

NEWS VS. FAKE NEL SISTEMA DELL'INFORMAZIONE

INTERIM REPORT
INDAGINE CONOSCITIVA
DEL. 309/16/CONS



AUTORITÀ PER LE
GARANZIE NELLE
COMUNICAZIONI

SERVIZIO ECONOMICO-STATISTICO



SERVIZIO ECONOMICO-STATISTICO

Anno 2018, mese di novembre

SOMMARIO

EXECUTIVE SUMMARY	i
PREMessa	1
1. LA PRODUZIONE DI INFORMAZIONE	6
1.1. INTRODUZIONE	7
1.2. LA QUANTITÀ E LA QUALITÀ DELL'INFORMAZIONE PRODOTTA IN ITALIA	9
1.3. LE TEMATICHE DELL'INFORMAZIONE	19
2. LA PRODUZIONE DI DISINFORMAZIONE	24
2.1. INTRODUZIONE	25
2.2. L'ENTITÀ DELLA DISINFORMAZIONE PRODOTTA IN ITALIA	28
2.3. LE TEMATICHE DELLA DISINFORMAZIONE	33
3. LA DIFFUSIONE DI NOTIZIE REALI E FALSE	47
3.1. INTRODUZIONE	48
3.2. IL CICLO DI VITA E LE MODALITÀ DI DIFFUSIONE DELLE NOTIZIE REALI	51
3.3. LE MODALITÀ DI DIFFUSIONE DELLA DISINFORMAZIONE E IL CICLO DI VITA DELLE NOTIZIE FALSE: AMBITO SOGGETTIVO E OGGETTIVO	62
4. LA PROPAGAZIONE DI NOTIZIE REALI E FALSE SULLE PIATTAFORME ONLINE	74
4.1. INTRODUZIONE	75
4.2. IL CONSUMO DI NOTIZIE SULLE PIATTAFORME ONLINE E LA POLARIZZAZIONE	78
4.3. L'ATTUAZIONE DELLE STRATEGIE DI DISINFORMAZIONE SULLE PIATTAFORME ONLINE	88
4.4. I MECCANISMI DI INTERAZIONE CON I CONTENUTI DI DISINFORMAZIONE SULLE PIATTAFORME ONLINE	94
APPENDICE METODOLOGICA	107

INDICE DELLE FIGURE

Figura 1.1 - Andamento mensile dell'informazione prodotta in Italia	11
Figura 1.2 - Funzione di produzione dell'informazione (stima, valori medi mensili)	12
Figura 1.3 - Intensità produttiva dei giornalisti nel mese medio	14
Figura 1.4 - Relazione tra intensità produttiva del giornalista e reputazione del mezzo	16
Figura 1.5 - Principale fonte di informazione per i cittadini italiani (% popolazione)	17
Figura 1.6 - Stima del consumo effettivo di informazione tra i fruitori dei mezzi	18
Figura 1.7 - Varietà e livello di conoscenza specialistica delle tematiche trattate	20
Figura 1.8 - Distribuzione dell'offerta e della domanda di informazione, per categoria	22
Figura 1.9 - Livello di conoscenza specialistica dei giornalisti e domanda potenziale dei cittadini, per categoria	23
Figura 2.1 - La filiera dei contenuti <i>fake</i>	26
Figura 2.2 - Andamento mensile della disinformazione prodotta in Italia	29
Figura 2.3 - Quantità incrementale di contenuti <i>fake</i> prodotti nel mese	30
Figura 2.4 - Incidenza dei contenuti <i>fake</i> nel sistema informativo	31
Figura 2.5 - Distribuzione dell'offerta di contenuti <i>fake</i> per categoria	34
Figura 2.6 - Principali tematiche oggetto dei contenuti <i>fake</i> (2018)	37
Figura 2.7 - Termini salienti dell'argomento "Politica/Governo" (2018)	38
Figura 2.8 - Termini salienti dell'argomento "Diritti" (2018)	39
Figura 2.9 - Termini salienti dell'argomento "Economia" (2018)	40
Figura 2.10 - Termini salienti dell'argomento "Salute e Ambiente" (2018)	41
Figura 2.11 - Termini salienti dell'argomento "Famiglia e Fede" (2018)	42
Figura 2.12 - Termini salienti dell'argomento "Cronaca" (2018)	43
Figura 2.13 - Termini salienti dell'argomento "Esteri" (2018)	44
Figura 2.14 - Termini salienti dell'argomento "Scienza" (2018)	45
Figura 2.15 - Termini salienti dell'argomento "Immigrazione" (2018)	46
Figura 3.1 - Ciclo di vita di una notizia	52
Figura 3.2 - Ciclo di vita della notizia "terremoto centro Italia"	53
Figura 3.3 - Ciclo di vita della notizia "Trump eletto presidente USA"	54
Figura 3.4 - Indici di diffusione di una notizia, per mezzo	57
Figura 3.5 - Concentrazione della distribuzione di una notizia nel tempo, per mezzo	59
Figura 3.6 - Indici di diffusione di una notizia, per categoria	61
Figura 3.7 - Indici di diffusione delle notizie reali e false su fonti di disinformazione	64
Figura 3.8 - Ciclo di vita di una notizia falsa	66
Figura 3.9 - Ciclo di vita della <i>fake news</i> "nuova era glaciale nel 2019"	67
Figura 3.10 - Ciclo di vita della <i>fake news</i> "incidente centrale nucleare di Krško"	68
Figura 3.11 - Indici di diffusione delle notizie reali e false nel sistema dell'informazione	70
Figura 3.12 - Indici di diffusione di una notizia falsa, per categoria	71
Figura 4.1 - Eterogeneità del consumo informativo, per durata di interazione (<i>lifetime</i> ; 4.1a) e livello di coinvolgimento degli utenti (<i>activity</i> ; 4.1b)	81
Figura 4.2 - Struttura delle comunità di pagine, per tipologia di azione informativa	82
Figura 4.3 - Polarizzazione degli utenti	83
Figura 4.4 - Struttura delle comunità di "pagine Brexit"	85

Figura 4.5 - Distribuzione della polarizzazione degli utenti verso le comunità di “pagine Brexit”	86
Figura 4.6 - Andamento mensile delle interazioni degli utenti con i <i>post</i> di disinformazione	91
Figura 4.7 - Andamento mensile dei <i>post</i> di disinformazione e delle relative interazioni degli utenti, per tipologia di contenuto	92
Figura 4.8 - Distribuzione della polarizzazione degli utenti di Facebook per “scienza” e “cospirazione” (Italia)	100
Figura 4.9 - Distribuzione della polarizzazione degli utenti di Facebook per “scienza” e “cospirazione” (USA)	100
Figura 4.10 - Distribuzione della polarizzazione degli utenti di Facebook e YouTube per “scienza” e “cospirazione” (USA)	101
Figura 4.11 - Stati emotivi in funzione del numero di commenti ai <i>post</i> discussi da entrambe le narrazioni	104
Figura 4.12 - Interazioni degli utenti con notizie di <i>debunking</i>	105
Figura 4.13 – Interazioni dei cospirazionisti prima e dopo il <i>debunking</i>	106
Figura A.1 – Funzionamento della piattaforma	109
Figura A.2 – Classificatore automatico dei contenuti informativi in categorie	111



EXECUTIVE SUMMARY

L'adozione di una metodologia innovativa basata sull'osservazione diretta di milioni di dati

Il presente Rapporto, che costituisce un primo esito dell'Indagine conoscitiva su "piattaforme digitali e sistema dell'informazione", ha l'obiettivo di delineare, grazie all'osservazione diretta di milioni di dati, gli scenari informativi che caratterizzano il sistema italiano.

Tenendo a mente la tutela del pluralismo come motivazione esplorativa, il Rapporto persegue l'intento di svolgere una valutazione circa la robustezza del sistema informativo e le *performance* delle piattaforme online e degli altri mezzi nella produzione e divulgazione di informazione, nonché di compiere un ulteriore passo in avanti nella comprensione dell'entità, delle peculiarità e modalità di diffusione della disinformazione.

Uno studio di questo tipo ha presupposto l'impiego di una metodologia sperimentale, incentrata sull'utilizzo e l'integrazione di una molteplicità di fonti di dati sulle diverse componenti dei sistemi dell'informazione e della disinformazione: testate informative, fonti di disinformazione, giornalisti, e cittadini/fruitori di notizie.

In particolare, l'analisi complessiva dell'intero ecosistema è stata realizzata su enormi moli di dati, provenienti da *dataset* differenti e unici al mondo, riconducibili a:

- l'intero contenuto testuale estrapolato, attraverso la piattaforma (sviluppata da *Volocom Technology*) di cui si è dotata l'Autorità, da circa 35 milioni di documenti generati in Italia, in un arco temporale di più due anni, da 1.800 fonti informative (canali televisivi e radiofonici, quotidiani, siti web di editori tradizionali, testate esclusivamente online, e relative pagine e *account* di social network), e fonti di disinformazione (siti web e pagine/*account* social) individuate come tali da soggetti esterni specializzati in attività di *debunking*;
- un campione di circa 700 principali notizie (reali e false), rappresentativo di tutte le categorie di generi trattati, considerando un intervallo temporale di un anno;
- informazioni, raccolte nell'ambito della II edizione dell'Osservatorio Agcom sul giornalismo, sulle modalità di produzione di contenuti informativi da parte di un ampio e rappresentativo campione di professionisti (circa 2.000) attivi in Italia;

- i risultati di una survey sul consumo di informazione svolta per l'Autorità da GfK Italia su un campione di oltre 14.000 individui, rappresentativo della popolazione italiana;
- decine di milioni di *account social* pubblici su cui sono stati applicati modelli di *big data analytics* per esaminare le modalità di fruizione dell'informazione e i meccanismi di interazione degli utenti attraverso le piattaforme online. La parte del Rapporto dedicata a questi specifici aspetti è stata redatta sulla base di un incarico di ricerca affidato dall'Autorità a Walter Quattrociocchi, che da diversi anni si occupa dell'analisi scientifica del fenomeno, pubblicando studi nazionali e internazionali.

Il Rapporto, fondato su tale impostazione metodologica, si inserisce tra le attività di vigilanza e monitoraggio del settore dei media, partendo dalle evidenze rilevate dall'Autorità in altre occasioni. Lo studio, dunque, in considerazione delle già riscontrate relazioni di interdipendenza tra sistema informativo e sistema della disinformazione, non può che soffermarsi sull'analisi approfondita di entrambe le componenti.

Più volte, infatti, l'Autorità ha osservato come, a livello nazionale e globale, fenomeni patologici di disinformazione, tendano ad annidarsi lì dove il sistema dell'informazione fallisce:

- la difficoltà di monetizzazione dei contenuti e la perdurante riduzione degli investimenti in informazione; l'evidenziata e connessa regressione nell'uso di meccanismi di verifica nell'ambito della professione giornalistica; la ristrettezza dei tempi dell'informazione online, sia nella fase di produzione sia in quella di consumo (che rischia di tradursi in una minor attendibilità nella raccolta di informazioni, oltre che in un minor livello di attenzione dell'utente) sono atte a compromettere l'adeguatezza dell'offerta informativa sul piano dell'accuratezza, dell'approfondimento e della copertura delle notizie. In termini più ampi, sono alla base della diffusa perdita di reputazione e fiducia accordata dai cittadini al sistema informativo tradizionale.

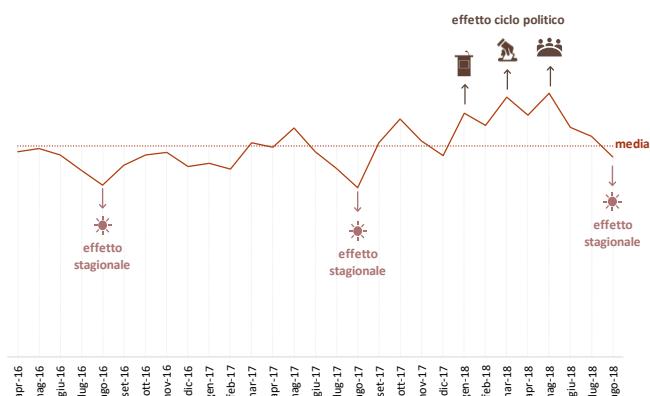
È in questo contesto che i cittadini rischiano di affidarsi sempre più a fonti informative alternative e non qualificate, che sono spesso alla base di strategie di disinformazione.

Le principali evidenze sulla produzione e diffusione di informazione

L'analisi dinamica sulla quantità di informazione prodotta in Italia rivela un innalzamento dell'ammontare medio del volume informativo immesso nel sistema, che raggiunge il valore massimo nel 2018. D'altra parte, l'andamento mensile suggerisce come sia possibile rintracciare gli effetti determinati dall'azione di almeno due fattori sulla quantità di offerta informativa prodotta dai media:

- il primo, di natura ricorrente (stagionalità), che si manifesta nella riduzione subita ogni anno dal volume informativo nel periodo estivo;
- il secondo, di carattere contingente (ciclo politico), che si rinvie nell'incremento di informazione prodotta in concomitanza del periodo che copre la campagna elettorale, lo svolgimento delle ultime elezioni politiche svoltesi in Italia (nel marzo 2018) e la successiva formazione del nuovo governo.

Andamento mensile dell'informazione prodotta in Italia



L'andamento medio giornaliero della quantità di contenuti informativi prodotti in merito a una singola notizia individua il ciclo di vita di quest'ultima, che può avere inizio anche prima del giorno esatto (t_0) in cui avviene il fatto relativo alla medesima. Tuttavia, è nei giorni che vanno dal t_0 in poi che si verifica la maggiore distribuzione dell'offerta informativa:

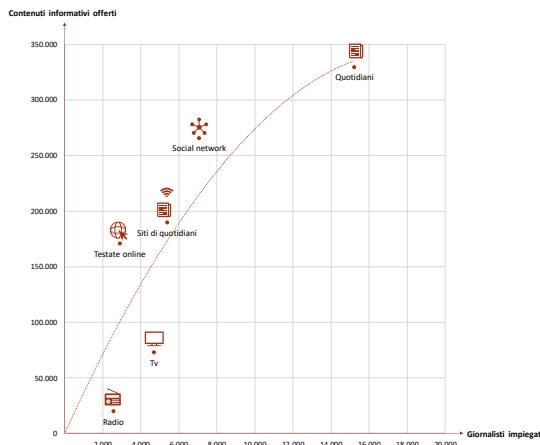
- la concentrazione di contenuti sulla notizia, che tendenzialmente subisce un'impennata in corrispondenza del t_0 , registra il suo massimo nei due giorni successivi a quello in cui ha avuto luogo l'accadimento, quando tutti i mezzi arrivano a diffonderla.



L'esame comparato della quantità di contenuti informativi prodotti e delle risorse giornalistiche impiegate, inoltre, consente di operare una distinzione tra mezzi:

- da un lato, le fonti informative online (siti di quotidiani, testate online, social network), che, a parità di risorse professionali utilizzate e di altre condizioni, producono un'offerta di informazione maggiore, evidenziando un sovrautilizzo della forza giornalistica impiegata, atto a riverberarsi negativamente sull'accuratezza e l'approfondimento del contenuto generato attorno alla notizia. Peraltro, nel caso dei social network, emerge chiaramente la rapidità con cui si passa dalla trattazione di una notizia ad un'altra, indice di una maggiore superficialità nell'esposizione degli accadimenti rispetto agli altri mezzi;
- dall'altro lato, quotidiani, Tv e radio, che presentano una minore intensità produttiva dei giornalisti impiegati in relazione alla quantità di contenuti informativi offerti, e, quindi, un maggior livello atteso di qualità e approfondimento dell'informazione prodotta, che, nel caso delle testate quotidiane e delle emittenti televisive, si accompagna a una diffusione e una copertura giornaliera della singola notizia superiori alla media.

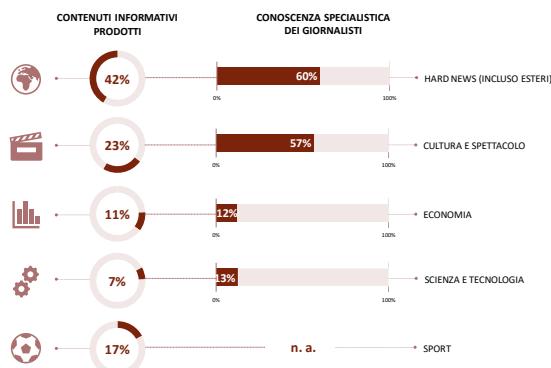
Funzione di produzione dell'informazione



Per quel che attiene alle categorie tematiche dell'informazione veicolata, si nota che:

- non soltanto l'informazione specializzata viene offerta in misura più contenuta rispetto a quella più generalista, ma è anche prodotta per la gran parte da professionalità che non possiedono competenze specialistiche nelle materie di riferimento.

Varietà e livello di conoscenza specialistica delle tematiche trattate



Peraltro, paragonando la distribuzione dell'offerta informativa a quella della domanda, si evidenzia l'esistenza di significativi squilibri:

- se per le "hard news" (politica, cronaca, esteri) e le notizie di "cultura e spettacolo" si osserva un eccesso di offerta, le categorie legate all'informazione specializzata, "scienza e tecnologia" in particolar modo, presentano, viceversa, un eccesso di domanda.

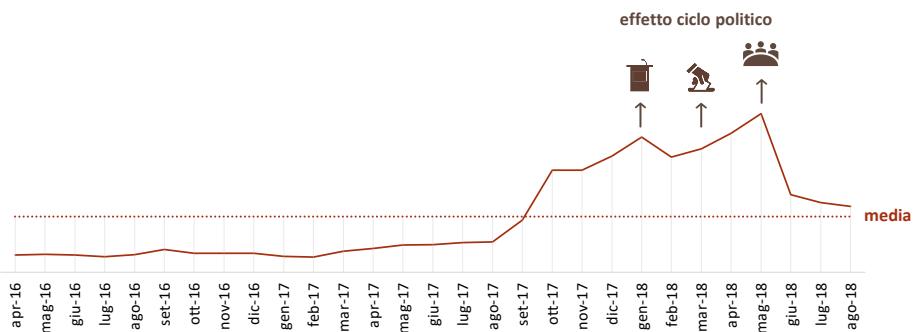
L'attuale proposta informativa appare, quindi, carente soprattutto nell'offerta di contenuti più specializzati o di nuovo interesse, laddove la forte richiesta da parte dei cittadini non risulta efficientemente soddisfatta né sotto il profilo quantitativo né sotto quello qualitativo dei contenuti prodotti.

Le principali evidenze sulla produzione, diffusione e propagazione di disinformazione

Le analisi dinamiche compiute sulla produzione di disinformazione mostrano un sistema nazionale che soffre la presenza di un volume di contenuti *fake* che:

- sembra essersi assestato su un valore mediamente più alto rispetto al passato,
- ha raggiunto il livello massimo durante l'ultimo ciclo politico, in corrispondenza delle elezioni politiche del 4 marzo 2018.

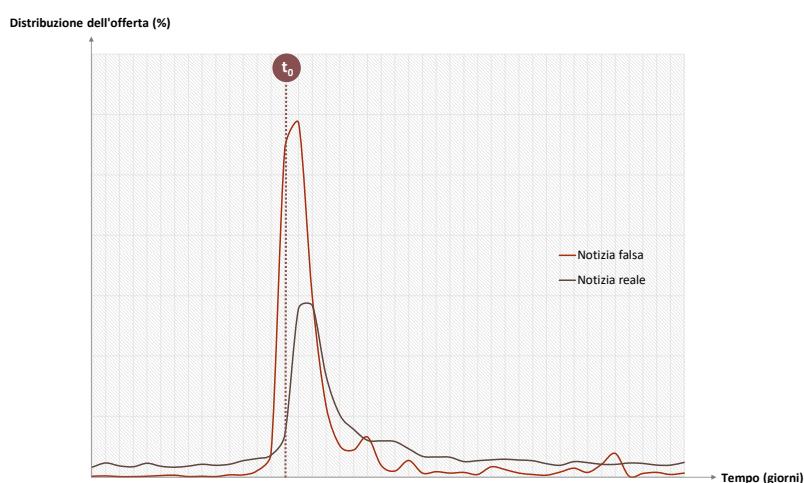
Andamento mensile della disinformazione prodotta in Italia



Il ciclo di vita di una singola notizia falsa si caratterizza essenzialmente per:

- la pressoché totale assenza di anticipazioni (ossia, contenuti *fake* diffusi prima del t_0) sui fatti oggetto della notizia falsa;
- una durata sensibilmente inferiore rispetto al ciclo di vita di una notizia reale, con una concentrazione decisamente più accentuata attorno al t_0 , che raggiunge il punto di massimo il giorno successivo al t_0 , per poi scendere velocemente verso valori prossimi allo zero.

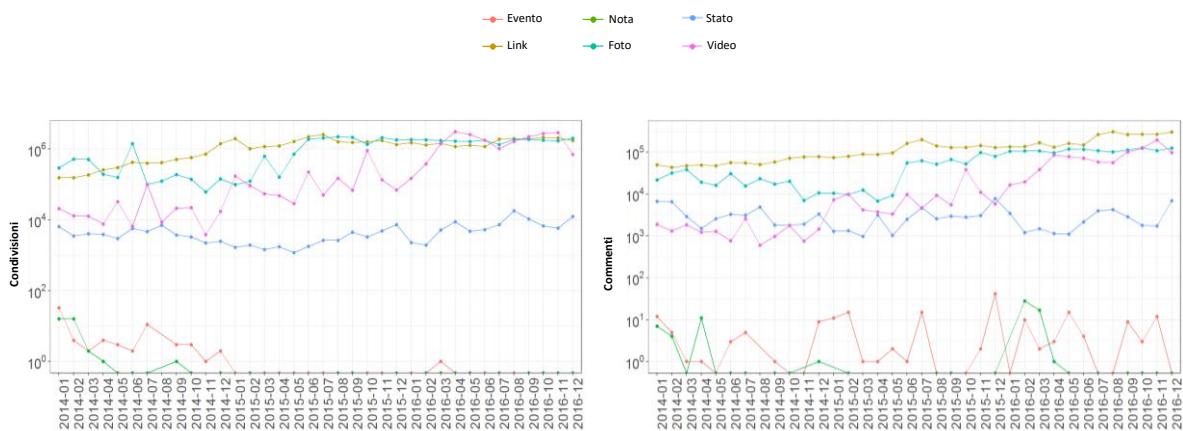
Ciclo di vita di una notizia falsa



In questo contesto, si riscontra che:

- le fonti di disinformazione (siti web, pagine e *account* social individuati come tali dai *debunker*) assumono un ruolo che consiste prevalentemente nel conferire impulso ai contenuti *fake* all'interno del sistema nazionale. La brevità del ciclo di vita delle singole notizie false diffuse tramite queste fonti e la concentrazione in pochi giorni della distribuzione dei relativi contenuti sono la spia stessa dell'intento di mettere in atto una strategia di disinformazione, prediligendo la trattazione di tante notizie diverse, evitando di approfondirne i contenuti;
- una volta innescata, la notizia falsa viene immessa e rilanciata nel sistema delle piattaforme online, anche attraverso l'inconsapevole contributo degli utenti, che la condividono e commentano sui social network.

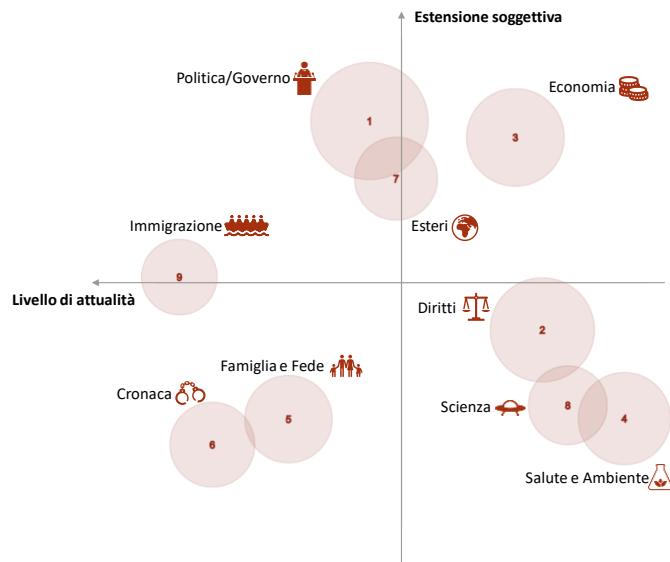
Interazioni degli utenti con i post di disinformazione, per tipologia di contenuto



Anche la scelta delle tematiche trattate dalle fonti di disinformazione si rivela collimante con l'intento di attivare meccanismi di propagazione virale sulle piattaforme online:

- in Italia, il 57% della produzione di contenuti *fake* riguarda argomenti di politica e cronaca, mentre circa il 20% tematiche di carattere scientifico; tutte tematiche che presentano un forte impatto emotivo e che possono essere divisive;
- l'analisi testuale dei contenuti *fake* prodotti dai siti di disinformazione nel 2018 conduce a individuare 9 argomenti principali (politica, diritti, economia, salute e ambiente, famiglia e fede, cronaca, esteri, scienza, immigrazione), la trattazione dei quali è peculiare: i temi sono affrontati in modo superficiale e impressionistico, mirando a stimolare gli stati d'animo delle persone.

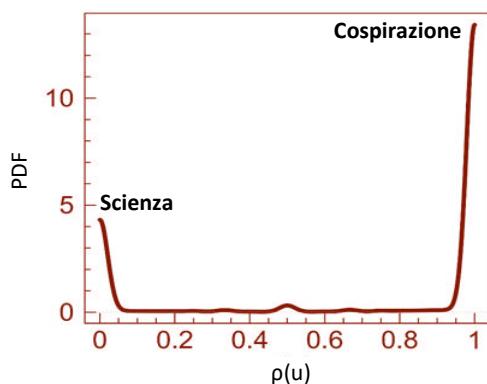
Principali tematiche oggetto dei contenuti *fake* nel 2018



La viralizzazione del contenuto *fake* è resa possibile dalle modalità di consumo informativo sulle piattaforme online, che, specie per le tematiche individuate come oggetto di disinformazione, avviene nell'ambito di comunità chiuse e distinte:

- i modelli di consumo informativo e l'interazione degli utenti con le notizie sulle piattaforme online sono caratterizzati da tendenza alla polarizzazione, esposizione selettiva, omofilia, e insorgenza di *echo chamber*. Gli utenti, pertanto, tendono a selezionare le informazioni che sono coerenti con il proprio sistema di credenze, formando gruppi polarizzati di persone con idee simili su narrazioni condivise, in cui le informazioni discordanti vengono ignorate;
- quanto emerge suggerisce come il processo che guida i meccanismi di acquisizione ed elaborazione delle informazioni non sia dettato soltanto da fenomeni cognitivi ma anche, e soprattutto, da pratiche culturali e sociali.

Distribuzione della polarizzazione degli utenti italiani di Facebook per “scienza” e “cospirazione”



PREMESSA

Il pluralismo informativo e la libertà di informazione costituiscono i principi essenziali della società democratica. Sono gli stessi, infatti, a rappresentare il presupposto imprescindibile per l'esercizio del diritto fondamentale alla libertà di espressione, riconosciuto tanto dalle tradizioni costituzionali dei diversi Paesi, quanto a livello comunitario.

Proprio la tutela di tali principi è tra gli obiettivi fondanti dell'azione regolamentare che l'Autorità ha mandato di espletare nel settore dei media, osservandone costantemente l'evoluzione tecnologica e di mercato. In virtù del proprio ruolo istituzionale, dunque, l'Autorità conduce da anni un'intensa attività di vigilanza e monitoraggio del sistema dell'informazione. Parallelamente all'attività procedimentale, ha promosso dotandosi a partire dal 2013 del Servizio economico-statistico, osservatori¹, iniziative (convegni, seminari, *workshop*) e istituito Tavoli Tecnici² su molteplici aspetti del panorama informativo, utilizzando tecniche e strumenti di valutazione via via più sofisticati, e avvalendosi della collaborazione del mondo accademico e scientifico nazionale e internazionale.

Proseguendo nell'adozione di un approccio quanto più orientato alla conoscenza profonda dei fenomeni per poter adeguatamente affrontare le sfide imposte dalle dinamiche evolutive dei mercati dell'informazione e dall'emergere di forme patologiche di disinformazione, l'Autorità svolge regolarmente specifiche analisi³, indagini

¹ Tra questi si annoverano l'[Osservatorio sulle comunicazioni](#), con una apposita sezione di approfondimento sui media; l'[Osservatorio sul giornalismo](#), focalizzato sull'evoluzione e le criticità della professione giornalistica; e l'[Osservatorio sulle testate online](#), dedicato all'analisi del settore degli editori che operano esclusivamente su internet.

² L'Autorità ha recentemente istituito il [Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali](#), che ha l'obiettivo di promuovere l'autoregolamentazione delle piattaforme e lo scambio di buone prassi per l'individuazione e il contrasto dei fenomeni di disinformazione *online*. Al Tavolo, che rappresenta un *unicum* mondiale, partecipano rappresentanti delle piattaforme *online* (Google, Facebook, Wikipedia), di tutti i maggiori editori nazionali (di quotidiani, radio e Tv), della professione giornalistica, della componente pubblicitaria, delle associazioni di categoria, comprese quelle che rappresentano i consumatori, nonché di istituzioni accademiche e di centri ricerca. L'intento del Tavolo è, infatti, favorire la condivisione di informazioni, il confronto, l'emersione di idonee metodologie di rilevazione, e l'individuazione degli strumenti di trasparenza, delle regole e delle forme di intervento più adeguate. In tal senso e con riferimento al contrasto della disinformazione online in campagna elettorale, nell'ambito del Tavolo, sono state adottate le [Linee guida per la parità di accesso alle piattaforme online durante la campagna elettorale 2018](#).

³ Si vedano, ad esempio, le attività di [monitoraggio televisivo e radiofonico in materia di pluralismo politico e sociale](#), di analisi dei dati relativi alla [tiratura delle testate quotidiane](#), di monitoraggio dei

conoscitive⁴ e Rapporti⁵ aventi ad oggetto le varie componenti del sistema informativo: i fruitori dell'informazione (in ultima istanza tutti i cittadini), gli editori (di tutti i mezzi di comunicazione di massa), la professione giornalistica (in tutte le sue anime), i contenuti (con le relative caratteristiche in termini di qualità e tipologia dell'offerta), le fonti di finanziamento e i relativi modelli di business (sia tradizionali che innovativi), le nuove piattaforme (social network e motori di ricerca) e i meccanismi di creazione e distribuzione di notizie (inclusi gli algoritmi).

È nel solco di queste attività che si inserisce l'indagine conoscitiva in materia di “[Piattaforme digitali e sistema dell'informazione](#)”, e in particolare il presente Rapporto, che ne costituisce un primo esito. Il Rapporto è volto a esaminare l'informazione da un'insolita prospettiva, che parte dall'analisi statica e dinamica dell'informazione prodotta in Italia, sia sotto il profilo della quantità di contenuti informativi offerti e della relativa distribuzione per mezzo, sia sotto quello della qualità dell'informazione e delle tematiche affrontate. L'analisi combina dati di “input”, sui giornalisti e la loro attività, e informazioni sull’“output” informativo (Capitolo 1).

In altri termini, sono i contenuti informativi stessi e le loro modalità produttive a costituire l'oggetto centrale dell'approfondimento condotto dall'Autorità in questa sede. L'attenzione è posta non soltanto sui prodotti informativi in senso stretto, riguardanti notizie e fatti reali, ma anche su quelli relativi a notizie false, infondate, manipolate o riportate in maniera non veritiera (per una loro definizione si rimanda al recente Rapporto dell'Autorità, “[Le strategie di disinformazione online e la filiera dei contenuti fake](#)”). Conseguentemente, uno specifico approfondimento è dedicato alla disinformazione, al volume e alla varietà di contenuti generati da fonti *fake* attive in Italia (Capitolo 2).

La parte successiva del Rapporto (Capitolo 3) riguarda l'aspetto strettamente successivo alla produzione di contenuti informativi, quale il ciclo di vita (momento d'ingresso e permanenza all'interno del sistema informativo), le modalità di trattazione e la

mercati dei media attraverso l'[Informativa economica di sistema \(IES\)](#) e il [Registro degli Operatori di Comunicazione \(ROC\)](#).

⁴ Cfr. le indagini conoscitive dell'Autorità sul “[Settore dei servizi Internet e sulla pubblicità online](#)” (2014), su “[Informazione e Internet in Italia. Modelli di business, consumi, professioni](#)” (2015) e quelle attualmente in corso in materia di “[Informazione locale](#)” e di “[Big data](#)” (di cui l'Autorità ha pubblicato le prime evidenze in un *interim report*).

⁵ Si vedano, tra gli altri, il Rapporto “[L'informazione locale in Italia: il ruolo del servizio pubblico](#)” (2017) e i Rapporti pubblicati dall'Autorità sulla domanda di informazione: “[Il consumo di informazione e la comunicazione politica in campagna elettorale](#)” (2016) e, più recentemente, il “[Rapporto sul consumo di informazione](#)” (2018).

diffusione delle notizie (reali e false). È proprio dal confronto tra notizie reali e contenuti *fake* riguardo alle modalità di diffusione dei contenuti che emergono rilevanti peculiarità di questi ultimi.

A tal fine, il quarto e ultimo Capitolo si concentra sulle modalità di propagazione nel tempo dei prodotti informativi, specie quelli *fake*. In particolare, la disinformazione si nutre di elementi quali la polarizzazione e la viralizzazione che connotano il consumo informativo dei cittadini sulle piattaforme online (e i social *in primis*).

Uno studio di questo tipo ha presupposto l'impiego di una metodologia sperimentale (di cui si propone una dettagliata descrizione in Appendice), incentrata sull'utilizzo e l'integrazione di una molteplicità di fonti di dati sulle diverse componenti dei sistemi dell'informazione e della disinformazione: testate informative, fonti di disinformazione, giornalisti, e cittadini/fruitori di notizie.

La prima fonte di dati è rappresentata da un enorme *database*, composto dall'intero contenuto testuale di circa 35 milioni di documenti generati, in un arco temporale di più due anni (29 mesi, da aprile 2016 ad agosto 2018), da 1.800 fonti informative riconducibili a canali televisivi e radiofonici, quotidiani, siti web di editori tradizionali, testate esclusivamente online e relative pagine e *account* di social network, nonché a siti web (e *account* social) inclusi nelle liste di fonti "di *fake news*"⁶ redatte da fonti esterne specializzate in attività di *debunking*. Per la costruzione dei *dataset* su cui sono state realizzate gran parte delle analisi che seguono è stata utilizzata la piattaforma sviluppata da *Volocom Technology*, che offre la possibilità di estrarre la ricorrenza di qualsiasi parola chiave (o insieme di termini) e di esaminare il testo di interi documenti relativi a notizie su qualsiasi argomento e accadimento.

Al fine di compiere un'analisi rigorosa sulla diffusione dell'informazione e della disinformazione, è stato altresì individuato un campione di circa 700 principali notizie

⁶ Per l'individuazione di questi siti, posto che il ruolo dell'Autorità esula dall'accertamento circa la natura reale o falsa delle singole notizie, si è fatto riferimento alle liste predisposte da fonti esterne (in particolare, siti di *debunking* quali *butac.it* e *bufale.net*) utilizzate da svariati studi scientifici sulla disinformazione online (cfr. tra gli altri, M. Del Vicario, W. Quattrociocchi, A. Scala, F. Zollo (2018), "Polarization and Fake News: Early Warning of Potential Misinformation Targets", *arXiv preprint arXiv:1802.01400*).

(reali e false⁷), rappresentativo di tutte le categorie di generi trattati, considerando un intervallo temporale di un anno.

Inoltre, lo studio si avvale delle informazioni acquisite dal Servizio economico-statistico di Agcom circa le modalità di produzione delle informazioni da parte di un ampio e rappresentativo campione di giornalisti attivi in Italia⁸ (circa 2.000 professionisti, pari a oltre il 5% della relativa popolazione di riferimento).

Quanto alle caratteristiche di consumo informativo da parte dei cittadini, ci si è innanzitutto avvalsi dei risultati di una survey su un campione di oltre 14.000 individui rappresentativo della popolazione italiana⁹. Tale indagine riguarda il consumo di informazione dei cittadini italiani su tutti i mezzi (Tv, radio, quotidiani, siti web, search, social). Con riferimento invece all'analisi delle modalità di fruizione dell'informazione attraverso le piattaforme online, i risultati derivano da modelli di *big data analytics* svolti su decine di milioni di *account social* (v. Capitolo 4).

In tal senso, la metodologia si avvale di enormi moli di dati, provenienti da *dataset* differenti e unici al mondo, integrandoli all'interno di un'analisi complessiva dell'intero ecosistema dell'informazione, sia nella fase di produzione dell'informazione – nella duplice accezione di attività giornalistica e di offerta del prodotto finale – che in quella di consumo.

Per l'implementazione di metodi idonei al trattamento di grandi masse di dati, il Servizio economico-statistico dell'Autorità che ha condotto la presente analisi si è giovato, sulla base della Convenzione che lega oramai da anni le due istituzioni, della collaborazione scientifica dell'Università degli Studi di Napoli "Federico II"¹⁰. Inoltre, la parte relativa al consumo di informazione e disinformazione sulle piattaforme online (di cui al Capitolo 4) è stata redatta sulla base di un incarico di ricerca affidato dall'Autorità a [Walter](#)

⁷ Anche in questo caso, per l'individuazione del campione di *fake news* da includere nell'analisi, si è fatto riferimento a fonti esterne, in particolare agli articoli di *debunking* pubblicati da siti specializzati, come [butac.it](#).

⁸ Dati dalla II edizione de l'[Osservatorio sul giornalismo](#), cit.

⁹ Cfr. il "[Rapporto sul consumo di informazione](#)", cit.

¹⁰ Si ringrazia, in particolare, la Prof.ssa Maria Gabriella Grassia per la collaborazione scientifica nella definizione delle metodologie di analisi della diffusione di notizie reali e *fake news*. Si ringraziano, inoltre, la Prof.ssa Marina Marino e il Dott. Alessandro Grieco.

Quattrocchi¹¹, che da diversi anni si occupa dell'analisi scientifica del fenomeno, pubblicando studi nazionali e internazionali.

Nelle prossime pagine, dunque, si procederà a illustrare i risultati delle predette analisi innovative, nell'intento di delineare, grazie all'osservazione diretta di milioni di dati, gli scenari informativi che caratterizzano il sistema italiano. Tenendo a mente la tutela del pluralismo come motivazione esplorativa, il Rapporto persegue l'obiettivo di svolgere una valutazione circa la robustezza del sistema informativo e le *performance* dei vari mezzi nella produzione e divulgazione di informazione, nonché di compiere un ulteriore passo in avanti nella comprensione dell'entità, delle peculiarità e modalità di diffusione della disinformazione. Valutazioni, queste ultime, a loro volta indispensabili per orientare opportunamente il percorso regolamentare e di policy dell'Autorità a garanzia di un'informazione corretta, libera e plurale.

