Ca' Foscari University of Venice

Department of Environmental Sciences, Informatics and Statistics

Master's Degree in Computer Science

Master's Thesis

Polarization on social media: A quantitative investigation of pro-science and conspiracy communities on Twitter

Supervisor

Prof. Fabiana Zollo

Co-supervisor

Dr. Alessandro Galeazzi

Candidate

Davide Rossato
Student ID 870530

Academic Year

2021-2022

This page intentionally left blank.

Acknowledgments

I would like to give my sincere thanks to my supervisor, Prof. Fabiana Zollo, for her support and patience during these very intense months. I would also like to thank Prof. Walter Quattrociocchi for being a reference and inspiration during these academic years; I hope to see him again in the near future.

I would like to thank all the people (Alessandro in particular) who helped me in writing and processing this thesis, giving me the right feedback and guidance to get the job done and not get discouraged.

I thank my lifelong friends who have accompanied me through these difficult 5 years of university, my family and relatives who always believed in me and gave me the strength I needed to make it this far.

Finally, I want to thank my fellow classmates; Giulia, Eleonora, Mattia, Giovanni and Veronica, your help and support was crucial and I will always be grateful, you are special.

This page intentionally left blank.

Abstract

In the last 20 years, online social networks have revolutionized many sectors. Politics, advertising, job search, commerce (from the multinational to the local shop), economics, social interactions and information are all fields that have undergone profound permanent changes, thanks to the advent of social media platforms. This thesis aims to validate behaviors observed in past studies on online social networks focusing on two opposite communities of users: pro-science and pseudo-science. In particular, we perform a quantitative analysis of 100,000 Twitter posts, measuring users' polarization and studying the communities in terms of network structure. Moreover, we investigate users' response to debunking content in the two communities, trying to confirm the expected behavior among users and the performance of the community networks described by previous studies. Our results shows that 1.2% of the conspiracy community moved to the scientific community, while only 0.2% made the opposite shift, but this shift doesn't occur with the same timing: in fact, it took in average 25 months for conspiracists and only 10 months for those in the scientific community.

Contents

Introduction

Social networks in recent years are changing the way people interact with each other and the way people create, consolidate or break relationships with other individuals; two-thirds (64%) of teens who have made a new friend online say they have met new friends on a social media platform and nearly all (94%) say they spend time with friends on social media [1]. Social media have helped people keep in touch with friends and to stay informed about the latest world news and share it with others; this behaviour has been documented in many situations, in a recent study has been reported that in 2018 (as in 2017) about two-thirds of American adults (68%) say they at least occasionally get news on social media, even though most people (57%) expect the news they see on social media to be largely inaccurate [2].

A recent report in 2021 by Censis, an italian institute, revealed that as many as 14 million Italians (30%) use Facebook to keep themselves informed, along with other media; however, as many as 4.5 million Italians inform themselves solely via social networks [3].

The availability of a lot of news, opinions and viewpoints in social networks allow them to play a key role in the development of people's opinions and the exchange of information due to its wide use [4, 5, 6]. This dynamic certainly brings great benefits to people, as they can share news and stay in touch much more easily than before, however it has been observed how people should be better informed and at the same time how people seek and avoid information in social networks [7].

The current online ecosystem is polluted with different types of non-genuine information, the so-called *dis-*, *mis-*, and *mal-* information [8] are very problematic for social dynamics as it fuels the population's rate of misinformation, creating opinions that greatly influence the population's thinking and can create unjustified popular discontent with economic, social, environmental policies; the World

Economic Forum has listed massive digital misinformation as one of the main risks for the modern society [9].

A first step in counteracting this type of information was taken by using search strategies and customised search engines within social media aiming to help users in retrieving information which is relevant with respect to their interests [10].

Recently, social network companies such as Facebook and Twitter are developing algorithms that try to prevent the dissemination of such information or share action by the user who has only read the scandalous headline without even having opened the article; however, many past studies have attempted to develop efficient algorithms for preventing the spread of misinformation [11, 12].

One of the most relevant means of transmission of information is certainly the presence of bots, i.e. accounts that automatically republish news, share it and sometimes comment on it. Although they follow very simple behaviours and are easy for an experienced user to spot, but they still manage to reach a segment of the population that interacts with them, commenting on posts and sharing them again. It has been observed that automated accounts diffusing fake news are much more active than the automated accounts diffusing other types of news; in addition, these super-spreaders may also use more sophisticated methods to increase their credibility, such as commenting on celebrity news (tagging them multiple times to exploit their visibility) or tend to target users with a higher median number of followers and lower variance [13]. In this way bots expose influential people, such as journalists and politicians, to a claim, creating the appearance that it is widely shared and the chance that the targets will spread it [14]. It has also been observed that social media real users had indirect exposure to bots (i.e. through account following) compared to direct interaction (i.e. by actively sharing, mentions and commenting on news) [15].

