

SUPSI

Comprensione del comportamento scorretto nella discussione di COVID19 sui social network

Studente/i	Relatore	Correlatore
Nogara Gianluca	Giordano Cremonese Silvia	Luceri Luca
Corso di laurea	Modulo / Codice Progetto	Anno
Ingegneria Informatica TP	C10344 Progetto di diploma	2020/2021
Committente	Data	
-	16/09/2021	

Indice

- Introduzione
- Contestualizzazione del problema
- Obiettivi
- Stato dell'arte
- Approccio al problema
- Risultati ottenuti
- Problemi riscontrati
- Sviluppi futuri
- Conclusione

Introduzione



- La pandemia da Covid-19 ha fatto crescere l'utilizzo dei social network
- La natura di Twitter fa sì che abbia giocato il ruolo di prima piattaforma di informazione
- Questo ha alzato l'attenzione sulle informazioni condivise

Contestualizzazione del problema

- L'infodemia e la proliferazione delle fake news ha raggiunto livelli preoccupanti, mettendo a rischio la salute delle persone
- È importante sviluppare approcci per contenere e combattere la disinformazione

Informati!

Per assicurarti di avere informazioni ufficiali ed accurate sul COVID-19 e sui vaccini anti COVID-19 consulta il sito del Ministero della Salute.

Ministero della Salute

Informazioni sui vaccini anti COVID-19

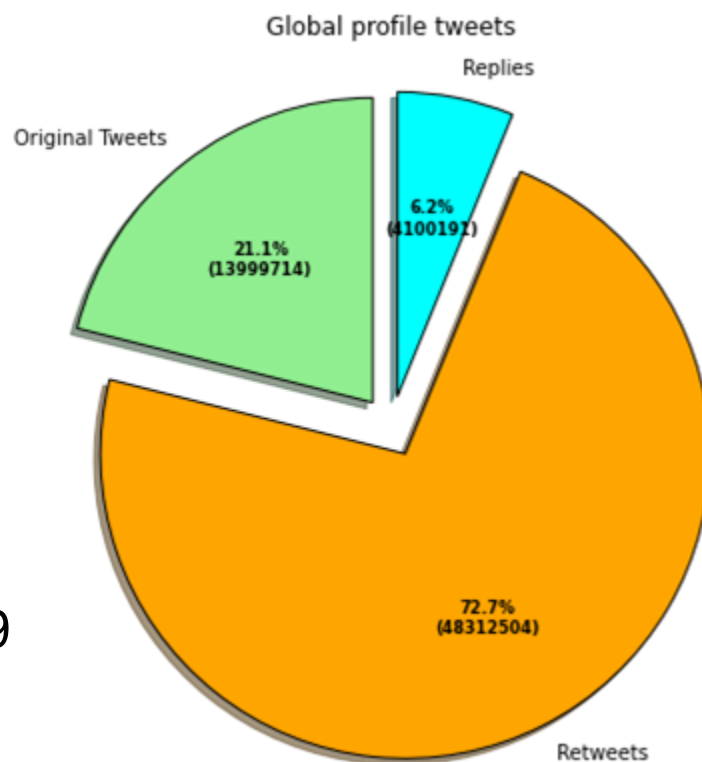
Obiettivi

- Caratterizzare entità/azioni corrette e malevoli nel dibattito Covid-19
- Identificare i mezzi e comportamenti che contraddistinguono le entità
- Studiare hashtags, link condivisi, attività e le altre strategie di condivisione delle informazioni false



Stato dell'arte – Dati a disposizione

- I dati a disposizione sono i tweet che coprono 2020 e 2021
- 54,09 GB di dati
- Dati:
 - 66,412,411 di attività
 - 13,999,715 di tweets
 - 48,312,504 di retweets
 - 4,100,192 di risposte
- Dati ottenuti da un set di keywords relative al Covid19



