere d'eco in cui siamo immersi sono destinate ad avere pareti sempre più spesse dovute alle sofisticate capacità di profilazione delle piattaforme; in questo scenario la nostra dieta informativa sarà sempre più allineata al pensiero omofilo della nostra rete amicale e rischieremo di imbatterci con probabilità crescente in contenuti capaci di alterare o falsificare la percezione che abbiamo della realtà.

In questo tipo di ambienti, gli elementi dell'informazione subiscono un continuo processo di riformulazione, confermando ancora una volta la riflessione di McLuhan che il messaggio trasmesso è costituito dalla natura del *medium* stesso. I criteri strutturali, stilistici e di profitto in base ai quali le piattaforme organizzano la diffusione e la fruizione dell'informazione le rendono non neutrali e particolarmente influenti nel processo di confezionamento delle notizie e nell'assunzione di determinati comportamenti e posizioni degli utenti.

Riguardo quest'ultimo punto, l'*opinion mining* ha evidenziato quanto le posizioni politiche populiste, ad un diverso grado di convinzione ed emotività, diventano il meccanismo guida degli utenti nel complesso universo disinformativo. Per gestire la credibilità e l'autorevolezza delle posizioni che si affermano in questi spazi chiusi, i gruppi di utenti costruiscono una sorta di "metatesto" caratterizzato da contenuti e lessici specifici attraverso il quale si esplicitano le ragioni di rilievo della comunità che lo assume come riferimento. Questa tendenza discorsiva è emersa con il postmoderno che ha reso gli individui valutatori della realtà in ragione delle proprie credenze e pulsioni. L'elemento tecnologico della piattaforma assume, in questi discorsi, una funzione che è tradizionalmente associata al linguaggio, quella *fàtica* (Lorusso, 2018), ossia finalizzata a instaurare e tenere vivo il contatto tra produttore e destinatario della comunicazione. Appare chiaro quanto l'elemento politico populista tragga importanti vantaggi dalla confusione informativa e quanto riesca a spingere situazioni problematiche verso una vera e propria crisi comunicativa (oltre che politica). Sebbene il populismo e la disinformazione siano il risultato dello stesso processo di disintermediazione avvenuto con l'avvento del *web* 2.0, sono pochi gli studi che considerano il populismo come elemento attivo nella promozione della disinformazione, nel consolidamento di posizioni cospirative e nella radicalizzazione di narrative intorno a temi particolarmente divisivi.

È emerso con prepotenza quanto il problema della disinformazione riguardi in modo particolare gli attori populisti di destra, i quali si contraddistinguono per opinioni convinte, fortemente radicalizzate e frequenti richiami all'odio. Queste posizioni sono anche quelle che hanno prodotto più *engagement* e quindi contribuito alla viralizzazione dei contenuti. Le motivazioni di questa viralità dipendono dall'utilizzo di una forte emotività: sono le emozioni che ci portano a commentare e condividere un contenuto; lo stupore, la sorpresa, l'ansia, la paura e la rabbia sono tra le più influenti. Lo studio ha dimostrato che esiste una connessione tra emotività amplificata e disponibilità a credere alle false notizie, ed è per questo che la disinformazione (così come il populismo) fanno leva sulle emozioni che al tempo stesso confondono la valutazione del falso e inducono a condividere il messaggio.
L'emotività indirizza la comunicazione al contatto più che allo scambio modificando così il paradigma *"many-to-many"* in *"me-to-me"* poiché l'obiettivo non è più il confronto su esperienze collettive ma il rafforzamento della propria identità (di singolo o gruppo).
È inevitabile che l'alto valore di *engagement* sia imputabile anche al funzionamento del *news feed* di Facebook che, seguendo logiche commerciali, predilige la diffusione di contenuti che generano rabbia e polarizzazione perché più proficui. Riguardo il ruolo delle piattaforme lo studio ha empiricamente rilevato il ruolo di Facebook come *hub* di smistamento degli utenti verso altri spazi digitali ricoprendo molte più responsabilità come facilitatore di questo ecosistema che come vero e proprio produttore.

Dal punto di vista metodologico, l'ibridazione tra tecniche classiche di analisi del contenuto e tecniche computazionali di apprendimento automatico ha permesso allo studio un ampio ventaglio interpretativo del fenomeno disinformativo in una prospettiva dinamico-evolutiva e con la possibilità di generalizzazione dei risultati.
Affinché l'ibridazione funzionasse è stato necessario approfondire anticipatamente il tipo di *user-generated data* raccolto attraverso l'API della piattaforma e seguire l'evoluzione della loro produzione in riferimento al modificarsi della piattaforma stessa. La conoscenza preliminare di questi dati ha consentito di massimizzare i vantaggi delle diverse tecniche in base al tipo di post da studiare. Riguardo ai post, un altro elemento messo in chiaro da questo lavoro è che la procedura necessaria per etichettare i diversi contenuti disinformativi rende meno rigidi i confini dei metodi digitali nella scelta dell'unità di analisi. L'unità di analisi individuata per la ricerca è il post ma la rilevazione degli stati di proprietà deve necessariamente adattarsi alle diverse forme che esso può assumere. Per essere più chiari: considerando di dover categorizzare un post di tipo *link,* è necessario che il *link* sia aperto e l'articolo sia letto; dunque, l'attribuzione della categoria a questo tipo di post sarà data dal risultato della revisione dell'articolo nel modo in cui esso è condiviso nel post.