This type of account has always been opposed by social networks because they are not commercially profitable and reduce the possibility of expanding the network of people (this is one of the reason of the recent cancellation of Elon Musk's acquisition of Twitter [16]), yet it is not usual to question them for their ability to spread disinformation.

On the other hand, social networks are increasingly discouraging the dissemination of information from journalistic sources, in fact it has been seen that the reduced link exposure suggests that Twitter's algorithm may impede users' goal of retrieving news on the platform. While news may still circulate in the absence of external links, the reduced connections to the provenance of information in credible primary sources could hinder users' ability to assess information quality [17].

Recent work has shown how the increased exposure of users to unverified rumours leads them to become more credulous like in [18] where they highlight a higher level of commitment of consumers of conspiracy news neglected by main stream media and scientific news, consequently it very difficult to verify. This behaviour opens to interesting about the pervasiveness of unsubstantiated rumors in online social media; in an another study they find out that users with strong preferences for alternative information sources, perhaps motivated by the will to avoid the manipulation played by mainstream media controlled by the government, are more susceptible to false information [19] and this could be very harm full for modern society.

In recent years, people are increasingly exposed to an ever-increasing amount of information, in the work, social and entertainment fields. Companies adopt aggressive advertising campaigns both in real life through billboards, magazines and flyers and, more recently, on the Internet and in online social networks. The latter are also adapting to this type of advertising, creating social networks that give less space to the user in the construction of their news feed and the choice of content is more algorithm-driven (the "for you" section of TikTok is the most striking example), with suggestion algorithms targeted at individual users, but also with increasingly pervasive and insistent advertising [20]. Every commercial entity tries to get our attention and, due to the fact that the human attention span is limited [21], it is not possible to widen this window of attention and they are therefore forced to be more and more competitive with each other.

Social networks are adapting to suggest similar and less heterogeneous content [22] in order to build user loyalty to a specific of them.

Twitter, which is the social network of reference in this study, has a very interesting algorithm for suggesting friends called the Who-to-follow system. In 2015 it has been observed that this algorithm based on personal interactions allows the circle of friends to be considerably enlarged, thus creating a chain effect as soon as the person registers for the first time (more than 35% of new users use this the first day) and it has a direct impact on Twitter's growth and quality of user engagement; each month, it leads to the creation of more than 500 million new connections and 15% of Twitter's active users use this system at least once a month [23].

People tend to socialise with individuals they consider similar to them or with interests or topics in common: examples in real life are associations and clubs, which in social networks are translated into groups and pages; sometimes these may be closed and private, but very often they are public and seek to include as many people as possible. Just to give an example, according to data released by Facebook [24], of the 2.9 billion registered users, 1.8 billion use the groups function every month; there are tens of millions of active communities. This user communities that are created within social networks can be seen as closed boxes where the users within them share the same preferences and style of language. They also generate worldviews of society that are prone to the collective vision of the community itself: in fact, polarized users reinforce their preexisting beliefs by leveraging the activity of their like-minded neighbors, and this trend grows with the user engagement suggesting how peer influence acts as a support for reinforcement seeking [25].

In this study, we defined a polarized user as the individual who which their like activity (positive feedback) is almost (95%) exclusively on the pages of one category of people, according to the definition given in [26, 18].

However, there are many other methods of measuring user polarization, such as in the study by [27] that proposes a measure, different from a one-subject calculation, which simultaneously measures homophily and antagonism between groups, this new approach focus on the existence (or absence) of antagonism between the groups. It considers nodes' decisions towards connecting to users who belong to the other (potentially opposing) group, in comparison to connect to members of its own group. Furthermore, they have shown that polarized networks tend to exhibit a low concentration of high-degree nodes in the boundary between two communities [27].

This mechanism, which recognizes the homophily in the interaction networks, together with the bias in the information diffusion toward like-minded peers make up the two ingredients to quantify what scientist defined a echo chamber [28].

The special feature of echo chambers is their ability to feed themselves through the confirmation bias [29], which is the tendency to favor information that reinforces existing beliefs. Users belonging to an echo chamber feel more and more a part of it because they find feedback in what they think, supported by other like-minded users. In social networks, thanks to the communities that form, individuals can find confirmation in many other people who validate their ideas. Already in 2014, scientist had come to these conclusions, stating that even on the internet (there was not yet such distinct talk of social networks) opinion-reinforcing information is a more important predictor and promotes news story exposure while opinion-challenging information makes exposure only marginally less likely [30].