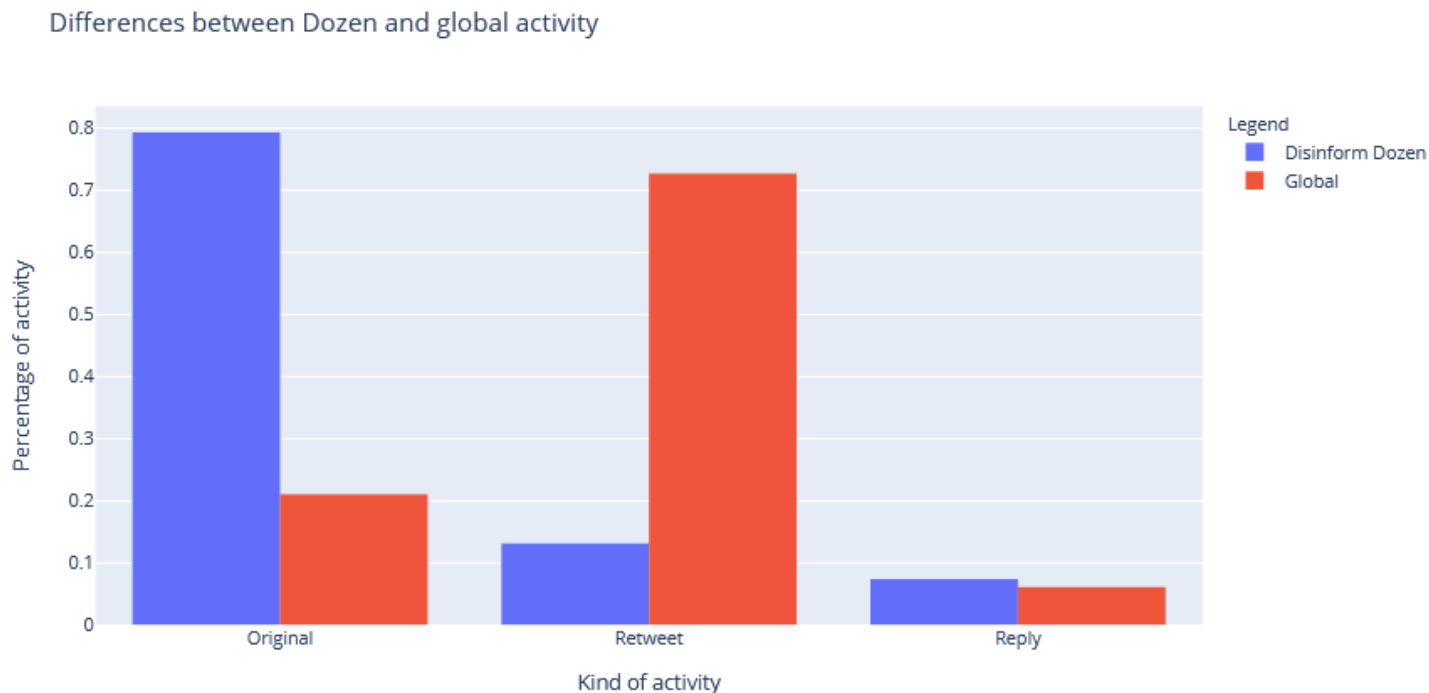
Approccio al problema – Studio dei Disinformation Dozen

- Center for Counter Digital Hate ha stilato una lista di 12 utenti responsabili del 65% dei contenuti anti-vaccini

Nome	Mansione	Stato	Followers
Joseph Mercola	medico osteopata	Attivo	314.302
Robert Kennedy, Jr.	pseudo sostenitore ambientalista	Attivo	311.481
Ty and Charlene Bollinger	attivista della medicina alternativa	Rimosso	-
Sherri Tenpenny	medico osteopata	Sospeso	-
Rizza Islam	attivista e social media influencer novax	Rimosso	-
Rashid Buttar	medico osteopata	Attivo	86.368
Erin Elizabeth	attivista della medicina alternativa	Attivo	39.384
Sayer Ji	attivista della medicina alternativa	Attivo	11.285
Kelly Brogan	attivista della medicina alternativa	Attivo	18.618
Christiane Northrup	ostetrica e ginecologia	Attivo	115.215
Ben Tapper	chiropratico	Sospeso	-
Kevin Jenkins	CEO di un gruppo novax	Attivo	897