¹¹ Cfr. Esiti di cui alla procedura comparativa per l'affidamento di un incarico individuale di ricerca su "Informazione e piattaforme digitali"

LA PRODUZIONE DI INFORMAZIONE



1.1. INTRODUZIONE

In linea generale, la produzione di informazione identifica il processo di realizzazione e offerta al pubblico di contenuti informativi aventi ad oggetto fatti, accadimenti, fenomeni, in altre parole notizie di qualsiasi genere. È, quindi, il processo da cui dipendono la quantità, la varietà e la qualità dell'informazione che raggiunge i cittadini, sulla base della quale gli stessi formano le proprie opinioni e punti di vista.

La produzione di informazione di ciascun mezzo di comunicazione (Tv, radio, quotidiani, internet) avviene ad opera degli editori e delle figure professionali occupate nelle relative strutture redazionali, primi fra tutti i giornalisti.

Nel dettaglio, gli editori, proprietari delle fonti informative, finanziano l'attività di produzione e diffusione di informazione attraverso modelli di business che prevedono soltanto in taluni casi (ad esempio, per i quotidiani) la vendita agli utenti finali del prodotto informativo. Per lo più, è la vendita di spazi pubblicitari agli inserzionisti a costituire la fonte di finanziamento prevalente, cui possono affiancarsi alcune forme di contributi pubblici e privati¹².

I giornalisti costituiscono il fattore produttivo ("input") principale impiegato nel processo di produzione dell'informazione. L'attività del giornalista si sostanzia nel reperimento, analisi, approfondimento delle notizie, cui segue la concreta composizione del contenuto informativo (articoli, servizi televisivi e radiofonici, ma anche *post/tweet* sulle piattaforme online). Il lavoro del giornalista, inoltre, ha le proprie specificità legate al mezzo utilizzato, che vanno di pari passo con le peculiarità del contenuto informativo offerto sul medesimo mezzo¹³.

Sotto il profilo tecnico, infatti, il prodotto informativo può essere destinato a supporti diversi (carta stampata, televisore, apparecchio radiofonico, pc/tablet/smartphone) e avere pertanto formati differenti, anche misti (testo, video, audio, foto/immagine). Ciascun contenuto informativo, inoltre, deve adattarsi allo spazio messo a disposizione dal mezzo. Di conseguenza, i prodotti informativi possono differenziarsi nella lunghezza, nella frequenza di aggiornamento, nello stile, nell'impostazione grafica.

¹² Nel 2017, nel settore televisione la pubblicità è stata la prima fonte di ricavi con il 41% del fatturato complessivo; percentuale che sale al 43% per i quotidiani e al 78% per la radio. Cfr. "[Relazione Annuale Agcom 2018](#)".

¹³ Per una descrizione dettagliata della professione comprensiva delle attività del giornalista, dei mezzi e degli strumenti utilizzati, delle competenze necessarie, si rimanda all'[Osservatorio sul giornalismo](#), cit.

Sotto il profilo contenutistico, i prodotti informativi si contraddistinguono altresì per linea editoriale, grado di approfondimento, registro linguistico.

Infine, sotto il profilo economico, l'informazione assume natura di "bene meritorio" (ossia meritevole, anche in assenza di fallimenti di mercato, di una speciale tutela da parte dello Stato), di "prodotto creativo" (contenendo in sé un sostanziale sforzo creativo), nonché di "bene di esperienza" (potendo essere valutato solo successivamente al suo effettivo consumo). Peraltro, l'informazione ha assunto nel tempo la connotazione di "bene (quasi) pubblico", presentando le due caratteristiche dell'assenza di rivalità e, sempre più, della non escludibilità nel consumo. L'avvento del digitale ha, infatti, reso il contenuto informativo un bene riproducibile infinite volte, consumato in gruppo, spesso gratuitamente, slegandolo con ciò dal supporto fisico e dai vincoli di scarsità¹⁴.

Tenendo conto delle caratteristiche del processo di produzione dell'informazione e della natura tecnica ed economica del prodotto informativo, nel prosieguo del Capitolo, si procederà a presentare e discutere i risultati delle analisi condotte sulla quantità, qualità e varietà dell'informazione prodotta in Italia da tutti i mezzi di comunicazione di massa.

¹⁴ Cfr. Agcom, (2015), "[Informazione e Internet in Italia. Modelli di business, consumi, professioni](#)", Capitolo 1 – L'Informazione.

1.2. LA QUANTITÀ E LA QUALITÀ DELL'INFORMAZIONE PRODOTTA IN ITALIA

In un'ottica di tutela del pluralismo, posto che è necessario che l'informazione prodotta possa effettivamente raggiungere le persone, ossia che le persone abbiano la possibilità di accedere ai media per informarsi¹⁵, vi è consenso generale sul fatto che una maggiore quantità di informazione disponibile in un sistema sia un bene per i cittadini¹⁶.

In tal senso, la quantità di informazione prodotta (e, almeno in potenza, disponibile per i cittadini) rappresenta un primo indicatore dello stato di salute del sistema informativo di un Paese. Una maggiore quantità di contenuti offerti può essere indice (o quanto meno innalza la probabilità) di una maggiore trasparenza, pluralità, copertura di fatti, argomenti e visioni, conferendo ai cittadini una più ampia opportunità di apprendere e confrontare notizie e fonti diverse anche all'interno di uno stesso mezzo.

Un'analisi delle caratteristiche dell'informazione prodotta nel sistema informativo italiano, quindi, non può prescindere da una cognizione preliminare dell'aspetto quantitativo e dei possibili fattori atti ad incidere sul processo di produzione dell'informazione.

Un primo studio in proposito è riportato nella Figura 1.1, che mostra l'evoluzione dinamica del volume di informazione prodotto in Italia negli ultimi tre anni. Nell'analisi svolta si è tenuto conto del numero complessivo di documenti informativi prodotti mensilmente da tutti i mezzi di comunicazione: dai quotidiani ai canali televisivi, dai canali radiofonici ai siti web di informazione (sia di editori attivi esclusivamente online sia di editori operanti anche su altri mezzi), fino ad arrivare alle piattaforme online (pagine e *account* di editori e *influencer*). Nel dettaglio, i documenti informativi afferiscono a ciascun articolo (nel caso di quotidiani e siti web di informazione), servizio televisivo e radiofonico, *post/tweet* (nel caso delle piattaforme online) rilasciato da circa 1.800 fonti di informazione (cfr. APPENDICE METODOLOGICA).

Si è fatto riferimento, inoltre, all'informazione nella sua accezione più generale, ricomprensivo non solo i prodotti informativi aventi ad oggetto le c.d. *hard news* (come quelle di cronaca e politica che, per la loro importanza, richiedono la pubblicazione immediata), ma anche le c.d. *soft news*, ossia quelle notizie legate a

¹⁵ Sulle criticità legate ai rischi di esclusione o marginalizzazione mediale di determinate fasce della popolazione, cfr. da ultimo Agcom, "[Rapporto sul consumo di informazione](#)", cit.

¹⁶ Al riguardo v. A. Prat, D. Stromberg, (2013), *The Political Economy of Mass Media*, in a cura di D. Acemoglu, M- Arellano, E. Dekel, *Advances in Economics and Econometrics: Volume 2, Applied Economics*, Cambridge University Press.

tematiche contraddistinte da un minor grado di urgenza informativa che non richiedono necessariamente una pubblicazione tempestiva. Pertanto, sono state incluse nell'analisi tutte le categorie di informazione prodotta: cronaca, politica, economia, esteri, scienza e tecnologia, cultura, spettacolo e sport.

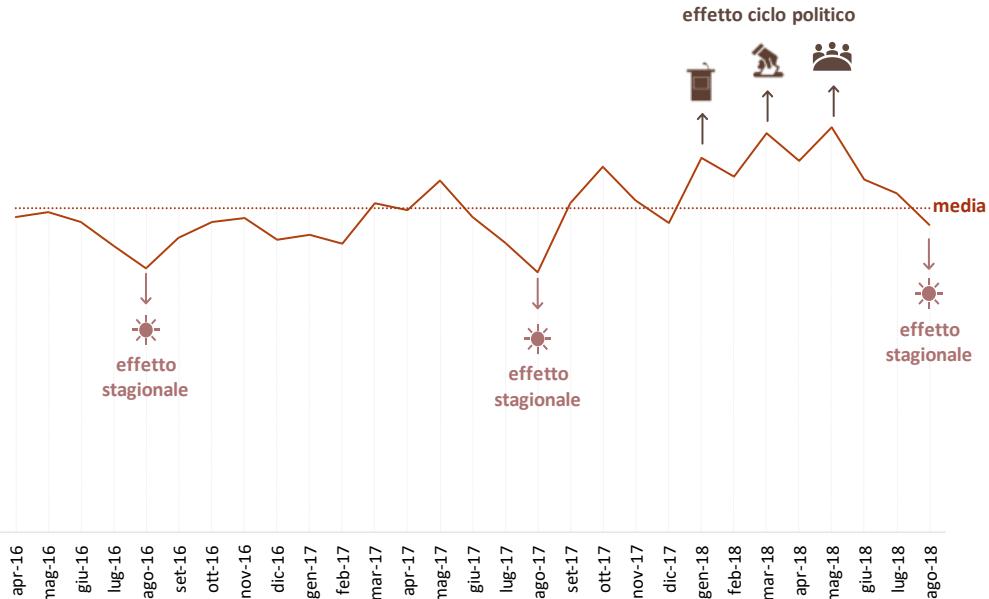
Sulla base delle sopra esposte premesse metodologiche, la Figura 1.1 evidenzia un andamento del volume informativo prodotto che, per quanto non appaia costante da un mese all'altro, non subisce scostamenti (in alto o in basso) di notevole entità rispetto alla media dell'intero periodo. Al riguardo, occorre comunque tener presente dell'esistenza, per alcuni mezzi in special modo, di vincoli (di spazio, di tempo di realizzazione e pubblicazione del prodotto, ...) alla flessibilità di adeguamento (verso valori più alti) della quantità offerta rispetto alla domanda e alle esigenze di copertura richieste dagli accadimenti che possono verificarsi ogni giorno. In ogni caso, se si guarda al *trend* annuale, si osserva un innalzamento dell'ammontare medio di informazione immessa nel sistema, che raggiunge il valore massimo nel 2018.

D'altra parte, l'analisi dinamica mensile suggerisce come sia possibile rintracciare gli effetti determinati dall'azione di almeno due fattori sulla quantità di offerta informativa prodotta dai media: il primo di natura ricorrente (stagionalità), il secondo di carattere contingente (ciclo politico).

Nel primo caso, si tratta dell'effetto negativo subito ogni anno dal volume informativo nel periodo estivo, in particolare nel mese di agosto, momento tipicamente caratterizzato da una minore domanda da parte del pubblico, dalla chiusura di diverse strutture pubbliche e aziendali e da un impiego meno intensivo della forza lavorativa in vari settori, incluso quello editoriale.

Un secondo effetto, di segno opposto, si rinvie in concomitanza dell'avvio del nuovo ciclo politico. Infatti, nel periodo che copre la campagna elettorale, lo svolgimento delle ultime elezioni politiche svoltesi in Italia (nel marzo 2018) e la formazione del nuovo governo (avvenuta a maggio 2018), si registrano i picchi più alti di informazione prodotta. Questa evidenza segnala la grande rilevanza attribuita dai media alla politica e alle vicende istituzionali, anche in risposta alle esigenze dei cittadini, che proprio dai media attingono le informazioni su cui fondano le proprie opinioni e le trasformano in espressioni di voto. Basti pensare che, come rilevato dall'Autorità nel recente [Rapporto sul consumo di informazione](#), l'85% della popolazione italiana maggiorenne dichiara di ricercare sui media notizie attinenti alla politica.

Figura 1.1 - Andamento mensile dell'informazione prodotta in Italia



Fonte: elaborazioni Agcom su dati Volocom

Volendo operare una distinzione per mezzo, rimanendo sul piano meramente quantitativo, la Figura 1.2 rivela il differente apporto conferito dai media al sistema nazionale dell'informazione. Più specificamente, il maggior contributo in termini di output prodotto deriva dai quotidiani (nazionali e locali), peraltro l'unico mezzo di comunicazione che, al netto delle inserzioni pubblicitarie, offre contenuti interamente dedicati all'informazione generalista o specializzata. Segue internet, per quel che attiene sia alla componente editoriale sia a quella social. Un minor contributo quantitativo deriva invece dai canali televisivi e, in maniera decisamente più esigua, dalle emittenti radiofoniche, mezzi per i quali i contenuti informativi costituiscono soltanto una parte dell'offerta proposta, a favore dei programmi di intrattenimento.

Tuttavia, i valori quantitativi, specie quando il prodotto in questione è l'informazione, devono essere letti in un'ottica più ampia, che chiama in causa anche altri aspetti della produzione, quali l'efficienza redazionale e la qualità dei contenuti offerti. Per rispondere a questa esigenza analitica, nella medesima Figura 1.2, è stata ricavata una stima della funzione di produzione dell'informazione, intesa come la curva che, a parità di altre condizioni, esprime in ogni punto la relazione tra le unità impiegate di fattore produttivo (l'input, sull'asse orizzontale) e il numero di contenuti informativi offerti (l'output prodotto, riportato sull'asse verticale). La stima della funzione di produzione

dell'informazione, più precisamente, è stata derivata dall'interpolazione dei valori noti del numero di contenuti informativi prodotti e giornalisti impiegati mediamente in un mese da ciascun mezzo.

Figura 1.2 - Funzione di produzione dell'informazione (stima, valori medi mensili)



Note: per Tv e radio sono state considerate le emittenti nazionali. Per tali mezzi, l'ammontare di contenuti informativi offerti, tenendo conto delle ore di programmazione dedicate ai Tg/Gr e agli altri programmi di informazione, è calcolato sulla base della durata media di un servizio. Per i quotidiani e i siti web di informazione, i contenuti informativi si identificano con gli articoli pubblicati, mentre per i social network si fa riferimento ai *post/tweet* pubblicati.

Fonte: elaborazioni Agcom su dati Volocom e aziendali (per i contenuti informativi offerti), e *Osservatorio Agcom sul giornalismo - II edizione* (per i giornalisti impiegati)

Innanzitutto, è possibile osservare che la funzione di produzione dell'informazione è crescente ed esibisce rendimenti di scala decrescenti, per cui un incremento del numero di giornalisti impiegati determina un incremento meno che proporzionale della quantità di contenuti informativi prodotti. Ne deriva che, in linea con quanto già rilevato dall'Autorità nell'ambito dell'indagine conoscitiva su "[Informazione e internet in Italia](#)".

Modelli di business, consumi, professioni", la dimensione ottima minima di una redazione – in corrispondenza della quale, in condizioni di pieno utilizzo della capacità produttiva, il costo medio unitario dell'input è minimo (non decresce più al crescere della dimensione) – assume un valore contenuto (che è diminuito negli ultimi anni¹⁷ e varia da mezzo a mezzo¹⁸), prevedendo l'impiego di un numero di giornalisti non elevato.

In secondo luogo, il posizionamento dei punti che individuano le combinazioni di input/output dei singoli mezzi rispetto alla curva della funzione di produzione fornisce indicazioni circa l'intensità di utilizzo dei giornalisti per la produzione di informazione (cfr. anche Figura 1.3) e, dunque, indirettamente anche sul livello qualitativo del contenuto informativo generato.

Al riguardo, si distinguono tre tipologie di mezzi, a seconda che si collochino in linea, al di sopra o al di sotto della funzione di produzione.

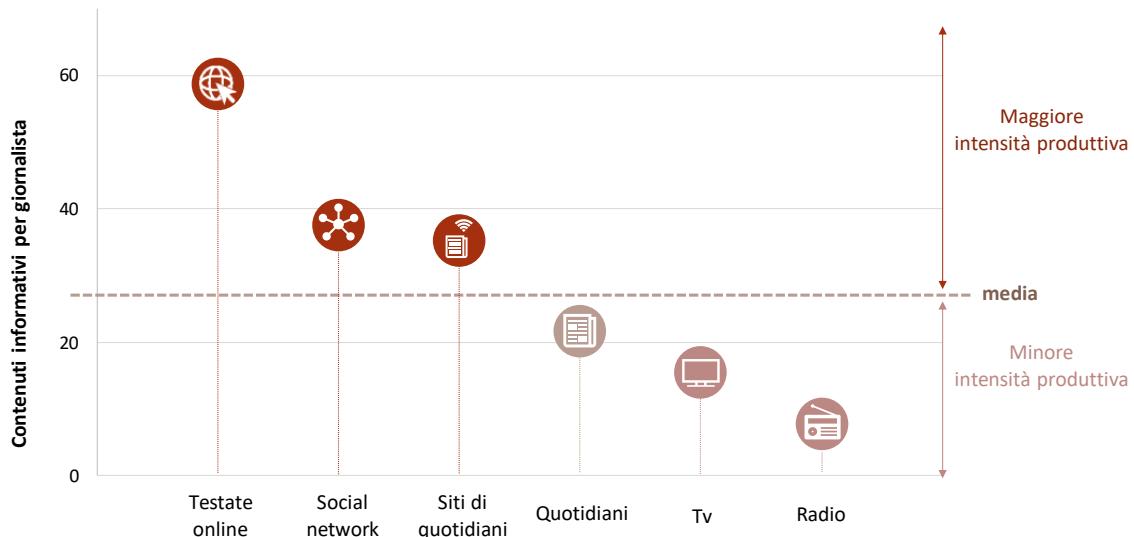
Nello specifico, le fonti informative riconducibili a internet (siti di quotidiani, testate online, social network) si trovano sopra la curva di produzione, per cui, a parità di giornalisti impiegati e a parità di altre condizioni, producono un'offerta di informazione maggiore. Tale circostanza segnala un sovrautilizzo del fattore produttivo (giornalisti) impiegato, atto a riverberarsi negativamente sulla qualità del prodotto finale. In questo caso, infatti, i giornalisti, nel produrre i contenuti informativi, si trovano a dover conciliare lo svolgimento delle proprie mansioni, da un lato con la velocità dei tempi di aggiornamento richiesti dall'informazione online e, dall'altro, con le logiche di monetizzazione dei contenuti online spesso basate sull'attrazione del maggior numero di click. La ristrettezza dei tempi rischia di tradursi, in termini pratici, in minor attenzione nella raccolta di informazioni e nella verifica delle fonti, oltre che in un minor

¹⁷ In tal senso, "La riorganizzazione del sistema informativo ha riguardato anche l'aspetto organizzativo complessivo e quindi, a livello editoriale, le redazioni. In tutto il mondo, la riorganizzazione ha comportato la nascita di nuove figure professionali (quali i web designer e i web editor) che si affiancano e supportano i giornalisti, il superamento di alcune figure tradizionali, e soprattutto la riorganizzazione di tutte le professionalità che operano per la produzione del prodotto informativo. Ciò ha implicato un ripensamento complessivo dell'organizzazione della filiera informativa (che è arrivato a ridefinire anche da un punto di vista architettonico e logistico la redazione), che si è accompagnato ad un processo di ridimensionamento delle strutture produttive (c.d. downsizing). Le redazioni si sono, da un lato, ridimensionate, mentre, dall'altro lato, hanno incorporato in un unico luogo le diverse medialità del gruppo di riferimento." (Agcom, "Indagine conoscitiva Informazione e internet in Italia", cit., par. 513, p. 215).

¹⁸ "...la dimensione media [delle redazioni] passa dai 53 addetti (di cui 16 giornalisti, 35 collaboratori esterni e 2 altre figure professionali) dei quotidiani, ai 21 della Tv, fino ad arrivare ai 9 della radio". (Agcom, "Indagine conoscitiva Informazione e internet in Italia", cit., par. 514, p. 215).

livello di approfondimento, a scapito della qualità stessa dell'informazione offerta (v. anche infra par. 3.3).

Figura 1.3 - Intensità produttiva dei giornalisti nel mese medio



Note: l'intensità produttiva dei giornalisti è calcolata, per ciascun mezzo, come il rapporto tra contenuti informativi offerti e giornalisti impiegati, nel mese medio.

Per Tv e radio sono state considerate le emittenti nazionali. Per tali mezzi, l'ammontare di contenuti informativi offerti, tenendo conto delle ore di programmazione dedicate ai Tg/Gr e agli altri programmi di informazione, è calcolato sulla base della durata media di un servizio. Per i quotidiani e i siti web di informazione, i contenuti informativi si identificano con gli articoli pubblicati, mentre per i social network si fa riferimento ai *post/tweet* pubblicati.

Fonte: elaborazioni Agcom su dati Volocom e aziendali (per i contenuti informativi offerti), e *Osservatorio Agcom sul giornalismo - II edizione* (per i giornalisti impiegati)

Un secondo gruppo di mezzi include le emittenti televisive e radiofoniche, che si collocano al di sotto della funzione di produzione, presentando una situazione opposta rispetto alle fonti online. I giornalisti presentano una minore intensità produttiva che, se per una parte (data la natura non solo informativa dei mezzi) può essere attribuita alla loro assegnazione anche a mansioni diverse da quelle routinarie giornalistiche, per altra parte può essere indice di maggiore accuratezza e approfondimento dell'informazione trasmessa.

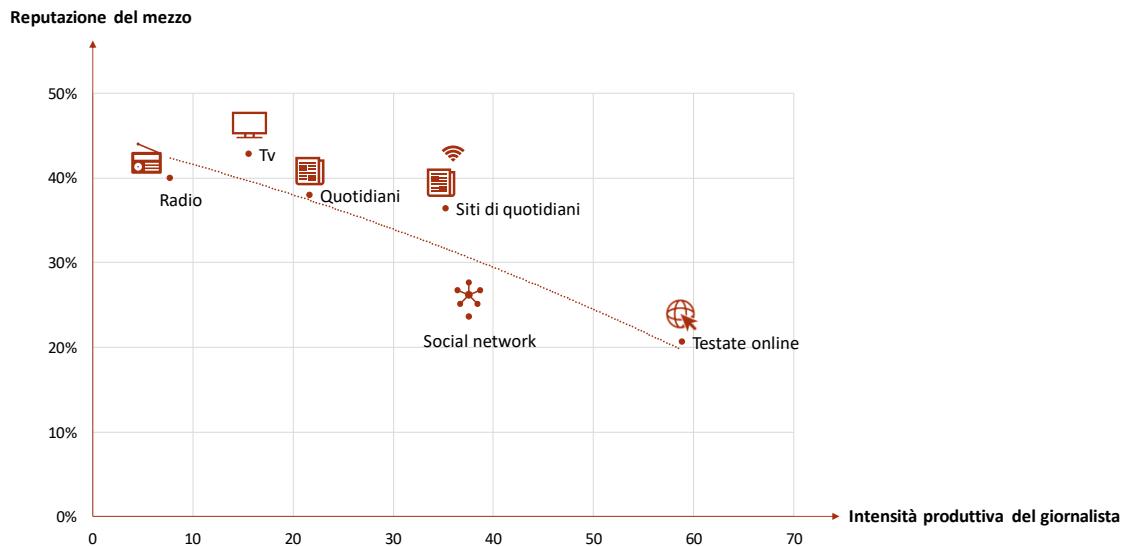
Diversamente dai mezzi precedenti, il posizionamento dei quotidiani è pressoché in linea (lievemente al di sotto) con la funzione di produzione. I quotidiani, quindi, si rivelano il mezzo con un'intensità produttiva dei giornalisti più vicina al valore medio,

come si evince anche dalla Figura 1.3. All'interno del mezzo, i giornalisti sono dedicati essenzialmente alle mansioni più tipiche della professione (essendo i quotidiani interamente dedicati all'informazione) e non scontano la ristrettezza dei tempi dettata dalla necessità di pubblicazione costante nell'arco della giornata, potendo così dedicare maggiore cura e attenzione all'aspetto qualitativo.

Le considerazioni sopra esposte sono ulteriormente avvalorate dalla relazione che si riscontra tra l'intensità produttiva (intesa come numero di contenuti informativi mediamente prodotti in un mese) dei giornalisti impiegati da un mezzo e la reputazione di cui lo stesso mezzo gode presso il pubblico, ossia l'affidabilità percepita da coloro che lo utilizzano per reperirvi notizie. Il grado di affidabilità riconosciuto a ciascun mezzo può essere considerato un indicatore di soddisfazione degli utenti circa il consumo che questi ultimi effettivamente fanno dell'informazione diffusa dallo stesso. Poiché, come anticipato, l'informazione è un "bene di esperienza", il cui valore può essere determinato solo a seguito della sua fruizione, una valutazione soggettiva circa l'affidabilità di una fonte informativa rispecchia direttamente il livello di qualità percepito.

La Figura 1.4 mostra chiaramente come all'aumentare dell'intensità produttiva del giornalista, la qualità percepita dell'informazione veicolata e di conseguenza la reputazione del mezzo diminuiscano, con i media tradizionali che, a fronte di una minore quantità di informazione offerta, mantengono la reputazione più alta e, viceversa, le fonti online, che producono il maggior numero di contenuti informativi godendo però di una più scarsa affidabilità.

Figura 1.4 - Relazione tra intensità produttiva del giornalista e reputazione del mezzo



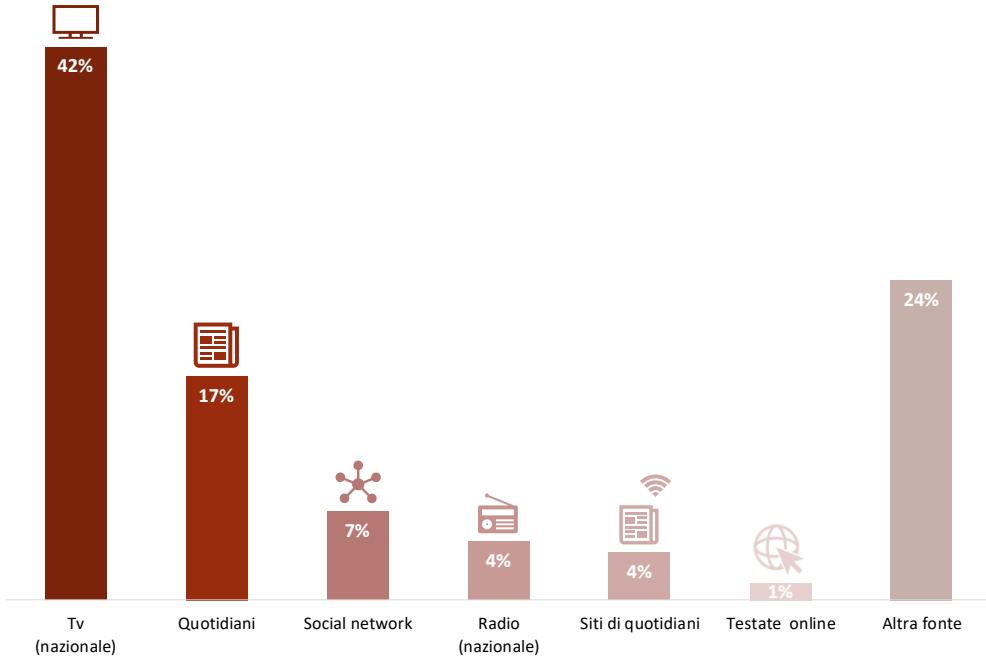
Note: l'intensità produttiva dei giornalisti è calcolata, per ciascun mezzo, come il rapporto tra contenuti informativi offerti e giornalisti impiegati, nel mese medio.

La reputazione del mezzo è calcolata come percentuale dei fruitori del mezzo a scopi informativi che ritengono lo stesso "affidabile" o "molto affidabile".

Fonte: elaborazioni Agcom su dati Volocom e aziendali; *Osservatorio Agcom sul giornalismo - II edizione* e AGCOM (2018), *Rapporto sul consumo di informazione*

Quando gli individui scelgono di accedere a un mezzo per informarsi è probabile che lo facciano anche perché ritengono la fonte attendibile, sulla base della reputazione di cui gode quest'ultima e della propria personale esperienza. Non a caso, nella ricerca di informazioni, la fonte ritenuta più importante dagli individui è riconducibile, nella maggior parte dei casi, a un mezzo tradizionale. Nell'ultimo [Rapporto sul consumo di informazione](#) pubblicato dall'Autorità (cfr. Figura 1.5), si osserva come i canali televisivi (nazionali) siano considerati la fonte più importante per informarsi dal 42% della popolazione italiana, seguiti dai quotidiani, indicati dal 17% dei cittadini, laddove le fonti online segnano valori di gradimento decisamente inferiori, compresi tra l'1% e il 7%.

Figura 1.5 - Principale fonte di informazione per i cittadini italiani (% popolazione)



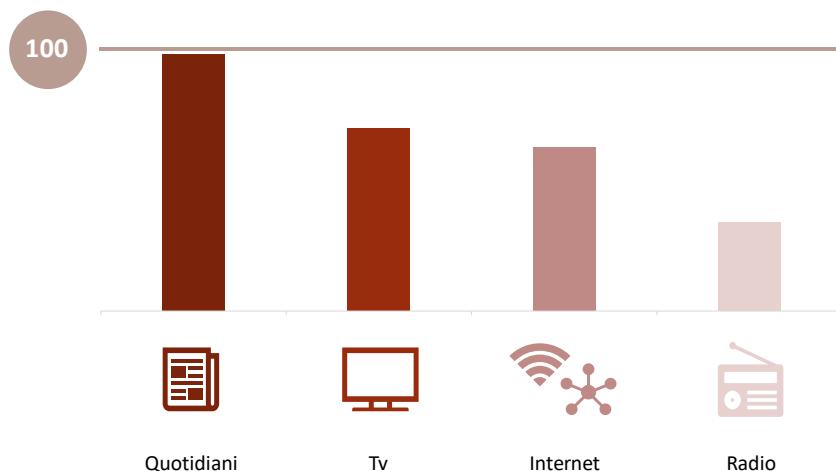
Nota: tra le altre fonti di informazione sono inclusi i canali televisivi e radiofonici locali e altre fonti online, quali motori di ricerca, aggregatori di notizie e portali, blog.

Fonte: AGCOM (2018), *Rapporto sul consumo di informazione*

Anche se si guarda al consumo effettivo, ossia attento e non superficiale, dei media per informarsi, la valenza informativa attribuita ai quotidiani in particolar modo e alla Tv emerge distintamente. La Figura 1.6, che riporta una stima del consumo effettivo di informazione, calcolato come il rapporto tra la percentuale di coloro che reputano un mezzo la fonte più importante per informarsi e la percentuale di coloro che accedono al medesimo mezzo a scopi informativi, registra per i quotidiani un valore prossimo a quello massimo (100), denotando la grande attenzione e fiducia riposta dai lettori nella fruizione degli articoli presentati.

Al contrario, alla grande e sempre maggiore affermazione di internet come mezzo di informazione di massa, si associano spesso bassi livelli di attenzione e credibilità, tanto che un consumo effettivo di informazione sul mezzo riguarda meno dei due terzi degli utenti che si informano online.

Figura 1.6 - Stima del consumo effettivo di informazione tra i fruitori dei mezzi



Nota: il consumo effettivo di informazione è stimato come il rapporto tra la percentuale di coloro che reputano un mezzo la fonte più importante per informarsi e la percentuale di coloro che accedono al medesimo mezzo a scopi informativi.

Fonte: AGCOM (2018), *Rapporto sul consumo di informazione*

1.3. LE TEMATICHE DELL'INFORMAZIONE

Per delineare in maniera più esaustiva il quadro delle caratteristiche che contraddistinguono la produzione di informazione nel sistema italiano, un aspetto ulteriore da indagare attiene alla varietà di generi e tematiche affrontate dai media e offerte al pubblico.

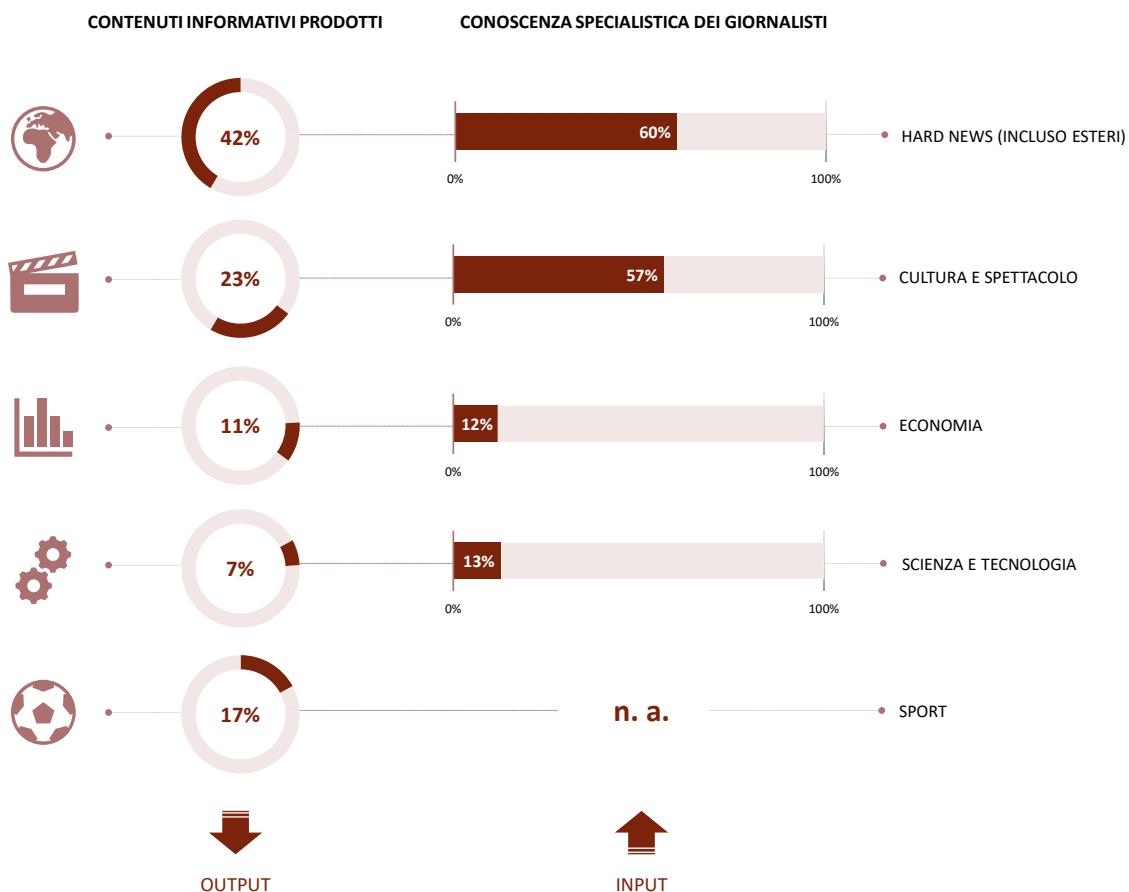
Al riguardo, l'analisi di milioni di contenuti informativi prodotti dai mezzi di comunicazione ha consentito di tracciare la distribuzione dell'offerta di informazione in base a cinque categorie tematiche, riconducibili a: *"hard news"*, *"cultura e spettacolo"*, *"economia"*, *"scienza e tecnologia"*, e *"sport"*.

Nello specifico (cfr. Figura 1.7), si rileva che, in un mese medio, oltre il 40% dell'informazione prodotta in Italia riguarda *hard news*, ossia notizie di cronaca (come quelle relative ad accadimenti di attualità, ad esempio gli eventi naturali, e alla cronaca nera e giudiziaria: reati, omicidi, rapine, violenze, incidenti, ...), politica (incluse le notizie riferite alle elezioni, ai referendum, alle questioni istituzionali o partitiche) e fatti di rilevanza internazionale (tra cui guerre, diritti umani, politica estera, ...).

Seguono le categorie più legate all'intrattenimento, quali *"cultura e spettacolo"*, che rappresenta quasi un quarto del volume informativo complessivo, e *"sport"*, che costituisce il 17% dell'informazione offerta.

Il minor apporto quantitativo all'insieme di contenuti immessi nel sistema informativo deriva dalle tipologie associate a un'informazione più specializzata, *"economia"* da un lato e *"scienza e tecnologia"* dall'altro. Si tratta di categorie informative che comunque concernono tematiche di una certa rilevanza, idonee a influire sulla sfera ideologica e sulle decisioni di investimento e consumo dei cittadini. L'informazione economica investe, infatti, argomenti come operazioni societarie, finanza, andamento macroeconomico; l'informazione in materia scientifica può comprendere, fra l'altro, notizie di carattere sanitario (vaccini, scoperte in campo medico, trattamenti terapeutici, ...), notizie su ambiente e cambiamento climatico, notizie sulle innovazioni tecnologiche.

Figura 1.7 - Varietà e livello di conoscenza specialistica delle tematiche trattate



Note: la categoria "hard news" include le notizie di cronaca, politica e quelle di rilevanza internazionale.

Il livello di conoscenza specialistica dei giornalisti è dato dalla percentuale di coloro che hanno una formazione (intesa come percorso universitario) specialistica attinente a una data tematica sul totale dei giornalisti che producono contenuti aventi ad oggetto quella stessa tematica.

Fonte: elaborazioni Agcom su dati Volocom e aziendali; *Osservatorio Agcom sul giornalismo - II edizione*

La distribuzione dell'offerta informativa per categoria può essere letta congiuntamente al livello di conoscenza specialistica posseduto, in base al proprio percorso universitario, dai giornalisti che si occupano di comporre i relativi contenuti informativi. Il livello di conoscenza specialistica, che riguarda l'input produttivo (i giornalisti appunto), assume rilievo sotto il profilo della qualità attesa dell'output prodotto (l'informazione), presumendo che una maggiore preparazione e formazione in un determinato campo sia atta a tradursi in una maggiore qualità dell'articolo o servizio realizzato in quel medesimo campo dell'informazione.

In tal senso, dalla Figura 1.7 si evince come le categorie in cui si riscontra una maggiore attinenza (prossima al 60%) tra il titolo di studio posseduto e i temi trattati dai giornalisti nella quotidianità della loro professione siano quelle relative a *"hard news"* e *"cultura e spettacolo"*, categorie per le quali, come visto, si rileva anche il maggior volume di contenuti offerti.

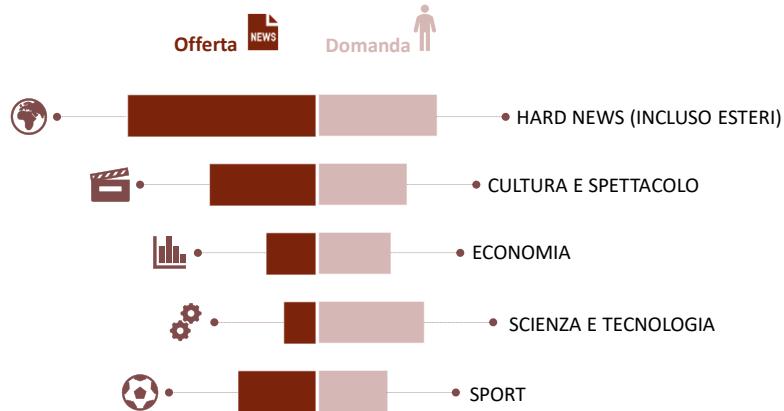
Al contempo, emerge che la trattazione di argomenti quali *"economia"* e *"scienza e tecnologia"*, che richiedono un maggior bagaglio di competenze specifiche, è quella in cui la presenza di giornalisti specializzati è invece meno significativa (poco superiore al 10%).

