# A Social Network Analysis of COVID-19 Vaccines Tweets

Arbër Ceni[1], Alda Kika[1] and Denada Çollaku[1]

[1]University of Tirana, Faculty of Natural Sciences, Bulevardi Zogu I, Tiranë 1001, Albania

Abstract

The COVID-19 pandemic has been characterized by many controversies regarding the illness itself as well as the vaccination process. Social media platforms play a major role in spreading both valuable and scarce information. This paper aims to conduct a social network analysis of Twitter posts mentioning one of the three major vaccine producers. Twitter was chosen because of its user base, open API access and the vast amount of information spread. Data were collected daily from Twitter API 1.1 over a period of nine months from November 2021 to July 2022. Graph metrics, groups and node metrics were calculated using SNAP. The analysis is focused on the most important nodes in the network ranked by betweenness centrality. For the highest ranked user, the content of his posts and the amount of engagements was analyzed. The results show that the highest ranked users are usually (not always) non-professionals who mainly post misinformation, fake-news or only negative true information about vaccines. Another interesting fact revealed is that some users are present in different datasets and their posts get engagements from users not speaking the same language. A more thorough analysis of this data using other techniques such as calculating the path of the information flow may reveal further valuable information.

**Keywords**

COVID-19, Twitter, Social network analysis, Pfizer, Moderna, AstraZeneca, vaccine

## 1. Introduction

The internet era we live in has massively contributed to the information spread and reach. A concerning result is the attraction of users to fake news. Fake news is defined to be “fabricated information that mimics news media content in form but not in organizational process or intent” [1]. Other terms used in the same context are “misinformation” and “disinformation”. Misinformation can be regarded as simple false information, while disinformation is a false information created and spread with the purpose of misleading people [1]. Different studies have pointed out that these information disorders pose a challenge to the public health [2, 3, 4] and [5].

Infodemic is a new term used extensively during the COVID-19 pandemic and is defined by WHO as false information related to a disease [6]. After a period of non-pharmaceutical interventions such as lockdowns, restrictions of movement and social gatherings, face masking, physical distancing [7] the development and widespread of different kinds of COVID-19 vaccines have brought great hope for the world. Very soon anti-vaccination movements and vaccination hesitancy arose among several groups [8, 9]. Individuals that preferred media platforms such as social media, online forums or blogs reported more frequently hesitance toward vaccination than those that relied on source-verified platforms information [9].

Social media platforms have a major role in the spread of (mis/dis) information, fake news or anti-vaccine movements that spread rumors about the alleged dangers of vaccines [10]. In this paper, Twitter tweets associated with COVID-19 vaccines from a period from November 2021 until July 2022 are analyzed using graphs for each month and type of vaccine. For each graph that is created betweenness centrality is used to determine the top 10 users. Betweenness centrality is the number of shortest paths passing through a node, showing its influence over the flow of information in the network [11, 12] and [13]. Betweenness centrality is very important in the analysis of social networks. It can be used to measure the influence of the vertex or the person in a social network.

The paper consists of four further sections: Section 2 reviews recent studies on social network analysis related to vaccination for COVID-19. The data collection, structure and storage that is used in this study, as well as the methodology that is used to create graphs, is presented in Section 3. In Section 4, the network analysis metrics and results are discussed. The conclusion and a concise summary conclude this paper.

## 2. Related work

The COVID-19 pandemic has caused a global infodemic through social media networks. Social network analysis is used to understand the structure and dynamics of these

*Proceedings of RTA-CSIT 2023, April 26–27, 2023 Tirana, Albania*
arber.ceni@fshn.edu.al (A. Ceni); alda.kika@fshn.edu.al (A. Kika); denada.xhaja@fshn.edu.al (D. Çollaku)
https://fshn.edu.al/departments/departamenti-i-informatikes/personel/msc-arber-ceni (A. Ceni); https://fshn.edu.al/departments/departamenti-i-informatikes/personel/prof-asoc-dr-alda-kika (A. Kika); https://fshn.edu.al/departments/departamenti-i-informatikes/personel/dr-denada-collaku (D. Çollaku)
0000-0003-1373-1438 (A. Ceni); 0000-0002-3598-7285 (A. Kika); 0000-0002-6398-5156 (D. Çollaku)

**Table 1**
Terms and their respective number of collected posts

| Term | Number of posts |
|---|---|
| moderna | 3009645 |
| pfizer | 3311945 |
| AstraZeneca | 1224705 |
| vaccine | 3138370 |
| coronavirus | 3138830 |
| COVID-19 | 2903138 |
| covid | 3567153 |
| vaksina | 4170 |
| **TOTAL** | **20 297 956** |

networks and how misinformation spreads within them.

Carvalho et.al [14], applied topic discovery to tweets with geo information extracted from the COVID-19 vaccination theme. They used the Latent Dirichlet Allocation for topic modeling, and clustering for visualization, with the t-SNE, enabling a more detailed view of the distribution of topics and polarities, according to the number of tweets, in time and Brazilian geographic space. The analytical process provided a framework containing a set of tools to deliver information that can help authorities to understand the evolution of public opinion on vaccination and identify cities with significant numbers of posts according to the extracted topics [14].

A study by Olszowski et al [15], conducted a social network analysis on Twitter discussions regarding mandatory COVID-19 vaccination in the polish speaking community by using NodeXL to generate betweenness centrality and network clusters, and Clauset–Newman–Moore algorithm to identify two important groups of users. The results of this study revealed a substantial degree of polarization, a high intensity of the discussion, and a high degree of involvement of Twitter users [15].

## 3. Data and Methods

In the following sections we show how and what data are downloaded from Twitter and how they are stored so they can later be retrieved and analyzed.

### 3.1. Data

Motivated by the large amount of social media discussion that the COVID-19 pandemic and vaccines have produced [15, 16], we focused our research on terms related to these topics. We identified eight terms to be of interest (Table 1).

#### 3.1.1. Data Collection

For these terms, a twitter search need to be conducted in order to identify all posts (tweets) that contain the term. The Twitter API 1.1 was used to search Twitter and download the data [17]. The API had different access levels ranging from free to Premium [1]. The free access level allowed to search for posts created in the last 7 days up to a maximum of around 18000 posts per search. To gather as much data as possible without being rate limited, we decided to perform data collection for each of the terms once per day every day. For this purpose, we built a console application which would authenticate with Twitter API using OAuth 1.0a [18], search for tweets containing the specified search term, get information for the users involved in the conversation and store everything into a database. We setup a task in Task Scheduler to run this application for each of the terms, so in total we configured eight tasks equally distanced through the 24 hour period.

