Scuola universitaria professionale della Svizzera italiana
Dipartimento tecnologie innovative

SUPSI

Comprensione del comportamento scorretto nella discussione di COVID19 sui social network e sui media online

Studente/i
Nogara Gianluca

Relatore
Giordano Cremonese Silvia

Correlatore
Luceri Luca

Committente
-

Corso di laurea
Ingegneria Informatica TP

Modulo
C10344 Progetto di diploma

Anno
2020/2021

Data
16/09/2021

Indice

Indice delle figure

Indice delle tabelle

Abstract

I social network online (OSN) sono tecnologie basate su computer che consentono agli utenti di creare contenuti e intrattenere relazioni sociali.

Le OSN sono rapidamente cresciute dall'essere semplicemente un canale di aggregazione all'essere un fenomeno globale che ha suscitato un cambio di paradigma della nostra società, trasformando il modo in cui le persone accedono alle notizie, condividono opinioni, fanno affari e fanno politica.
In uno scenario del genere, l'accuratezza, la veridicità e l'autenticità del contenuto condiviso sono ingredienti necessari per mantenere una sana discussione online.

Tuttavia, negli ultimi tempi, le OSN hanno dovuto affrontare una notevole crescita di account e attività dannose, che minano intenzionalmente l'integrità delle conversazioni online condividendo, nelle piattaforme OSN, informazioni false e provocatorie per influenzare l'opinione pubblica e creare conflitti su questioni sociali o politiche.

Ciò è stato, e continua ad essere, drammaticamente vero per la discussione sul COVID19, che è stata minata da campagne di manipolazione e disinformazione a livello globale. È quindi necessario studiare i comportamenti peculiari delle entità dannose nel dibattito COVID19 per fornire un cambiamento sociale verso una migliore comprensione (alfabetizzazione) delle reti sociali.

Questo include il rilevamento di malintenzionati, l'identificazione di informazioni di scarsa credibilità e notizie false, la consapevolezza degli utenti e il controllo dei comportamenti scorretti.

Online social networks (OSNs) are computer-based technologies that allow users to create content and engage in social relationships.

OSNs have rapidly grown from being simply a channel of aggregation to being a global phenomenon that has sparked a paradigm shift in our society, transforming the way people access news, share opinions, do business, and make policy.
In such a scenario, accuracy, truthfulness, and authenticity of shared content are necessary ingredients to maintain a healthy online discussion.

However, in recent times, OSNs have faced a significant growth of malicious accounts and activities, which intentionally undermine the integrity of online conversations by sharing, in OSN platforms, false and provocative information to influence public opinion and create conflict on social or political issues.

This has been, and continues to be, dramatically true of the COVID19 discussion, which has been undermined by global campaigns of manipulation and misinformation. Therefore, it is necessary to study the peculiar behaviors of malicious entities in the COVID19 debate to provide social change towards a better understanding (literacy) of social networks.

This includes the detection of malicious actors, the identification of information of low credibility and fake news and the users' awareness and control of misbehaviors.

Progetto assegnato

Lo scopo di questo progetto è quello di analizzare entità malevoli presenti in un dataset pubblico su COVID19. L' OSN di riferimento è Twitter e l'attività di analisi e classificazione delle entità malevoli (come bot e trolls) viene fatta tramite diversi algoritmi esistenti utilizzando Data Science e Network Science.
Il progetto in questione consente quindi di analizzare il comportamento all'interno della discussione e l'impatto che hanno tali entità malevole nella discussione globale.

COMPITI

- Identificazione degli strumenti usati per la propagazione di notizie
- Identificazione e classificazione delle entità nella discussione
- Analisi comportamentale delle entità malevoli e confronto con entità competenti

OBIETTIVI

L'obiettivo principale di questa tesi è quello di arrivare a distinguere le strategie e l'impatto di entità/azioni corrette e quelle malevoli nel dibattito COVID19.

Identificare quindi tramite quali mezzi e comportamenti le varie entità malevoli influiscono nelle discussioni, analizzando hashtags, link condivisi e attività, facendo una comparazione tra chi ha l'obiettivo di informare correttamente e chi intenzionalmente diffonde informazioni non veritiere.

TECNOLOGIE

- Python
- Complex networks

1 Analisi

1.1 Contesto

Il progetto si inserisce nel contesto del Dipartimento Tecnologie Innovative della SUPSI e in particolare nel corso di Ingegneria Informatica. Si tratta di un progetto interno alla SUPSI finanziato dal fondo nazionale svizzero (SNSF) ed intitolato "Detecting Troll Activity in Online Social Networks", di grande importanza sociale dal momento che le informazioni raccolte possono essere utilizzate per limitare la propagazione di entità malevoli sui social network e sensibilizzare su un tema tanto delicato e attuale.

1.2 Fruitori

Il progetto in questione è destinato sia ad un pubblico scientifico interessato al dibattito, in particolare su come le varie entità interagiscono sulle OSN nella discussione sul Covid-19, sia ad un pubblico più ampio, interessato a sapere di più sull'attendibilità delle informazioni che circolano sulle OSN relativa al Covid-19.
La finalità scientifica del progetto lo rende pubblicabile ed espandibile per nuovi studi di vario genere e operazioni di aggiornamento su campagne di misinformazione e detezione di comportamenti scorretti su OSN.

1.3 Requisiti

1.3.1 Contenuti

Il principale requisito a livello contenutisco è quello di avere dei costanti feedback visivi dell’analisi effettuata sui dati, in modo da effettuare operazioni di ricerca sempre più dettagliata in funzione di risultati.
Una volta identificate le strategie di condivisione delle informazioni è quindi possibile andare a stabilire di quali tipologie di utenti si tratta e definire quindi nuove metriche di valutazione per la credibilità di tali utenti.
Identificando utenti malevoli è inoltre possibile effettuare studi sulle loro interazioni attive e passive, definendo delle community potenzialmente utili per ricavare ulteriori

informazioni utili per l'identificazione di comportamenti coordinati per la diffusione di informazioni fuorvianti e/o non veritiere.

1.3.2 Rispetto della privacy

Per lo svolgimento del progetto sono state utilizzate grosse quantità di dati pubblici ottenuti tramite un repository che contiene l'id dei tweets coinvolti nella discussione sul COVID19 divisi per mese.[1]
I dati pubblicati nel repository in questione sono solo id (per questioni di policy di Twitter), pertanto non contengono informazioni ulteriori sugli utenti e/o sui tweets.

Dal momento che i dati così non avrebbero alcuna informazione, per avere accesso a informazioni complete è stato eseguito un processo di idratazione, che consiste nel richiedere a Twitter, via API, di avere i tweets completi a partire da una serie di tweets id.

Poiché i dati in questione sono informazioni pubbliche su una piattaforma aperta che possono essere visionati anche senza registrazione su Twitter non ci sono problemi legati alla privacy nella condivisione di tali. Degli esempi di dati possono essere visualizzati nelle figure sottostanti (Figura 1 e Figura 2).

user_screen_name	rt_user_screen_name	hashtags	urls
Huerconetzin	AnneKPIX	[{'text': 'coronavirus', 'indices': [94, 106]}]	[]
JustAnotherAme4	cnni	[]	[]
HHSRegion8	CDCgov	[{'text': 'coronavirus', 'indices': [47, 59]}]	[]
Paxman42	OurWarOnCancer	[{'text': 'lungcancer', 'indices': [103, 114]}]	[]
beerhowell	nytimes	[]	[]
Gambiste1	StocksUnhinged	[{'text': 'CDC', 'indices': [89, 93]}, {'text'...	[]
IAmTonyaNash	CDCgov	[{'text': 'publichealth', 'indices': [30, 43]}]	[{'url': 'https://t.co/8kzRD1IADl', 'expanded_...
AndiUgo	NaN	[]	[{'url': 'https://t.co/2u7F1xZZYM', 'expanded_...
_ayychris	PMBreakingNews	[]	[]
ParkvilleMOm	Reuters	[]	[]

Figura 1: Struttura del DataFrame dei tweets filtrata su poche colonne

[1] https://github.com/echen102/COVID-19-TweetIDs

Le colonne mostrate nella Figura 1 indicano:

- **user_screen_name**: l'username dell'utente che effettua l'azione (tweet, retweet o risposta)
- **rt_user_screen_name**: l'username dell'utente che ricevere il retweet
- **hashtags:** insieme di hastags usati nel tweet
- **urls:** domini condivisi nel tweet

name	screen_name	location	description	followers_count	verified
Anissa Primafidyanti	prmfdynt	INA	NaN	615	False
Alma	allmitaquino	NaN	adicta a las series y la fotografía	39	False
Gregory Tribiani	GregoryTribiani	NaN	Don't kill our kids 😷	84	False
Nicole 🐈🐈	DenisseeNLB	NaN	Sin limites, sin miedo, pasion y amor♥ Sagitar...	669	False
Alex Scott Samuel	ASAMUELFEATURE	North Carolina, USA	COMING: My Collection of 13 Southern Horror St...	211	False
-NEW ME-	AmourVirus	มักกะสัน, ประเทศไทย	Girls' Generation | Jessica Jung | Krystal | Y...	202	False
🖤	d_rek02	NaN	Why so serious? | they/them	156	False
Madhyamam	madhyamam	Kozhikode	Madhyamam is India's first international newsp...	31765	False
マリちゃん🐬	Lomo_kanit	bangkok	การเรียนรู้จบสิ้นเมื่อเราได้จากโลกนี้ไปตลอดกาล...	331	False
constitutionalistmln	mlnconstitution	Oregon, USA	God first, Pro Life, Defy socialism, law & ord...	88	False

Figura 2: Struttura del DataFrame dei profili filtrata su poche colonne

Le colonne mostrate nella Figura 2 indicano:

- **name**: l'username dell'utente che effettua l'azione (tweet, retweet o risposta)
- **screen_name**: l'username dell'utente che ricevere il retweet
- **location**: posizione impostata dall'utente
- **description:** breve descrizione che un utente dà al proprio profilo
- **followers_count:** domini condivisi nel tweet
- **verified:** flag che indica se il profilo sia verificato o meno

1.4 Obiettivi

Gli obiettivi di questo progetto sono i seguenti:

- Classificazione degli utenti/entità malevoli
- Investigazioni sulle attività degli utenti/entità malevoli
- Investigazioni sulle attività degli utenti/enità attendibili
- Investigazione sulle strategie di condivisione di utenti/entità malevole
- Studio degli utenti che interagiscono con tali utenti/entità
- Studio della credibilità dei domini condivisi dagli utenti/entità, effettuando una classificazione binaria per la credibilità del dominio condiviso (“low” o “high”)
- Comparazione delle attività tra utenti/entità malevoli e attendibili

2 Stato dell'arte

2.1 Analisi dei dati

Come situazione di partenza è stato messo a disposizione un piccolo dataset di 50,000 tweets.

I dataset in questione riguardano i profili utente e i singoli tweet, la struttura dei due dataset è completamente differente, in particolare quello dedicato agli utenti è molto ricco di colonne (ben 46), mentre quello dedicato agli utenti ne ha poco meno della metà (18).
Questa differenza è data dal fatto che, per quanto riguarda i tweet, abbiamo a disposizione molte più informazioni rispetto a dei semplici profili utente.
Di seguito sono mostrate le colonne più utilizzate dei due dataset:

Tweets

Campo	Tipologia	Descrizione
id	int64	Id univoco del tweet
created_at	object	Data di creazione del tweet
user_id	int64	Id univoco del profilo
user_screen_name	object	Nome utente del profilo
text	object	Contenuto del tweet
retweet_count	int64	Numero di retweet del tweet
in_reply_to_user_id	float64	Id dell'utente che riceve la risposta
in_reply_to_screen_name	object	Username dell'utente che riceve la risposta
rt_user_id	float64	Id dell'utente che viene il retwittato
rt_user_screen_name	object	Username dell'utente che viene retwittato
hashtags	object	Hashtag usati nel tweet
urls	object	Link condivisi nel tweet

Tabella 1: Colonne più importanti del DataFrame dei tweets

Profili

Campo	Tipologia	Descrizione
id	int64	Id univoco dell'utente
Screen_name	object	Username dell'utente
location	object	Località dell'utente
follower_count	int64	Numero di follower dell'utente
verified	bool	Booleano che indica se il profilo è verificato
geo_enabled	bool	Booleano che indica se il profilo mostra la posizione quando twitta

Tabella 2: Colonne più importanti del DataFrame degli utenti

I dati da analizzare risultano eterogenei, sono presenti infatti valori booleani, valori interi, oggetti e valori in virgola mobile, è quindi importante effettuare un preprocessamento per un corretto studio.

2.2 Linguaggi di programmazione e ambiente di sviluppo

Dal momento che i dati su cui lavorare sono file con formato .csv di grossa dimensione, si parla di cinque file da 54,09Gb in totale, la scelta del linguaggio è ricaduta su Python, questo anche perché fornisce una grande varietà di librerie e tools per data analisi e complex networks.

Anaconda ha svolto un compito molto importante per quanto riguarda la programmazione in Python e la realizzazione di feedback visivi; infatti, dal momento che per un'analisi esauriente si ha bisogno di generare report grafici, Anaconda contiene il software Jupyter Notebook.
Quest'ultimo è un'applicazione Web open source che rappresenta lo strumento perfetto per effettuare operazioni di data analisi grazie alle celle che consentono di salvare rappresentazioni grafiche e testo senza dover rieseguire necessariamente tutto il codice.

2.3 Librerie

Durante lo sviluppo del progetto si è fatto uso di una serie di librerie che hanno semplificato lo svolgimento del progetto, le più importanti hanno dato un contributo significativo e sono state utilizzate con una certa frequenza.

2.3.1 Pandas

Pandas è tra le librerie più importanti in Python dal momento che è stato sviluppato per la manipolazione e l'analisi dei dati.
In particolare, offre strutture dati e operazioni per manipolare tabelle numeriche e serie temporali.

L'utilizzo di Pandas ha consentito lo sfruttamento dei DataFrame per la lettura dei file .csv e la gestione dei relativi dati contenuti.
I DataFrame hanno rappresentato il cuore pulsante del lavoro svolto dal momento che sono altamente gestibili e semplici nel filtraggio.

Lo sfruttamento della libreria ha permesso inoltre l'utilizzo delle Series, ndarray monodimensionali che espongono un'infinità di metodi per la manipolazione dei dati. Questa tipologia di strutture è stata largamente utilizzata per effettuare classificazioni basate sul numero di occorrenze.

2.3.2 Matplotlib e Plotly

Matplotlib è una libreria per la creazione di grafici ed è stata largamente utilizzata per la rappresentazione grafica di diverse tipologie di classificazioni, quali istogrammi, bar plot e distribuzioni varie. In Figura 3 possiamo vederne un esempio.

