

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## La qualità alla sfida dei Big Data

La ricerca nel Web sociale

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

13 Settembre, 2023

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- 1 BigData & Social Network
  - Streaming & Search
- 2 Dati raccolti & metodologie/tecnologie di analisi
  - Stima del numero di account italiani di Twitter
  - Topic Modelling con i Big Data Social
- 3 La Disinformazione & le Echo Chamber QCPS-II
- 4 La Qualità della Comunicazione Pubblica Social QCPS-I
- 5 Conclusioni

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## BigData & Social Network

# Timeline della nascita dei Big Data

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network

Streaming & Search

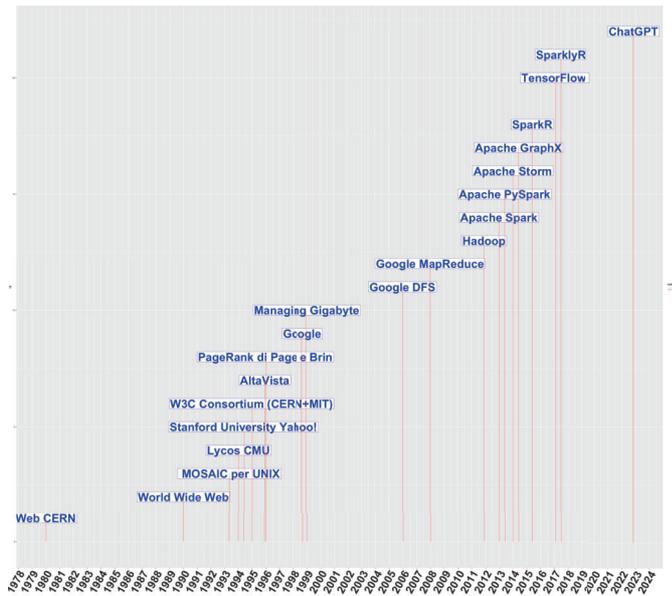
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- I Big Data sono spesso identificati come l'insieme dei nostri dati sensibili in rete, delle nostre interazioni pubbliche o private nei SN, utilizzati per profilare e predire le nostre azioni di utenti, cittadini o consumatori.
- I sensori di un aereo Boeing producono 8.000 dati al secondo. Sono circa 4 Petabyte di dati al giorno per migliorare le sue rotte e minimizzare i guasti.
- Con il 5G l'IoT produrrà enormi quantità di dati.
- Big Data + AI: Zero Touch Network per le reti 5G , 6G ...

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

### ● Big Data & AI

- I Big Data riducono i dati a Small (paradigma map reduce)
- I dati vengono distribuiti su più macchine. Su ciascuna macchina viene eseguito lo stesso codice. Non richiede CPU pregiate, piuttosto il loro numero (scalabilità orizzontale).
- Molti algoritmi di AI e ML possono essere eseguiti seguendo questi principi.
- AI come sinonimo di Deep Learning, genera milioni/miliardi di parametri da apprendere da Big Data. La distribuzione dei dati e degli algoritmi è più complicata. Ha bisogno di risorse di capacità di calcolo più performanti con GPU e TPU.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordon  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnolo-  
gie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

È possibile con una piccola infrastruttura HW & SW poter risolvere problemi anche significativi, grazie alla "*magia*" delle metodologie e tecnologie di Big Data.



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordononi  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

### 3 cluster di server

- 2 cluster di 8 server Fujitsu ciascuno: 512 core e **512 GB** di RAM, 24 TB di hard disk e 7.7 TB di SSD
- 1 cluster di 9 macchine Fujitsu: 36 core e **72 GB** di RAM, 18 TB
- 3 workstation, 2 UPS (Uninterruptible Power Supply - Gruppo di continuità), 2 condizionatori, 1 router, 1 switch a 64 porte, 1 monitor, 1 armadio rack

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie/  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- **Monitoraggio delle piattaforme sociali:**
- Si è misurata **quantitativamente e qualitativamente** la **capacità di presidio e di interazione** delle PA su Twitter.
- Misurata la **diffusione** della disinformazione nella rete italiana di Twitter
- Risultati del progetto QCPS in: G. Amati, S. Angelini, A. Cruciani, G. Fusco, G. Gaudino, Daniele Pasquini, P. Vocca, *Topic modeling by community detection algorithms* [Amati et al., 2021]
- Risultati si QCPS-II in: G. Amati, S. Angelini, A. Cruciani, G. Fusco, G. Gaudino, D. Pasquini, P. Vocca, G. De Luca *Classification and analysis of polarized Echo Chambers on the Italian Twitter Network [in preparazione]*

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## Streaming & Search

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordononi  
gamati@fub.it

BigData & Social Network

Streaming & Search

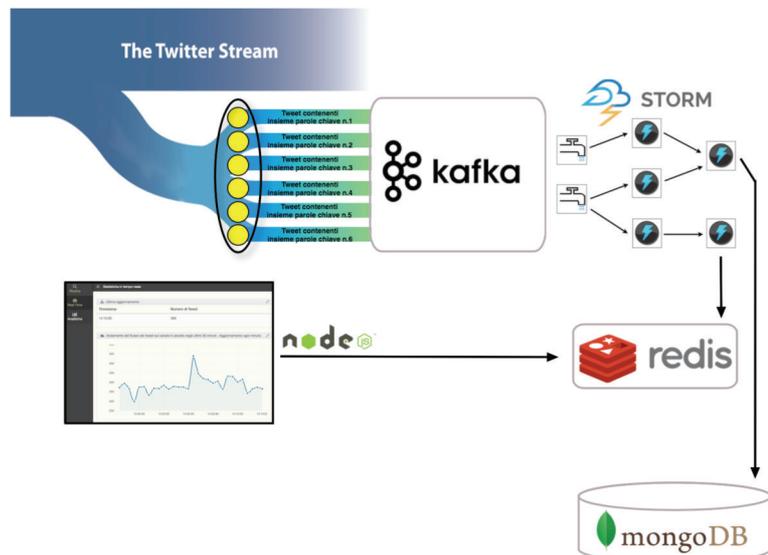
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-



La qualità alla sfida dei Big Data

Giambattista Amati  
 Fondazione Ugo Bordoni  
 gamati@fub.it

BigData & Social Network

Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-

Classifica degli hashtag più utilizzati negli ultimi 30 minuti sul canale in ascolto -