#### 3.1.2. Data structure

Twitter returns the queried data using the JSON format. For each of the posts that contain the search term, the API returns a JSON object, whose fields were used to build the graph. Below we show some of the fields used. For a complete list of fields and their detailed description, please refer to [19].

- **created_at** - Date/time when the post was created.
- **entities** - Hashtags, users and URLs contained in the post.
- **id** – Unique identifier of the post.
- **in_reply_to_screen_name** - The user which this posts replies to.
- **in_reply_to_status_id** - The post ID which this post replies to.
- **is_quote_status** - This post is a quote of another post.
- **lang** - Language of the post.
- **retweeted_status** – The original post that this post is a retweet of.
- **text** - The post content.
- **user** - A nested JSON object containing information about the author of the post.

For the user, the main fields used during graph creation were:

- **id** – Unique identifier of the user.

[1]At the time the data importer was built, Twitter allowed a free access level. At the time this paper was written, Twitter introduced the new access level system and revoked the free access level.

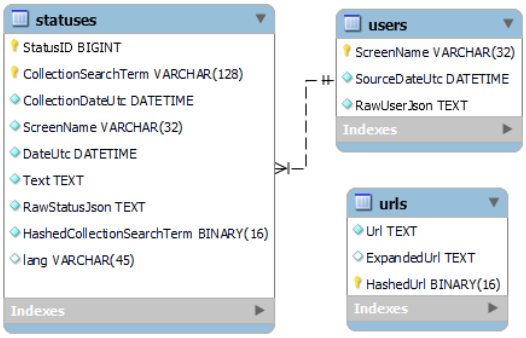


**Figure 1:** The data model

- **screen_name** – A username that the user likes to identify himself with.
- **name** – The display name of the user

For a complete list of fields for the user JSON object, please refer to [20].

#### 3.1.3. Data storage

To store the collected data into a database, we had two possible solutions:

1. Extract all the fields from the JSON object and save their values to their respective columns in the database table.
2. Save the whole JSON object in a text column.

The first solution would allow us to run faster and better-built queries. The second solution is more flexible regarding changes done by the API in the JSON structure. Given the fact that the database would be used mainly as a storing mechanism, most of the queries would use either the date or the search term, and the introduction of Twitter API 2.0 which in fact changed the JSON structure, we chose to apply solution number 2. MySQL was chosen as a database management system for its flexibility, price and the ability to partition tables. The structure of the database tables is shown in Figure 1.

The amount of collected data and the fact that the whole JSON object is stored in a single text column, has a negative effect on the performance of the queries. To compensate for this performance loss, we partitioned the table using HASH partitioning [21] on the "Hashed-CollectionSearchTerm" column since most of the search queries would be performed on this column. A subpartitioning by date would further benefit the performance since "DateUtc" is the second most used column in the search queries, but HASH partitions cannot be subpartitioned in MySQL [22].

### 3.2. Methods

In this section we show the technique used to build a network (graph) from the collected and stored data. Throughout the section we present challenges and solutions we have adopted.

#### 3.2.1. Relationships

By analyzing the JSON object of a post, we identified five different kinds of interactions or relationships between users.

1. **Mention** – A "mention" relationship will exist between User A and User B if User A has mentioned User B's username in the post that he has authored. Mentioning in Twitter is done by writing another user's username (i.e. @userb) in the posts' text.
2. **ReplyTo** – A "replyto" relationship will exist between User A and User B if User A has replied to a post authored by User B. A user can reply to another user by clicking the "Reply" button in a post or by preceding his post with a mention (i.e @userB...).
3. **Retweet** – A "retweet" relationship will exist between User A and User B if User A has retweeted User B's post by using the Retweet option in Twitter.
4. **MentionInRetweet** – A "mentioninretweet" relationship is a special case of the "retweet" relationship. Using the current classification technique, in the case where User A has retweeted User B and the post contains a mention of User C, then a "mention" relationship will be created between User A and User C. This is not entirely correct as User A has not directly mentioned User C, although there is a mention. To fix this type of problems, we introduced the "mentioninretweet" relationship which will have as source User A and as destination User C.
5. **Tweet** – A "tweet" relationship will exist between User A and himself if User A has authored a post which doesn't contain a "mention", is not a "replyto", "retweet" or "mentioninretweet". This relationship is considered as self-loop, where the source and the destination is the same.

An important thing to notice here is that a single post may generate multiple relationships.

#### 3.2.2. The Graph

Each of the relationships described above constitutes an edge with equal weight in our graph. The nodes of our graph are the users. Beside the required information to build a graph (*node* for nodes and *source* and *destination* for edges), other important information is included as well.

The data were collected from November 2021 until July 2022[2] and the number of posts collected for each of the terms are reported in Table 1. Given the vast amount of data collected, we decided to split them in a "by term" and "by month" basis. Building a huge graph consisting of all the collected posts would not allow us to run metrics and analyze the graph. The graph generation was done on an Intel Xeon Silver 4110 with 16GB of RAM.

A number of possible graph file formats were considered to be used for storing the generated graph. In particular we investigated GEXF [23], GDF [24], GraphML [25] and adjacency matrix. An adjacency matrix would not be suitable in our case for two fundamental reasons: the first is that with an adjacency matrix representation we are not able to include different attributes for nodes and edges and the second reason is that real-world graphs are known to be sparse [12] so we would end up with a sparse matrix which would need a lot of storage space. For these reasons we considered the other graph file formats which basically represent the graph as a list of edges, hence allowing to have attributes as well as a lower need for storage space. GDF is an open text file format used by the graph manipulation software GUESS [26]. It is well supported by other software like Gephi [27] and NodeXL [28]. The drawback of using this format is that not many graph libraries support it and has not been regularly maintained and updated. GraphML and GEXF are both XML-based file formats and both are supported by many graph manipulation software and libraries. We chose GraphML as the go-to format to save our graphs.