In generale è stato il lavoro dettagliato e non esaustivo di categorizzazione a restituire un'immagine completa e profonda del fenomeno indagato, migliorando la comprensione e la descrizione dell'oggetto di studio in riferimento alla letteratura e all'evolversi dell'analisi. I diversi stati della disinformazione emersi dall'analisi fattoriale e di classificazione si prestano alla possibilità di una riconcettualizzazione teorica del fenomeno e delle strategie su cui si basa: abbandonando il fattore d'intenzionalità presente in letteratura e considerando la disinformazione in riferimento al suo rapporto con la piattaforma e con il tipo di messaggio veicolato. Per non rinunciare alla profondità dell'analisi il ruolo del ricercatore è stato indispensabile, ma molto dispendioso.

L'ibridazione è stata utilizzata in prospettiva pratica volta cioè a trovare risposte adeguate alle domande di ricerca iniziali e a quelle emerse durante le diverse fasi d'indagine. Nello specifico, l'integrazione è stata utilizzata per colmare i limiti delle tecniche singolarmente considerate con l'obiettivo di elaborare un nuovo strumento analitico che diviene esso stesso il risultato. La creazione di uno strumento in grado di rilevare la complessità di un fenomeno prevalentemente digitale richiama la necessità espressa dalla Lupton (2015) di ripensare alle pratiche di ricerca sociale in modo innovativo, critico e aperto. Per ciò che concerne gli sviluppi futuri del lavoro presentato, la volontà è quella di lavorare sulla riconcettualizzazione teorica della disinformazione, di analizzarla in combinazione alla comunicazione populista aprendosi al confronto multipiattaforma in modo da definire il ruolo dei singoli algoritmi e il loro effetto nel frammentare lo spazio virtuale in tanti piccoli gruppi che non ascoltano altro se non le proprie convinzioni.

BIBLIOGRAFIA

Agcom (2017), *Le strategie di disinformazione online e la filiera dei contenuti fake.* Rapporto Tecnico 423/17/CONS

Agcom (2017), *News vs. Fake nel Sistema dell'informazione*, Interim Report. Indagine conoscitiva 309/2017/CONS

Ahmed, H., Traore, I., & Saad, S. (2017, October). Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments (pp. 127-138).* Springer, Cham.

Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. *Information Processing & Management, 39(1), 45-65.*

Albertson, B., & Gadarian, S. K. (2016). Did that scare you? Tips on creating emotion in experimental subjects. *Political Analysis, 24(4), 485-491.*

Alleman, J. H., Baranes, E., Rappoport, P. (2019). Multisided Markets and Platform Dominance. *The 47th Research Conference on Communication, Information and Internet Policy*.

Amaturo, E. (1988). L'analisi delle corrispondenze lessicali: una proposta per il trattamento automatico di dati testuali. In Livolsi, M. e Rositi, F. (a cura di) *La ricerca sull'industria culturale*, Roma: La Nuova Italia Scientifica. pp. 95-109

Amaturo, E. (1989). *Analyse Des Donnees & Analisi Dei Dati Nelle Scienze Scoiali.* Torino: Centro Scientifico Editore.

Amaturo, E. (1993). *Messaggio, simbolo, comunicazione. Introduzione all'analisi del contenuto*, Roma: Nuova Italia Scientifica.

Amaturo E., Aragona B. (2016), La "rivoluzione" dei nuovi dati: quale metodo per il futuro, quale futuro per il metodo? In Corbisiero F., Ruspini E. (a cura di), *Sociologia del futuro. Studiare la società del ventunesimo secolo,* Lavis, Wolters Kluwer, pp. 25-50.

Amaturo, E., Aragona, B. (2019). Per un'epistemologia del digitale: note sull'uso di big data e computazione nella ricerca sociale. In *Quaderni di Sociologia, 81- LXIII, 71-90*

Amaturo, E., Aragona, B. (2021). Critical Optimism: A Methodological Posture to Shape the Future of Digital Social Research. In *Italian Sociological Review, 11 (4S), 167-182.*

Amaturo, E., Aragona, B., Grassia M. G., Lauro C. N., Marino M. (2018). *Statistica per le scienze sociali*. Milano: UTET Università.

Amaturo, E., Punziano, G. (2013). *Content Analysis. Tra comunicazione e politica*. Milano: Ledizioni.

