

OSSERVATORIO SULLA DISINFORMAZIONE ONLINE



Produzione di informazione e disinformazione



Categorie dell'informazione e della disinformazione



Tematiche della disinformazione



Informazione e disinformazione sulle elezioni europee



Informazione e disinformazione sui temi di rilevanza europea



Appendice - Termini salienti del primo bimestre 2019

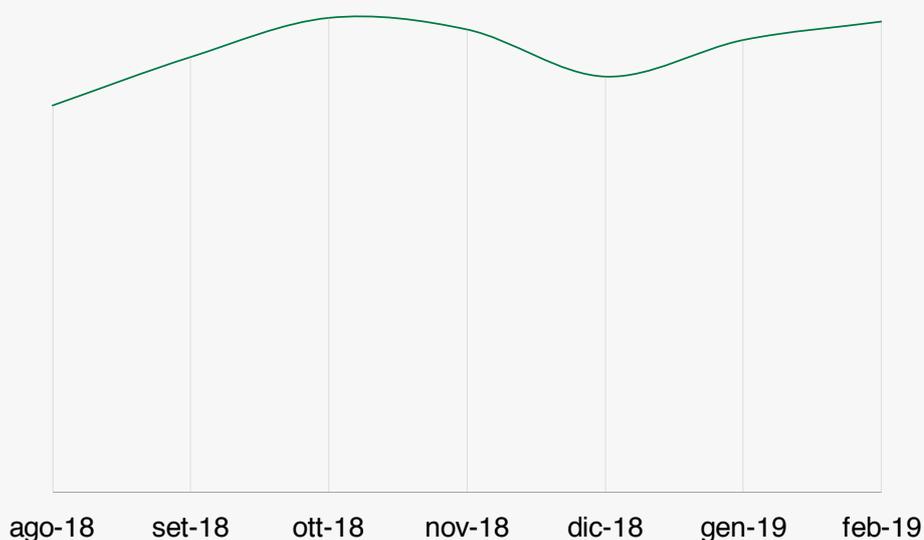


Nota metodologica



Contenuti informativi prodotti nel giorno medio

I bimestre 2019



INFORMAZIONE PRODOTTA IN ITALIA NEL PRIMO BIMESTRE 2019

Dall'analisi dinamica del volume informativo offerto dai media (Tv, radio, quotidiani, agenzie di stampa, siti web di informazione, relative pagine e account social) in Italia si rileva:

- un incremento dei contenuti di informazione prodotti nel giorno medio del primo bimestre del 2019, pari al 5% rispetto al bimestre precedente
- un'incidenza dei contenuti online sul totale dell'informazione prodotta prossima al 66%

DISINFORMAZIONE ONLINE PRODOTTA IN ITALIA NEL PRIMO BIMESTRE 2019

La disinformazione prodotta in Italia (da siti web, pagine e account social di disinformazione) nel primo bimestre del 2019 mostra un andamento crescente. In particolare:

- nei 6 mesi analizzati, la quantità di contenuti di disinformazione prodotti nel giorno medio raggiunge il valore massimo a febbraio 2019
- la disinformazione prodotta giornalmente nel primo bimestre 2019 rappresenta il 7% dei contenuti online
- si stima che un sito di disinformazione pubblici, in media, 5 nuovi articoli al giorno

Contenuti di disinformazione prodotti nel giorno medio

I bimestre 2019



+10% su dicembre 2018

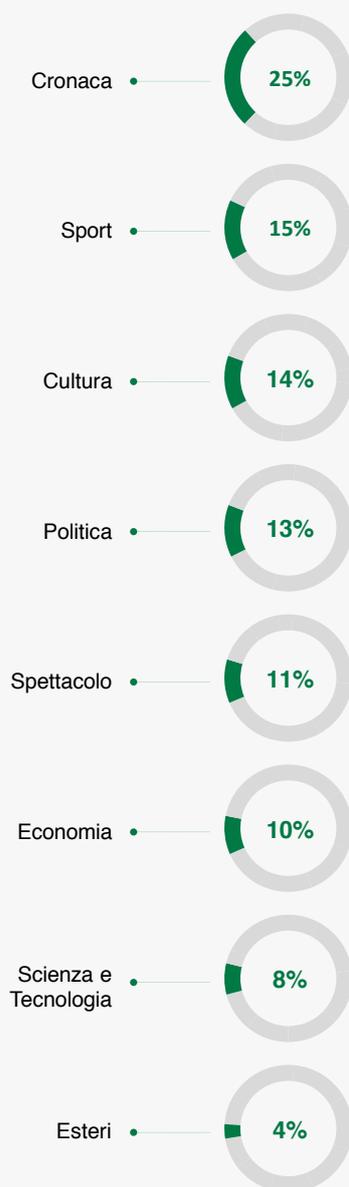
DISINFORMAZIONE ONLINE NEL GIORNO MEDIO DI GENNAIO 2019

+4% su gennaio 2019

DISINFORMAZIONE ONLINE NEL GIORNO MEDIO DI FEBBRAIO 2019

Categorie dell'informazione e della disinformazione

Ripartizione dell'offerta di informazione per categoria (I bimestre 2019)