The general idea behind the graph generation algorithm is to get from the database all the records matching the search term provided by the user. From those records extract the authors of the posts and store them in a dictionary in memory. This is done to prevent duplicate users because there cannot be duplicate nodes in a graph. For each of the records, extract users that have been mentioned or replied to and add them to the dictionary as well if they are not present. Last but not least, for each record determine if the there is a "ReplyTo", "Retweet", "Mention", "Tweet" (self-loop) or it is a "MentionInRetweet" and create the appropriate edges.

Some challenges were encountered while executing this algorithm. A straightforward select query from the database resulted in a huge response time because of the RawStatusJson field which is set to be of type TEXT. This also posed a second challenge: it required more RAM memory than the machine had available just to load the data. Working with the data to create the graph would need even more memory. To overcome these challenges we implemented a technique to query smaller chunks of data from the database. This is controlled by the limit parameter. This technique introduced another challenge: a unique list of users needed to be kept in memory all the time in order to prevent duplicate users which would result in duplicate nodes and that is not allowed on a graph. Storing the whole JSON object in memory required a large amount of memory which was fixed by deserializing the JSON object into programming language object representations called TwitterUser and TwitterStatus. The last challenge was related to the time required by the algorithm to run. Our initial code used a single core from a 16-core CPU to run. To make the algorithm run faster, we made use of the parallelization features that C# offers and changed all the loops, except the loops responsible for creating the GraphML file because read-write is not a thread-safe operation, to run in parallel. After these changes, a single graph could be created in a matter of hours compared to never completing.

[2]The collection is still active. The dates reported here belong to the data analyzed.

**Table 2**
Graphs size for "**Pfizer**"

| Month | Nodes | Edges |
|---|---|---|
| November 2021 | 367294 | 731875 |
| December 2021 | 278266 | 543490 |
| January 2022 | 198258 | 451691 |
| February 2022 | 229197 | 673992 |
| March 2022 | 244193 | 717114 |
| April 2022 | 203168 | 608647 |
| May 2022 | 210783 | 626985 |
| June 2022 | 172249 | 526720 |
| July 2022 | 40060 | 75511 |

#### 3.2.3. Results

To further study the user polarization and (mis)information spread, we selected three terms: **Pfizer**, **Moderna** and **AstraZeneca**. For these terms and for the nine months period under consideration (November 2021 – July 2022), we constructed 27 graphs in total. Table 2, 3 and 4 show the size of the constructed graphs.

## 4. Analysis

In this section and the following subsections we present different metrics regarding the generated graphs. Each subsection contains an analysis of the graphs for the terms Pfizer, Moderna and AstraZeneca respectively. All calculations were done using SNAP [29].

**Table 3**
Graphs size for "**Moderna**"

| Month | Nodes | Edges |
|---|---|---|
| November 2021 | 358701 | 602031 |
| December 2021 | 267581 | 430523 |
| January 2022 | 247518 | 413592 |
| February 2022 | 298231 | 600610 |
| March 2022 | 323143 | 634862 |
| April 2022 | 275389 | 569098 |
| May 2022 | 273315 | 629752 |
| June 2022 | 238628 | 494884 |
| July 2022 | 56759 | 79718 |

**Table 4**
Graphs size for "**AstraZeneca**"

| Month | Nodes | Edges |
|---|---|---|
| November 2021 | 216976 | 390200 |
| December 2021 | 167942 | 288595 |
| January 2022 | 110906 | 192169 |
| February 2022 | 117994 | 213708 |
| March 2022 | 93629 | 176314 |
| April 2022 | 66175 | 201440 |
| May 2022 | 79209 | 222855 |
| June 2022 | 56726 | 130264 |
| July 2022 | 23721 | 39747 |

## 4.1. Pfizer

In Table 5 some overall graph metrics are presented. In Table 6, 7 and 8 the top 10 users (nodes) ranked by betweenness centrality are shown. A node with high betweenness centrality means that it is central to the flow of information in this network. The official pfizer account is expected to be in this list as everyone is talking about them so it gets mentioned or replied to very often. It is interesting to analyze what makes the other nodes so important.

In the November 2021 dataset user processic has the second highest betweenness centrality. As we will see in future sections, this user is part of other datasets as well. During this period the user has created 7 posts where 1 is a tweet, 5 reply-to and 1 retweet. The user posts useful information about vaccination centers and their location in his country. The user is from Thailand, has joined Twitter in 2008 and has 44000 followers. His posts created 19047 interactions, where 19041 were retweets, 1 mention and 5 reply-to. This means that nearly half of the user's followers engage with him.

In December 2022 user reuters is second in the list. The user posted 52 times during this period and their content were vaccine related news. These posts were retweeted 792 times, replied-to 60 times, mentioned 79 times and mentioned-in-retweet 170 times for a total of 1106 engagements. As we can see the engagements in this case are more evenly spread in regards to the different types of engagement than in other datasets. This might be because the user (reuters) is a news agency and is considered trustworthy. It would be interesting to perform a deeper analysis for paths [30] in this dataset and determine if there are more shares down the tree.

In January 2022 user erictopol is the second highest ranked user. He has authored 8 posts during this time, of which 5 were tweets and 3 were tweets with mentions. His posts mainly discuss the vaccine efficacy for the new (at that time) COVID-19 variant Omicron and always contain a reference to an article. These posts have produced 199 engagements, of which 149 were retweets, 9 reply-to, 34 mention and 7 mention-in-retweet. An interesting fact is that there are not many direct interactions with this user to justify the high betwenness. We believe the high betwenness might be attributed to the possibility that there might be second or third level spreaders of the information. To confirm or deny this belief a further analysis using paths is needed.

In the February 2022 dataset the user with the second highest betweenness centrality is pokrath. He is a doctor and has a Twitter verified account. He has tweeted 9 times and the content of those tweets is about vaccine efficacy for the Omicron variant. 156 other users have retweeted these tweets and 1 has replied to. Also in this case we can notice that a small amount of engagement has produced a high betweenness.

In March 2022 user lakovosjustice is ranked second in the list. The account is now suspended, but at the time of the import it had 28166 followers, was created in October 2021 and had authored more than 3000 tweets in such a short time. In the dataset, this user authored 9 tweets and all of them contain anti-vaccine claims with most not having any source of information. These tweets generated 4792 engagements, with 4747 retweets, 27 reply-to, 16 mentions and 3 mention-in-retweet.

In April 2022 we notice user jakeshieldsajj being second in the list. He is a form MMA/UFC world champion and he has a verified account with more than 300k followers. During this period he has authored 3 tweets complaining about Twitter suspending user accounts that created negative posts about the Pfizer vaccine. These tweets got retweeted 4928 times, mentioned 2 times and replied to 16 times.