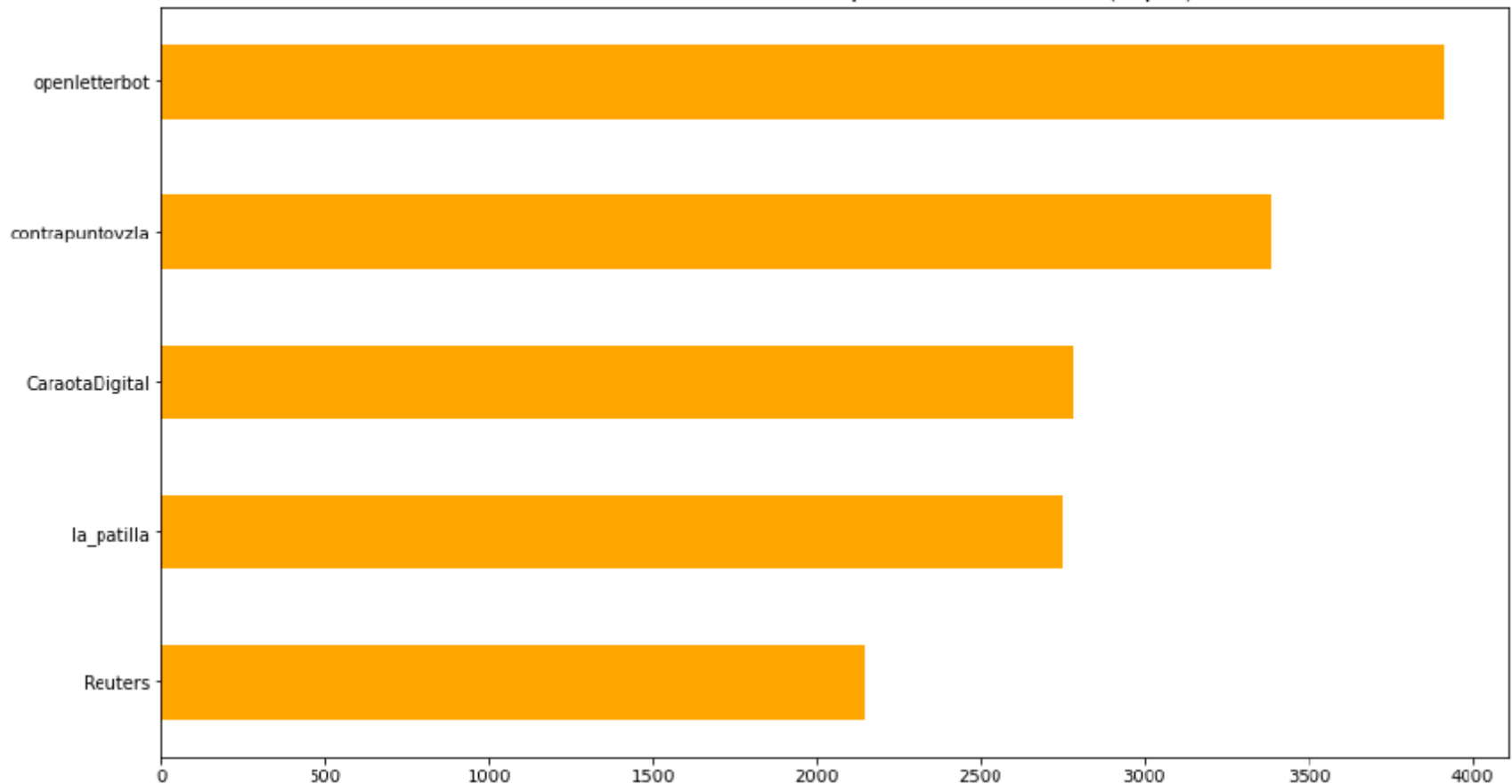
Risultati – Comportamento e strategie dei Disinformation Dozen

- Strategia di informazione orientata al tweet



Approccio al problema – Utenti verificati

Number of ORIGINAL tweets per VERIFIED users (top 5)