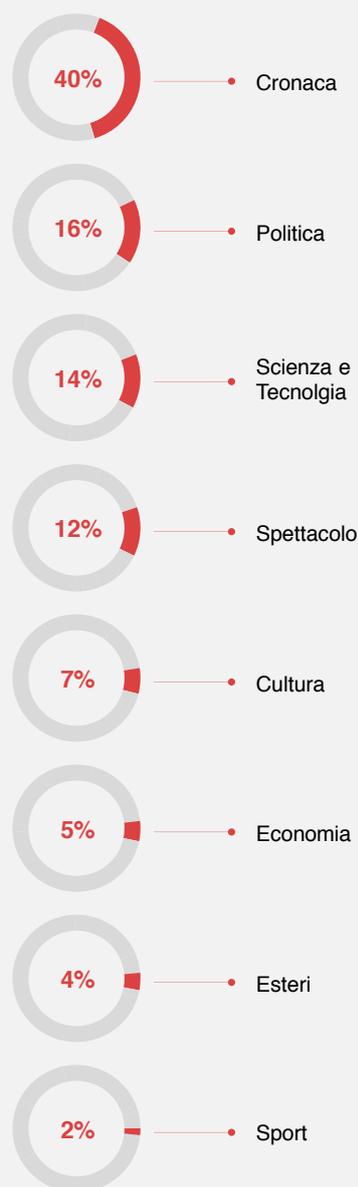
CATEGORIE DELL'INFORMAZIONE NEL PRIMO BIMESTRE 2019

Nel primo bimestre del 2019, la ripartizione dell'offerta informativa per categoria non subisce significativi cambiamenti rispetto alla media del 2018:

- cronaca rimane la prima categoria di genere informativo trattato (25% del totale). Le altre categorie c.d. *hard news* (politica ed esteri) rappresentano congiuntamente il 17%
- largo spazio è dedicato alle notizie di sport (15%), cultura (14%) e spettacolo (11%)

PERSISTENTE CARENZA DI INFORMAZIONE SU ARGOMENTI DI SCIENZA E TECNOLOGIA

Ripartizione dell'offerta di disinformazione online per categoria (I bimestre 2019)



CATEGORIE DELLA DISINFORMAZIONE ONLINE NEL PRIMO BIMESTRE 2019

La distribuzione per categoria degli articoli di disinformazione online nei primi mesi del 2019 mostra:

- una crescita rispetto al 2018 dei contenuti dedicati ai fatti di cronaca, che si collocano al primo posto. Seguono politica, che si conferma al secondo posto, e scienza e tecnologia, in leggera riduzione
- una spiccata concentrazione della disinformazione su cronaca e politica (56% del totale, 3 punti percentuali in più rispetto alla media del 2018)