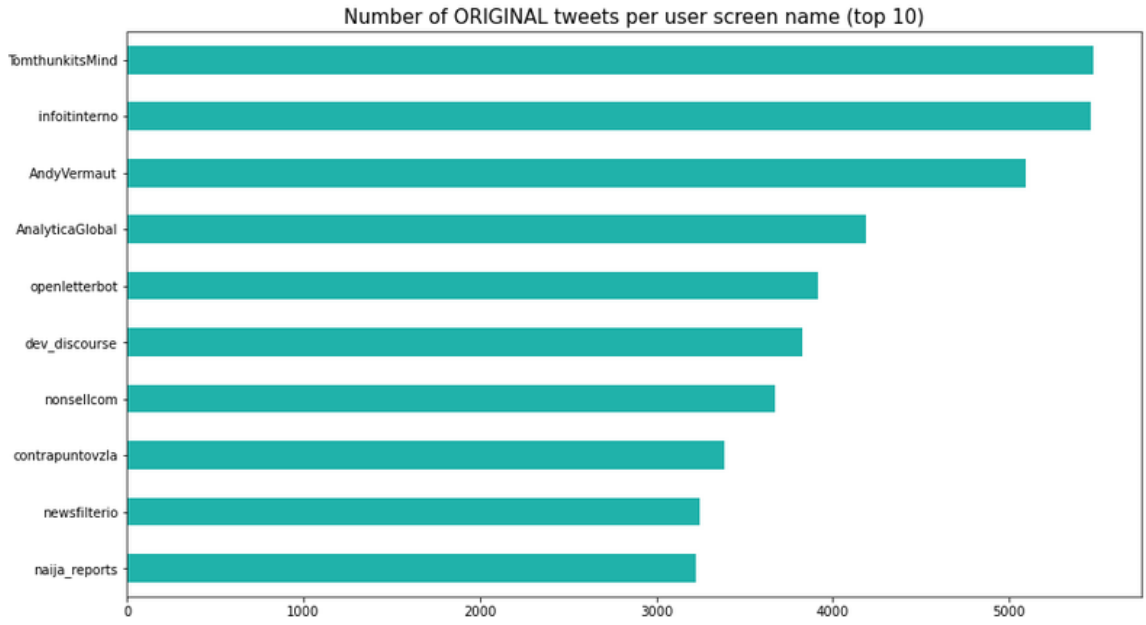


Figura 3: Esempio di plot con Matplotlib (10 utenti che effettuano più tweet)

Dal momento che i plot generati sono statici essi risultano molto veloci da generare e leggeri da salvare, anche per quantità molto grosse di dati.

In contrapposizione vi è Plotly, libreria di grafici che viene creare grafici interattivi browser-based, usata per generare report dinamici ma che, in presenza di una quantità di dati significativa, risulta poco efficiente e pesante. Un esempio di mappa interattiva realizzata con Plotly, basata sul sample di tweets di utenti italiani, è in Figura 4, qui riportata.

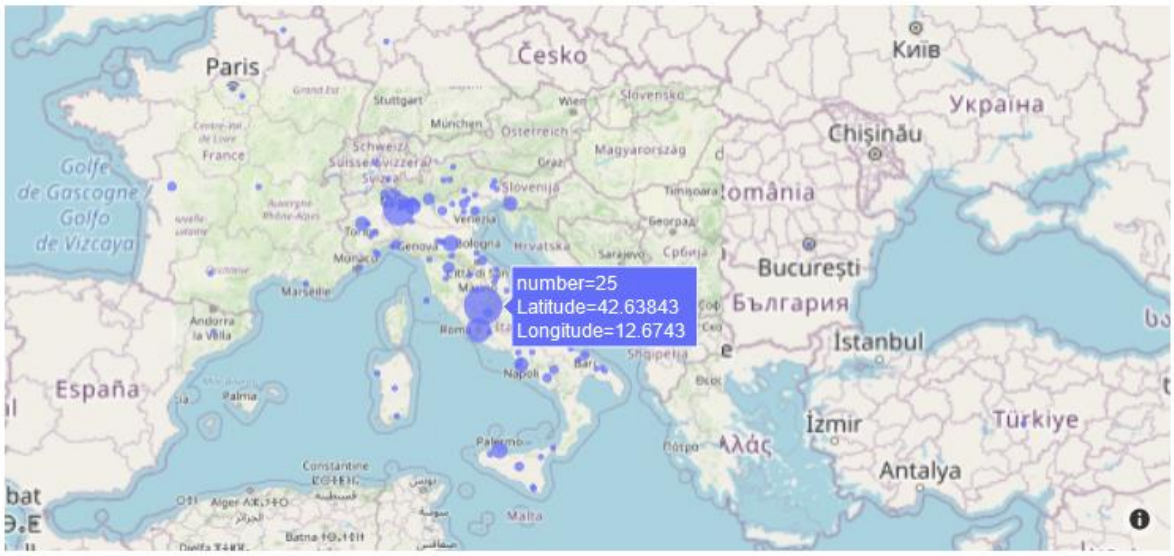


Figura 4: Esempio di mappa interattiva generata con Plotly

2.3.3 NetworkX

NetworkX è una libreria realizzata per semplificare la generazione e lo studio di grafi e reti, è stata largamente utilizzata per raggruppare gli utenti in community e capire quindi i legami che hanno l'un l'altro.

NetworkX è una libreria adatta per il funzionamento su dati molto consistenti: ad esempio, grafici con più di 10 milioni di nodi e 100 milioni di archi, dal momento che è realizzata in puro Python, infatti è ragionevolmente efficiente, molto scalabile e altamente portabile per l'analisi di reti e social network.

2.4 Strumenti utilizzati

Durante lo svolgimento del progetto vi è stata la necessità di utilizzare strumenti realizzati da terzi per poter dare una validità ai dati generati o ottenere determinate informazioni.

2.4.1 Botometer

Botometer è un algoritmo di machine learning addestrato per calcolare un punteggio ad un utente, più il punteggio è basso e più è probabile che l'account sia un umano, analogamente i punteggi alti indicano probabili account bot.
Per calcolare il punteggio, Botometer confronta un account con decine di migliaia di esempi etichettati passando per la Twitter API.

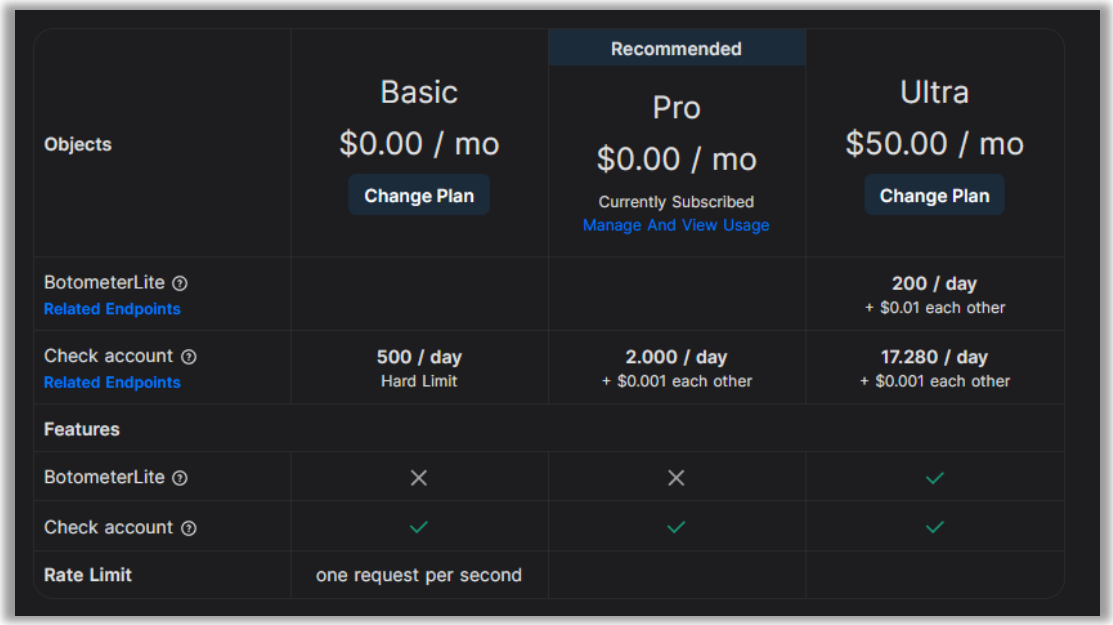
La risposta che Botometer fornisce non ci dà un singolo valore ma una serie di informazioni suddivise nella lingua principale che viene usata da quel profilo e in inglese. Le informazioni che l'API restituisce riguardano anche dei valori per poter identificare la tipologia di bot che ci troviamo di fronte.

Si tratta di un progetto del gruppo OsoMe (Observatory on Social Media) sviluppato all'Indiana University.
OSoMe è una collaborazione tra il Network Science Institute (IUNI), il Center for Complex Networks and Systems Research (CNetS) e la Media School dell'Indiana University.

Botometer è facilmente fruibile tramite l'apposito sito web,[2] ma espone anche un'API pubblica.
Tutte le informazioni per l'integrazione di Botometer sono disponibili sulla pagina GitHub del progetto.[3]

Per poter utilizzare l'API è necessario passare per RapidAPI e da Twitter.
RapidAPI è l'hub di API più grande al mondo, registrandosi gratuitamente è possibile accedere a Botometer Pro e sottoscrivere il piano adatto per il lavoro tra quelli disponibili, mostrati nella Figura 5.

[2] https://botometer.osome.iu.edu/
[3] https://github.com/IUNetSci/botometer-python



Objects	Basic $0.00 / mo Change Plan	Recommended Pro $0.00 / mo Currently Subscribed Manage And View Usage	Ultra $50.00 / mo Change Plan
BotometerLite Related Endpoints			200 / day + $0.01 each other
Check account Related Endpoints	500 / day Hard Limit	2.000 / day + $0.001 each other	17.280 / day + $0.001 each other
Features			
BotometerLite	✗	✗	✓
Check account	✓	✓	✓
Rate Limit	one request per second		

Figura 5: Piani offerti da Botometer Pro su RapidAPI

Una volta sottoscritto il piano si ottengono due key, rispettivamente una key_rapidapi e una consumer_key, per poter accedere appunto all'API.

Una volta ottenuti i dati da RapidAPI è necessario accedere alla Twitter API e ottenere i permessi da sviluppatore e realizzare quindi una Twitter app.
Anche in questo caso si otterranno delle keys: consumer_secret, bearer_token, access_token e access_token_secret.
Tutte queste keys sono essenziali dal momento che vengono utilizzate da Botometer per poter effettuare lo score sugli utenti passati.

È possibile utilizzare Botometer tramite un'apposita libreria:

```
import botometer

rapidapi_key = "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
twitter_app_auth = {

    'consumer_key': 'xxxxxxxx',
    'consumer_secret': 'xxxxxxxxxx',
    'access_token': 'xxxxxxxxx',
    'access_token_secret': 'xxxxxxxxxxx',
  }
blt_twitter = botometer.BotometerLite(rapidapi_key=rapidapi_key,
**twitter_app_auth)
screen_name_list = ['example']
blt_scores = blt_twitter.check_accounts_from_screen_names(screen_name_list)
user_id_list = [1234567890]
blt_scores = blt_twitter.check_accounts_from_user_ids(user_id_list)
```

Per quanto riguarda lo svolgimento del progetto si è utilizzato uno script già esistente che richiede in input un file .csv formato da campi id_utente,nome_utente ritornando un file .json contente gli score di Botometer.

2.4.2 I bot

Per quanto riguarda il termine “bot”, nel contesto in questione, fa riferimento ai social bot.
Dal momento che i bot possono avere una vasta gamma di comportamenti diversi, non esiste una definizione universalmente di bot.
Alcuni sono innocui o addirittura utili, tuttavia i bot dannosi possono essere utilizzati per manipolare gli utenti dei social media amplificando la disinformazione, creando l'impressione che alcune persone, idee o prodotti siano più popolari di quanto non siano, usando una strategia di “infodemia” mettendo in circolo una quantità eccessiva di informazioni, talvolta non vagliate con accuratezza, che rendono difficile orientarsi su un determinato argomento per la difficoltà di individuare fonti affidabili..

Un social bot viene definito quindi come un account di social media controllato almeno in parte tramite software.
Le API dei social forniscono metodi per eseguire azioni in modo programmatico, come delle schedule che postano automaticamente un determinato post a una determinata ora, il tutto in modo automatizzato.
La maggior parte dei bot automatizzati effettua operazioni basilari, come una risposta, follow, retweet, in base a un’azione; pertanto, sono facilmente riconoscibili e poco nocivi nelle discussioni.

2.4.3 Media Bias/Fact Check

Media Bias / Fact Check è un sito Web ed un servizio di fact-checking che valuta qualitativamente i media (online, cartacei e radiofonici) in termini di credibilità, verificando le news e le fonti di ogni notizia.

Il sito è fruibile gratuitamente in lingua inglese sulla pagina web dedicata.[4]
Si tratta di un sito utilizzato per verificare la credibilità di un dominio e non solo si occupa infatti di fornire informazioni anche su orientamento politico-religioso ed

4 https://mediabiasfactcheck.com/

un'eventuale manipolazione informativa, essenziale per lo studio di fattibilità delle fonti che vengono condivise sui social.

Nonostante il metodo utilizzato dal sito Web non ha un rigore scientifico, la sua attendibilità è inopinabile. Il sito è una fonte largamente citata per notizie e studi sulla disinformazione, è stato infatti utilizzato dai ricercatori dell'Università del Michigan, e non solo, per tracciare la diffusione delle "fake news" e fonti opinabili sui social media. Media Bias / Fact Check fornisce una grande quantità di informazioni, chiaramente visibili in Figura 6, che possono essere usate per motivare la classificazione.