Pertanto, non solo l'informazione specializzata viene offerta in misura più contenuta rispetto a quella più generalista, ma è anche prodotta per la gran parte da professionalità che non possiedono competenze specialistiche nelle materie di riferimento.

Peraltro, paragonando la distribuzione dell'offerta informativa a quella della domanda, appare evidente l'esistenza di significativi squilibri (cfr.

Figura 1.8). Se per le *"hard news"* e le notizie di *"cultura e spettacolo"* si osserva un eccesso di offerta, le categorie legate all'informazione specializzata, *"scienza e tecnologia"* in particolar modo, presentano, viceversa, un eccesso di domanda.

Figura 1.8 - Distribuzione dell'offerta e della domanda di informazione, per categoria



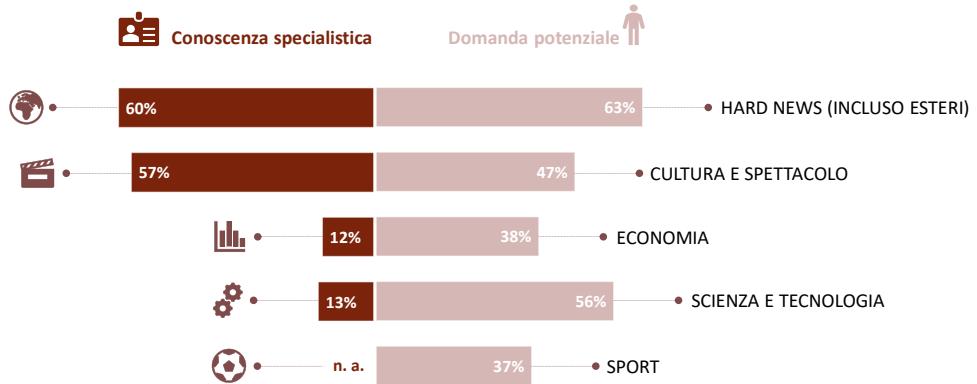
Nota: la categoria "hard news" include le notizie di cronaca, politica e quelle di rilevanza internazionale.

Fonte: elaborazioni Agcom su dati Volocom e aziendali (per l'offerta) e Reuters Institute for the Study of Journalism, *Digital News Report 2017* (per la domanda)

L'attuale proposta informativa si rivela, quindi, carente soprattutto nell'offerta di contenuti più specializzati o di nuovo interesse, laddove la forte richiesta da parte dei cittadini non appare efficientemente soddisfatta né sotto il profilo quantitativo né sotto quello qualitativo dei contenuti prodotti.

Come mostra chiaramente la Figura 1.9, il livello di conoscenza specialistica per contenuti informativi che riguardano economia, finanza, scienze e tecnologia appare insufficiente. A una domanda potenziale, pari al 56% della popolazione nel caso delle notizie di scienza e tecnologia e a poco meno del 40% nel caso delle notizie economiche, corrisponde una bassa qualificazione dell'offerta, per cui gli argomenti specialistici sono spesso trattati da giornalisti che non hanno una specifica preparazione di base. E questa criticità emerge con ancor più evidenza in un momento quale quello attuale in cui il lavoro giornalistico è sottoposto costantemente al vaglio dei lettori, sempre più dotati essi stessi di specifiche professionalità ed interessi.

Figura 1.9 - Livello di conoscenza specialistica dei giornalisti e domanda potenziale dei cittadini, per categoria



Note: la categoria “hard news” include le notizie di cronaca, politica e quelle di rilevanza internazionale.

Il livello di conoscenza specialistica dei giornalisti è dato dalla percentuale di coloro che hanno una formazione specialistica (intesa come percorso universitario) attinente a una tematica sul totale dei giornalisti che producono contenuti aventi ad oggetto quella stessa tematica.

La domanda potenziale per ciascuna categoria informativa è calcolata come percentuale della popolazione che dichiara di essere molto o estremamente interessata alle notizie di quella categoria.

Fonte: *Osservatorio Agcom sul giornalismo - II edizione* ed elaborazioni Agcom su dati Reuters Institute for the Study of Journalism, *Digital News Report 2017*

LA PRODUZIONE DI DISINFORMAZIONE

2



2.1. INTRODUZIONE

Quando si parla di disinformazione ci si riferisce alla divulgazione di contenuti informativi falsi, infondati, manipolati o riportati in maniera non veritiera, creati ad arte in modo da risultare verosimili nel contesto mediatico.

Per quanto la disinformazione possa riguardare tutti i media e la diffusione di questo tipo di contenuti sia sempre stata considerata una possibile leva di condizionamento dell'opinione pubblica, è con l'affermarsi di internet (delle piattaforme online in particolare) che il fenomeno ha assunto una nuova e così ampia portata.

Le analisi riportate nel presente Capitolo, pertanto, sono incentrate proprio sulla disamina della produzione di disinformazione online. Nel recente Rapporto "[Le strategie di disinformazione online e la filiera di contenuti fake](#)", redatto nell'ambito delle attività del [Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali](#), l'Autorità ha definito la disinformazione online individuandone gli elementi distintivi. Si tratta di elementi che attengono sia alla sfera oggettiva (ossia, relativa all'oggetto del contenuto veicolato) che soggettiva (ovvero i soggetti coinvolti nella creazione, produzione e distribuzione del contenuto) e interessano tutte le fasi della filiera che conduce alla messa in atto della strategia di disinformazione.

In sostanza, l'esistenza di disinformazione online si connota per la compresenza di sei elementi principali:

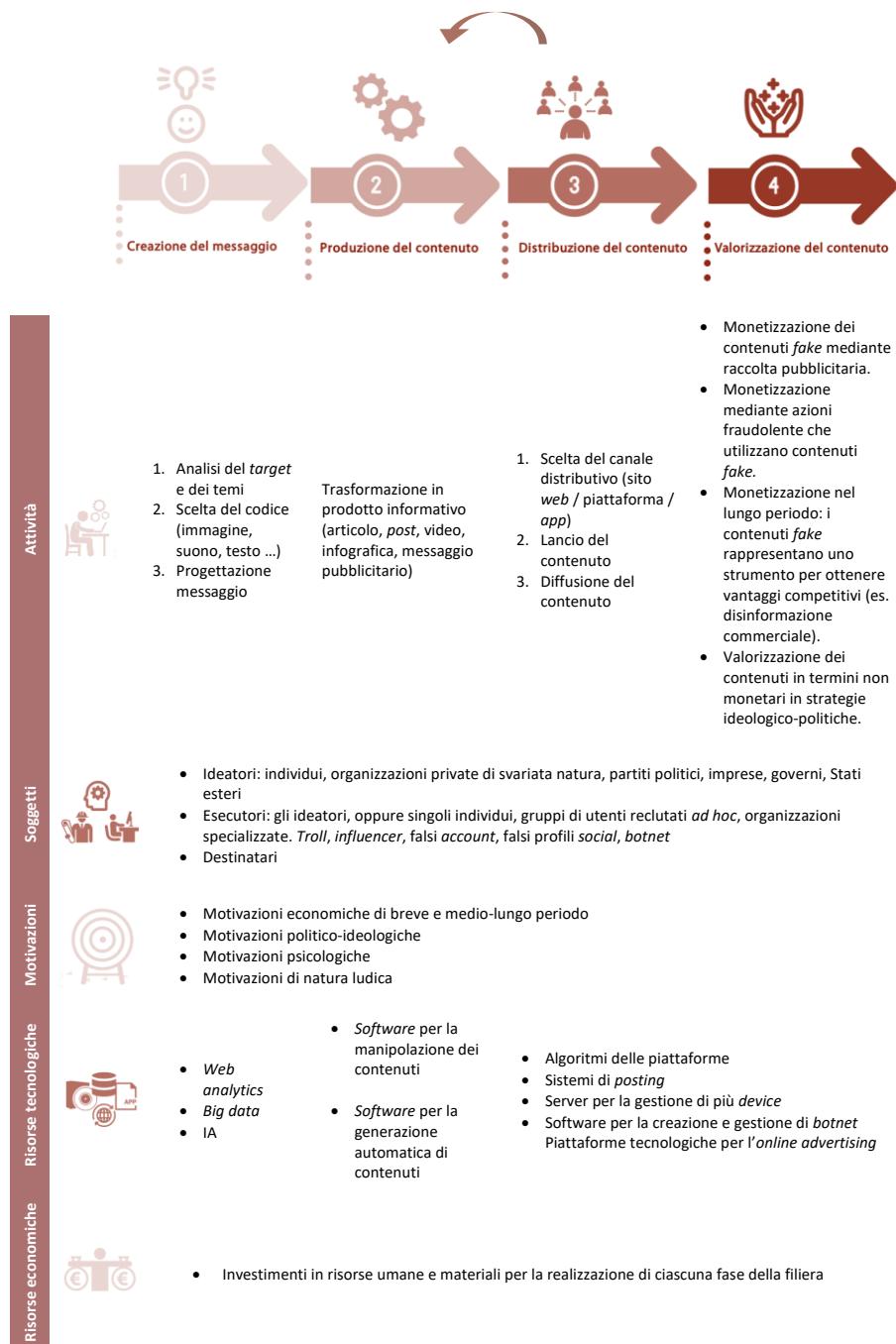
- la falsità dei contenuti (c.d. contenuti *fake*);
- la contagiosità degli stessi (ossia, l'attitudine a trasferire stati emotivi e percezioni tra gli utenti, ovvero a condizionare il comportamento dei riceventi);
- l'intento doloso sottostante alla loro creazione;
- la motivazione politico/ideologica o economica di chi li crea per poi diffonderli;
- la diffusione degli stessi in maniera massiva;
- l'attitudine a produrre un impatto per il pluralismo informativo (quindi, a generare effetti sulla formazione dell'opinione dei cittadini)¹⁹.

L'immissione nel sistema informativo di contenuti *fake* avviene essenzialmente in tre passaggi (v. Figura 2.1), ossia la creazione del messaggio che si vuole trasmettere, la

¹⁹ Sulla base dei sei succitati elementi sono chiaramente esclusi tutti quei contenuti di natura satirica, o anche falsi ma che non hanno alcun intento doloso, e che caratterizzano molto spesso le conversazioni sui social. Questo tipo di contenuti falsi rientra nel concetto di mis-informazione, cfr. Capitolo 2 del Rapporto "[Le strategie di disinformazione online e la filiera di contenuti fake](#)", cit.

produzione del contenuto in cui il messaggio viene incorporato e trasformato in un prodotto informativo, e, quindi, la distribuzione di quest'ultimo (a queste fasi si aggiunge poi la valorizzazione).

Figura 2.1 - La filiera dei contenuti *fake*



Nella fase di creazione, viene elaborato il messaggio da veicolare mediante il contenuto *fake*; questo assume caratteristiche diverse in ragione dell'obiettivo degli ideatori e a seconda del target cui è destinata la strategia di disinformazione. In generale, il messaggio viene ideato in modo da attivare l'audience cui è rivolto, coinvolgendola anche nella diffusione ulteriore del contenuto.

Nella fase di produzione del contenuto, il messaggio viene trasformato in un prodotto informativo, che può assumere la forma di un testo (ad esempio, un *post* o un articolo), di un'immagine, di un video, oppure una combinazione di questi.

Infine, nella fase di distribuzione, il contenuto *fake* viene pubblicato online e reso quindi disponibile. La distribuzione si concretizza attraverso uno o più canali, tipicamente un sito web o una piattaforma online (in special modo di social network), e consente di collocare il contenuto *fake* nel contesto mediatico prescelto (vale a dire, la rete di contenuti che circolano sui diversi media online e offline attorno e/o insieme ai contenuti *fake*), che svolge un ruolo importante soprattutto nel conferire attendibilità al messaggio.

Alla realizzazione di contenuti *fake* partecipano generalmente diversi soggetti. Si distinguono in particolare gli ideatori del contenuto o di un'intera campagna di disinformazione (singoli individui; imprese editoriali e non; organizzazioni con finalità culturali, ideologiche, politiche, criminali; servizi di intelligence; governi; Stati), e gli esecutori delle diverse attività lungo la filiera. Gli esecutori sono coloro che contribuiscono direttamente alla creazione e produzione del contenuto *fake*, e talvolta coincidono con gli ideatori dell'iniziativa. Possono essere singoli individui, gruppi di utenti reclutati *ad hoc*, organizzazioni vere e proprie specializzate nella progettazione e implementazione di campagne di disinformazione.

Molto spesso, inoltre, i soggetti che persegono strategie di disinformazione possono agire con l'ausilio di meccanismi automatici come i bot, che consentono la pubblicazione e distribuzione dei contenuti *fake* attraverso una molteplicità di *account* falsi o falsi profili social. Alla divulgazione dei contenuti *fake* possono concorrere altresì gli stessi destinatari dei contenuti, laddove, anche inconsapevolmente, si trovano a rilanciarli favorendone la diffusione.

Considerando le caratteristiche della filiera che conduce all'introduzione di disinformazione nell'ecosistema nazionale, in questo Capitolo viene condotta un'analisi sulla quantità di contenuti *fake* prodotti e sulla varietà delle tematiche oggetto di disinformazione in Italia.

2.2. L'ENTITÀ DELLA DISINFORMAZIONE PRODOTTA IN ITALIA

A prescindere dalle altre condizioni, l'efficienza di un sistema informativo è tanto maggiore quanto meno esteso è il volume di disinformazione presente nello stesso. In altri termini, dal momento che la disinformazione conduce il cittadino ad assumere visioni distorte, formulare congetture errate e prendere, pertanto, decisioni economiche, politiche e sociali non informate, la quantità socialmente ottimale di disinformazione immessa nel sistema informativo è pari zero. Infatti, se l'informazione è un bene economico e sociale con le caratteristiche descritte nel Capitolo precedente (cfr. par. 1.1), la disinformazione è di per sé un “male”, ossia un “bene” la cui maggiore disponibilità fa diminuire la soddisfazione (utilità) del consumatore.

Conseguentemente, non essendo valutabile l'aspetto qualitativo (i contenuti *fake* hanno già in partenza una connotazione negativa), il dato quantitativo diviene un elemento fondamentale dell'analisi sulla produzione di disinformazione in un'ottica di tutela del pluralismo, un indicatore immediato delle dimensioni del fenomeno e della propria dinamica evolutiva.

Le successive analisi sul volume di disinformazione online prodotto in Italia tengono conto del numero complessivo di documenti informativi generati mensilmente da siti web, e pagine/*account* di social network (Facebook e Twitter) inclusi nelle liste redatte da fonti esterne specializzate nell'attività di *debunking* e usualmente utilizzate negli studi scientifici in materia. Nello specifico, i documenti considerati afferiscono a ciascun articolo (nel caso dei siti web), *post* o *tweet* (nel caso dei social network) pubblicato da 335 fonti (cfr. APPENDICE METODOLOGICA)²⁰.

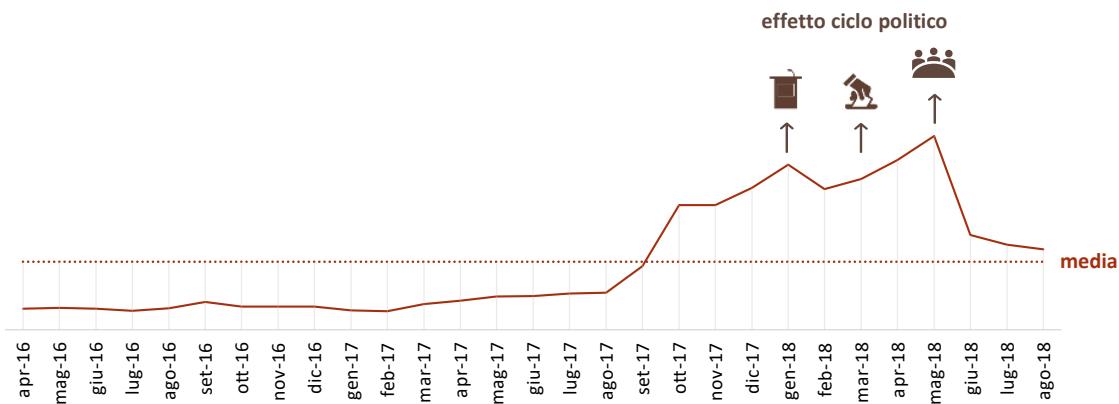
La Figura 2.2 mostra l'andamento mensile della suddetta stima della quantità di contenuti *fake* prodotti in Italia da aprile 2016 ad agosto 2018. Il *trend* registrato consente di suddividere idealmente il grafico in due parti: la prima, che arriva fino ad agosto 2017, in cui la stima del livello di disinformazione prodotta nel mese, comunque

²⁰Tale metodologia permette di pervenire a una stima della produzione di contenuti *fake* prodotti in Italia in un mese medio. Il valore così calcolato rappresenta la produzione di notizie da parte di tutte quelle fonti che sono state individuate da addetti ai lavori come soggetti creatori di contenuti falsi. In tal senso, è una stima di carattere soggettivo (ossia calcolata partendo dai soggetti produttori di *fake*) e, probabilmente, per difetto (non essendo possibile includere l'universo di tutti i produttori di notizie false in Italia). Analogamente, questo valore non include i casi di disinformazione di carattere incidentale prodotta da testate che producono informazione reale. D'altra parte, in questa stima rientrano tutti i contenuti informativi provenienti da fonti *fake*, anche quelli che potrebbero invece presentare caratteristiche di veridicità. Nel Capitolo 3, invece, l'esercizio si soffermerà sul carattere oggettivo della disinformazione, analizzando un campione di notizie reali e uno di *fake news*, a prescindere dal soggetto che ha prodotto e distribuito i relativi contenuti.

positivo, si attesta su un valore piuttosto costante e inferiore alla media dell'intero periodo; la seconda, che comprende l'ultimo anno, in cui la stima della quantità di contenuti *fake* prodotti segna un'impennata e assume in ogni mese un valore superiore alla media dell'intero periodo.

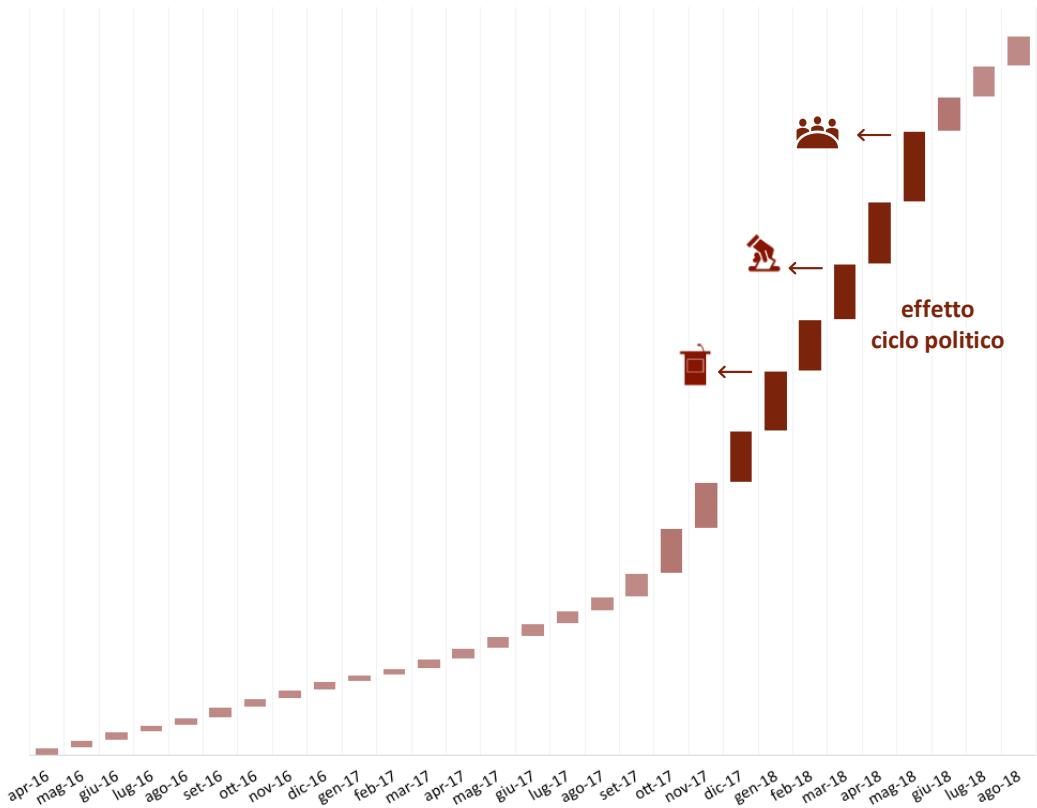
Il dato denota una evidente criticità del sistema, poiché non soltanto il volume di disinformazione prodotto non è nullo ma nel tempo presenta una crescita del proprio livello medio.

Figura 2.2 - Andamento mensile della disinformazione prodotta in Italia



Fonte: elaborazioni Agcom su dati Volocom

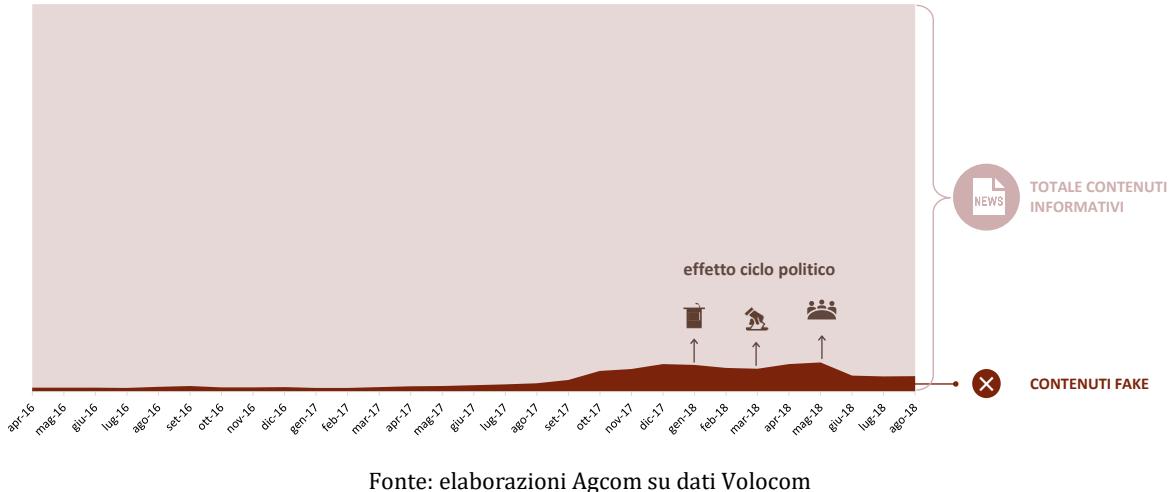
In sostanza, se si guarda all'ammontare di contenuti *fake* cumulativamente prodotto nel periodo indagato (cfr. Figura 2.3), si osserva chiaramente una curva sempre crescente, espressione della somma di quantità mensili mai pari a zero.

Figura 2.3 - Quantità incrementale di contenuti *fake* prodotti nel mese

Fonte: elaborazioni Agcom su dati Volocom

Per di più, nell'ultimo anno, il volume di disinformazione risulta nel complesso crescente sia in termini assoluti sia in termini relativi. Al riguardo, come evidenzia la Figura 2.4, la stima dell'incidenza media dei contenuti *fake* sul totale dei contenuti informativi del sistema nazionale subisce un incremento, passando dall'1% del periodo che arriva fino ad agosto 2017 (2% se si considerano soltanto i contenuti online) a circa il 6% degli ultimi dodici mesi (10% dei contenuti online).

Tali considerazioni, peraltro, devono tener conto del fatto che, nelle analisi compiute, non sono inclusi i documenti messi in circolazione nel sistema informativo da singoli utenti né da testate tradizionali (cfr. nota 20), che, nelle attività di condivisione svolte attraverso le piattaforme online, possono contribuire non solo alla diffusione di contenuti *fake* prodotti da altre fonti (anche modificandoli e trasformandoli in altri contenuti) ma anche introdurne di nuovi, accrescendo ulteriormente il volume di disinformazione (sul ruolo svolto dagli utenti online v. Capitolo 4).

Figura 2.4 - Incidenza dei contenuti *fake* nel sistema informativo

Fonte: elaborazioni Agcom su dati Volocom

La lettura dei risultati esposti nelle figure precedenti evidenzia una ancora maggiore problematicità dal punto di vista del pluralismo informativo, se si pone l'accento sulla coincidenza dell'incremento di disinformazione con gli accadimenti politici del Paese. L'avvicendarsi del nuovo ciclo politico (dallo scioglimento delle Camere del precedente governo fino all'insediamento del governo attuale, passando per le elezioni politiche e la relativa campagna elettorale) ha prodotto un effetto propulsivo sulla quantità di disinformazione rilasciata, che registra dei picchi proprio in concomitanza dell'avvio della campagna elettorale, dello svolgimento delle elezioni e della formazione del nuovo governo (cfr. Figura 2.2 e Figura 2.3).

Come osservato nel par. 1.2, anche nel caso dell'informazione la contingenza del ciclo politico ha generato un aumento del volume di contenuti offerti. Tuttavia, la crescita dell'incidenza dei contenuti *fake* sul sistema informativo complessivo (cfr. Figura 2.4) proprio in corrispondenza del ciclo politico è indice di un aumento della disinformazione più che proporzionale rispetto all'incremento dell'informazione.

Come si avrà modo di approfondire nel par. 2.3, la criticità dello scenario delineato sotto l'aspetto quantitativo trova conferma anche sotto quello contenutistico, laddove le tematiche inerenti alla politica e al governo rappresentano l'oggetto più ricorrente dei contenuti *fake* prodotti nel 2018 (cfr. Figura 2.6).

In definitiva, le evidenze emerse descrivono un sistema informativo che soffre la presenza di un volume di disinformazione che, da un lato, sembra essersi assestato su

un valore mediamente più alto e, dall'altro, ha raggiunto il livello massimo durante l'ultimo ciclo politico. I rischi associati a tali tendenze sono evidentemente legati all'attitudine dei contenuti *fake*, al pari di quelli aventi ad oggetto notizie reali, ad incidere sulla conoscenza, sulle opinioni e sui punti di vista dei cittadini in merito alle questioni più varie. E quando gli argomenti trattati sono quelli afferenti alla politica, specie nel periodo elettorale, il rischio è che informazioni false, manipolate o infondate concorrono a formare le espressioni di voto dei cittadini, e, per mezzo delle stesse, a influire sugli esiti elettorali, da cui dipendono le linee politiche adottate a livello nazionale.

Particolare rilevanza assume, dunque, l'esame concreto dei generi e delle tematiche dei contenuti *fake*, i quali, considerando le motivazioni (politiche/ideologiche/economiche) che generalmente si celano dietro la messa in atto di una strategia di disinformazione, producono un impatto negativo sul pluralismo tanto maggiore quanto più gli argomenti coinvolti sono quelli in grado di influenzare l'opinione pubblica. In ragione di ciò, si è scelto quindi di dedicare il prossimo paragrafo all'analisi specifica delle tematiche della disinformazione prodotta in Italia.

2.3. LE TEMATICHE DELLA DISINFORMAZIONE

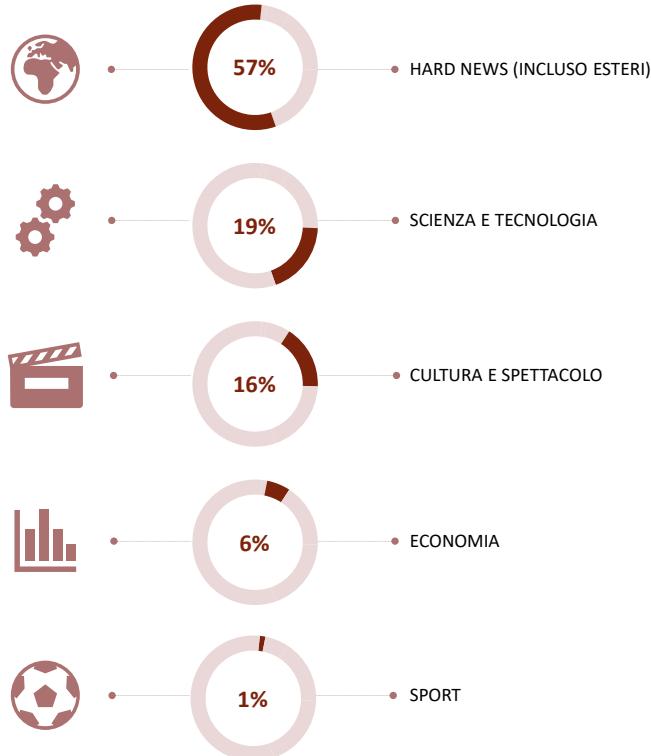
Una prima indicazione sugli argomenti che in Italia sono più frequentemente oggetto di contenuti *fake* è offerta dalla distribuzione per categorie della disinformazione prodotta.

In proposito, la Figura 2.5 mostra come mediamente, nell'intero periodo analizzato, più della metà dei contenuti divulgati dalle fonti *fake* considerate riguardi proprio le notizie contraddistinte da una maggiore valenza sul piano del pluralismo (*"hard news"*), riconducibili a fatti di cronaca, politica e accadimenti di rilevanza internazionale.

Segue la categoria delle notizie di carattere scientifico e tecnologico, oggetto del 19% dei contenuti *fake*. Si tratta di notizie idonee a produrre effetti sulla sfera ideologica dei cittadini, e che peraltro suscitano grande interesse nella popolazione, come dimostrano le Figura 1.8 e Figura 1.9. Interesse, come anticipato, non adeguatamente soddisfatto dall'offerta di informazione né per quantità (riguardando soltanto il 7% dell'output prodotto, con una significativa carenza rispetto alla domanda, cfr. Figura 1.8) né per livello di preparazione specialistica posseduto dai giornalisti che se ne occupano (cfr. Figura 1.7), e che quindi rischia più delle altre categorie di essere colmato attraverso la fruizione di contenuti *fake*.

Quanto alle restanti categorie di contenuti (“cultura e spettacolo”, “economia” e “sport”), le medesime coprono congiuntamente meno di un quarto dell'offerta complessiva di disinformazione, con le notizie sportive che rappresentano la porzione decisamente più esigua anche in ragione della loro più facile verificabilità da parte dei cittadini e, dunque, della loro minor idoneità ad essere manipolate o artatamente falsificate.

Figura 2.5 - Distribuzione dell'offerta di contenuti *fake* per categoria



Nota: la categoria "hard news" include le notizie di cronaca, politica e quelle di rilevanza internazionale.

Fonte: elaborazioni Agcom su dati Volocom

In ogni caso, accanto alla distribuzione per categorie, l'importanza delle implicazioni associate al diffondersi di disinformazione richiede uno studio più specifico delle tematiche oggetto dei contenuti *fake*, che consenta di verificare quali sono gli argomenti che registrano una maggior concentrazione di false informazioni.

A tal fine, a partire dall'insieme di tutti i documenti prodotti nel 2018 (dal 1° gennaio al 31 agosto) dai siti web²¹ individuati come fonti di contenuti *fake* e inclusi nelle presenti analisi, è stata adottata un'impostazione metodologica riconducibile al c.d. *topic modelling*. Un *topic model* è un modello statistico per l'individuazione automatica degli argomenti che compaiono in una raccolta di documenti. In particolare, la classificazione del testo dell'insieme di contenuti *fake* in determinati argomenti è stata ottenuta

²¹ Nelle analisi svolte, non sono stati considerati i documenti afferenti a pagine e *account* di social network, in quanto la lunghezza ridotta dei contenuti (*post*, *tweet*) rende gli stessi poco adatti ad essere trattati con le tecniche del *topic modelling*.

attraverso l'utilizzo di un modello prodotto via LDA (*Latent Dirichlet Allocation*) - un algoritmo di apprendimento automatico non supervisionato che tiene conto della frequenza e del contesto dei termini utilizzati nella raccolta di documenti. Lo stesso è poi alla base del sistema interattivo di visualizzazione LDavis²², con cui sono stati prodotti i risultati esposti nelle figure successive. Più precisamente, il sistema LDavis fornisce una visione complessiva degli argomenti individuati (e del modo in cui differiscono l'uno dall'altro), consentendo allo stesso tempo un'analisi approfondita dei termini salienti associati a ciascuno di essi. Il sistema, infatti, offre una visualizzazione compatta, in cui, da una parte, vengono mostrati gli argomenti prevalenti, tenendo conto della loro frequenza e del loro posizionamento nel piano, e, dall'altra, sono evidenziati i termini salienti (i primi 30), che presentano cioè il maggior numero di occorrenze all'interno dei relativi argomenti, riportati in ordine decrescente di frequenza.

L'applicazione di queste tecniche di analisi testuale ai contenuti *fake* ha consentito di compiere una mappatura delle principali aree tematiche oggetto di disinformazione in Italia nel 2018. Più specificamente, le figure successive offrono una rappresentazione dei 9 argomenti per i quali è stata riscontrata la maggior presenza nei contenuti *fake*. Ogni tematica è rappresentata da un cerchio, la cui ampiezza segnala la frequenza di occorrenza nell'insieme complessivo dei documenti esaminati. Le tematiche, inoltre, assumono una posizione nello spazio cartesiano, collocandosi attorno agli assi, orizzontale e verticale, i quali individuano le due dimensioni principali che spiegano la variabilità dei dati. Nel dettaglio, in questo caso, la prima dimensione è riportata sull'asse orizzontale ed esprime il livello di attualità delle notizie riconducibili a una tematica, che assume un valore crescente andando da destra verso sinistra (da minore attualità a stretta attualità). La seconda dimensione, riportata sull'asse verticale, esprime l'estensione soggettiva delle notizie inerenti a una data tematica e assume valore crescente procedendo dal basso verso l'alto (da un'estensione molto piccola, circoscritta a singoli soggetti, a un'estensione molto ampia, che arriva a coinvolgere istituzioni o Paesi).

La Figura 2.6 propone una visione complessiva delle tematiche predominanti nei contenuti *fake*, molte delle quali appartenenti alla categoria "hard news". Spicca in

²² Cfr. C. Sievert, K. E. Shirley (2014), "LDavis: A method for visualizing and interpreting topics", *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pp. 63–70. Cfr. anche D. M. Blei, John Lafferty (2009), "Visualizing Topics with Multi-Word Expressions", *arXiv:0907.1013v1 [stat.ML]*; J. M. Bischof, E. M. Airola (2012), "Summarizing topical content with word frequency and exclusivity", ICML; J. Chuang, C. D. Manning, J. Heer (2012), "Termite: Visualization Techniques for Assessing Textual Topic Models", *Advanced Visual Interfaces*.

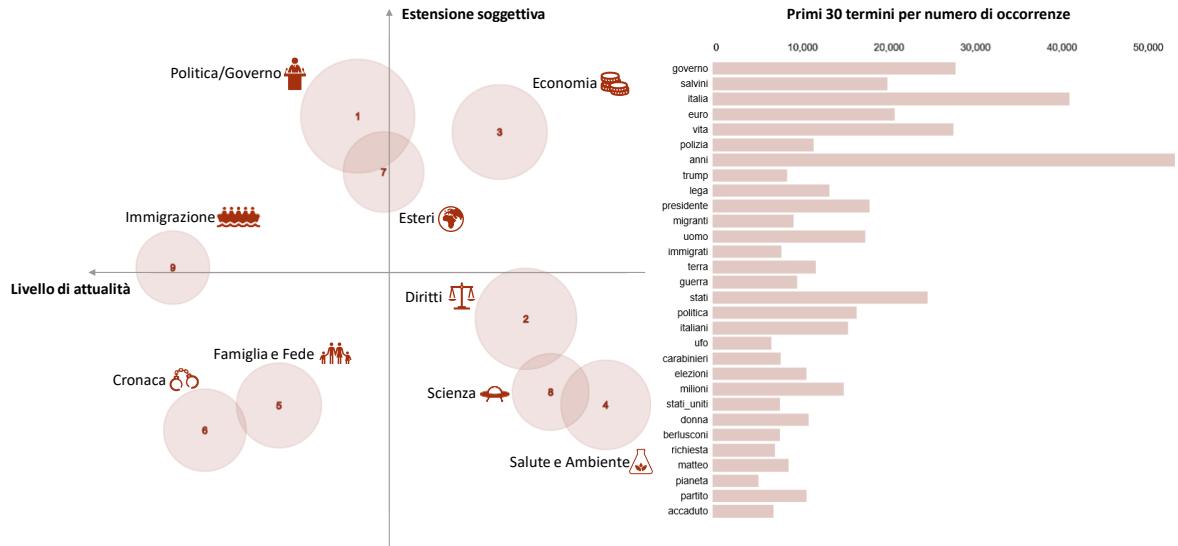
particolar modo l'argomento “Politica/Governo”, che assume le dimensioni maggiori, tanto che molti dei termini che compaiono tra i 30 più frequenti dell'intero *dataset* sono legati alle vicende politiche e istituzionali del Paese. Questa evidenza denota senz'altro una certa concentrazione di disinformazione attorno alle questioni politiche nei primi 8 mesi del 2018 e contribuisce ad acuire le criticità già emerse nel paragrafo precedente laddove si rilevava una maggiore produzione di contenuti *fake* durante il periodo elettorale. Il risultato di fatto suggerisce che, non soltanto nei mesi immediatamente precedenti e seguenti alla chiamata alle urne degli italiani la disinformazione ha visto un deciso incremento, ma anche che i contenuti *fake* divulgati hanno avuto ad oggetto notizie strettamente connesse alla sfera politica.

Oltre alla politica, la Figura 2.6 identifica tra le macro-tematiche prevalenti della disinformazione in Italia ulteriori argomenti rientranti nella categoria delle *hard news*, quali “Cronaca”, “Immigrazione” ed “Esteri”.

Nel complesso dei contenuti *fake*, grande rilevanza, in termini di distribuzione di frequenza, assumono altresì gli argomenti connessi all'ambito sociale, come “Diritti” e “Famiglia e Fede”, e a quello dell’“Economia”. Inoltre, coerentemente a quanto rilevato dalla Figura 2.5, un'ampia porzione di disinformazione riguarda le materie scientifico-tecnologiche, con particolare riguardo a quelle che investono il campo sanitario e ambientale.

In linea generale, è possibile osservare come le 9 macro-tematiche individuate abbiano come elemento comune il fatto di richiamare argomenti polarizzanti²³, ossia in grado di creare o accentuare la separazione degli individui in gruppi distinti (questo argomento sarà trattato nell'ambito del Capitolo 4 a cui si rimanda). Sono quindi argomenti dibattuti e di particolare interesse, dotati di un certo livello di contagiosità, quindi atti a trasferire stati emotivi e percezioni tra gli utenti. Di conseguenza, si tratta di tematiche che si prestano a divenire oggetto di propagazione virale soprattutto per mezzo delle piattaforme online (anche ad opera degli utenti stessi che rilanciano e condividono i contenuti).