In May 2022 user kwagular has the second highest betweenness centrality. The user has made 2 posts stating that a new document has emerged from Pfizer that suggests to not breastfeed after vaccination and baby formula is running out. A simple fact check in fact checking sites[3] reveals that these claims are not true. However, these

[3]We used https://www.factcheck.org/ and https://toolbox.google.com/factcheck/explorer/

**Table 5**
Overall metrics for "**Pfizer**"

| Month | Connected Components | Diameter | Avg. Geodesic Distance | Graph Density | Modularity | Number of Groups |
|---|---|---|---|---|---|---|
| November 2021 | 60854 | 25 | 5.885073 | $4.3 \cdot 10^{-6}$ | 0.174675 | N/A |
| December 2021 | 28186 | 22 | 6.054619 | $5.5 \cdot 10^{-6}$ | 0.709607 | 1715 |
| January 2022 | 23052 | 23 | 5.97937 | $8.7 \cdot 10^{-6}$ | 0.657333 | 10325 |
| February 2022 | 22076 | 23 | 5.41669 | $9.3 \cdot 10^{-6}$ | 0.204962 | 9995 |
| March 2022 | 19081 | 22 | 5.310234 | $9.3 \cdot 10^{-6}$ | 0.630077 | 8991 |
| April 2022 | 14574 | 20 | 5.161476 | $1.1 \cdot 10^{-5}$ | 0.620909 | 7422 |
| May 2022 | 12958 | 18 | 4.845637 | $1.1 \cdot 10^{-5}$ | 0.600138 | 5710 |
| June 2022 | 10271 | 18 | 4.919369 | $1.3 \cdot 10^{-5}$ | 0.610528 | 4641 |
| July 2022 | 4026 | 22 | 6.327091 | $3.7 \cdot 10^{-5}$ | 0.7011 | 2380 |

**Table 6**
Top 10 users by Betweenness Centrality for "**Pfizer**"

| Nov 2021 | Dec 2021 | Jan 2022 |
|---|---|---|
| pfizer | pfizer | pfizer |
| processic | reuters | erictopol |
| dominatkung | rwmalonemd | chonabisy |
| ariehkovler | plobjai | drjohnb2 |
| abdulmalig | chain_plus | jaysaran |
| bokuwa_kumaa | manopsi | michaelpsenger |
| bmj_latest | eig_banphot | pdubdev |
| nytimes | pokrath | france3provence |
| reuters | ezralevant | f_philippot |
| us_fda | sasha_twt | us_fda |

**Table 7**
Top 10 users by Betweenness Centrality for "**Pfizer**"

| Feb 2022 | March 2022 | April 2022 |
|---|---|---|
| pfizer | pfizer | pfizer |
| pokrath | lakovosjustice | jakeshieldsajj |
| suddhi2 | techarp | f_philippot |
| bokuwa_kumaa | manopsi | techarp |
| f_philippot | somorangi_e | blemontd |
| drsimonegold | f_philippot | drpacomoreno1 |
| lakovosjustice | afshineemrani | merissahansen17 |
| lereveildatlas | follforfight | manopsi |
| disclosetv | bubblesaii_ | cotrsmo |
| koathegreat | verity_france | mazagan_ft |

**Table 8**
Top 10 users by Betweenness Centrality for "**Pfizer**"

| May 2022 | June 2022 | July 2022 |
|---|---|---|
| pfizer | pfizer | pfizer |
| kwagular | bluewoodhomes | tuckercarlson |
| loffredojeremy | clarecraigpath | julesbw58 |
| f_philippot | maajidnawaz | feriglesias |
| inconforme75 | disclosetv | blemontd |
| lakovosjustice_ | osmosis8989 | marycbanegas2 |
| osmosis8989 | rmconservative | f_philippot |
| jamiesale | toadmeister | sergiodde |
| chriscottonstat | vprasadmdmph | robertkennedyjr |
| theofleury14 | prisonplanet | anibinani |

posts got 5792 engagements, of which 5698 are retweets, 43 reply-to, 49 mentions and 4 mention-in-retweet. The user has less than 1500 followers and has joined Twitter since 2013, so it is interesting to know who the users that produced so many engagements are?

In the June 2022 dataset, user bluewoodhomes is second in the list. It is interesting to notice that the account is not active anymore, but at the time of import the user had 996 followers and joined Twitter in 2011. The user has authored 3 tweets in the dataset and is claiming that his son had serious heart-related adverse events after the second dose of the Pfizer vaccine. These posts got 5755 engagements, of which 5590 were retweets, 73 reply-to, 81 mentions and 11 mention-in-retweet. Given the low number of followers for this user, it is interesting to know how the other users engaged with this tweet. Although there is no way to determine if the user's claim is true or false, the fact that the account does not exist anymore make these claims suspicious.

In July 2022 user tuckercarlsson has the second highest betweenness centrality. The user is a well-known journalist. He has authored a single post where he mentions Pfizer, but is not related to the vaccine but to medications in general. This tweet has produced 739 engagements with 686 being retweets, 27 mentions, 11 reply-to and 15 mention-in-retweet.

As we can see from this analysis, betweenness centrality is a good indicator of importance in a network. Users with high betweenness centrality mean that other users are engaging with them many times. We can notice that among these users are professionals who responsibly share true data, but there are also other users who like to share and amplify fake-news or disinformation. Another interesting fact is the presence of user drjohnb2 in the top 10 users list for January 2022. We will see this user being present in other datasets as well.