The existence of these echo chambers has been questioned several times [31, 32] however, most recent studies confirm their existence and they have described their most peculiar properties and effects[29].

Several echo chamber studies have revealed how political manipulation can influence how people vote in real life from political orientation [33] to U.S. and French presidential elections [34]. A relevant study performed on the extent to which echo chambers can influence political opinion showing that the spreadability of users (i.e. the efficiency of single users to disseminate information) is strongly correlated with their political orientation: information sent by pro-impeachment individuals spreads throughout the network much better than messages sent by other users, thus confirming the fact that conspiracy news spreads faster than others. Furthermore, they discover that users with larger spreadability are able to reach individuals with more diverse sentiments, actually escaping their echo chamber [35].

In more recent studies [36, 37], it has been seen that there is a link between the spread of misinformation and polarization; certain studies state that online misinformation and fake news polarized society developing an interesting 3-stage model (creation, spreading, polarizing) that illustrate how fake news polarized society; they also identify the catalyst and barriers that have an impact on polarization process, like cognitive biases (i.e. systematic error of judgement because they based their judge only from their perception of the input), incentives and online discussion [36]. Finally they try to formulate different solution or remedies to the spreading of fake news such as making people more aware and educated in order to identify sources of spreading disinformation [36]. Other researchers claim that it is polarization that fosters the spread of misinformation; they use a classifier that try to predict future fake news topics on social media [37]. In any case, both agree that there is a link between polarization and spread of misinformation.

In particular, the communities we are going to analyse in this study are communities that have been much studied [26, 18] in recent years and they are one the opposite of the other: pro-science community and conspiracy community.

While the pro-science community constitutes the majority of the population and is therefore the predominant one in social networks, the conspiracy community is the community that is found in social networks and that carries out an intense networking activity. One of the things that most distinguishes the two communities in the content with which they interact is that the former (pro-science) make use of verified scientific articles published by established scientists or recognised organisations, while conspiracists spread news that is 'neglected' by the mainstream media, effectively making it impossible to identify the sources. Another characteristic of conspiracists is the tendency to simplify all scientific, social and economic explanations in order to spread the message as widely as possible, to create buzz, basically to bring the community into the limelight [38].

This work, wants at first retrace the path of past studies, thus identifying the two types of communities examined (pro-science and conspiracy), what types of interaction patterns develop within them and how people are connected to each other. An analysis will be conducted concerning the relationship that different communities have with the various types of interaction provided by the social network (in our case the social network is Twitter which has 3 type of interaction: likes, retweets and replies). Finally, we conducted a quantitative study on the interaction of these two communities in the common field of debunk news: what we went to analyse was the presence or absence of patterns that can identify the behaviour of pro-science users from conspiracy users and vice versa, identifying what they have in common and what they differ from each other. A study was therefore conducted to see if it is possible to say that debunk news actually works and depolarizes conspiracy-minded users, leading them to more moderate and conciliatory positions; it has been shown that online debunking campaigns create a backfire effect in usual consumers of conspiracy stories [18], however, we want to find new feedback and understand more precisely whether there is a threshold of interaction with debunk news such that users tend to depolarize.

Chapter 1 the first part introduces some fundamental concepts of Graph Theory, reporting the most important definitions, theorems and algorithms that give context to the research carried out, then the second part wants to describe in detail all the methods and tools that have been used in the project;

Chapter 2 describes the final results of applying the methods described in the previous chapter;

1 | Foundations of Network Science

The purpose of this chapter is to introduce the theoretical concepts of what is called Graph Theory, a field of mathematics developed to study the interactions between subjects and objects, on which Social Network Analysis is based.

Starting from the definitions of the basic elements that make up a network, we then pass to formalize the main properties, its measurement methods and models of known networks.

Graph Theory has been applied for years to many fields of study such as biology, chemistry, sociology, electronics and telecommunications; in recent years the definitions have been generalized in such a way that they could also be suitable for the field of computer science.

The application of these graph models to the real world (they are called networks) allows us to define their characteristics and study their present behaviors, trying to predict the growth of the network and the interaction between the subjects that belong to.

1.1 Graph and Network

A graph can be described as the set of multiple components (called nodes or vertices) that relate to each other through some connections (called edges).

Usually the components of the node are graphically represented by a dot, while the edges are represented as straight lines (except in the case in which the same edge must reconnect to the same starting node). Before going any further, it is necessary to make a distinction between two types of graphs: **undirected** graphs and **directed** ones.