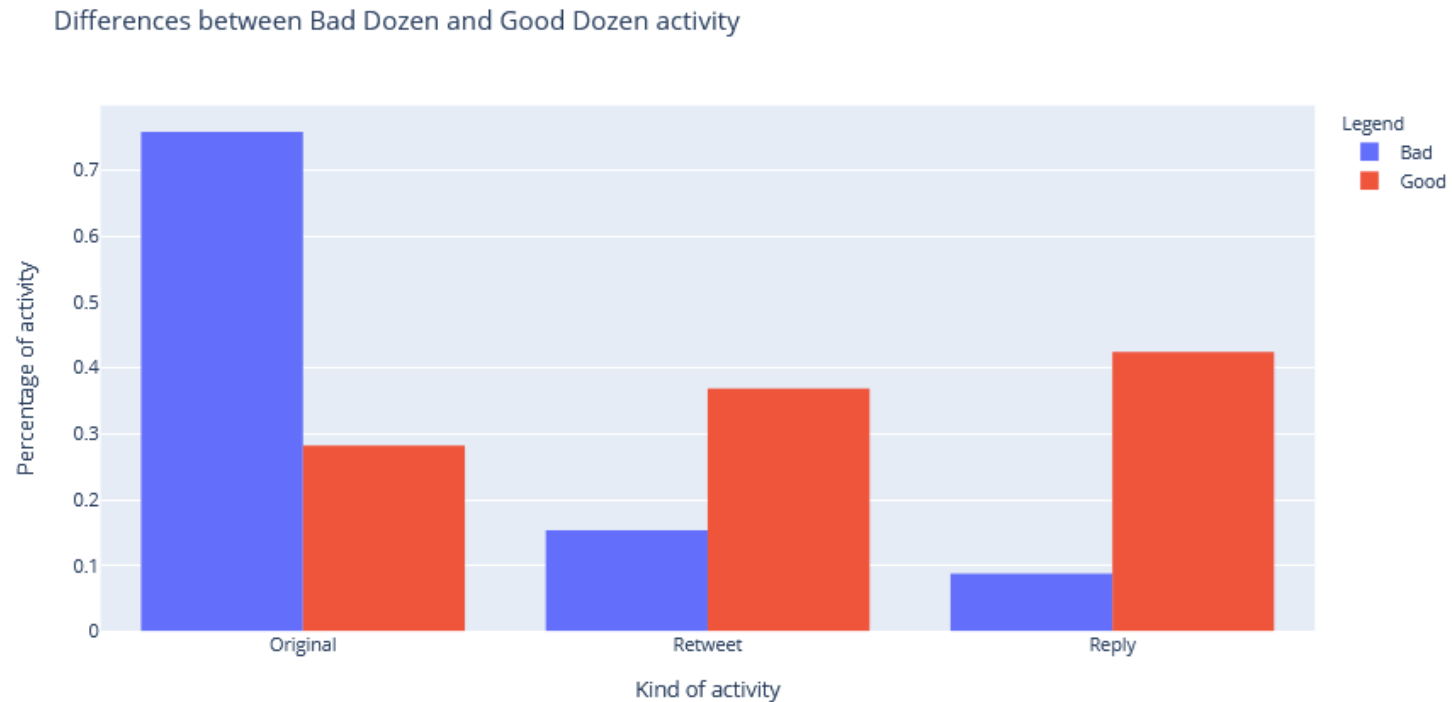
Approccio al problema – Studio dei Good Dozen

- Da uno studio di Jonathan Oppenheim sono stati ottenuti attori che fanno corretta informazione

Nome	Mansione	Stato	Followers
Dena Grayson	medico, ricercatore ed esperta di Ebola	Attivo	326.532
Ian M. Mackay	virologo	Attivo	120.086
Eric Feigl-Ding	epidemiologo ed economista sanitario	Attivo	588.577
Ilona Kickbusch	fondatrice del Global Health Centre	Attivo	30.435
Ashish Jha	medico e ricercatore	Attivo	264.490
Helen Branswell	reporter di malattie infettive e salute globale	Attivo	216.153
Marc Lipsitch	epidemiologo e microbiologo di malattie infettive	Attivo	234.412
Trevor Bedford	ricercatore in virus, evoluzione e immunità	Attivo	328.185
Kai Kupferschmidt	giornalista scientifico e biologo molecolare	Attivo	134.596
Ed Yong	scrittore scientifico per The Atlantic	Attivo	354.278
Rochelle Walensky	direttrice del CDC del governo degli USA	Attivo	332.234
Tedros Ghebreyesus	direttore generale del WHO	Attivo	1.530.289