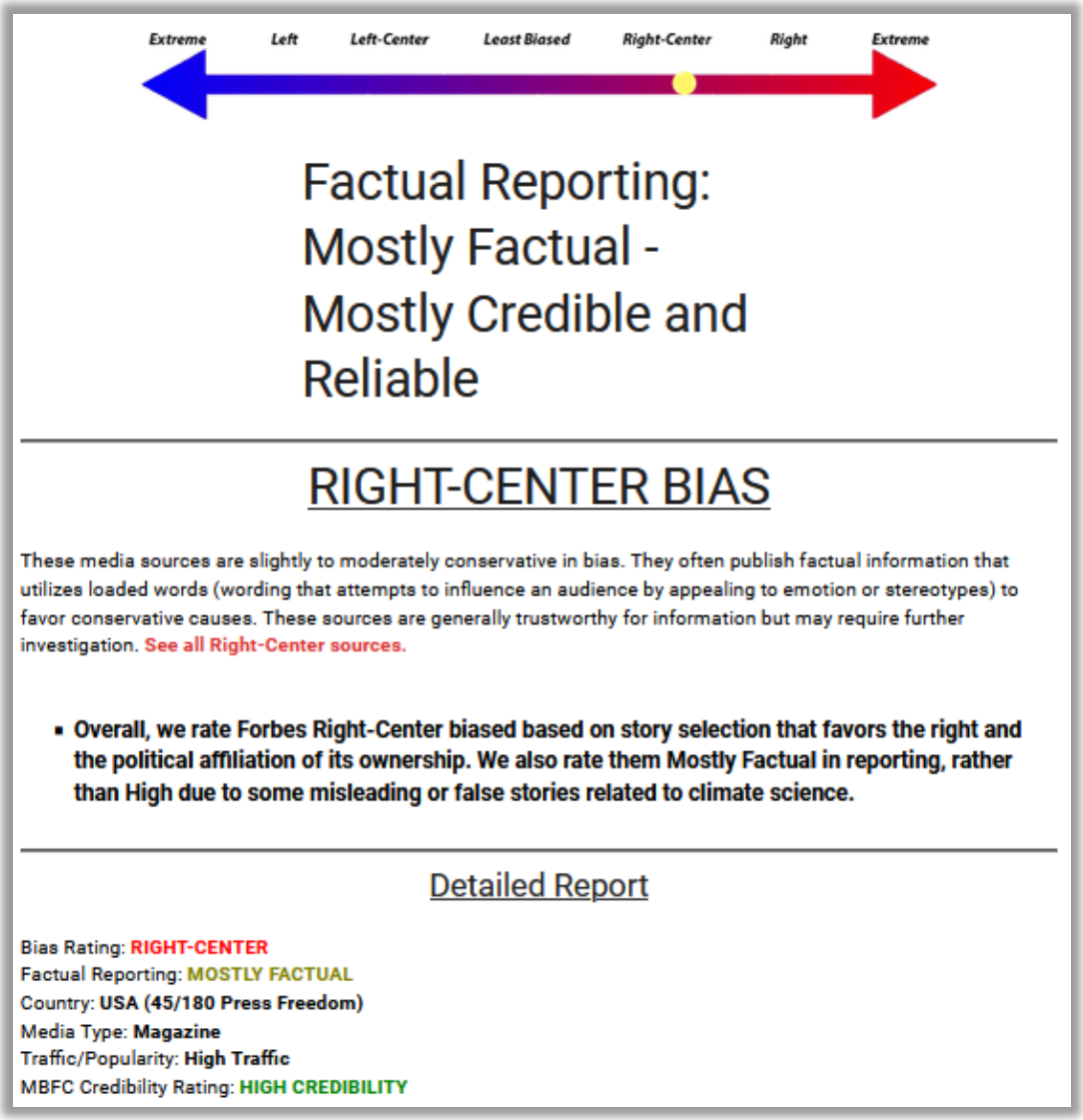


Figura 6: Esempio di classificazione della nota rivista statunitense "Forbes"

Dal momento che il sito non espone API per l'integrazione in Python, sono stati classificati manualmente oltre 800 domini, assegnando ad ogni nome uno score di credibilità ("high" o "low") e una spiegazione della classificazione.

3. Implementazione

3.1 Familiarizzare con i dati

Per poter condurre correttamente uno studio efficace su dati grandi è necessario prendere confidenza con campioni più piccoli con la stessa struttura.
Il primo task svolto per la realizzazione del progetto è stato un breve studio generico su un sample di dati che comprendono entrambi i DataFrame menzionati precedentemente: uno dedicato ai tweets e uno sugli utenti.
che ha riguardato sostanzialmente tutte le colonne di entrambi i DataFrame.
Le prime analisi condotte, seppur molto superficiali, hanno avuto un certo riscontro con analisi precedentemente condotte nel progetto precedentemente menzionato, dal momento che il sample è un subset di un DataFrame molto più grande in analisi.

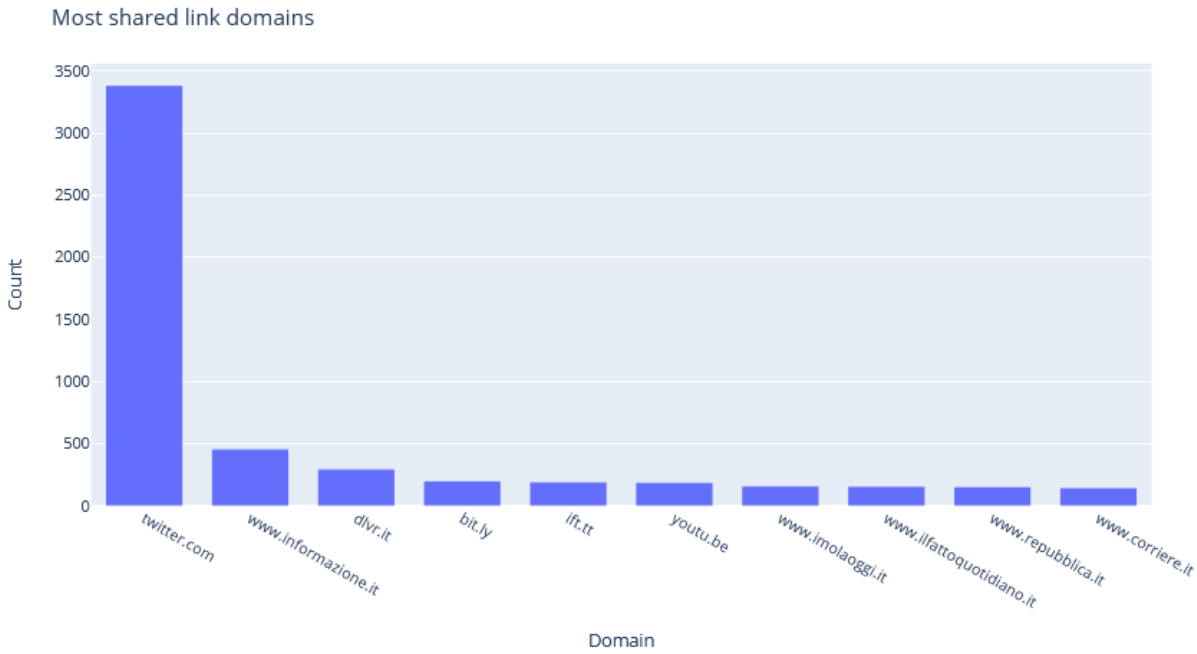


Figura 7: Esempio di analisi sui dati condotta sul sample (Domini più condivisi)

Le operazioni svolte sul set dei tweets sono state le seguenti:

- Divisione del DataFrame in SubDataFrame per tipologia di tweet (retweet, risposte e post originali)
- Plot degli utenti più attivi per tipologia di tweet
- Plot timeline dei tweets
- Plot domini più condivisi

- Ricerca di tweet contenenti una serie di hashtags sospetti (#iononmivaccino, #bigpharma, #nessunacorrelazione, #TuttiComplici e #iononmivaccinerò)

Le operazioni svolte sul set degli utenti sono invece le seguenti:

- Plot timeline della creazione dei profili
- Plot pie chart degli utenti verificati e non
- Plot pie chart degli utenti con una località impostata e non
- Ricerca di potenziali utenti con poche informazioni, come un'immagine di profilo, una descrizione, il profilo verificato, o una location che possano rappresentare potenziali bot.
- Ricerca delle persone con la parola "novax" in descrizione
- Plot di una mappa interattiva sulla località degli utenti

3.2 Dataset

Una volta presa confidenza con i dati campione, si è passati allo studio sul dataset completo. Si tratta di un insieme di dieci file: un file .csv per i tweets e uno per i profili suddivisi mensilmente per cinque mesi. La quantità dei dati è rilevante, in particolare vi sono:

- 66,412,411 di attività
- 13,999,715 di tweets
- 48,312,504 di retweets
- 4,100,192 di risposte

Che possiamo identificare meglio in Figura 8.

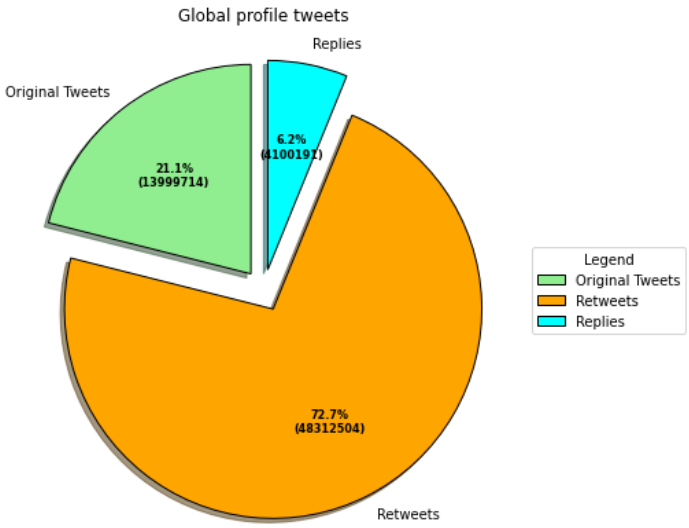


Figura 8: Statistiche in funzione della tipologia di tweet sull'intero DataFrame

Dal diagramma a torta riportato nella Figura 8 possiamo notare subito come l'attività maggiormente svolta riguarda i retweet, che compongono il 72,7% dell'intero Dataset.
I tweet originali, quindi i tweet veri e propri che non sono generati in risposta ad un'attività, si prendono tuttavia una fetta piuttosto ampia andando a comporre il 21,1% mentre le risposte solo il 6,2% del Dataset.
Da questo si deduce che gli utenti tendono a effettuare retweet piuttosto che scrivere un post.

3.2.1 Parallelizzazione

Una volta presa visione dei dati è stato necessario effettuare un'integrazione con gli studi iniziali per generare un punto di partenza per le successive analisi.

Il primo approccio è stato considerare ogni mese come un semplice DataFrame, operazione che ha però portato alla saturazione del disco e crash.

Per ovviare a questo problema si è pensato di suddividere i grossi DataFrame in SubDataFrame più piccoli e processarli in parallelamente.
Si è usato l'attributo `chunksize` della funzione `read_csv()` del modulo Pandas, in grado di prendere un numero di righe pari alla dimensione del parametro passato.
Si è poi processato in modo asincrono e parallelo i dati sfruttando il modulo `concurrent.futures`. Il modulo consente di lanciare task in parallelo in modo asincrono (usando `ThreadPoolExecutor` o `ProcessPoolExecutor`).

Per scelta implementativa si è utilizzata una `ProcessPoolExecutor`, con numero di processi massimi uguali al numero di processori logici disponibili (nel mio caso otto).

Per poter effettuare un'operazione di lettura e processamento dei dati sono stati destinati sette processori logici al processamento delle informazioni (quindi filtraggio sui dati, operazioni di conversione, operazioni logiche, ecc…), mentre un processore logico si occupa di leggere il file successivo.

La funzione `process_all_data` è la responsabile di quanto spiegato:

```
def process_all_data(filename, cols, list_name=None, chunksize=chunksize,
workers=workers):
    c = 1
    executor = ProcessPoolExecutor(max_workers=workers)
    futures = []
    partial_results = []
    results = []
    chunks = pd.read_csv(filename, lineterminator='\n', chunksize=chunksize,
usecols=cols, low_memory=False)
    chunk = None
    try:
        chunk = next(chunks)
    except StopIteration:
        chunk = None
    i = 0
    while chunk is not None:
        print(f"Processing chunk {c}")
        subchunks = np.array_split(chunk, workers)
        for sc in subchunks:
            try:
                futures.append(executor.submit(process_data_tweets, sc))
            except Exception as e:
                logger.exception("Error", e)
            i += 1
        try:
            chunk = next(chunks)
        except StopIteration:
            chunk = None
        c += 1
        futures_wait(futures)
        try:
            partial_results = [fut.result() for fut in futures]
        except Exception as e:
            logger.exception("Error", e)
    results.append(partial_results)
    return results
```

In questo modo si va a generare un corretto load balancing sulla CPU migliorando le prestazioni in modo non lineare, per la lettura dei files contenenti le informazioni degli utenti si passa da circa 1000 secondi a 390, e senza rischiare la saturazione del disco. La differenza è chiaramente visibile nella Figura 9.

Figura 9: Differenza tra esecuzione in parallelo (a sinistra) e seriale (a destra)

3.3 Score con Botometer

Una volta integrati i dati nuovi con gli studi iniziali è iniziato il processo di raccolta degli score degli utenti con Botometer.
Il criterio di scelta degli utenti “to be scored” è stato dato in funzione dell’attività degli utenti e della tipologia di attività compiuta da tali, in particolari è andato a dividere il DataFrame in cinque diverse tipologie di attività:

- Tweet originali
- Retweet effettuati
- Retweet ricevuti
- Risposte effettuate
- Risposte ricevute

Una volta ottenute queste informazioni sono stati filtrati i 20.000 utenti più attivi per tipologia di attività, raggiungendo un numero di utenti “to be scored” di 100.000 unità.

Dopo aver definito il criterio di scelta di score è stato lanciato lo script responsabile della raccolta di tali informazioni.
La risposta di Botometer indica, secondo una serie di campi, delle informazioni sia sulla possibilità che l’utente sia o meno un bot, sia sulla tipologia di bot che rappresenterebbe.

Questo dato è piuttosto lungo e ricco di informazioni, per capirne meglio la struttura è necessario visionare un esempio in Figura 10.


```
{'cap': {'english': 0.797075908158591, 'universal': 0.8052478284119403},
 'display_scores': {'english': {'astroturf': 0.4,
   'fake_follower': 0.9,
   'financial': 0.0,
   'other': 3.4,
   'overall': 3.4,
   'self_declared': 1.4,
   'spammer': 0.2},
  'universal': {'astroturf': 0.9,
   'fake_follower': 0.8,
   'financial': 0.1,
   'other': 3.2,
   'overall': 3.2,
   'self_declared': 1.6,
   'spammer': 0.0}},
 'raw_scores': {'english': {'astroturf': 0.09,
   'fake_follower': 0.18,
   'financial': 0.0,
   'other': 0.68,
   'overall': 0.68,
   'self_declared': 0.29,
   'spammer': 0.05},
  'universal': {'astroturf': 0.18,
   'fake_follower': 0.15,
   'financial': 0.02,
   'other': 0.64,
   'overall': 0.64,
   'self_declared': 0.33,
   'spammer': 0.01}},
```


Figura 10: Risposta di Botometer API alla richiesta di score per un utente

L'oggetto in figura 10 viene trattato come dizionario in Python e contenente categorie e sottocategorie con i relativi score, suddivisi come segue:

Categorie:

- cap: probabilità condizionale che un utente con un punteggio maggiore o uguale a questo sia automatizzato (sia nella lingua dell'utente che in inglese)
- display scores: come raw score ma con un range [0,5]
- raw scores: bot score con range [0,1], sia in inglese che nella lingua dell'utente, contiene sottocategorie
- user: Utente Twitter con informazioni sull'id, nome e lingua dedotta dalla maggioranza dei tweet

Sottocategorie:

- fake_follower: bot acquistati per aumentare il numero di follower
- self_declared: bot ottenuti su botwiki.org
- astroturf: bot politici etichettati manualmente e account coinvolti in follow trains (diversi account si seguono a vicenda per aumentare il numero di followers) che cancellano sistematicamente i propri contenuti
- spammer: account etichettati come spambot
- financial: bot che usando i cashtag
- other: vari altri bot ottenuti da annotazioni manuali, feedback degli utenti, ecc.