Undirected graphs has edges without a specific direction, consequently the nodes will be connected to each other in both ways. Directed graphs, on the other hand, specify the direction of the relationship by limiting it.

In the real world, an example of an undirected graph can be represented by a social network of friendship of an individual, while for a directed graph it can be represented by the number of links that a certain web page has: therefore all internet pages that lead to this web page (edges with the arrow pointing to the page) are called **in-degree** links, while all the links of the page that lead to other internet pages are called **out-degree** links.
A graphical example of these two types of graphs are shown in **Figure 1.1**.

Using a formal notation, we give the definition of graph in the following way:

> **Definition 1.1.1** (Graph)**.** A graph G consists of a collection V of vertices and a collection edges E, for which we write $G = (V, E)$. Each edge $e \in E$ is said to join two vertices, which are called its **end points**. If e joins u, $v \in V$, we write $e = \langle u, v \rangle$.
> Vertex u and v in this case are said to be **adjacent**. Edge e is said to be **incident** with vertices u and v, respectively.

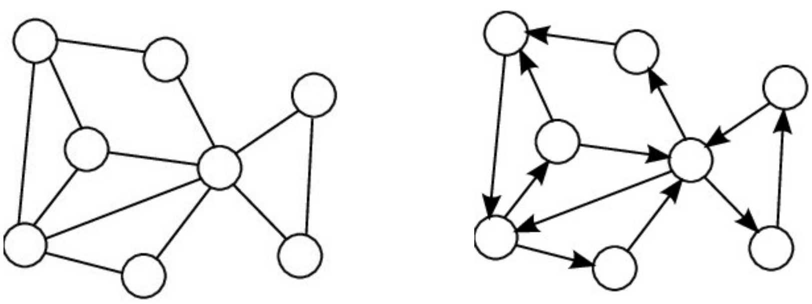

Figure 1.1: An example of undirected and directed graph

To indicate the set of vertices of the graph we use the writing $V(G)$, while to indicate the set of edges we use $E(G)$.

Another important set in graph theory is the neighboring vertices; given a vertex v, we define with $N(v)$ the set of vertices that are connected by an edge to the vertex v. More formally, this set can be defined as follows:

> **Definition 1.1.2** (Neighbor Set)**.** For any graph G and vertex $v \in V(G)$, the neighbor set $N(v)$ of v is the set of vertices (other than v) adjacent to v, that:
>
> $$N(v) \stackrel{\text{def}}{=} \{w \in V(G) \mid v \neq w, \exists e \in E(G) : e = \langle u, v \rangle\}$$

In order to work mathematically on graphs it is necessary to represent the nodes and the connections between them through an ideal data structure, also because generally the size of the graphs is very large.
The best data structure is the matrix and we can distinguish mainly two types:

- Adjacency Matrix: encode the relation of vertex-vertex pairs; it is an $i \times j$ matrix (where i and j are the number of vertices) such that:

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between vertices } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$

- Incident Matrix: encode the relation of vertex-edge pairs; it is an an $n \times m$ matrix (where n and m are respectively the number of vertices and edges) such that:

$$B_{nm} = \begin{cases} 1 & \text{if the vertex is connected to the edge} \\ 0 & \text{otherwise} \end{cases}$$

These data structures are also functional for weighted graphs, i.e. graphs that have a weight attributed to each edges. In that case it will be enough to modify the value 1 with the weight of the edge.

1.1.1 Properties of Graph

Degree of a Graph

The most important property of graphs is certainly the definition of the degree of a graph, as it allows us to study the structure of the network, identifying the most important nodes and the most evident connections on a very large network. Furthermore, this concept is the basis of many measurements that will be formalized later on. The degree of a graph is defined as follows:

Definition 1.1.3. The number of edges incident with a vertex i is called the degree of i, denoted as k_i (or in some cases with $\delta(i)$). Loops are counted twice.

The first important theorem concerning the degree of a graph is the following[1]:

Theorem 1.1.4. *For all graphs G, the sum of the vertex degrees is twice the number of edges, that is,*

$$\sum_{i \in V(G)} k_i = 2 \cdot L$$

[1] from now on, we report $|E(G)|$ with the letter L

Consequently to Theorem 1.1.4 it is possible to calculate the number of links L in an undirected network with the following formula:

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i$$

where k_i is the degree of the i^{th} node of the network

Another important property related to the definition of degree is that of the average degree, where for an undirected network it is calculated with the following formula:

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$$

While if we are using directed network, we must distinguish two types of average degree; one for the incoming degree (k_{in}), and one for the outgoing degree (k_{out}).