²³ Per polarizzazione ideologica si intende il risultato individuale di quel processo sociale di separazione e frammentazione della popolazione in gruppi distinti, separati e non comunicanti tra loro su tematiche divisive. Cfr. Sunstein, C. R. (2002), “The Law of Group Polarization”, *Journal of Political Philosophy*, 10(2), pp. 175–195; Idem (2017), *#Republic. Divided Democracy in the Age of Social Media*, Princeton University Press. Il grado di polarizzazione di un contenuto può essere chiaramente associato alla sua capacità di generare fenomeni di *echo chamber* o *confirmation bias* in specifici gruppi di utenti (cfr. Quattrociocchi, W., Scala, A. & Sunstein, C. (2016), “Echo Chambers on Facebook”, *working paper*).

Figura 2.6 - Principali tematiche oggetto dei contenuti *fake* (2018)

Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

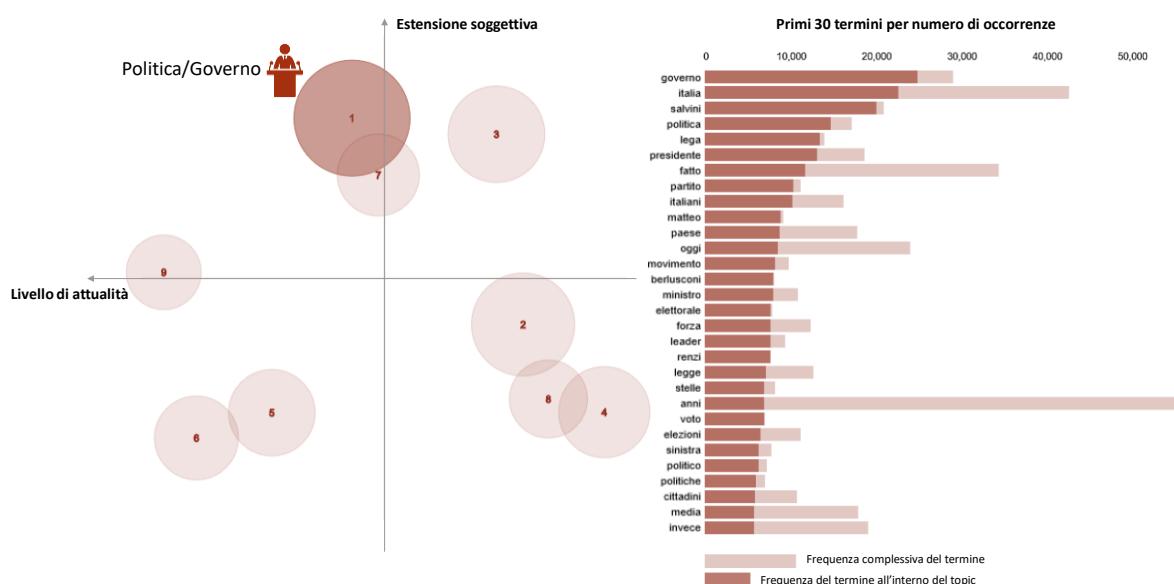
Aumentando ulteriormente il grado di granularità dell’analisi, è possibile esaminare la composizione di ciascuna tematica della disinformazione, ricavando, in base ai termini salienti associati alle stesse, indicazioni circa l’oggetto specifico delle notizie diffuse.

In tal senso, la Figura 2.7 esplora la composizione dell’argomento “Politica/Governo”. Dall’analisi dei termini più ricorrenti, si evince come l’attenzione dei contenuti *fake* ascrivibili a tale tematica sia stata posta su tre aspetti principali: *i*) i soggetti politici, in special modo partiti (si vedano ad esempio i termini “partito”, “lega”, “movimento”, “stelle”, “forza”, “italia”), esponenti e candidati alle elezioni del 4 marzo 2018 (“salvini”, “renzi”, ...); *ii*) i soggetti istituzionali (“governo”, “ministri”, “presidente”); *iii*) lo svolgimento delle elezioni politiche del 4 marzo e la partecipazione politica (“elettorale”, “voto”, “elezioni”, “politiche”, ...).

Tali aspetti sono certamente quelli che assumo la rilevanza più elevata sotto il profilo del pluralismo, ancor più considerando che gran parte del periodo cui si riferisce l’analisi è stato interessato dall’esplicarsi del nuovo ciclo politico, con i cittadini chiamati

alle urne ed esposti al rischio di ricavare informazioni distorte su cui fondare le proprie espressioni di voto.

Figura 2.7 - Termini salienti dell'argomento “Politica/Governo” (2018)



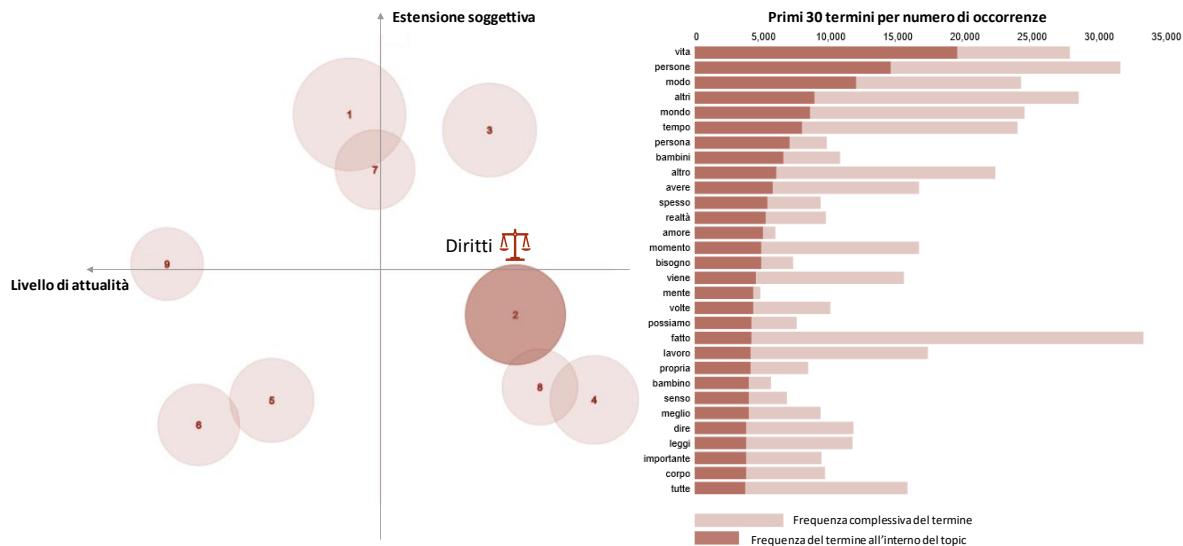
Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

La seconda tematica più estesa nel panorama dei contenuti *fake* è quella che attiene ai "Diritti", intesi in senso ampio.

La disamina dei termini salienti offerta dalla

Figura 2.8 pone l'accento sulla constatazione che molto spesso sono le notizie legate ai diritti e bisogni della persona ("vita", "persone", "bambini", "amore", "lavoro", ...) e all'esigenza di vederne riconosciuta la tutela ("legge", "bisogno", ...) a costituire la leva della disinformazione. Disinformazione che, in casi come questo, si rivela frutto di strategie mirate a far presa sulla sfera emozionale e sul coinvolgimento diretto dei cittadini.

Figura 2.8 - Termini salienti dell'argomento "Diritti" (2018)

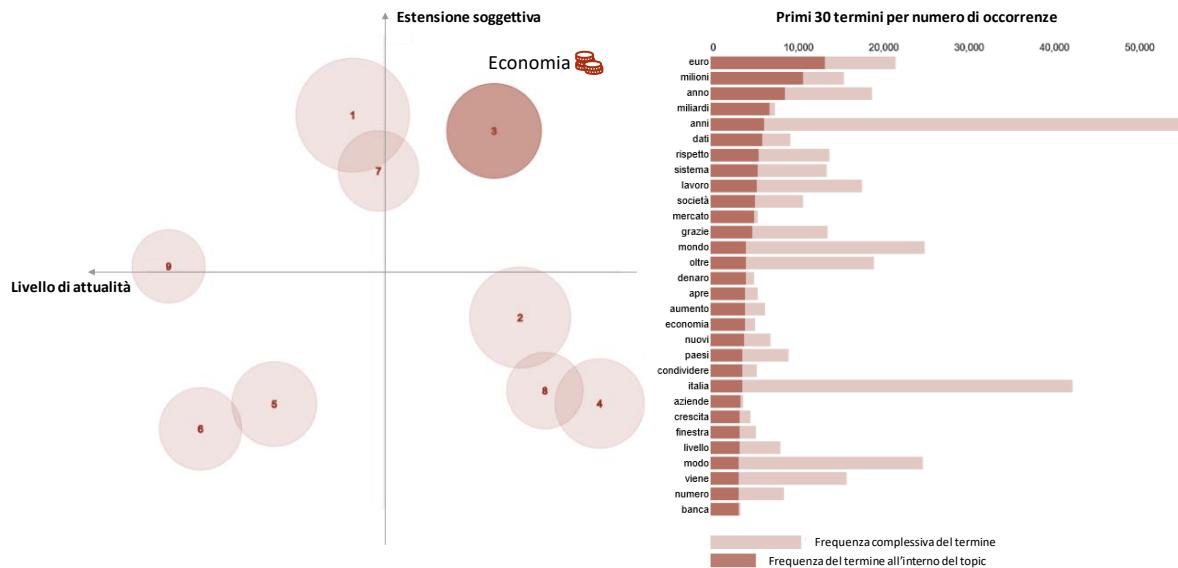


Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

I contenuti *fake* della tematica "Economia", come mostra la Figura 2.9, appaiono focalizzati, da un lato, sull'andamento economico generale ("sistema", "crescita", ...) e sugli aspetti di natura finanziaria ("euro", "milioni", "miliardi", "sistema", "mercato", "banche", ...), dall'altro, su questioni legate al contesto aziendale ("società", "aziende", ...) e lavorativo ("lavoro", "livello", "aumento", "finestra", ...).

È interessante notare come solitamente questi argomenti siano avvertiti con particolare enfasi dai cittadini, chiamando in causa la propria condizione economica e le sorti dei propri risparmi, e siano posti al centro di intenti e programmi politici anche in campagna elettorale, divenendo variabili decisionali fondanti per le preferenze degli individui. Circostanza che indubbiamente tende ad aggravare ancor di più i rischi di ricadute negative associate alla disinformazione.

Figura 2.9 - Termini salienti dell'argomento “Economia” (2018)



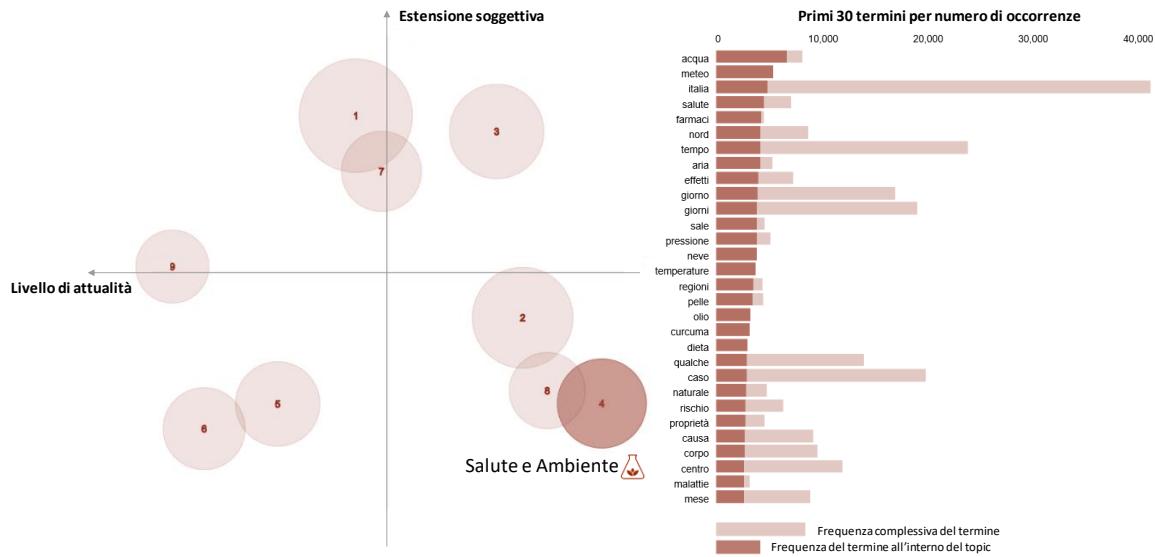
Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

Ulteriore impulso per la produzione di contenuti *fake* proviene da temi che, pur riguardando materie diverse, coinvolgono l’ideologia delle persone.

Al riguardo, l’argomento “Salute e Ambiente” (cfr. Figura 2.10) abbraccia sicuramente notizie atte ad influenzare gli orientamenti e i punti di vista dei cittadini relativamente a fatti che suscitano apprensione, come la salute e le condizioni climatiche e ambientali (avverse).

Testimonianza ne è la frequenza elevata di termini che richiamano questioni sulle quali si riscontra tipicamente uno stato di preoccupazione degli individui: “salute”, “farmaci”, “dieta”, “pelle”, “corpo”, “malattie”, oppure “acqua”, “tempo”, “aria”, “temperature”, peraltro associati a parole quali “effetti”, “rischio”, “causa”, che ne accentuano ulteriormente la connotazione negativa.

Figura 2.10 - Termini salienti dell'argomento "Salute e Ambiente" (2018)

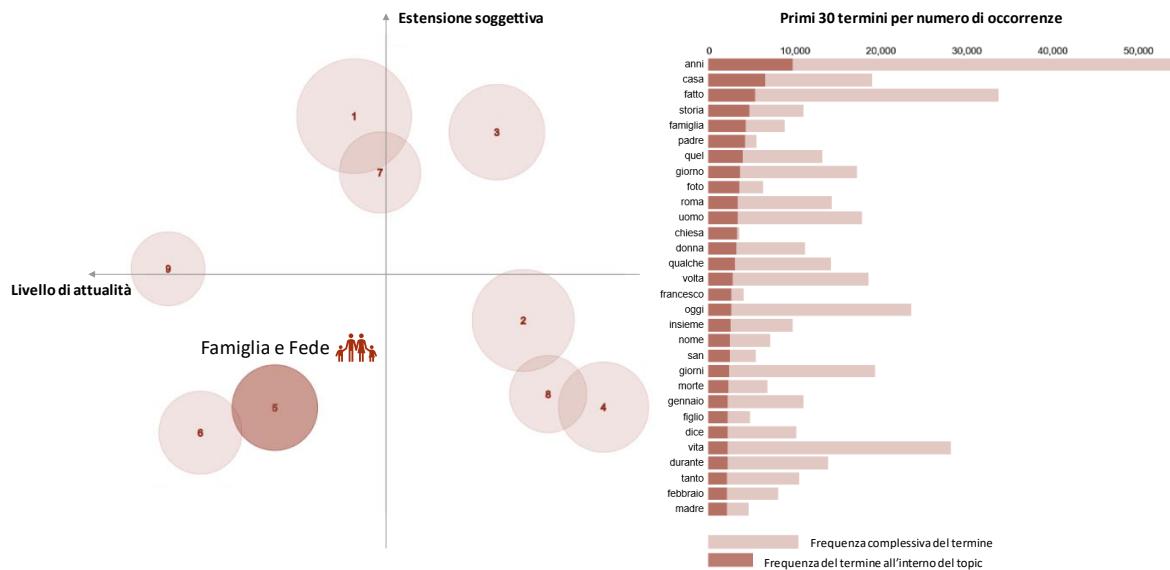


Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

Analogamente, il profilo ideologico è alla base di strategie di disinformazione inerenti alla fede religiosa. Molti contenuti *fake* nel 2018 hanno avuto ad oggetto notizie su "Famiglia e Fede", tanto che l'argomento è risultato il quinto più trattato dai siti web individuati come fonti di disinformazione.

In particolare, il maggior numero di notizie false, manipolate o distorte, afferenti alla tematica in questione, ha riguardato la chiesa ("chiesa", "francesco", "san", ...), ma anche questioni esistenziali (come "vita" e "morte"), oltre che la famiglia ("casa", "famiglia", "uomo", "donna", "insieme", ...), da sempre riconosciuto tra i valori fondanti delle comunità religiose (cfr. Figura 2.11).

Figura 2.11 - Termini salienti dell'argomento “Famiglia e Fede” (2018)

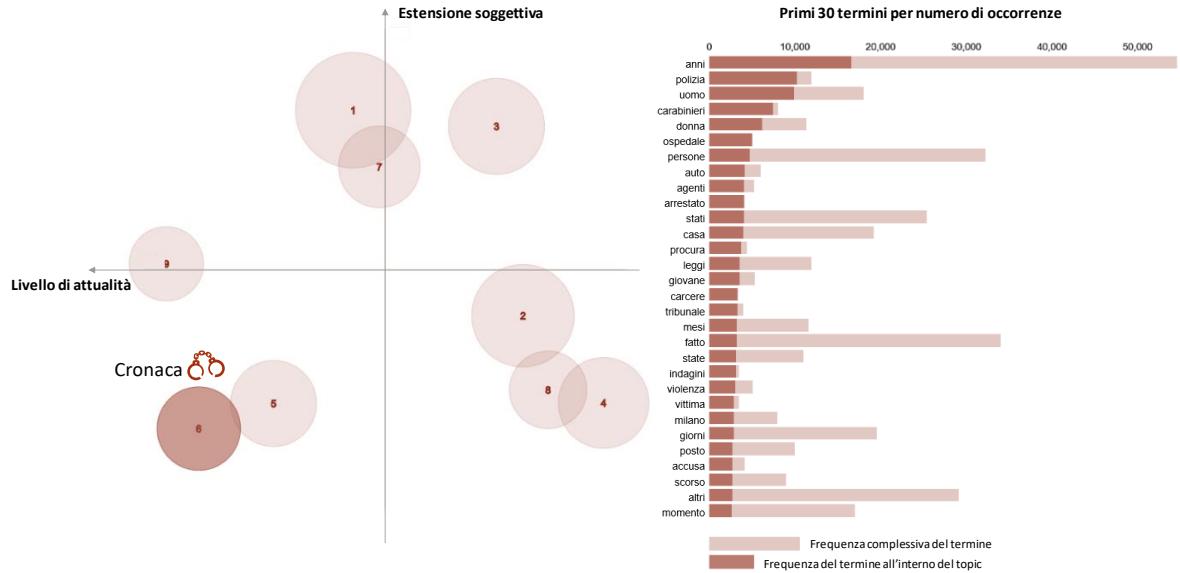


Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

Quanto alla disinformazione prodotta in merito alla tematica “Cronaca”, si osserva che la stessa, pur non sembrando concentrata su uno o pochi fatti specifici che hanno assunto rilevanza a livello nazionale nel 2018, si rivela incentrata prevalentemente sugli accadimenti di cronaca nera e giudiziaria.

Dall’analisi dei termini salienti riportati nella Figura 2.12, infatti, risulta evidente l’abbondanza di notizie che riguardano tanto le modalità con cui può essere avvenuto un reato (“persone”, “ospedale”, “violenza”, “vittima”, ...), quanto l’intervento delle forze dell’ordine (“polizia”, “carabinieri”, “agenti”, “arrestato”, “indagini” ...) e le vicende giudiziarie connesse (“procura”, “leggi”, “tribunale”, “accusa”, “carcere”, ...).

Figura 2.12 - Termini salienti dell'argomento "Cronaca" (2018)

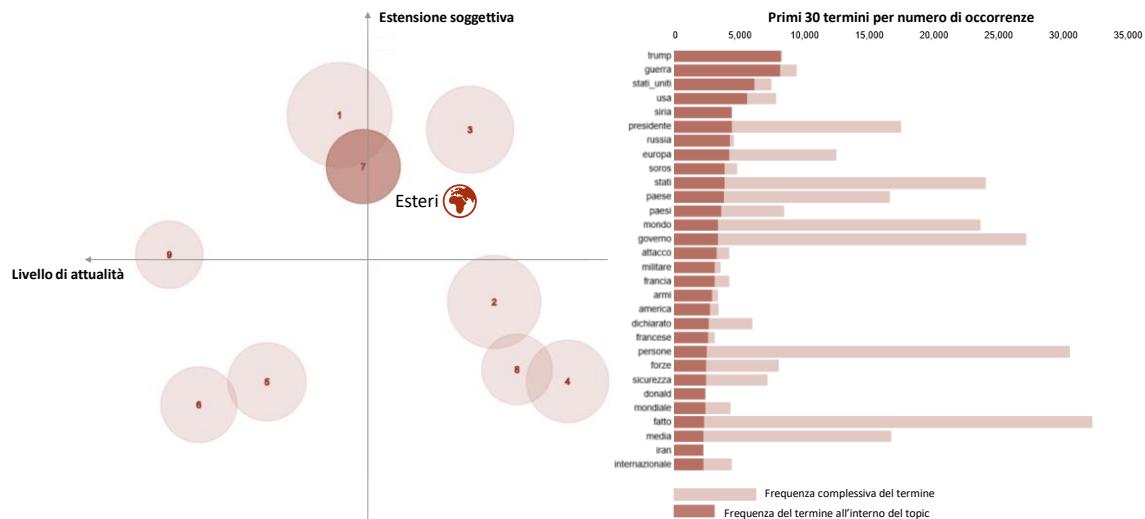


Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDavis (Sievert e Shirley; 2014)

Passando agli avvenimenti internazionali ("Esteri"), la Figura 2.13 mostra come i contenuti *fake* prodotti in Italia nei primi otto mesi del 2018 abbiano avuto ad oggetto in prevalenza notizie relative a conflitti e attacchi militari ("guerra", "siria", "attacco", "militare", "armi", "iran", ...).

Particolare risalto, inoltre, è stato conferito alle vicende degli Stati Uniti, come dimostra la frequenza di occorrenza di termini salienti quali "trump", "stati_uniti", "usa", "america", "donald", ...).

Figura 2.13 - Termini salienti dell'argomento “Esteri” (2018)



Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

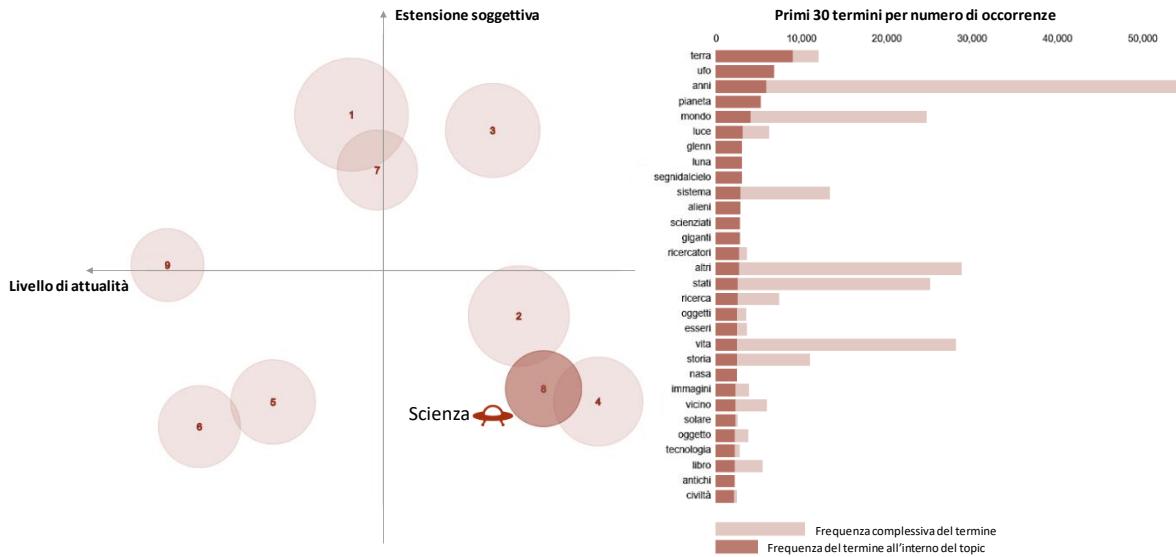
In merito alla disinformazione in campo scientifico, emerge che, oltre a riguardare notizie su salute e ambiente, l'argomento appare fortemente impernato anche sulle questioni e i fenomeni che attengono all'astronomia e allo spazio.

In tal senso, la Figura 2.14 evidenzia come l'argomento “Scienza” includa in larga parte notizie su pianeti e corpi celesti (“terra”, “pianeta”, “sistema”, “solare”, ...); ufo e presenze aliene nello spazio (“ufo”, “alieni”, “vita”, ...).

Si rileva, peraltro, come tra quelli salienti compaiano altresì termini che richiamano direttamente il lavoro di ricerca scientifica, presumibilmente invocato nel tentativo di conferire veridicità alle teorie esposte negli articoli e spesso sottostanti a strategie di disinformazione che seguono logiche complottiste e cospirazioniste²⁴.

²⁴ Sullo studio della diffusione online di contenuti *fake* aventi ad oggetto notizie scientifiche, cfr., tra gli altri, A. Bessi et al. (2015), “Science vs Conspiracy: Collective Narratives in the Age of Misinformation”, *PLoS ONE* 10(2); A. Bessi et al. (2016), “Homophily and Polarization in the Age of Misinformation”, *The European Physical Journal Special Topics*, 225(10); M. Del Vicario et al. (2016), “The Spreading of Misinformation Online”, *Proceedings of the National Academy of Science* 113(3); F. Zollo et al. (2017), “Debunking in a World of Tribes”, *PLoS ONE* 12(7).

Figura 2.14 - Termini salienti dell'argomento "Scienza" (2018)



Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

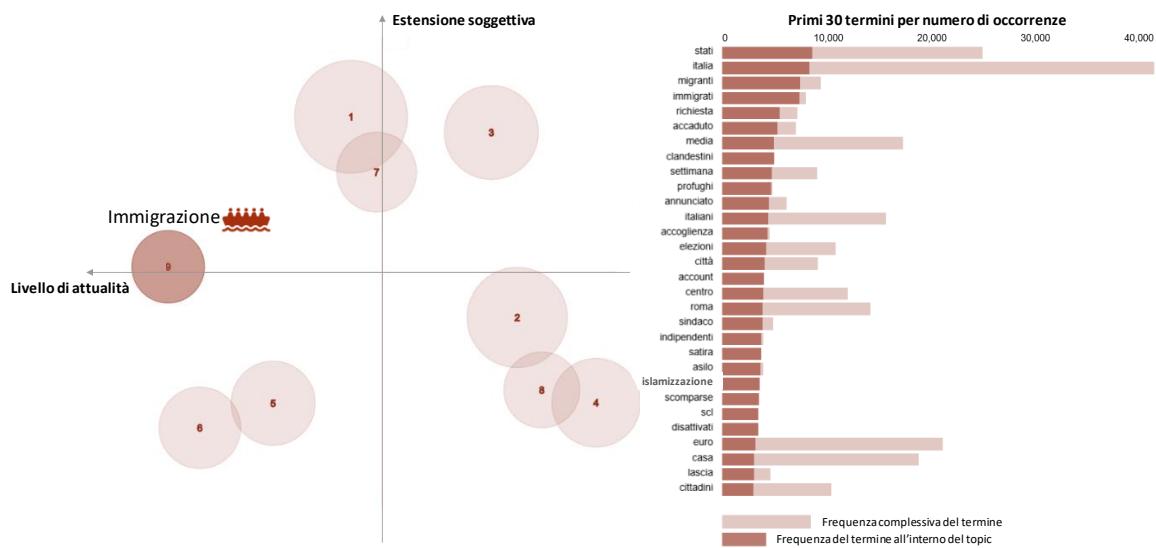
Infine, tra le tematiche principali oggetto di disinformazione figura, come anticipato, l’“Immigrazione”. Nel corso del 2018, molteplici notizie sull’argomento hanno caratterizzato il dibattito politico e acceso i toni della campagna elettorale. I fatti di cronaca (tra cui quello emblematico degli spari contro i migranti a Macerata) e le misure adottate dal nuovo governo (come la disposizione della chiusura dei porti alle navi umanitarie) hanno diviso l’opinione pubblica.

In questo scenario, ha trovato terreno fertile la proliferazione di contenuti *fake*. Il più delle volte, le motivazioni sottostanti alla produzione di disinformazione sul tema dell’immigrazione sono di natura ideologica e i contenuti veicolati sono atti a innescare campagne d’odio (soprattutto a sfondo razziale) e infondere intolleranza.

Come rappresentato dalla Figura 2.15, la disinformazione sull’immigrazione nel periodo esaminato è stata costruita per lo più attorno a notizie legate all’accoglienza dei migranti da parte dell’Italia e degli altri Stati. Tra i termini salienti della tematica, infatti, molti

sono riconducibili a parole evocative di tale questione, come “stati”, “italia”, “migranti”, “immigrati”, “richiesta”, “asilo”, “profughi”, “clandestini”.

Figura 2.15 - Termini salienti dell'argomento “Immigrazione” (2018)



Fonte: elaborazioni Agcom su dati Volocom con sistema di visualizzazione LDAvis (Sievert e Shirley; 2014)

LA DIFFUSIONE DI NOTIZIE REALI E FALSE

3



3.1. INTRODUZIONE

Nei capitoli precedenti, il sistema informativo è stato esaminato sotto l'aspetto della produzione, intesa come quantità, qualità e varietà dei contenuti generati dalle fonti attive in Italia sui diversi mezzi, quali quotidiani, Tv, radio, siti web e piattaforme online (in particolare, social network).

A partire da una metodologia di carattere soggettivo (ossia che parte dall'individuazione delle fonti di informazione e disinformazione) sono state individuate le caratteristiche e messe in luce le criticità dell'offerta informativa proposta da ciascun mezzo (anche in relazione alla domanda e alle risorse produttive impiegate, cfr. Capitolo 1), nonché della disinformazione prodotta e rilasciata nel sistema nazionale (cfr. Capitolo 2).

Cambiando necessariamente la prospettiva di analisi, ossia partendo dalle notizie, quindi dall'oggetto dei contenuti informativi e falsi prodotti, l'obiettivo di questo capitolo è identificare le peculiarità che contraddistinguono la diffusione dell'informazione e della disinformazione ad opera delle predette fonti.

Nei paragrafi che seguono, pertanto, saranno indagate le modalità di propagazione delle singole notizie, reali e false, nel tempo e nello spazio definito dall'insieme dei media che compongono il sistema informativo. L'attenzione sarà rivolta in particolare alla definizione del ciclo di vita delle notizie (permanenza nel sistema informativo), e alla comprensione degli elementi distintivi (durata, presenza effettiva, copertura) delle medesime, in base alla tipologia (genere) di notizia e al mezzo di diffusione.

Sotto il profilo metodologico, come detto, le analisi esposte nel prosieguo sono state compiute applicando un criterio di carattere oggettivo. In tal senso, sono stati costruiti *dataset* per due campioni di notizie: uno per le notizie reali, uno per le quelle false²⁵.

²⁵ Nel passaggio dal primo al secondo metodo di analisi, è implicito un cambiamento che deve essere ben compreso dal lettore. Nel metodo soggettivo (capitoli 1 e 2), ad essere censite sono state le fonti di informazione e disinformazione. Essendo le prime ben definite, l'analisi dell'informazione è stata effettuata a livello pressoché censuale. Sulle seconde, invece, come detto, ci si è affidati a liste di siti e pagine/*account* di disinformazione individuati dalla comunità scientifica. Questi ultimi rappresentano, necessariamente, un sotto-insieme di tutte le fonti di disinformazione attive. In questo capitolo, che ha l'obiettivo di analizzare il percorso dinamico delle singole notizie, si è necessariamente individuato un campione di notizie reali e false, utilizzando la procedura descritta nel testo. L'analisi, quindi, non solo fa riferimento al carattere oggettivo dell'informazione e della disinformazione, ma utilizza anche un campione, per forza di cose limitato (anche se, come sarà illustrato, altamente rappresentativo), di tutte le notizie presenti nell'ecosistema. Di conseguenza, l'analisi in esame è condizionata dal campione selezionato e dal periodo considerato: infatti, ogni anno, le principali notizie possono appartenere a categorie diverse a seconda degli eventi accaduti nel periodo (elezioni politiche, notizie di cronaca, eventi internazionali, eventi sportivi, ...).

In modo da assicurare la robustezza e la significatività dei risultati, particolare cura è stata dedicata alla definizione dei campioni di notizie, le quali sono state selezionate tenendo conto della loro rilevanza nel panorama informativo e della rappresentatività per categoria di genere trattato, a prescindere dai soggetti che hanno prodotto contenuti informativi o falsi aventi ad oggetto le stesse.

Nello specifico, per quel che concerne le notizie reali, sono stati individuati i principali accadimenti avvenuti in un arco temporale di 12 mesi. Per ogni mese, sono state incluse nel campione le 5 notizie più rilevanti di ciascuna categoria (cronaca, politica, esteri, economia, scienza, cultura, spettacolo e sport), annotando la data esatta (t_0) in cui si è verificato il fatto e definendo per ognuna di esse l'insieme di parole chiave più idoneo a costituire la *query*²⁶ di ricerca per estrarre le relative occorrenze giornaliere dal *database* documentale utilizzato (cfr. APPENDICE METODOLOGICA).

Analogamente a quanto si rinviene diffusamente in letteratura, la rilevanza delle notizie è stata valutata attraverso lo studio dei *trend* delle ricerche effettuate dagli utenti su internet, interpretando le crescite improvvise di ricerche (e quindi di interesse) per un determinato argomento da parte degli utenti come il riflesso dell'andamento di una notizia, inseguitrice di un fatto o avvenimento che stava avendo luogo in Italia o nel mondo. In generale, per ogni mese, sono state inserite nel campione le notizie che, nell'ambito di una categoria, hanno registrato i più alti tassi di ricerca²⁷.

²⁶ Per “*query*” si intende l’elaborazione di una sequenza di parole chiave attraverso connettivi logici (operatori booleani), secondo le regole descritte dal linguaggio *Apache Lucene Query Parser Syntax*.

²⁷ In particolare, per le analisi dei *trend* di ricerca, si è fatto riferimento allo strumento *Google Trends*, che consente di conoscere la frequenza di ricerca online di una determinata parola o insieme di termini, da cui risalire alla notizia di riferimento. Nel dettaglio, i risultati sulle tendenze di ricerca sono stati estratti attraverso la funzione “Esplora” di *Google Trends*, che permette di circoscrivere l’analisi ad una specifica area geografica (in questo caso, l’Italia), ad uno specifico arco temporale (in questo caso, il singolo mese), e di selezionare le categorie tematiche in base alle quali sono aggregate le tendenze di ricerca, oltre che la fonte mediante la quale gli utenti effettuano la ricerca, ossia “Ricerca Google” e “Google News”.

Nella letteratura scientifica è ampio oramai il ricorso a questa metodologia, e in particolare all’uso di *Google Trends*, per quantificare e/o prevedere fenomeni riguardanti l’economia (v., ad esempio, H. Choi e H. Varian, 2012, “Predicting the Present with Google Trends”, *Economic Record*), la finanza (T. Preis, 2013, “Quantifying Trading Behavior in Financial Markets Using Google Trends”, *Scientific Records*), la medicina (S. V. Nuti, B. Wayda, I. Ranasinghe, S. Wang, R. P. Dreyer, S. I. Chen, K. Murugiah, 2014, “The Use of Google Trends in Health Care Research: A Systematic Review”, *Plos one*), l’ecosistema digitale (L. Kristoufek 2013, “BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era”, *Scientific Reports*). Recentemente, la metodologia è stata estesa alla misurazione dell’interesse degli utenti per le notizie (v., ad esempio, L. T. P. Nghiêm, S. K. Papworth, F. K. S. Lim, L. R. Carrasco, 2016, “Analysis of the Capacity of Google Trends to Measure Interest in Conservation Topics and the Role of Online News”, *Plos one*).

Diversamente, per la selezione del campione di notizie false, sono state utilizzate fonti esterne di *debunking*, specializzate nell'applicazione di metodologie scientifiche per la scoperta e la smentita di *fake news*²⁸. Più precisamente, per l'individuazione delle notizie false, sono stati raccolti gli articoli pubblicati nel corso di un anno che fossero relativi a fatti falsi, infondati, manipolati o riportati in maniera non veritiera, dotati di effettiva “notiziabilità”, ossia atti ad essere trasformati in notizia e recepiti dal contesto mediatico e dal pubblico come tali. Nella scelta delle notizie da includere nel campione, inoltre, si è tenuto conto della rappresentatività per categoria di genere trattato, in base a quanto emerso dall'esame della distribuzione dell'offerta (cfr. Figura 2.5). Come per le notizie reali, anche per ciascuna di quelle false del campione è stata annotata la data di origine del fatto trattato (t_0) e formulata la *query* di ricerca più adatta per estrarre le corrispondenti occorrenze dal *database* documentale impiegato per le analisi del presente Rapporto.

Una volta definiti i campioni di notizie reali e false secondo le modalità sopra illustrate, sono stati costruiti i due *dataset* contenenti le occorrenze giornaliere registrate da ciascuna notizia su ogni fonte informativa, nell'intervallo temporale di 45 giorni comprendente i 15 giorni prima del fatto (t_0) e i 30 giorni successivi²⁹.

La definizione del t_0 ha consentito di centrare tutte le notizie rispetto a una coordinata, così da rendere possibile il confronto tra le differenti notizie e trarre indicazioni dirette circa il ciclo di vita caratteristico delle notizie, la loro persistenza nel sistema informativo e la loro copertura mediatica nel tempo.

Le evidenze emerse dalle analisi condotte sui *dataset* di notizie, reali e false, sono presentate nei paragrafi successivi, dedicati rispettivamente all'esame della diffusione dell'informazione e della disinformazione.

²⁸ Per l'individuazione delle *fake news* si è fatto riferimento in particolare al sito butac.it, rilevando, in ogni caso, come gran parte delle notizie false oggetto degli articoli prodotti da questa fonte fossero generalmente oggetto di trattazione anche da parte di diversi altri siti web di *debunking*.

²⁹ Si consideri che molto spesso prima del fatto in sé è possibile comunque trovare un certo “rumore” informativo sull'argomento (v. infra, par. 3.2, Figura 3.1).