Per la categorizzazione bot/human, come da stato dell'arte nella sezione "Data analysis",[5] nel capitolo "Automation", si è andato ad individuare come effettuare la distinzione tra entità che sono sicuramente umani e le entità che sono sicuramente dei bot.

In particolare, l'articolo fa riferimento alla figura statistica del decile, ovvero uno dei nove valori medi che divide una successione di numeri, in dieci parti uguali.
L'articolo afferma che gli utenti con cap score inferiori al primo decile sono umani, mentre quelli successivi con cap score superiore all'ultimo decile sono dei bot.

È possibili visualizzare la distribuzione del cap score con due linee ad indicare la posizione di entrambi i decili nella Figura 11.

5 https://journals.uic.edu/ojs/index.php/fm/article/download/11431/9993

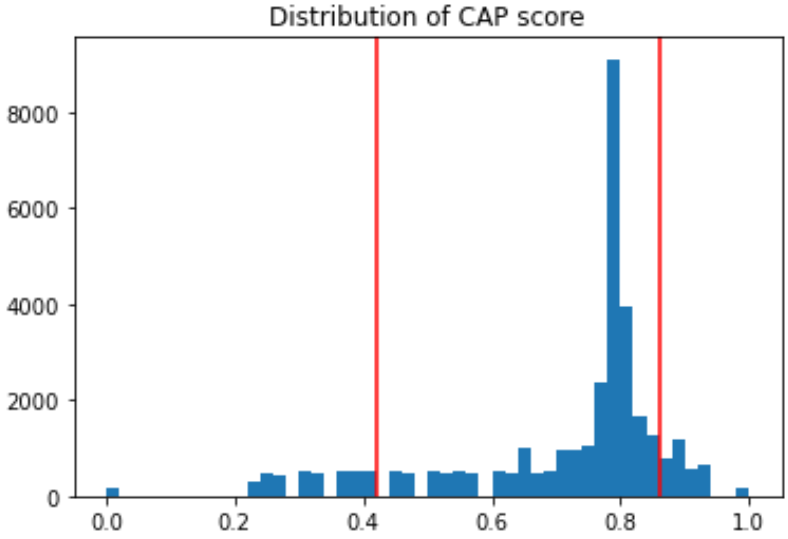


Figura 11: Distribuzione del CAP score con indicazione del primo e ultimo decile

Una volta effettuata la classificazione bot/human si è suddiviso per tipologia di bot in modo tale da verificare se i risultati avessero un riscontro con la tipologia di entità prevista.

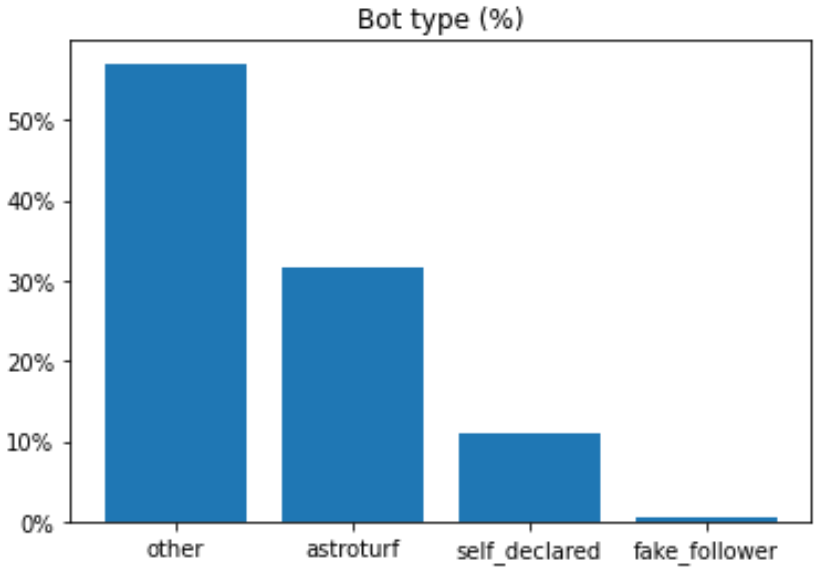


Figura 12: Tipologie di Bot ottenute

Come ci si aspettava non vi sono bot di tipo “financial” e “spammer”, tipologie di bot non inerenti all’argomento trattato.

Una volta effettuato questo genere di studi gli utenti con i relativi botscore sono stati riportati su un file per eventuali confronti incrociati nel corso del progetto.

3.3 Disinformation Dozen

Un'analisi complementare a quella effettuata sulle attività dei bot riguarda i cosiddetti "Disinformation Dozen", questo termine indica una categoria di utenti responsabili del 65% delle notizie anti-vaccino sui social.
La notizia della presenza di uno studio già esistente che ha prodotto i “Disinformation Dozen” è stata appresa da un articolo molto interessante di Repubblica.[6]
L’articolo in questione è un report del “Center for Countering Digital Hate”, organizzazione non-profit con sede a Londra che ha come scopo “distruggere l'architettura di odio e disinformazione online”.
L’organizzazione ha pubblicato il report il 24 marzo 2021,[7] il quale afferma che solo dodici utenti sono responsabili di quasi due terzi dei contenuti anti-vaccini che circolano sulle piattaforme di social media.
Questa nuova analisi porta alla luce come un piccolo gruppo di determinati novax sia responsabile di un'ondata di disinformazione.
Lo studio mette in mostra, inoltre, di come gli OSN siano i responsabili della proliferazione di notizie di qualsiasi genere e validità scientifica.

I personaggi in questione sono riportati in Tabella 3.

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

Tabella 3: Disinformation Dozen con relative informazioni

(Aggiornato a settembre 2021, stato e followers sono in riferimento a Twitter)

[6] https://www.repubblica.it/esteri/2021/07/18/news/coronavirus_quella_sporca_dozzina_d_influencer_dietro_a_oltre_meta_delle_fake_news_sui_vaccini-310811514/
[7] https://www.counterhate.com/disinformationdozen

Salta subito all'occhio come nessuno degli utenti in questione abbia una competenza scientifica in merito all'argomento trattato, tra questi utenti infatti figurano attivisti, osteopati, chiropratici, avvocati ma nessun mestiere che abbia a che fare con la medicina o approvato dalla comunità scientifica o dalla medicina.

La prima attività svolta con queste preziose informazioni è stata una ricerca sull'effettiva presenza di questi "Disinformation Dozen" all'interno del nostro DataFrame di lavoro.
L'esito positivo di questa operazione ha rivelato poche centinaia di tweets dal momento che i dati a nostra disposizione fanno riferimento al periodo gennaio 2020 – maggio 2020 e le notizie sui vaccini sono ipoteticamente più recenti.
Si è quindi passati tramite la Twitter API per poter raccogliere tutti i tweets relativi a questi personaggi, nello specifico le attività di questi profili (quindi anche i retweet ricevuti, quando sono citati e le risposte).
Una volta ottenute le informazioni, una prima analisi dei dati ha mostrato subito pattern interessanti da analizzare: si nota come le strategie di attività dei "Disinformation Dozen" siano differenti rispetto alle attività ottenute nella rete globale.
Nella Figura 13 è possibile vedere tale differenza.

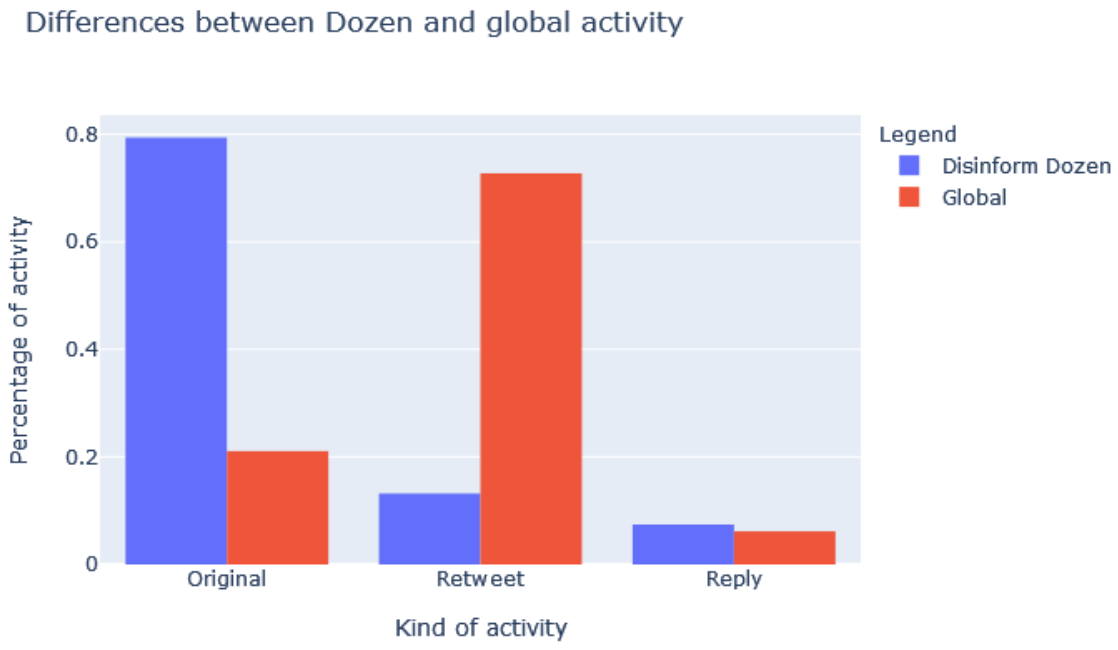


Figura 13: Confronto tra le attività degli utenti e i "Disinformation Dozen"

Una prima idea ipotizzata di questa differenza nelle attività è che i "Disinformation Dozen" hanno come finalità la produzione di fake news o notizie di dubbia veridicità, fungendo come fonte di informazione per chi interagisce con loro, in contrapposizione molti utenti tendono a

retwittare un contenuto piuttosto che produrne uno.
Per capire la portata della disinformazione creata da questi 12 utenti basti pensare che coinvolgono oltre 150.000 utenti diversi in termini di retweet e risposte, ricevendo oltre 2.000.000 di retweet nel periodo che va da gennaio 2020 a luglio 2021.
Il più popolare in assoluto è Robert Francis Kennedy Junior, nonché nipote di John Fitzgerald Kennedy, presidente americano ucciso nel 1963.

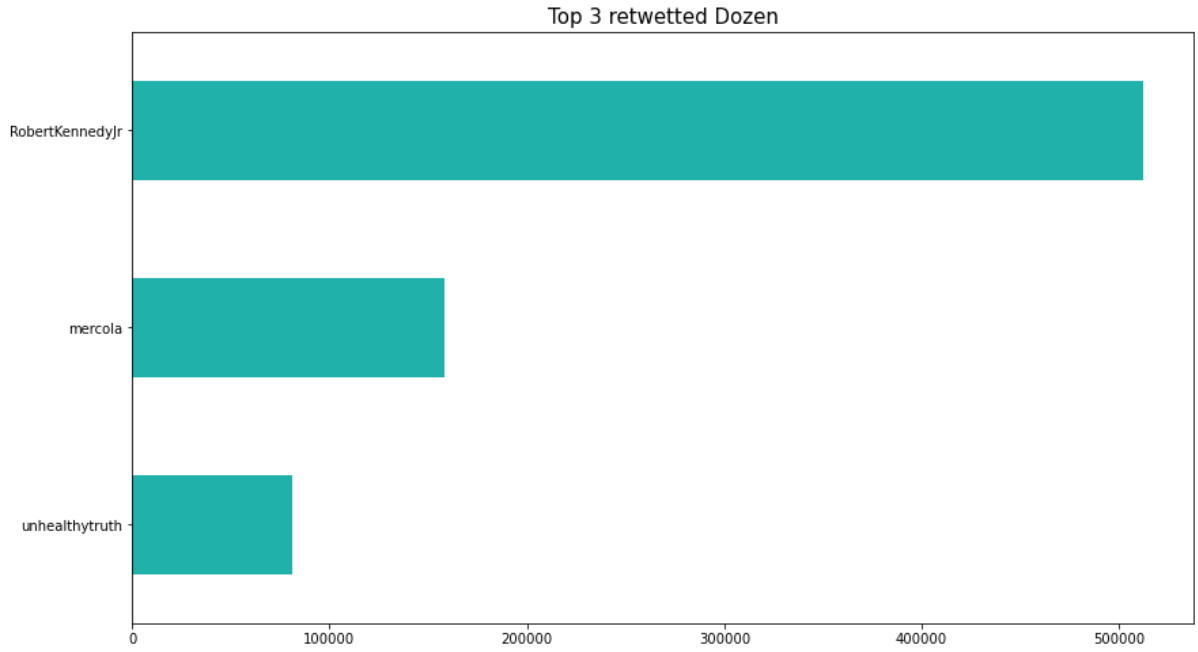


Figura 14: Dozen più retwittati nel DataFrame in utilizzo

Come è possibile notare in figura ottiene da solo oltre 500.000 di retweet, il suo legame con l'ex presidente legato al cognome, la sua attività politica e l'attivismo in campo ambientale e umanitario (presidente e fondatore di un'organizzazione non governativa, la "childrenshealthdefense") lo rendono agli occhi di molti una fonte attendibile da cui ottenere informazioni e giustificano la sua popolarità nel contesto in analisi.

3.4 Disinformation Dozen e utenti verificati

Dal momento che il confronto dei "Disinformation Dozen" con dei semplici utenti qualsiasi, come potrebbero essere giornali, medici o gente comune non rappresenta un metro di paragone valido si è scelto di effettuare una scrematura degli utenti con cui effettuare un paragone.

La prima scelta è stata effettuare un'operazione di filtering sul DataFrame generale estraendo tutti i tweets degli utenti verificati che rappresentano entità "autentiche, notorie e attive".[8]
Andando nel dettaglio un account verificato deve rispettare i seguenti parametri:

- **Autentico**: verificato tramite documento, sito ufficiale o indirizzo e-mail ufficiale
- **Notorio**: l'account deve rappresentare un individuo o un brand illustre oppure esservi in qualche modo associato
- **Attivo**: l'account deve essere attivo e vantare una storia di rispetto delle Regole di Twitter

Per ottenere i tweets degli utenti attivi si è filtrato il DataFrame degli utenti secondo la colonna "verified" e, una volta ottenuti gli id e il nome utente di questi, sono stati ottenuti i tweets di questi utenti dal DataFrame dei tweets.
A questo punto è stata effettuata un'attività di calcolo della credibilità delle fonti condivise dai singoli utenti di entrambi i DataFrame per farne un paragone.

