Università degli Studi di Padova

Department of Mathematics 'Tullio Levi-Civita'

Master of Science in Data Science

Analyzing social dynamics and linguistic features of online extremism: the case of Gab

Gabriele Etta

Supervisor: Prof. Mauro Conti
University of Padua

Co-Supervisor: Prof. Walter Quattrociocchi
Ca' Foscari University of Venice

October 2020

Abstract

The introduction of stricter restrictions imposed by social media platforms to contrast the arising of misinformation and hate speech produced a change in how extremism acts online, with the result of a migration of their communities to new niche platforms that impose fewer restrictions about the reliability of the posts produced and their political correctness. Among all the available niches, one in particular has experienced the biggest arising and coverage from news media: Gab. Gab is well known to be an extremist-friendly platform that performs little control on the posted content. Thus it represents an ideal benchmark for studying phenomena potentially related to polarization such as misinformation spreading. The combination of these factors may lead to hate as well as to episodes of harm in the real world. In this work[1], we provide a characterization of the interaction patterns within Gab around the COVID-19 topic. To assess the spreading of different content types, we analyze consumption patterns based on both interaction type and source reliability. Overall we find that there are no strong statistical differences in the social response to questionable and reliable content, both following a power law distribution. However, questionable and reliable sources display structural and topical differences in the use of hashtags. The commenting behavior of users in terms of both lifetime and sentiment reveals that questionable and reliable posts are perceived in the same manner. We can conclude that despite evident differences between questionable and reliable posts Gab users do not perform such a differentiation thus treating them as a whole. Our results provide insights toward the understanding of coordinated inauthentic behavior and on the early-warning of information operation.

[1]An extract of this work has been published as a research paper in SocInfo 2020 conference.

Contents

Chapter 1

Introduction

The evolution of the Internet ecosystem through the introduction of social media has massively changed how our society is currently living. Indeed, they offered a tool for reaching out to people more easily, being able to aggregate with others around specific topics of interest. In addition to this, the access to information and knowledge broke down many barriers, observing a decentralization of the news sources to the point that everyone can propose himself as a potential journalism authority [10, 44, 54]. This phenomenon, combined with the tendency of humans to behave irrationally, brought to light the problem of *misinformation*, to the point that the World Economic Forum listed in 2013 the misinformation as one of the main threats to our society [34]. The presence of this problem in the current historical moment, where data is becoming one of the primary sources for our everyday life, has already shown how it can be dangerous for our society, also due to the creation of communities under narratives that can be harmful [39]. The regulations imposed by social media platforms to contrast the arising of these communities that promote hate speech, racism, and other extreme behaviors led to the restriction and ban of many users, who reorganized themselves to find a new online platform in which they can act more freely. The arising of these places has become a serious problem for the online and offline ecosystem due to the freedom they have in promoting, sharing, and discussing ideas that can be dangerous for the political and social stability [57]. In many cases, those discussions both took the form of cyber attacks to specific entities or real acts of violence such as shootings, kidnappings, and similar [37, 50]. For these reasons, the arising of *terrorism* and *cyberterrorism* is becoming one of the biggest challenges of this century [25, 7], with the necessity to maintain the security for both online and offline realities.

In this thesis, we characterize the users' behavior in an extremist social media platform, called Gab. All social media platforms, indeed, present distinct features such as the type of content as well as user interaction options [15, 16]. Along with mainstream platforms like Facebook, Twitter, YouTube, and Instagram, niche ones such as Gab or Reddit have been created. These platforms strongly differ for various aspects mainly related to the political leaning of the user base and the content regulation policy implemented. The latter element has crucial importance, being strongly linked to the risk of opinion polarization and to the development of hating contents [15]. Within such niche media, biased information proliferates [17] and users, either interested in joining the community or banned by other social media, tend to develop the feeling of belonging to a group of like-minded individuals, i.e., an echo-chamber [24]. Hence, studying users' interaction patterns in platforms like Gab becomes of primary interest to shed light on the dynamics of content production and information spreading in such a segregated and unregulated environment. We then explore different aspects related to the spreading of both questionable and reliable contents in Gab by taking into account several aspects, such as the users' reactions, the topics embedded in hashtag networks, and the users' sentiment. In more detail, we first focus on the differences between interaction types in terms of frequency and time. Then we build statistically validated *hashtag co-occurrence networks* assessing the topological differences between questionable and reliable contents. Finally, we analyze the interplay between the sentiment of comments and commenting behavior concerning information source questionability.

The thesis is organized as follows:

- Chapter 2 describes the current state of the art of non-mainstream social media, focusing in particular on Gab;

- Chapter 3 describes the data set used for this work and explores its structure;
- Chapter 4 describes the mathematical tools behind the analysis;
- Chapter 5 describes how results are obtained;
- Chapter 6 summarizes the results and discusses the possible applications associated.

Chapter 2

Background and State of the Art

2.1 Misinformation and Social Media

Social media platforms play a crucial role in the public sphere, influencing the public debate on a wide range of topics including politics, health, climate change, economics, migration [6, 13, 8, 23]. Users online have shown a tendency a) to acquire information adhering to their system of beliefs [5], b) to ignore dissenting information [58], c) to form polarized groups around a shared narrative [21]. One of the dominating traits of online social dynamics, indeed, is polarization [51]. Divided into echo chambers, users account for the coherence with their preferred narrative rather than the true value of the information [24, 20]. This scenario creates the perfect incubator for information operations [14]. Among the most pressing issues is the spread of fictitious and low-quality information (e.g., fake news, rumors, hoaxes). Questionable means are often used to influence the public opinion toward polarization or to burst distrust in governments and institutions [24]. The spread of low-quality information is sometimes carried out by groups of coordinated or automated accounts that pollute our social environments by injecting and sharing a large number of targeted messages [26], i.e., what Facebook calls "coordinated inauthentic behavior" [30]. Although some studies focused on the interplay between false and real information [52], the main point to understand is how information fits into a larger disinformation campaign [48, 14]. Most of the information operations involve users which are not aware of their role but which may foster polarization and distrust toward science and mainstream journalism [4, 32, 46, 22].

2.2 Social Security and the growing of Niches

A wide research effort has been spent to characterize online information operations [14, 47, 29, 55, 33] especially in the case of terrorism [49, 36, 11]. Several works addressed the analysis of social behaviors to detect features to anticipate and thus inhibit information disorders [56, 51, 41, 12]. Most of the tactics tend to exploit the confirmation bias [24] of users to foment heated debates [53, 2]. Niche social media performing little regulation on their contents seem to be the ideal environment for triggering polarization dynamics that can turn into actions of harm in the offline world [19].

2.3 Aggregating online extremism: the case of Gab

Gab is an online social platform, describing itself as "A social network that champions free speech, individual liberty, and the free flow of information online. All are welcome" [27]. Such a claim, together with the political leaning of its founders and developers, made Gab the "safe haven" for the alt-right movement. Moreover, low moderation and regulation of contents have resulted in a widespread of hate speech and fake news. For these reasons, it has been repeatedly suspended by its service provider, and its mobile app has been banned from both App and Play stores [57]. In particular, Gab attracted the interest of researchers due to its permissive content regulatory policy and the political leaning of its users. In [40] authors analyze the content shared on Gab and the leaning of users, finding a rather homogeneous environment prone to share right biased content. Authors of [57] characterize

Gab in terms of user leaning and content shared, suggesting that it is more similar to a safe place for right-wing extremists rather than an environment where free speech is protected. Moreover, a topological analysis performed by authors of [15] reveals that Gab users appear as one quite homogeneous cluster biased to the right. Further, differently from other platforms such as Twitter and Facebook, in Gab there is a lack of users with leaning opposite to the most popular one. Overall, all these studies suggest that Gab can be considered as a homogeneous environment where biased content and misinformation may easily proliferate.

Chapter 3

Data set description and exploration

This chapter describes the data set used for this work and the analysis performed to extract meaningful insights. All the data refers to the period 01/01/2020 – 31/03/2020.
The structure of this chapter is the following:

- We describe the mechanisms behind the data set collection from Gab and the architecture that we implemented to perform this operation;

- We introduce an additional data set that provides a categorization of the news outlets based on the quality of the information they share;

- We provide an explanation of the preprocessing phase to prepare the data for the analysis and experiments;

- We conduct an Exploratory Data Analysis (EDA) on the collected data sets, showing the results obtained.

3.1 Data collection

3.1.1 Gab Corpus

The obtainment of the data sets started from the definition of the topic that we want to consider which, in our case, has chosen to be the COVID-19. The reason for this particular choice is two-fold. First, it has been demonstrated [45] how health-related topics are one of the most common scenarios for polarization, bringing a series of point of views in which, for the conspiracy faction, the narratives proposed may seriously put at risk the life of people due to their contents, as recently experienced with the COVID-19 where some pseudoscience sources proposed bleach shots as a ”miracle cure” for the virus [31]. This is strictly correlated with the second point that refers to the environment in which those conspiracy theories proliferates. Indeed, Gab's user-base and structure offer a place where there is a tendency to take these alternative theories as ground-truth, inverting the debate for which the opposition is composed of all those science-based opinions and facts.

After having defined the topic of our study, it is necessary to look for the most popular research keywords related to start the data collection process. To achieve this, we perform some researches on Google Trends by looking for the most common terms associated with the keyword *coronavirus*, retrieving the following words: *wuhan, china, coronavirusoutbreak, covid, covid19.*

The next step consists of retrieving all the posts from Gab concerning the topic expressed by the terms previously obtained from Google Trends. To achieve this, we build an API oriented pipeline, described in Figure 3.1, which is structured in the following way. First, we perform a Gab search with each term, obtaining a list of associated hashtags. We then inspect the results which are filtered based on the hashtags meaning. Then, we iterate over each hashtag to retrieve all the posts with their related comments. We end up with 116343 posts, associated with 26136 different hashtags, that received 96757 likes, 60007 comments and 60563 reblogs by 4293 users.

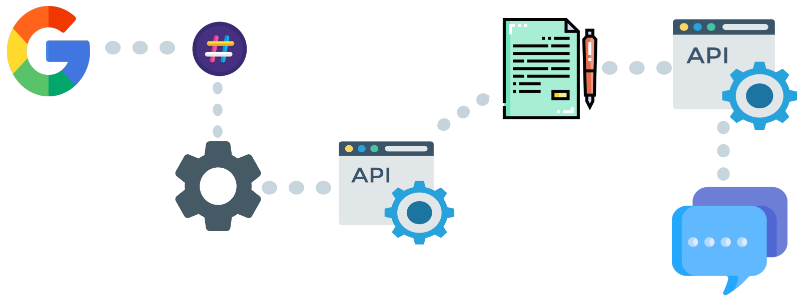


Figure 3.1: Representation of the API oriented pipeline. It starts by retrieving all the most trending research keywords associated with *coronavirus* term. Those keywords were used to obtain all the meaningful hashtags associated with them. After having obtained the list of hashtags, a SNAT technique was applied to parallelize the API calls for retrieving the posts associated with each hashtag. Iteratively, for each post obtained we called again another specific API that returned all the comments associated. In the end, we obtained two data sets: the first co mposed of all posts associated with the COVID-19 related hashtags, whilst the second is composed of all the comments associated with each post obtained.