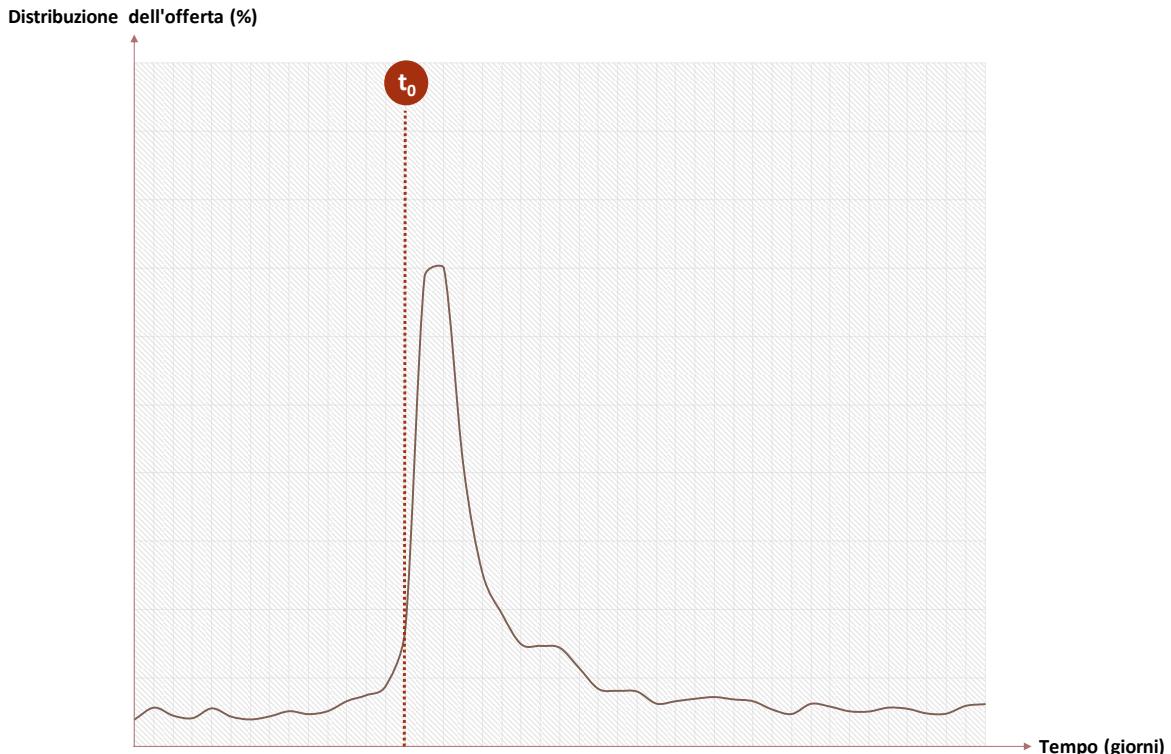
3.2. IL CICLO DI VITA E LE MODALITÀ DI DIFFUSIONE DELLE NOTIZIE REALI

L'ampiezza del sistema italiano dell'informazione, in termini di contenuti informativi prodotti, è stata oggetto di stima nel Capitolo 1, nell'ambito del quale sono state analizzate le caratteristiche che attengono specificamente alla produzione di informazione (non solo quantità, ma anche qualità e varietà). Stante quanto esaminato e discusso nella prima parte del Rapporto, in questo paragrafo, seguendo l'andamento di un numeroso campione di notizie principali (selezionato sulla base dei criteri sopra esposti, cfr. par. 3.1), si passa ad esaminare nel dettaglio come, mediamente, una singola notizia viene trattata e diffusa all'interno del sistema informativo.

Una prima indicazione sulle modalità di diffusione delle notizie viene offerta dalla rappresentazione del ciclo di vita delle stesse, inteso come l'andamento medio giornaliero della quantità di contenuti informativi riguardanti la notizia, divulgati dai diversi mezzi di comunicazione.

Come mostra la Figura 3.1, la propagazione di una notizia può avere inizio anche prima del giorno esatto (t_0) in cui avviene il fatto relativo alla medesima. Molto dipende dalla tipologia di notizia, che può essere legata a un accadimento improvviso o imprevedibile, ovvero ad un evento programmato o prevedibile. In quest'ultimo caso, chiaramente, è facile riscontrare la copertura della notizia anche nei giorni antecedenti al t_0 . Tuttavia, è nei giorni che vanno dal t_0 in poi che si verifica la maggiore distribuzione dell'offerta informativa. In particolare, la concentrazione di contenuti sulla notizia, che tendenzialmente subisce un'impennata in corrispondenza del t_0 , registra il suo massimo nei due giorni successivi a quello in cui ha avuto luogo l'accadimento (assumendo, in ciascuno dei due giorni, un valore prossimo al 14% del totale), quando tutti i mezzi arrivano a diffonderla. Come si approfondirà nel prosieguo, infatti, ciascun mezzo ha le proprie specificità e copre la notizia in tempi e modi che possono differire tra loro anche in base alla frequenza di aggiornamento della propria offerta (i quotidiani, ad esempio, essendo pubblicati una volta al giorno, possono coprire la notizia a partire dal giorno successivo al t_0 , tranne nei casi in cui il fatto avvenga entro le prime ore della giornata), oltre che alla linea editoriale prescelta e allo spazio destinato all'informazione.

Figura 3.1 - Ciclo di vita di una notizia



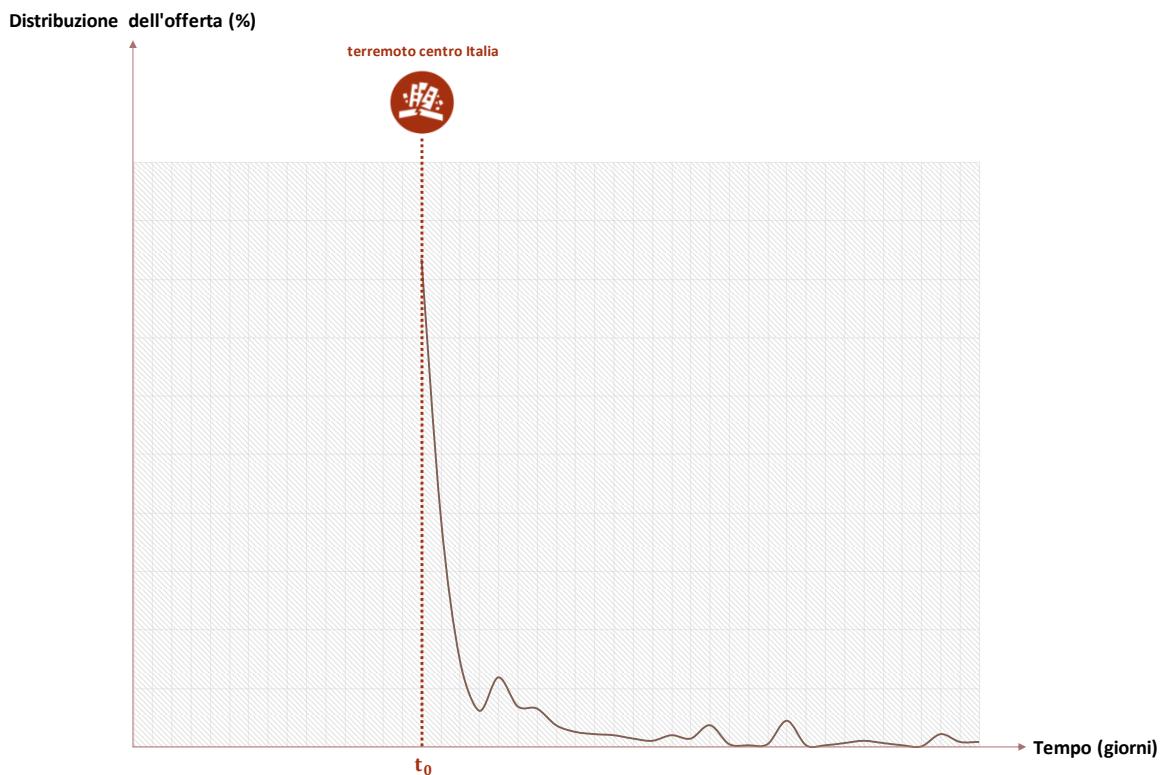
Fonte: elaborazioni Agcom su dati Volocom

Il t_0 , dunque, segna sicuramente il momento essenziale attorno a cui si sviluppa il ciclo di vita della notizia. Più in generale, si distinguono le notizie per le quali il t_0 è “assoluto” da quelle per le quali è “relativo”. Nel primo caso, il t_0 segna l’inizio del ciclo di vita della notizia. Questo avviene per notizie che presentano un andamento come quello riportato, a titolo esemplificativo, nella figura successiva.

Nel dettaglio, la Figura 3.2 espone il ciclo di vita della notizia sul terremoto che nella notte del 24 agosto 2016 colpì il centro Italia. In questo caso, si tratta di un fatto non prevedibile, per cui la notizia non riporta anticipazioni ma viene immessa nel sistema informativo negli istanti immediatamente successivi all’accadimento, peraltro con una copertura che, raggiungendo il punto di massimo direttamente nel giorno t_0 , lascia presumere la trattazione da parte di tutti i media in concomitanza del t_0 .

Un ciclo di vita con caratteristiche analoghe si rinviene tipicamente per notizie quali fatti di cronaca nera, calamità naturali, attacchi terroristici, ...

Figura 3.2 - Ciclo di vita della notizia “terremoto centro Italia”
(agosto-settembre 2016)

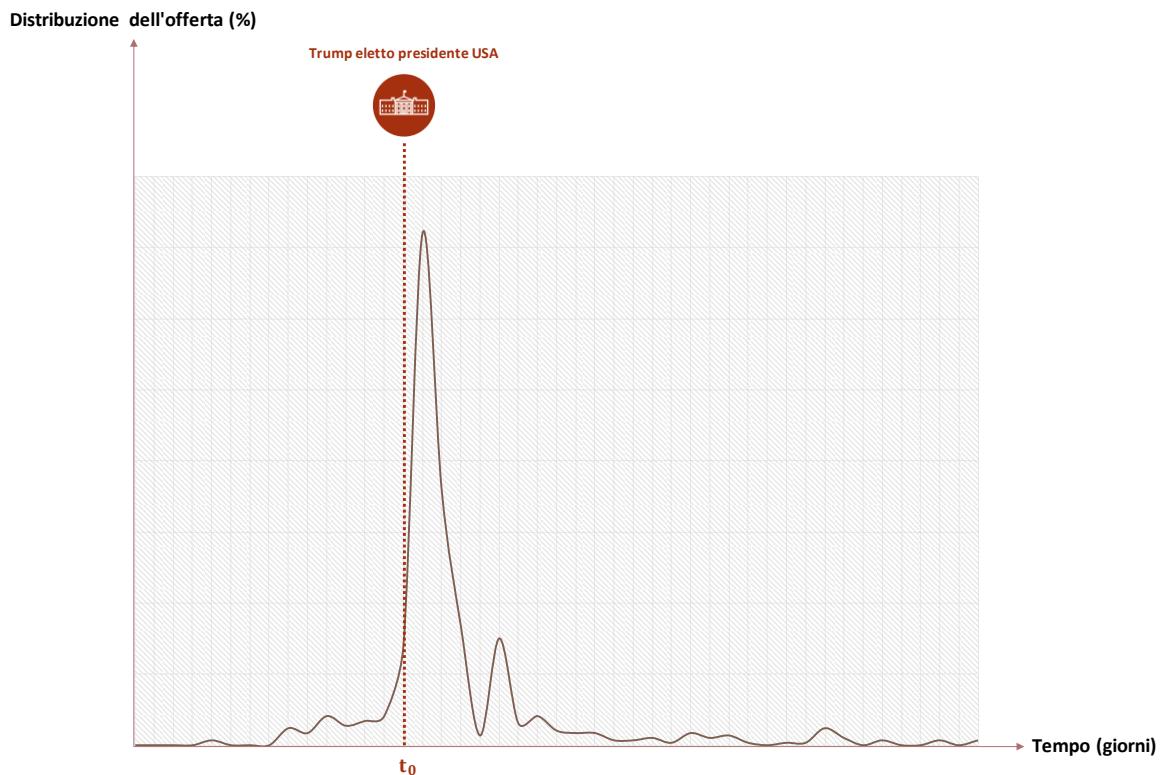


Fonte: elaborazioni Agcom su dati Volocom

Quando il t_0 è relativo, invece, il ciclo di vita di una notizia inizia prima che l'evento di cui la stessa tratta si verifichi. Un esempio di notizia che mostra occorrenze nel sistema informativo prima del t_0 è riportato nella Figura 3.3, che si riferisce all'elezione di Donald Trump a presidente degli Stati Uniti. Chiaramente, la notizia, essendo connessa a un accadimento prevedibile, presenta anticipazioni legate soprattutto alla trattazione dello scenario prospettato in caso di vittoria del candidato alle elezioni.

È frequente riscontrare anticipazioni nella produzione di contenuti informativi su una notizia nel caso di competizioni politiche, ma anche di eventi sportivi, culturali o di manifestazioni. Lo stesso può valere per operazioni societarie o eventi astronomici e, in generale, per tutti quegli accadimenti prefissati o comunque attesi, in cui la copertura mediatica prima del t_0 può essere proprio finalizzata a preannunciare l'avvenimento.

Figura 3.3 - Ciclo di vita della notizia “Trump eletto presidente USA”
(novembre-dicembre 2016)



Fonte: elaborazioni Agcom su dati Volocom

Aumentando il grado di profondità dell’analisi, è possibile individuare degli indicatori specifici di diffusione delle notizie, in modo da delineare in maniera più puntuale le modalità di propagazione dell’informazione ad opera dei diversi mezzi (cfr. Figura 3.4 eFigura 3.5).

Nello specifico, sono stati considerati tre indicatori principali, riconducibili a:

- la “durata media di una notizia”, corrispondente alla distanza che mediamente intercorre tra il primo e l’ultimo giorno in cui una notizia registra almeno una occorrenza, ossia è oggetto di almeno un contenuto informativo. Tale indicatore, in sostanza, quantifica la lunghezza del ciclo di vita della notizia, intesa come giorni di permanenza dell’informazione nel sistema nazionale;
- la “presenza effettiva di una notizia” all’interno del sistema informativo, corrispondente al numero medio di giorni, anche non consecutivi, in cui una

notizia è trattata da almeno un documento informativo. Tale misura, rapportata alla durata media, fornisce un'indicazione circa l'ampiezza della porzione del ciclo di vita della notizia concretamente interessata dalla presenza di contenuti informativi su un fatto. L'indicatore costituisce, pertanto, un elemento importante per quantificare il lasso di tempo in cui il cittadino può potenzialmente essere esposto all'informazione diretta su un accadimento e, dunque, entrarvi in contatto attraverso la fruizione dei mezzi che la veicolano;

- la “copertura di una notizia” nel giorno medio, corrispondente all'indice che esprime il rapporto (moltiplicato per 100) tra il numero di occorrenze giornaliere registrate mediamente da una notizia su uno specifico mezzo e il numero di occorrenze giornaliere registrate mediamente da una notizia su un qualsiasi mezzo. Di conseguenza, valori superiori a 100 indicano una copertura più alta della media, valori pari a 100 indicano un livello di copertura in linea con la media complessiva dei mezzi di informazione, e valori inferiori a 100 indicano una copertura più bassa della media. La copertura così definita costituisce, quindi, una misura relativa di quanto una notizia sia seguita da un mezzo, a seconda della linea editoriale, della rapidità di aggiornamento, dell'interesse e della rilevanza dei fatti per il pubblico.

I tre indicatori consentono di qualificare e quantificare la diffusione dell'informazione seguendo la trattazione delle notizie sia da un punto di vista statico (quanto viene trattata in un giorno medio) che dinamico (quanti giorni la notizia rimane nel sistema informativo e per quanti giorni effettivamente viene trattata).

Al riguardo, la Figura 3.4 mostra come, in media, tra tutti i mezzi, le fonti editoriali online presentino il maggior numero di giorni di permanenza di una notizia (25), nel corso dei quali l'accadimento è concretamente oggetto di trattazione per circa il 60% del periodo. Peraltro, per le medesime fonti, anche in considerazione dell'elevata frequenza di aggiornamento che le caratterizza, si osserva il più alto tasso di copertura giornaliera della notizia. In ogni caso, vale rilevare, per un verso, che il dato è il risultato dell'aggregazione di diverse fonti (quali siti web di quotidiani, di radio, Tv e testate esclusivamente online) e che, come specificato nel par. 3.1, la selezione del campione di notizie utilizzato per l'analisi è basata sui *trend* delle ricerche effettuate su internet, circostanza atta a incidere positivamente sui valori riferiti alle fonti online (che potrebbe condurre a una parziale sovra-rappresentazione del fenomeno).

Per altro verso, gli indici di diffusione devono essere letti congiuntamente a quanto rilevato nel Capitolo 1 (cfr. Figure Figura 1.2, Figura 1.3Figura 1.4), laddove, per le fonti

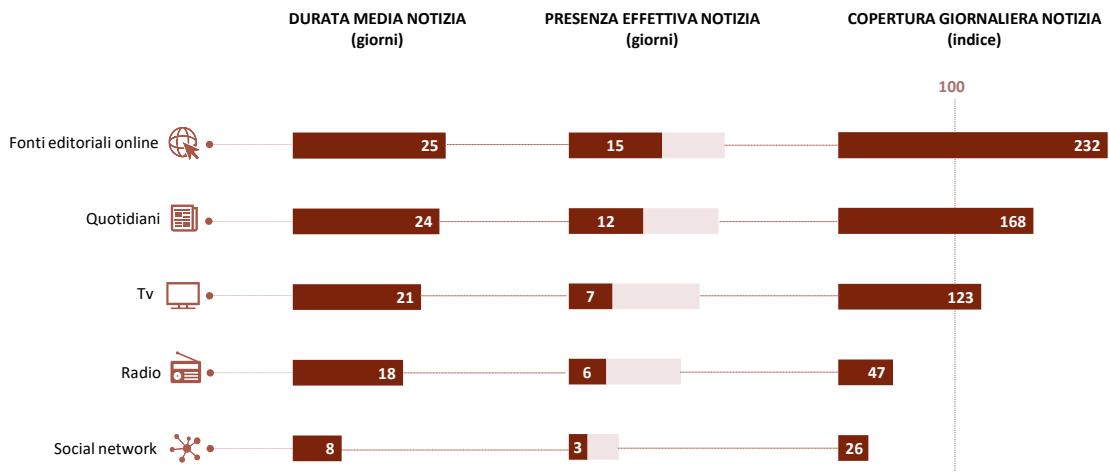
editoriali online, a fronte di una copiosa produzione di contenuti informativi, è stato riscontrato un sovrautilizzo della forza giornalistica impiegata, con il conseguente rischio di ricadute negative sull'accuratezza e l'approfondimento del contenuto generato attorno alla notizia.

Per i quotidiani e la Tv, che invece presentano una minore intensità produttiva dei giornalisti in relazione alla quantità di contenuti informativi offerti, un maggior livello di qualità attesa è accompagnato da una copertura giornaliera della notizia superiore alla media e una lunghezza del ciclo di vita della notizia mediamente superiore ai 20 giorni. Tra i due mezzi, i quotidiani, anche in ragione della loro vocazione pressoché interamente informativa, mostrano valori più alti soprattutto con riferimento ai giorni di presenza (trattazione) effettiva della notizia: 12 (oltre la metà della durata media) contro i 7 della televisione (un terzo della rispettiva durata media).

Nel caso della radio, a un'offerta informativa complessivamente bassa è associata anche una scarsa copertura della singola notizia, sia giornaliera che nel tempo. Tuttavia, come visto in precedenza (cfr. Figura 1.2), il mezzo, analogamente alla Tv, registra anche una bassa intensità produttiva delle risorse giornalistiche.

Per i social network (pagine e *account* di fonti informative e *influencer*), la diffusione della notizia avviene in un arco temporale decisamente più corto. La durata media di una notizia è di appena 8 giorni, in 3 dei quali soltanto si rileva una trattazione effettiva della notizia, unitamente a una copertura media giornaliera di molto inferiore agli altri mezzi. Questi risultati, se confrontati con gli elevati valori della quantità di contenuti informativi complessivamente offerti, forniscono un'indicazione circa la rapidità con cui i *post* e i *tweet* passano dalla trattazione di una notizia ad un'altra. Si tratta, dunque, anche per la natura stessa del prodotto informativo tipica delle piattaforme online e dello spazio entro cui è circoscritto, di una trattazione degli accadimenti più superficiale rispetto agli altri mezzi, come dimostra peraltro il basso grado di qualità atteso appurato nel primo capitolo.

Figura 3.4 - Indici di diffusione di una notizia, per mezzo



La "durata della notizia" è calcolata come distanza media tra il primo e l'ultimo giorno in cui una notizia registra almeno un'occorrenza.

La "presenza effettiva della notizia" esprime il numero medio di giorni, anche non consecutivi, in cui una notizia registra almeno un'occorrenza.

L' "indice di copertura" esprime il rapporto (moltiplicato per 100) tra il numero di occorrenze giornaliere registrate mediamente da una notizia su uno specifico mezzo e il numero di occorrenze giornaliere registrate mediamente da una notizia su un qualsiasi mezzo. Pertanto, valori superiori a 100 indicano una copertura più alta della media, valori pari a 100 indicano un livello di copertura in linea con la media complessiva dei mezzi di informazione, e valori inferiori a 100 indicano una copertura più bassa della media.

Fonte: elaborazioni Agcom su dati Volocom

L'analisi degli indicatori di diffusione dell'informazione da parte dei singoli mezzi può essere ulteriormente approfondita esaminando come la distribuzione di una notizia si concentra nel tempo. A tale scopo è stata utilizzata un'apposita rappresentazione grafica, c.d. *heat map* (cfr. Figura 3.5), che contemporaneamente consente di visualizzare l'aggregazione tra distribuzioni statisticamente omogenee ottenuta tramite procedure di *clustering* (parte sinistra del grafico) e il diverso grado di concentrazione delle occorrenze giornaliere riscontrate per una notizia su ciascun mezzo. Quest'ultimo aspetto viene rilevato attraverso la colorazione assunta in ogni singolo giorno, per cui il livello di concentrazione aumenta man mano che si passa dal rosso vivo (minimo) al verde brillante (massimo).

In linea generale, la Figura 3.5 contribuisce a confermare e qualificare ulteriormente le caratteristiche emerse dallo studio degli indicatori precedenti.

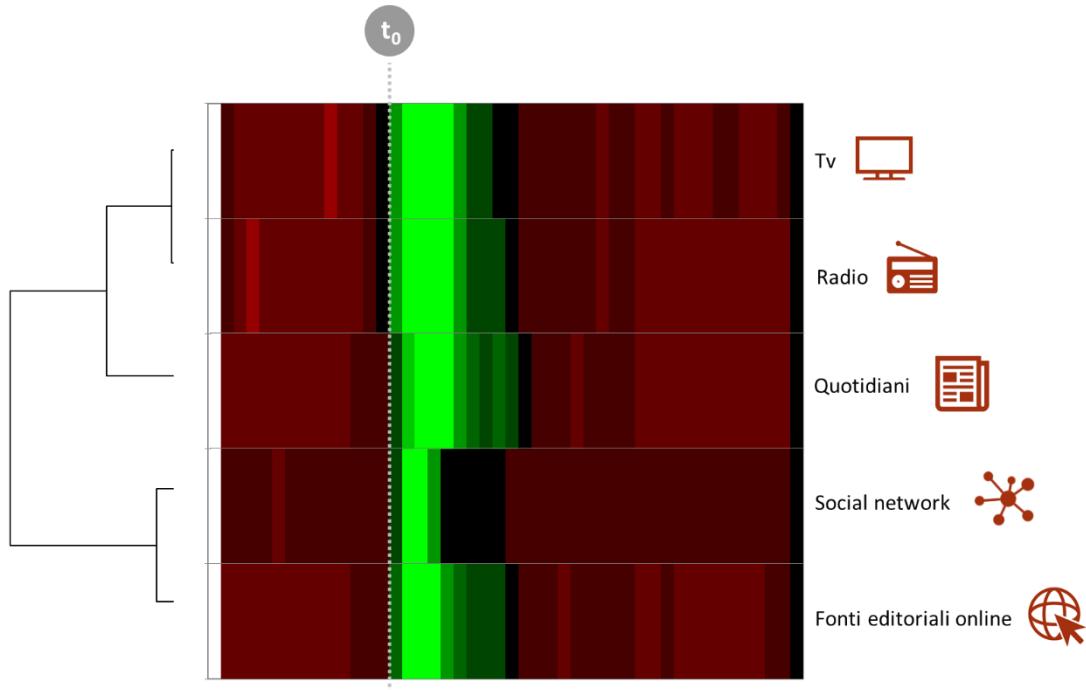
Nello specifico, si evidenzia un andamento simile per le emittenti televisive e radiofoniche, con una distribuzione delle occorrenze per notizia che inizia a intensificarsi in concomitanza del t_0 per poi concentrarsi maggiormente nei 4 giorni successivi.

I quotidiani, scontando, come noto, una frequenza di aggiornamento con cadenza giornaliera, mostrano una distribuzione della notizia nel tempo peculiare, che mediamente inizia ad accentuarsi con un giorno di ritardo rispetto agli altri mezzi, per poi assumere il valore massimo (rispetto al totale del mezzo) nei 3 giorni seguenti. I quotidiani, inoltre, rispetto agli altri mezzi, presentano il maggior numero di giorni contraddistinti da una sfumatura di verde, ad indicare che la distribuzione della notizia è più diluita nel tempo, con una varianza meno accentuata.

Viceversa, le distribuzioni con cui si propagano le notizie su internet sono accomunate da una forte varianza tra i livelli di concentrazione raggiunti a ridosso del t_0 e quelli dei giorni successivi (e sono aggregate quindi dal processo statistico di clusterizzazione). Nel giorno seguente al t_0 , infatti, le due distribuzioni raggiungono un picco di concentrazione che assume un valore più alto rispetto a quello registrato dagli altri mezzi in corrispondenza del proprio massimo, per poi rimanere su una percentuale comunque elevata di occorrenze anche il giorno dopo. La tendenza appare ancor più accentuata per i social network, nella misura in cui il passaggio tra la massima e la minima concentrazione di contenuti informativi riguardanti una notizia avviene in maniera sensibilmente più repentina che negli altri mezzi di informazione.

I medesimi indicatori (durata media, presenza effettiva e copertura media giornaliera della notizia) utilizzati per l'analisi della diffusione di una notizia sui vari mezzi possono essere impiegati per indagare gli elementi caratterizzanti della divulgazione delle varie categorie di notizie, tenendo comunque a mente, nella lettura dei dati che seguono, che i valori riportati sono espressione di tendenze medie ottenute per un campione di notizie principali (sebbene ampio e adeguatamente rappresentativo).

Figura 3.5 - Concentrazione della distribuzione di una notizia nel tempo, per mezzo



Note: per Tv e radio sono state considerate le emittenti nazionali. Le fonti editoriali online includono siti web di quotidiani, di radio, Tv e testate esclusivamente online; mentre i social network comprendono pagine e *account* di fonti informative e *influencer*

La figura (c.d. *heat map*) rappresenta la concentrazione della distribuzione delle occorrenze giornaliere registrate mediamente per una notizia su un mezzo. Il livello di concentrazione aumenta man mano che si passa dal rosso vivo (minimo) al verde brillante (massimo).

Fonte: elaborazioni Agcom su dati Volocom

In tal senso, la Figura 3.6 mostra come mediamente non si registrino differenze molto significative tra *hard news* (considerate complessivamente) e notizie specializzate per quel che attiene ai primi due indicatori: durata media della notizia e ampiezza del periodo di effettiva trattazione.

A distinguersi per maggiore lunghezza del ciclo di vita sono essenzialmente le notizie di politica e di economia, che, rispetto alle altre, iniziano ad essere oggetto di trattazione ben prima del t_0 e permangono nel sistema informativo fino a diversi giorni dopo il t_0 , con le notizie di carattere politico che presentano (in valore assoluto e relativo) la maggiore presenza effettiva all'interno dei contenuti informativi veicolati.

All'opposto, le notizie di cronaca si differenziano per la brevità del loro ciclo di vita che, chiaramente, per la maggior parte dei fatti cui si riferiscono, inizia in corrispondenza del

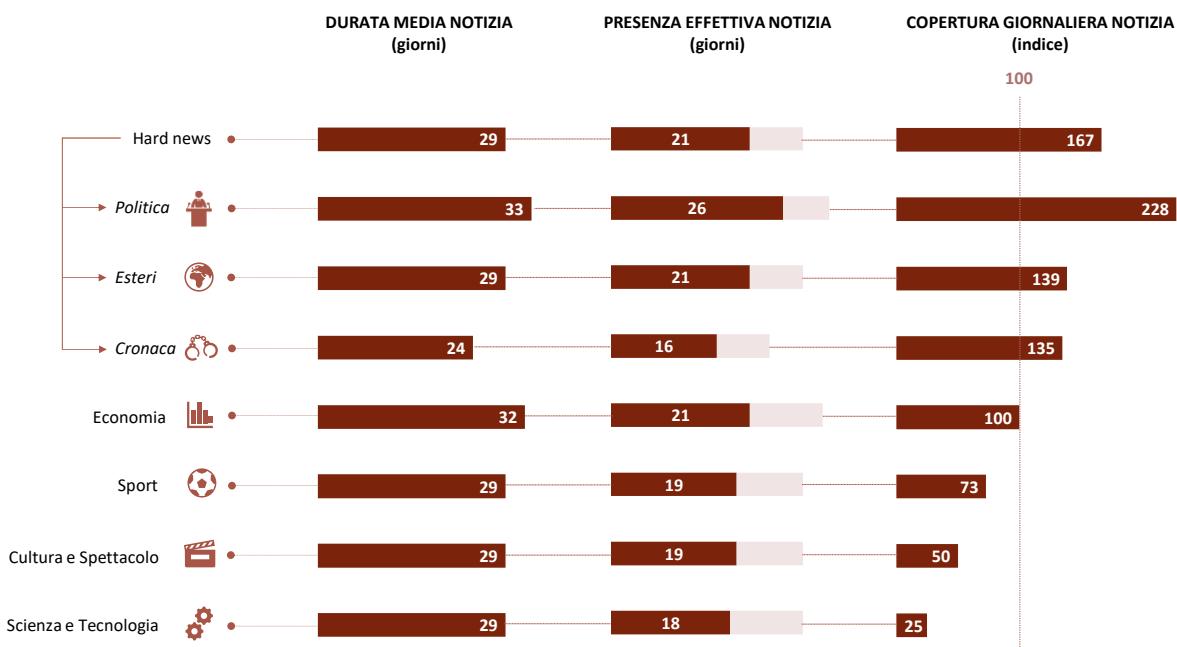
t_0 per terminare in un momento meno distante dal fatto stesso rispetto agli altri generi di notizie.

Il confronto tra le categorie di notizie (mostrato in Figura 3.6) restituisce, invece, una diversificazione più marcata relativamente al terzo indicatore, ossia il grado di copertura conferito ai vari argomenti. In particolare, per i tre generi di *hard news* (politica, cronaca ed esteri), che sotto il profilo della tutela del pluralismo hanno una rilevanza più spiccata, si riscontra una copertura più elevata della media, in special modo per le notizie di politica. Il dato conferma, anche a livello di singola notizia e di singolo giorno, l'abbondanza di informazione prodotta per queste categorie che già era stata osservata a livello di offerta informativa complessiva nel Capitolo 1 (cfr. Figura 1.7), tra l'altro rilevando il maggior grado di specializzazione dei giornalisti che se ne occupano.

Le notizie di economia, se tendenzialmente non trovano largo spazio nell'offerta informativa e vengono affrontate da giornalisti con un basso livello di preparazione specialistica, mostrano comunque una copertura giornaliera in linea con la media.

Scarsa copertura giornaliera rispetto alla media si rileva invece per le restanti categorie di notizie, sia sport e cultura e spettacolo - che pure, come anticipato nel primo capitolo, incidono sull'offerta complessiva di contenuti informativi ciascuna per una quota prossima al 20% - sia soprattutto scienza e tecnologia. Quest'ultima categoria, per la quale l'offerta totale di contenuti informativi mostra scarsità sul piano quantitativo (carenza di offerta rispetto alla domanda) e qualitativo (bassa specializzazione da parte dei giornalisti), si rileva dunque scarsamente trattata anche in termini di diffusione di contenuti informativi aventi ad oggetto la singola notizia.

Figura 3.6 - Indici di diffusione di una notizia, per categoria



Note: la “durata della notizia” è calcolata come distanza media tra il primo e l’ultimo giorno in cui una notizia registra almeno un’occorrenza.

La “presenza effettiva” della notizia esprime il numero medio di giorni, anche non consecutivi, in cui una notizia registra almeno un’occorrenza.

L’ “indice di copertura” esprime il rapporto (moltiplicato per 100) tra il numero di occorrenze giornalieri registrate mediamente da una notizia appartenente a una determinata categoria e il numero di occorrenze giornalieri registrate mediamente da una notizia appartenente a qualsiasi categoria. Pertanto, valori superiori a 100 indicano una copertura più alta della media, valori pari a 100 indicano un livello di copertura in linea con la media, e valori inferiori a 100 indicano una copertura più bassa della media.

Fonte: elaborazioni Agcom su dati Volocom

3.3. LE MODALITÀ DI DIFFUSIONE DELLA DISINFORMAZIONE E IL CICLO DI VITA DELLE NOTIZIE FALSE: AMBITO SOGGETTIVO E OGGETTIVO

Nel paragrafo precedente sono state esaminate le modalità con cui una notizia reale si diffonde nel tempo all'interno del sistema nazionale dell'informazione. I risultati emersi, letti congiuntamente allo scenario delineato sotto il profilo della produzione (cfr. Capitolo 1), hanno consentito di individuare più compiutamente le peculiarità, gli aspetti soddisfacenti e le criticità sotto il profilo del pluralismo dell'offerta informativa, sia sul piano generale dei contenuti complessivamente generati, sia sul piano particolare della copertura conferita ai singoli accadimenti.

In maniera analoga, l'obiettivo di questo paragrafo è completare il quadro definito nel Capitolo 2 relativamente alla produzione di disinformazione, approfondendo come le notizie false si diffondono nel sistema per mezzo delle varie fonti, che, è bene sottolineare, non tengono conto dell'azione di condivisione di contenuti svolta dagli utenti attraverso pagine e profili personali (la cui trattazione sarà oggetto del prossimo capitolo).

Vale anzitutto osservare come la diffusione di disinformazione nel sistema nazionale debba essere studiata in entrambe le sue componenti, che attengono rispettivamente a:

- l'*ambito soggettivo*, ossia, il comportamento adottato nella trattazione delle notizie dai soggetti individuati quali fonti di disinformazione (cfr. Capitolo 2);
- l'*ambito oggettivo*, ossia, la distribuzione nel tempo delle singole notizie false.

L'*ambito soggettivo* può essere indagato attraverso l'esame dei medesimi indicatori di diffusione introdotti nel par. 3.2 (durata media, presenza effettiva e copertura giornaliera della notizia). In particolare, circoscrivendo l'analisi alle sole fonti di contenuti *fake* (specifici siti web, pagine e *account* di social network individuate come tali da fonti esterne), è possibile valutare come e quanto vengono seguite dalle stesse le notizie reali e quelle false. Al riguardo, occorre evidenziare che i siti che producono disinformazione non contengono soltanto contenuti *fake*, ma combinano notizie reali e notizie false proprio al fine di ingannare e manipolare il lettore della notizia.

L'esercizio è riportato nella Figura 3.7. Per quel che concerne la diffusione delle notizie reali sulle fonti di disinformazione, il primo confronto può essere operato rispetto ai valori assunti dagli stessi indicatori per gli altri media (cfr. Figura 3.4). Se paragonate alle altre fonti informative, si osserva come quelle di contenuti *fake* trattino le notizie più superficialmente (per meno tempo e con minore intensità di copertura) rispetto a

tutti gli altri mezzi (ad eccezione dei social network, che tuttavia hanno finalità informative e struttura diffusiva distintiva). Per le fonti di contenuti *fake*, a differenza delle altre, la produzione dell'offerta non è frutto di un lavoro giornalistico, per cui, anche quando i contenuti veicolati riguardano notizie reali, l'informazione che ne deriva non è caratterizzata né da accuratezza né, tanto meno, da un approfondimento. A ciò si aggiunga che, non di rado, da parte di queste fonti, contenuti *fake* vengono costruiti attorno a fatti reali, magari inseriti in false contestualizzazioni, costituendo fenomeni di mala-informazione.

Spostando il confronto sulle diverse tipologie di notizie diffuse sulle fonti di disinformazione, la Figura 3.7 evidenzia come la durata media esibita dalle notizie false sia ancora minore di quella osservata per le notizie reali (3 giorni per le prime, rispetto ai 9 giorni delle seconde), a fronte però di un numero di giorni di presenza effettiva molto basso per entrambi i tipi di notizie. Questo suggerisce che la distribuzione di contenuti generati attorno alle notizie è molto concentrata e quando l'oggetto dei contenuti divulgati è una *fake news* mediamente la totalità delle occorrenze si verifica in un solo giorno.

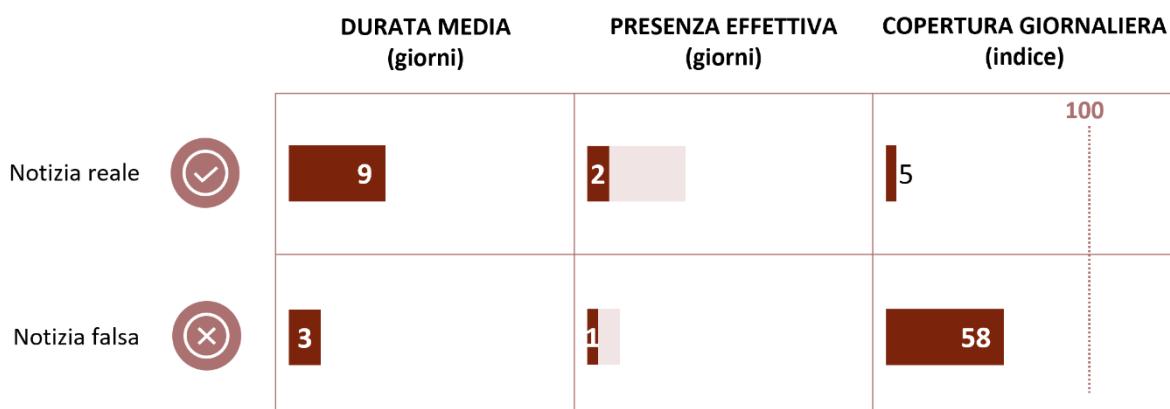
La differenza più rilevante tra notizie reali e false si registra in merito all'indice di copertura media giornaliera conferita dalle fonti disinformazione, che risulta decisamente più alto per i contenuti *fake*, segnalando come l'offerta complessiva proposta e diffusa da tali fonti si componga di poche notizie reali (trattate superficialmente o per veicolare intenzionalmente messaggi di mala-informazione) e di molte notizie false.

Peraltro, la rapidità con cui vengono trattate le singole notizie (reali e false) e un'intensità giornaliera di copertura delle medesime che, in ogni caso, rimane bassa in relazione all'intero sistema informativo, se rapportate all'elevata quantità offerta complessivamente dalle fonti di disinformazione (cfr. Capitolo 2), forniscono una prima evidenza di come il ruolo di queste ultime sia soprattutto quello di dare impulso a molte notizie diverse, lasciando poi che le stesse trovino diffusione massiva attraverso altri canali, fino a viralizzarsi. Si pensi, al riguardo, all'azione combinata dei meccanismi automatici (algoritmi) sottostanti al funzionamento delle piattaforme online e della condivisione dei contenuti messa in atto dagli utenti ma anche, sempre più spesso, da *bot*.