Risultati – Confronto tra Bad e Good Dozen

- Strategie di condivisione differenti



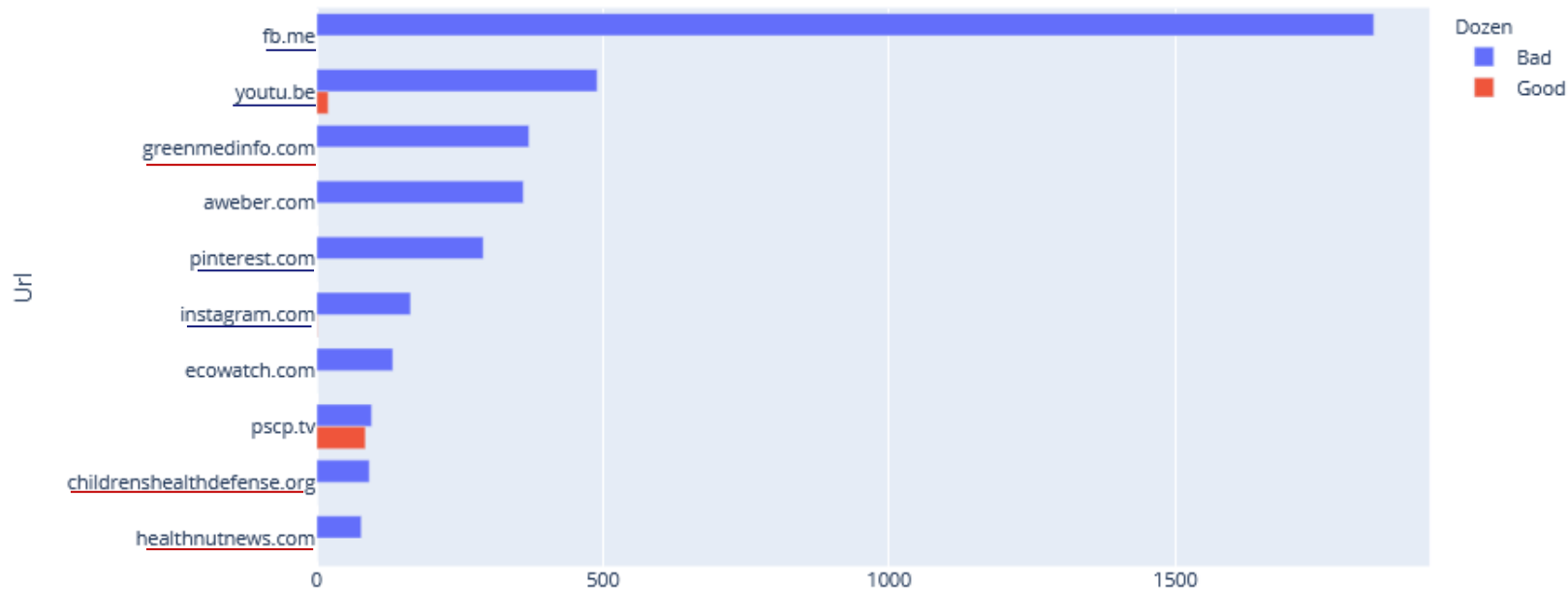
Risultati – Conversation threading



Risultati – Domini condivisi dai Bad Dozen

- Condivisione di domini di loro proprietà e credibilità dubbia

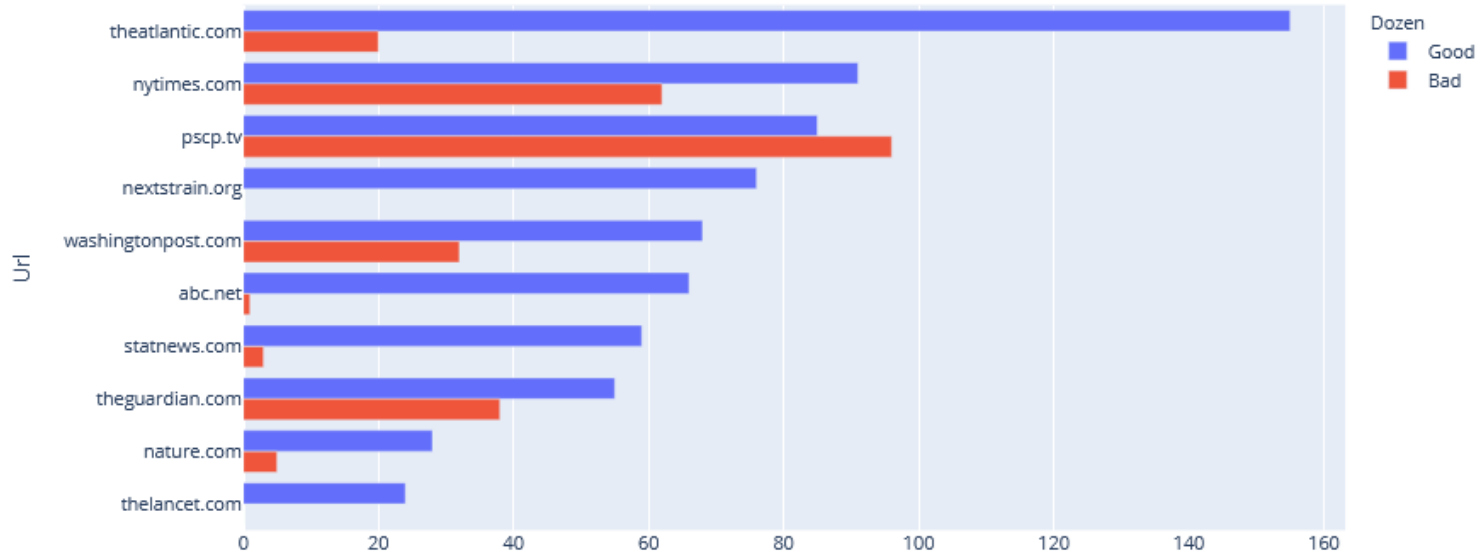
Comparison Bad and Good Dozen link shared (From top 10 Bad Dozen)