Given the restriction imposed by Gab on the number of requests that can be accomplished from the same IP address, we make use of *Source Network Address Translation (SNAT)* technique which allows the modifying of the source IP address in order to be recognized from the receiver as a sender with a new identity. In this way, the collection of the information regarding the different elements can be done in parallel, thus being more faster and avoiding all those limitations that may affect the composition of our data set.

3.1.2 Questionable and Reliable Sources

To evaluate the quality of the information circulating on Gab, we employ a source-based approach by building a data set of news website domains, known with the term of *news outlet*, where each element is labeled either as "questionable" or "reliable" employing a classification procedure provided by a fact-checking organization called Media Bias/Fact Check (MBFC). MBFC has been frequently used for source classification [9, 3, 15]. It provides a classification determined by ranking bias in four different categories that are: Biased Wording/Headlines, Factual/Sourcing, Story Choices and Political Affiliation. A score is assigned to each category per each news outlet and the average score determined the bias of the outlet, as explained in the Methodology section of the website [42].

Each news outlet has associated a label that refers either to a political bias, namely, Right, Right-Center, Least-Biased, Left-Center and Left or to its reliability that is expressed in three labels, namely, Conspiracy-Pseudoscience, Pro-Science or Questionable. Noticeably, also the Questionable set includes a wide range of political biases, from Extreme Left to Extreme Right. For instance, the Right label is associated with Fox News, the Questionable label to Breitbart (the well-known extreme right outlet) and the Pro-Science label to Science. Using such a classification, we divide the news outlets into Questionable outlets and Reliable outlets. All the outlets already classified as Questionable or belonging to the category Conspiracy-Pseudoscience are labeled as Questionable, the rest is labeled as Reliable.

Considering all the 2637 news outlets that we retrieve from the list provided by MBFC we end up with 800 outlets classified as Questionable 1837 outlets classified as Reliable.

3.2 Data Preprocessing

The data sets obtained from the previous operations were not suitable for any kind of analysis at their initial state. As a consequence of this, a preprocessing phase is then necessary to clean and manipulate the data.

We start by transforming the text contained in posts and comments. After having converted all the letters into lowercase, we remove each element that is an HTML tag, email, website, or word composed only by digits, special

	Posts	**Comments**
Max	1977	1398
Min	25	1
Standard Deviation	604	314
Mean	640	454

Table 3.1: Summary of general descriptive statistics for the daily evolution of Posts and Comments.

characters referring to mentions or hashtags and stopping words. Furthermore, all the URLs in a short form are extended to their original formulation. In the end, we perform an expansion of all those verbs that are written in a contracted form (e.g: *isn't* is then transformed into *is not*). As a result, we obtain a data set of posts and comments whose text is only composed of meaningful words.
For what concerns the hashtags, we perform specific filtering on their names to remove all of those that are not identified as English. In addition to this, we also filter all those hashtags that were not similar with those used for the initial search by making use of the Levenshtein Distance, setting the number of maximum edits to 4.
After having obtained cleaned text contents, we associate each news outlet with their corresponding domain provided by MBFC dataset, therefore obtaining the outlet category. In the end, the final data set is composed of all the comments related to the different posts with the reliability of the associated news outlet.

3.3 Exploratory Data Analysis

The distribution of comments and posts over the analysis period, described in Figure 3.2, shows an interest from users that increases over time which is comparable to the growth of the pandemic all over the globe. Table 3.1 shows how users tend to prefer posting rather than commenting, with a daily average of 640 for the first and 454 for the second.

The fact that posting is preferred over commenting may be a proxy for

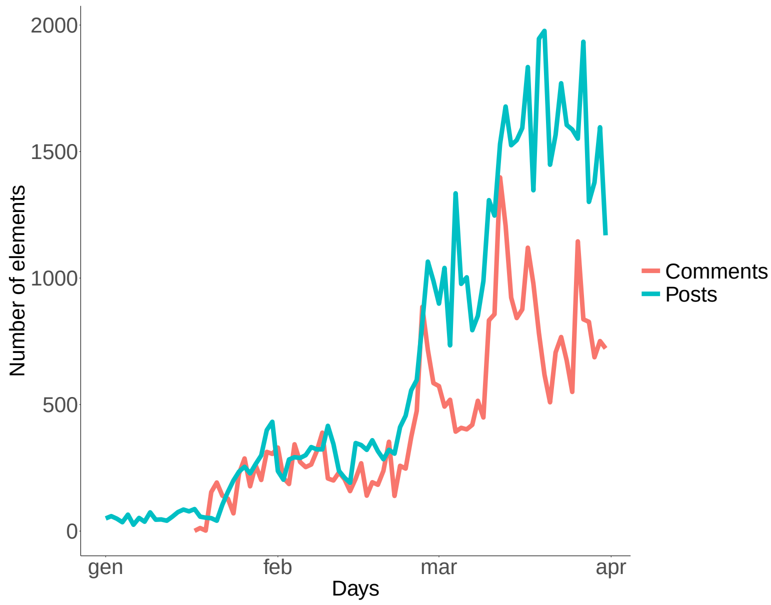
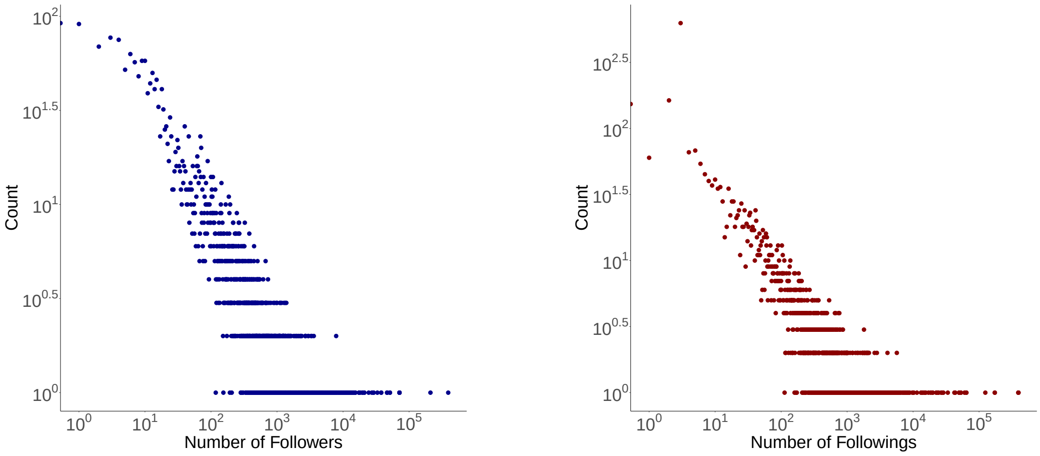


Figure 3.2: Daily distribution of posts and comments over the analysis period. The evolution of both types of contents is according to the spreading of the COVID-19 virus, providing evidence of the interest in the topic. The distribution also shows how posting is the preferred way of spreading content.

the propagation of narratives, which is strictly correlated with the number of *followers*, i.e., the number of users that follow another one, and *followings*, i.e., the number of users that a user is following. Therefore, Figure 3.3 describes the distribution of these two entities, which can be approximated with a power law distribution. Results show how the majority of users rely upon the interval of $[10, 10000]$ following and followers, being potentially influential concerning the message they share inside their posts.

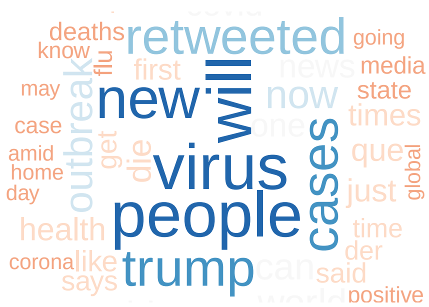
To further extend the analysis, we investigate the lexical content from a qualitative perspective, as described in Figure 3.4. We can see how comments and posts share a considerable amount of frequent words that revolve around the narrative of the pandemic. Words like *virus, people, trump* and *corona* reflect a common narrative that is likely based around the current political



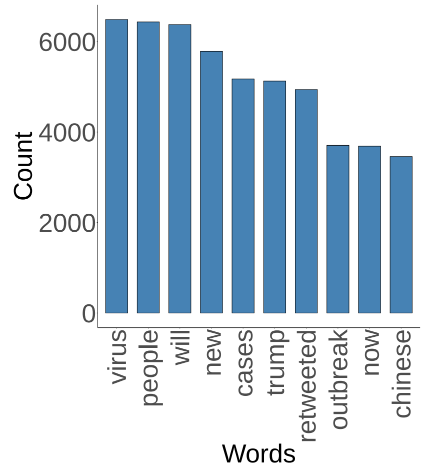
(a) (b)

Figure 3.3: Distribution of user's Followers (left) and Followings (right). For both of them, the interval $[100, 1000]$ is the most representative one in terms of users followed and for which they are being followed.

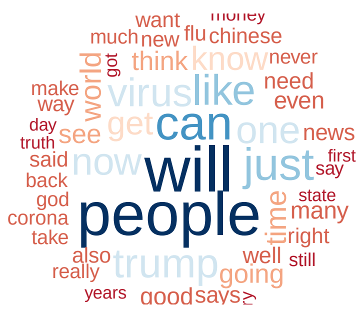
and social situation, especially in the USA since Gab user base is mainly composed of people from the US. This is also confirmed by looking at the most frequent words from both sources, showing how posts propose nouns that are more technical and more explanatory for the topic, such as *outbreak, chinese, cases.* Comments, instead, propose simpler terms that do not seem to describe any particular context.



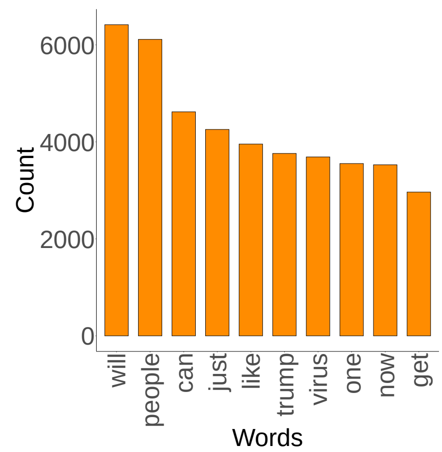
(a)



(b)



(c)



(d)

Figure 3.4: Qualitative and quantitative representation of most common terms used both in Posts and Comments. Upper Panel: Wordcloud representation of the most 50 frequent words used in posts (left) together with the frequency values of the first ten (right). Lower panel: same representation as to the upper one, but for comments. The analysis shows how most frequent words are strongly related to the social and political situation harmed by the growth of the pandemic. The presence of terms like *trump* or *chinese* may be correlated with the Gab community which is strongly USA based, referring to political debates which sees China as the intentional creator of the COVID-19 virus. In addition to this, the ten most frequent words in Posts are more topic explanatory compared with the one used in Comments, which are lexical simpler. This is due to the nature of the two sources: comments, which are written by users with generally no competence in journalism, will have a more elementary syntax and lexicon, whilst posts have to be more technical and professional due to the expected competence of the creators behind them.