Rank	Hashtag	Numero di hashtag
1	#CyberSecurity	158
2	#cybersecurity	133
3	#Facebook	131
4	#infosec	102
5	#Cybersecurity	98
6	#Bitcoin	90
7	#CambridgeAnalytica	82
8	#Security	70
9	#IoT	62
10	#security	42
11	#AI	40
12	#databreach	37
13	#InfoSec	35
14	#Fraud	33
15	#blockchain	32
16	#GDPR	32
17	#malware	31
18	#Blockchain	29
19	#fintech	27
20	#BigData	27

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network

Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

🔍 Classifica degli autori più attivi negli ultimi 30 minuti sul canale in ascolto - Aggiornamento\*

Rank	Autore	Numero di tweet
1	@CyberToolsBooks	15
2	@CyberDomain	15
3	@imsmartinc	13
4	@KeoXes	13
5	@DNAProfileLaw	12
6	@IModeSixx	11
7	@AzurFlorian	11
8	@abdullah1212191	10
9	@sectest9	9
10	@blogger_boss	9
11	@ParadigmHQ	9
12	@FabrizioBustama	9
13	@Andrewwise911	9
14	@mrscribable	7
15	@theautomatski	6
16	@ReKnutecom	5
17	@NightWatcher6	5
18	@AnaGranger21	5
19	@he_mu_yu	4
20	@fernandocuena	4

La qualità alla sfida dei Big Data

Giambattista Amati  
 Fondazione Ugo Bordoni  
 gamati@fub.it

BigData & Social Network

Streaming & Search

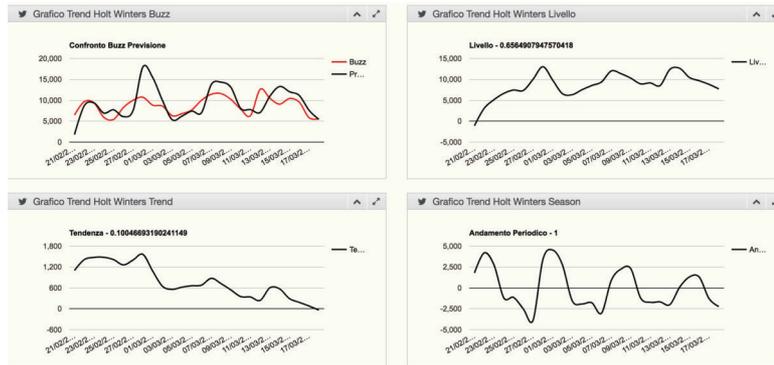
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
 QCPS-II

La Qualità della Comuni-



La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network

Streaming & Search

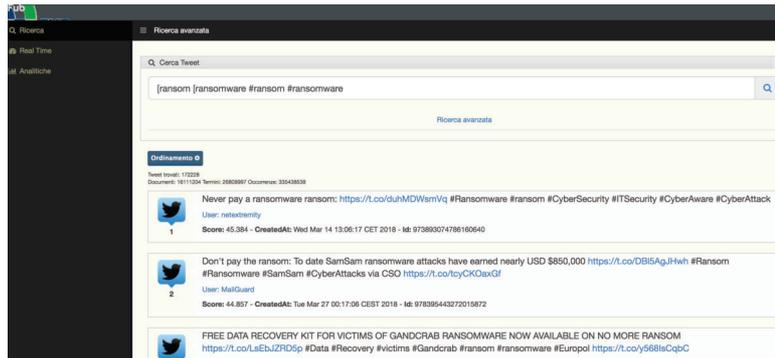
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCP5-II

La Qualità della Comuni-



The screenshot shows a Twitter search interface with the following details:

- Search bar: [ransom] [ransomware] #ransom #ransomware
- Search type: Ricerca avanzata
- Results list:
  - Item 1: User: netalexcherry, Score: 45.384, CreatedAt: Wed Mar 14 13:06:17 CET 2018 - Id: 973892074786160640. Text: "Never pay a ransomware ransom: <https://t.co/duhMDWsmVq> #Ransomware #ransom #CyberSecurity #ITSecurity #CyberAware #CyberAttack"
  - Item 2: User: MailGuard, Score: 44.857, CreatedAt: Tue Mar 27 00:17:04 CEST 2018 - Id: 976395443272015872. Text: "Don't pay the ransom: To date SamSam ransomware attacks have earned nearly USD \$850,000 <https://t.co/DBi5AgJhwh> #Ransom #Ransomware #SamSam #CyberAttacks via CSO <https://t.co/tyCkOaxGf>"
  - Item 3: User: MailGuard, Score: 44.857, CreatedAt: Tue Mar 27 00:17:04 CEST 2018 - Id: 976395443272015872. Text: "FREE DATA RECOVERY KIT FOR VICTIMS OF GANDCRAB RANSOMWARE NOW AVAILABLE ON NO MORE RANSOM <https://t.co/LaEbJZRD5p> #Data #Recovery #Victims #Gandcrab #ransom #ransomware #Europol <https://t.co/y568sCqbG>"

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordononi  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

Ricerca
Ricerca entità
Grafici

**Ordinamento** ▾

Menzioni trovate: 71442  
Numero totale di account: 5064825

1	 <p>User: <a href="#">@zdnet</a> Score: 2170.391</p>
2	 <p>User: <a href="#">@malwrhunterteam</a> Score: 2143.287</p>
3	 <p>User: <a href="#">@theybersechub</a> Score: 2132.782</p>
4	 <p>User: <a href="#">@scmagazine</a> Score: 1961.256</p>
	 <p>User: <a href="#">@darkreading</a> Score: 1888.251</p>

**malwrhunterteam**

```

friends_count: 50
favourites_count: 16899
created_at: 1414431991000
description: Official MHT Twitter account. CyberTracker is closed. Check out ID Ransomware (created by @demonstay335).
screenName: malwrhunterteam
geo_enabled: false
id: 2847021941
lang: en
time_zone: Vienna
statuses_count: 29467
followers_count: 49958
name: MalwareHunterTeam
                    
```

*Ultimo Aggiornamento: Mon Mar 19 10:55:25 CET 2018*

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network

Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-

author\_demonslay335 Q

Ricerca avanzata

Indici:

Data inizio:

Data fine:

Tipo ricerca:  Rilevanza  Freschezza  Sentiment

Nazione:

Ricerca   Ricerca entità   Grafici   Grafici Holt Winters   Concetti latenti

Ordinamento ▾

Tweet trovati: 85  
Documenti: 24896413 Termini: 38149860 Occorrenze: 50639789

1  RT @campusodi: Author of Polski, Vortex, and Flotera Ransomware Families Arrested in Poland <https://t.co/mXRVs869nF>  
#ransomware #poland #i...  
User: demonslay335  
Score: 1521332627000 - CreatedAt: Sun Mar 18 01:23:47 CET 2018 - Id: 97516683436717057

2  RT @BleepinComputer: Zenis Ransomware Encrypts Your Data & Deletes Your Backups - by @LawrenceAbrams  
<https://t.co/PDrBN8YzKE>  
User: demonslay335

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

**Dati raccolti  
& metodolo-  
gie/tecnolo-  
gie di analisi**

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## Dati raccolti & metodologie/tecnologie di analisi

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

- Twitter permette un **Filtro di 400 parole**. Abbiamo catturato il **Firehose italiano** con le 400 parole di maggiore uso comune (**stop-word**) + filtro della lingua italiana di Twitter
- Abbiamo 1,45 miliardi di Tweet:

Collezione	Nr Docs	Nr Termini Unici	Nr Occorrenze	Indici Dimensioni	Nr indexes
SNOOPI	53.643.416	82.937.329	1.095.045.889	83GB	111
QCPS-I (26/05/2020-25/04/2021)	902.380.246	706.600.254	14.874.506.326	966GB	1071
QCPS-II (23/02/2022-01/09/2022)	322.673.309	250.508.192	5.454.542.841	498GB	368
(23/02/2022-14/12/2022)	494.786.038	386.788.877	8.397.659.420	776GB	574

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordon  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

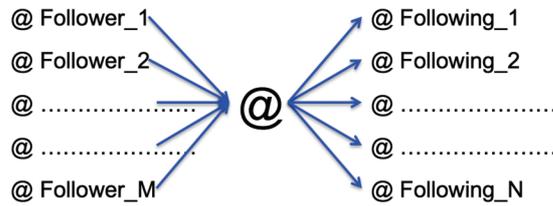
Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

- **Individuazione di 5,000 account pubblici.** Da 248 account istituzionali ufficiali iniziali si sono selezionati i 5,000 account più autorevoli tra i loro following.



La qualità alla sfida dei Big Data

Giambattista Amati  
 Fondazione Ugo Bordononi  
 gamati@fub.it

BigData & Social Network  
 Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

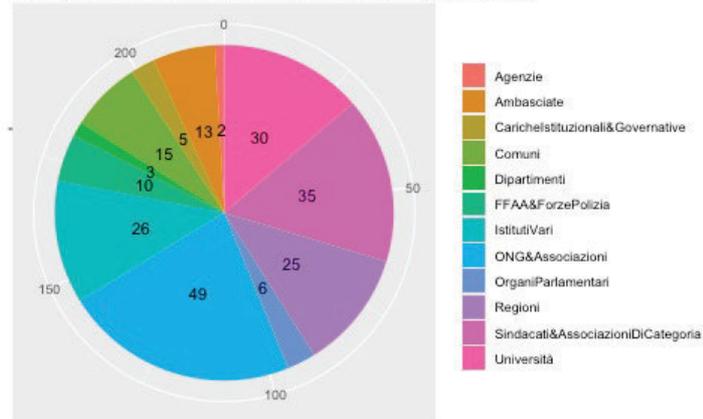
Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
 QCPS-II

La Qualità della Comuni-

Account pubblici o istituzionali tra i 5,000 account monitorati



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## Stima del numero di account italiani di Twitter

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Abbiamo monitorato per un giorno alcuni account italiani e confrontato il numero di tweet filtrati sugli account con quelli presenti nel flusso principale e valutato il tasso di falsi negativi uguale all'**6,86%** (140 falsi negativi ogni 2042 filtrati )
- Corrispondono a circa 40K account mancanti nel database.
- La stima degli utenti attivi italiani di Twitter inclusiva dei falsi positivi e falsi negativi del flusso è dunque 3.070.912
- Poiché si stima che il 32,5% degli utenti di Twitter non posti ma sia utente passivo, abbiamo circa **4,5 milioni di account italiani di Twitter**,

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- se ad ogni account dovesse corrispondere un unico utente, allora un estremo superiore del numero di utenti italiani sarebbe circa l'8,6% della popolazione italiana (ma non sappiamo quanti bot o utenti hanno account multipli).
  - Twitter afferma che esistono solo 2,8 milioni utenti italiani mensili raggiungibili dalla loro pubblicità.
  - 12,4 milioni di utenti nella fascia di età tra i 13 e i 64 anni (share del 32,8%) rapporto Digital 2021 di Datareportal.
  - 4,39% della popolazione totale [Statcounter].
  - 8,2 milioni nel 2017 [AGCOM 2017].
  - share del 9,9% [AGCOM 2020].

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordon  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

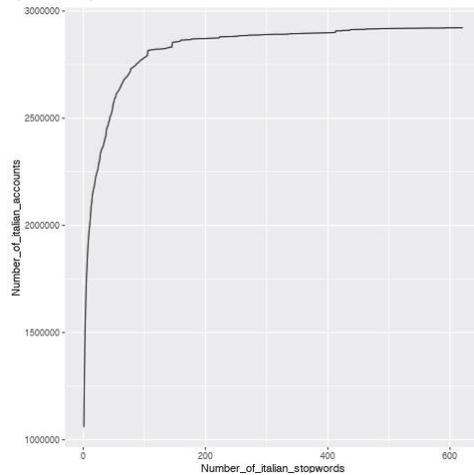
Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

La stima [Amati et al., 2021] è stata confermata in QCPS-II. Abbiamo considerato tutte le parole italiane di maggiore uso e abbiamo selezionato tutti gli account che abbiano utilizzato quelle parole. Sono stati individuati circa 2.7 milioni di utenti:



La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network  
Streaming & Search

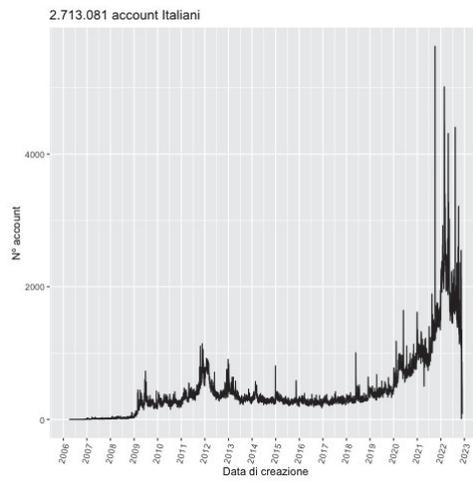
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## Topic Modelling con i Big Data Social