Amaturo, E., Punziano, G. (2016). *I Mixed Methods nella ricerca sociale.* Roma: Carocci.

Amaturo, E., Punziano, G. (2017). Blurry Boundaries: Internet, Big-New Data, and Mixed-Method Approach, Data Science and Social Research. In Lauro, N.C., Amaturo, E., Grassia, M.G., Aragona, B., & Marino, M. (a cura di), *Data Science and Social Research. Epistemology, Methods, Technology and Applications.* Springer: Cham, 35-55.

Amazeen, M. A. (2015). Revisiting the epistemology of fact-checking. *Critical Review, (27 1), 1-22.*

Amazeen, M. A. (2016). Checking the fact-checkers in 2008: Predicting political ad scrutiny and assessing consistency. In *Journal of Political Marketing, (15 4), 433-464*

Amazeen, M. A. (2015). *Sometimes political fact-checking works. Sometimes it doesn't: Here's what can make the difference.* The Washington Post.

Aragona, B., Arvidsson, A., & Felaco, C. (2020). Introduction. Ethnography of algorithms. The cultural analysis of a sociotechnical construct. Etnografia e ricerca qualitativa, 13(3), 335-349.

Aragona, B., Felaco, C. (2018). La costruzione socio-tecnica degli algoritmi. Una ricerca nelle infrastrutture dei dati. In *The Lab's Quarterly, 20(4), 97-115.*

Aragona, B., & Felaco, C. (2020). Understanding algorithms. Spaces, expert communities, and cultural artifacts. Etnografia e ricerca qualitativa, 13(3), 423-439.

Aragona, B., Felaco, C., Marino, M. (2018), The Politics of Big Data Assemblages. In *Partecipazione e conflitto, XI, 2, pp. 448-471.*

Baden C., Kligler-Vilenchik, N., Yarchi, M. (2020), Hybrid Content Analysis: Toward a Strategy for the Theory-driven, Computer-assisted Classification of Large Text Corpora. In *Communication Methods and Measures, 14:3, 165-183.*

Baudrillard, J. (1976), *La società dei consumi*, (trad. italiana). Bologna: Il Mulino.

Baudrillard, J. (1991). *La Guerre du Golfe n 'a pas eu lieu.* Editions Galilee

Beaudouin, V. (2016). Retour aux origines de la statistique textuelle: Benzécri et l'école française d'analyse des données. *In JADT 2016 (pp. 17-27).*

Beauchamp, N. (2017), Predicting and Interpolating State-Level Polls Using Twitter Textual Data. In *American Journal of Political Science, 61: 490-503.*

Beck, U. (1997), *I rischi della libertà. L'individuo nell'epoca della globalizzazione* (trad. italiana). Bologna: Il Mulino

Benjamin, W. (1973). *L'opera d'arte nell'epoca della sua riproducibilità tecnica,* Donzelli editore.

Bell, D. (1973). *The coming of post-industrial society*. Basic Books, Inc., New York

Bennato, D. (2015). *Il computer come macroscopio. Big data e approccio computazione per comprendere i cambiamenti sociali e culturali*. Milano: Franco Angeli.

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. In *European journal of communication, (33 2), 122-139*

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

Bentivegna, S., e Boccia Artieri, G. (2019). *Le teorie delle comunicazioni di massa e la sfida digitale.* Bari: Editori Laterza.

Benzécri, J.P. (1968). La place de l'a priori, "Organum". In *Encyclopedia Universalis. pp. 11–24.*

Benzécri, J.P. (1980). *Pratique de l'analyse des données. Analyse des correspondances & classification*. Paris: Exposé élémentaire.
Benzécri, J.P. (1981). *Pratique de l'analyse des données, Linguistique et lexicologie*. Paris: Dunod.

Benzécri, J.P. (1982). *Histoire et préhistoire de l'analyse des données.* Paris: Dunod.

Benzécri, J. P. (1992). *Correspondence analysis handbook*. CRC Press LLC.

Berelson, B. (1952). *Content Analysis in Communication Research*. Michigan: Free Press.

Bergmann, E. (2018). *Conspiracy & populism: The politics of misinformation.* Cham: Springer International Publishing.

Bergmann, E. (2020). Populism and the politics of misinformation. In *Safundi: The Journal of South African and American Studies 21(3), 251–265.*

Berselli, E. (1999). Un giornale tra due fuochi. Ortodossi in politica, eccentrici altrove? In *Problemi dell'Informazione. Rivista Quadrimestrale, 1/1999, pp. 54-60.*

Bessi, A., Caldarelli, G., Del Vicario, M., Scala, A., Quattrociocchi, W. (2014), Social determinants of content selection in the age of (mis) information. In *International Conference on Social Informatics (pp. 259-268).* Springer, Cham.

Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs conspiracy: Collective narratives in the age of misinformation. In *PloS one, 10(2), e0118093.*

Bland, J. M., & Altman, D. G. (1994). Correlation, regression, and repeated data. In *British Medical Journal, 308(6933), 896.*

Blei, D. M., Laerty, J. D. (2009). Topic models. In *Text Mining FF (pp.101-124).* Chapman and Hall/CRC.