Chapter 4

Preliminaries and Definitions

4.1 Networks

The basis for the conceptualization of a network is a graph $G = (V, E)$, being V the set of n nodes and E the set of m edges. The nodes are denoted as $i, j \in V$ or, similarly, $i, j = 1, \ldots, n$, and the edge that formalizes the connection between i and j is denoted as (i, j) .

With $\mathbf{A}$ we denote the adjacency matrix, a n-squared binary matrix taking values 0 or 1, where the element $A_{ij} = 1$ if nodes i and j are connected and $A_{ij} = 0$ otherwise; the degree of the node i is $k_i = \sum_j A_{ij}$, and it quantifies the number of neighbours of the node i; the number of links in the graph G is $m = \frac{1}{2} \sum_{ij} A_{ij}$. A graph that respects the last formalized equality, called Handshaking Lemma, is an indirect graph. Another instance that we take into account is the bipartite graph, which is a graph in which the vertex set V is the union of two disjoint independent sets called the partitions of G, as represented in Figure 4.1. The equivalent of an adjacency matrix for a bipartite graph is a $h \times p$ rectangular matrix called incidence matrix $\mathbf{B}$ that takes values 0 or 1, where the element $B_{ij} = 1$ if nodes i and j are connected.

A bipartite graph can be easily projected onto one of its partitions by performing an operation called one-mode projection that can be formalized in terms of the product $\mathbf{P} = \mathbf{B}^T\mathbf{B}$, in the case we are projecting onto the partition of size p, and $\mathbf{P} = \mathbf{B}\mathbf{B}^T$ if we are projecting onto the partition of

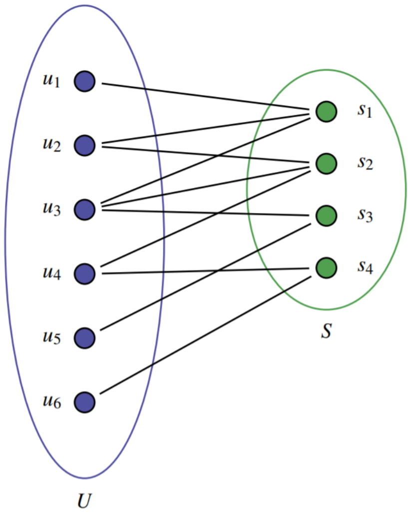


Figure 4.1: Representation of a bipartite graph. U and S represent two independent partitions of the original vertex set V.

size h. An example of a result from this operation is described in Figure 4.2. $\mathbf{P}$ is a symmetric matrix whose elements P_{ij} are nonnegative numbers that represent, in the case of off-diagonal elements, the number of common links of the nodes i and j to the partition of size h or p. The diagonal elements of the matrix $\mathbf{P}$ are also nonnegative numbers that represent the degree of the node in the bipartite graph. Since the elements on the diagonal of the matrix $\mathbf{P}$ have a different meaning with respect to the elements away from the diagonal, it is common practice to set the diagonal elements $P_{ii} = 0$. The matrix $\mathbf{P}$ after such a treatment can be also called the co-occurrence matrix since two elements are interconnected if they are co-connected to at least one node of the partition they don't belong to. Also, the number of co-connections between i and j is represented by the link weight, i.e., by the element P_{ij} of the matrix $\mathbf{P}$.

4.2 Hashtag Co-occurrence Network

Starting from the set $\mathcal{P}$ of all posts and the set $\mathcal{H}$ of the hashtags found in those posts, we create the bipartite network T, whose nodes are posts and

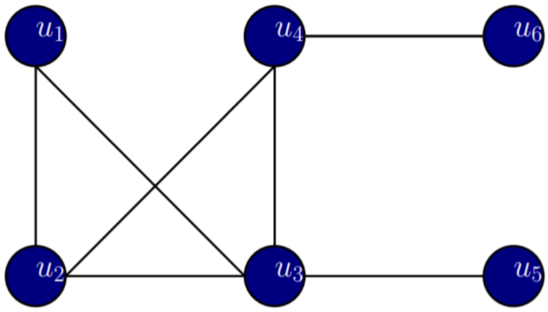


Figure 4.2: Projection of a bipartite graph composed by partitions U and S. The one-mode projection, in this case, projects U over S.

hashtags. A link between an hashtag h_i and a post p_i exists if the hashtag h_i is used inside the post p_i. From the bipartite network T we create the hashtag co-occurrence weighted network H by projecting the bipartite network T onto the partition $\mathcal{H}$. Additionally, we need to assign significance to the co-occurrence of two hashtags to represent only meaningful associations. This is strictly correlated to the computation of a statistical hypothesis test. In fact, in our case we want to test how likely it is that the association between two hashtags is observed more than a specific number of times, Therefore, given two hashtags α and β that independently occur in posts n_α n_β, we want to test how likely it is that they appear in more than r posts. This is equal to the computation of the probability

$$p = \sum_{k \geq r} p(k) \tag{4.1}$$

that we would expect to observe more than r co-occurrences by chance.

From the general idea of hypothesis testing between two hashtags, we define the probability function $p(\cdot)$. Let then $N = |P|$ be the number of posts and n_1 and n_2 the number of them in which the first and the second hashtags are found, respectively. To perform the statistical test, we first define the null model which in this case refers to the random and independent selection of these hashtags among the N total posts, obtaining the probability of k, i.e., the number of posts in which both hashtags appear. This random selection of n_1 and n_2 sets of posts out of N can be done in

$$\binom{N}{n_1}\binom{N}{n_2} \tag{4.2}$$

different ways. The number of coincidences k is therefore bounded between

$$\max\{0, n_1 + n_2 - N\} \leq k \leq \min\{n_1, n_2\}. \tag{4.3}$$

This means that we can have zero coincidences at least if the sum of $n_1 + n_2$ does not exceed N, obtaining $n_1 + n_2 - N$ otherwise, whilst the largest number of coincidences cannot exceed the smallest number of posts where a hashtag appeared, i.e., n_1 or n_2. To assess how many of these choices there are exactly k coincidences, we can classify four kinds of posts:

- k posts showing a coincidence
- $n_1 - k$ posts selected only in the first choice
- $n_2 - k$ posts selected only in the second choice
- $N - n_1 - n_2 + k$ posts not selected in any of the two choices

Therefore, the exact number is obtained from the computation of the multinomial coefficient

$$\binom{N}{k, n_1 - k, n_2 - k}, \tag{4.4}$$

which makes use of the definition $\binom{p}{q_1,\ldots,q_n} \equiv \frac{p!}{q_1!\ldots q_n!(p-q_1-\ldots q_n)!}$. Therefore, the probability $p(k)$ that exactly k hashtags coincide when we choose n_1 and n_2 posts randomly and independently among the total N is

$$p(k) = \binom{N}{n_1}^{-1}\binom{N}{n_2}^{-1}\binom{N}{k, n_1 - k, n_2 - k} \tag{4.5}$$

if the condition expressed in Equation 4.3 holds, otherwise we obtain $p(k) = 0$.

For computational reasons, we can reformulate Equation 4.5 in a different way, introducing the notation $(a)_b \equiv a(a-1)\dots(a-b+1)$, for any $a \geq b$, and assuming that $n_1 \geq n_2 \geq k$. Therefore, we obtain:

$$\begin{aligned} p(k) &= \frac{(n_1)_k (n_2)_k (N-n_1)_{n_2-k}}{(N)_{n_2} (k)_k} \\ &= \frac{(n_1)_k (n_2)_k (N-n_1)_{n_2-k}}{(N)_{n_2-k} (N-n_2+k)_k (k)_k}, \end{aligned} \tag{4.6}$$

where in the second form we have used the identity $(a)_b = (a)_c (a-c)_{b-c}$ valid for $a \geq b \geq c$. This equation can be better written as

$$p(k) = \prod_{j=0}^{n_2-k-1} (1 - \frac{n_1}{N-j}) \prod_{j=0}^{k-1} \frac{(n_1-j)(n_2-j)}{(N-n_2+k-j)(k-j)}. \tag{4.7}$$

The way to proceed from here is standard: given the co-occurrence matrix between the different hashtags and the information about the posts, we set a threshold p_0 such that we keep only those occurrences that appear with probability $p < p_0$. The result is hence a co-occurrence hashtag network where two hashtags are connected if their co-occurrence in the set of posts is statistically significant.

4.3 Hashtag Questionability Index

In order to measure the extent to which a hashtag is used in posts associated with either reliable or questionable content, we introduce a measure called *questionability*. The measure is defined in the range $q \in [0, 1]$ and it equals 0 when a certain hashtag is used exclusively in posts associated with reliable sources while it equals 1 when a certain hashtag is used only in posts associated with questionable sources.

Formally, hashtag questionability can be defined as follows: let $\mathcal{P}$ be the set of all posts with a url matching a domain in our data set and $\mathcal{H}$ the set containing all the hashtags. At each element $p_j \in \mathcal{P}$ is associated a binary value $l_j \in \{0, 1\}$ based on the domain of the link contained: if the url refers to

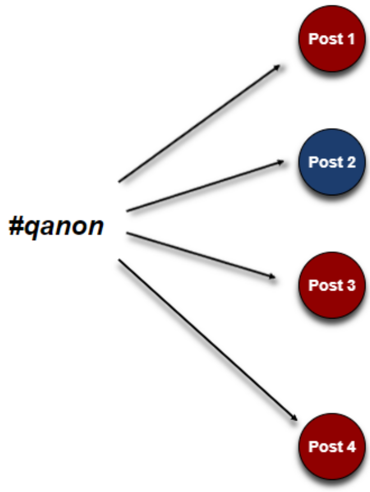


Figure 4.3: An example of questionability for the hashtag *qanon*. In all the posts where it was mentioned, 3 of them are identified as Questionable, whilst the remaining one is Reliable. Therefore, the hashtag questionability index will be $q = \frac{3}{4} = 0.75$

.

a domain classified as questionable then $l_j = 1$, otherwise $l_j = 0$. Considering and hashtag h_i in the bipartite network T then the questionability index q_i of hashtag h_i can be defined as

$$q_i = \frac{1}{k_i} \sum_{j=1}^{k_i} l_j \quad , \tag{4.8}$$

where l_j is the questionability score of the j-th neighbour of the hashtag h_i.