Dal momento che i "Disinformation Dozen" sono solo dodici si è scelto di considerare, per questo studio, anche chi retwitta i loro tweet, creando due gruppi di utenti: i supporter dei "Disinformation Dozen", compresi "Disinformation Dozen stessi" e gli utenti verificati.

L'idea è quella di iterare su tutti gli utenti, ottenendo tutti i domini condivisi da ciascuno utente, a questo punto è possibile verificare quali di questi domini condivisi sia ad alta credibilità o bassa.

Effettuando un rapporto tra il numero di domini ad alta credibilità condivisi e il numero totale di domini condivisi si va a creare un valore [0,1] che rappresenta lo score di high credibility per ogni utente.
Al termine di questa operazione viene a realizzarsi una lista di utenti con associato high score credibility.

8 https://help.twitter.com/it/managing-your-account/about-twitter-verified-accounts

La funzione che si occupa di questa classificazione è la seguente:

```
def percentage_high_credibility(users, df, credibility_url):
    perc_list = []
    for i in users:
        x = df[df["user_screen_name"]==i]
        urls = tweets_utils.format_urls(x["urls"])
        cred_list = []
        value = 0
        value_h = 0
        if(len(urls) > 0):
            for j in range(len(credibility_url["Domain"])):
                if credibility_url["Domain"][j] in urls:
                    value = value + 1
                    if(credibility_url["Class"][j] == "high"):
                        value_h = value_h + 1
            if(value > 0) and (value_h) > 0:
                perc_high = value_h / value
            elif (value_h) == 0 and (value == 0):
                perc_high = -1
            elif (value_h == 0) and (value > 0):
                perc_high = 0
            perc_list.append((i, perc_high))
    return perc_list
```

Il risultato è una lista di tuple che contengono il nome dell'utente e un valore di "high credibility shared" che va da 0 a 1, questo rappresenta per il singolo utente la percentuale di credibilità delle fonti condivise.

Per capire meglio questo valore, lo 0 significa che tutte le fonti condivise hanno credibilità bassa, 0.5 significa che l'utente condivide una metà di fonti con alta credibilità e l'altra metà di bassa credibilità (quindi al 50% sono "high credibility") e 1 significa che le fonti condivise sono tutte fonti credibili (quindi al 100% sono "high credibility").

Il risultato ottenuto è molto interessante, infatti la distribuzione di high credibility dei "Disinformation Dozen" e dei loro supporters è distribuita in modo quasi omogeneo tra lo 0 e l'1, con un picco verso lo 0; mentre per gli utenti verificati la concentrazione maggiore si ha verso l'1, con dei valori molto bassi tra lo 0,8 e lo 0.

Questo indica che i "Disinformation Dozen" e i loro supporters tendono a condividere e ricondividere domini con scarsa credibilità, mentre gli utenti verificati pubblicano principalmente da articoli con alta credibilità

Il risultato è riportato nel graficamente nella Figura 15.

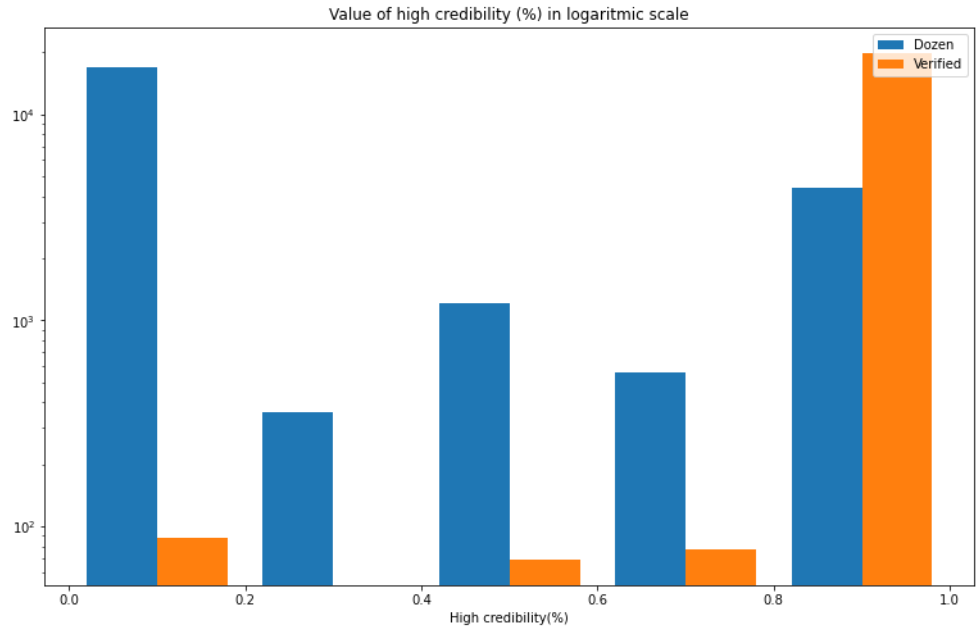


Figura 15: Confronto di credibility score tra supporters dei i Dozen e utenti verificati

L'analisi condotta fornisce risultati interessanti, una volta però effettuato uno studio sugli utenti più attivi che componevano questo DataFrame di utenti "Verified" è emerso come la maggior parte di questi utenti fossero testate giornalistiche o fonti di informazione; quindi, non si ha la certezza che siano utenti che portano buona informazione. La Figura 16 mostra i cinque utenti verificati che effettuano il maggior numero di tweets.

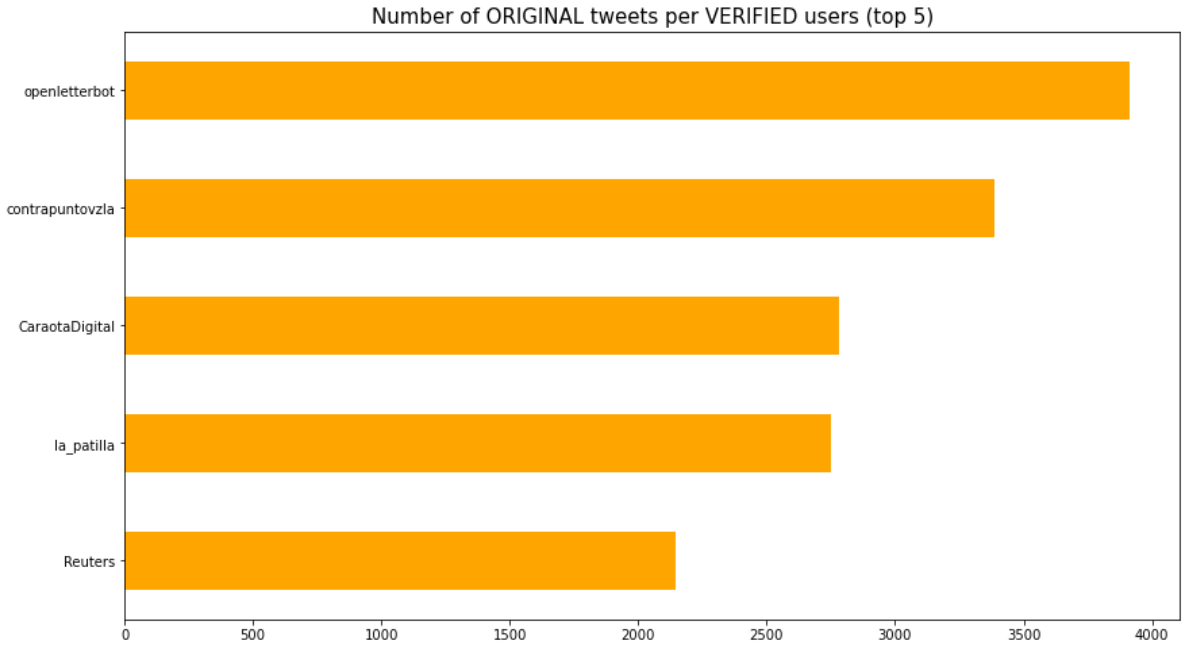


Figura 16: 10 utenti verificati che effettuano più tweets

Come anticipato salta immediatamente all'occhio come gli utenti verificati non siano un buon punto di partenza per effettuare un confronto sensato ma, anzi, hanno proprio come utente più attivo un bot.
Queste osservazioni ci hanno portato alla conclusione di dover filtrare ulteriormente gli account verificati da dover confrontare con i Dozen, come spiegato nella prossima sezione.

3.4 Good e Bad Dozen

L'esigenza è stata quindi quella di trovare una dozzina di utenti responsabili di informazione di qualità o ad alta affidabilità da contrapporre ai Dozen già trovati. Per portare avanti questa ricerca si è ispezionata manualmente la lista degli utenti verificati più attivi, andando a classificare ogni in base all'affidabilità delle fonti condivise.
Dopo aver trovato pochi utenti che potessero corrispondere alla figura di interesse e aver appurato che la ricerca fosse troppo onerosa in termini di tempo, si è scelto di verificare se ci fossero studi già esistenti per trovare questi utenti.
Fortunatamente è stata trovata una pubblicazione di Jonathan Oppenheim (Professore di Teoria Quantistica e ricercatore della Royal Society, University College London) e Kelly Truelove (ricercatore della University of California) riguardante i "Twitter accounts to follow on Covid-19", i migliori account da seguire per ottenere informazioni scientificamente attendibili sul Covid-19.
La pubblicazione raccoglie ben 300 differenti account, scelti in base alla loro competenza e popolarità.[9]

Prima di prendere i primi 12 utenti di questa lista composta da 300 nomi è stato verificato manualmente che ogni utente fosse effettivamente attivo e presente nel DataFrame in utilizzo fino a questo momento.
Una volta verificato che i nuovi utenti fossero presenti si è passati alla definizione di due nuove categorie: si è deciso di definire come "Bad Dozen" gli utenti precedentemente individuati, responsabili del 65% delle fake news, e "Good Dozen" i primi 12 utenti ottenuti dallo studio condotto.

Per analizzare se gli utenti fossero attivi almeno quanto i "Bad Dozen" sono stati calcolati il numero di post e i retweet ricevuti dai "Good Dozen", ottenendo come risultato 14.348.506, dimostrando la loro popolarità.

9 https://www.ucl.ac.uk/oppenheim/Covid-19_tweeps.shtml

La tabella che fa riferimento ai “Good Dozen” è la seguente:

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

Tabella 4: Good Dozen con le relative informazioni

(Aggiornato a settembre 2021, stato e followers sono in riferimento a Twitter)

Dando un’occhiata alle tabelle appare subito lampante la differenza abissale di competenze tra gli attori dell’informazione in questione, in particolare possiamo constatare come per i “Bad Dozen” sia molto comune la medicina alternativa, mentre per i “Good Dozen” la componente chiave lo svolge la medicina tradizionale.

È importante sottolineare tuttavia che chi ha competenze specifiche maggiori non necessariamente ha un comportamento corretto nella discussione e non si ha la certezza che i domini condivisi siano ad alta credibilità.
In questo caso è bene ricordare che la scelta dei “Good Dozen” non è affatto casuale e, oltre alle competenze scientifiche si ha la certezza del corretto comportamento nel dibattito.

Una volta ottenute le interazioni di queste due categorie di utenti con il resto dell’utenza di Twitter si è andato ad effettuare un confronto per capire come le strategie di informazione differiscano.
In modo analogo a quanto precedentemente fatto con gli utenti verificati sono state analizzate le statistiche relative alla tipologia di attività svolta, andando a fare un confronto tra le due tipologie di Dozen.

In Figura 17 notiamo subito una differenza di attività tra i Dozen netta.

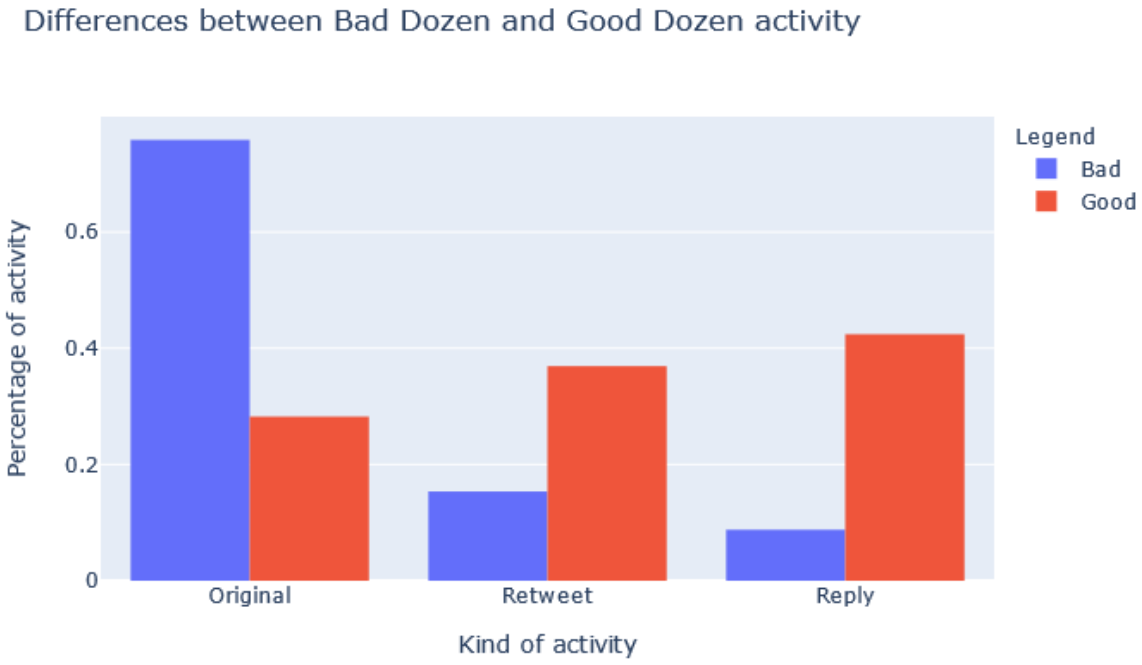


Figura 17: Confronto tra le attività dei "Good" e dei "Bad" Dozen

La differenza netta tra queste due tipologie di utenti suggerisce un comportamento diametralmente opposto nella strategia di condivisione delle informazioni, i “Bad Dozen” infatti fungono come “fonte di verità” per i loro seguaci; pertanto, sono autori di post originali e tendono a non dare particolare peso all’attività di retweet e risposte. Per quanto riguarda i “Good Dozen” invece abbiamo, per la prima volta, una categoria che sembra attenta alle risposte e mantiene un certo equilibrio nelle varie attività.

Un’analisi più approfondita ha permesso di stabilire come non ci siano interazione tra “Bad Dozen” e “Good Dozen”, andando ad effettuare uno studio delle risposte dei “Good Dozen” è emerso come il 52.50% siano auto risposte. Nella figura 18 abbiamo un riscontro grafico.