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

- **clustering dei dati** + **classificazione dei cluster in argomenti** (topic)
  - topic modelling convenzionale (testuale) come LDA o K-means, sono di fatto intrattabili
  - clustering massivo come LSH (lineare): più adatto a risolvere il problema del Near Duplicate Detection.
  - Abbiamo ricondotto il task a un problema di Community detection su grandi grafi, scegliendo tre algoritmi lineari
    - **CoDA**, ma è attualmente in versione **standalone**
    - **Algoritmo di Louvain** **versione distribuita**
    - **Algoritmo di Leiden** molto efficiente anche se non ancora in versione distribuita

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network

Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

- Grafo pesato degli *hashtag*
  - 2.726.034 *hashtag*, considerati come i nodi del grafo;
  - 34.769.762 *co-occorrenze*, considerati come gli archi del grafo.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCP5-II

La Qualità  
della Comuni-

- Louvain è estremamente veloce
  - 64,338 comunità di hashtag individuate
  - valutazione manuale delle comunità più grandi
  - valutazione manuale degli hashtag più frequenti
  - ogni tweet è classificato in base alla topic che riceve più voti con gli hashtag che contiene (come fa Random Forest)
  - Abbiamo un algoritmo proprietario **Propagate-S** e distribuito (7 volte più efficiente nel calcolare il ground truth di grandi grafi) che calcola la centralità dei nodi e degli archi. È possibile applicarlo all'interno di questi cluster per eliminare gli archi centrali e ottenere una ripartizione più fine. [Amati et al., 2023]

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordononi  
gamati@fub.it

BigData & Social Network

Streaming & Search

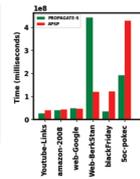
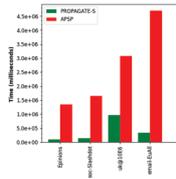
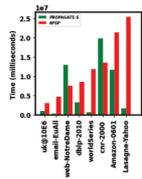
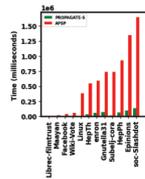
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCP5-II

La Qualità della Comuni-



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordononi  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCP5-II

La Qualità  
della Comuni-

Classi	Nr hashtag	Citazioni	%
Tv & Radio	54.092	52.257.298	35,09%
Politica	478.869	34.433.579	23,12%
Sport	223.304	17.727.610	11,90%
Salute	33.538	7.666.946	5,15%
Arte & Cultura	104.351	5.801.901	3,90%
Scienza & Tecnologia	182.622	5.169.674	3,47%
Mondo & Europa	7.666	3.441.604	2,31%
Notizie	1012	4.159.098	2,79%
Musica	17.982	2.940.716	1,97%
Turismo	20.888	2566472	1,72%
Economia & Lavoro	10.915	2468186	1,66%
Sesso	48.307	1905758	1,28%
Giochi	49.062	1255619	0,84%

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

Cinema & spettacolo	8.461	996.074	0,67%
Cucina	22.148	765.366	0,51%
Social Network	6.791	722.901	0,49%
Istruzione	7.452	700.12	0,47%
PA & Istituzioni	2.741	579.316	0,39%
Pubblicità	7.652	394.309	0,26%
Meteo & Traffico	6.604	373.956	0,25%
ambiente	5.954	243.52	0,16%
fakenews	1.828	135.52	0,09%
Altro	72	2.236.543	1,50%

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## La Disinformazione & le Echo Chamber QCPS-II

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordon  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnolo-  
gie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

La manipolazione su larga scala sui social media da parte della disinformazione ha due caratteristiche [Hristakieva et al., 2022]:

- (i) usa la propaganda per influenzare gli altri;
- (ii) adotta o crea delle azioni coordinate per diffondere la propaganda e amplificare l'impatto.

Una Echo Chamber [Jamieson and Cappella, 2008] ha le seguenti caratteristiche:

- (i) si manifesta attraverso più canali multi-mediali (uno o più programmi radio e televisivi, uno o più giornali, uno o più utenti di una piattaforma social);

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

(ii) obiettivo/risultato è di rinforzare i bias cognitivi dei partecipanti (omofilia degli utenti)

(iii) screditare gli altri punti di vista

Twitter è la piattaforma social "perfetta" per coordinare e costituire una Echo Chamber.

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordon  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

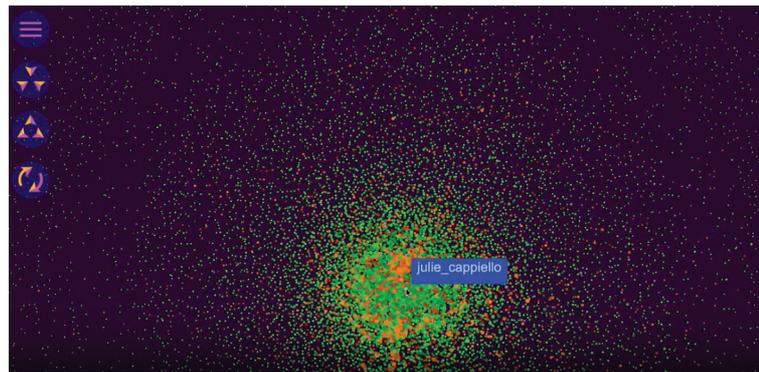


Figura 1: L'Echo Chamber di Trump: molti bot o profili falsi popolari(verdi centrali)

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordon  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

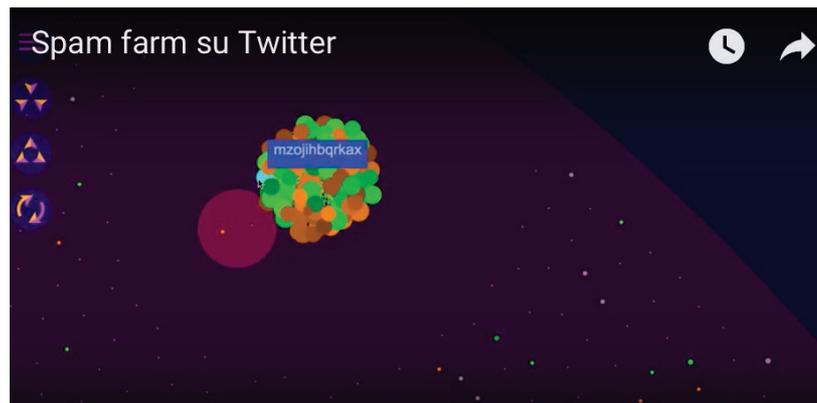


Figura 2: Comunità di bot (spam farm).