4.4 Power Law Distribution

A power law distribution characterises a continuous or discrete random variable. The Probability Density Function (PDF) associated with the first is defined as

$$p(x) = \frac{\alpha - 1}{x_{min}} (\frac{x}{x_{min}})^{-\alpha}, \tag{4.9}$$

where $\alpha > 1$ and $x_{min} > 0$. For the discrete case instead, we define the Probability Mass Function (PMF) as

$$P(X = x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})}, \tag{4.10}$$

where

$$\zeta(\alpha, x_{min}) = \sum_{n=0}^{\infty} (n + x_{min})^{-\alpha+1} \tag{4.11}$$

is the generalized zeta function [1], which becomes standardized if $x_{min} = 1$. Accordingly, we define the Cumulative Density Function (CDF) for the continuous version as

$$P(X \leq x) = 1 - (\frac{x}{x_{min}})^{-\alpha+1}. \tag{4.12}$$

For the discrete version, instead, we have

$$P(X \leq x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})}. \tag{4.13}$$

4.5 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov Test (abbreviated as KS test from now on), is a statistical test used to compare a sample with a specific probability distribution used as a reference or to compare two samples together. It is known for being a non parametric test, therefore it is not based on data drawn from a given family of probability distributions. The KS statistic computed from a test gives a value which indicates the distance between the Empirical Distribution Function (EDF) and the Cumulative Distribution Function (CDF) of the distribution used as a reference or, for the two-sample case, the distance between the two samples.

In order to compute the test, suppose that we have observations $X_1, \ldots, X_n$, which we want to know if they come from a distribution P. The KS test is based on the null hypothesis

$$H_0 : \text{the samples come from } P, \tag{4.14}$$

compared against the alternative hypothesis

$$H_1 : \text{the samples do not come from } P. \tag{4.15}$$

To achieve this, we want to compare the EDF of the data, F_{obs} with the CDF associated to the null hypothesis, F_{exp}. Given the observations $x_1, \ldots, x_n$ the EDF $F_{obs}(x)$ gives the proportion of the data whose values are lower or equal than x. Therefore:

$$F_{obs}(x) = \frac{\text{no. observations lower or equal than } x}{\text{no. observations of the sample}}. \tag{4.16}$$

The CDF of the null hypothesis, F_{exp}, is defined as

$$F_{exp}(x) = P(X \leq x). \tag{4.17}$$

The following definitions are included in the computation of the KS statistic

$$D_n = \max_x |F_{exp}(x) - F_{obs}(x)|. \tag{4.18}$$

In the case of two samples, we substitute the CDF of the theoretical distribution with the EDF of the second sample.

4.6 Kaplan Meier Estimator

The Kaplan Meier estimator is a non-parametric statistic for estimating a survival function defined on discrete interval times. Let then $S(t)$ be a function representing the probability of having a lifetime greater than the time t, such that

$$S(t) = P(\rho > t), \tag{4.19}$$

where $t = 0, 1, \ldots$. However, in real-life cases the true survival function $S(t)$ is never known. Therefore, we define an estimator which is the fraction of observations who survived for a specific amount of time t_i, where $i = 1, \ldots, n$. This results in the following definition

$$\hat{S}(t) = \prod_{i:\ t_i \leq t} (1 - \frac{d_i}{n_i}), \tag{4.20}$$

where t_i is the time when at least one event happened, d_i is the number of events (e.g., deaths) that happened at time t_i and n_i represents the number of observation at risk, i.e., the individuals known to have survived up to time t_i, which means that they did not die or they have been censored instead. To summarize, this estimator computes, at each time t_i, the product of the survival until that time.

4.7 Log Rank Test

When computing the Kaplan Meier estimator for the different groups of users, we want to assess if there is a difference between the two survival distributions that we obtained. For these purposes, a popular test that is implemented is the Log Rank Test, whose null hypothesis states that there is no difference in survival between the two groups at any time t or, more formally

$$H_0 : S_{1,t} = S_{2,t}. \tag{4.21}$$

The alternative hypothesis, instead, states that the two survival curves are not identical, stated as

$$H_1 : S_{1,t} \neq S_{2,t}. \tag{4.22}$$

The Log Rank statistic is distributed approximately as a chi-square distribution. The test statistic is defined as

$$Z^2 = \frac{\sum_{t=1}^{T} (O_{i,t} - E_{i,t})^2}{\sum_{t=1}^{T} V_{i,t}}, \tag{4.23}$$

where $O_{i,t}$ is the observed number of events in the groups at time t, such that $O_t = \sum_{i=1}^{G} O_{i,t}$ is the sum of the observed events in all groups at each time t. Similarly, $E_{i,t}$ is the expected number of events in the groups at time t, such that $E_{i,t} = N_{i,t}\frac{O_t}{N_t}$ with $N_t = \sum_{i=1}^{G} N_{i,t}$ the number of subjects who did not report an event considered from the survival function at each time t. In the end, the term $V_{i,t} = E_{i,t}(\frac{N_t - O_t}{N_t})(\frac{N_t - N_{i,t}}{N_t - 1})$ is the variance of each group at time t.

Chapter 5

Results and Discussion

In this chapter we analyze and compare how users perceive news in terms of reactions to posts, topics embedded in hashtag networks, and users' sentiment.

5.1 Consumption Patterns

We investigate the news consumption and the activity of Gab users by considering a data set of posts related to the COVID-19 pandemics. As shown in Figure 5.1, users tend to prefer a type of interaction that is more immediate and less cognitive-demanding [38]. Indeed, the left panel of Figure 5.1 shows how *Likes* are the most active way to engage, consequently followed by *Reblogs* and *Replies*. The same behavior is also confirmed by the cumulative number of interactions during the analyzed period. In this case, the difference between *Likes* and *Reblogs* is less accentuated until the beginning of February, with all distributions following an incremental trend that is comparable with consumption patterns from other social media [17].

The consumption pattern can also be analyzed considering the categorization of posts into questionable and reliable. Panel 5.4a of Figure 5.4, displays the distribution of like reactions to questionable and reliable contents that, overall, show rather similar behavior.

Panel 5.4b of Figure 5.4 shows the probability distributions of the number

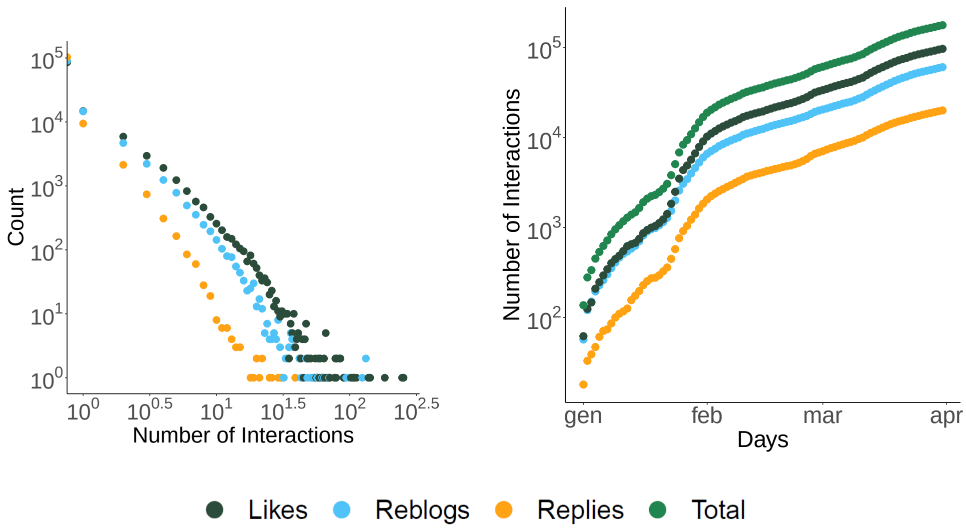


Figure 5.1: Frequency distribution of interactions with posts (left) and their cumulative engagement (right). A like is usually a positive feedback on a news item. A reblog indicates a desire to spread a news item to friends. A reply can have multiple features and meanings and can generate collective debate. The left panel display that every kind of reaction follows an heavy-tailed distribution that allows room for large deviations, i.e., some posts go viral. The right panel displays the evolution of the cumulative number of interactions over time. The trend is always increasing with a rapid increase at the beginning of February that is likely to be connected to the beginning of the COVID-19 infodemic. Both plots show how likes are the preferred type of interaction and how their frequencies are inversely proportional to the amount of cognitive effort required.

of posts by category with their corresponding fits. To check that those fits follow a power law distribution [18], we perform a bootstrap procedure [28] for both categories to estimate the scaling parameter α and x_{min}. The procedure for estimating the first relies on the computation of the Maximum Likelihood Estimator (MLE) which, for our discrete case, is not available. Therefore, we rely on the following approximation

$$\hat{\alpha} \simeq 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min} - 0.5}\right]^{-1} \tag{5.1}$$

where x_i are the observations such that $x_i \geq x_{min}$. For what concerns the estimation of x_{min}, it is conditioned by the calculation of $\hat{\alpha}$. That is because, in practice, power laws are only followed by the tails of a given distribution, so a minimum value x_{min} must be defined to consider an interval for the fitting procedure. An approach for accomplishing this goal is the estimation of this lower threshold by making use of the KS test which, in this case, calculates the distance between the data and the fitted model CDFs

$$D = \max_{x \geq x_{min}} |S(x) - P(x)|, \tag{5.2}$$

where $S(x)$ and $P(x)$ are the CDFs of each category of likes. The estimation of x_{min} is the value of x_{min} that minimizes D. After having described the way x_{min} and α are estimated, we implement those notions to test if the distribution of likes for each category follows a power law. Indeed, since it is possible to fit a power law distribution to any data set, we need to perform a test that indicates the goodness of the fitting that we obtain. To achieve this, we implement a procedure proposed in [18] and described in Algorithm 1.

Algorithm 1: Testing the power law hypothesis

Calculate point estimates for x_{min} and the scaling parameter α;
Calculate the Kolmogorov-Smirnov statistic, KS_d, for the original data set;
Set n_1 equal to the number of values below x_{min};
Set $n_2 = n - n_1$ and $P = 0$;
for *i in 1:B* **do**
  Simulate n_1 values from a uniform distribution $U(1, x_{min})$ and n_1 values from a power law distribution (with parameter α);
  Calculate the associated Kolmogorov-Smirnov statistic, KS_{sim};
  If $KS_d > KS_{sim}$ then $P = P + 1$;
end
$P = \frac{P}{B}$

We start by estimating α and x_{min} for the original data set calculating the KS statistic KS_d. The estimations allow us to compute n_1, i.e., the number of values below x_{min} from the initial data set, and n_2, i.e the difference with the total cardinality n and the n_1 previously calculated. Consequently, we computed 5000 randomized sampled where, for each of them, we performed a Kolmogorov-Smirnov test between the uniform distribution $U(1, x_{min})$, containing n_1 values, and the power law distribution with parameter a, containing n_2 values. We end up with obtaining a p-value that ranges from 0.23 to 0.31 for the distribution of likes to questionable, and 0.986 to 0.992 for the reliable ones. Thus we can accept the hypothesis that data is generated by a power law distribution.