CONCENTRAZIONE DELLA DISINFORMAZIONE SU FATTI DI CRONACA E POLITICA

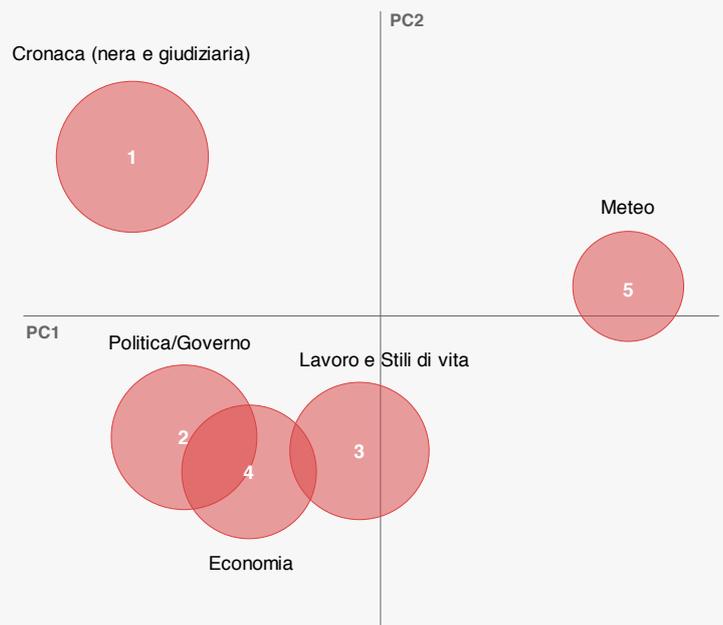
Tematiche della disinformazione

PRINCIPALI TEMATICHE DELLA DISINFORMAZIONE ONLINE NEL PRIMO BIMESTRE 2019

Dall'analisi testuale (*topic modeling*) dei contenuti dei siti di disinformazione, che consente di individuare delle macrotematiche sulla base della co-occorrenza di termini all'interno dei documenti, si riscontra che:

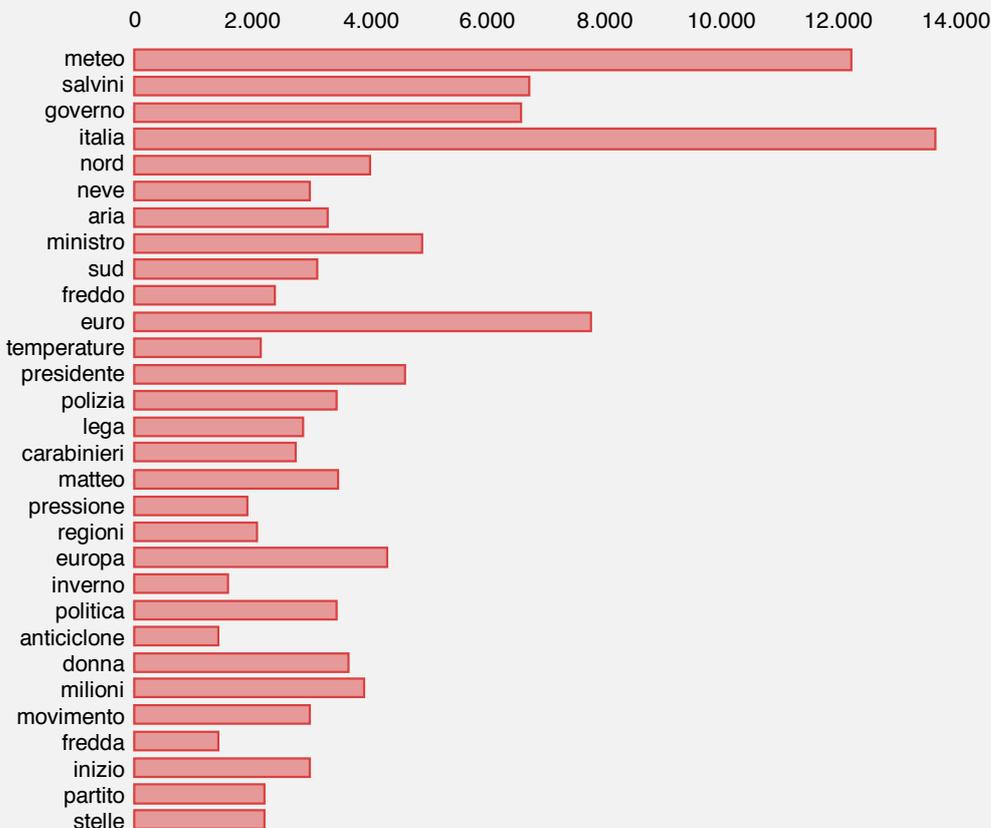
- i 5 principali argomenti oggetto di trattazione nel primo bimestre del 2019 riguardano la cronaca nera e giudiziaria, la politica e gli affari di governo, il lavoro e gli stili di vita, le questioni economiche, e le condizioni meteorologiche

Principali tematiche oggetto di disinformazione online (I bimestre 2019)



L'elenco dei 30 termini più frequenti riconducibili a ciascuna tematica è riportato in [Appendice](#)

Primi 30 termini per numero di occorrenze nei siti di disinformazione (I bimestre 2019)



TERMINI SALIENTI DELLA DISINFORMAZIONE ONLINE NEL PRIMO BIMESTRE 2019

La distribuzione di frequenza dei termini contenuti negli articoli dei siti di disinformazione mostra come:

- "italia", "meteo", "euro" siano le parole più ricorrenti nel primo bimestre del 2019
- molti dei 30 termini più frequenti siano legati a partiti/esponenti politici e soggetti istituzionali
- tra gli altri termini salienti, diversi siano riconducibili agli eventi climatici, oltre che all'intervento delle forze dell'ordine

CRONACA NERA, POLITICA/GOVERNO E LAVORO SONO LE TEMATICHE PRINCIPALI DELLA DISINFORMAZIONE ONLINE NEL PRIMO BIMESTRE 2019

SPAZIO DEDICATO ALLE ELEZIONI EUROPEE NEL PRIMO BIMESTRE 2019

Con riferimento al primo bimestre del 2019, si osserva che:

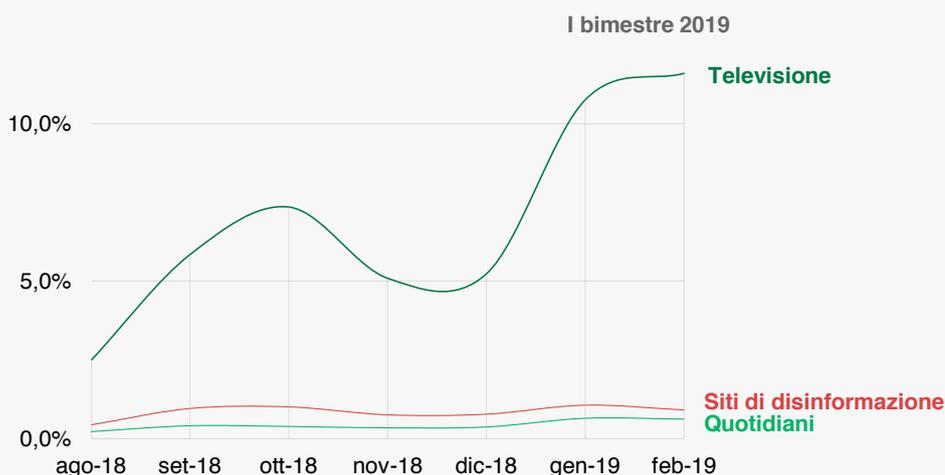
- lo spazio dedicato alle elezioni europee sia dalle fonti di informazione che da quelle di disinformazione si attesta su valori complessivamente superiori rispetto ai bimestri precedenti
- i contenuti di disinformazione, rispetto a quelli informativi (soprattutto di Tv e radio), tendono a concentrarsi sulle **tematiche** con un forte impatto emotivo (incluse quelle di **rilevanza europea**) piuttosto che sullo specifico argomento delle elezioni europee

Incidenza dell'argomento "elezioni europee" sul totale informazione e disinformazione



SPAZIO DEDICATO ALLE ELEZIONI EUROPEE DAI SINGOLI MEZZI NEL PRIMO BIMESTRE 2019

Incidenza dell'argomento "elezioni europee" sul totale dei contenuti prodotti dal mezzo



Concentrando l'analisi sulle singole fonti di informazione e disinformazione, si riscontra come, nel primo bimestre del 2019:

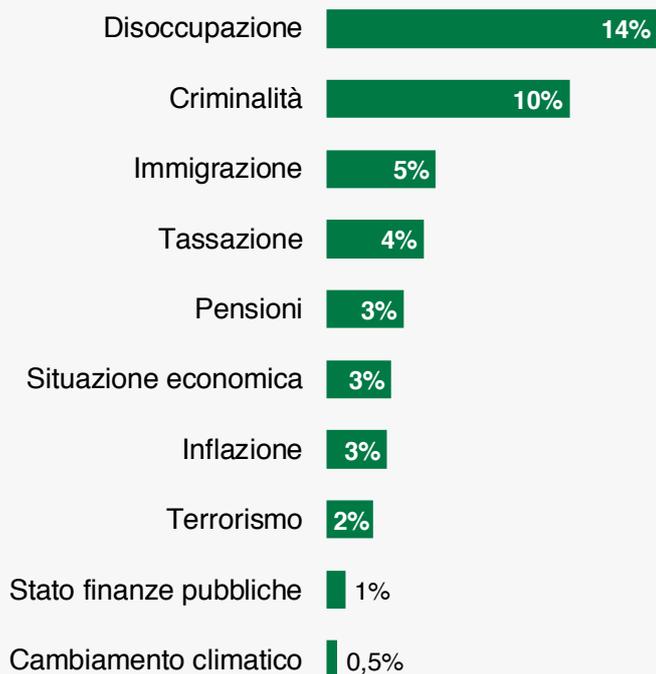
- la televisione si conferma il mezzo che attribuisce il maggior spazio alle elezioni europee rispetto al totale della propria offerta informativa
- i siti di disinformazione continuano comunque a dedicare alle elezioni europee una quota sul totale dei propri contenuti mediamente maggiore rispetto ai quotidiani (e ai siti di informazione)