Degree Distribution

The notions just mentioned are important to understand the definition of the degree distribution of a network; this is certainly very useful to figure out the general structure of the network we are looking at.

As we will see later, the network conformation which are based on relationships between people (especially on internet, but also in real world) follow a certain trend in terms of growth of the network and interaction between users. Finally we can state that degree distribution determines many network phenomena, from the robustness of a network to the dissemination of information; for now we will limit ourselves to stating the definition:

> **Definition 1.1.5.** The degree distribution, p_k, provides the probability that a randomly selected node in the network has degree k. Since p_k is a probability, it must be normalized i.e.
>
> $$\sum_{k=1}^{\infty} p_k = 1$$
>
> For a network with N nodes the degree distribution is the normalized histogram, given by:
>
> $$p_k = \frac{N_k}{N}$$
>
> where N_k is the number of degree-k nodes.

Finally, if the degree distribution p_k of a network is known, it is possible to obtain the average degree of the network with the following formula:

$$\langle k \rangle = \sum_{k=0}^{\infty} k \cdot p_k$$

Clustering Coefficient

The clustering coefficient is another parameter that is usually calculated when analyzing a graph, and it measures the network's local link density; considered a node i, it represents the degree of connection between the neighbors of the nodes i. More formally, for a node i of degree k, the local clustering coefficient is defined as follows:

$$C_i = \frac{2L_i}{k(k_i - 1)}$$

where L_i represents the number of links between the k_i neighbors of node i.

We can therefore state that the more densely interconnected the neighborhood of node i, the higher is its local clustering coefficient.

Finally, we can represent the degree of clustering of the entire network using the average clustering coefficient $\langle C \rangle$ calculated on the average of all the local clustering coefficients of all the nodes of the network:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

1.1.2 Sparsity of Large Network

One of the distinguishing properties of a large network graph is its sparsity, which is an indication of the extent of its deviation from a fully connected graph. The more the deviation the higher is the sparsity [39].

Consequently, when we say that a network is sparse, it implies that also its adjacency matrix is sparse, thus creating problems in terms of information storage. When approaching a real large network, it is better to keep a list of node links with respect to the adjacency matrix, since an overwhelming fraction of the matrix elements are zero.

In graph theoretic literature, sparsity is a measure of the extent of a graph's deviation from the corresponding fully connected graph. But in many applied fields like signal processing or economics or sociology, sparsity is a measure to indicate relative diversity among related entities with respect to a certain quantity of interest [39].

In accordance with the results reported in [39], we report the formula and the consequent explanation of the best sparsity index of a network, that is the GINI index.

This concentration index is defined as the ratio between the area between the Line of Equality and the Lorenz curve (A) and the total area under the Line of Equality (A + B). **Figure 1.2** can help identify such areas. It follows that:

$$\text{GINI index} = \frac{A}{(A+B)}$$

If a representation is made up of a small number of entities that hold most of the total quantity, then the GINI index will tend to the value 1, indicating a sparse representation; vice versa if the distribution of real quantities approaches that of the Line of Equality, the GINI index will tend towards the value 0.

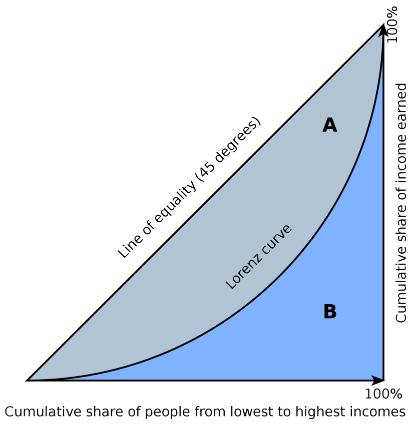

Figure 1.2: GINI Index calculating respect to Lorenz Curve and the line of perfect equality; this index was originally thought to be applied in economic field, but find some application also in social network analysis [40].

1.1.3 Bipartite Network

A bipartite graph is one of the most common types for representing large graphs whose nodes can be divided into two disjonted sets. This representation is used to display links between nodes belonging to the same set indirectly passing through the second set.