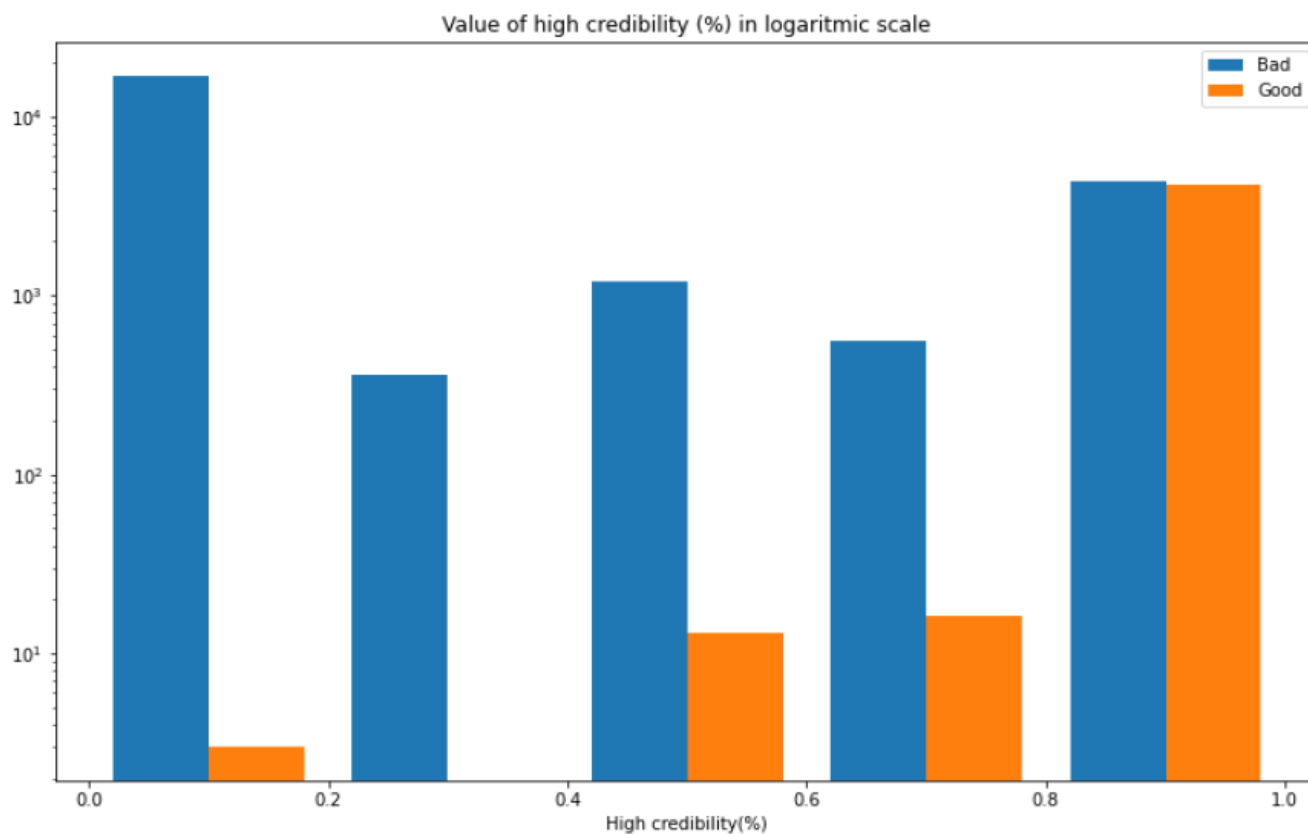
Risultati – Domini condivisi dai Good Dozen

- Meno domini condivisi ma con maggiore credibilità

Comparison Bad and Good Dozen link shared (From top 10 Good Dozen)