+28%
sul bimestre precedente

DISINFORMAZIONE ONLINE SULLE
ELEZIONI EUROPEE NEL GIORNO
MEDIO DEL PRIMO BIMESTRE 2019

Informazione e disinformazione sui temi di rilevanza europea

Incidenza dei temi di rilevanza europea sul totale informazione (I bimestre 2019)



INFORMAZIONE SUI TEMI RILEVANTI PER L'UE NEL PRIMO BIMESTRE 2019

Focalizzando l'analisi sull'offerta informativa prodotta nel primo bimestre del 2019 sui temi di rilevanza europea individuati da Eurobarometro, si osserva che in Italia:

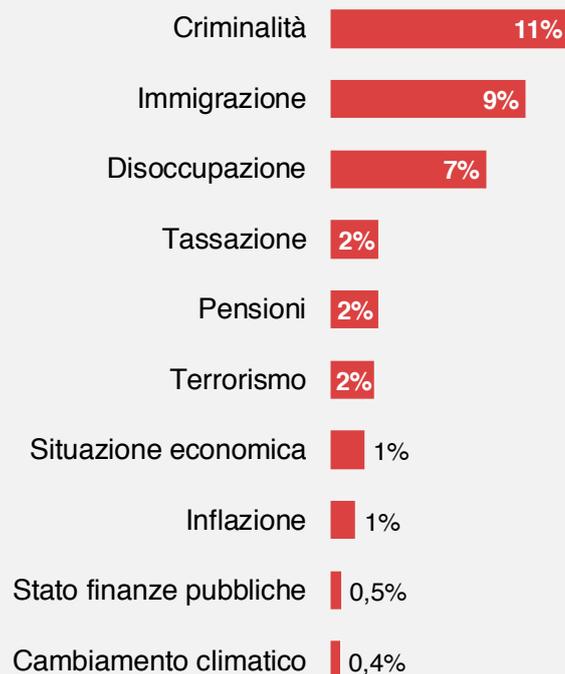
- disoccupazione e criminalità si confermano gli argomenti con la maggiore copertura da parte dei mezzi di informazione, con un'incidenza sul totale pari rispettivamente al 14% e al 10%
- l'immigrazione (5%) sale al terzo posto, mentre il minor spazio è dedicato al cambiamento climatico (meno dell'1%)

TRA I TEMI DI RILEVANZA EUROPEA:

DISOCCUPAZIONE È IL PIÙ TRATTATO DA TELEVISIONE, RADIO, QUOTIDIANI E FONTI EDITORIALI ONLINE

CRIMINALITÀ È IL PIÙ TRATTATO DA PAGINE/ACCOUNT SOCIAL DELLE FONTI DI INFORMAZIONE

Incidenza dei temi di rilevanza europea sul totale disinformazione (I bimestre 2019)



DISINFORMAZIONE SUI TEMI RILEVANTI PER L'UE NEL PRIMO BIMESTRE 2019

La distribuzione della disinformazione online prodotta nei primi 2 mesi dell'anno sugli argomenti di rilevanza europea consente di rilevare come:

- la criminalità rimanga il tema più trattato dalle fonti di disinformazione in Italia, con un'incidenza sul totale pari all'11%
- seguono, con quote non distanti tra loro, l'immigrazione (9% dei contenuti complessivamente prodotti dalle fonti di disinformazione) e la disoccupazione (7%)

TRA I TEMI DI RILEVANZA EUROPEA:

CRIMINALITÀ È IL PIÙ TRATTATO SIA DAI SITI CHE DALLE PAGINE/ACCOUNT SOCIAL DI DISINFORMAZIONE

IMMIGRAZIONE È QUELLO SU CUI LA DISINFORMAZIONE INCIDE MAGGIORMENTE

TERMINI SALIENTI DI CIASCUNA TEMATICA DELLA DISINFORMAZIONE ONLINE NEL PRIMO BIMESTRE 2019

La composizione delle singole tematiche della disinformazione consente di ricavare, in base ai termini salienti associati alle stesse, indicazioni circa l'oggetto specifico delle notizie false diffuse.