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter  
Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

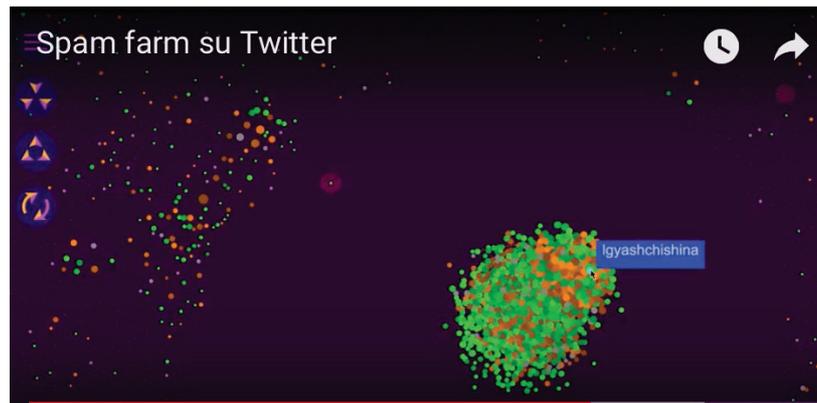


Figura 3: Comunità di 80.000 bot russi (spam farm)

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-

- Individuazione di un insieme di account (seed), della disinformazione novax o filorussa. A tale scopo abbiamo utilizzato un insieme abbastanza esaustivo di hashtag polarizzati su ciascun argomento.
- Costruzione del grafo dei Retweet come grafo di omofilia [Conover et al., 2021].

Misure	Valori
Diametro (lower bound)	44
Diametro effettivo (90%)	8.46
Average distance	6.31
Numero di coppie raggiungibili	8.64 bilioni
Numero di coppie raggiungibili al 90%	7.78 bilioni
Numero di coppie raggiungibili ad hop 2	12.43 miliardi

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordononi  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-

- Inclusionione degli hashtag come parte integrante del grafo: due tipologie di nodi (utenti e hashtag) e tre tipologie di archi (tra utenti, tra hashtag e tra utente e hashtag).
  - 10.613.744 nodi (2.465.333 nodi di tipo hashtag e 8.148.411 nodi di tipo user)
  - 343.766.707 archi (166.543.447 interazioni totali di tipo retweet e 177.223.260 interazioni totali di tipo hashtag)
- L'algoritmo di clustering è stato applicato sull'intero grafo pesato dalle frequenze occorrenti nelle relazioni indipendentemente dalle diverse tipologie e contiene anche il valore di importanza dei nodi (PageRank).
- Il risultato è stato quello di costruire cluster finali contenenti sia utenti sia gli hashtag utilizzati da quegli stessi utenti.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Abbiamo considerato solo i cluster costituiti da almeno 100 utenti che contenessero un numero significativo di utenti novax o utenti filorussi.
- Sono stati individuati in entrambi i casi 11 cluster novax e 11 cluster filorussi, 6 di questi erano in comune. Abbiamo analizzato quindi 16 cluster complessivi.
- Abbiamo quindi già il primo dato preoccupante: ben il 15% dei cluster più numerosi sono popolati da un numero significativo di account della disinformazione. Il primo cluster più numeroso e con maggior grado di omofilia (CPM) è riconducibile tutto alla propaganda sia novax sia filorussa.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Questa sottorete costituisce circa 11% circa del traffico complessivo di Twitter (ATTENZIONE è un lower bound!)

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordon  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCP5-II

La Qualità  
della Comuni-

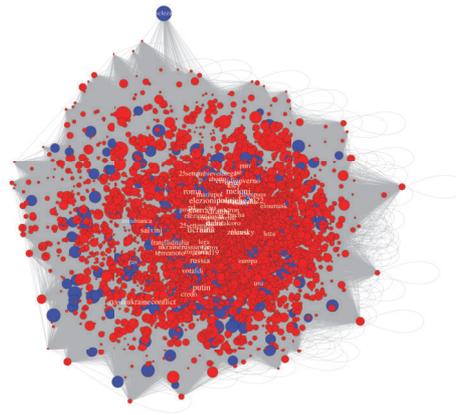


Figura 4: Comunità di novax e filorussi (cluster 3-0)

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordon  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnolo-  
gie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

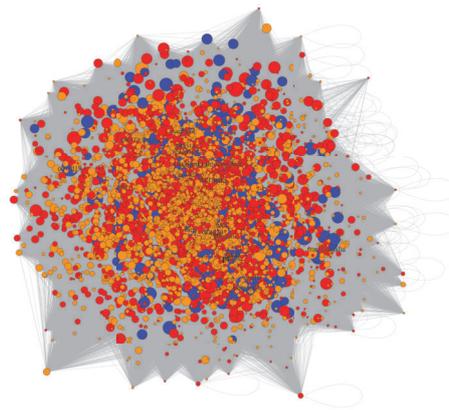


Figura 5: Comunità di novax e filorussi (cluster 3-0). In giallo i profili novax e filorussi.

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordononi  
gamati@fub.it

BigData & Social Network

Streaming & Search

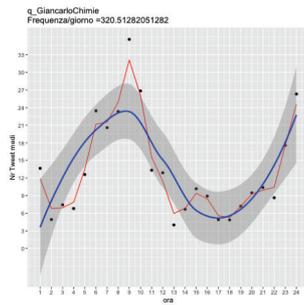
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-



← Profilo



@giancarlochimie

### Account sospeso

Sospendiamo gli account che violano le Regole di X.  
[Scopri di più](#)



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

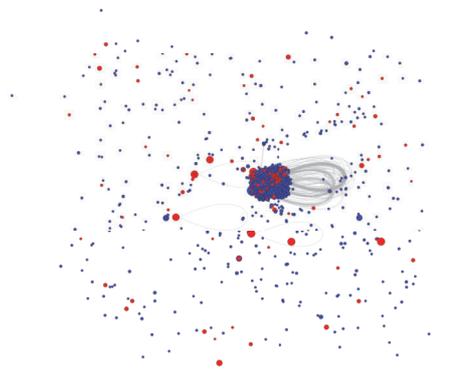


Figura 7: Comunità di novax (cluster 2-31)