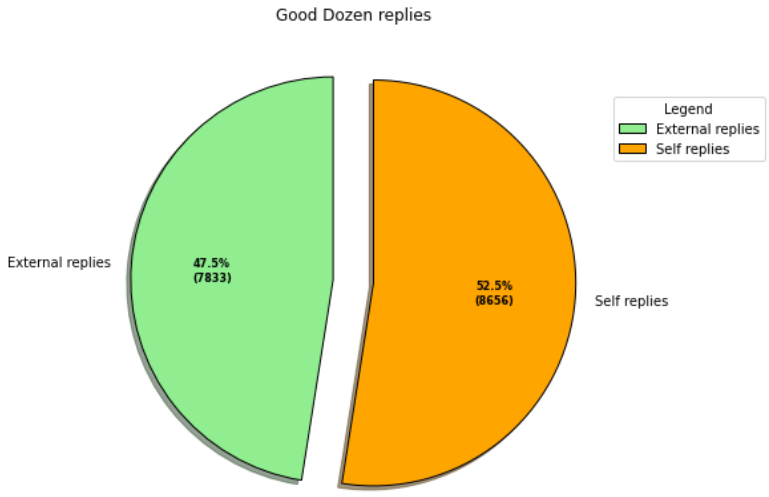


Figura 18: Tipologie di risposte dei "Good Dozen"

Facendo un controllo manuale è emerso comunque che questa categoria di utenti risulta essere molto attiva in ambito risposte, spesso effettuate in risposta a scettici e curiosi, dal momento che con un tweet tendono a creare un “Conversation threading”: una volta stabilito il topic, l'interazione che ne segue assume la forma di un una conversazione tra l’utente che ha effettuato il post e followers.

Questa tipologia di strategia consente agli utenti interessati di effettuare una pseudo chat con l’autore del post che può dare diverse informazioni con un singolo tweet ed effettuare anche Q&A.

Un esempio di questa strategia di informazione è stato riportato nella Figura 19, che mostra un tweet di Kai Kupferschmidt, giornalista tedesco che ha studiato biomedicina molecolare all’Università di Bonn.

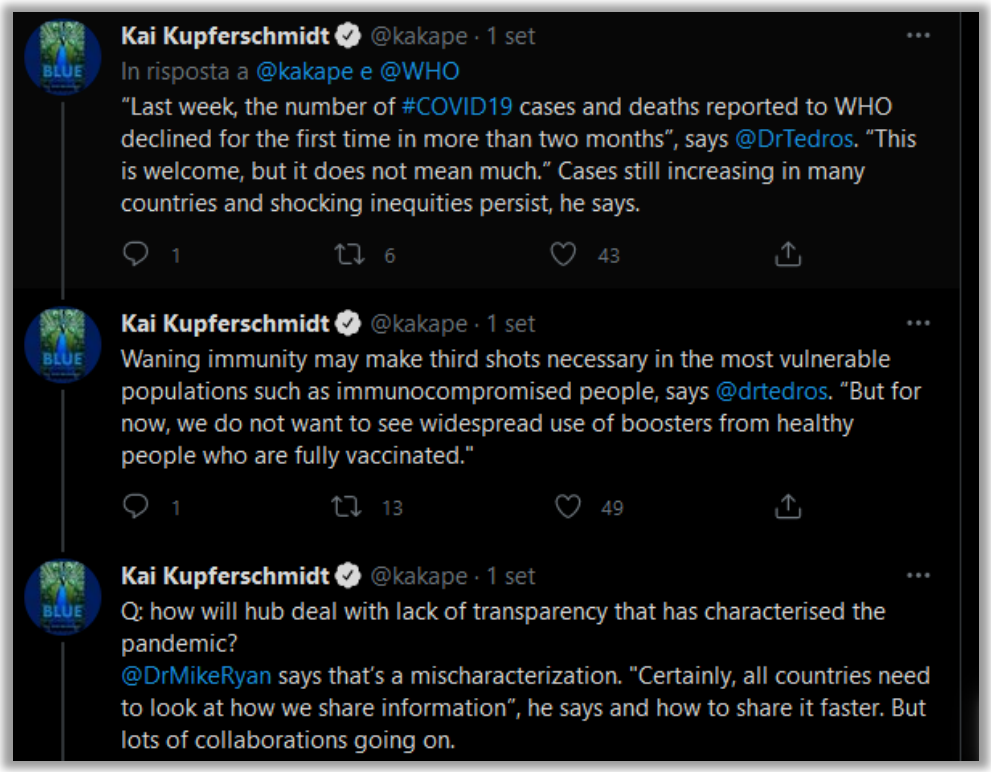


Figura 19: Esempio di “Conversation threading”

3.4.1 Hashtag

Gli hashtags sono gli strumenti di informazione per eccellenza, introdotto proprio da Twitter nel 2007, è un aggregatore tematico che semplifica la ricerca di un contenuto o un tema. Individuando gli hashtags più usati dalle due categorie è possibile identificare un contesto sul quale si basa la propagazione delle informazioni. Per quanto riguarda i "Bad Dozen" gli hashtag maggiormente utilizzati sono stati riportati graficamente nella Figura 20 sottostante.

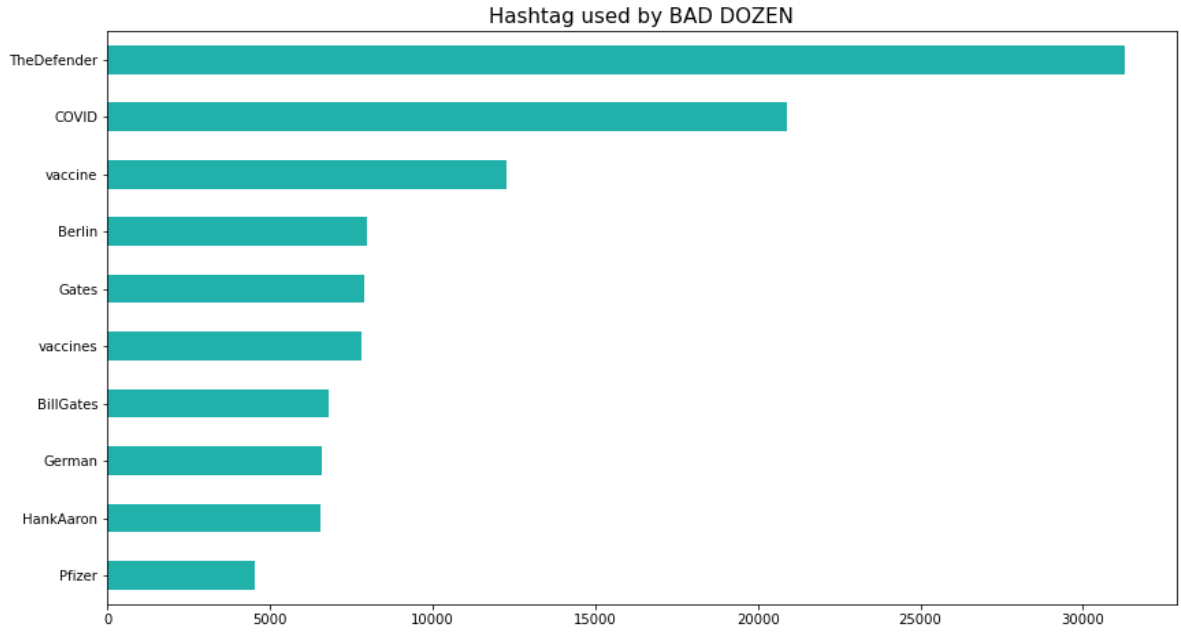


Figura 20: Hashtag più usati dai "Bad Dozen" e seguaci

Analizzando gli hashtag più interessanti otteniamo:

- **#TheDefender**: fa riferimento al giornale dell'organizzazione "Children's Health Defense", di proprietà di Robert F. Kennedy Jr.
- **#Berlin**, **#German**: fa riferimento ad una grossa protesta avvenuta proprio nella capitale tedesca, in cui Robert F. Kennedy Jr. è salito sul palco per tenere un discorso
- **#Gates**, **#BillGates**: fa riferimento al noto imprenditore americano accusati da molti novax di essere al centro degli interessi economici sui vaccini e altre teorie complottiste
- **#HankAaron**: fa riferimento all'ex giocatore di baseball americano che, secondo molti, sarebbe morto poco dopo la prima dose di vaccino a causa del vaccino stesso; teoria poi smentita dalle cause di morte che indicano sia morto di morte naturale all'età di 86 anni

Notevole differenza si ha negli hashtags usati dai “Good Dozen” e supporters, come mostrato nella Figura 21.

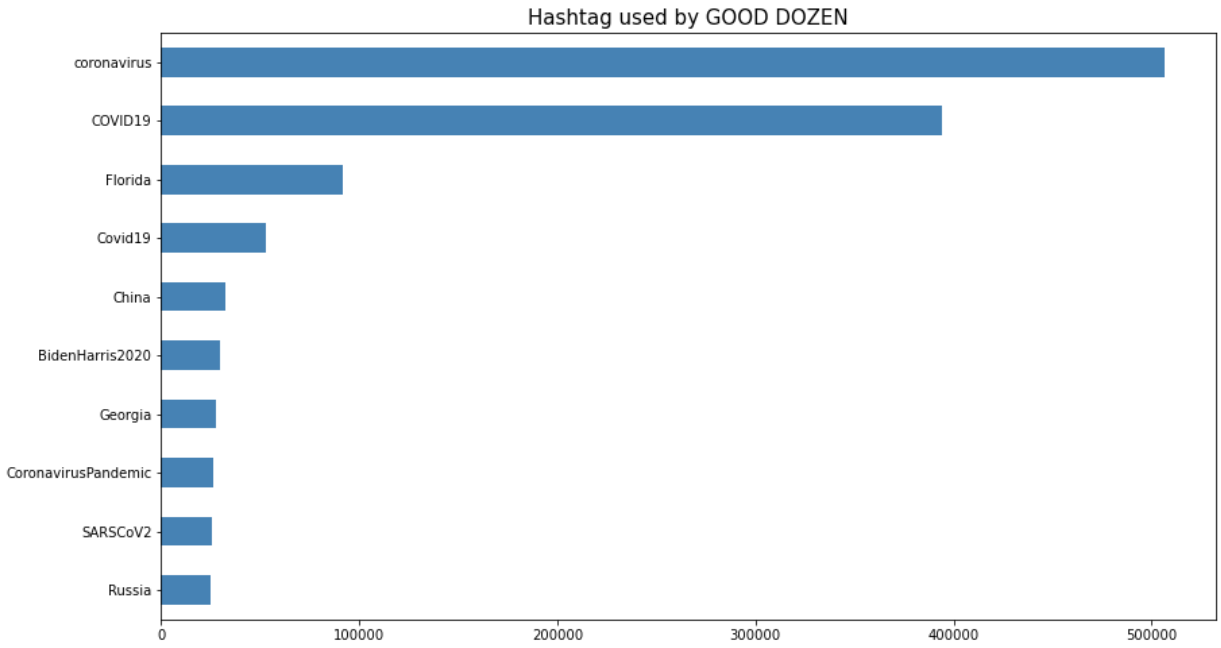


Figura 21: Hashtag più usati dai “Good Dozen” e seguaci

Per quanto riguarda questo grafico c’è ben poco da dire, gli hashtag sono più o meno tutti inerenti alla discussione sul Covid-19, la presenza di stati americani e dell’hashtag #BidenHarris2020 è data dall’inclinazione politica dei “Good Dozen”.

3.4.2 Domini

Dopo aver effettuato un’analisi sugli hashtags si è passato ai domini più condivisi, che rappresentano la principale fonte di informazione.

Per poter definire la tipologia di domini condivisi si è andati a prendere i più condivisi, stando però attenti ad alcuni domini non classificabili correttamente, ovvero tutti i domini compressi di cui non si è effettuata l’operazione di decompressione (es. bit.ly e ow.ly) e il dominio twitter.com, dal momento che se si cita un tweet è come se stessimo condividendo un link.

Si è quindi svolta un’operazione di comparazione tra i domini più condivisi per le due tipologie di utenti, nella Figura 22 possiamo vedere i dieci domini condivisi dai “Bad Dozen” e quanto questi siano condivisi dai “Good Dozen”.

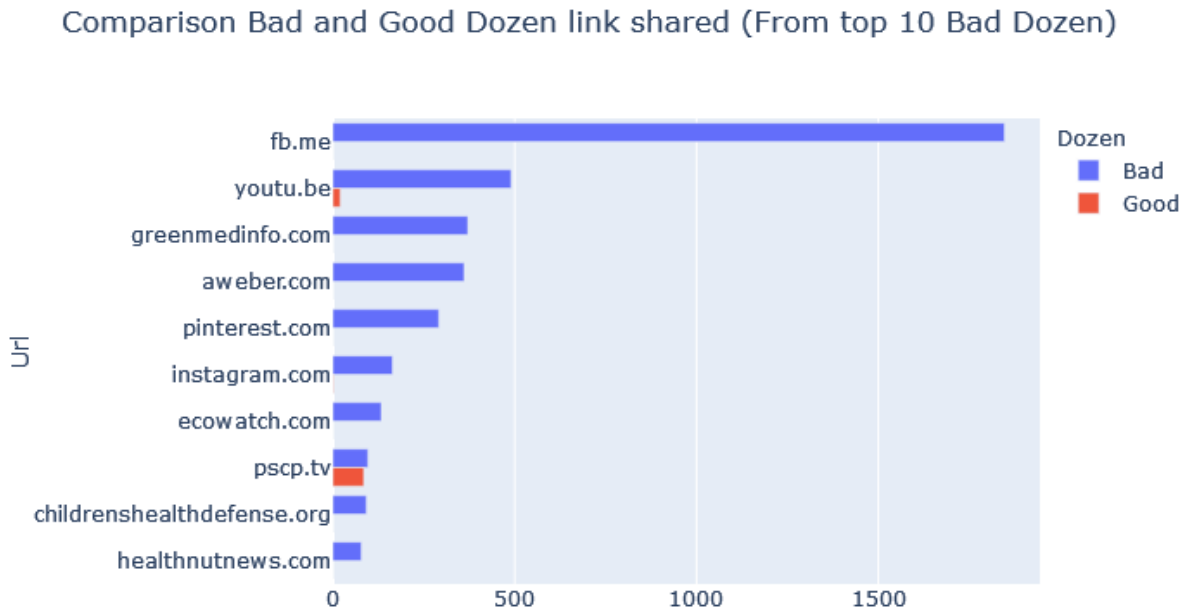


Figura 22: Comparazione con i 10 domini condivisi dai "Bad Dozen"

Spiccano i social network, primo tra tutti Facebook, poi YouTube e Instagram, che difficilmente sono mezzi di informazione affidabili.