The estimated exponent is 3.36 and 3.34 for questionable and reliable sources respectively, implying the presence of a very large deviation in the number of likes for both categories.

Panels 5.4c and 5.4d of Figure 5.4 show the temporal evolution of the cumulative and average number of likes to questionable and reliable contents. The matching between the two curves observed in the panel 5.4c is due to an increase in the number of reliable posts rather than an increase in the users' endorsement to such posts. Indeed, as confirmed by panel 5.4d of Figure 5.4, questionable posts receive on average an higher number of likes. It is nonetheless interesting the inflation in the number of reliable posts happened at the beginning of February that could be related either to a growing concern about the global pandemic or to a growing debate around reliable news. However, this inflation does not reflect in a correspondent growth in the number of likes, showing a constant interest of Gab users towards questionable sources.

5.2 Comparing Questionable and Reliable Hashtags

Hashtags are a good proxy for describing the semantic and topical elements of posts. Therefore, investigating the interplay between the use of hashtags

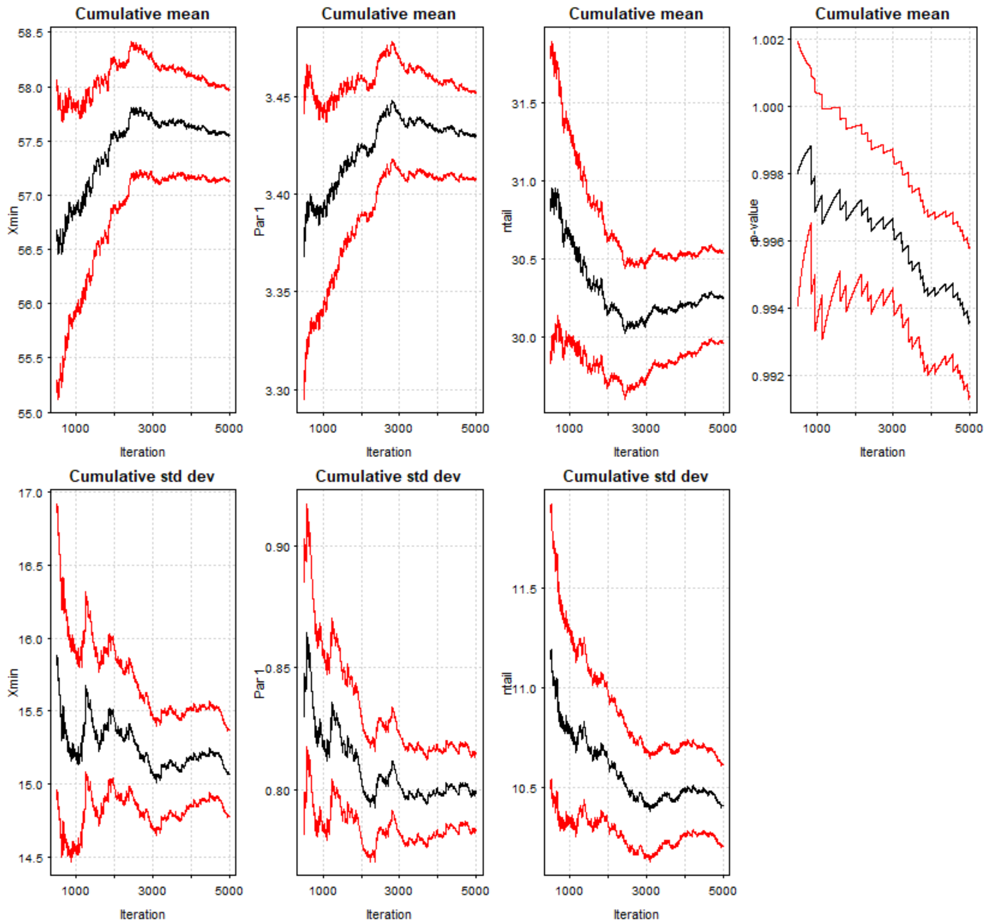


Figure 5.2: Results from the standard bootstrap procedure to test the relationship between Reliable likes distribution and the power law. Upper panel: evolution of the mean estimate of parameters x_{min}, α (Par1), the number of tail values greater than x_{min} (n_{tail}) and the result of the KS test with between the sampled uniform distribution and power law distribution. Lower panel: evolution of the standard deviation in the estimation of x_{min}, α and n_{tail}. The dashed-lines give approximate 95% confidence intervals.

and the diversity of information sources may unveil the narratives related to questionable and reliable contents. To achieve this goal, we consider 17996 hashtags appeared in labeled posts. The hashtags can be divided into three categories: those appearing mainly in questionable posts ($q_i \in [0.95, 1]$), those appearing mainly in reliable posts ($q_i \in [0, 0.05]$) and those appearing in both types of posts ($q_i \in (0.05, 0.95)$). The first subset is made up of 1332

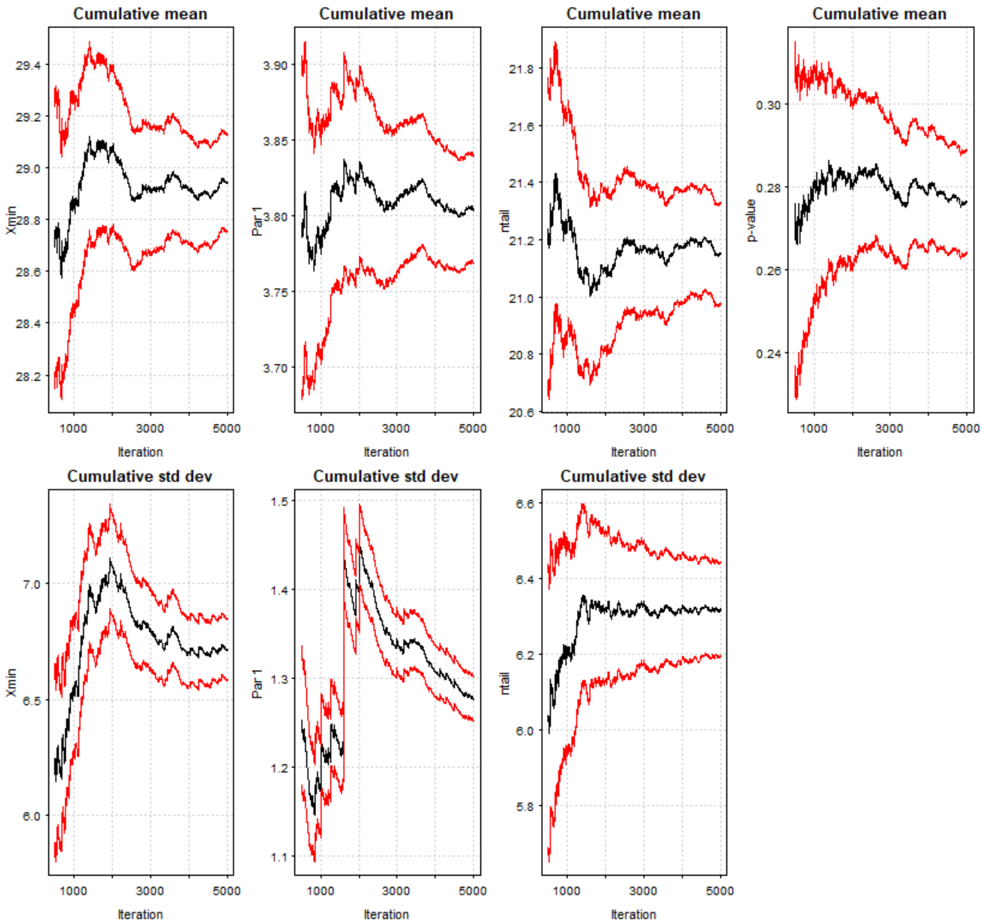


Figure 5.3: Results from the standard bootstrap procedure to test the relationship between Questionable likes distribution and the power law. Upper panel: evolution of the mean estimate of parameters x_{min}, α (Par1), the number of tail values greater than x_{min} (n_{tail}) and the result of the KS test with between the sampled uniform distribution and power law distribution. Lower panel: evolution of the standard deviation in the estimation of x_{min}, α and n_{tail}. The dashed-lines give approximate 95% confidence intervals.

hashtags, the second is made up of 14565 hashtags and the third is made up of 2099 hashtags. The hashtag questionability, described in Equation 4.8, follows a multimodal distribution with peaks located at extreme values, as shown in Figure 5.5.

We use these sets to build their corresponding co-occurrence networks using the procedure described in chapter 4.2, which are represented in Figure