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordononi  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

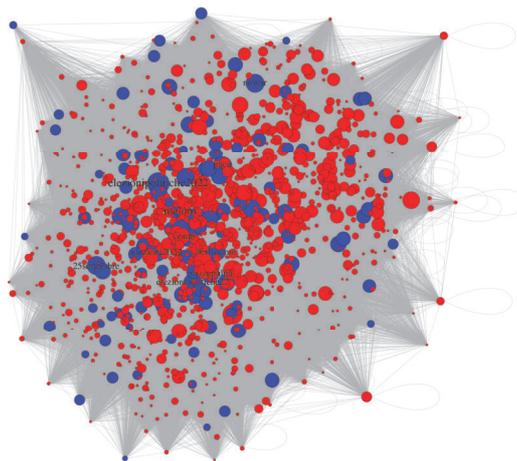


Figura 8: Comunità di novax (cluster 0-2)

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordononi  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnolo-  
gie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCP5-II

La Qualità  
della Comuni-

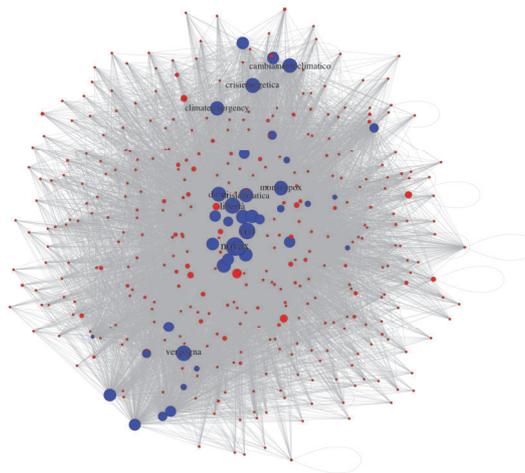


Figura 9: Comunità di novax (cluster 0-18)

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordononi  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

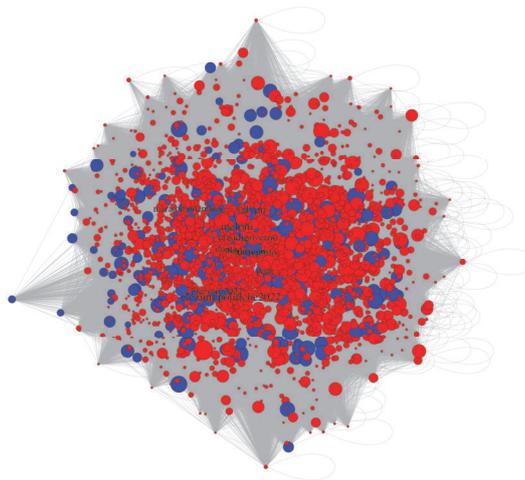


Figura 10: Comunità di novax (cluster 3-1)

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## La Qualità della Comunicazione Pubblica Social QCPS-I

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordonì  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

Il contenuto medio di sentiment di un tweet normalizzato mediante lo **Z score** ci permette di rapportarlo ad altre grandezze che altrimenti non sarebbero state confrontabili.

Riportiamo su un piano le due grandezze

- Lo **Z score del sentiment** di un tweet di interazione
- Lo **Z score del logaritmo del numero di interazioni** in ordinata

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network  
Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-

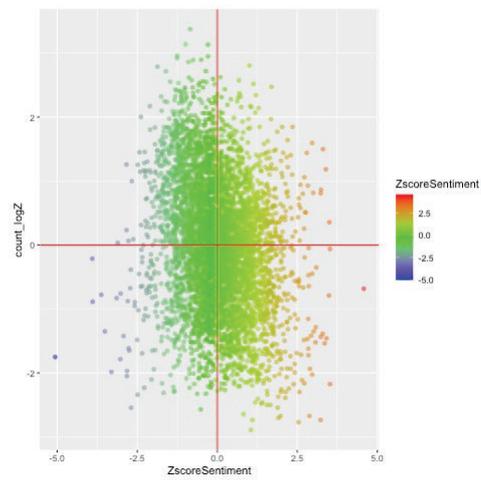


Figura 11: Distribuzione dei 5000 account per sentiment e interazioni con gli utenti (log scale)



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCP5-II

La Qualità  
della Comuni-

- Il numero delle **interazioni** con gli utenti
- Il numero dei **tweet emessi** dall'autore
- Il valore del contenuto del **sentiment medio** nelle interazioni con gli utenti
- Il numero dei **like** effettuati dall'account
- Il numero delle **liste** in cui l'account compare
- Il numero dei **followers** dell'account
- Il numero dei **following** dell'account

$SIndex = \text{mean}(Z(\text{Sentiment}), Z(\text{following}), Z(\text{followers}), Z(\text{citazioni}), Z(\text{tweets}), Z(\text{liste}), Z(\text{like}))$

# Indice sintetico di posizionamento social degli account istituzionali I

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordononi  
gamati@fub.it

BigData & Social Network

Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-

Show 10 entries

Search:

	categoria2	N
1	ONG&Associazioni	49
2	Sindacati&AssociazioniDICategoria	35
3	IstitutiVari	34
4	Università	30
5	Regioni	25
6	Comuni	15
7	Ministeri	14
8	Ambasciate	13
9	Dipartimenti	11
10	FFAA&ForzePolizia	11