In tal senso, per ognuna delle 5 tematiche principali della disinformazione nel primo bimestre del 2019, si riporta l'elenco dei 30 termini più frequenti che le compongono:

1. CRONACA (NERA E GIUDIZIARIA)

- anni
- fatto
- uomo
- polizia
- casa
- donna
- carabinieri
- stati
- napoli
- giorni
- persone
- giovane
- vita
- famiglia
- durante
- momento
- carcere
- donne
- figlio
- procura
- corso
- roma
- ospedale
- oggi
- moglie
- uomini
- arrestato
- auto
- mamma
- indagini

2. POLITICA/GOVERNO

- salvini
- governo
- italia
- ministro
- fatto
- presidente
- matteo
- politica
- lega
- paese
- movimento
- anni
- oggi
- italiani
- partito
- stelle
- caso
- voto
- francia
- cittadini
- fonte
- leader
- interno
- legge
- europa
- sinistra
- renzi
- nazionale
- paesi
- persone

3. LAVORO E STILI DI VITA

- anni
- vita
- lavoro
- mondo
- tempo
- persone
- amore
- anno
- terra
- avere
- spesso
- caso
- corpo
- tutte
- ricerca
- volta
- oggi
- momento
- altre
- fatto
- possibile
- nuovo
- pensione
- volte
- attraverso
- viene
- studio
- giorno
- sistema
- mesi

4. ECONOMIA

- euro
- anni
- milioni
- italia
- legge
- fatto
- tutte
- anno
- oggi
- stati
- caso
- nazionale
- mondo
- rimanere
- persone
- lavoro
- sistema
- momento
- società
- qualsiasi
- tempo
- miliardi
- rispetto
- roma
- mercato
- dati
- bambini
- banca
- articolo
- grazie

5. METEO

- meteo
- italia
- nord
- neve
- aria
- sud
- freddo
- temperature
- giorni
- europa
- tempo
- pressione
- inizio
- regioni
- inverno
- settimana
- anticiclone
- centro
- fredda
- pubblicato
- primavera
- caldo
- condizioni
- aree
- acqua
- venti
- nevicata
- zone
- gelo
- precipitazioni

COMPOSIZIONE DEL DATABASE

Le informazioni riportate nel presente numero dell'Osservatorio sono il risultato di elaborazioni svolte dall'Autorità su un database documentale costruito a partire dai dati estrapolati attraverso la piattaforma sviluppata da *Volocom Technology*.

In particolare, è stato analizzato l'intero contenuto testuale estrapolato da 8,4 milioni di documenti generati in Italia, da agosto 2018 a febbraio 2019, da oltre 1.900 fonti informative (canali televisivi e radiofonici nazionali, quotidiani, agenzie di stampa, 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. Il volume di disinformazione online prodotto in Italia è stato quindi stimato con una metodologia di tipo soggettivo, ossia considerando il numero complessivo di documenti generati mensilmente dalle predette fonti di disinformazione.

Il database utilizzato si compone di tutti i documenti generati nell'arco di una giornata da ciascuna fonte informativa e di disinformazione. 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.

ANALISI TESTUALE ATTRAVERSO TOPIC MODELING

Per lo studio delle tematiche principali della disinformazione online, sono state condotte analisi sull'intero contenuto divulgato dai siti di disinformazione nel primo bimestre del 2019, adottando un'impostazione metodologica riconducibile al c.d. *topic modeling*.

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 di disinformazione in determinati argomenti è stata ottenuta 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 della co-occorrenza dei termini utilizzati nella raccolta di documenti. Lo stesso è poi alla base del sistema interattivo LDAvis (Sievert e Shirley, 2014), che 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. Nello specifico, 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.

APPROFONDIMENTI

Per una trattazione più dettagliata sugli aspetti metodologici, si rimanda al Rapporto Agcom (2018), [*News vs. fake nel sistema dell'informazione*](#)

Per approfondimenti di carattere definitorio e tecnico in merito alla disinformazione online, si rinvia al Rapporto Agcom (2018), [*Le strategie di disinformazione online e la filiera dei contenuti fake*](#)

Per ulteriori approfondimenti e aggiornamenti sulle attività dell'Autorità in materia di disinformazione online, è possibile consultare l'area dedicata del [*Tavolo pluralismo e piattaforme online*](#)

Tutti i numeri dell'Osservatorio sulla disinformazione online sono consultabili nella sezione [*Osservatori*](#) del sito web di Agcom



AUTORITÀ PER LE
GARANZIE NELLE
COMUNICAZIONI



www.agcom.it



ses@agcom.it

SERVIZIO ECONOMICO STATISTICO