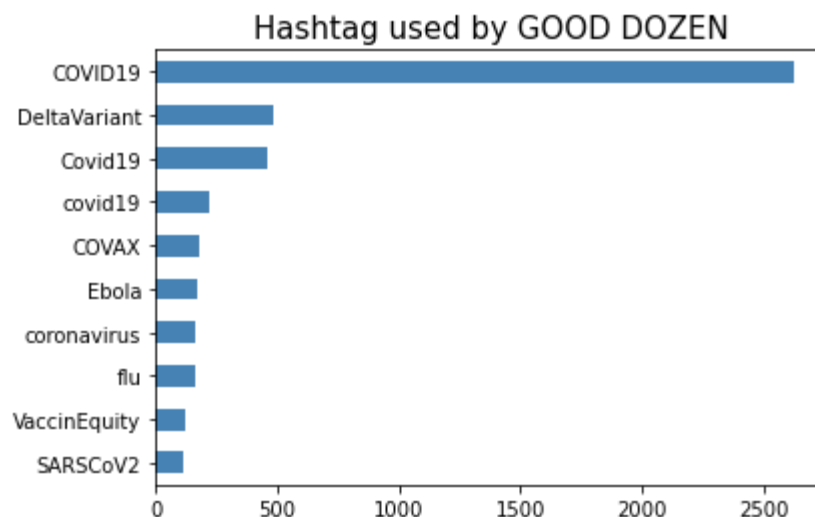
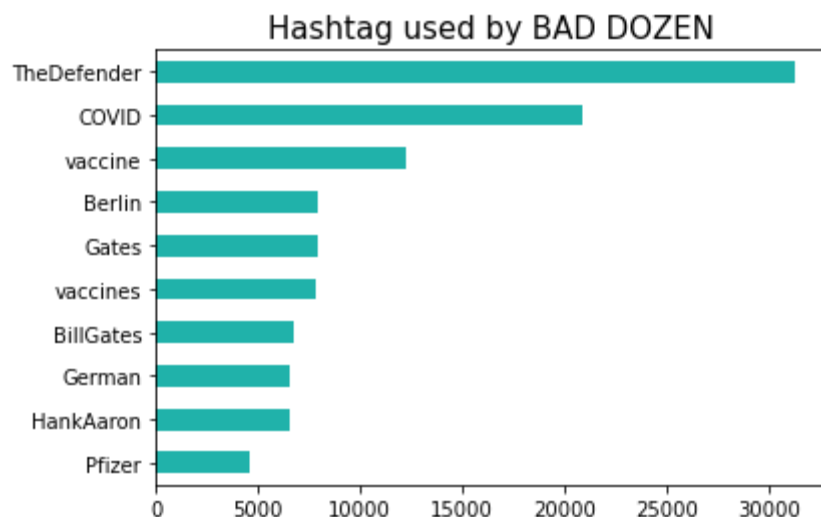