Showing 1 to 10 of 13 entries

Previous 1 2 Next

id	name	account	categoria2	Stocks	ZscoreSettlement	countZ	followers_countZ	retweets_countZ	likes_countZ	friends_countZ
1	Papa Francesco	@pontifex_8	CaricheIstituzionali&Governative	8.28935995682809	-1.17748100796258	4.43271376115242	7.21763802982204	-0.295179127338197	2.85019146320064	-0.19314882584713
2	Stefano Ponticorno	@ponticornoita	CaricheIstituzionali&Governative	6.45820094516532	2.0697232878905	-0.231044631511447	0.0351319297223813	0.418074821183021	0.0313890079287383	0.0153025637527428
3	CISL Nazionale	@cislazionale	Sindacati&AssociazioniDICategoria	2.90796633563331	0.522693319406884	1.11780720748993	-0.0147350610088732	1.43156699736266	0.0713116451264304	0.0842558728729367
4	Regione Lazio	@regionelazio	Regioni	2.80768382796211	0.043671645948043	0.32670787800533	0.02814639185812177	0.135920476070815	0.157115622020257	-0.0541238587623254
5	Giulio Terzi	@giulioTerzi	CaricheIstituzionali&Governative	2.76959624461138	0.89799484827641	0.117055598979383	0.830003188070383	1.7238306748117	0.261779443109615	0.416317711094968
6	ILCOM&MCDIA	@ilcomcdia	Regioni	2.73861511082432	0.82328862182058	-0.17493125732711	-0.0182243361451915	0.38785946231433	0.091570953545333	0.12823715882363
7	Caritas Ambrosiana	@caritas_milano	ONG&Associazioni	2.48614683037197	-0.637091645183662	0.342803684761172	0.0181253465339948	-0.174837392074624	0.017083617656619	4.90469091654488
8	Ministero della cultura	@mibac	Ministeri	2.48718059617425	0.87754344664807	0.15506363294889	0.349332061800101	0.000135181082186	0.10070840321535	0.700771124745138
9	Operadomunitarie	@operadomunitarie	ONG&Associazioni	2.071659108295016	1.38365789531414	-0.267879580346207	0.0170827017910381	0.16795488641148	0.852096568883863	-0.16886644628997
10	Espa Druas	@espadrusas	ONG&Associazioni	1.8912388527005	1.873830051948	1.05973185406207	-0.0213742515860768	0.103540911182411	-0.025430135680754	-0.10713186440323

Showing 1 to 10 of 138 entries

Previous 1 2 3 4 5 ... 28 Next

La qualità alla sfida dei Big Data

Giambattista Amati  
 Fondazione Ugo Bordoni  
 gamati@fub.it

BigData & Social Network

Streaming & Search

Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-

Show  entries

Search:

	name	screenName	Index
1	JuventusFC	juventusfc	20.2822763810622
2	AC Milan	acmilan	19.8208051425957
3	Valentino Rossi	ValeYellow46	15.7620372727374
4	Laura Pausini	LauraPausini	10.3709261738785
5	Lorenzo Jovanotti	lorenzojova	9.9867022402005
6	European Commission	EU_Commission	9.4008698737428
7	Gianluigi Buffon	gianluigibuffon	8.79578501821457
8	Inter	Inter	8.34681035091051
9	Papa Francesco	Pontifex_it	8.28935995682869
10	Giorgio Chiellini	chiellini	8.17021325073813

Showing 1 to 10 of 5,058 entries

Previous      ...  Next

## Evoluzione del posizionamento social per alcuni account istituzionali I

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network

Streaming & Search

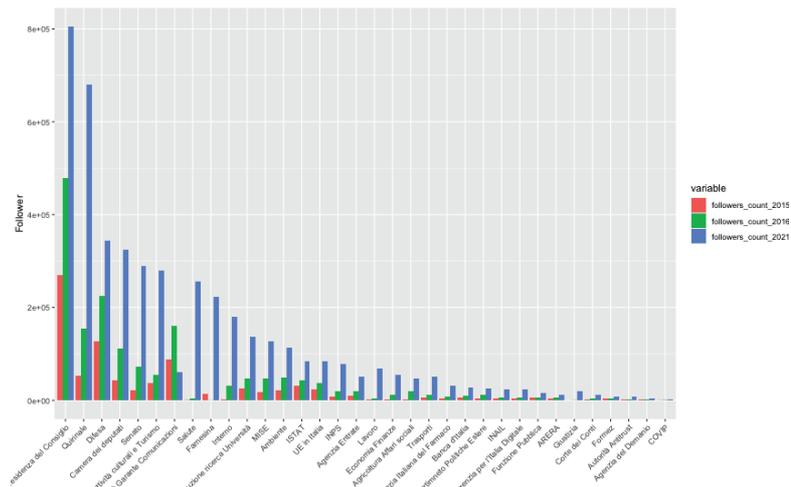
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber  
QCPS-II

La Qualità della Comuni-



## Evoluzione del posizionamento social per alcuni account istituzionali I

La qualità alla sfida dei Big Data

Giambattista Amati  
Fondazione Ugo Bordoni  
gamati@fub.it

BigData & Social Network

Streaming & Search

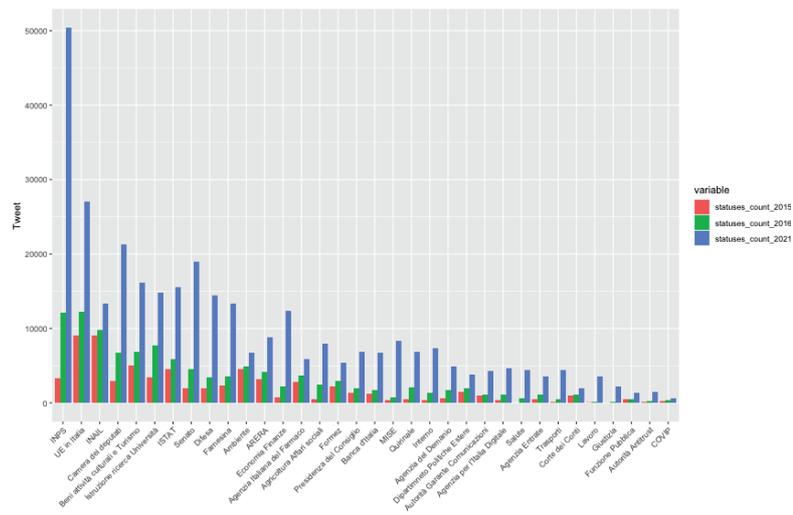
Dati raccolti & metodologie/tecnologie di analisi

Stima del numero di account italiani di Twitter

Topic Modelling con i Big Data Social

La Disinformazione & le Echo Chamber QCP5-II

La Qualità della Comuni-



La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

## Conclusioni

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Abbiamo mostrato che Twitter è facilmente esposto per la sua natura aperta alle interazioni di tutti con tutti al fenomeno della disinformazione.
- Il lavoro dimostra che esiste un'attività non residuale ma consistente della diffusione della disinformazione sui social in Italia.
- I primi cluster più numerosi e più attivi nell'azione di retweeting afferiscono ad attività di propaganda sia novax sia filorussa, costituiscono da soli, escludendo tutte le altre interazioni con il resto della TwitterSphere circa l'11% circa del traffico complessivo di Twitter di un anno.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Ciò dimostra che la disinformazione sfrutta ogni argomento divisivo per inquinare il dibattito politico italiano allo scopo di minare l'autorità esistenti, governi, partiti politici, media e istituzioni, attraverso i quali in precedenza passava tutta l'informazione.
- La crescita esponenziale dell'offerta di informazione dei nuovi media digitali comincia a sovraccaricare cognitivamente le persone e l'esclusione delle notizie di cattiva qualità, false o manipolate è sempre più difficile[Gurri, 2018].