Blei, D. M., Lafferty, J. D. (2007). A correlated topic model of science. In *The Annals of Applied Statistics, (1 1), 17-35.*

Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation. In *Journal of machine Learning research, (3 Jan), 993-1022.*

Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic Topic Models: A focus on graphical model design and applications to document and image analysis. *In IEEE Signal Processing Magazine, 27(6), 55–65.*

Blommaert, J., Collins, J., & Slembrouck, S. (2005). Spaces of multilingualism. In *Language & Communication, 25(3), 197-216.*

Boccia Artieri, G. (2012). *Stati di connessione. Pubblici, cittadini e consumatori nella (Social) Network Society*. Roma: Franco Angeli.

Boccia Artieri, G. (2017). Fenomenologia dei social network: Presenza, relazioni e consumi mediali degli italiani online. In *Fenomenologia dei social network, 1-181.*

Bolasco, S. (2005). Statistica testuale e Text Mining: alcuni paradigmi applicativi. In *Quaderni di Statistica, 7, 17-53.*

Bolasco, S., Della Ratta-Rinaldi, F. (2004). Experiments on semantic categorisation of texts: analysis of positive and negative dimension. In *Le poids des mots, Actes des 7es journées Internationales d'Analyse Statistique des Données Textuelles*, UCL, Presses Universitaires de Louvain, 202-210.

Bolasco, S., De Mauro, T. (2013). *L'analisi automatica dei testi: fare ricerca con il Text Mining.* Roma: Carrocci Editore

Boulianne, S., Koc-Michalska, K., & Bimber, B. (2020). Right-wing populism, social media and echo chambers. In *Western democracies. New Media & Society, 22(4), 683–699.*

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital journalism, 4(1), 8-23.*

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15(5), 662-679.*

Briggs, A., & Burke, P. (2002). *Storia sociale dei media: da Gutenberg a Internet.* Bologna: Il Mulino.

Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research. In *Information, Communication & Society, (22 11), 1544-1566.*

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage* (Vol. 45). Peter Lang.

Buenaño-Fernandez, D., González, M., Gil, D., & Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. In *IEEE Access, 8, (35318-35330)*

Caliandro, A., Gandini A. (2019), *I metodi digitali nella ricerca sociale*. Roma: Carocci Editore.

Castells, M. (2017). *Comunicazione e potere*. Nuova edizione. EGEA spa.

Castells, M. (2000). The rise of the network society. In *The Information Age Economy, Society, and Culture (Vol. 1)*. Blackwell Publishing Ltd (2e)

Castells, M. (2008), *La nascita della società in rete.* Milano: Bocconi Editore.

Censis (2017). *Quattordicesimo Rapporto sulla comunicazione. I media e il nuovo immaginario collettivo.* Roma: Franco Angeli.

Cinelli, M., Cresci, S., Galeazzi, A., Quattrociocchi, W., & Tesconi, M. (2020). The limited reach of fake news on Twitter during 2019 European elections. In *PloS one, 15(6), e0234689.*

Cinelli, M., Quattrociocchi, W., Galeazzi, A. et al. (2020), The COVID-19 social media infodemic. In *Sci Rep 10, 16598*

Conroy, N.J., Rubin, V.L., Chen, Y. (2015), Automatic deception detection: Methods for finding fake news. In *Proceedings of the Association for Information Science and Technology, 52 (1), pp. 1-4.*

Corbetta, P. (1992), *Metodi di analisi multivariata per le scienze sociali*. Bologna: Il Mulino.

Corbu, N. e Negrea-Busuioc, E. (2020). Populism Meets Fake News: social media, Stereotypes, and Emotions. In B. Krämer e Holtz-Bacha, C. (a cura di) *Perspectives on Populism and the Media.* Avenues for Research, Baden-Baden: Nomos (181-201).

Creech, B. (2020). Fake news and the discursive construction of technology companies' social power. In *Media, Culture & Society, 42(6), 952-968.*

Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., & Quattrociocchi, W. (2017). Mapping social dynamics on Facebook: The Brexit debate. In *Social Networks, 50, 6-16.*

Dey, A., Rafi, R. Z., Parash, S. H., Arko, S. K., & Chakrabarty, A. (2018). Fake news pattern recognition using linguistic analysis. In *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 305-309).* IEEE.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. In *Digital journalism, 3(3), 398-415.*

Di Franco, G. (2010). *Il campionamento nelle scienze umane. Teoria e pratica*. Milano: Franco Angeli.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. In *Machine Learning, 29(2), 103-130.*

Engesser, S., Ernst, N., Esser, F., & Büchel, F. (2017). Populism and social media: How politicians spread a fragmented ideology. *Information, communication & society, 20(8), 1109-1126.*

Erosheva, E., Bayesian estimation of the grade of membership model. *Bayesian Statistics, vol. 7, pp. 501-510, 2003.*

Faccani, R. e Marzaduri, M. (a cura di) (1975). *Lotman. Tipologia della cultura*. Milano: Bompiani.