**Table 9**
Overall metrics for "**Moderna**"

| Month | Connected Components | Diameter | Avg. Geodesic Distance | Graph Density | Modularity | Number of Groups |
|---|---|---|---|---|---|---|
| November 2021 | 50752 | 30 | 6.319129 | $3.7 \cdot 10^{-6}$ | 0.726029 | 18031 |
| December 2021 | 35512 | 27 | 6.444243 | $4.7 \cdot 10^{-6}$ | 0.745883 | 15802 |
| January 2022 | 37432 | 26 | 6.545534 | $5.2 \cdot 10^{-6}$ | 0.727311 | 17053 |
| February 2022 | 39358 | 30 | 6.03299 | $5.3 \cdot 10^{-6}$ | 0.6857 | 17683 |
| March 2022 | 48520 | 30 | 6.409782 | $4.4 \cdot 10^{-6}$ | 0.677765 | 19957 |
| April 2022 | 37561 | 38 | 6.592767 | $5.5 \cdot 10^{-6}$ | 0.681876 | 16784 |
| May 2022 | 38615 | 30 | 6.360262 | $5.9 \cdot 10^{-6}$ | 0.643267 | 17674 |
| June 2022 | 31384 | 30 | 6.206134 | $6.6 \cdot 10^{-6}$ | 0.684548 | 13790 |
| July 2022 | 7294 | 30 | 7.004717 | $2.1 \cdot 10^{-5}$ | 0.793278 | 4076 |

**Table 10**
Top 10 users by Betweenness Centrality for "**Moderna**"

| Nov 2021 | Dec 2021 | Jan 2022 |
|---|---|---|
| moderna_tx | moderna_tx | moderna_tx |
| processic | plobjai | nomoretimecafe |
| allaboutsadden | manopsi | sarapayarom |
| bokuwa_kumaa | drericding | jaysaran |
| disclosetv | ezralevant | zornitsaxx |
| plobjai | gobgbby | realcandaceo |
| sputnikvaccine | cieseksandra | pokrath |
| pran2844 | pfizer | jordanschachtel |
| angelwansa66 | bolardear | jikkykjj |
| reuters | jordanschachtel | fahyadak |

**Table 11**
Top 10 users by Betweenness Centrality for "**Moderna**"

| Feb 2022 | March 2022 | April 2022 |
|---|---|---|
| moderna_tx | moderna_tx | moderna_tx |
| jordanschachtel | louietraub | merissahansen17 |
| orwells_ghost_ | follforfight | yuzawn |
| craig_a_spencer | sensanders | elonmusk |
| p_mcculloughmd | manopsi | ryan_wigand |
| perpetualmaniac | reuters | tomtsec |
| sbancel | faesq3639 | manopsi |
| donaldjtrumpjr | theirberge | disclosetv |
| pfizer | pfizer | zimermanricardo |
| manopsi | nytimes | pfizer |

**Table 12**
Top 10 users by Betweenness Centrality for "**Moderna**"

| May 2022 | June 2022 | July 2022 |
|---|---|---|
| inconforme75 | moderna_tx | moderna_tx |
| pkolding | us_fda | ratchakorn |
| moderna_tx | unrulycat2511 | seoul_cafe |
| pfizer | itsmylifech | navynblue |
| robertkennedyjr | calcarneiro85 | buscadorjos |
| adversereports | sensanders | maelvirasalazar |
| realmonsanto | tracklist | oregakitaworld |
| us_fda | pfizer | l_rinanz__ |
| mariolysosap | sailorrooscout | agusantonetti |
| jtrianat | beeddos | vegatorressas |

## 4.2. Moderna

Some overall metrics for the constructed graphs for Moderna are shown in Table 9. We will now consider the top 10 users ranked by betweenness centrality for the generated graphs for the "Moderna" term displayed in Table 10, 11 and 12.

In the November 2021 dataset, we once again have user processic ranked second, same as in the Pfizer dataset for the same period. During this period this user has authored 39 posts mainly sharing information regarding mRNA vaccines. These tweets have produced 22196 reactions, of which 22144 are retweets, 12 reply-to, 3 mention, and 5 mention-in-retweet. As we can see once again the main type of engagement in this case is retweet.

In December 2021 another user from Thailand, plobjai, is second in the list of highest betweennes centrality. This user has authored 6 tweets mainly giving information about how the vaccination process works in Thailand and notifying his followers that he is going to get vaccinated. These tweets got 16655 reactions of which, 16422 are retweets, 226 reply-to and 1 mention. In this case the reply-to number is higher than for user processic in the previous dataset.

In January 2022 user nomoretimecafe has the second highest betweenness centrality. The interesting fact in this case is that this user has only 374 followers and has authored 3 posts. In one of his tweets the user asks for people who have extra Moderna vaccines to donate them. This tweet got 11027 engagements, all of them retweets.

For the February 2022 dataset we will consider the second user in the list given the fact that the first user in the list is moderna_tx and it is expected to be so since the conversation is around them and many users engage with them. User jordanschachtel has authored 21 posts during this period. He has a verified account with 255000 followers and mainly posts against big pharma companies and vaccines. This user claims that the approved vaccines are not the same as the emergency approved ones and no citizen has access to the approved vaccines. His posts produced 9303 engagements of which 9105 retweets, 107 reply-to, 31 mentions and 61 mention-in-retweet.

In March 2022 user louietraub has the second highest betweenness centrality. He describes himself as an advocate for vaccines injury and during this period he authored 22 posts talking about his own injuries after the second dose of the Moderna vaccine. His Twitter account is shadow-banned (not all his posts are visible) and his Facebook account is restricted. He has a total of 16000 followers and 8338 people engaged with his 22 posts. 8196 were retweets, 123 reply-to, 18 mentions and 1 mention-in-retweet.

In the April 2022 dataset user merissahansen17 is second in the list of users with highest betwenness centrality. During this period the user has authored 2 tweets. The first tweet is news about the CFOs of Pfizer and Moderna both resigning within 72 hours over vaccine safety. Performing a fact check on this claim we can notice that this news is false[4]. However, these tweets got retweeted 6461 times, replied-to 77 times, mentioned 15 times and mentioned in a retweet 16 times.

In May 2022 there are two users with betweenness centrality higher than moderna_tx. The first user is inconforme75. He authored 2 posts where the second one is a retweet of his first tweet. This post reports an accusation of Russia towards high USA officials, Pfizer and Moderna regarding bioweapons in Ukraine. A simple search on Google about this topic yields results from trustworthy media that this is fake-news[5]. This post got 4474 engagements, of which 4397 are retweets, 55 reply-to, 14 mentions and 9 mention-in-retweet. The second user in the list is pkolding. He has authored 4 posts during this period and in all of them he is pointing out the restrictions towards Moderna vaccine because of the heart-related problems. These tweets produced 9823 engagements, of which 9553 are retweets, 216 reply-to, 42 mention and 14 mention-in-retweet. What is interesting to notice here is the fact that this user got a high number of reply-to compared to other cases.