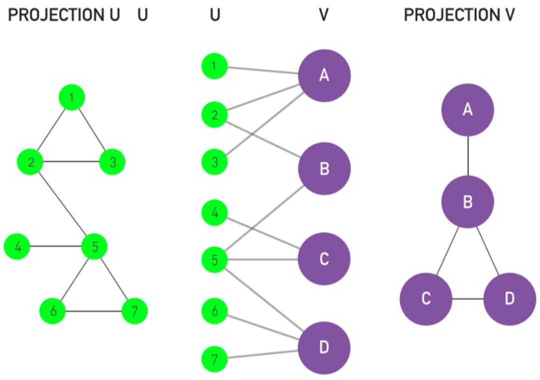

Figure 1.3: Bipartite visualization of a simple graph and projection of V set [41]

We can take the Hollywood actors' network as an example; each actor (who belongs to set A) has acted in one or more films (set F) and consequently will have one or more connections with the nodes of set F.

By visualizing the network as a bipartite graph it is possible to see which actors have acted in the same films and therefore which actors have been in contact.

An example of a bipartite graph visualization can be seen in **Figure 1.3**. As you can see, the nodes belonging to the same set do not communicate directly with each other; this allows us to create a second type of visualization, eliminating one of the two sets and thus transforming the indirect links between the elements into direct links.

1.2 Random Networks

Random networks are a category of graphs that until the last century remained unknown and not very detailed due to the little interest they aroused. When researchers discovered that many natural or social phenomena could be explained (and simulated) through the use of random networks, they began to study them and formalize their properties.

1.2.1 The Random Network Model

In 1950 the two Erdős-Rènyi researchers developed a model that can be formalized in the following way:

> **Definition 1.2.1.** An Erdős-Rènyi model of a random network on n vertices, also referred to as an **ER random graph**, is an undirected graph $G_{n,p}$ in which

each two (distinct) vertices are connected by an edge with probability p. For a given number M of edges, the ER random graph $G_{n,M}$ is an undirected graph in which each of the M edges is incident to randomly chosen pairs of vertices.

For each vertex u of an ER network there are at most $n-1$ vertices that can be neighbors of u, consequently there are $\frac{n-1}{n}$ possibilities to choose k different vertices adjacent to it.

The probability of having exactly j vertices (i.e. the degree of the node of a certain value) is equivalent to $p^j \cdot (1-p)^{n-1-j}$, so that

$$P[\delta(u) = j] = \binom{n-1}{j} p^j (1-p)^{n-1-j}$$

We can conclude that the degree distribution of an ER network is binomial.

For completeness, we state the theorem concerning the clustering coefficient:

Theorem 1.2.2. *The clustering coefficient of any $ER(n,p)$ is equal to p.*

1.3 Small World Property

The small world property (also known as Six degree of separation [42]) is one of the most important properties that characterize modern social networks. Its name derives from research carried out in past years on the average path by asking to a person to send letter to another person in the world. To their surprise, they found that on average 5.5 hops were needed to deliver the letter.

As you can expect, this property is even more important and marked in social networks, where the number of friendships and above all the possibility of getting in touch with people who are geographically very distant is certainly greater.

The researchers therefore tried to replicate this property also on an artificially generated Random Network (such as the ER-Model); Watts and Strogatz were the first to propose a similar model, now collectively referred to as small-world networks.

The model works as follows:

Algorithm 1.3.1 (Watts-Strogatz)**.** Consider a set of n vertices $v_1, v_2, \ldots, v_n$ and an (even) number k. In order to ensure that the graph will have relatively few edges (i.e., it is sparse), choose n and k such that $n \gg k \gg \ln(n) \gg 1$.

1. Order the n vertices into a ring and connect each vertex to its first $k/2$ left-hand (clockwise) neighbors, and to its $k/2$ right-hand (counterclockwise) neighbors, leading to graph G.
2. With probability p, replace each edge $\langle u, v\rangle$ with an edge $\langle u, w\rangle$ where w is a randomly chosen vertex from $V(G)$ other than u, and such that $\langle u, w\rangle$ is not already contained in edge set of(the modified) G.

1.4 The 80/20 rule

The 80/20 rule (also called the Pareto principle, so named in honor of who formulated this principle, Vilfredo Pareto, an Italian economist scientist who lived in the 19th century) highlights a characteristic to which many networks are prone; he observed that a small amount of people earned most of the money, while the majority of the population earned rather small amounts.

This trend has also been seen on the Internet (most links point to a small percentage of sites), or 80 percent of article citations go to 38 percent of scientists, etc. This principle only points out that most networks follow what is referred to as Power Law distribution, which we will formalize later.

1.5 Scale-Free Networks

The Watts-Strogatz model is generally considered to be the model that manages to reproduce the phenomenon of the small world. In 1999 Barabasi and Albert developed a model that best replicates the behavior evidenced by the Pareto principle, that is, that there are a few high-degree nodes, but that the number of nodes with a high degree decreases exponentially [43].