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- L'uomo non segue il modello di ragionamento razionale [Haidt, 2013] ma parte sempre da una marcata preferenza (bias) per una realtà di suo piacimento e utilizza poi le sue capacità cognitive per selezionare i dati empirici e formulare teorie a sostegno di quella visione della realtà, con un processo definito di "ragionamento motivato".
- Le piattaforme digitali, per finalità economiche, fanno largo uso del ragionamento motivato.
- Le piattaforme sociali favoriscono il formarsi delle echo chamber che diventano un punto di riferimento per molti utenti della rete anche senza entrare a farne parte in modo attivo ma in qualità di semplici follower.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Maggiore è la diffusione di queste echo chamber maggiore è l'esposizione dell'utente comune ad accedere a questi cluster seppure come utenti passivi.
- La situazione preoccupante è stato il cambio recente di policy di sicurezza adottate da Twitter dopo l'acquisizione di Musk. L'intero reparto di sicurezza rivolto a far rispettare degli standard condivisi di sicurezza e rispetto delle regole di interazione social è stato smantellato in nome di una direttiva che segue il "free speech" sulla rete, la cui definizione non sembra ancora stata chiarita e soprattutto in quale modo venga declinata se non in quella di eliminare gran parte delle regole di condotta fin qui adottate.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Inoltre, si arriva al dilemma di come bilanciare il diritto di tutti gli utenti alla privacy insieme al diritto di parola da un lato, soprattutto nell'incertezza delle regole adottate dalle piattaforme stesse ospitanti, e il diritto alla trasparenza della rete nelle modalità di diffusione e di generazione dell'informazione dall'altro.
- Siamo stati in grado di fornire una dimensione (un lower bound) del traffico relativo alla disinformazione in Italia che ha riguardato principalmente se non del tutto nel 2022 la propaganda in favore della guerra contro l'Ucraina e l'efficacia dei vaccini.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

- Con le restrizioni di Musk di accesso alle API di Twitter non è più possibile utilizzare la nostra metodologia e comunque anche uno studio limitato avrebbe oggi un costo proibitivo di circa 500K di dollari.
- Di fatto, oggi, non è più possibile una verifica da terze parti del livello di diffusione della disinformazione su Twitter.
- Sarebbe opportuno che i 1.45 miliardi di dati possano essere utilizzati anonimizzandoli e organizzandoli in modo opportuno per fornire anche da parte di altri istituti terzi di ricerca uno studio completo ed esaustivo della disinformazione in Italia.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

 Amati, G., Angelini, S., Capri, F., Gambosi, G., Rossi, G., and Vocca, P. (2016).

Modelling the temporal evolution of the retweet graph.

*IADIS International Journal on Computer Science & Information Systems*, 11(2, ISSN: 1646-3692):19–30.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di  
analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

 Amati, G., Angelini, S., Cruciani, A., Fusco, G., Gaudino, G., Pasquini, D., and Vocca, P. (2021).

Topic modeling by community detection algorithms.

In Guidi, B., Michienzi, A., and Ricci, L., editors, *OASIS@HT 2021: Proceedings of the 2021 Workshop on Open Challenges in Online Social Networks, Virtual Event, Ireland, 30 August 2021*, pages 15–20. ACM.

 Amati, G., Cruciani, A., Pasquini, D., Vocca, P., and Angelini, S. (2023).

Propagate: a seed propagation framework to compute distance-based metrics on very large graphs.

In *ECML-PKDD 2023*, page to appear.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

 Boldi, P., Rosa, M., and Vigna, S. (2011).

HyperANF: Approximating the neighbourhood function of very large graphs on a budget.

In *Proc. 20th International Conference on World Wide Web*, pages 625–634, Hyderabad, India.

 Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., and Flammini, A. (2021).

Political polarization on twitter.

*Proceedings of the International AAAI Conference on Web and Social Media*, 5(1):89–96.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodolo-  
gie/tecnologie  
di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinforma-  
zione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-



Eppstein, D. and Wang, J. (2001).

Fast approximation of centrality.

*In Proc. 12th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 228–229, Washington, D.C., USA.



Flajolet, P. and Martin, G. N. (1985).

Probabilistic counting algorithms for data base applications.

*J. Comput. Syst. Sci.*, 31(2):182–209.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

 Gionis, A., Indyk, P., and Motwani, R. (1999).

Similarity search in high dimensions via hashing.

*In Proc. 25th Int. Conf. on Very Large Data Bases*, pages  
518–529, Edinburgh, UK.

 Gurri, M. (2018).

*The Revolt of the Public and the Crisis of Authority in the  
New Millennium.*

Stripe Press, San Francisco.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network  
Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

 Haidt, J. (2013).

*Menti tribali. Perché le brave persone si dividono su politica e religione.*

Torino.

 Hristakieva, K., Cresci, S., Da San Martino, G., Conti, M., and Nakov, P. (2022).

The spread of propaganda by coordinated communities on social media.

In *14th ACM Web Science Conference 2022, WebSci '22*, pages 191–201, New York, NY, USA. Association for Computing Machinery.

La qualità alla  
sfida dei Big  
Data

Giambattista  
Amati  
Fondazione  
Ugo Bordoni  
gamati@fub.it

BigData &  
Social  
Network

Streaming & Search

Dati raccolti  
& metodologie/  
tecnologie di analisi

Stima del numero di  
account italiani di  
Twitter

Topic Modelling con  
i Big Data Social

La Disinformazione & le  
Echo Chamber  
QCPS-II

La Qualità  
della Comuni-

 Jamieson, K. H. and Cappella, J. N. (2008).

*Echo chamber: Rush Limbaugh and the conservative media establishment.*

Oxford University Press.

 Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005).

Graphs over time: Densification laws, shrinking diameters and possible explanations.

*In Proc. 11th Int. Conf. ACM SIGKDD*, pages 177–187, Chicago, IL, USA.