The June 2022 dataset captures a nice feature of the betweenness centrality. In this dataset we can notice user us_fda being second in the list and user unrulycat2511 being in third place. Us_fda is the official FDA Twitter account and has tweeted about a committee meeting regarding emergency authorization for the Moderna vaccine. These tweets have produced 4430 engagements. User unrulycat2511 has tweeted about the new improved Moderna vaccine that claims to provide better and longer-lasting immunity response. This single tweet has produced 9145 engagements, double the number of us_fda, but still in the list user unrulycat2511 is positioned below us_fda. This is because betweenness centrality is not a mere representation of the amount of connections in a network, but rather a representation of a node's importance in the whole graph structure.

User ratchakorn is ranked second in the top users list for July 2022. He has authored 2 posts where he shows concern about some expired Moderna vaccines being used. On his second tweet he shares a document which states that the expiration date of the vaccines has been extended by 2 months. These tweets have produced 11825 engagements where 11823 are retweets and 2 are reply-to, despite the fact that the user has only 1300 followers.

### 4.3. AstraZeneca

Table 13 shows overall metrics about each constructed graph. In Table 14, 15 and 16 are shown the top 10 users (nodes) ranked by betweenness centrality. A node with high betweenness centrality means that it is central to the flow of information in this network. The official astrazeneca account is expected to be in this list as everyone is talking about them so it gets mentioned or replied to very often. It is interesting to analyze what makes the other nodes so important.

User processic has authored 38 posts in the November 2021 dataset, 37 of which are simple posts and one is a reply to another user. These posts have produced 26522 interactions where 26520 are retweets and only 2 are reply-to. As we can see the interaction is a form of a broadcast [31] where many users share a single user's post, but do not interact much with each-other. This user's posts are mainly informative posts with references to actual news[6].

In the December 2021 dataset we notice user drericding has the second highest betweenness centrality. By analyzing his interactions we notice that he has authored 21 posts in this dataset where 15 were tweets, 3 retweets of his earlier posts, 2 mentions and 1 reply-to. His tweets contain information about vaccine efficacy drop related to the newest (at that time) COVID variant, Omicron. Users interacted with these posts 6302 times, 6167 of which were retweets. 64 times users replied to, 40 times mentioned the user and 16 times retweeted his tweets while mentioning other users.

In January 2021 the first user in the list is drjohnb2. He is part of the top 10 users list in other datasets as well. For this dataset it is interesting the fact that this user has a higher betweenness centrality than astrazeneca. He authored 25 posts, of which 22 were tweets and 3 were

[4]https://www.newsweek.com/fact-check-pfizer-moderna-cfos-quit-within-72-hours-over-vaccine-safety-1701912

[5]https://www.npr.org/2022/03/25/1087910880/biological-weapons-far-right-russia-ukraine

[6]The user is from Thailand so we used Google Translate to translate his posts.

**Table 13**
Overall metrics for "**AstraZeneca**"

| Month | Connected Components | Diameter | Avg. Geodesic Distance | Graph Density | Modularity | Number of Groups |
|---|---|---|---|---|---|---|
| November 2021 | 25495 | 24 | 6.347963 | $6.4 \cdot 10^{-6}$ | 0.73373 | 8512 |
| December 2021 | 20387 | 25 | 6.079106 | $7.9 \cdot 10^{-6}$ | 0.727288 | 7372 |
| January 2022 | 16354 | 23 | 6.163895 | $1.2 \cdot 10^{-5}$ | 0.734273 | 6526 |
| February 2022 | 23857 | 27 | 6.450127 | $1 \cdot 10^{-5}$ | 0.694757 | 7221 |
| March 2022 | 26088 | 25 | 6.183913 | $1.3 \cdot 10^{-5}$ | 0.654513 | 6743 |
| April 2022 | 11668 | 23 | 6.244373 | $2.2 \cdot 10^{-5}$ | 0.463935 | 3868 |
| May 2022 | 6808 | 25 | 5.088193 | $2.6 \cdot 10^{-5}$ | 0.62252 | 2930 |
| June 2022 | 5979 | 25 | 5.439821 | $3.1 \cdot 10^{-5}$ | 0.648769 | 2530 |
| July 2022 | 4494 | 22 | 6.103577 | $5.1 \cdot 10^{-5}$ | 0.709285 | 1747 |

**Table 14**
Top 10 users by Betweenness Centrality for "**AstraZeneca**"

| Nov 2021 | Dec 2021 | Jan 2022 |
|---|---|---|
| astrazeneca | astrazeneca | drjohnb2 |
| processic | drericding | astrazeneca |
| reuters | ake2306 | alberto_yepes |
| viralvideovlogs | pokrath | tonyhinton2016 |
| drjohnb2 | viralvideovlogs | reuters |
| spectatorindex | joncoopertweets | liomont |
| mthai | drjohnb2 | otwieramy |
| danadjuto | telegraph | nuicemedia |
| kkgpst | hermanntertsch | chris_80f |
| dilleydilley8 | m_ebrard | momotchiii_ |

**Table 15**
Top 10 users by Betweenness Centrality for "**AstraZeneca**"

| Feb 2022 | March 2022 | April 2022 |
|---|---|---|
| astrazeneca | astrazeneca | astrazeneca |
| drjohnb2 | mateo85966574 | marianoalbert2 |
| tsererak | gaditanasinmor1 | funesta |
| rhodofansie | m_ebrard | santacarmelac |
| k4ats | arwen0506 | vaxreports1 |
| jeffyindy | robertkennedyjr | ex_infirmier |
| nathrnetn | tonyhinton2016 | ake2306 |
| drpaulofaria22 | drjohnb2 | julesserkin |
| spongebobcatz | nhs100k | ezralevant |
| follforfight | drhoenderkamp | storiesofinjury |

**Table 16**
Top 10 users by Betweenness Centrality for "**AstraZeneca**"

| May 2022 | June 2022 | July 2022 |
|---|---|---|
| astrazeneca | astrazeneca | astrazeneca |
| adversereports | buckyouhorses | bfntvv |
| rasekrasek | christo19725794 | chaknorris93 |
| jimenezlessons4 | gbnews | agoizs |
| stopnwo666 | storiesofinjury | papanours60 |
| thefreds | le_general_off_ | pe_1801 |
| gbnews | maajidnawaz | leeloo2022 |
| robertkennedyjr | oscasosraros | dean44559496 |
| bereaguilarv | vickyvarosario | bbcnews |
| buzzard89 | hannah_chesh_83 | pfizer |

retweets of his previous tweets. All of the tweets contain adverse events following vaccination and a reference to a published paper explaining the case. A total of 22 papers were cited, 14 of them were single report cases, 7 referred to thrombotic events[7]. There were 3106 engagements with these tweets and all of them were retweets. The creation date of his account is November 2021.