It has become common practice to call a network scale free if the distribution of vertex degrees follows a power law. Roughly speaking, this means that the probability that an arbitrary node has degree k is proportional to $(1/k)^{\alpha}$ for some number $\alpha > 1$ called the scaling exponent. In mathematical terms, $P[k] \propto k^{-\alpha}$. For most real-world scale-free networks, it turns out that $2 < \alpha < 3$ [44].

For convenience, the power law distribution is reported in a log-log plot, thus resulting:

$$\log P[k] \propto -\alpha \log k$$

1.5.1 Preferential attachment

What most characterizes the Barabasi-Albert model (hereinafter abbreviated as BA model) is the way in which the network grows; unlike other models, the network grows both in the number of nodes and in that of edges, connecting these new nodes following the principle of preferential attachment. This principle consists in privileging the connection of new nodes with nodes that already have many connections.

This practice can be summarized in the phrase "Rich get richer" as this mechanism auto-feeds the growth (exponentially) of nodes with more links, at the expense of those with few links. The algorithm works as follows:

> **Algorithm 1.5.1** (Barabàsi-Albert)**.** Consider a (relatively small) ER random graph G with n_0 vertices V_0. At each step $s > 0$:
>
> 1. Add a new vertex v_s to V_{s-1} (i.e., $V_s \leftarrow V_{s-1} \cup \{v_s\}$).
> 2. Add $m \leq n_0$ edges to the graph, each edge being incident with v_s and a vertex u from V_{s-1} chosen with probability
> $$\mathbb{P}[\text{select } u] = \frac{\delta(u)}{\sum_{w \in V_{s-1}} \delta(w)}$$
> that is, choosing a vertex u is proportional to the current vertex degree of u. Vertex u must not have been previously chosen during this step.
> 3. Stop when n vertices have been added, otherwise repeat the previous two steps.

A consequence of the principle of preferential attachment, which clearly distinguishes the conformation of a scale-free network from a random network, is precisely the constitution of **hubs**. With the term hub we indicate all the nodes of the network that have many connections (high-degree nodes); it has been proven that if two networks (one scale-free and one random) with the same average degree and number of nodes are compared, those nodes in the first are several orders of magnitude larger than the biggest node in a random network [41].

We can therefore conclude that while in a random network hubs are not conceived as many nodes have a comparable degree, as the size of the network grows polynomially (and the size of the largest nodes grows logarithmically or in any case more slowly than the number of nodes) , in scale-free networks the size of hubs grows exponentially and consequently it is very important that they exist and that they are studied.

This chapter will deal with the procedural and executive part of the project, starting with the tools used (with definitions and theoretical concepts necessary to understand the workflow), the software used and what kind of data was collected and how it was manipulated.

However, it is necessary to make a single premise: the type of data we will be dealing with are continuous variables, remarking the definition[2]:

> **Definition 1.5.2** (Discrete and Continuous variables)**.** A **discrete** random variable is an random variable whose possible values constitute either a finite set or a countably infinite set (e.g., the set of all integers, or the set of all positive integers).
> A random variable is **continuous** if *both* of the following apply:
>
> 1. Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).
> 2. No possible value of the variable has positive probability, that is, $P(\mathcal{X} = c) = 0$ for any possible value c.

1.6 Statistical Methods

This section describes all the tools used to analyse the degree distribution (for more details see Chapter 1.1) of the data collected concerning interactions with posts (likes, retweets and replies).

The first important thing that characterises a random variable (continuous) is the probability distribution function, which gives the probability that a random variable $\mathcal{X}$ takes on a certain value in a specific interval. More formally:

> **Definition 1.6.1** (Probability density function)**.** Let $\mathcal{X}$ be a continuous random variable. Then a probability distribution or probability density function (PDF) of $\mathcal{X}$ is a function $f(x)$ such that for any two numbers a and b with $a \leq b$,
>
> $$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

[2]All definitions and formal theoretical concepts of Section 1.6 and 1.6.1 are taken by [45]

That is, the probability that $\mathcal{X}$ takes on a value in the interval $[a, b]$ is the area above this interval and under the graph of the density function, as illustrated in Figure 1.4. The graph of $f(x)$ is often referred to as the *density curve.*

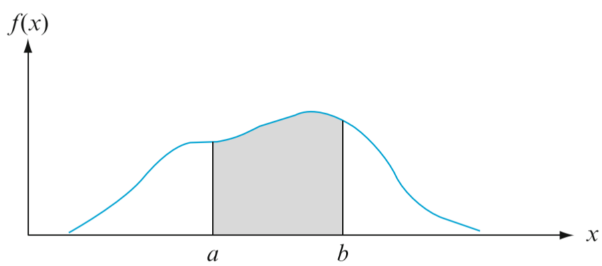

Figure 1.4: $P(a \leq X \leq b)$ = the area under the density curve between a and b