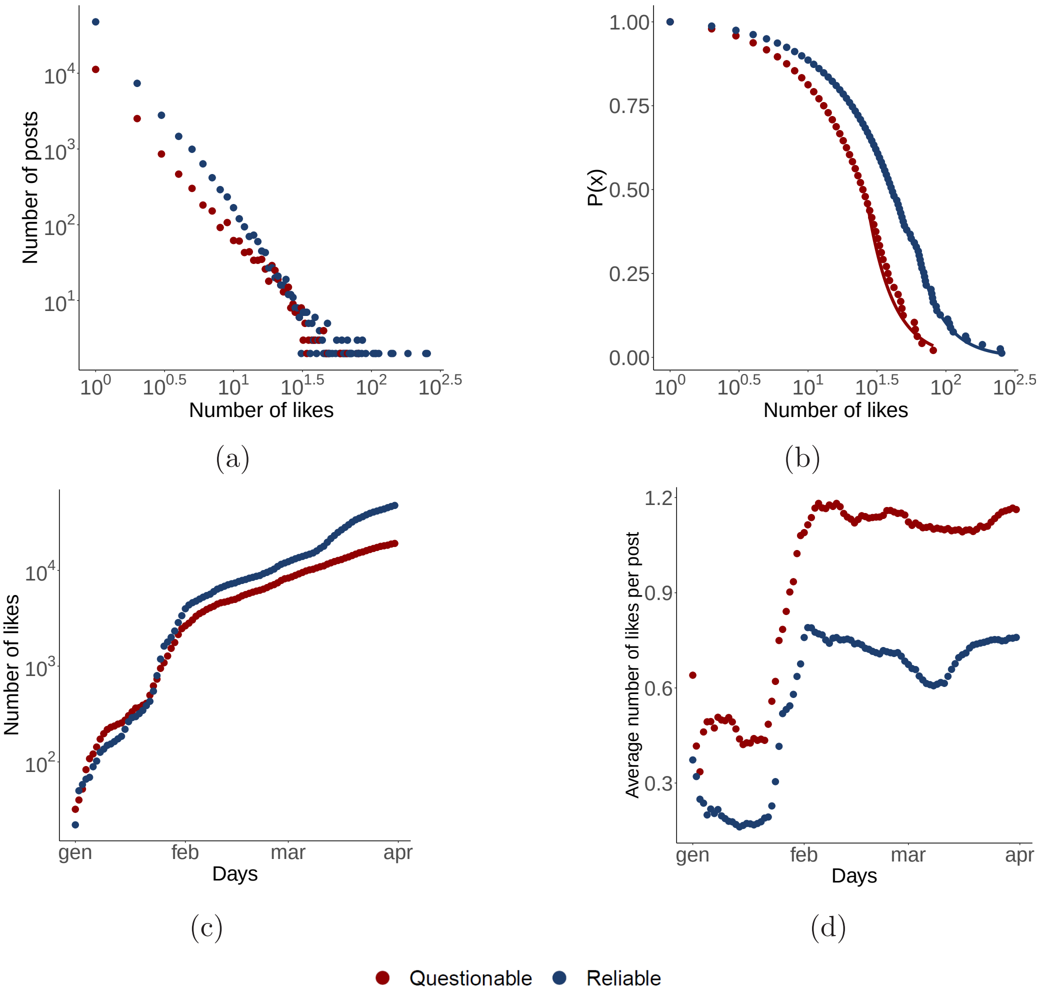


Figure 5.4: Panel a: distribution of the frequency of likes obtained by posts related to questionable and reliable sources. Panel b: Probability distribution of the number of likes to posts related to questionable and reliable sources. Panel c: Cumulative number of like over the time for questionable and reliable sources. Panel d: Average number of likes per post over time for questionable e reliable sources. The average is computed using a time window that contains all the posts since January the 1^{st}. Posts from both sources are similar in terms of likes' distribution, whilst their temporal evolution has shown a differentiation starting at the beginning of February.

5.6. The ten hashtags of the largest connected component with the highest prominence in terms of degree are reported in Table 5.1. The networks related to purely questionable and purely reliable contents display a strongly discon-

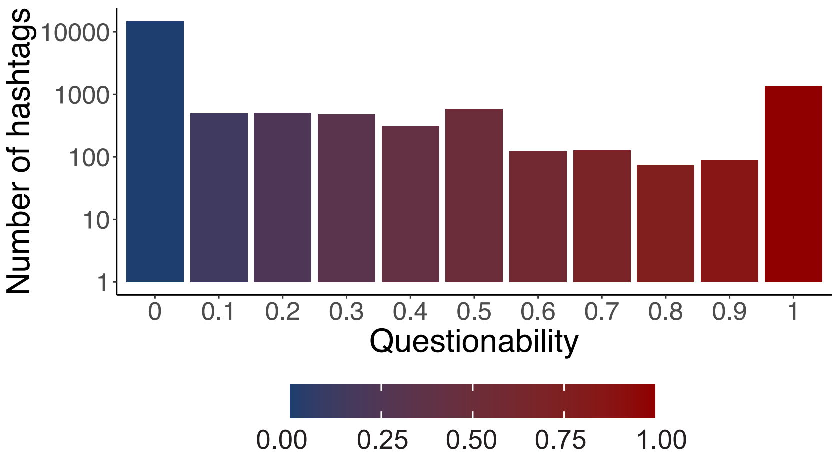


Figure 5.5: Distribution of questionability between hashtags. The two peaks at the extremes suggest there are recurrent hashtags for posts belonging to questionable or reliable sources.

nected structure made up of multiple connected components. In the case of questionable sources, we have a decentralized structure with the largest connected component of 8 vertices, accounting for the 2% of the total network. We also notice how most of the other connected components are organized as cliques. This highlights that questionable news have their dialect in terms of hashtags. In the case of reliable sources, we have a more centralized structure due to the contribution of its largest connected component that consists of 8527 vertices. Noticeably, the largest connected component accounts for 77% of the number of reliable hashtags, revealing a different structure with respect to purely Questionable hashtags.

The investigation of those two networks is then extended by looking at the most central hashtags. For the purely questionable network, the hashtags with the highest degree value are mostly associated with political facts and frustration about the current pandemic. Hashtags referring to *Dominic Raab* (the first secretary of state in the U.K.) were used, as well as hashtags like *dominance* or *shitholecountry*. For the purely reliable network the most central hashtags are mostly pandemic-related, e.g. *outbreak* and *school*. However, alt-right hashtags such as *wwg1wga*, which is generally associated with the Q-Anon movement have an important role.

Questionable		Reliable		Intersection	
hashtag	**degree**	**hashtag**	**degree**	**hashtag**	**degree**
brands	7	outbreak	763	pandemic	235
apology	7	tc	422	cdc	200
dominic	7	wwg1wga	236	who	195
raab	7	kag	173	news	183
thanks	7	donaldjohntrump	163	cia	166
dominance	7	Coronavirus*	147	trump	164
key	7	boingboing	141	health	156
shitholecountry	7	startups	137	virus	155
-	-	walkaway	133	democrats	148
-	-	school	122	maga	147

Table 5.1: Top 10 hashtags in the largest connected component.* Translation from Tamil language.

The right panel of Figure 5.6 displays the co-occurrence network related to hashtags that are significantly used in both types of posts. In this case, the network has the largest connected components of 2054 nodes that are higher than in the previous cases being about 98% of the total number of hashtags in the subset. The set of hashtags used in this case is also more general and related either to COVID-19 (e.g., pandemics, WHO, health) or to politics (e.g., trump, maga, democrats).

5.3 Characterizing Commenting Behaviour for Questionable and Reliable posts

In order to understand how news is perceived, we investigate the commenting behavior of users utilizing the sentiment expressed in the comments on questionable and reliable posts. We first pre-process the text of the comments via lemmatization and we use the Bing Lexicon [35], a list containing around 6800 terms related to opinions and sentiments divided by category (Positive or Negative), to obtain the sentiment of each comment. The sentiment s_i can be simply computed considering the number of positive and negative terms,

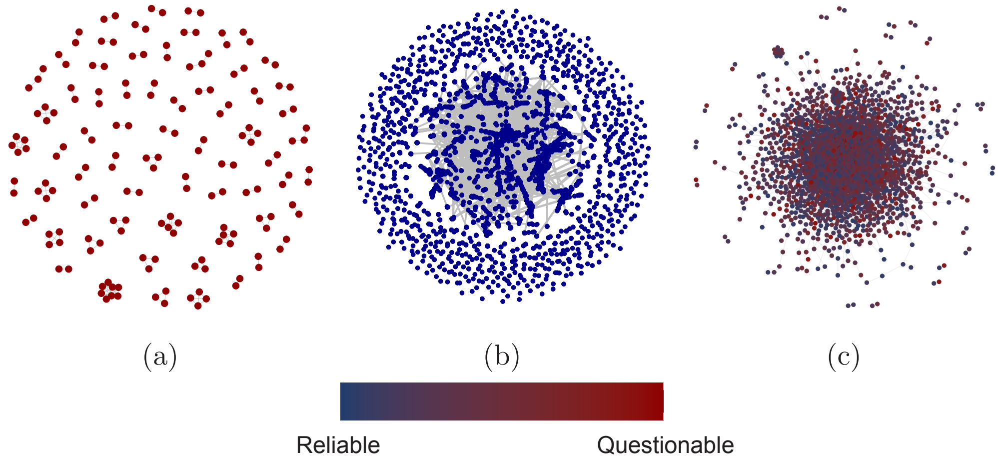


Figure 5.6: Projections of Post-Hashtag bipartite networks. Top Left: representation of the projection that contains only questionable hashtags, i.e., they are only used in posts from questionable sources. Top Right: representation of the projection that contains only reliable hashtags, i.e., they are only used in posts from reliable sources. Bottom: representation of the projection containing questionable and reliable posts which have at least one hashtag in common.

c_i^+ and c_i^- respectively, by means of the following equation:

$$s_i \equiv \frac{c_i^+ - c_i^-}{c_i^+ + c_i^-} \quad . \tag{5.3}$$

Notice that $s_i \in [-1, 1]$ for every i, where -1 means that the comment contains only negative terms, 0 that terms are equally distributed between positives and negatives and $+1$ that the comment contains only positive terms.

By computing the sentiment of comments on questionable and reliable posts, we obtain the distribution shown in Figure 5.7. Noticeably, there is no difference between the sentiment of comments under Questionable and Reliable sources. Furthermore, negative sentiment is what regulates user comments, with less pronounced peaks in correspondence of positive and neutrals. To assess the difference between the two distributions, we perform a Kolmogorov-Smirnov test [18] that reveals no significant difference between the twos (p=0.73).

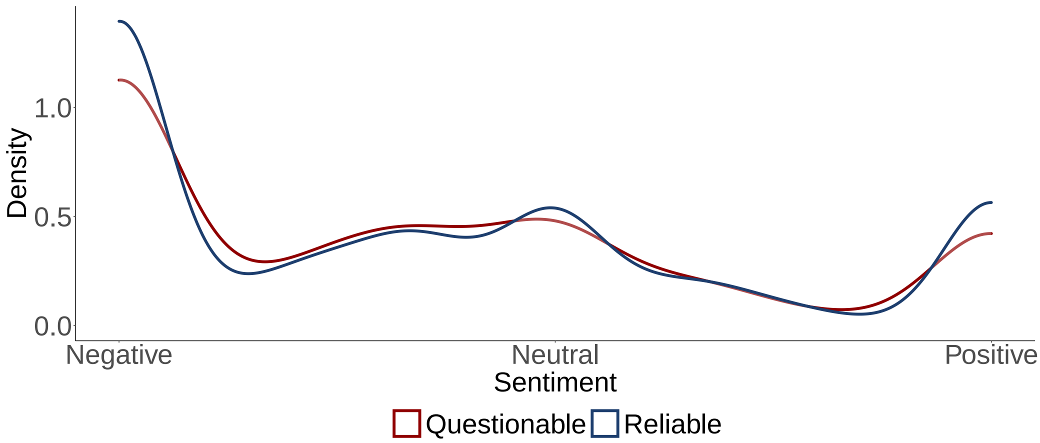


Figure 5.7: Sentiment distribution for Questionable (red) and Reliable (blue) post's comments.