Faggiano, M. P. (2016). *L'analisi del contenuto di oggi e di ieri: testi e contesti on e offline*. Roma: Franco Angeli.

Festinger, L. (1957), *Teoria della dissonanza cognitiva*, (trad. italiana). Roma: Franco Angeli.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. In *Communications of the ACM, 59(7), 96-104.*

Ferraris, M. (2012) *Manifesto del nuovo realismo*. Bari: Laterza Editori

Foucault, M. (1971) *L'ordine del discorso e altri interventi*. Edizione Italiana (2004). Torino: Einaudi Editore.

Franklin B. (1997), *Newszak and News Media*, London: Hodder Education.

Fuchs, C. (2012). La politica economica dei social media. In *Sociologia della Comunicazione 43/2012*. Roma: Franco Angeli.

Fuchs, C. (2014). Social media and the public sphere. In *Triple C 12(1): 57–101.*

Gabielkov M., Ramachandran A., Chaintreau A., Legout A. (2016), Social Clicks: What and Who Gets Read on Twitter? In *ACM SIGMETRICS/IFIP Performance 2016, Antibes Juan-les-Pins, France.*

Gagliano, G. (2017). *Deception. Disinformazione e propaganda nelle moderne società di massa*. Fuoco Edizioni

Gherghi, M., & Lauro, C. (2004). *Appunti di analisi dei dati multidimensionali. Metodologia ed Esempi.* Napoli: RCE Multimedia.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* London*:* Harvard University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis, 21(3), 267-297.*

Grzesiak-Feldman, M. (2013), The Effect of High-Anxiety Situations on Conspiracy Thinking. In *Current Psychology, 32(1), pp. 100–118.*

Habermas, J., Lennox, S., & Lennox, F. (1974). The public sphere. An encyclopedia article (1964). In *New German Critique, (3), 49-55.*

Hagey, K. and Horwitz, J. (2021). Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. In *The facebook files. A Wall Street Journal Investigation*. Wall Street Journal.

Hartley, J. (2009). Journalism and Popular Culture. In *Wahl-Jorgensen, Hanitzsch, pp.310-325*

Hasan, K. A., Sabuj, M. S., & Afrin, Z. (2015). Opinion mining using naive bayes. *In 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 511-514.*

Hay, A. M. (1988). The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing, 9(8), 1395-1398.*

Heilmann, L. (a cura di) (2008). *Roman Jakobson. Saggi di linguistica generale*. Milano: Feltrinelli.

Hermida, A., & Thurman, N. (2008). A clash of cultures: The integration of user-generated content within professional journalistic frameworks at British newspaper websites. *Journalism practice, 2(3), 343-356.*

Hochman, N., & Manovich, L. (2013). Zooming into an Instagram City: Reading the local through social media. *First Monday, 18(7).*

Hong, L. E Davison, B. D. (2010), Empirical study of Topic Modeling in Twitter. *In Proceedings of the first workshop on social media analytics, pp. 80-88.*

Horne, B. D., Adali, S. (2017). This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *The 2nd International Workshop on News and Public Opinion at ICWSM.*

Horwitz, J. e al. (2021). *The facebook files. A Wall Street Journal Investigation,* Wall Street Journal.

Jagers, J., & Walgrave, S. (2007). Populism as political communication style: An empirical study of political parties' discourse in Belgium. *European journal of political research, 46(3), 319-345.*

Jamieson, A., Solon, O., (2016). *Facebook to begin flagging fake news in response to mounting criticism.* The Guardian.

Karlsson, M., & Sjøvaag, H. (2016). Content Analysis and online news. In *Digital Journalism, 4(1), 177–192.*

Khaldarova, I., & Pantti, M. (2016). Fake news: The narrative battle over the Ukrainian conflict. *Journalism practice, 10(7), 891-901.*

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American political science Review, 107(2), 326-343.*

Kiriya I. (2021), From "Troll Factories" to "Littering the Information Space": Control Strategies Over the Russian Internet. In *Media Control Revisited. Challenges, Bottom-Up Resistance and Agency in the Digital Age. Vol. 9, No. 4.*

Klapp, O. E. (1986). Overload and Boredom: Essays on the Quality of Life. In *The Information Societv.* Westport, CT: Greenwood Press.

Kligler-Vilenchik, N., Baden, C., & Yarchi, M. (2020). Interpretative polarization across platforms: How political disagreement develops over time on Facebook, Twitter, and WhatsApp. In *Social Media + Society 6(3).*

Knorr Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge.* Harvard: Harvard University Press.