In February 2022 user drjohnb2 is again at the top of the list. During this time 34 posts were authored where 28 were simple tweets and 6 were retweets (shares) of his own earlier tweets. All of the tweets contain adverse events following vaccination and a reference to a published paper explaining the case. A total of 25 papers were cited, 12 of them contained reports of single cases. Twelve papers were referring to thrombotic events and one was not peer-reviewed. These posts produced 4341 interactions. 4287 of them are retweets. 2 are reply-to (one being in French), 14 mentions (one in Spanish and 12 from the same user) and 8 mention-in-retweet.

In March 2022 user mateo85966574 has the second highest betweennes centrality. During this period he has authored a single tweet in which he claims that the doctors have diagnosed a tumor in his body right after vaccination and he has retweeted a tweet from an account which is suspended. A total of 2239 users engaged with these 2 posts, of which 2209 were retweets, 29 reply-to and 1 mention. Another interesting fact is that this user account was created on December 2021 and the user describes himself as anti-COVID vaccination.

In the April 2022 dataset user marianoalbert2 is second in the list. He has authored a single tweet commenting his COVID-19 symptoms after three doses of the AstraZeneca vaccine. A total of 1261 users reacted to this

[7]By that time a number of countries had stopped the use of AstraZeneca COVID-19 vaccines.

tweet, 1245 of which retweeted it and 15 replied to. The account creation date is November 2018 and the user has posted tweets related to other subjects as well.

In May 2022 user adversereports can be spotted being second in the list. This user account was created in April 2022 and mainly reports about adverse events with the vaccines. During this time the user has authored 200 posts, usually within 5 min from each-other. These tweets produced 2719 engagements, 2486 of which are retweets, 132 reply-to, 11 mentions and 90 mention-in-retweet.

In the June 2022 dataset user buckyouhorses is second in the list of highest ranking users. This user has produced 10 posts with 86 interactions. The user account is now suspended, but the account creation date was 2009. In the description the user states that is a widow because her partner deceased after an AstraZeneca vaccination. These posts created 4134 engagements, 4017 were mention-in-retweet (because the original tweet contained mentions), 11 were reply-to and 102 were mentions. Unfortunately it is impossible to know the reason of an account suspension.

In July 2022 we notice user bfntvv has the second highest betweenness centrality. The user has tweeted one time stating that a BBC radio presenter has been deceased due to vaccination and a retweet of this tweet. This tweet has produced 1074 user engagements, of which 1065 are retweets, 5 reply-to and 4 mentions. A curious fact about this user account is that it is suspended, but the reported creation date in the dataset is April 2022.

As we can see most of these accounts had a recent creation date which makes one think that they were deliberately created to spread information for a certain purpose. Some of the accounts are already suspended and some are not. We noticed that the accounts not suspended are indeed sharing true information with proper referencing, but those cases are not the whole picture and while it is not dis/mis-information, it seems like the purpose of sharing only that information is to stop people from getting the vaccine. In the analyzed datasets there is a noticeable difference between the information shared by verified users (i.e. drericding) and other users (i.e. drjohnb2, adversereports, mateo85966574 etc.) even though both information can be regarded as "negative" with regard to vaccines. The same analysis can be further extended to include other users (nodes) in the top 10 list and check the information that those users share.

# 5. Conclusions

Twitter has been monitored for a period of nine months from November 2021 until July 2022 and posts regarding COVID-19 and vaccines have been collected and stored in a database. The main focus of this study have been posts related to the three main vaccine producers. From this collection of data we built 27 graphs by analyzing user engagements such as retweet, reply-to, mention and mention-in-retweet. For each of the constructed graphs we calculated overall metrics as well as betweenness centrality for each of the nodes in a graph. Betweenness centrality was used to rank users from the most to the least important in regards to his position in the network. An analysis of the shared content by the top ranked user has been conducted.

From this analysis we conclude that betweenness centrality is a good metric to distinguish the most central users in a network as it does not rely on the amount of engagements alone, but captures the network structure as well. The main form of engagement in Twitter was found to be retweet. Among the highest ranked users we found professionals who responsibly share trustworthy information, common individuals who try to help others as well as suspicious accounts who share fake-news, misinformation or real news noticing only the negative effect of the vaccines. From the later group of users we found that they are present in different datasets (i.e. drjohnb2). While most of the accounts that share fake-news are now suspended, the accounts that do share real but negative news are still active. Given the fact that they share real news they should not be suspended, but the information they share is not all the information there is. There is also a positive side of the story. Can deliberately sharing negative only information around a topic be classified as an information disorder? Should there be another term for this behavior? Should there be efforts to stop this behavior?

Suspending an account is a reactive response, meaning that those accounts got a lot of engagement at first and then got suspended. A proactive response would be more desirable. Being able to predict if a post is going to attract a lot of engagement and checking if that information is true or false so we can stop it before it spreads is a powerful feature to have. We aim to study the possibility of achieving this by using graph neural networks in future work.

# Acknowledgments

This research was funded by Agjencia Kombëtare e Kërkimit Shkencor dhe Inovacionit (National Agency for Scientific Research and Innovation). Its content is the author's responsibility and the opinions here expressed are not necessarily the opinions of NASRI.

# References

[1] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan,

G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, J. L. Zittrain, The science of fake news, Science 359 (2018) 1094–1096. URL: https://www.science.org/doi/abs/10.1126/science.aao2998. doi:10.1126/science.aao2998.

[2] I. J. Borges do Nascimento, A. Beatriz Pizarro, J. Almeida, N. Azzopardi-Muscat, M. André Gonçalves, M. Björklund, D. Novillo-Ortiz, Infodemics and health misinformation: A systematic review of reviews, Bulletin of the World Health Organization 100 (2022) 544–561. doi:10.2471/blt.21.287654.

[3] S. C. Briand, M. Cinelli, T. Nguyen, R. Lewis, D. Prybylski, C. M. Valensise, V. Colizza, A. E. Tozzi, N. Perra, A. Baronchelli, et al., Infodemics: A new challenge for public health, Cell 184 (2021) 6010–6014. doi:10.1016/j.cell.2021.10.031.