Risultati – Confronto credibilità sui domini condivisi dai seguaci dei Dozen



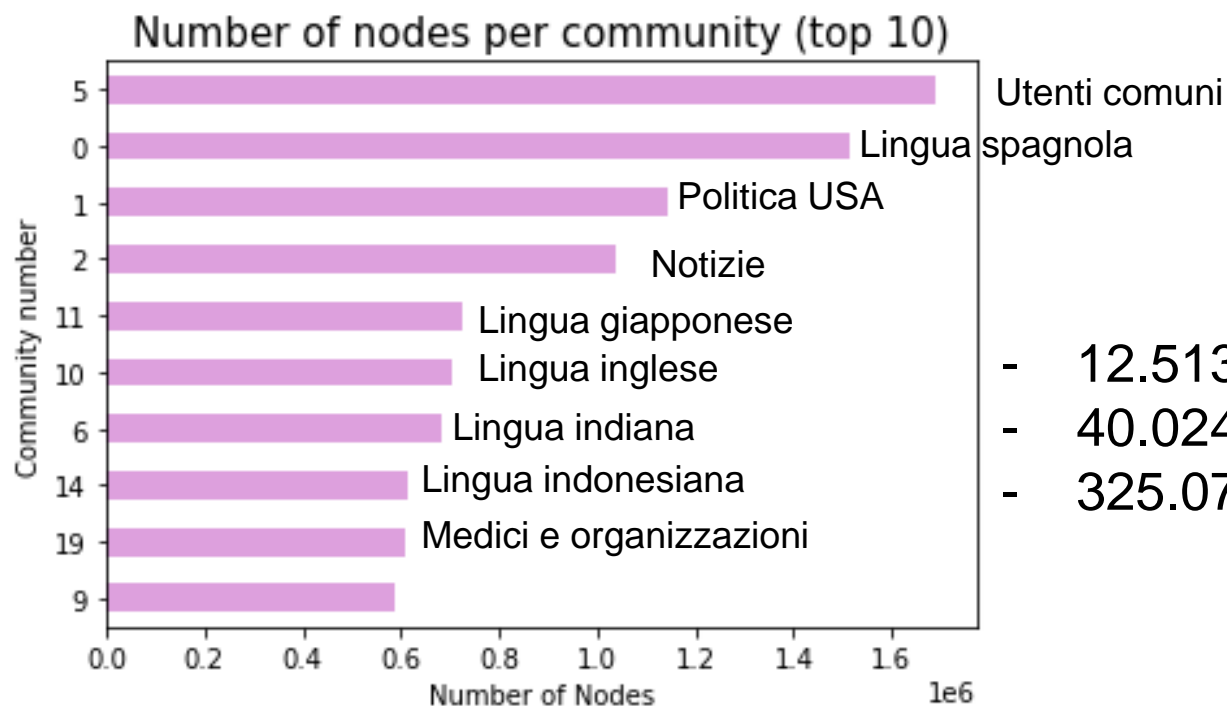
Risultati – Confronto hashtags

- La differenza degli hashtags usati indica una strategia completamente diversa nelle tematiche



Approccio al problema – Retweet Network

- Il retweet è lo strumento più significativo per mettere in relazione due utenti



- 12.513.502 nodi
- 40.024.282 archi
- 325.073 communities

Risultati – Network communities

Nome	Community	Dozen
DrDenaGrayson	1	Good
ashishkjha	1	Good
edyong209	1	Good
HelenBranswell	1	Good
trvrb	2	Good
CDCDirector	2	Good
MackayIM	2	Good
IlonaKickbusch	2	Good
DrEricDing	2	Good
kakape	2	Good
mlipsitch	2	Good
unhealthytruth	2	Bad
BusyDrT	2	Bad
RobertKennedyJr	2	Bad
mercola	2	Bad
kevdjenkins1	2	Bad
DrChrisNorthrup	2	Bad
DrButtar	2	Bad
DrTedros	19	Good

Problemi riscontrati

- Limitazioni hardware
- Limitazioni sulla chiamate Botometer
- Problemi nello scoring degli utenti
- Tempi di processamento lunghi



Sviluppi futuri

- Integrazione di dati più recenti
- Studio più approfondito di subnetworks
- Studio degli utenti sospesi e rimossi
- Timeline dell'informazione: come le persone sono state influenzate dalla discussione



Conclusione

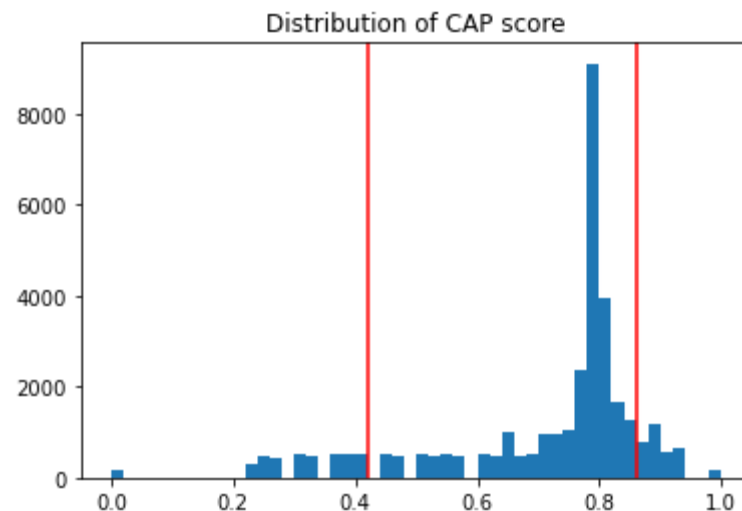
- Definizione di due fronti: «Good» e «Bad» Dozen
- Strategie di condivisione differenti
 - Hashtag
 - Tipologia di attività
- Strategie di informazione differenti
 - Domini di proprietà
 - Social network
- Appartenenza dello stesso gruppo nella retweet network

Domande



Approccio al problema – Identificazione dei bot

- Raccolta degli utenti più attivi per categoria di attività
 - original tweet
 - retweet
 - passive retweet
 - reply
 - passive reply
- Utilizzo dei decili per ottenere umani e bot



Risultati – Classificazione dei bot

- Classificazione dei bot in base alle sottocategorie

- **fake_follower**: bot acquistati
- **self_declared**: bot da botwiki.org
- **astroturf**: bot politici
- **spammer**: spambot
- **financial**: bot che usando i cashtag
- **other**: da annotazioni manuali, feedback degli utenti