Kohut, A., Morin, R., & Keeter, S. (2007). What Americans know: 1989-2007, Public knowledge of current affairs little changed by news and information revolutions. *PEW Research Center, April, 15.*

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.

Kushner S., Albertson B. G. (2013) Anxiety, Immigration, and the Search for Information. In *Political Psychology Volume35, Issue2.*

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Van Alstyne, M. (2009). Life in the network: the coming age of Computational Social Science. In *Science (New York, NY), 323(5915), 721*

Lebart, L., Salem A. (1988), *Analyse statistique des donnèes textuelles,* Dunod, Paris.

Lebart, L., Pincemin, B., & Poudat, C. (2019). *Analyse des données textuelles.* Presses de l'Université du Québec.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. *In European conference on machine learning (pp. 4-15).* Springer, Berlin, Heidelberg.

Levin, S. (2017a). *Facebook told advertisers it can identify teens feeling "insecure" and "worthless"*. The Guardian.

Levin, S. (2017b). *Facebook promised to tackle fake news. But the evidence shows it's not working*, The Guardian.

Lorusso, A. M. (2015). L'abito in Peirce. Una teoria non sociologica per la semiotica della cultura. In *Rivista italiana di filosofia del linguaggio.*

Lorusso, A. M. (2018). *Postverità: Fra reality tv, social media e storytelling*. Gius. Laterza & Figli Spa.

Losito, G. (1996). *L'analisi del contenuto nella ricerca sociale* (Vol. 1). Milano: Franco Angeli.

Lupton, D. (2014). *Digital sociology*. Routledge.

Lyotard, J. F. (1979). *La condition postmodeme: Rapport sur le savoir*. Paris: Minuit, 109.

Maddalena, G. (a cura di) (2008), *Peirce. Scritti scelti.* Torino: UTET Libreria.

Mancini, P. (2000). *Il sistema fragile.* Roma: Carocci Editori.

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Marietta, M., Barker, D. C., & Bowser, T. (2015). Fact-checking polarized politics: does the fact-check industry provide consistent guidance on disputed realities? In *The Forum (Vol. 13, No. 4, pp. 577-596).*

Marradi, A. (1996). Metodo come arte. In *Quaderni di Sociologia. XL, 10, pp. 71-92.*

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. *In AAAI-98 workshop on learning for text categorization (Vol. 752, No. 1, pp. 41-48).*

McMillan, S. J. (2000). The microscope and the moving target: The challenge of applying content analysis to the World Wide Web. *Journalism & Mass Communication Quarterly, 77(1), 80-98.*

Merrill, J.B., Oremus W. (2021). *Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation.* The Washington Post.

Metsis, V., Androutsopoulos, I., & Paliouras, G. (2006). Spam filtering with naive bayes-which naive bayes? In *CEAS, Vol. 17, pp. 28-69.*

Mocanu, D., Rossi, L., Zhang, Q., Karsai, M., & Quattrociocchi, W. (2015). Collective attention in the age of (mis)information. In *Computers in Human Behavior, 51, 1198-1204.*

Morris, C. (2020). *Coronavirus: False claims by politicians debunked.* BBC News

Mourão, R. R., & Robertson, C. T. (2019). Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information. In *Journalism Studies. 1,19.*

Mueller, H., & Rauh, C. (2018). Reading between the lines: Prediction of political violence using newspaper text. American *Political Science Review, 112(2), 358-375.*

Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. In *International Conference on Intelligent Data Engineering and Automated Learning (194-201).* Berlin: Springer.

Natale, P., & Airoldi, M. (2017). *Web & Social Media: le tecniche di analisi*. Maggioli Editore.

Neuman, W.R. (2007). *The affect effect: Dynamics of emotion in political thinking and behavior*. University of Chicago Press.

Noelle-Neumann, E. (2002). *La spirale del silenzio. Per una teoria dell'opinione pubblica.* Milano: Meltemi Editore.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* Penguin UK.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. In *Annual Review of Psychology, 54(1), 547-577.*

Pierri, F., Artoni, A., & Ceri, S. (2020). Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections. In *PloS one, 15(1), e0227821.*

Pira, F., & Altinier, A. (2018). *Giornalismi: la difficile convivenza con fake news e misinformation*. Libreriauniversitaria Edizioni.

Provost, F., & Kohavi, R. (1998). Guest editors' introduction: On applied research in machine learning. In *Machine Learning, 30(2), 127-132.*

Quattrociocchi, W. (2021), L'era della (dis)informazione. *I Quaderni de le Scienze, Nr. 14, pp. 3-9*

Quattrociocchi, W., & Vicini, A. (2016). *Misinformation. Guida alla società dell'informazione e della credulità*. Roma: Franco Angeli.

Reinert, A. (1983). Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. In *Cahiers de l'analyse des données 8.2. 187-198.*

Renjith, R. (2017). The effect of information overload in digital media news content. In *Communication and Media Studies, 6(1), 73-85.*

Rieder, B. (2020). *Engines of order: A mechanology of algorithmic techniques*. Amsterdam: Amsterdam University Press.

Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence Vol. 3, No. 22. (41-46).*

Ritzer, G. (2017). *La McDonaldizzazione della produzione,* Roma: Lit Edizioni.

Ritzer, G., & Jurgenson, N. (2010). Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer'. In *Journal of consumer culture, 10(1), 13-36.*

Rogers, R. (2013). *Digital Methods*. MIT Press.

Rogers, R. (2021). *Mainstreaming the Fringe: How Misinformation Propagates on social media.* Amsterdam: Amsterdam University Press.

Rogers, R. (2018). Social Media Research After the Fake News Debacle. In *Partecipazione e Conflitto, The Open Journal of Sociopolitical Studies, (11, 2).*

Rogers, R. (2020). The scale of Facebook's problem depends upon how 'fake news' is classified. In *Misinformation Review, 1(6).* The Harvard Kennedy School (HKS)

Rogers, R. (2021). Marginalizing the Mainstream: How Social Media Privilege Political Information. In *Frontiers in big Data, 4.*

Rubin, V.L., Chen, Y., Conroy, N.J. (2015), Deception detection for news: three types of fakes. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (ASIST 2015).* Article 83, p. 4, American Society for Information Science, Silver Springs

Rubin, V. L., et al (2016). Fake news or truth? Using satirical cues to detect potentially misleading news. In *Proceedings of NAACL-HLT*

Rouvroy, A., & Berns, T. (2013). Gouvernementalité algorithmique et perspectives d'émancipation. *Réseaux, (1), 163-196.*

Sartre, J. P. (1943). *L'essere e il nulla. La condizione umana secondo l'esistenzialismo,* Vol. 8. (ed. italiana). Milano: Il saggiatore.

Schulte, S. (2013). *Cached: The Internet in Global Popular Culture.* New York: New York University Press.

Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R. In *Journal of Open Source Software, 1(3), 37.*

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach.* O'Reilly Media, Inc.

Sjøvaag, H., & Stavelin, E. (2012). Web media and the quantitative content analysis: Methodological challenges in measuring online news content. In *Convergence, 18(2), 215-229.*

Sjøvaag, H., Moe, H., & Stavelin, E. (2012). Public service news on the Web: A large-scale Content Analysis of the Norwegian Broadcasting Corporation's online news. In *Journalism Studies, 13(1), 90-106.*

Sorrentino, C. (2003). *Il giornalismo in Italia. Aspetti, processi produttivi, tendenze.* Roma: Carrocci Editore.

Splendore, S. (2017). *Giornalismo ibrido. Come cambia la cultura giornalistica italiana*. Roma: Carrocci Editore.

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. In *Business Information Review, 34(3), 150-160.*

Sunstein, C. R. & Vermeule, A. (2009). Conspiracy theories: causes and cures. In *Journal of Political Philosophy. 17 (2): 202–227.*

Sunstein, C. R. (2014). *Conspiracy theories and other dangerous ideas*. Simon and Schuster.

Tandoc, Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. In *Digital journalism, 6(2), 137-153.*

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. In *Journal of language and social psychology, 29(1), 24-54.*

Tipaldo, G. (2014). *L'analisi del contenuto e i mass media*. Bologna: Il Mulino.

Thomas, S. L., Nafus, D., & Sherman, J. (2018). Algorithms as fetish: Faith and possibility. *Algorithmic work. Big Data & Society, 5(1).*

Thompson, J.B. (1995). *The Media and Modernity.* Cambridge: Polity Press

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PloS one, 13(9), e0203958.*

Tsur, O., Calacci, D., & Lazer, D. (2015). A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd Annual Meeting of the ACL & the 7th International Joint Conference on NLP,* Beijing, China.

Twitter (2019). *Glorification of violence policy.* Twitter Help Center. Available at: https://help.twitter.com/en/rules-and-policies/glorification-of-violence.

Capecchi, V., Livolsi. M. (1971). *La stampa quotidiana in Italia.* Milano: Bompiani.

Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media.* Oxford University Press.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world.* Oxford University Press.

Venturini, T. & Rogers, R. (2019). 'Api-based Research' or how can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica. In *Data Breach, Digital Journalism, 7:4, 532-540.*

Venturini, T. (2019). From fake to junk news: The data politics of online virality. In *Data Politics (pp. 123-144).* Routledge.

Vicario, M. D., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and fake news: Early warning of potential misinformation targets. In *ACM Transactions on the Web (TWEB), (13 2), 1-22.*

Walker, S., Mercea D., Bastos, M. (2019). The disinformation landscape and the lockdown of social platforms, Information. In *Communication & Society, (22 11), 1531-1543.*

Wardle, C. (2017). *Fake news. It's complicated.* First Draft. Disponibile al link: https://firstdraftnews.org/articles/fake-news-complicated/

Wardle, C. (2018). The Need for Smarter Definitions and Practical, Timely Empirical Research on Information Disorder. In *Digital Journalism, (68), 951-963.*

Wardle, C. (2018). *Information disorder: The essential glossary.* Harvard, MA: Shorenstein Center on Media, Politics, and Public Policy, Harvard Kennedy School.