[4] A. Chiolero, How infodemic intoxicates public health surveillance: From a big to a slow data culture, Journal of Epidemiology and Community Health 76 (2022) 623–625. doi:10.1136/jech-2021-216584.

[5] E. Wilhelm, I. Ballalai, M.-E. Belanger, P. Benjamin, C. Bertrand-Ferrandis, S. Bezbaruah, S. Briand, I. Brooks, R. Bruns, L. M. Bucci, N. Calleja, H. Chiou, A. Devaria, L. Dini, H. D'Souza, A. G. Dunn, J. C. Eichstaedt, S. M. A. A. Evers, N. Gobat, M. Gissler, I. C. Gonzales, A. Gruzd, S. Hess, A. Ishizumi, O. John, A. Joshi, B. Kaluza, N. Khamis, M. Kosinska, S. Kulkarni, D. Lingri, R. Ludolph, T. Mackey, S. Mandić-Rajčević, F. Menczer, V. Mudaliar, S. Murthy, S. Nazakat, T. Nguyen, J. Nilsen, E. Pallari, N. Pasternak Taschner, E. Petelos, M. J. Prinstein, J. Roozenbeek, A. Schneider, V. Srinivasan, A. Stevanović, B. Strahwald, S. Syed Abdul, S. Varaidzo Machiri, S. van der Linden, C. Voegeli, C. Wardle, O. Wegwarth, B. K. White, E. Willie, B. Yau, T. D. Purnat, Measuring the burden of infodemics: Summary of the methods and results of the fifth who infodemic management conference, JMIR Infodemiology 3 (2023) e44207. URL: https://infodemiology.jmir.org/2023/1/e44207. doi:10.2196/44207.

[6] Infodemic, 2023. URL: https://www.who.int/health-topics/infodemic#tab=tab_1.

[7] N. Perra, Non-pharmaceutical interventions during the covid-19 pandemic: A review, Physics Reports 913 (2021) 1–52. doi:10.1016/j.physrep.2021.02.001.

[8] O. V. Ebrahimi, M. S. Johnson, S. Ebling, O. M. Amundsen, Ø. Halsøy, A. Hoffart, N. Skjerdingstad, S. U. Johnson, Risk, trust, and flawed assumptions: Vaccine hesitancy during the covid-19 pandemic, Frontiers in Public Health 9 (2021). doi:10.3389/fpubh.2021.700213.

[9] S. A. Nossier, Vaccine hesitancy: The greatest threat to covid-19 vaccination programs, Journal of the Egyptian Public Health Association 96 (2021). doi:10.1186/s42506-021-00081-2.

[10] K. Gunaratne, E. A. Coomes, H. Haghbayan, Temporal trends in anti-vaccine discourse on twitter, Vaccine 37 (2019) 4867–4871. URL: https://www.sciencedirect.com/science/article/pii/S0264410X1930876X. doi:https://doi.org/10.1016/j.vaccine.2019.06.086.

[11] M. Needham, A. E. Hodler, Graph algorithms: Practical examples in Apache Spark and neo4j, O'Reilly, 2021.

[12] A.-L. Barabási, M. Pósfai, Network science, Cambridge University Press, 2017.

[13] L. C. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1977) 35. doi:10.2307/3033543.

[14] V. D. de Carvalho, T. C. Nepomuceno, T. Poleto, A. P. Costa, The covid-19 infodemic on twitter: A space and time topic analysis of the brazilian immunization program and public trust, Tropical Medicine and Infectious Disease 7 (2022) 425. doi:10.3390/tropicalmed7120425.

[15] R. Olszowski, M. Zabdyr-Jamróz, S. Baran, P. Pięta, W. Ahmed, A social network analysis of tweets related to mandatory covid-19 vaccination in poland, Vaccines 10 (2022) 750. doi:10.3390/vaccines10050750.

[16] W. Ahmed, J. Vidal-Alaball, J. Vilaseca, A social network analysis of twitter data related to blood clots and vaccines, International Journal of Environmental Research and Public Health 19 (2022) 4584. doi:10.3390/ijerph19084584.

[17] Standard v1.1 | docs | twitter developer platform, 2023. URL: https://developer.twitter.com/en/docs/twitter-api/v1.

[18] Oauth 1.0a | docs | twitter developer platform, 2023. URL: https://developer.twitter.com/en/docs/authentication/oauth-1-0a.

[19] Tweet object | docs | twitter developer platform, 2023. URL: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet.

[20] User object | docs | twitter developer platform, 2023. URL: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/user.

[21] Mysql 5.7 reference manual :: 22.2.4 hash partitioning, 2023. URL: https://dev.mysql.com/doc/refman/5.7/en/partitioning-hash.html.

[22] Mysql 5.7 reference manual :: 22.6 restrictions and limitations on partitioning, 2023. URL: https://dev.mysql.com/doc/refman/5.7/en/partitioning-limitations.html.

[23] Gexf file format, 2023. URL: https://gexf.net/.

[24] Gdf format, 2023. URL: https://gephi.org/users/supported-graph-formats/gdf-format/.

[25] The graphml file format, 2023. URL: http://graphml.graphdrawing.org/.

[26] Guess: The graph exploration system, 2023. URL: http://graphexploration.cond.org/.

[27] M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks, Proceedings of the International AAAI Conference on Web and Social Media 3 (2009) 361–362. doi:10.1609/icwsm.v3i1.13937.

[28] M. Smith, A. Ceni, N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, C. Dunne, Nodexl: a free and open network overview, discovery and exploration add-in for excel 2007/2010/2013/2016, from the social media research foundation (2010). URL: https://www.smrfoundation.org.

[29] J. Leskovec, R. Sosič, Snap: A general-purpose network analysis and graph-mining library, ACM Trans. Intell. Syst. Technol. 8 (2016). URL: https://doi.org/10.1145/2898361. doi:10.1145/2898361.

[30] S. Goel, A. Anderson, J. Hofman, D. J. Watts, The structural virality of online diffusion, Management Science 62 (2016) 180–196. doi:10.1287/mnsc.2015.2158.

[31] M. Smith, B. Shneiderman, L. Rainie, I. Himelboim, Mapping twitter topic networks: From polarized crowds to community clusters (2014).