Possiamo poi trovare Periscope e Aweber, utilizzati per videoconferenze e E-mail marketing con cui presumibilmente i “Bad Dozen” riescono a raggiungere l’utenza che li segue
Vi sono siti web di proprietà dei “Bad Dozen”, come greenmedinfo.com (di proprietà di Sayer Ji) e childrenshealthdefense.org (di proprietà di Robert F. Kennedy Jr.) e healthnutnews.com (di Erin Elizabeth) che sono chiaramente cospirazionisti e fonti assolutamente poco affidabili.
Questa informazione risulta essere tuttavia molto importante perché suggerisce come i “Bad Dozen” puntino ad ima strategia di condivisione delle informazioni legata ai propri siti web, evitando di sfruttare altre testate e domini, anche più popolari, facendo anche i propri interessi economici.

Vi è una sola fonte autorevole, ovvero ecowatch.com, rivista incentrata sull’ecologia e l’ambiente.
È interessante notare come non vi siano quasi domini condivisi tra i “Bad” e i “Good” Dozen, se non per qualche piccola eccezione composta da YouTube e Periscope, piattaforme usate per condividere video (nel caso di YouTube) o avere delle vere e proprie conferenze (nel caso di Periscope).

Riprendendo quanto fatto precedentemente è stato condotto uno studio analogo sui dieci domini più condivisi dai “Good Dozen”, le differenze di strategie adottate sono piuttosto visibili. Nella Figura 23 possiamo vedere lo studio in modo chiaro.

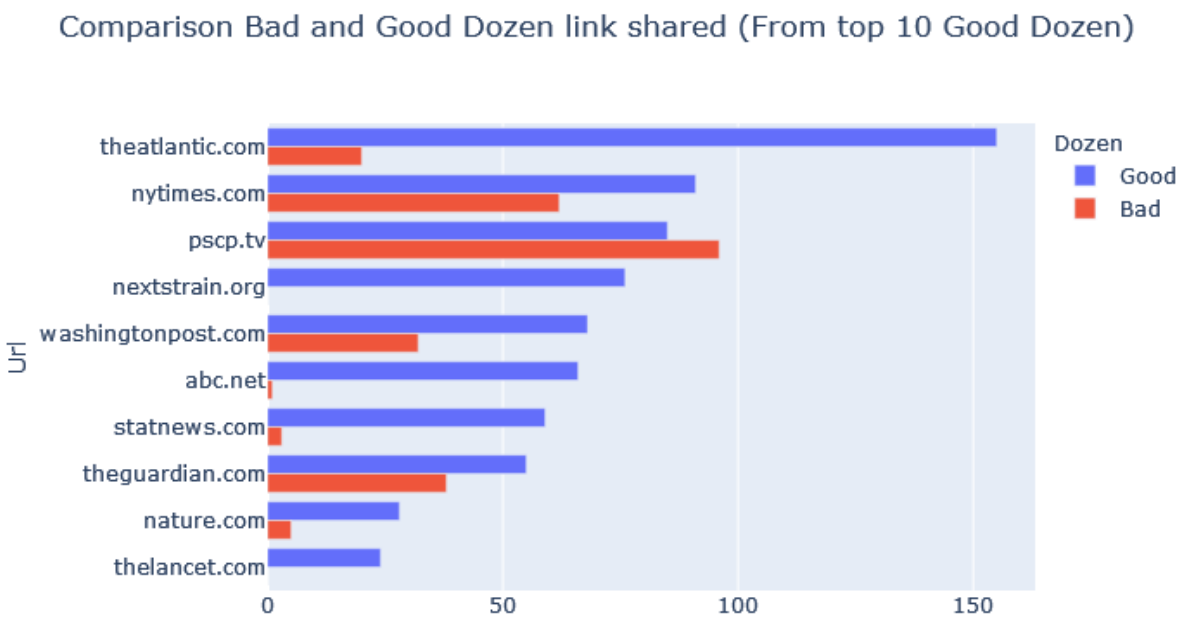


Figura 23: Comparazione con i 10 domini condivisi dai "Good Dozen"

La prima cosa che possiamo notare è la tipologia dei domini condivisi, si passa infatti da social network e servizi a testate scientifiche e giornalistiche, indice di una tipologia di informazione decisamente più orientata ad un approccio scientifico-informativo.
Un'altra nota da sottolineare è il numero di volte che i domini sono stati condivisi; infatti, il dominio più condiviso dai "Good Dozen" è theatlantic.com, condiviso solo 155 volte, mentre per i "Bad Dozen" Facebook è stato condiviso ben oltre le 1800 volte.

Differentemente da quanto accaduto prima abbiamo una grande quantità di domini in comune, non vi è nessuna condivisione comune con i domini who.int (World Health Organization) e nextstrain.org (progetto open-source per sfruttare il potenziale scientifico e di salute pubblica dei dati sul genoma dei patogeni), che rappresentano i domini scientifici più presenti.

3.4.3 Credibilità

Una volta identificati i domini è stato effettuato un lavoro di classificazione di credibilità per ogni utente (come visto nel capitolo 3.4 Disinformation Dozen e utenti verificati), per verificare se effettivamente i "Good Dozen" siano una fonte di informazione credibile.
L'equiparazione trai Dozen è netta, visibile nella Figura 23.

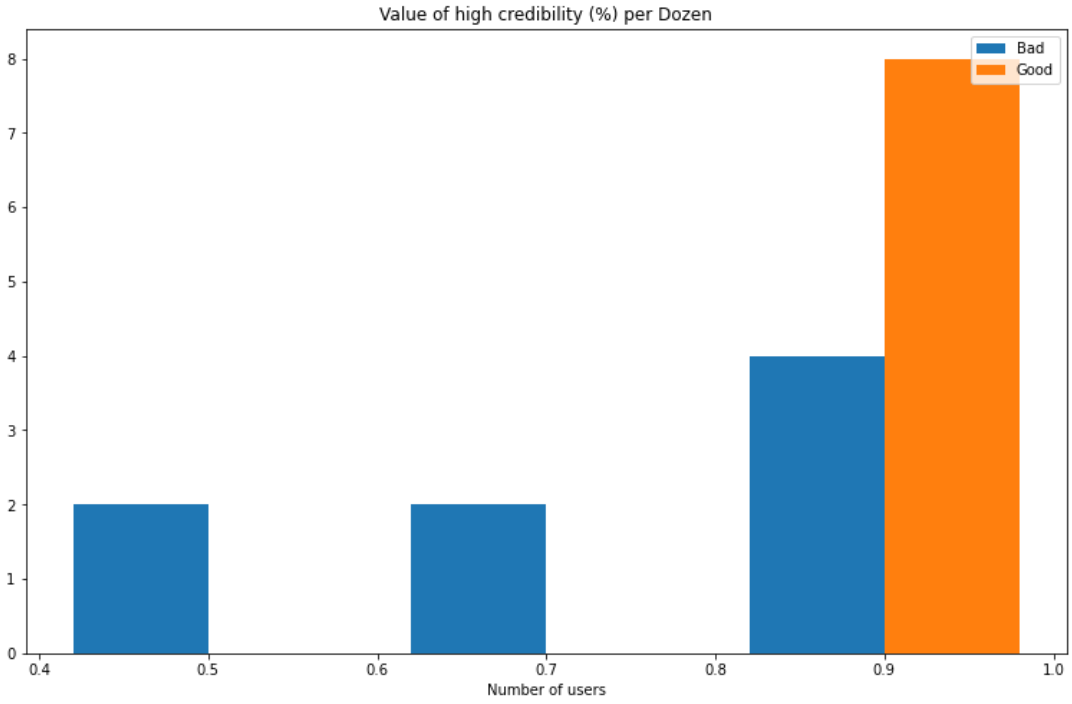


Figura 24: Comparazione di credibilità dei domini condivisi tra i Dozen

Dallo studio condotto sono 8 su 12 i "Good Dozen" che condividono dei domini di cui abbiamo informazioni, stesso risultato per i "Bad Dozen", con l'unica differenza che i "Good Dozen" condividono solo domini ad alta credibilità, mentre per i Bad questo risultato varia.
Per capire meglio il valore degli score dei "Bad Dozen" possiamo leggere la seguente tabella:

Dozen	Score
mercola	0.921053
RobertKennedyJr	0.921569
DrButtar	0.500000
unhealthytruth	0.933333
sayerjigmi	0.428571
DrChrisNorthrup	0.647059
kevdjenkins1	0.849315
BusyDrT	0.666667

Tabella 5: Bad Dozen con i relativi score di high credibility

Andando ad effettuare un lavoro analogo a quanto fatto sulla credibilità degli utenti verificati e i supporters dei “Bad Dozen”, sostituendo gli utenti verificati con i supporters dei “Good Dozen”, otteniamo un risultato molto simile al precedente (Figura 15).
Il grafico mostrato in Figura 24 mostra come la credibilità dei domini condivisi dai supporters dei “Good Dozen” sia molto alta, con una piccola eccezione che si colloca vicino lo 0.

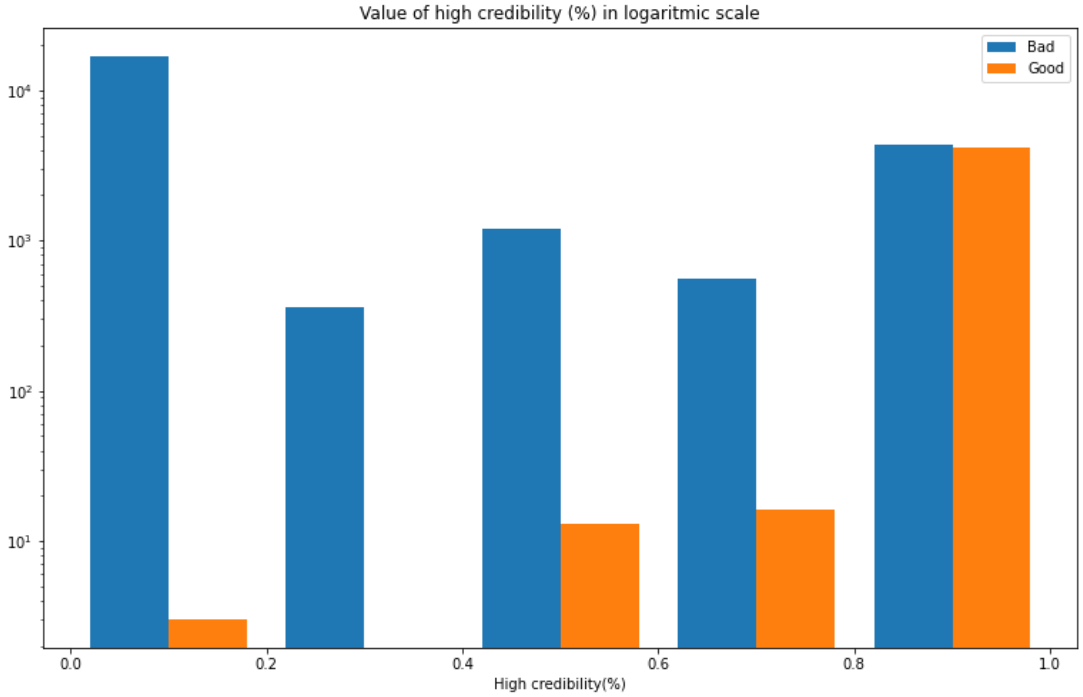


Figura 25: Confronto di credibility score tra gli utenti che interagiscono con i Dozen

Sebbene la distribuzione degli score per le “Bad Dozen interactions” sia equilibrata, si nota una grossa quantità di utenti con score pari a 0, discorso diametralmente opposto nel caso delle “Good Dozen interactions”, che eccetto pochi utenti hanno score di credibilità molto alto.

3.4 Retweet Network

Una parte importante del progetto fa riferimento alle network, è necessario infatti stabilire una metrica per il raggruppamento di utenti in community per poterne identificare un’identità, un comportamento, una linea o un’idea comune.
Il modo migliore e generalmente riconosciuto in letteratura è quello di mettere in relazione due utenti con il retweet; questo semplice gesto, composto da un click, ci permette infatti di avere una relazione tra due utenti nel modo più semplice possibile.
Il retweet generalmente indica che l’utente che effettua l’azione è d’accordo o in interesse con l’argomento del tweet e vuole quindi diffonderne il contenuto.

Si è scelto di non considerare le “quotes”, si tratta di retweet con un contenuto testuale; pertanto, è molto difficile classificare il testo per capire se l’utente ritwitta perché è d’accordo o critica il tweet in sé.
Per la realizzazione della Network si sono ottenuti tutti i retweet dal DataFrame, generando un grosso file .csv composto dalle colonne “name” e “rt_name”, a indicare il nome dell’utente che retwitta e il nome dell’utente retwittato.