In tal senso, la produzione e la trattazione di disinformazione si differenziano sostanzialmente dal processo informativo, basato sul lavoro giornalistico (cfr. Capitolo 1). In questo caso, si producono tante notizie (cfr. Capitolo 2), che vengono trattate

brevemente e superficialmente. L'obiettivo non è quello di approfondire una tematica bensì di diffondere il più possibile (viralizzare) una falsa informazione. Nella disinformazione online assume perciò grande rilevanza il momento della distribuzione e condivisione attraverso i social di una notizia falsa. Il Capitolo 4 sarà, pertanto, dedicato ad approfondire questi aspetti.

Figura 3.7 - Indici di diffusione delle notizie reali e false su fonti di disinformazione



Note: la "durata della notizia" è calcolata come distanza media tra il primo e l'ultimo giorno in cui una notizia registra almeno un'occorrenza.

La "presenza effettiva della notizia" esprime il numero medio di giorni, anche non consecutivi, in cui una notizia registra almeno un'occorrenza.

L' "indice di copertura" esprime il rapporto (moltiplicato per 100) tra il numero di occorrenze giornaliere registrate mediamente da una notizia sulle fonti di contenuti *fake* e il numero di occorrenze giornaliere registrate mediamente da una notizia su un qualsiasi mezzo. Pertanto, valori superiori a 100 indicano una copertura più alta della media, valori pari a 100 indicano un livello di copertura in linea con la media complessiva dei mezzi di informazione, e valori inferiori a 100 indicano una copertura più bassa della media.

Fonte: elaborazioni Agcom su dati Volocom

Trasferendo l'attenzione dalla componente soggettiva a quella oggettiva, attraverso le figure che seguono, si passa ad analizzare l'altro aspetto della diffusione di disinformazione (in questo caso comprensiva anche di fenomeni di mis- e mala-information³⁰, in quanto alcuni contenuti ritenuti *fake* dai *debunker* nascono, o

³⁰ Cfr. Capitolo 1, Rapporto "[Le strategie di disinformazione online e la filiera di contenuti fake](#)", cit.

transitano, da fonti tradizionali), ossia il ciclo di vita delle singole *fake news* e le modalità con cui si distribuiscono all'interno del sistema dell'informazione complessivo.

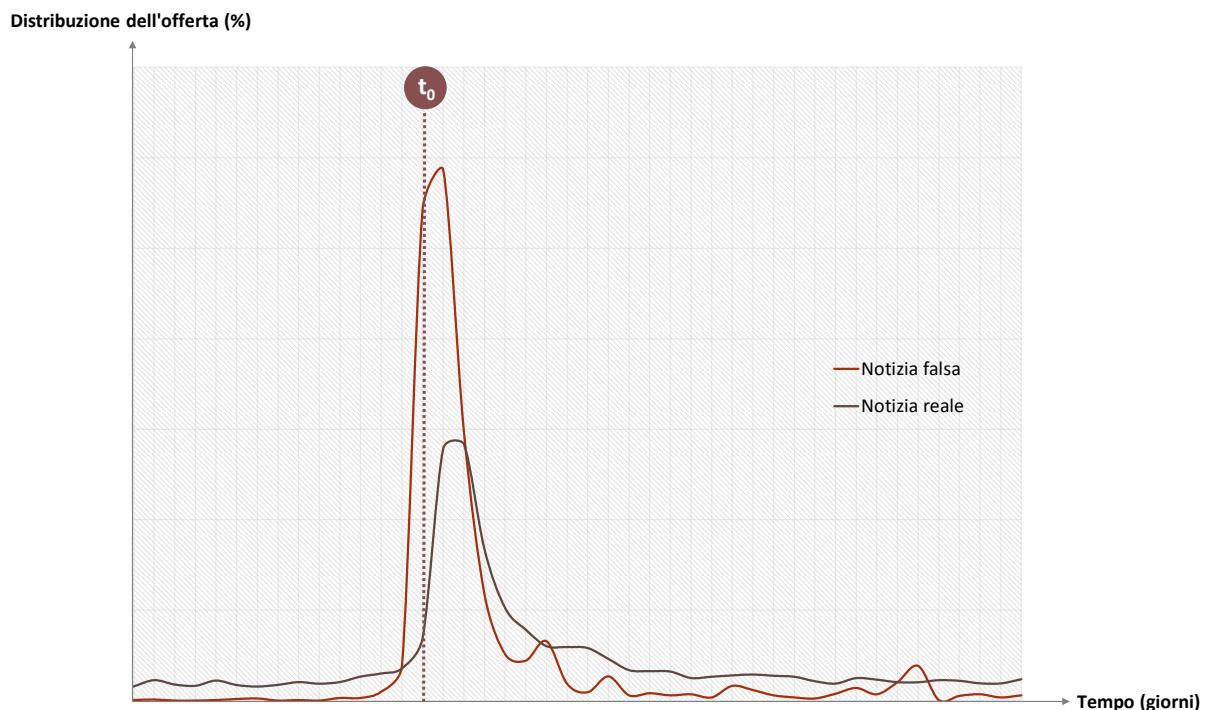
L'*ambito oggettivo*, dunque, può essere esaminato innanzitutto a partire dalla rappresentazione del ciclo di vita che in media presenta una notizia falsa. Al riguardo, la Figura 3.8 riporta la curva (in rosso) che esprime l'andamento medio giornaliero della distribuzione dei contenuti prodotti riguardanti una *fake news*. Le evidenze che emergono più chiaramente, anche in termini comparativi rispetto al ciclo di vita di una notizia reale, sono essenzialmente riconducibili a due constatazioni.

In primo luogo, la pressoché totale assenza di anticipazioni rispetto ai fatti di cui trattano le notizie false. Nel caso delle *fake news*, infatti, il t_0 tende ad essere assoluto e gli accadimenti, essendo falsi, difficilmente vengono preannunciati.

In secondo luogo, il ciclo di vita delle notizie false si rivela decisamente più concentrato attorno al t_0 , raggiungendo il punto di massimo il giorno successivo al t_0 per poi scendere velocemente verso valori prossimi allo zero. In altri termini, la rapidità di trattazione delle notizie osservata con riferimento alle fonti di contenuti *fake* (cfr. Figura 3.7) può essere generalizzata a livello di sistema. Infatti, anche qualora siano gli altri mezzi di informazione a diffondere contenuti aventi ad oggetto una falsa notizia, la trattazione della notizia tende ad essere abbandonata velocemente, non appena verificata nell'ambito dell'attività redazionale presso gli editori.

In definitiva, si può affermare che le due distribuzioni del ciclo di vita della notizia, reale e falsa, sono statisticamente differenti l'una dall'altra, e i contenuti *fake* sono pertanto facilmente riconoscibili perché caratterizzati da una funzione di distribuzione schiacciata ("skewed") e concentrata attorno al t_0 .

Figura 3.8 - Ciclo di vita di una notizia falsa

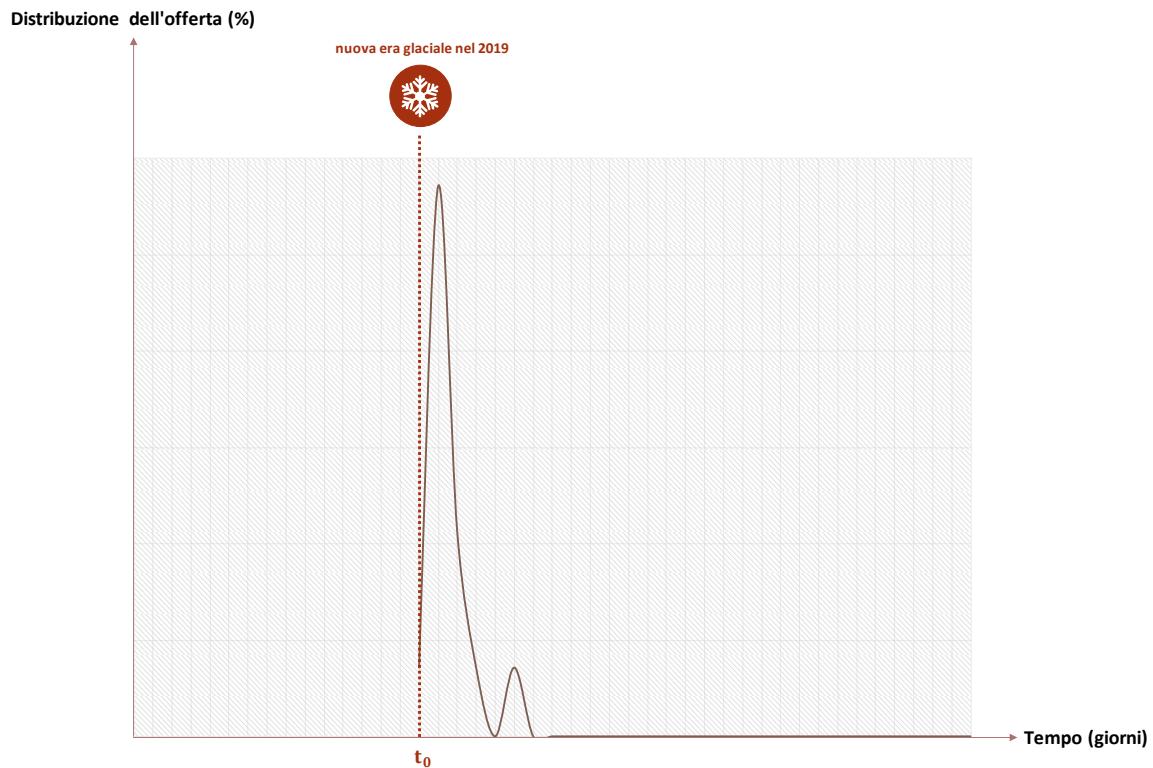


Fonte: elaborazioni Agcom su dati Volocom

Nella Figura 3.9 viene riportato l'esempio di una specifica notizia falsa, per la quale il ciclo di vita inizia in corrispondenza del t_0 , che coincide con il momento stesso in cui viene registrata la prima occorrenza sulla notizia.

Nel dettaglio, la notizia riguarda la pubblicazione di presunti studi scientifici che porterebbero a prevedere la glaciazione terrestre nel 2019. Come mostra la figura, il ciclo di vita della notizia, chiaramente infondata, si esaurisce nel giro di pochissimi giorni, pur avendo trovato spazio, nell'intervallo temporale immediatamente seguente al t_0 , sia tra le fonti di contenuti *fake* sia tra le altre fonti di informazione online e non.

Figura 3.9 - Ciclo di vita della *fake news* “nuova era glaciale nel 2019”



Fonte: elaborazioni Agcom su dati Volocom

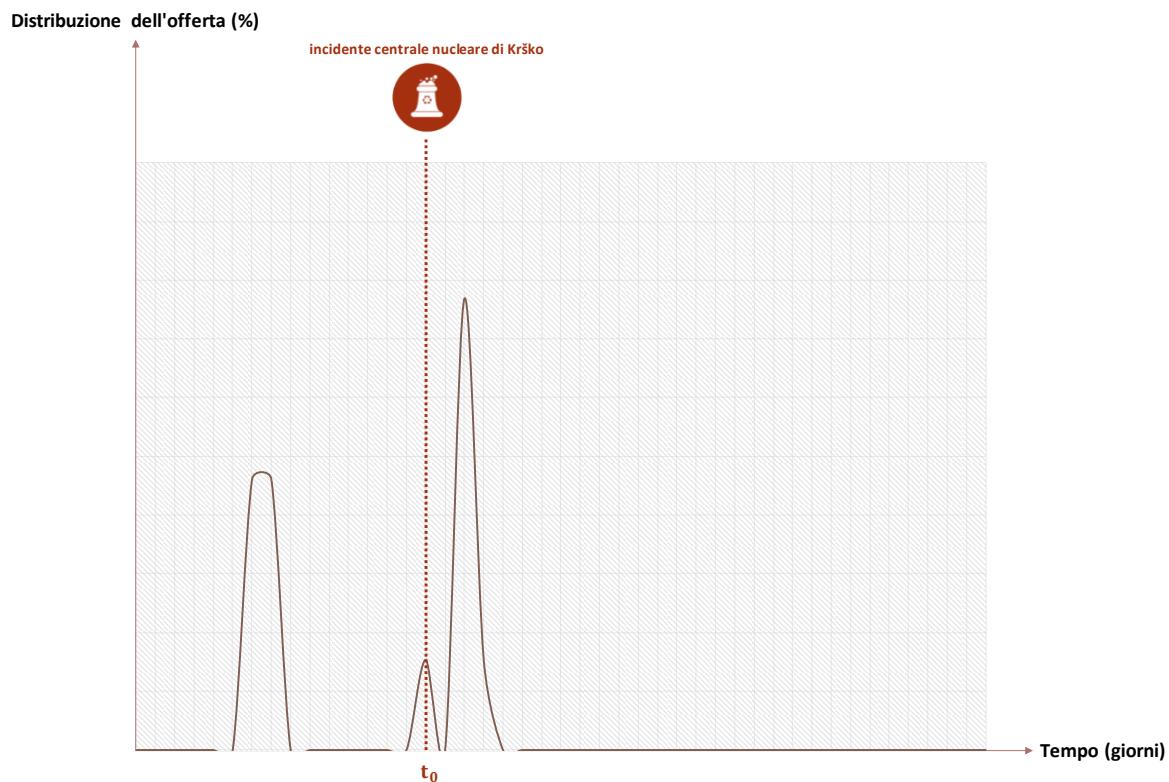
Un ulteriore esempio di diffusione di una notizia falsa nel sistema informativo è esposto nella Figura 3.10. La notizia si riferisce al fatto, nella realtà non verificatosi, di un incidente presso la centrale nucleare di Krško, in Slovenia, non distante dal confine con il territorio italiano.

Questa notizia falsa è stata lanciata online da una fonte di contenuti *fake*, con il chiaro intento di far leva sull'allarmismo che contraddistingue una parte di cittadini con determinate posizioni ideologiche sull'utilizzo dell'energia nucleare, e al contempo innescare un processo di *clickbaiting*.

La notizia, comunque ripresa anche da altre fonti informative, presenta un ciclo di vita peculiare. In effetti, si registrano occorrenze anche prima del t_0 . In questo caso, la ragione della presenza di anticipazioni è rinvenibile nel fatto che, alcuni giorni prima che si diffondesse la notizia falsa, si era verificato un evento reale: era stato arrestato, come la prassi richiede, un impianto della centrale nucleare di Krško per verifiche su un problema rilevato dai sistemi di sicurezza. I primi contenuti apparsi nel sistema

informativo menzionavano l'incidente in via ipotetica, creando lo spunto per l'ideazione dei contenuti *fake* proliferati in seguito.

Figura 3.10 - Ciclo di vita della *fake news* “incidente centrale nucleare di Krško”



Fonte: elaborazioni Agcom su dati Volocom

Quanto osservato mediante la rappresentazione grafica della curva identificativa del ciclo di vita delle notizie false trova conferma nel dato numerico riscontrato per gli indicatori di diffusione delle notizie nel sistema informativo.

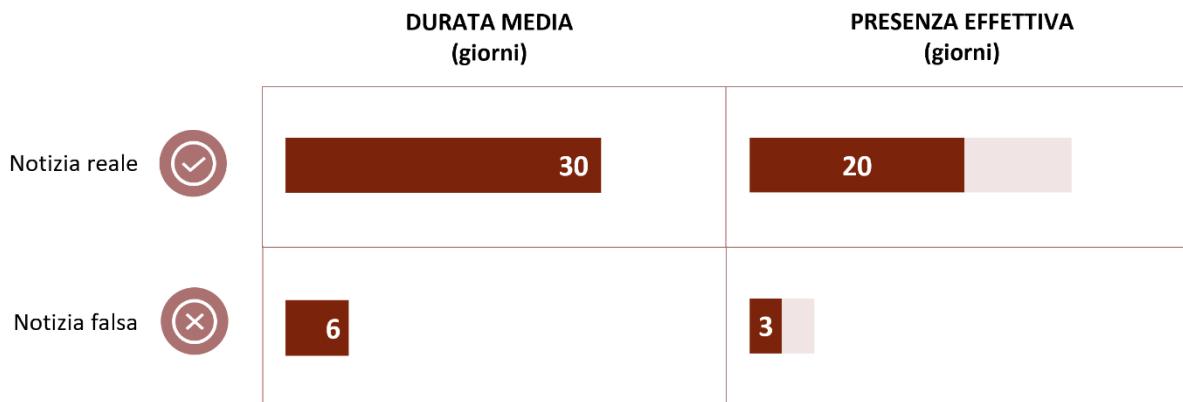
In tal senso, la Figura 3.11 riporta, in un'ottica comparativa, i valori di durata media e presenza effettiva di una notizia falsa rispetto a una reale. La differenza che ne emerge è notevole. Se per una notizia reale, la distanza media che intercorre tra la prima e l'ultima occorrenza è di 30 giorni, per una falsa il lasso temporale è 5 volte inferiore (6 giorni). Conseguentemente, l'intera distribuzione dell'offerta di una notizia falsa si concentra mediamente in 3 giorni, laddove l'effettiva divulgazione di contenuti concernenti una notizia reale si distribuisce nell'arco di 20 giorni.

È evidente che i risultati ottenuti riflettono anche la natura stessa delle strategie di informazione e disinformazione. Lo scopo della diffusione delle notizie reali è quello di informare il pubblico su determinati avvenimenti, seguendo una linea editoriale, l'interesse della domanda e la rilevanza per l'opinione pubblica degli eventi. Diversamente, la diffusione di notizie false – escludendo le inclusioni incidentali all'interno delle fonti informative – risponde a logiche e motivazioni (di ordine ideologico, politico, economico) differenti. Pur sfruttando gli stessi meccanismi cognitivi innescati dall'informazione, è mossa dall'intento, il più delle volte doloso, di confondere, screditare, instillare false convinzioni, che quindi mal si concilia con esigenze di approfondimento e necessità di dedicare più tempo e più spazio alla singola notizia, che rischierebbe di confutare più che confermare la notizia di per sé falsa.

Inoltre, anche quando la notizia falsa transita sulle fonti tradizionali (nella Figura 3.11, a differenza della Figura 3.7, sono, come detto, ricomprese tutte le fonti, non solo quelle di contenuti *fake*, in quanto il focus dell'analisi è l'oggetto – la notizia falsa – e non il soggetto) la durata è breve per l'attivarsi dei meccanismi di verifica delle fonti che regolano il sistema dell'informazione. In alcuni casi, infine, i contenuti *fake* assumono una valenza informativa in quanto tali e vengono trattati e discussi dal sistema informativo nella propria accezione di notizie false.

Ciò non toglie l'esistenza di un discreto numero di contenuti informativi *fake* anche nel sistema informativo tradizionale (mis-informazione), che sono spesso la spia di una regressione nell'uso dei predetti meccanismi di verifica; in tal senso nell'ambito dell'Osservatorio dell'Autorità sul giornalismo si è osservato come “*i giornalisti si sentono gravati particolarmente dalla necessità di operare in un contesto più precario, a retribuzione decrescente, senza tempo per approfondimenti e verifica delle fonti*” (cfr. Agcom, 2017, [Osservatorio sul giornalismo: II edizione](#), par. 194, p. 79).

Figura 3.11 - Indici di diffusione delle notizie reali e false nel sistema dell'informazione



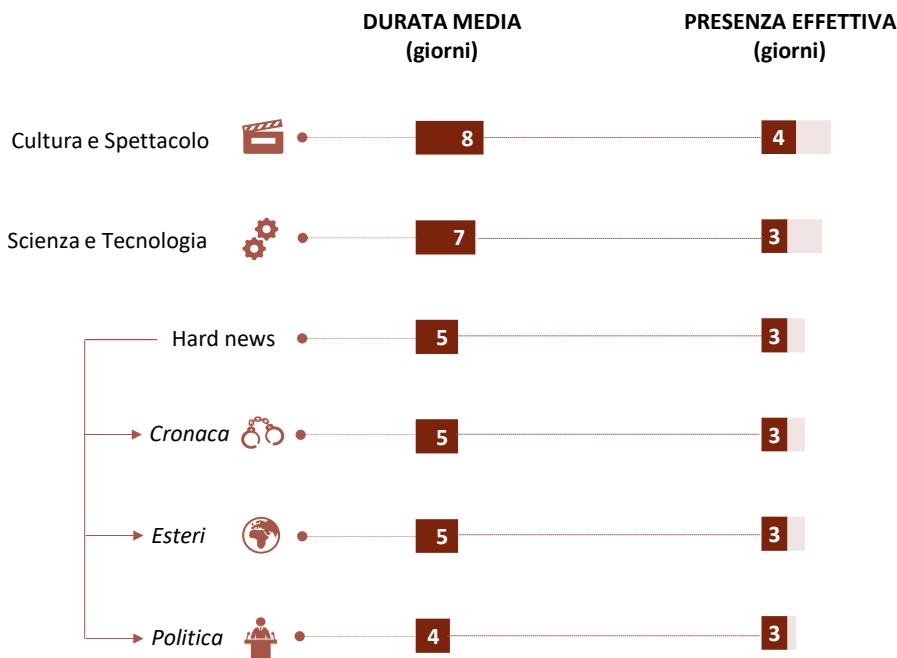
Note: la "durata della notizia" è calcolata come distanza media tra il primo e l'ultimo giorno in cui una notizia registra almeno un'occorrenza.

La "presenza effettiva della notizia" esprime il numero medio di giorni, anche non consecutivi, in cui una notizia registra almeno un'occorrenza.

Fonte: elaborazioni Agcom su dati Volocom

Peraltro, non si registrano rilevanti scostamenti rispetto al valor medio osservato, se si dettaglia l'analisi per categoria tematica di notizia falsa. La Figura 3.12 rileva come la durata media di un contenuto *fake* sia non di molto superiore alla media per cultura e spettacolo e scienza e tecnologia, arrivando a una permanenza rispettivamente di 8 e 7 giorni nel sistema informativo. Leggermente inferiore alla media è la lunghezza del ciclo di vita delle false *hard news*, che non supera i 5 giorni. Tuttavia, le stesse mostrano una presenza effettiva nel sistema informativo analoga alle altre (e, quindi, in percentuale, superiore se rapportata alla rispettiva durata media).

Figura 3.12 - Indici di diffusione di una notizia falsa, per categoria



Note: la "durata della notizia" è calcolata come distanza media tra il primo e l'ultimo giorno in cui una notizia registra almeno un'occorrenza.

La "presenza effettiva della notizia" esprime il numero medio di giorni, anche non consecutivi, in cui una notizia registra almeno un'occorrenza.

Fonte: elaborazioni Agcom su dati Volocom

In definitiva, l'analisi sulla disinformazione svolta in questo paragrafo e nel precedente capitolo ha messo in luce che, per quanto attiene all'aspetto soggettivo, le fonti di disinformazione assumono un ruolo che consiste prevalentemente nel conferire impulso ai contenuti *fake* all'interno del sistema nazionale.

Dal punto di vista oggettivo, la brevità del ciclo di vita delle singole *fake news* e la concentrazione della distribuzione delle relative occorrenze in pochi giorni sono la spia stessa dell'intento di mettere in atto una strategia di disinformazione, prediligendo la trattazione di tante notizie diverse, evitando di approfondirne i contenuti. Molto spesso, le *fake news* immesse nel sistema informativo, oltre che per la falsità del fatto narrato, si contraddistinguono per la loro contagiosità nel trasferire stati emotivi e percezioni, il che le porta, come anticipato nel par. 2.1, a poter agevolmente divenire oggetto di strategie di disinformazione. In relazione a tali strategie, la diffusione massiva, che di fatto conduce alla valorizzazione economica dei contenuti *fake*, avviene mediante

l'attivazione di ulteriori canali distributivi, che chiamano direttamente in causa il consumo e le azioni informative svolte dagli utenti sulle piattaforme sociali online. Lo studio di questo aspetto, introdotto in diversi recenti Rapporti dell'Autorità³¹, sarà affrontato dettagliatamente nel prossimo Capitolo, che dà conto dei risultati di analisi empiriche condotte su grandi masse di dati.

Un'ultima notazione riguarda le relazioni di interdipendenza tra sistema informativo e sistema della disinformazione.

La diffusione di contenuti *fake* appare legata anche a diverse criticità che riguardano, sempre più, il primo e che sono state oggetto di analisi e valutazione da parte dell'Autorità in altre occasioni. In quest'ambito, è opportuno richiamare in maniera sintetica le principali questioni collegate ai fenomeni patologici in esame:

- la *perdurante* riduzione degli investimenti in informazione che “*rischia ... di innescare una spirale negativa, in cui, a fronte di una diminuzione della qualità dell'informazione, si registra un'ulteriore contrazione dei ricavi*” (Agcom, 2015, [“Indagine conoscitiva su Informazione e Internet in Italia”](#), par. 520, p. 217);
- l'evidenziata e connessa regressione nell'uso di meccanismi di verifica nell'ambito della professione giornalistica (cfr. Agcom, 2017, [Osservatorio sul giornalismo: II edizione](#));
- la ristrettezza dei tempi dell'informazione online, sia nella fase di produzione (cfr. Capitolo 1), sia in quella di consumo (cfr. Agcom, 2018, [Rapporto sul consumo di informazione](#)), che rischia di tradursi in una minor attendibilità nella raccolta di informazioni da parte delle redazioni, oltre che in un minor livello di attenzione dell'utente, a scapito del livello di qualità dell'informazione stessa;
- il mancato approfondimento (quantitativo e qualitativo) di talune tematiche, quali, in particolare, quelle di tipo scientifico e tecnologico, che scontano una perdurante carenza di offerta informativa nell'ambito dell'informazione tradizionale;

³¹ Cfr. Agcom, [“Rapporto sul consumo di informazione”](#), cit.; [“Big data Interim report nell'ambito dell'indagine conoscitiva di cui alla delibera n. 217/17/CONS”](#), cit.; [“Le strategie di disinformazione online e la filiera di contenuti fake”](#), cit.

- il conseguente costante declino della reputazione del sistema informativo tradizionale nel suo complesso (fenomeno che, come i precedenti, caratterizza non solo l'ecosistema italiano, ma, più in generale, quello mondiale)³².

In questo quadro, i cittadini rischiano di affidarsi sempre più a fonti informative alternative e non qualificate, che sono spesso alla base di strategie di disinformazione. Il prossimo Capitolo approfondirà pertanto le modalità di consumo informativo online e il ruolo svolto dai cittadini e dalle piattaforme online (e dai relativi algoritmi).

³² L'annuale Rapporto di Edelman sullo stato di fiducia dei cittadini (“[Trust Barometer](#)”) evidenzia come i media siano oramai l’istituzione con più bassa reputazione. Nel 2018, a livello mondiale (tra i 28 maggiori Paesi), solo il 43% della popolazione avrebbe fiducia nei media. In Italia, tale valore sarebbe pari al 45%, in calo di 3 punti percentuali rispetto all’anno precedente.

LA PROPAGAZIONE DI DI NOTIZIE REALI E FALSE SULLE PIATTAFORME ONLINE

4



4.1. INTRODUZIONE

L'avvento di internet e delle tecnologie proprie del web ha cambiato radicalmente il paradigma del consumo di notizie, portando alla formazione di un nuovo scenario in cui le persone partecipano attivamente non solo alla diffusione dei contenuti, ma anche alla loro produzione. Si è passati da un modello in cui le informazioni venivano fornite da un insieme definito di fonti ufficiali e mediate da esperti e giornalisti, all'attuale ambiente disintermediato e ri-intermediato da piattaforme algoritmiche, costituito da una massa eterogenea di fonti di notizie, che si affiancano e mescolano al flusso tradizionale.

In un tale contesto, le piattaforme online svolgono un ruolo cruciale, divenendo sempre più centrali per l'informazione degli utenti e, più in generale, per la vita sociale e il mondo politico e civile. Solo per fare qualche esempio, ogni 60 secondi, su Facebook, vengono creati 3,3 milioni di *post*, pubblicati 510.000 commenti e aggiornati 293.000 stati; su Twitter vengono inviati 350.000 *tweet*; su Whatsapp vengono scambiati 29 milioni di messaggi; su Google vengono effettuate 3,8 milioni di ricerche. In altre parole, gli utenti si rapportano con un flusso continuo e ininterrotto di informazioni, notizie reali e false che coesistono e possono confondersi. Sulle piattaforme online, accanto alle fonti informative ufficiali, si moltiplicano voci infondate e spesso non veritieri, in grado di influenzare in concreto la formazione dell'opinione pubblica. Non a caso, dal 2013 il World Economic Forum sta ponendo il pericolo globale di una massiccia disinformazione digitale al centro di altri rischi tecnologici e geopolitici, che vanno dal terrorismo, agli attacchi informatici, fino al fallimento di *governance globale*³³.

La disinformazione, come si è avuto modo di osservare in precedenza, tende ad annidarsi lì dove il sistema dell'informazione fallisce: difficoltà di monetizzazione dei contenuti e contrazione degli investimenti, scarsa preparazione specialistica delle risorse professionali in determinate materie, esigenze di velocità di aggiornamento dei contenuti informativi (specialmente online) sono atte a compromettere l'adeguatezza dell'offerta informativa sul piano dell'accuratezza, dell'approfondimento e della copertura delle notizie. In termini più ampi, sono alla base della diffusa perdita di reputazione e fiducia accordata dai cittadini al sistema informativo.

Quando le persone sono disinformate, appaiono inclini a difendere le proprie credenze, trascurando le prove concrete. Quando non hanno fiducia nel sistema informativo, il

³³ Cfr. W.L. Howell (2013), *Digital Wildfires in a Hyperconnected World. Tech. Rep. Global Risks*, World Economic Forum.

loro atteggiamento di diffidenza le conduce a mostrare resistenza ai fatti³⁴, e le correzioni veicolate da fonti ufficiali (seppur scientificamente fondate) possono non riuscire a ridurre le percezioni errate, talvolta agendo addirittura come effetto di ritorno³⁵.

È in una situazione di questo tipo che si registra la propensione degli individui ad informarsi affidandosi alla propria rete online di contatti, ad attribuire credibilità ai contenuti e alle fonti che confermano le proprie congetture, a condividere e affermare in prima persona il proprio punto di vista e orientamento ideologico. Ed è, dunque, su questi atteggiamenti che fanno leva i soggetti che mettono in atto le strategie di disinformazione, contando sull'effetto di viralizzazione che gli stessi possono contribuire a diffondere, unitamente alle caratteristiche tecnologiche e agli algoritmi di personalizzazione delle piattaforme online.

Dopo essersi soffermati sugli aspetti legati alla produzione (Capitoli 1 e 2) e alla diffusione (Capitolo 3) di notizie reali e di contenuti *fake* ad opera delle fonti informative e di quelle individuate dai *debunker* come fonti di disinformazione, i prossimi paragrafi saranno dedicati all'analisi: *i)* del consumo di notizie sulle piattaforme online e dei fattori che conducono all'insorgenza di fenomeni di polarizzazione; *ii)* delle modalità attraverso le quali vengono attuate e si diffondono massivamente le strategie di disinformazione sulle piattaforme online, anche sfruttando le azioni informative degli utenti; *iii)* della relazione tra la polarizzazione degli utenti attorno a determinati argomenti e le tematiche oggetto di disinformazione.

Per affrontare correttamente lo studio di tali temi e comprendere il dispiegarsi dei fenomeni sopra esposti, si è ritenuto necessario seguire un'impostazione metodologica basata sull'esame di grandi massi di dati, che al tempo stesso, tramite l'utilizzo di strumenti sperimentali, potesse tener conto dei fattori socio-cognitivi che intervengono nei processi indagati. In tal senso, gran parte delle analisi esposte in questo Capitolo sono il frutto della collaborazione di ricerca di Walter Quattrociocchi, che – nello svolgimento dell'incarico affidatogli dall'Autorità, nell'ambito dell'Indagine conoscitiva

³⁴ Cfr. J.H. Kuklinski et al. (2000), "Misinformation and the Currency of Democratic Citizenship." *The Journal of Politics*, 62 (3), pp. 790-816.

³⁵ Cfr. B. Nyhan, J. Reifler (2010), "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior*, 32 (2), pp. 303-330.

in cui si inserisce il presente Rapporto³⁶ – ha prodotto un primo contributo di sintesi sulle più recenti evidenze scientifiche emerse in materia di disinformazione online.

³⁶ Cfr. gli esiti della procedura comparativa per l'affidamento di un incarico individuale di ricerca su "Informazione e piattaforme digitali", ex determina n. 78/17/SG.

4.2. IL CONSUMO DI NOTIZIE SULLE PIATTAFORME ONLINE E LA POLARIZZAZIONE

A livello globale, le piattaforme online sono definitivamente divenute parte integrante della dieta informativa quotidiana dei cittadini. Al riguardo, nell'ultimo [Rapporto sul consumo di informazione](#), l'Autorità ha rilevato come gli Italiani che utilizzano internet per ricercare notizie accedano all'informazione prevalentemente attraverso le piattaforme online. La consultazione di almeno una piattaforma online per informarsi riguarda infatti il 55% della popolazione, laddove la fruizione delle fonti editoriali online si ferma al 39%. In particolare, social network e motori di ricerca raggiungono le porzioni più ampie di popolazione, pari ciascuna al 37%. Anche quando la finalità informativa è politica/elettorale, il reperimento di notizie e punti di vista su internet passa in maniera prioritaria attraverso le piattaforme online, piuttosto che attraverso i siti *web* o le *app* di quotidiani o altri siti di informazione *online*: 28% a fronte dell'8% della popolazione maggiorenne.

Più in generale, le piattaforme costituiscono sempre più frequentemente veri e propri *gatekeeper* per l'accesso all'informazione. In un contesto caratterizzato dallo "spacchettamento" del prodotto informativo e da una fruizione frammentata dei contenuti (articoli, commenti, video, *post*, ecc.), le piattaforme fungono da nuovi intermediari per l'accesso all'informazione online da parte dell'individuo, accesso che molto spesso è frutto anche dell'incidentalità e casualità della scoperta delle notizie da parte dello stesso cittadino, che peraltro rischia di non avere piena consapevolezza circa la natura e la provenienza dell'informazione³⁷.

Allo scopo di analizzare come concretamente avviene il consumo di notizie sulle piattaforme online da parte degli utenti e quali caratteristiche assume, si propongono di seguito i risultati di ricerche scientifiche condotte su *dataset* di milioni di dati. L'approccio metodologico seguito combina scienza delle reti e meccanica statistica per trovare, da un punto di vista macroscopico, caratteristiche fisiche dell'interazione degli utenti con le informazioni.

Una prima analisi è incentrata sulle modalità di consumo adottate dagli utenti nella fruizione sui social network di notizie prodotte da fonti informative ufficiali. Nello specifico, sono stati esaminati i modelli di consumo di 376 milioni di utenti su Facebook,

³⁷ Ad esempio, "l'utente che accede a Facebook può trovare notizie o, più probabilmente, imbattersi in esse in maniera incidentale, attraverso molteplici fonti: pagine di quotidiani, pagine native Facebook, link a notizie condivise da amici con relativi commenti e discussione", pp. 39-40; [Rapporto sul consumo di informazione](#), cit.

in un arco temporale di 6 anni³⁸. Per ogni pagina Facebook di una fonte di informazione sono stati estratti i dati relativi alle azioni informative svolte dagli individui (*like*, commenti) in relazione alle notizie (*post*) pubblicate dalle fonti, così come esposto nella Tabella 4.1.

Tabella 4.1 - Composizione del *dataset* per l'analisi dei modelli di consumo informativo sui social network

Variabili	N.ro osservazioni	Incidenza sul totale
Pagine Facebook di fonti informative	920	
Post	12.825.291	
Like	3.621.383.495	
Commenti	366.406.014	
Utenti	376.320.713	100%
Utenti che hanno messo "like"	360.303.021	96%
Utenti che hanno commentato	60.115.975	16%

I dati raccolti, riportati nella Tabella 4.1, forniscono una prima indicazione del volume di interazioni che mediamente gli utenti producono attorno a una notizia su una piattaforma online. La quasi totalità degli utenti del social network (96%) si è trovata ad esprimere almeno una reazione positiva per un *post* pubblicato da una fonte informativa, generando, nel periodo considerato, un ammontare complessivo di oltre 3,6 miliardi di *like*. Il 16% degli utenti della piattaforma è intervenuto direttamente nella discussione sulla notizia, commentandola, producendo nell'intero periodo, 366 milioni di commenti.

Osservando più approfonditamente le interazioni degli utenti con le pagine Facebook delle fonti informative, è possibile valutare il grado di eterogeneità del consumo informativo realizzato, ossia verificare con quante fonti informative diverse interagisce generalmente l'utente sulla piattaforma.

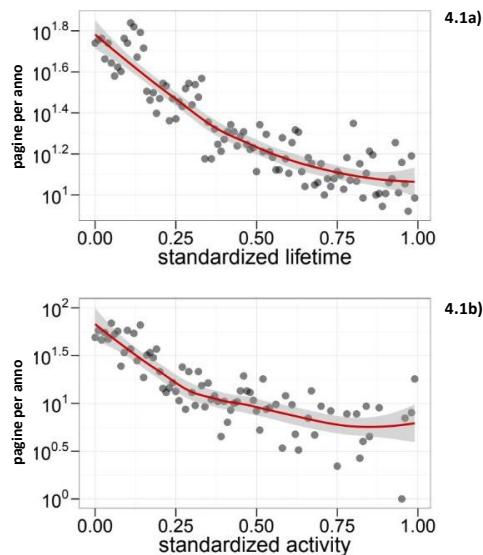
³⁸ Cfr. A.L. Schmidt et al. (2017), "Anatomy of News Consumption on Facebook", *PNAS*, 114 (12).

In tal senso, la Figura 4.1 mostra come varia il numero di pagine di fonti informative con cui un individuo interagisce (in un anno) al variare di due fattori: la durata dell'intervallo temporale (*lifetime*) compreso tra la prima e l'ultima interazione dell'utente (relativamente a un *post* delle fonti informative); e il livello di coinvolgimento dell'utente (in termini di quantità di *like* espressi per un *post*, c.d. *activity*). I valori di entrambi i fattori sono standardizzati (*standardized lifetime* e *standardized activity*), per cui sono compresi tra un minimo di 0 e un massimo di 1.

Nel dettaglio, la Figura 4.1 evidenzia come, sulla piattaforma online, un utente tende ad interagire con un insieme di ristretto di fonti informative. Per di più, all'aumentare della durata del periodo di interazione così come del livello di coinvolgimento, il numero di fonti informative con cui l'utente interagisce si riduce. Infatti, mentre gli utenti con una durata di interazione e un livello di coinvolgimento molto bassi interagiscono con circa 100 pagine in un anno (30 in un mese); gli utenti più attivi e con la maggiore durata di interazione interagiscono soltanto con 10 pagine in un anno (meno di 4 pagine in un mese).

Queste evidenze, ricavate dallo studio del comportamento di centinaia di milioni di utenti in diversi anni rispetto a un cospicuo numero di fonti informative, segnalano distintamente la diffusa tendenza degli utenti a limitare la propria attività informativa a un numero circoscritto di pagine social riconducibili a fonti di informazione, manifestando sostanzialmente l'emergenza di un fenomeno di esposizione selettiva nel consumo di notizie, che peraltro si acuisce tanto più sono attivi gli utenti.