To provide further insights for the commenting behavior of users under questionable and reliable posts, we model the persistence of users commenting repeatedly under a post of the same category. The modeling is performed by means of Kaplan Meier estimates of two survival functions: the first relies on the time span between the user's first and last comment, i.e., the lifetime of a user with respect to comments, whilst the second takes into account the number of comments of users. Figure 5.8 shows Kaplan Meier estimates of survival functions grouped by category for the two cases. Survival curves based on comments lifetime appear very similar (Figure 5.8a) , while the curves computed through the number of comments (Figure 5.8b) seem to present a slightly lower survival probability for comments to questionable posts. In spite of the latter observation, by performing a LogRank test [43] we detect no significant difference between the two survival functions (p-values equal to 0.81 and 0.54 respectively). Thus we can state that that the two categories are not significantly different in terms of survival probabilities: questionable and reliable sourced are perceived in the same way by users in Gab.

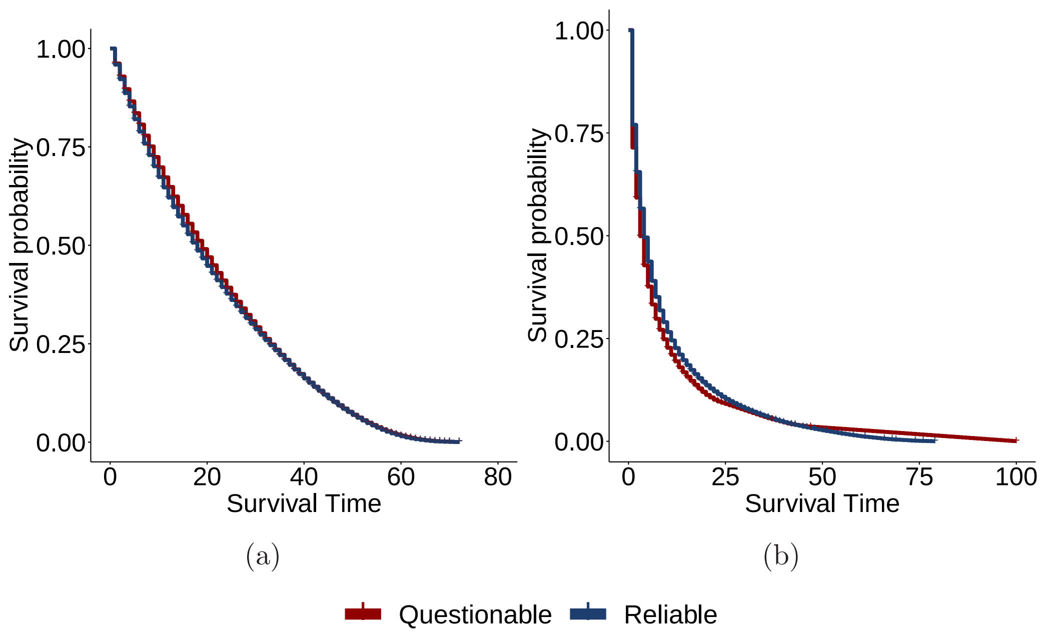


Figure 5.8: Panel a: Kaplan Meier estimates of survival functions computed using user lifetime, i.e., time span between user's first and last comment. Panel b: Kaplan Meier estimates of survival functions computed using the number of comments per user. Distributions are statistically indistinguishable in both cases, revealing the independence of comments' persistence from source questionability.

Chapter 6

Conclusions

In this thesis we presented an analysis concerning the users' online dynamics on Gab, a social media platform that is well known for being a safe place for people with extreme opinions, using the COVID-19 topic as use case due to its current importance and nature which, like many other health-related topics, can strongly polarize the communities into well-separated factions.

We started by introducing the problem, the background and the current state of the literature, underlying the importance of studying these dynamics due to the strong relationship between the problem of online and offline security and the spreading of misinformation and harmful ideas. We then described the data sets used for this study, which are composed of all the comments and posts from the period $01/01/2020 - 31/03/2020$ together with the classification of all the news outlets obtained from Media Bias/Fact Check, a well-known authority in the field of fact-checking. In addition to this, we described the data collection process with their API oriented structure, as well as the pre-processing operation in order to clean and organize the data in a suitable format for our analysis.

We described our analysis by introducing all the preliminaries and definitions useful for the experiments. We started from the concept of Networks and how they were useful to obtain a Bipartite Graph which, together with the statistical validation approach applied on the edges, made us able to create a network of hashtags that are connected if they appear on the same post. We also described the Kaplan Meier estimator and the statistical tests used during the analysis, i.e., Kolmogorov-Smirnov and Log Rank test.

After the introduction of the preliminary definitions, we described the experiment that we conducted, starting with the investigation of the consumption patterns of users. We characterized users' engagement on posts in terms of interaction and how it evolves during time as the pandemic arises. Furthermore, we classified posts into Questionable and Reliable categories depending on the questionability of the information source. We investigated users' endorsement to both categories in terms of social response and their evolution over time, focusing on differences and similarities of users' behavior. We analyzed the hashtag networks from a topological perspective and discussed the differences related to source type. Finally, we considered comments from both categories and study whether the questionability of the information source influences the distribution of the comments sentiment and the persistence in commenting of users.

Our analysis showed that users prefer less cognitive demanding interactions such as *Likes* and their attention respect to questionable and reliable sources changes over time, switching to the latter as the pandemic advances. In terms of hashtag associations through posts, the topological analysis reveals significant differences between reliable and questionable sources in terms of both structure and semantic content. However, the distribution of the sentiment deriving from the analysis of the comments reveals a rather similar pattern between reliable and questionable sources. Indeed, both distributions showed their peak in correspondence of negative sentiments revealing that the perception of the news does not depend on the source type. Thus, our results show the way users process information in a segregated environment such as Gab is homogeneous and does not depend on the source. The unconcern of Gab users with respect to the source in terms of endorsement and sentiment dynamics seems to provide further evidence for a mechanism of reinforcement that tend to interpret every news within a collective narrative that is typically found in echo chambers.

A series of possible additions can be made starting from this work. As an example, it may be possible to study how different topics are perceived from the users, without focusing only on the COVID-19 use case, and analyzing how much they differ by accounting to the metrics proposed. Moreover, sentiment analysis can be extended by searching for more specific sentiments. In this way, the concept of *Positive* and *Negative* that we proposed can be better characterized through their subdivision.

Bibliography

[1] Milton Abramowitz. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,* USA: Dover Publications, Inc., 1974. ISBN: 0486612724.

[2] Abdallah Alsaad, Abdallah Taamneh, and Mohamad Noor Al-Jedaiah. "Does social media increase racist behavior? An examination of confirmation bias theory". In: *Technology in Society* 55 (2018), pp. 41–46.

[3] Atanas Atanasov, Gianmarco De Francisci Morales, and Preslav Nakov. "Predicting the Role of Political Trolls in Social Media". In: *arXiv preprint arXiv:1910.02001* (2019).

[4] Christopher A Bail et al. "Exposure to opposing views on social media can increase political polarization". In: *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9216–9221.

[5] Alessandro Bessi et al. "Science vs conspiracy: Collective narratives in the age of misinformation". In: *PloS one* 10.2 (2015), e0118093.

[6] Alessandro Bessi et al. "Trend of narratives in the age of misinformation". In: *PloS one* 10.8 (2015).

[7] A. D. Borchgrave. "Cyber threats and information security : meeting the 21st century challenge : a report of the CSIS Homeland Defense Project". In: 2001.

[8] Alexandre Bovet and Hernán A Makse. "Influence of fake news in Twitter during the 2016 US presidential election". In: *Nature communications* 10.1 (2019), pp. 1–14.

[9] Alexandre Bovet, Flaviano Morone, and Hernán A Makse. "Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump". In: *Scientific reports* 8.1 (2018), pp. 1–16.

[10] Jo Brown, Amanda J. Broderick, and Nick Lee. "Word of mouth communication within online communities: Conceptualizing the online social network". In: *Journal of Interactive Marketing* 21.3 (2007), pp. 2–20. ISSN: 1094-9968. DOI: https://doi.org/10.1002/dir.20082. URL: http://www.sciencedirect.com/science/article/pii/S1094996807700300.

[11] Alex Burns and Ben Eltham. "Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 election crisis". In: (2009).

[12] Peter A Chew and Jessica G Turnley. "Understanding Russian information operations using unsupervised multilingual topic modeling". In: *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer. 2017, pp. 102–107.

[13] Wen-Ying Sylvia Chou, April Oh, and William MP Klein. "Addressing health-related misinformation on social media". In: *Jama* 320.23 (2018), pp. 2417–2418.

[14] Matteo Cinelli et al. "(Mis) Information Operations: An Integrated Perspective". In: *Journal of Information Warfare* (2019).

[15] Matteo Cinelli et al. "Echo Chambers on Social Media: A comparative analysis". In: *arXiv preprint arXiv:2004.09603* (2020).

[16] Matteo Cinelli et al. "Selective exposure shapes the facebook news diet". In: *PloS one* 15.3 (2020), e0229129.

[17] Matteo Cinelli et al. "The covid-19 social media infodemic". In: *arXiv preprint arXiv:2003.05004* (2020).

[18] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. “Power-Law Distributions in Empirical Data”. In: *SIAM Review* 51.4 (2009), pp. 661–703. DOI: `10.1137/070710111`. eprint: `https://doi.org/10.1137/070710111`. URL: `https://doi.org/10.1137/070710111`.

[19] CNN. “Gab, the social network used by the Pittsburgh suspect, has been taken offline”. In: (). URL: `https://edition.cnn.com/2018/10/29/tech/gab-offline-pittsburgh/index.html`.

[20] Mauro Conti et al. “It’s always April fools’ day!: On the difficulty of social network misinformation classification via propagation features”. In: *2017 IEEE Workshop on Information Forensics and Security (WIFS)*. IEEE. 2017, pp. 1–6.

[21] Michela Del Vicario et al. “Echo chambers: Emotional contagion and group polarization on facebook”. In: *Scientific reports* 6 (2016), p. 37825.

[22] Michela Del Vicario et al. “Mapping social dynamics on Facebook: The Brexit debate”. In: *Social Networks* 50 (2017), pp. 6–16.

[23] Michela Del Vicario et al. “News consumption during the Italian referendum: A cross-platform analysis on facebook and twitter”. In: *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. 2017, pp. 648–657.

[24] Michela Del Vicario et al. “The spreading of misinformation online”. In: *Proceedings of the National Academy of Sciences* 113.3 (2016), pp. 554–559. ISSN: 0027-8424. DOI: `10.1073/pnas.1517441113`. eprint: `https://www.pnas.org/content/113/3/554.full.pdf`. URL: `https://www.pnas.org/content/113/3/554`.

[25] I. Duić, V. Cvrtila, and T. Ivanjko. “International cyber security challenges”. In: *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 2017, pp. 1309–1313.

[26] Emilio Ferrara et al. “The rise of social bots”. In: *Communications of the ACM* 59.7 (2016), pp. 96–104.