Il codice utilizzato invece è il seguente:

```
retweets_graph = nx.from_pandas_edgelist(df,'name','rt_name',
create_using=nx.DiGraph())
print(nx.info(retweets_graph))
degree_dict = dict(retweets_graph.degree(retweets_graph.nodes()))
sorted_degree = sorted(degree_dict.items(), key=itemgetter(1), reverse=True)
in_degree_dict = dict(retweets_graph.in_degree(retweets_graph.nodes()))
sorted_in_degree = sorted(in_degree_dict.items(), key=itemgetter(1),
reverse=True)
out_degree_dict = dict(retweets_graph.out_degree(retweets_graph.nodes()))
sorted_out_degree = sorted(out_degree_dict.items(), key=itemgetter(1),
reverse=True)

#### Communities with Louvain modularity
retweets_graph_und = nx.from_pandas_edgelist(df,'name','rt_name')
best_part = cm.best_partition(retweets_graph_und, random_state=42)
```

dove `sorted_degree`, `sorted_in_degree` e `sorted_out_degree` indicano la cardinalità dei nodi rispettivamente in generale, in ingresso e in uscita.
Si tratta di un dizionario, composto da nome_utente : cardinalità_nodo.
Un esempio dei cinque utenti con il grado più alto è il seguente:

```
{'CNN': 92923,
 'OH_mes2': 82234,
 'BarackObama': 71822,
 'spectatorindex': 68714,
 'tedlieu': 62764}
```

Il dizionario `best_part` è composto invece da una coppia chiave : valore secondo lo scema nome_utente : numero_community, un esempio è il seguente:

```
{'Huerconetzin': 0,
 'AnneKPIX': 1,
 'JustAnotherAme4': 2,
 'cnni': 2,
 'HHSRegion8': 2}
```

La rete ottenuta risulta di notevoli dimensioni, abbiamo infatti:

- Nodi 12.513.502
- Archi: 40.027.282
- Grado medio: 6.40

Per la classificazione delle community si è usato il metodo di Louvain, si tratta di un metodo di ottimizzazione greedy adatto alle grosse reti ed efficiente (che funziona in tempo O (n · log(n)) con n numero di nodi della rete).
La prima attività svolta è stata quella di identificazione delle community per poter comprendere il numero di community e la composizione. Per far ciò è stato necessario mettere i valori del dizionario `best_part` in una serie di Pandas e fare un plot, ottenendo il seguente grafico:

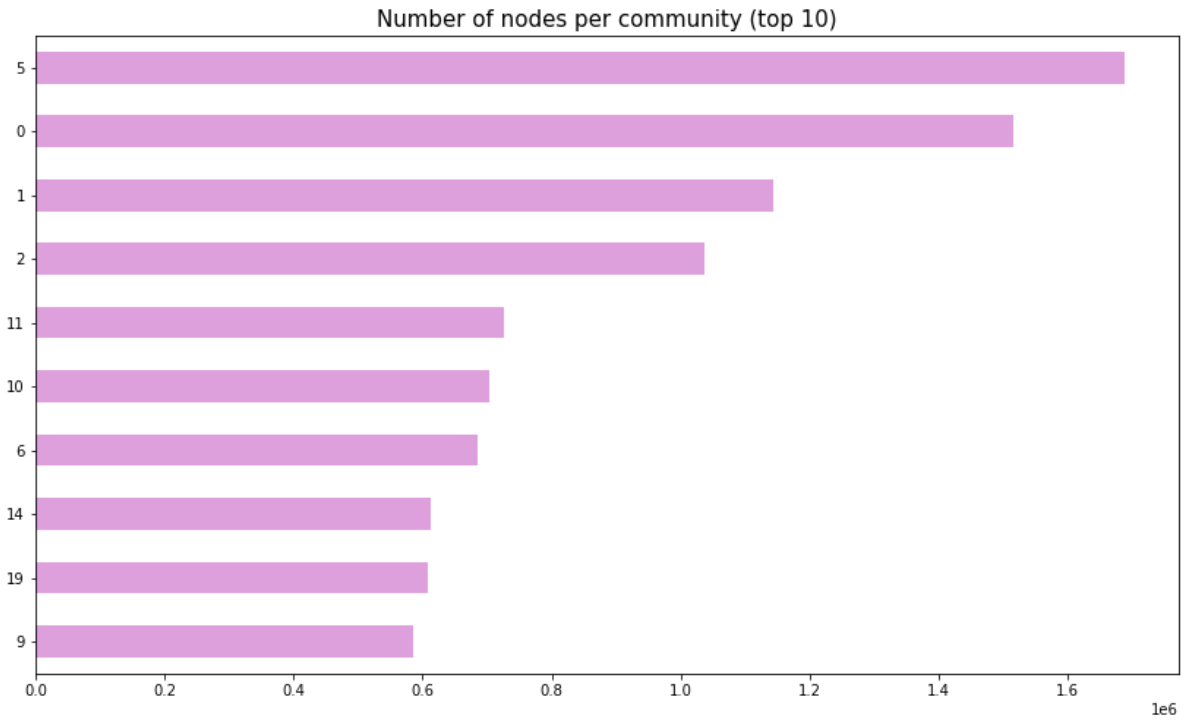


Figura 26: Community con il relativo numero di nodi che la compongono

Dal grafico si nota quanto siano numerose le prime dieci community, approfondendo gli studi abbiamo ottenuto che le prime 10 community contengono il 74.35% dei nodi, su un totale 325.073 communities diverse.
Per poter dare un contesto alle community sono state classificate secondo un approccio manuale e soggettivo: sono stati presi i 10 utenti più attivi per community e 10 nel mezzo, dopodiché i 20 utenti sono stati ricercati su Twitter, in funzione del contenuto da loro pubblicato è stata data una descrizione per ogni community.

Questa operazione, seppur molto manuale, ha dato degli ottimi risultati in linea con quanto ci si aspettasse, sono state identificate diverse community di lingue diverse (lingue presenti sulla documentazione del repository GitHub da cui sono stati ottenuti i dati), utenti con diversi interessi, ruoli, e provenienza.

Per comprendere meglio la composizione e dare una descrizione delle vare community si è scelto di rappresentarle nella Tabella 6 sottostante.

Community	Descrizione	Nodi
5	Utenti comuni, no personaggi famosi, politici, giornalisti (se non piccole eccezioni)	1.687.953
0	Notizie in lingua spagnola	1.516.564
1	Politici americani, giornalisti (tendenzialmente di sinistra)	1.142.470
2	Notizie e informazioni, testate internazionali	1.036.651
11	Notizie in lingua giapponese	725.867
10	Notizie in lingua inglese inerenti al Regno Unito	703.335
6	Notizie in lingua indiana	685.690
14	Notizie in lingua indonesiana	612.418
19	Importanti enti e medici	607.295

Tabella 6: Classificazione delle community

Molto interessanti sono diverse informazioni da questa classificazione: la prima sono le numerose community linguistiche, ben 5 sulle 9 analizzate, è interessante inoltre notare come la community più numerosa sia di utenti normali e non abbia un'impronta sociale o politica.

L'informazione più importante lo otteniamo però da tre community: la 1, la 2 e la 19.
Queste community sono incentrate sulle notizie e informazioni di pubblica utilità; pertanto, è plausibile che contengano secondo i Dozen, sia buoni che cattivi.

Andando a iterare attraverso queste Community otteniamo come la presenza dei Dozen sia distribuita all'interno di queste community in modo più o meno equo.

Le informazioni sulle community e i Dozen sono riportate nella Tabella 7.

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

Tabella 7: Dozen con le rispettive community

Per quanto riguarda la classificazione degli utenti che interagiscono con i Dozen otteniamo un importante risultato, andando a definire quindi due community che vanno necessariamente tenute d'occhio: la 2 e la 1.

Questo perché, in modo quasi analogo ai Dozen, le community contengono gli utenti che li retwittano, come mostrato nella figura seguente:

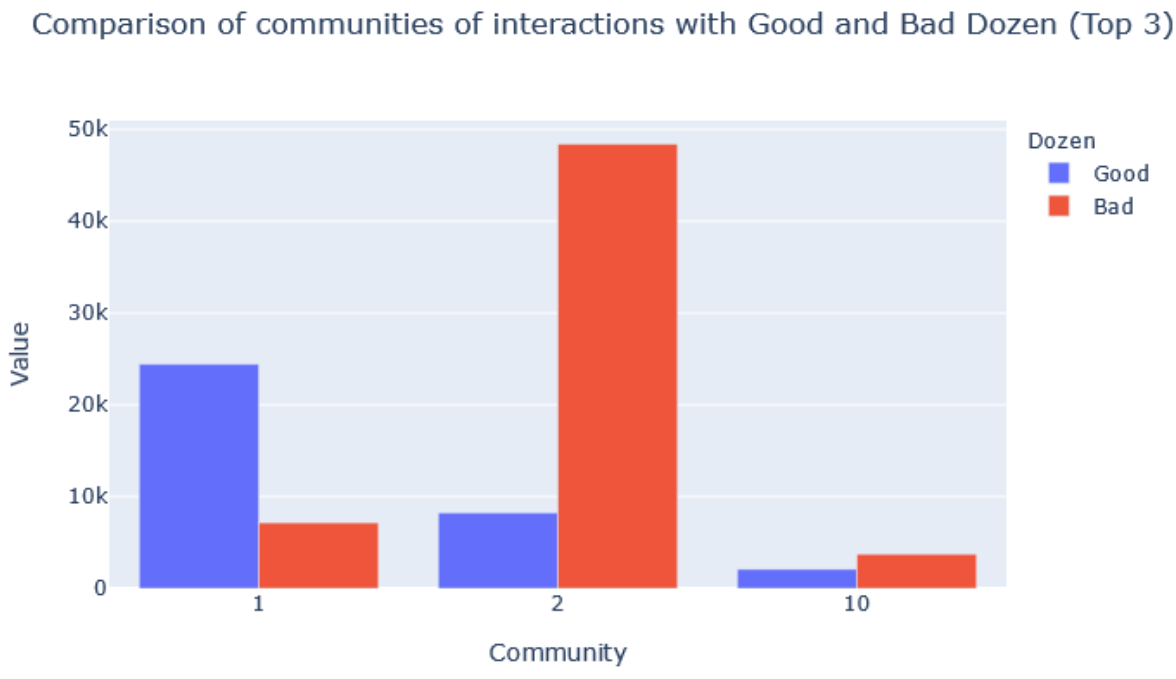


Figura 27: Differenza delle community degli utenti che retwittano i Dozen

Questo suggerisce che le due community contengano utenti che effettuano corretta informazione (nel caso della community 1) e cattiva informazione (nel caso della community 2) nonostante la stragrande maggioranza dei “Good Dozen” sia appunto nella community numero 2.

4 Piani di lavoro

Il lavoro è stato svolto con controllo periodico da parte del relatore e correlatore, effettuando meeting a cadenza settimanale o bimensile in funzione dei risultati e delle attività svolte.
Per mostrare il codice e gli eventuali grafici sono stati organizzati dei report sottoforma di presentazione, realizzati con Google Presentazioni.

4.1 Slack

Per poter mantenere un ottimo grado di conversazione si è fatto uso del software Slack, strumento di collaborazione utilizzato per inviare messaggi in modo istantaneo ai membri del team.
Grazie all'opportunità di inviare messaggi diretti, usare gruppi e condividere qualsiasi tipo di file, Slack è stato essenziale per lo svolgimento del progetto fornendo supporto.

4.2 GitHub

Il repository è stato uno strumento cruciale per quanto riguarda la sincronizzazione del lavoro e il mantenimento del codice aggiornato e ordinato.
Si è scelto di usare GitHub in quanto servizio di hosting per progetti software, usando come strumento di controllo versione Git.
L'utilizzo del repository è iniziato dopo la presa di confidenza con i dati, in particolare il 07/06/2021 è avvenuto il "first push".
Le attività hanno avuto una cadenza giornaliera, in funzione del lavoro svolto e delle attività non ancora svolte.
Il codice e i dati non sono pubblici, il repository è attualmente privato e l'accesso è consentito solo a chi ha l'autorizzazione di accedervi.

5 Conclusioni

5.1 Problemi riscontrati

Purtroppo, durante lo sviluppo del progetto sono stati innumerevoli i problemi che sono stati riscontrati, principalmente per limitazioni hardware legate alla grossa mole di dati utilizzata:

- **Saturazione del disco** dovuta all'enorme quantità di dati da elaborare durante la lettura dei file relativi ai tweets
- **Errori di out of memory** nell'elaborazione di dati sul DataFrame dei tweets dovuta alla notevole quantità di memoria che le varie liste e oggetti vanno ad occupare quando vengono utilizzate.
- **Limiti di Botometer** legati al numero di chiamate giornaliere, sono infatti limitate a 2.000 richieste ogni 24 ore, questo ha portato a limitare lo score di utenti a un ridotto gruppo secondo dei parametri scelti in fase di implementazione.
- **Modifiche della risposta di Botometer** che ha portato una serie di rallentamenti durante l'operazione di raccolta degli score, è stata infatti aggiornato un campo nella risposta http portando una serie di errori che hanno richiesto la modifica dello script che si occupa della raccolta degli score.
- **Dimensione della retweet network** troppo grande per l'hardware a disposizione, infatti la generazione della network con relativa classificazione di community ha occupato oltre 50 gigabyte di RAM. L'operazione è stata svolta su una macchina dedicata ad alte prestazioni.

5.2 Risultati

Tramite questo studio è stato possibile poter classificare gli utenti in funzione della loro posizione all'interno della discussione da Covid-19 su Twitter. Sono stati individuati i principali attori della disinformazione e i corrispettivi attori della buona informazione, andando a ottenere preziose informazioni sui domini condivisi, le strategie di informazione e gli hashtag utilizzati.
È stato possibile individuare i bot in gioco, andando a dare un interessante classificazione di questi e identificarli nella discussione, sono infatti principalmente impegni in operazioni di condivisione delle informazioni senza particolare interesse nell'influenzare la discussione.

Partendo da queste informazioni è possibile poter individuare e agire contro le azioni che portano alla propagazione e proliferazione di fake news di qualsiasi genere legate alla discussione sul Covid-19. Questo darebbe la possibilità di migliorare la qualità delle informazioni in circolo, limitando pertanto la visibilità che gli utenti scorretti hanno. Tutto ciò può giocare un ruolo importante anche sulla salute dei cittadini che usano i social network per informarsi.

5.3 Implementazioni future

Il progetto è destinato ad essere portato avanti, sarà parte infatti di ulteriori studi, tra cui alcuni attualmente in corso.

È di indubbia importanza lo studio delle singole network per capire dove si posizionano “Good Dozen” e “Bad Dozen”, dal momento che condividono la stessa community.

Lo studio è attualmente limitato al periodo di tempo che va da gennaio 2020 a maggio 2020, sarà quindi indispensabile effettuare la raccolta dei dati del periodo successivo fino ai giorni d’oggi per capire come le varie entità mutano il loro comportamento, analizzando quindi l’evoluzione della retweet network.

Sarà possibile capire se ci sono persone che hanno cambiato idea, in quale direzione si sono spostati e quando.

Un’altra attività, legata alla Twitter API, riguarda la ricerca degli utenti sospesi o rimossi: non sono stati presi in considerazione, infatti, questi utenti che possono potenzialmente dare informazioni interessanti per quanto riguarda i motivi di tali limitazioni da parte di Twitter e identificare quindi le fonti da cui vengono prese le informazioni da questi utenti.

6 Fonti

https://botometer.osome.iu.edu/api
https://pypi.org/project/botometer/#description
https://github.com/IUNetSci/botometer-python
https://cnets.indiana.edu/blog/2020/09/01/botometer-v4/
https://blog.quantinsti.com/detecting-bots-twitter-botometer/
https://rapidapi.com/developer/dashboard
https://networkx.org/documentation/stable/index.html
https://networkx.org/
https://docs.python.org/3/library/concurrent.futures.html
https://help.twitter.com/it/safety-and-security/public-and-protected-tweets
https://link.springer.com/article/10.1007/s42001-021-00139-3
https://www.counterhate.com/disinformationdozen
https://www.repubblica.it/esteri/2021/07/18/news/coronavirus_quella_sporca_dozzina_d_influencer_dietro_a_oltre_meta_delle_fake_news_sui_vaccini-310811514/
https://github.com/echen102/COVID-19-TweetIDs
https://help.twitter.com/it/managing-your-account/about-twitter-verified-accounts
https://en.wikipedia.org/wiki/Louvain_method
https://python-louvain.readthedocs.io/en/latest/
https://pythonspeed.com/articles/chunking-pandas/
https://www.kite.com/python/answers/how-to-read-specific-column-from-csv-file-in-python
https://pandas.pydata.org/pandas-docs/stable/user_guide/scale.html
https://pandas.pydata.org/pandas-docs/dev/reference/api/pandas.read_csv.html
https://publichealth.jmir.org/2020/2/e19273/