Figura 4.1 - Eterogeneità del consumo informativo, per durata di interazione (*lifetime*; 4.1a) e livello di coinvolgimento degli utenti (*activity*; 4.1b)



Note: la “durata di interazione” (*lifetime*) rappresenta l’intervallo di tempo compreso tra la prima e l’ultima interazione dell’utente relativamente a un *post* delle fonti informative.

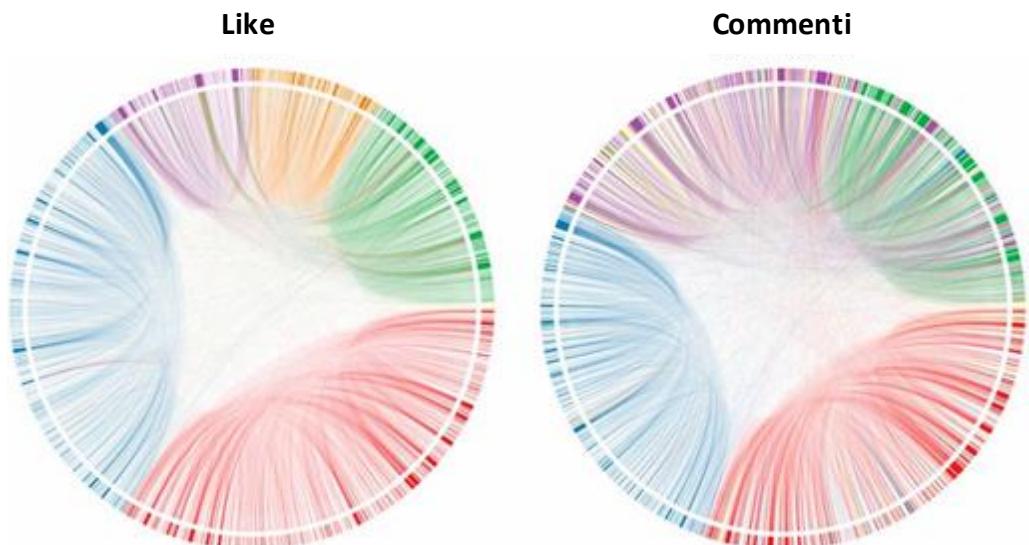
Il “livello di coinvolgimento” (*activity*) rappresenta la quantità di *like* espressi dall’utente per un *post* delle fonti informative.

La propensione ad interagire con poche pagine di fonti informative può determinare l’aggregazione delle stesse in gruppi (c.d. comunità), su cui gli utenti concentrano la propria attenzione e da cui attingono informazione.

Al fine di verificare questa ipotesi, si è proceduto, in primo luogo, a rappresentare le interazioni degli utenti con le fonti informative per definire la struttura delle comunità di pagine che emergono in base alle attività degli utenti. In tal senso, nei grafici che compongono la Figura 4.2, i nodi (lungo il cerchio) rappresentano le pagine, e due pagine sono legate se un utente esprime apprezzamento (o commenta) almeno un *post* di entrambe. Il peso di un legame, ossia la dimensione dell’arco, è determinato dal numero di utenti che le due pagine hanno in comune. I colori identificano l’appartenenza di un nodo (pagina) a una specifica comunità³⁹.

³⁹ Le comunità sono state rilevate dall’algoritmo Fast Greedy (FG). L’algoritmo di rilevazione delle comunità è usato per identificare gruppi di nodi in una rete. La strategia si basa sulla modularità che quantifica la suddivisione di una rete in *cluster* separati, e un’elevata modularità corrisponde a una densa connettività tra i nodi in una comunità e connessioni sparse tra i moduli. Cfr. A. Clauset, M.E.J. Newman, C. Moore (2004), “Finding Community Structure in Very Large Networks”, *Physical review E* (APS), 70 (6).

Figura 4.2 - Struttura delle comunità di pagine, per tipologia di azione informativa



Nota: le comunità sono state rilevate dall'algoritmo Fast Greedy (FG). I colori identificano l'appartenenza di un nodo (pagina) a una specifica comunità.

Esaminando nel dettaglio le attività degli utenti tra le varie comunità e misurando come si estendono tra le pagine, si riscontra in effetti come la maggior parte degli utenti rimanga confinata all'interno di specifiche comunità.

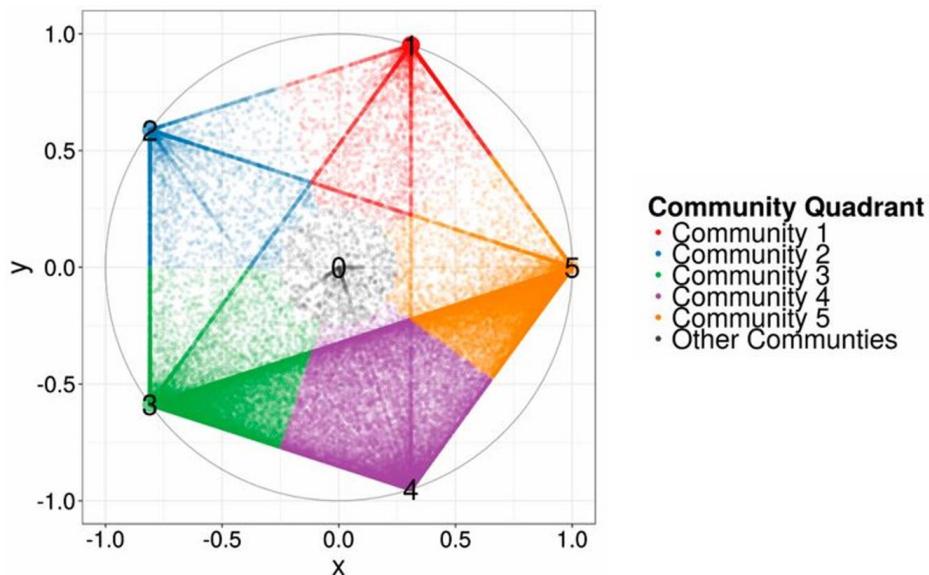
Per comprendere meglio la relazione tra i raggruppamenti di pagine e il comportamento degli utenti, è possibile quantificare la frazione di attività che un utente ha nelle comunità più grandi e quella che ha in qualsiasi altra comunità.

La Figura 4.3 rappresenta l'attività degli utenti nelle comunità più grandi. Più precisamente, i vertici del pentagono rappresentano le 5 comunità più grandi, mentre il punto centrale rappresenta tutto il resto. La posizione di ciascun punto è determinata dal numero di comunità con cui gli utenti interagiscono; la dimensione e la trasparenza indicano il numero di utenti in quella posizione.

Dalla figura si evince come gli utenti siano fortemente polarizzati e come la loro attenzione sia limitata a una singola comunità di pagine. In definitiva, l'interazione degli utenti di Facebook con le fonti di notizie denota una struttura di comunità di pagine alle quali corrispondono gruppi di utenti ben definiti. Poiché gli utenti tendono a

concentrare l'interazione verso un numero limitato di pagine, l'insieme di fonti informative sul social network risulta raggruppato in una precisa struttura comunitaria.

Figura 4.3 - Polarizzazione degli utenti



Nota: i vertici del pentagono rappresentano le 5 comunità più grandi e il punto centrale rappresenta tutto il resto. La posizione di ogni punto è determinata dal numero di comunità con cui l'utente interagisce. La dimensione e la trasparenza indicano il numero di utenti in quella posizione.

Concentrando l'analisi sui modelli di consumo esibiti dagli utenti rispetto a una determinata tematica, è stato poi condotto uno studio specifico con riferimento all'informazione sulla Brexit.

La ricerca ha tenuto conto di tutti i *post* pubblicati su Facebook dalle pagine di fonti informative ufficiali. Per ogni pagina, sono stati raccolti i tutti i *post* dal 1 ° gennaio al 15 luglio 2016 (il referendum sulla Brexit si è tenuto il 23 giugno 2016), e tutte le relative espressioni di apprezzamento (*like*) e i relativi commenti.

Tutte le pagine delle fonti di informazione sono state divise in due gruppi: "pagine Brexit", che includono le pagine impegnate nel dibattito intorno alla Brexit, e tutte le altre. Su 81 pagine (che hanno generato 303 mila post), 38 hanno pubblicato almeno una notizia sulla Brexit (per un totale di 5 mila post). La composizione del *dataset* è riportata nella Tabella 4.2.

Tabella 4.2 - Composizione del *dataset* per l'analisi dei modelli di consumo informativo della tematica “Brexit”

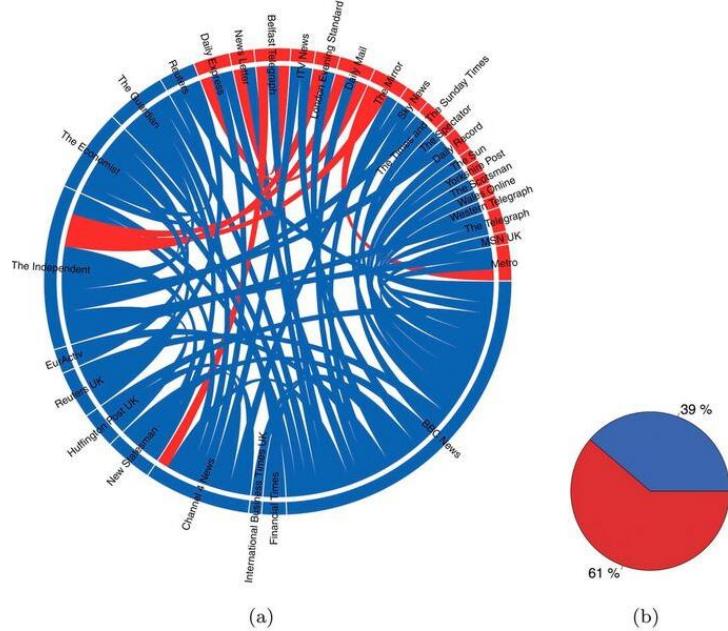
Variabili	Totale	(di cui) Brexit
Pagine Facebook di fonti informative	81	38
Post	303.428	5.039
Like	186.947.027	2.504.956
Commenti	38.182.541	469.397
Utenti che hanno messo “like”	30.932.388	1.365.821
Utenti che hanno commentato	7.222.273	259.078

Come primo passo, sono state individuate le proprietà strutturali delle interazioni rispetto alle “pagine Brexit”, in modo da verificare se anche in questo caso specifico si registra un fenomeno di polarizzazione degli utenti.

La Figura 4.4 mostra il grafico delle interazioni tra pagine e utenti, in cui i nodi sono pagine Brexit e due pagine hanno un arco che le connette se un utente è attivo (compie un’azione informativa) su entrambe. Il peso di un collegamento è determinato dal numero di utenti in comune tra le due pagine.

Senza operare alcuna categorizzazione a priori delle “pagine Brexit”, si osserva chiaramente l’emergere spontaneo di due comunità distinte e separate (C1 e C2), i cui utenti non interagiscono mai tra loro. I colori (rispettivamente, blu e rosso) rappresentano l’appartenenza a una delle due comunità.

Figura 4.4 - Struttura delle comunità di “pagine Brexit”

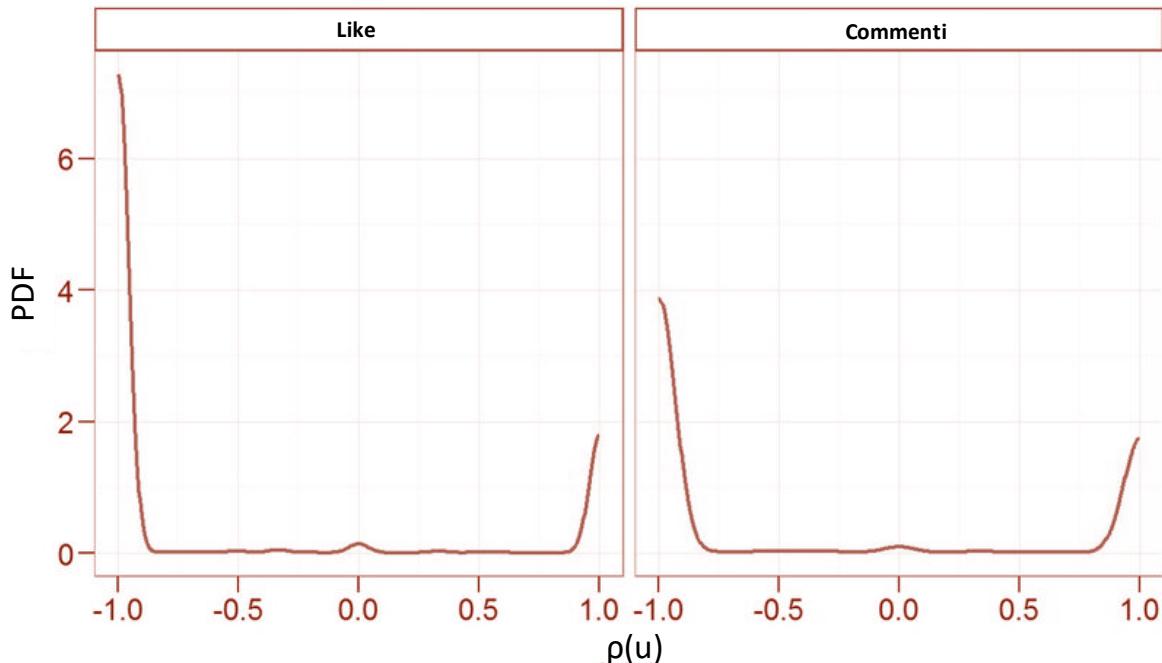


Nota: le comunità sono state rilevate dall'algoritmo Fast Greedy (FG). I colori (blu e rosso) identificano l'appartenenza di un nodo (pagina) a una specifica comunità.

Volendo indagare le modalità attraverso le quali i contenuti delle due comunità vengono consumati dagli utenti di Facebook, si può definire la polarizzazione degli utenti in base al numero di apprezzamenti o commenti rilasciati dagli stessi rispetto ai *post* della comunità C1 e della comunità C2. La distribuzione della polarizzazione definita attraverso il numero di *like* (pannello di sinistra) e di commenti (pannello di destra) è rappresentata nella Figura 4.5.

In entrambi i casi, la distribuzione della polarizzazione si rivela fortemente bimodale (con due picchi visibili), indicando che la maggior parte degli utenti si divide in due gruppi principali che si riferiscono alle due comunità di “pagine Brexit”, C1 e C2. I due gruppi di utenti appaiono divisi e rivolgono la loro attenzione a un insieme limitato di pagine specifiche, tendendo a concentrarsi su una sola narrazione e ignorando l'altra.

Figura 4.5 - Distribuzione della polarizzazione degli utenti verso le comunità di “pagine Brexit”



In conclusione, gli studi svolti su enormi quantità di dati circa il consumo informativo dei cittadini attraverso piattaforme algoritmiche quali i social network, dimostrano che

- nonostante la presenza di un numero elevatissimo di fonti informative (peraltro gratuite), gli utenti tendono ad accedere ad un numero limitato di esse;
- gli utenti più attivi sui social sono anche quelli che accedono a meno fonti informative;
- tale fenomeno, che sembrerebbe paradossale, si spiega con l'emergere sui social di comunità molto polarizzate, distinte e separate l'una dall'altra.

Questi elementi producono un effetto assai rilevante per il sistema informativo nel suo complesso: nonostante la pluralità di fonti informative presenti in rete, i modelli di consumo online non conducono ad un allargamento dello spazio cognitivo dell'individuo, che anzi, rischia di rinchiudersi in stanze dell'eco (c.d. *eco chamber*) molto polarizzate. Come dire che il pluralismo informativo rischia in rete, e sui social in particolare, di non produrre gli attesi effetti positivi connessi all'allargamento dei punti di vista a disposizione del cittadino.

Peraltro, l'Autorità, nell'ambito del [Rapporto sul consumo di informazione](#), ha osservato come la polarizzazione operi già a livello di selezione del mezzo di accesso all'informazione: gli utenti internet sono infatti più polarizzati (a livello ideologico) di quelli che usano la Tv come mezzo informativo.

Inoltre, nell'esaminare la relazione sussistente tra la polarizzazione ideologica degli utenti e le attività informative che essi svolgono in rete, si è rilevato come la polarizzazione possa avere un effetto significativo sul maggior impegno (*engagement*) nei confronti delle notizie divulgate dai social network, con evidenti riflessi sul concretizzarsi di fenomeni di diffusione di posizioni radicalizzate e creazione di bolle ideologiche.

Nel paragrafo seguente, in cui l'attenzione sarà posta sulla disinformazione, verrà approfondita empiricamente la rilevanza del fenomeno della polarizzazione nell'attuazione delle strategie di disinformazione attraverso le piattaforme online.

4.3. L'ATTUAZIONE DELLE STRATEGIE DI DISINFORMAZIONE SULLE PIATTAFORME ONLINE

Il concetto di disinformazione e i tratti distintivi del fenomeno sono stati definiti nel Capitolo 2, dove è stato descritto come l'immissione nel sistema di contenuti falsi e la diffusione di strategie di disinformazione avvenga, con il coinvolgimento di diversi soggetti (ideatori ed esecutori), attraverso quattro fasi successive (per una descrizione dettagliata della filiera, si rimanda al Rapporto dell'Autorità [“Le strategie di disinformazione online e la filiera dei contenuti fake”](#); cfr. anche Figura 2.1). Come anticipato, le fasi principali che costituiscono la filiera dei contenuti *fake* online sono riconducibili alla creazione del messaggio; alla produzione del contenuto in cui lo stesso viene incorporato; alla distribuzione del contenuto; e alla valorizzazione, monetaria o non, di quest'ultimo.

Nella fase di creazione, il messaggio, per risultare efficace, deve essere costruito in maniera tale da raggiungere l'audience e attivarla, coinvolgendola anche nella diffusione ulteriore del contenuto. Nella fase di produzione del contenuto, il codice comunicativo viene affinato ulteriormente, tanto che si possono rinvenire diversi “generi” di contenuti *fake*, ad esempio, in base al diverso grado di “manipolazione” conferito al messaggio (idealmente, si va dal contenuto completamente falso, quindi fabbricato *ex novo*, a quello basato su un'informazione originaria vera ma manipolata). Nella fase di distribuzione, si scelgono uno o più canali attraverso i quali si vuole diffondere il contenuto e si individua il contesto mediatico in cui il contenuto si inserisce. Nell'ultima fase, i contenuti *fake* vengono valorizzati, ossia possono produrre guadagni monetari più o meno immediati attraverso l'adozione di una serie di strategie commerciali, oppure possono raggiungere gli scopi desiderati senza generare necessariamente un flusso di entrate monetarie, poiché rispondono ad altre motivazioni, come quelle di carattere ideologico-politico. In particolare, per ciò che riguarda i ricavi, soprattutto in strategie commerciali di breve-medio periodo, esistono due fonti principali di remunerazione per i produttori: le risorse pubblicitarie e, in alcuni casi, il contributo diretto degli utenti ottenuto con azioni fraudolente. Nell'ambito di strategie di più lungo periodo, ritorni economici possono derivare ad esempio da campagne di disinformazione che, danneggiando l'immagine e la reputazione di un'impresa concorrente, mirano a sottrarre quote di mercato. Infine, si riscontra la presenza di strategie ibride, in cui coesistono finalità politico-ideologiche e finalità di natura economica, che possono produrre un'alterazione degli assetti di un mercato, tale da determinare un rafforzamento della posizione economica di alcuni soggetti d'impresa a scapito di altri, generando così dei vantaggi economici per gli ideatori.

Posta la ricostruzione della filiera della disinformazione appena illustrata, le analisi sulle fasi di produzione e diffusione di contenuti *fake* condotte nei capitoli precedenti hanno mostrato, a fronte di una elevata offerta complessiva delle fonti di disinformazione individuate come tali (siti web e pagine/account social), una bassa copertura nella diffusione delle singole notizie, sia in termini di contenuti prodotti sulla specifica notizia falsa sia in termini di durata del ciclo di vita della notizia. Ne è emerso, quindi, come il ruolo delle fonti di disinformazione (che non contemplano un lavoro giornalistico) sia soprattutto quello di dare impulso a molte notizie diverse, lasciando poi che le stesse trovino diffusione massiva attraverso altri canali, fino a viralizzarsi. A tale riguardo, in più occasioni, l'Autorità ha avuto modo di appurare come la conformazione e i meccanismi di funzionamento delle piattaforme online – presupponendo congiuntamente l'impiego di sistemi di personalizzazione automatica (che operano sulla base di algoritmi e dei *big data* acquisiti) e la possibilità di interazione degli utenti – facilitino la proliferazione di notizie false e la propagazione virale di contenuti *fake* polarizzanti.

Di conseguenza, i modelli di consumo informativo esibiti dagli utenti delle piattaforme online, l'intensità e le modalità delle loro interazioni rispetto alle notizie e alle narrazioni assumono centralità nell'ideazione e nella concreta attuazione delle strategie di disinformazione. Nello specifico, l'esposizione selettiva e la polarizzazione diffusamente mostrate dagli utenti (cfr. par. 4.2), naturali destinatari di tali strategie, possono essere abilmente sfruttate dai soggetti che intervengono in tutte le fasi della filiera della disinformazione per attivare i meccanismi cognitivi ed emotivi degli utenti, tanto nella predisposizione a ricevere il messaggio quanto nella propensione a condividere il contenuto.

Già nella fase di creazione del contenuto *fake*, l'attenzione viene posta dagli ideatori sull'utente, nel momento in cui viene individuato come facente parte dell'audience potenziale della strategia disinformativa. Più specificatamente, coloro che mettono in atto la strategia di disinformazione creano il messaggio, individuando tre aspetti precisi:

- il “profilo degli utenti e il *target* di riferimento”: è proprio la profilazione degli utenti online a consentire di predisporre messaggi e contenuti *fake* più efficaci e, in fase distributiva, di realizzare una diffusione mirata dei medesimi;

- i “temi da trattare”, che solitamente rientrano tra gli argomenti dibattuti, di particolare interesse, e polarizzanti (ossia in grado di creare o accentuare la separazione degli individui in gruppi distinti)⁴⁰;
- il “modo in cui gli individui elaborano le informazioni”, per poter così agire sui *bias cognitivi* degli individui e soddisfare non solo i loro bisogni di informazione, ma soprattutto le loro aspettative in termini di corrispondenza rispetto alle proprie convinzioni (c.d. *confirmation bias*), di coinvolgimento emotivo e di condivisione della visione del mondo.

Successivamente, nella fase di distribuzione, i soggetti che perseguono strategie di disinformazione – singoli o gruppi di individui oppure organizzazioni più o meno strutturate – possono agire con l’ausilio di meccanismi automatici come i *bot*, che consentono la pubblicazione e distribuzione dei contenuti di disinformazione attraverso una molteplicità di *account falsi* o falsi profili *social*. È in questo momento della filiera che possono attivarsi i meccanismi sopra descritti da parte degli utenti destinatari dei contenuti, che, anche inconsapevolmente, rilanciano i contenuti *fake* (magari modificandoli e quindi creandone di nuovi) e ne favoriscono la diffusione, concorrendo di fatto ad amplificare e accelerare il c.d. *snowball effect* artatamente indotto.

In generale, infatti, i meccanismi di interazione sulle piattaforme social consentono a qualunque individuo di farsi parte attiva nella promozione di un contenuto di disinformazione trasferendo altresì i propri stati emotivi, così contribuendo ai processi di viralizzazione.

Le figure che seguono forniscono una chiara indicazione dell’entità assunta nel tempo dalle interazioni degli utenti attorno ai contenuti di disinformazione sulle piattaforme online. I grafici riportati si riferiscono alle interazioni espresse dagli utenti di Facebook in Italia ai *post* generati da fonti di disinformazione individuate come tali da soggetti specializzati in *debunking* (ossia, attraverso un approccio soggettivo, v. par. 3.3).

Preliminarmente, vale osservare che ad ogni azione informativa svolta dagli utenti rispetto a una notizia è associato un significato specifico⁴¹: mentre un *like* rappresenta

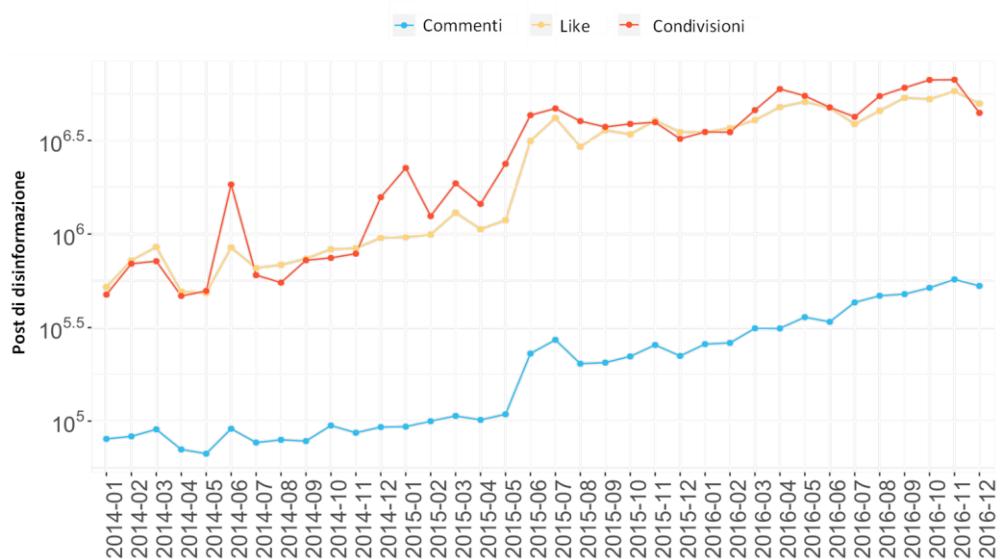
⁴⁰ Sul legame tra tematiche della disinformazione e polarizzazione, cfr. M. Del Vicario, W. Quattrociocchi, A. Scala, F. Zollo (2018), “Polarization and Fake News: Early Warning of Potential Misinformation Targets”, cit.

⁴¹ Cfr. N.B. Ellison, C. Steinfield, C. Lampe (2007), “The Benefits of Facebook “Friends”: Social Capital and College Students’ Use of Online Social Network Sites”, *Journal of Computer-Mediated Communication*, 12 (4).

una reazione positiva, di apprezzamento al *post*, una condivisione esprime il desiderio di aumentare la visibilità di una data informazione; infine, un commento è il modo in cui si innesca o si contribuisce ad accrescere il dibattito che prende forma attorno all'argomento del *post*.

Ciò premesso, la Figura 4.6 mostra una marcata crescita, rispetto al 2014, del volume di interazioni nei confronti delle notizie false, tendenza che si verifica per tutte e tre le tipologie di azioni. Particolarmente elevati e simili tra loro sono i valori riscontrati per *like* e condivisioni. Nello specifico, il dato delle condivisioni è di una certa rilevanza per le strategie di disinformazione, dal momento che le stesse sono atte ad accrescere la visibilità del contenuto *fake* e a favorirne la diffusione effettiva.

Figura 4.6 - Andamento mensile delle interazioni degli utenti con i *post* di disinformazione

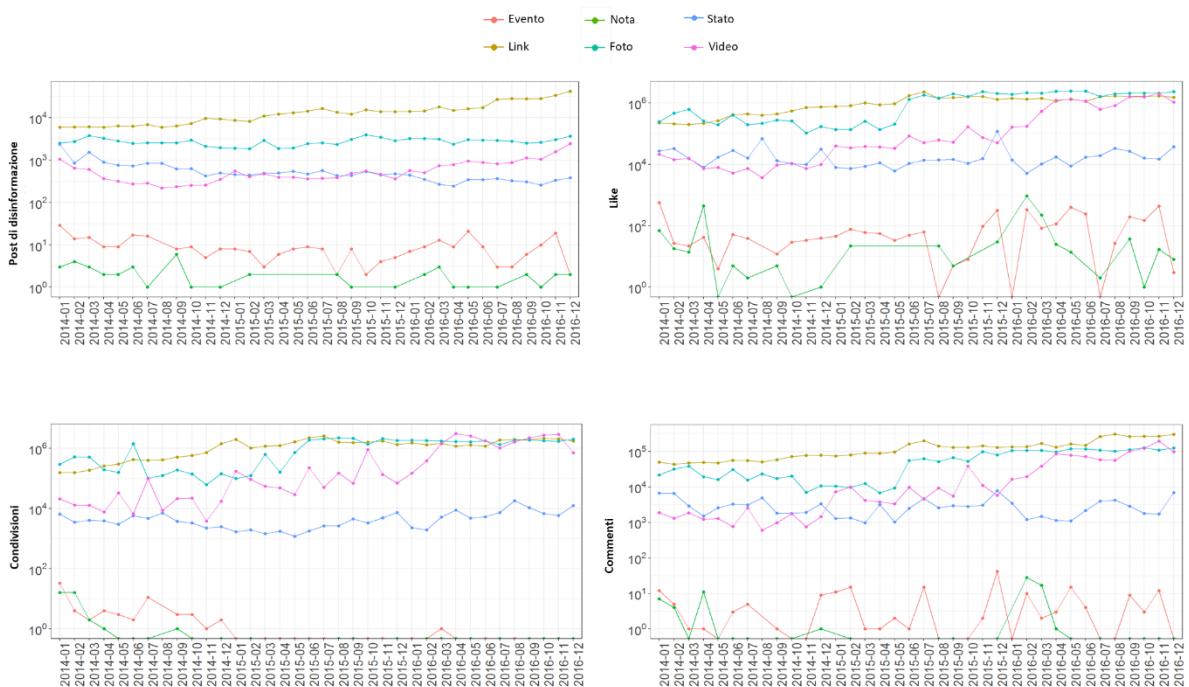


Nota: i *post* di disinformazione sono quelli generati in Italia su Facebook da pagine individuate come fonti di disinformazione da soggetti specializzati in attività di *debunking*.

Specificando l'analisi in base alla tipologia di contenuto veicolato (*link*, evento, nota personale, stato, foto, video) attraverso un *post* dalle pagine delle fonti di disinformazione, la Figura 4.7 riporta sia l'andamento dell'offerta prodotta (primo grafico in alto a sinistra), sia l'andamento delle interazioni generate per tali *post* dagli utenti (negli altri 3 grafici relativi rispettivamente a *like*, condivisioni e commenti).

I *link* che rimandano a una pagina esterna rappresentano la forma più diffusa di contenuto di disinformazione. In aggiunta, i *link*, insieme a foto e video, sono tra i formati che innescano il maggior numero di interazioni e i primi per quantità di commenti ricevuti. Questi risultati contribuiscono ad avvalorare quanto osservato nel Capitolo 3 e sopra richiamato, in merito alla funzione prevalentemente di impulso delle fonti di disinformazione rispetto alle singole notizie false. La scarsa copertura e diffusione conferita da queste fonti alle singole notizie appare, infatti, compensata e fortemente amplificata dalle interazioni degli utenti con i contenuti. È sulle piattaforme online, quindi, che l'impulso lanciato dalle fonti di disinformazione si propaga e raggiunge sempre più utenti che apprezzano, commentano e, soprattutto, condividono i *link* alle notizie (e gli altri formati di contenuti *fake*).

Figura 4.7 - Andamento mensile dei *post* di disinformazione e delle relative interazioni degli utenti, per tipologia di contenuto



Nota: i *post* di disinformazione sono quelli generati in Italia su Facebook da pagine individuate come fonti di disinformazione da soggetti specializzati in attività di *debunking*.

In altre parole, dall'analisi congiunta delle evidenze emerse in questo e nei precedenti capitoli, le principali fasi della filiera del sistema della disinformazione possono essere schematizzate nel seguente modo:

- la profilazione degli utenti e la selezione dei temi da trattare: in Italia, il 57% della produzione di contenuti *fake* riguarda argomenti di politica e cronaca, mentre circa il 20% tematiche di carattere scientifico; tutte tematiche che presentano un forte impatto emotivo, possono essere divisive, e spesso non vengono trattate adeguatamente dal sistema informativo tradizionale;
- in particolare, nel 2018, si è registrato un picco nella produzione di contenuti di disinformazione legato alle elezioni politiche del 4 marzo, in cui temi riguardanti i partiti, i candidati, le politiche proposte, le elezioni stesse, l'economia e la cronaca la hanno fatta da padrone. Accanto a queste, sono emerse tematiche legate alle scienze, alla salute e all'immigrazione;
- la trattazione di queste tematiche da parte dei siti di disinformazione è peculiare e distintiva rispetto a quella che avviene attraverso l'attività di informazione: si crea d'improvviso un contenuto *fake*, la cui trattazione è concentrata in pochi giorni (non più di 3), e i temi sono affrontati in modo superficiale e impressionistico, mirando a stimolare gli stati d'animo delle persone;
- una volta innescata, la notizia falsa viene immessa e rilanciata nel sistema delle piattaforme online anche attraverso l'inconsapevole contributo degli utenti che condividono e commentano la notizia stessa sui social;
- la viralizzazione del contenuto *fake* è resa possibile dalle modalità di consumo dell'informazione sui social (specie per le tematiche individuate come oggetto di disinformazione), che avviene nell'ambito di comunità chiuse, distinte e polarizzate.

Il prossimo e ultimo paragrafo è pertanto dedicato ad approfondire i meccanismi attraverso cui le false notizie vengono consumate e condivise nel mondo delle piattaforme online.

4.4. I MECCANISMI DI INTERAZIONE CON I CONTENUTI DI DISINFORMAZIONE SULLE PIATTAFORME ONLINE

In considerazione della rilevanza che le interazioni degli utenti con le notizie false assumono (per il volume raggiunto e per il livello di coinvolgimento degli individui che sottintendono) nella definizione prima e nel concreto diffondersi della strategia di disinformazione poi, è di primaria importanza comprenderne i meccanismi. Le analisi che seguono sono, dunque, orientate a esaminare, attraverso l'impiego di milioni di dati, le modalità mediante le quali si realizza la diffusione massiva della disinformazione sulle piattaforme online e come la polarizzazione degli utenti è coinvolta in questo processo.

L'obiettivo è caratterizzare il ruolo assunto dalle determinanti cognitive (selezione espositiva, polarizzazione, *confirmation bias*) che governano le azioni informative individuali sulle piattaforme online nella propagazione della disinformazione.

In tal senso, è stato compiuto uno studio empirico⁴² sul comportamento degli utenti rispetto a narrazioni diverse e specifiche, riconducibili a: teorie complottiste (*cospirazione*) e teorie scientifiche (*scienza*), quindi, tematiche di disinformazione da un lato, e di informazione dall'altro. Gli argomenti riguardano temi che, come visto in precedenza, non sono trattati adeguatamente (in termini sia quantitativi sia qualitativi) dal sistema tradizionale dell'informazione, e vengono largamente affrontati nel contesto del sistema della disinformazione.

Nel dettaglio, lo studio si concentra sull'analisi di tre aspetti:

- i. l'interazione degli utenti con le pagine Facebook appartenenti alle due narrazioni, in un arco temporale di cinque anni, nel contesto italiano e in quello statunitense;
- ii. la comparazione dei modelli di consumo e interazione degli utenti rispetto a contenuti appartenenti a entrambe le categorie (cospirazione e scienza) su social media diversi (Facebook e YouTube);

⁴² La descrizione dello studio e la presentazione dei risultati si basano sulla relazione tecnica del Prof. Walter Quattrochiocchi. Cfr. anche A. Bessi et al. (2015), "Science vs Conspiracy: Collective Narratives in the Age of Misinformation", *PLoS ONE*, 10 (2); F. Zollo et al. (2015), "Emotional Dynamics in the Age of Misinformation", *PLoS ONE*, 10 (9); A. Bessi et al. (2016), "Homophily and Polarization in the Age of Misinformation", *The European Physical Journal Special Topics*, 225 (10); M. Del Vicario et al. (2016), "The Spreading of Misinformation Online", *Proceedings of the National Academy of Sciences*, 113 (3); F. Zollo et al. (2017), "Debunking in a World of Tribes", *PLoS ONE*, 12 (7).

- iii. la risposta degli utenti a tre tipi di notizie, ossia quelle coerenti con le proprie convinzioni, quelle satiriche/caricaturali/derisorie rispetto alle teorie complottiste, e quelle discordanti rispetto alle proprie convinzioni (ad esempio, i tentativi di *debunking*).

Da un punto di vista metodologico, per ciascuna categoria di narrazioni, sono state individuate le pagine che diffondono le relative notizie. La categoria “cospirazione” comprende tutte le pagine che veicolano notizie sulle teorie complottiste (che vanno dagli attacchi alieni, alle scie chimiche, al geocentrismo, fino agli effetti paradossali derivanti dalle vaccinazioni). La categoria “scienza” include le pagine di divulgazione scientifica, tra cui quelle di istituzioni, organizzazioni, editoria scientifica. In aggiunta, sono state individuate altre due categorie di pagine: quelle di “satira sulle teorie complottiste” (per la composizione del *dataset* sul caso italiano) e quelle di “debunking sulle teorie complottiste” volte a correggere teorie e notizie false o non veritiero (per il *dataset* sul caso statunitense).

L’individuazione delle pagine da includere nell’analisi è avvenuta con il coinvolgimento di diversi esperti attivi nell’attività di *debunking* soprattutto in materia notizie false su tematiche scientifiche e teorie del complotto. Tutte le pagine sono state quindi verificate manualmente osservando la loro auto-descrizione e il tipo di contenuto promosso.

Le composizioni dei *dataset* per l’Italia e per gli USA sono riportate rispettivamente nelle tabelle seguenti. L’intera procedura di raccolta dei dati è stata eseguita tramite l’API di Facebook Graph (Facebook 2017), utilizzando esclusivamente le informazioni disponibili pubblicamente.

Vale innanzitutto osservare come l’esame dei dati raccolti per la costruzione dei *dataset* consenta di rilevare alcune tendenze generali. In primo luogo, per quel che attiene al contesto italiano, la Tabella 4.3 evidenzia che, in proporzione alle pagine considerate, le fonti di disinformazione scientifica diffondono una quantità di contenuti 3 volte superiore alle fonti di informazione scientifica, laddove negli USA, si verifica la situazione esattamente opposta (cfr. Tabella 4.4).

Tabella 4.3 - Composizione del *dataset* per il caso italiano

	Scienza (informazione)	Cospirazione (disinformazione)	Satira su cospirazione
Pagine	34	39	2
Post	62.705	208.591	4.709
Like	2.505.399	6.659.382	40.341
Commenti	180.918	836.591	58.686
Utenti che hanno messo “like”	332.357	864.047	15.209
Utenti che hanno commentato	53.438	226.534	43.102

In secondo luogo, in Italia, per l'informazione e la disinformazione scientifica si registra un livello di interazione analogo da parte degli utenti. In media, i singoli utenti attivi esprimono 8 *like* e 3 commenti, nel caso dei contenuti informativi, e 8 *like* e 4 commenti nel caso dei contenuti di disinformazione.

Negli USA, oltre alla produzione, anche la reazione in termini di *like* suscitata negli utenti risulta maggiore per i contenuti di “scienza” (mediamente, 11 *like* per utente attivo) rispetto a quelli di “cospirazione” (7 *like* per utente attivo), a fronte di un numero medio di commenti generato da un utente attivo del tutto analogo (3 commenti per utente attivo).