[27] *Gab*. URL: `https://gab.com/`.

[28] Colin Gillespie. "Fitting Heavy Tailed Distributions: The poweRlaw Package". In: *Journal of Statistical Software, Articles* 64.2 (2015), pp. 1–16. ISSN: 1548-7660. DOI: 10.18637/jss.v064.i02. URL: https://www.jstatsoft.org/v064/i02.

[29] Njagi Dennis Gitari et al. "A lexicon-based approach for hate speech detection". In: *International Journal of Multimedia and Ubiquitous Engineering* 10.4 (2015), pp. 215–230.

[30] Nathaniel Gleicher. "Taking Down More Coordinated Inauthentic Behavior: What We've Found So Far". In: *Facebook Newsroom* (2018).

[31] The Guardian. *Bleach touted as 'miracle cure' for Covid being sold on Amazon*. 2020. URL: https://www.theguardian.com/world/2020/sep/19/bleach-miracle-cure-amazon-covid.

[32] Loni Hagen et al. "Rise of the Machines? Examining the Influence of Social Bots on a Political Discussion Network". In: *Social Science Computer Review* (2020), p. 0894439320908190.

[33] Roland Heickerö. *Emerging cyber threats and Russian views on Information warfare and Information operations*. Defence Analysis, Swedish Defence Research Agency (FOI) Stockholm, 2010.

[34] L. Howell. "Digital wildfires in a hyperconnected world. WEF Report 2013." In: *WEF Report 2013* (2013). URL: http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/.

[35] Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge* (2004).

[36] Haroro J Ingram. "The strategic logic of Islamic State information operations". In: *Australian Journal of International Affairs* 69.6 (2015), pp. 729–752.

[37] Mother Jones. "Men Who Allegedly Plotted to Kidnap Michigan's Governor Celebrated Violent Far-Right Extremism". In: (2020). URL: https://www.motherjones.com/politics/2020/10/men-who-allegedly-plotted-to-kidnap-michigan-governor-celebrated-violent-far-right-extremism/.

[38] Sara Levens et al. "Using Information Processing Strategies to Predict Message Level Contagion in Social Media". In: *Advances in Artificial Intelligence, Software and Systems Engineering*. Ed. by Tareq Ahram. Cham: Springer International Publishing, 2020, pp. 3–13. ISBN: 978-3-030-20454-9.

[39] Stephan Lewandowsky et al. "Misinformation, Disinformation, and Violent Conflict: From Iraq and the "War on Terror" to Future Threats to Peace". English. In: *American Psychologist* 68.7 (2013), pp. 487–501. ISSN: 0003-066X. DOI: 10.1037/a0034515.

[40] Lucas Lima et al. "Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system". In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE. 2018, pp. 515–522.

[41] CAR MacNulty and JJCH Ryan. "Using Values-Based Cultural Data to Shape Information Operations Strategies". In: *Journal of Information Warfare* 15.3 (2016), pp. 1–6.

[42] *Media Bias / Fact Check*. URL: https://mediabiasfactcheck.com/methodology/.

[43] Richard Peto and Julian Peto. "Asymptotically Efficient Rank Invariant Test Procedures". In: *Journal of the Royal Statistical Society: Series A (General)* 135.2 (1972), pp. 185–198. DOI: 10.2307/2344317. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.2307/2344317. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2344317.

[44] Walter Quattrociocchi, Rosaria Conte, and Elena Lodi. "Opinions Manipulation: Media, Power and Gossip". In: *Advances in Complex Systems* 14.04 (2011), pp. 567–586.

[45] Ana Lucía Schmidt et al. "Polarization of the vaccination debate on Facebook". In: *Vaccine* 36.25 (2018), pp. 3606–3612. ISSN: 0264-410X. DOI: https://doi.org/10.1016/j.vaccine.2018.05.040. URL: http://www.sciencedirect.com/science/article/pii/S0264410X18306601.

[46] Ana Lucıa Schmidt et al. "Anatomy of news consumption on Facebook". In: *Proceedings of the National Academy of Sciences* 114.12 (2017), pp. 3035–3039.

[47] Kai Shu et al. "Fake news detection on social media: A data mining perspective". In: *ACM SIGKDD Explorations Newsletter* 19.1 (2017), pp. 22–36.

[48] Kate Starbird. "Disinformation's spread: bots, trolls and all of us". In: *Nature* 571.7766 (2019), p. 449.

[49] Catherine A Theohary. *Terrorist use of the internet: Information operations in cyberspace*. DIANE Publishing, 2011.

[50] New York Times. "On Gab, an Extremist-Friendly Site, Pittsburgh Shooting Suspect Aired His Hatred in Full". In: (2018). URL: https://www.nytimes.com/2018/10/28/us/gab-robert-bowers-pittsburgh-synagogue-shootings.html.

[51] Michela Del Vicario et al. "Polarization and fake news: Early warning of potential misinformation targets". In: *ACM Transactions on the Web (TWEB)* 13.2 (2019), pp. 1–22.

[52] Soroush Vosoughi, Deb Roy, and Sinan Aral. "The spread of true and false news online". In: *Science* 359.6380 (2018), pp. 1146–1151.

[53] Jeremy Waldron. *The harm in hate speech*. Harvard University Press, 2012.

[54] G. Caldarelli Walter Quattrociocchi and A. Scala. "Opinion dynamics on interacting networks: media competition and social influence". In: *Scientific Reports* 4 (2014).

[55] Edward Waltz. *Information warfare: Principles and operations*. Artech House Boston, 1998.

[56] Claire Wardle and Hossein Derakhshan. "Information disorder: Toward an interdisciplinary framework for research and policy making". In: *Council of Europe report* 27 (2017).

[57] Savvas Zannettou et al. "What is Gab". In: *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18* (2018). DOI: `10.1145/3184558.3191531`. URL: `http://dx.doi.org/10.1145/3184558.3191531`.

[58] Fabiana Zollo et al. "Debunking in a world of tribes". In: *PloS one* 12.7 (2017).

Chapter 7

Appendix

7.1 Data Sets

7.1.1 Gab Posts Attributes

Column	Description
id	Post ID (Primary Key).
created_at	Date of the creation of the post.
in_reply_to_id	NULL if the post is not part of a thread, otherwise valorised with the post for which it is replying to.
in_reply_to_account_id	NULL if the post is not part of a thread, otherwise valorised with the user account id for which it is replying to.
language	Language of the post.
url	URL of the post.
replies_count	Number of replies to the post.
reblogs_count	Number of reblogs to the post.
favourites_count	Number of counts to the post.
content	HTML content of the post
account.id	id of the poster.
account.username	Username of the poster.

account.acct	Account name of the poster.
account.display_name	Username displayed of the poster.
account.bot	TRUE if the user is a bot, FALSE otherwise.
account.followers_count	Number of user followers.
account.following_count	Number of users being followed.
account.statuses_count	Account status id.
mentions.id	NULL if the post is not part of a thread, otherwise it contains the user id for which the post is replying to.
mentions.username	NULL if the post is not part of a thread, otherwise it contains the username for which the post is replying to.
tags.name	List of hastags for which the post was tagged with.
card.url	NULL if the post is not part of any card in Gab, otherwise it is valorised with its url.
card.type	NULL if the post is not part of any card in Gab, otherwise it is valorised with its type.
card.provider_name	NULL if the post is not part of any card in Gab, othwerwise it is valorised with its provider name.
group.id	NULL if the post was not published inside any group, otherwise it is valorised with its id.
group.title	NULL if the post was not published inside any group, otherwise it is valorised with its title.
group.description	NULL if the post was not published inside any group, otherwise it is valorised with its description.
group.member_count	NULL if the post was not published inside any group, otherwise it is valorised with the number of users within.
expanded_url	Extended URL of the post.
domain	Website domain of the post, i.e., the news outlet.
questionability	Questionability assigned for the news outlet.

7.1.2 Gab comments data set attributes

Column	Description
id	Comment ID (Primary Key).
created_at	Date of the creation of the comment.
in_reply_to_id	NULL if the comment is not part of a thread, otherwise valorised with the comment for which it is replying to.
in_reply_to_account_id	NULL if the comment is not part of a thread, otherwise valorised with the user account id for which it is replying to.
language	Language of the comment.
url	URL of the comment.
replies_count	Number of replies to the comment.
reblogs_count	Number of reblogs to the comment.
favourites_count	Number of counts to the comment.
content	HTML content of the comment
account.id	id of the commenter.
account.username	Username of the commenter.
account.acct	Account name of the commenter.
account.display_name	Username displayed of the commenter.
account.bot	TRUE if the user is a bot, FALSE otherwise.
account.followers_count	Number of user followers.
account.following_count	Number of users being followed.
account.statuses_count	Account status id.
mentions.id	NULL if the co is not part of a thread, otherwise it contains the user id for which the comment is replying to.
mentions.username	NULL if the comment is not part of a thread, otherwise it contains the username for which the comment is replying to.

tags.name	List of hastags for which the comment was tagged with.
card.url	NULL if the comment is not part of any card in Gab, otherwise it is valorised with its url.
card.type	NULL if the comment is not part of any card in Gab, otherwise it is valorised with its type.
card.provider_name	NULL if the comment is not part of any card in Gab, othwerwise it is valorised with its provider name.
group.id	NULL if the comment was not published inside any group, otherwise it is valorised with its id.
group.title	NULL if the comment was not published inside any group, otherwise it is valorised with its title.
group.description	NULL if the comment was not published inside any group, otherwise it is valorised with its description.
Column	**Description**
group.member_count	NULL if the comment was not published inside any group, otherwise it is valorised with the number of users within.
expanded_url	Extended URL of the comment.
domain	Website domain of the comment, i.e., the news outlet.
questionability	Questionability assigned for the news outlet.

7.1.3 Media Bias Fact Check Attributes

Column	**Description**
domain	News outlet name.
url	URL of the news outlet.
reliability	It can be questionable or reliable.

7.2 Gab API

7.2.1 Getting the list of hashtags related with the one given as input

Column	**Parameters**
url	`https://gab.com/api/v1/timelines/tag/{hashtag}`
method	GET
Description	The following method returns a list of hashtag which are correlated with the one given in input.
response	A list of hashtags in string format.

7.2.2 Getting the details of a Post

Column	**Parameters**
url	`https://gab.com/api/posts/{post_id}`
method	GET
Description	The following method returns all the details about the post identified with the input id.
response	A JSON object describing the posts whose fields are in the format described in Section 7.1.1.

7.2.3 Getting the comments of a given Post

Column	**Parameters**
url	`https://gab.com/api/v1/statuses/{post_id}/context`
method	GET
Description	The following method returns all the details about the post identified with the input id.

response	A list of JSON objects, each one containing all the fields about a comment in the format described in Section 7.1.2.