Wardle, C., & Derakhshan, H. (2017). Information Disorder: Toward an interdisciplinary framework for rsearch and policy making. In *Council of Europe Report, 27.*

Wardle, C., & Derakhshan, H. (2018). Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. In *Journalism 'fake news'& disinformation. 43-54.* Paris: Unesco.

Weare, C., & Lin, W. Y. (2000). Content analysis of the World Wide Web: Opportunities and challenges. *Social science computer review, 18(3), 272-292.*

Wickham, H., & Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data.* O'Reilly Media, Inc.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. In *IEEE transactions on evolutionary computation, 1(1), 67-82.*

Urbano, L. (2019). Segreti visibili. Riflessioni sul complottismo nell'era dei social media. In *Psiche. Rivista di cultura psicoanalitica. 2/2019, pp. 415-424.*

Zhang, H. (2004). The optimality of naive Bayes. In *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, Miami Beach, Florida, USA 1(2), 3.*

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., & Quattrociocchi, W. (2017*).* Debunking in a world of tribes. In *PloS one, 12(7), e0181821.*

INDICE GRAFICI

INDICE TABELLE

INDICE FIGURE

ALLEGATO 1 – Scheda di analisi dei contenuti disinformativi

RISK	DISINFORMATION SPECTRUM	
HIGH RISK	10	Deleted content
	1	Fabricated content
	2	Manipulated content
	3	Impostor content
	4	False context
MEDIUM RISK	5	Misleading content
	6	False connection
LOW RISK	7	Satire or Parody
	8	Nothing to Register
	9	Other

CONTENT CONSTRUCTION			MULTIPLE CHOICE
1	FALSE	a	False fact
		b	False title
		c	False statistics / data
		d	False expert testimony
		e	False at the date of publication
		f	False accusations
2	INACCURATE	g	Inaccurate fact
		h	Inaccurate title
		i	Inaccurate statistics / data
		l	Inaccurate expert testimony
		m	Fact inaccurate at the date of publication
		n	Unfounded accusations
3	OUT OF CONTEXT	o	Manipulated image
		p	Opposite image
		q	Image out of context
		r	Selective copying (false and / or misleading material selectively combined with true information)
		s	Out of context use of reliable sources
		t	Out of context use of parody content
		u	Isolated case used as a general question
		v	Recycled news
		z	Time out of context, false news at the time of
		w	Place out of context
4	CONSPIRACY THEORY		
5	OTHER		
6	NOTHING TO REGISTER		

ALLEGATO 2 – Scheda di Analisi del Contenuto per immagini

IMAGE CONTENT ANALYSIS SHEET					
1	Years	1. Years 2016 2. Years 2017 3. Years 2018	4. Years 2019 5. Years 2020		
2	Month	1. January 2. February 3. March 4. April 5. May 6. June	7. July 8. August 9. September 10. October 11. November 12. December		
3	Key Word				
4	Engagement	1. from 0 to 5.000 2. from 5001 to 10.000 3. from 10.001 to 20.000 4. from 20.001 to 50.000 5. > di 50.000			
5	Disinformative Risk	1.High	2. Medium		
6	Disinformative spectrum	1. Fabricated content 2. Manipulated content 3. Impostor content	4. False context 5. Misleading content 6. False connection	7. Other	
7	Representation type	1. Landscape 2. Public Place 3. Private Place	4. Screenshot 5. Graphic image 6. Neutral background	7. Other	
8	Number of characters	1. One 2. Two 3. Small group	4. Multitude 5. No characters		
9	Character type (1)	1. Political World 2. Sports World 3. Show Biz 4. Religious World	6. Scientific World 7. Influencer 8. Chronicle Protagonist (indicate) 9. Other (Indicate)		
9.b	Character Type (2)	5. Cultural World 1. Political World 2. Sports World 3. Show Biz 4. Religious World 5. Cultural World	6. Scientific World 7. Influencer 8. Chronicle Protagonist (indicate) 9. Other (Indicate)		
10	Using the image	1. Instrumental	2. Emotional	3. Promoted	4. Informative
11	Presence of text	1. YES	2. NO		
12	Text				
13	Textual Expedient	1. Politic 2. Technical 3. Cultural	4. Aesthetic 5. Scientific 6. Conspiracy	7. Religious 8. Sportif 9. Other	
14	Style and tone of communication	1. Assertive	2. Argumentative	3. Aggressive	4. Neutral
15	Lens of image text	1. Get done	2. Let know	3. Make feel	
16	Political position expressed	1. Predominantly right-wing Political Position 2. predominantly left- wing Political Position		3 Predominantly in the center 4. Political Position Rejected	5. No political position expressed