Tabella 4.4 - Composizione del *dataset* per il caso statunitense

	Scienza (informazione)	Cospirazione (disinformazione)	Debunking su cospirazione
Pagine	83	330	66
Post	262.815	369.420	47.780
Like	453.966.494	145.388.117	3.986.922
Commenti	22.093.692	8.304.644	429.204
Utenti che hanno messo “like”	39.854.663	19.386.131	702.122
Utenti che hanno commentato	7.223.473	3.166.726	118.996

In definitiva, si evidenzia una situazione opposta tra Italia e USA (cfr. Tabella 4.5), per tutti gli indicatori di coinvolgimento attivo della comunità dei complottisti rispetto a quella scientifica, laddove negli Stati Uniti la comunità scientifica è più attiva, mentre in Italia la situazione è più complessa.

Come emerge chiaramente dall'indicatore relativo ai *post* (che è pari a 2,9, quindi largamente superiore a 1; a fronte dello 0,4 degli USA), nel nostro Paese vi è una produzione di contenuti scientifici, a parità di fonti analizzate, molto maggiore di disinformazione rispetto all'informazione. Ciò avvalora le evidenze dei capitoli precedenti.

Tuttavia, se si considerano gli altri indicatori, mentre, come detto, negli USA la comunità scientifica è decisamente più attiva (in termini sia di *like* sia di commenti ai post prodotti), in Italia il grado di coinvolgimento è più omogeneo per le due comunità, con i cospirazionisti tendenzialmente più pronti a commentare (valori superiori a 1, per commenti e utenti che commentano) e la comunità che segue i siti scientifici più portata a interagire attraverso *like* (valori inferiori a 1, per *like* e utenti che mettono *like*).

Tabella 4.5 - Coinvolgimento della comunità dei cospirazionisti rispetto a quella scientifica: Italia vs. USA

Coinvolgimento attraverso:	Italia	USA
Post	2,9	0,4
Like	0,8	0,2
Commenti	1,4	0,3
Utenti che hanno messo "like"	0,8	0,3
Utenti che hanno commentato	1,3	0,3

Nota: il coinvolgimento è calcolato come rapporto tra il livello di attività (*post, like, ...*) per pagina (dei *post*) e per *post* (di *like, commenti e utenti*) della comunità dei cospirazionisti rispetto a quella scientifica. Valori superiori a 1 (inferiori a 1) segnalano un maggior (minor) coinvolgimento dei cospirazionisti.

Due ulteriori *dataset* sono stati costruiti (a partire dall'insieme di pagine di cui alla Tabella 4.4, ossia per il campione statunitense) con riferimento a due piattaforme online differenti, Facebook e YouTube. Nello specifico, sono stati raccolti tutti i *post* pubblicati su Facebook dalle fonti scientifiche di informazione e disinformazione contenenti un *link* a un video su YouTube, nonché i *like*, i commenti e le condivisioni di tali *post*. Tramite l'API di YouTube Data, per ogni video di YouTube collegato ai predetti *post* di Facebook, sono stati estratti i dati relativi alle interazioni degli utenti (*like, commenti e visualizzazioni*). La composizione dei due *dataset* è riportata nelle tabelle che seguono.

La Tabella 4.6 mostra come, nel caso di Facebook, le reazioni ai *post* che rimandano a video su YouTube si ripartiscono in maniera simile per informazione e disinformazione (anche se si conferma per gli USA il maggior coinvolgimento della comunità che segue siti scientifici; con il relativo valore, se calcolato, sempre ben al di sotto dell'unità), con i *like* che rappresentano le azioni più frequenti, seguite dalle condivisioni (azione che più delle altre è atta a contribuire alla viralizzazione dei contenuti).

Tabella 4.6 - Composizione del *dataset* relativo ai *post* di Facebook che rimandano a video di YouTube (USA)

	Scienza (informazione)	Cospirazione (disinformazione)
Post	4.388	16.689
Like	925.000	1.000.000
Commenti	86.000	127.000
Condivisioni	312.000	493.000

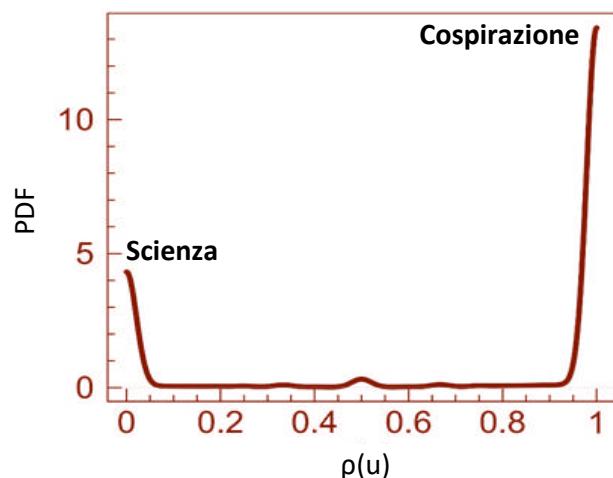
Passando a qualificare più approfonditamente le interazioni compiute dagli utenti rispetto alle narrazioni individuate, è possibile indagare se il livello di coinvolgimento (attività informativa) degli utenti per un tipo specifico di contenuto è indice del grado di polarizzazione degli utenti che condividono lo stesso sistema di credenze, (ossia dell'esistenza di *echo chamber*).

Si suppone che un utente sia polarizzato nei confronti delle narrazioni di “scienza” (informazione) o “cospirazione” (disinformazione) se ha lasciato, rispettivamente, più del 95% dei suoi *like* (o commenti) a *post* scientifici ovvero a *post* cospirazionisti.

Rappresentando la funzione di densità delle probabilità (PDF) della polarizzazione degli utenti (in base alle loro interazioni) su Facebook in Italia, si osserva una distribuzione fortemente bimodale, che presenta due picchi principali in corrispondenza di due valori (cfr. Figura 4.1). Tale risultato indica chiaramente come la maggior parte degli utenti sia polarizzata nei confronti di una o dell'altra narrazione, suscitando la formazione di due gruppi di utenti ben segregati che non interagiscono tra loro.

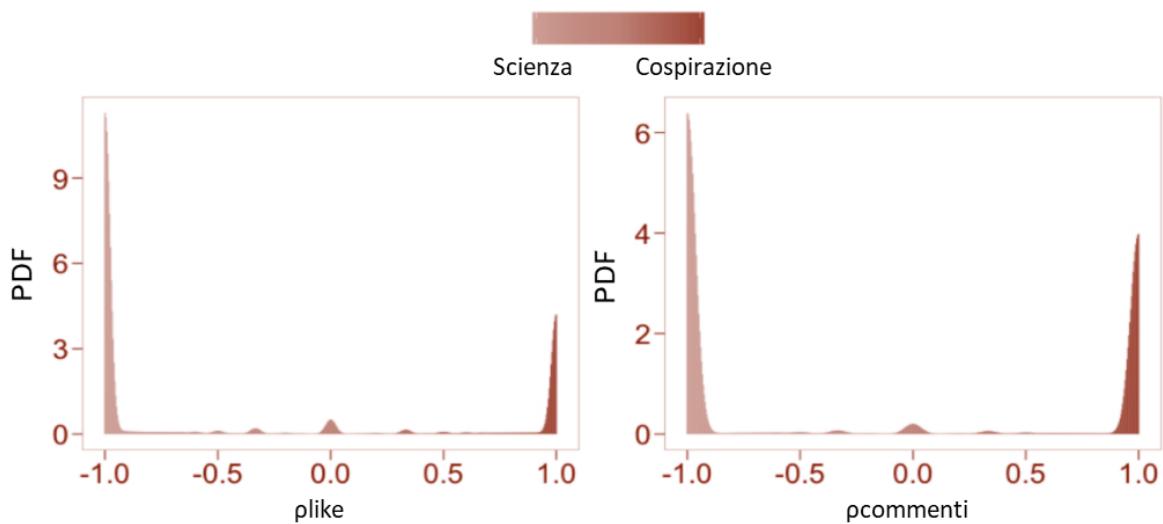
Peraltro, lo studio evidenzia che più un utente è attivo nella sua narrazione, più è circondato da amici sul social network che condividono lo stesso atteggiamento. Ne deriva un elevato grado di omofilia che caratterizza le interazioni sociali degli utenti di Facebook per cui gli individui tendono ad associarsi e legarsi a chi mostra preferenze e credenze simili. Infatti, si riscontra che per gli utenti polarizzati la frazione di amici con la stessa polarizzazione è molto alta e cresce con l'intensificarsi dell'attività informativa svolta nella propria comunità.

Figura 4.8 - Distribuzione della polarizzazione degli utenti di Facebook per “scienza” e “cospirazione” (Italia)



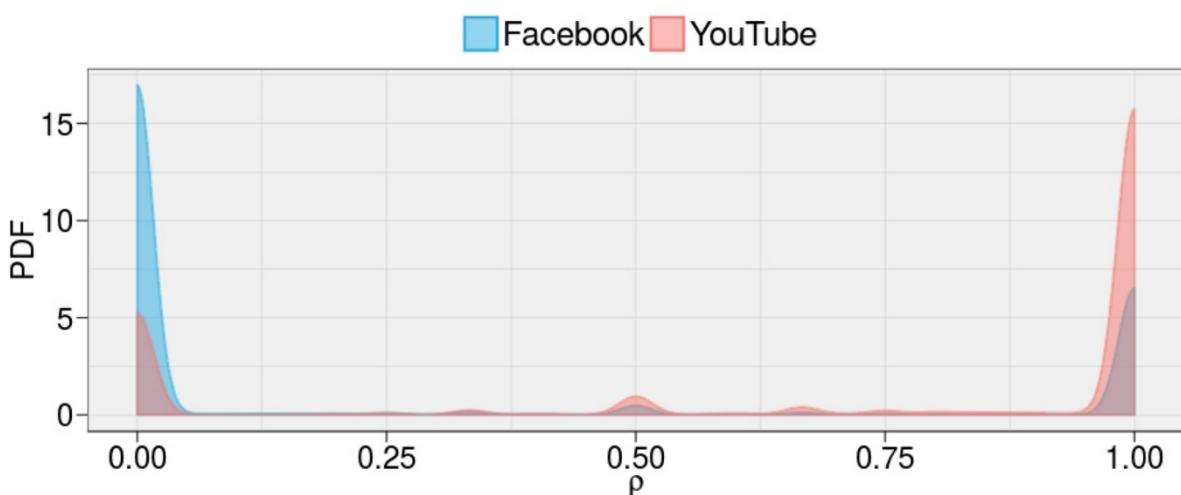
Modelli simili di consumo e interazione degli utenti rispetto ai contenuti di informazione e disinformazione emergono anche sulla piattaforma social negli Stati Uniti, con una distribuzione della polarizzazione fortemente bimodale in corrispondenza delle due diverse narrazioni, sia considerando i *like* che i commenti (cfr. Figura 4.9).

Figura 4.9 - Distribuzione della polarizzazione degli utenti di Facebook per “scienza” e “cospirazione” (USA)



Analogamente, la Figura 4.10 mostra che sia su Facebook che su YouTube la grande maggioranza degli utenti è polarizzata verso una delle due narrazioni. Ciò evidenzia come il fenomeno della polarizzazione valga in maniera piuttosto diffusa per i vari tipi di piattaforma online, siano essi social network (Facebook) o più in generale social media (YouTube).

Figura 4.10 - Distribuzione della polarizzazione degli utenti di Facebook e YouTube per “scienza” e “cospirazione” (USA)



In definitiva, le analisi precedenti hanno messo in luce l'emergenza di modelli di consumo informativo e interazione degli utenti con le notizie reali e false che presentano le medesime caratteristiche (tendenza alla polarizzazione, omofilia, insorgenza di *echo chamber*) indipendentemente dal tipo di narrazione (informazione e disinformazione), e della piattaforma online (Facebook e YouTube). I contenuti relativi a narrazioni diverse aggregano gli utenti in comunità distinte e polarizzate, in cui interagiscono con persone che condividono il loro stesso sistema di credenze.

In Italia, in particolare, la produzione di natura ascientifica e complotista appare assai florida, mentre risulta carente la trattazione approfondita di natura informativa delle medesime tematiche. La mancanza di un dibattito allargato (dovuto a fenomeni di polarizzazione) e di un approfondimento informativo creano i presupposti per un'opinione pubblica meno recettiva (ad esempio rispetto agli Stati Uniti) e più propensa ad apprezzare e condividere notizie anche palesemente false.

A questo punto, diviene cruciale esaminare come i *bias* cognitivi, in particolare il c.d. *confirmation bias*, agiscono nei processi virali di diffusione delle notizie. A tal fine, è necessario individuare, da un punto di vista statistico, le caratteristiche delle cascate di interazioni degli utenti a seconda della tipologia di narrazione (“scienza” e “cospirazione”), ossia informazione e disinformazione.

La prima caratteristica di una cascata è l’ampiezza dell’intervallo temporale (in ore) trascorso tra la prima e l’ultima condivisione di un *post*. I risultati mostrano che, per entrambe le tipologie, si registrano due picchi: uno in corrispondenza di una durata di circa 1-2 ore e uno in corrispondenza di una durata di circa 20 ore, il che denota che i modelli di condivisione temporale sono simili per informazione e disinformazione.

Tuttavia, se si considera la durata dell’interazione in funzione della dimensione della cascata (cioè, il numero di utenti che condividono il *post*), si rileva una differenziazione tra le tipologie di notizie. Infatti, per le notizie scientifiche, la variabilità della durata di interazione cresce con le dimensioni della cascata, e dimensioni maggiori della cascata corrispondono a una maggiore variabilità della durata, ossia a durate diverse a seconda della notizia scientifica. Per i contenuti relativi alle teorie complottiste, invece, la durata delle interazioni su un *post* mostra una crescita monotona rispetto alla dimensione della cascata. Tali risultati suggeriscono che l’informazione scientifica, di solito, viene assimilata (cioè raggiunge un più alto livello di diffusione) rapidamente e poi si evolve in maniera variabile. Al contrario, le notizie di cospirazione sono assimilate più lentamente e mostrano una correlazione positiva tra durata e dimensione per tutti i contenuti; quindi, i *post* che rimangono nel tempo (quelli longevi) sono discussi da comunità più grandi. In altre parole, mentre per le notizie scientifiche c’è un maggior grado di assimilazione e di passaggio a nuove notizie, per le teorie complottiste alcuni contenuti tendono a rimanere nel tempo e ad essere condivisi da comunità sempre più grandi.

Inoltre, è possibile osservare che la maggior parte dei legami tra utenti che condividono consecutivamente un *post* è omogenea, cioè entrambi gli utenti condividono la stessa polarizzazione e, quindi, appartengono alla stessa *echo chamber*. In particolare, le evidenze emerse suggeriscono che la diffusione delle informazioni avviene principalmente all’interno di gruppi omogenei in cui tutti gli utenti condividono la stessa polarizzazione. Quindi, i contenuti tendono a circolare ed essere confinati all’interno delle rispettive *echo chamber*.

In sintesi, sebbene i modelli di consumo rispetto alle narrazioni di informazione e disinformazione siano simili, le dinamiche a cascata delle interazioni differiscono.

L'esposizione selettiva si configura quale motore principale della diffusione dei contenuti e genera la formazione di gruppi polarizzati, ciascuno con la propria cascata di interazioni.

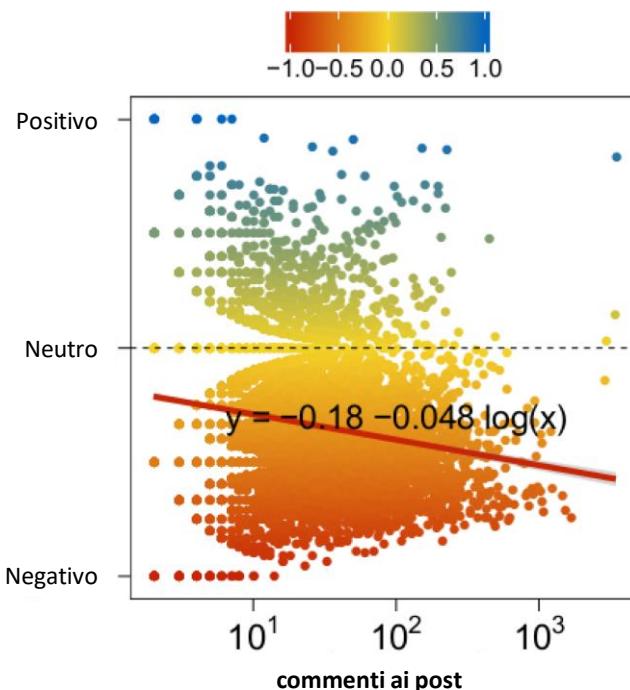
Estendendo ulteriormente lo studio, ci si può soffermare sulle dinamiche emotive che si sviluppano tra *echo chamber* differenti. Al riguardo, sono stati selezionati tutti i *post* che rappresentano il campo in cui si svolge il dibattito tra gli utenti di "scienza" e quelli di "cospirazione". In particolare, sono stati raccolti tutti i *post* commentati almeno una volta sia da un utente polarizzato nei confronti di "scienza" che da un utente polarizzato nei confronti di "cospirazione". Precisamente, emergono 7.751 *post* di questo tipo (su 315.567), avvalorando il fatto che le due comunità sono strettamente separate e spesso non interagiscono tra loro.

Procedendo ad esaminare come cambia lo stato emotivo quando aumenta il numero di commenti ai *post*, ossia quando la discussione diventa più lunga, la Figura 4.11 mostra lo stato emotivo (a livello aggregato) di tali *post* in funzione del loro numero di commenti. Chiaramente, con il crescere del numero di commenti, e il protrarsi della discussione, lo stato emotivo diventa sempre più negativo. La durata della discussione, dunque, influisce sulla negatività del sentimento degli utenti coinvolti nel dibattito.

Inoltre, l'esposizione degli utenti alle affermazioni non motivate (come quelle complottiste) influenza i loro criteri di selezione dei contenuti e aumenta la loro attitudine ad interagire con informazioni false. Il *confirmation bias* svolge, quindi, un ruolo cruciale nella selezione dei contenuti. Peraltro, i risultati appaiono coerenti con la letteratura che indica l'esistenza di una relazione tra credenze nelle teorie cospirative e la necessità di una chiusura cognitiva, per arrivare a conclusioni definite (anche irrazionali)⁴³. Più in generale, le persone che preferiscono un approccio euristico per valutare le evidenze e formare le proprie opinioni mostrano maggiori probabilità di trovare una spiegazione coerente con il proprio preesistente sistema di credenze.

⁴³ Cfr. tra gli altri, P.J. Leman, M. Cinnirella (2013), "Beliefs in Conspiracy Theories and the Need for Cognitive Closure", *Frontiers in psychology*, 4.

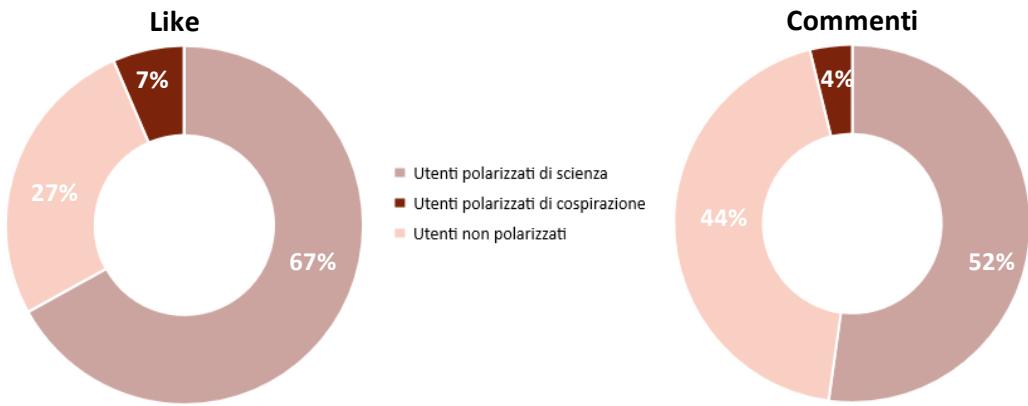
Figura 4.11 - Stati emotivi in funzione del numero di commenti ai *post* discussi da entrambe le narrazioni



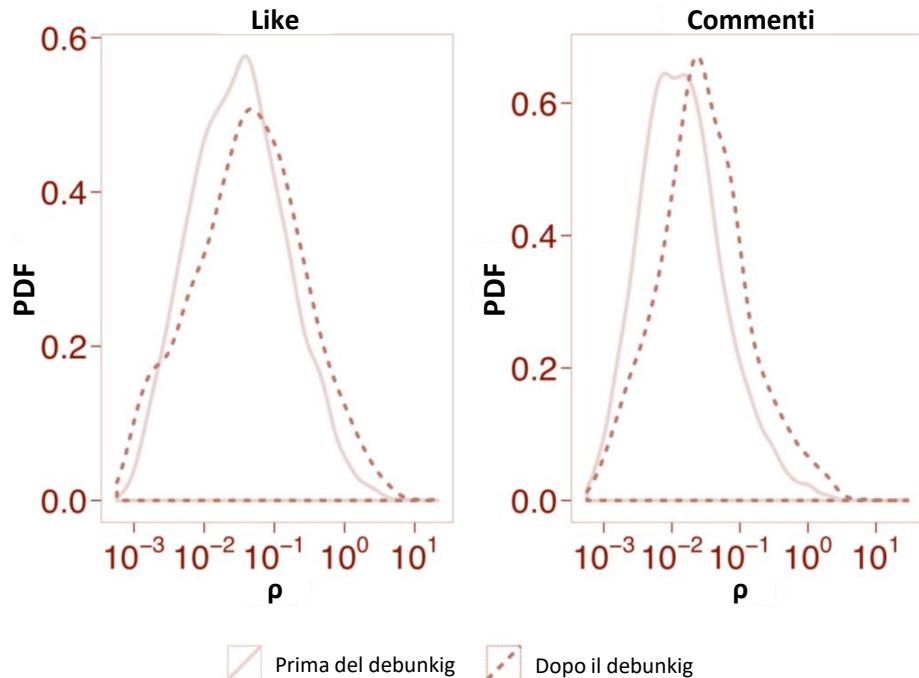
Infine, i modelli di interazione informativa degli utenti possono essere studiati con riferimento alle notizie discordanti rispetto alle proprie convinzioni.

In proposito, le pagine di *debunking* su Facebook tentano di contrastare la diffusione della disinformazione, fornendo informazioni verificate su fatti e argomenti specifici. Tuttavia, se il *confirmation bias* gioca un ruolo chiave nei criteri di selezione, è probabile che i contenuti di *debunking* si configurino per gli utenti solitamente esposti alla disinformazione (“cospirazione”) come notizie che dissentono dalla loro narrazione preferita.

In primo luogo, la Figura 4.12 mostra come le notizie di *debunking* vengono apprezzate e commentate in base alla polarizzazione degli utenti. Si osserva che la maggior parte di entrambe le azioni (*like* e commenti) è svolta dagli utenti polarizzati nella categoria “scienza” (che registra il 67% dei *like* complessivi alle notizie di *debunking* e il 52% di tutti i commenti), mentre solo una minoranza delle interazioni è composta da utenti polarizzati nella narrazione “cospirazione” (7% dei *like* e 4% dei commenti).

Figura 4.12 - Interazioni degli utenti con notizie di *debunking*

In secondo luogo, è possibile condurre un'ultima analisi comparativa tra il comportamento degli utenti appartenenti alla comunità di “cospirazione” prima e dopo la loro prima interazione con una notizia di *debunking*. La Figura 4.13 mostra il gradimento e il tasso di commento, ossia il numero medio di *like* e commenti su una notizia complottista nel giorno, prima e dopo la prima interazione con *post* di *debunking*. Ci si potrebbe aspettare che gli utenti riconoscano la correzione e riducano il loro impegno all'interno del gruppo cospirazionista. Al contrario, si evidenzia invece che i loro *like* e commenti sui *post* di “cospirazione” non cambiano sostanzialmente (*like*) o addirittura aumentano (commenti) dopo aver interagito con una notizia correttiva.

Figura 4.13 – Interazioni dei cospirazionisti prima e dopo il *debunking*

In conclusione, gli utenti tendono a selezionare le informazioni che sono coerenti con il loro sistema di credenze, formando gruppi polarizzati di persone con idee simili su narrazioni condivise (c.d. comunità), in cui le informazioni discordanti vengono ignorate, rendendo spesso inutili i tentativi di *debunking*.

Tale ultimo risultato suggerisce che il processo che guida i meccanismi di acquisizione e elaborazione delle informazioni non sia dettato soltanto da fenomeni cognitivi ma anche, e soprattutto, da pratiche culturali e sociali. In altre parole, “*facts remain robust only when they are supported by a common culture, by institutions that can be trusted, by a more or less decent public life, by more or less reliable media*”⁴⁴.

⁴⁴ È quanto sostiene il filosofo francese della post-verità Bruno Latour, v. A. Kofman, “Bruno Latour, the Post-Truth Philosopher, Mounts a Defense of Science”, *New York Times Magazine*, 25 ottobre 2018.

APPENDICE METODOLOGICA

Come evidenziato in Premessa, il Rapporto ha adottato una metodologia volta all'analisi integrata di dati provenienti da fonti e *dataset* differenti. In particolare, si è fatto ricorso a informazioni, qualitative e quantitative, sia sulla domanda che sull'offerta di informazione.

Sul primo aspetto, si è fatto riferimento prevalentemente a dati provenienti da una specifica survey su oltre 14.000 consumatori italiani (condotta per l'Autorità da GfK Italia), utilizzati, tra l'altro, nell'ambito del recente [Rapporto sul consumo di informazione](#) in Italia. Si rimanda alla relativa [Appendice metodologica](#) per una illustrazione della metodologia sottostante (numerosità campionaria, metodi di stratificazione e di riponderazione, ecc.).

Sempre con riferimento alla domanda di informazione, le analisi sul consumo di notizie (reali e false) mediante le piattaforme online sono state condotte attraverso modelli di *big data analytics* svolti da Walter Quattrociocchi (incaricato dall'Autorità), e alcuni co-autori, su decine di milioni di *account social* (tendenzialmente Facebook e Twitter). Questi *dataset* sono brevemente descritti nell'ambito del Capitolo 4.

Con riguardo all'offerta informativa, nel presente Rapporto sono stati utilizzati dati sia sui giornalisti, sia sulla produzione di documenti informativi. Sui giornalisti, le informazioni provengono dalla [II edizione dell'Osservatorio sul giornalismo](#) e riguardano un campione di circa 2.000 professionisti italiani. Anche in questo caso, si rimanda all'Allegato al documento che contiene una dettagliata illustrazione degli aspetti metodologici della ricerca e del campione sottostante.

Con riferimento, infine, all'analisi dei prodotti informativi offerti in Italia, il *database* documentale utilizzato per le elaborazioni del presente Rapporto è stato costruito a partire dai dati estratti attraverso la piattaforma (sviluppata da *Volocom Technology*) di cui si è dotata l'Autorità per compiere studi sulla diffusione e il contenuto delle notizie riportate da tutti i mezzi di comunicazione.

Stante la assoluta novità di questo *dataset* e della relativa metodologia, nel prosieguo di questa Appendice si illustreranno le principali caratteristiche della piattaforma e del connesso database documentale.

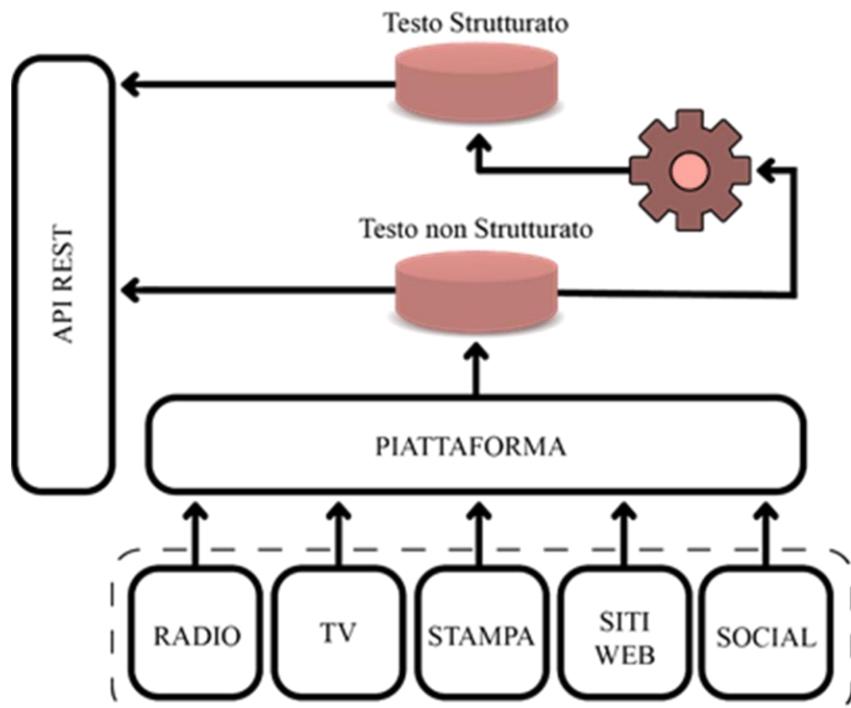
LA PIATTAFORMA

La Piattaforma utilizzata offre la possibilità di estrarre la ricorrenza di qualsiasi parola chiave (o insieme di termini) e di esaminare il testo di interi documenti riconducibili a notizie su qualsiasi argomento e accadimento.

Peraltro, la piattaforma, tenendo costantemente conto dell'aggiornamento dei flussi informativi delle fonti, è in grado di rilasciare serie storiche di dati e consente, quindi, lo svolgimento di analisi dinamiche.

Per quel che attiene al funzionamento della piattaforma (cfr. Figura A.0.1), l'output fornito dalla stessa consiste nel testo non strutturato proveniente dalla sorgente (rappresentata dai singoli mezzi di informazione) arricchito da metadati che caratterizzano il testo, quali fonte, data, categoria, ecc. Attraverso la piattaforma è possibile accedere allo storico delle sorgenti in modo da analizzarle nel tempo oppure lungo assi diversi che corrispondono ai metadati. Il sistema attuale di metadatazione prevede l'applicazione di opportuni algoritmi al testo non strutturato, in modo da estrarre i metadati che descrivono il testo.

Figura A.0.1 – Funzionamento della piattaforma

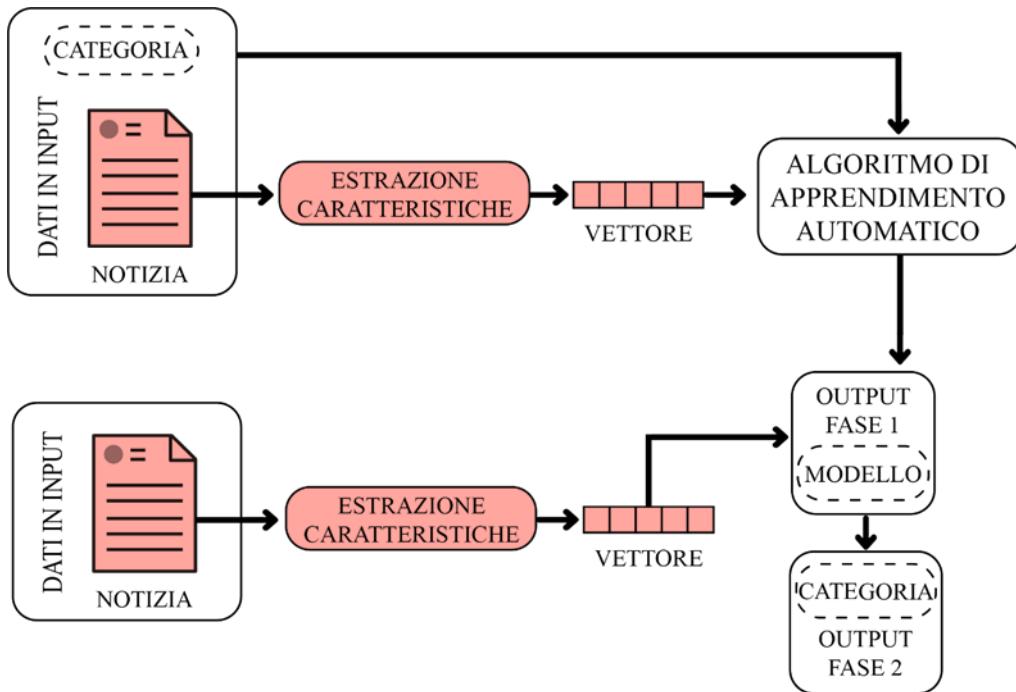


Sotto il profilo tecnico-informatico, i servizi della piattaforma sono resi disponibili da un'interfaccia programmatica basata su API REST (*REpresentational State Transfer*). Questo consente l'estrazione di dati dalla piattaforma con la massima granularità possibile, superando i limiti imposti da un'interfaccia utente predefinita e permettendo di procedere alle analisi sui dati con la più ampia flessibilità. Per l'interscambio dei dati, la piattaforma utilizza il formato JSON (*JavaScript Object Notation*).

Il funzionamento della piattaforma, inoltre, si avvale di due algoritmi rilevanti. Il primo è un algoritmo di *stemming*, che consente di estendere la ricerca alle diverse varianti di un termine prescelto. Il secondo è un algoritmo di intelligenza artificiale attraverso il quale avviene la classificazione automatica di ogni notizia rispetto ad una categoria di argomenti predefiniti (quali cronaca, politica, esteri, economia, esteri, scienza e tecnologia, cultura, spettacolo e sport).

Si tratta di un algoritmo ad apprendimento automatico di tipo supervisionato. Lo schema di funzionamento dello stesso, illustrato nella Figura A.0.2, consiste in una fase preliminare in cui l'algoritmo viene addestrato a partire da dati di input (*training dataset*), e in una fase operativa successiva in cui l'algoritmo, in seguito all'apprendimento, è in grado, per ogni nuovo documento ricevuto, di effettuare una predizione della categoria di appartenenza. Lo specifico algoritmo ad apprendimento automatico che viene utilizzato nella piattaforma è una macchina a vettori di supporto (o SVM - *Support Vector Machine*). Poiché l'algoritmo SVM è un classificatore lineare binario, la piattaforma esegue l'algoritmo con un procedimento iterativo in più passi ai fini di ottenere una classificazione del contenuto informativo in una delle categorie possibili.

Figura A.0.2 – Classificatore automatico dei contenuti informativi in categorie



LA COMPOSIZIONE DEL DATABASE DOCUMENTALE

Nella costruzione del *database* documentale sottostante alle elaborazioni del presente Rapporto, particolare cura è stata riposta nella preliminare definizione dell'insieme di fonti informative rappresentativo di ciascun mezzo di comunicazione. A questo aspetto è stata dedicata grande attenzione, in quanto dalla completezza e rappresentatività dell'insieme definito dipende la robustezza e la significatività dei risultati ottenuti nelle analisi effettuate.

Allo scopo di procedere alla corretta definizione dell'insieme di fonti, si è tenuto conto dell'intero sistema informativo, ricomprensivo i mezzi di comunicazione tradizionali (Tv, radio e quotidiani) e online (siti *web* di media tradizionali, testate esclusivamente online, piattaforme online).

Per ciascuno dei predetti mezzi, è stato poi selezionato il più ampio numero possibile di fonti informative in base alla disponibilità dei dati in formati atti ad essere acquisiti e trattati con le tecniche informatiche sopra descritte.

Da un punto di vista pratico, la selezione delle singole fonti informative è avvenuta tenendo conto di alcuni criteri specifici, riconducibili alla rilevanza assunta da ciascuna di esse in termini di:

- *audience* raggiunta;
- diffusione sul territorio;
- effettiva presenza di contenuto informativo all'interno dell'offerta proposta; così, ad esempio, nel caso delle piattaforme online, sono state selezionate le pagine Facebook e gli *account* Twitter riferibili a Tv, radio, quotidiani, siti di informazione e *influencer*.

Inoltre, alla luce dell'attuale conformazione del sistema informativo e dell'affermarsi sempre più diffusamente di forme patologiche di disinformazione online, si è ritenuto necessario, in modo da poter studiare empiricamente tali fenomeni, ricoprendere tra i mezzi che alimentano il *database* anche siti web e pagine/*account* di social network inclusi nelle liste di fonti "di fake news" predisposte da fonti esterne specializzate nell'attività di *debunking* (quali *butac.it* e *bufale.net*) utilizzate da svariati studi scientifici sulla disinformazione online.

Pertanto, in considerazione della struttura e delle criticità del sistema informativo italiano, della valenza (in termini di accesso da parte dei cittadini allo scopo di reperirvi informazioni) assunta dai predetti mezzi, e della tipologia di contenuti veicolati, nonché delle peculiarità della piattaforma di rilevazione utilizzata, l'insieme di fonti informative oggetto delle analisi condotte in questo Rapporto è composto come riportato nella successiva tabella.

Complessivamente, le fonti informative che alimentano il *database* documentale sono circa 1.800.

Nel dettaglio, il *database* si compone dell'intero contenuto testuale di tutti i documenti generati nell'arco di una giornata da ciascuna fonte informativa. Per documento si intende l'intero articolo, nel caso di quotidiani e siti *web* di informazione; la trascrizione di un segmento di trasmissione (servizio), nel caso di Tv e radio; l'intero *tweet/post*, nel caso delle piattaforme online.

Tabella A.1 – Insieme delle fonti informative che alimentano il *database*

MEZZI	FONTI INFORMATIVE
Televisione	Es. Rai 1, Rai 2, Rai 3, Rai News 24, Rete 4, Canale 5, Italia 1, TGCOM 24, La 7, SKYTG24, ...
Radio	Es. Radio 1, Radio 2, Radio 3, Radio 24, ...
Quotidiani	- Testate nazionali (es. La Repubblica, Corriere della Sera, Il Sole 24 Ore, La Stampa, ...) - Testate locali (es. L'Eco di Bergamo, Gazzetta di Mantova, ...)
Siti di canali/programmi televisivi	Es. rainews.it, tgcom24.mediaset.it,
Siti di canali/programmi radiofonici	Es. radio24.ilsole24ore.com, rtl.it/notizie, ...
Siti di quotidiani	Es. repubblica.it, corriere.it, ecodibergamo.it, ...
Testate esclusivamente online	Es. Il Post, Huffington Post, ...
Siti “di fake news”	Siti inclusi nelle liste redatte da fonti esterne specializzate in debunking
Piattaforme online (social network)	Pagine Facebook e account Twitter di Tv, radio, quotidiani, testate online, influencer e fonti di fake news

I dati che compongono il *database* fanno riferimento a un arco temporale molto esteso, che copre i 29 mesi che vanno da aprile 2016 ad agosto 2018. Pertanto, il *database* si caratterizza per dimensioni di notevole entità, conferendo grande robustezza ai risultati ottenuti. Le analisi svolte nel presente Rapporto, infatti, sono basate sull'esame di circa 35 milioni di documenti.