Another important tool used in the project is certainly the Comulative Distribution Function (CDF), which describes the probability that the random variable $\mathcal{X}$ takes from $-\infty$ to a certain value x; as the name implies, you have a graph obtained by cumulating the probabilities prior to the value. The formal definition is as follows:

Definition 1.6.2 (Cumulative distribution function)**.** The **cumulative distribution function** *F(x)* for a continuous random variable $\mathcal{X}$ is defined for every number x by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$$

For each x, *F(x)* is the area under the density curve to the left of x. This is illustrated in Figure 1.5, where $F(x)$ increases smoothly as x increases.

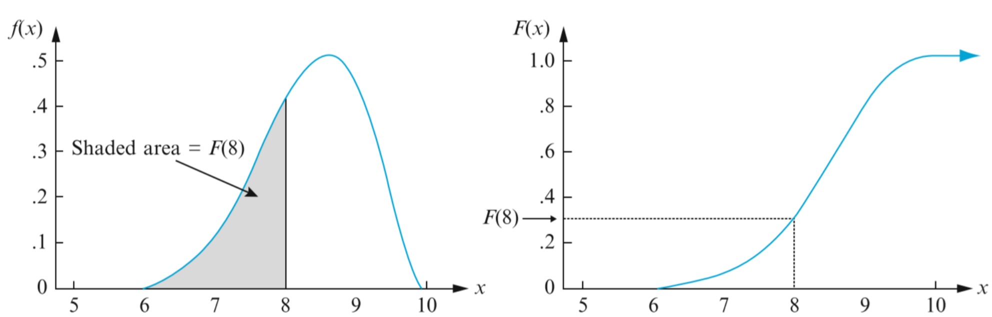

Figure 1.5: A PDF and associated CDF

In the project, we calculated the PDF of polarized users to see how users were distributed along all possible values of the polarity index.

If you would like to read more about how the polarity index is calculated for users, see Section 2.1 of this chapter.

The last tool we will use to represent the distributions is the **Complementary Cumulative Distribution Function (CCDF)**, which is simply calculated as $1 - F(x)$ (where $F(X)$ is the CDF mentioned in 1.6.2); it is used to better visualise the behaviour of empirical heavy-tailed distributions (by projecting the CCDF in a log-log plots).

Figure 1.6 shows the relationship between CDF, CCDF and CCDF in a log-log plot.

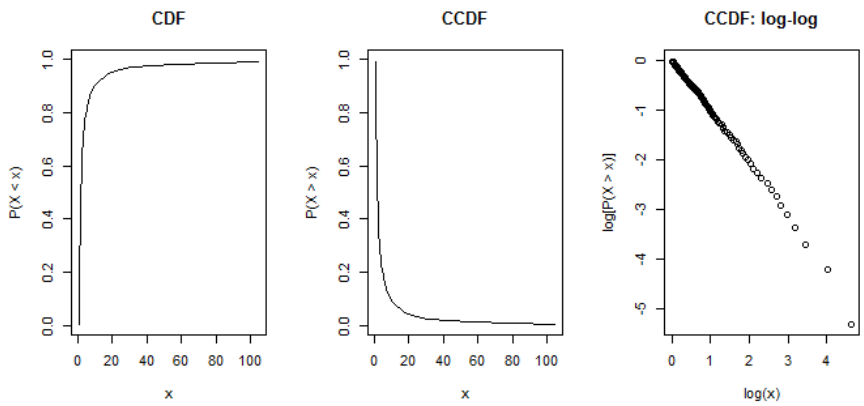

Figure 1.6: Comparison between a CDF, its CCDF and the CCDF in a log-log plot.

The CCDF (in log-log plot) came in handy to visualise the trend of the empirical distribution of total interactions, compared to the CCDFs of the individual classes of interactions (like, retweet, reply) for the community of pro-science and conspiracy users.

1.6.1 Known distributions

In this section, we will introduce two of the most popular common distributions related to Social Network Analysis; as we will see later, they have properties that are very close to the empirical distribution of user interactions in social networks.

For each of these distributions, work was carried out to find the parameters that best fit our empirical distribution (for both the science community and the conspiracy community).

Lognormal distribution

The log-normal distribution is a right skewed continuous probability distribution is used extensively in engineering, medical and financial analysis.

It is defined by the definition:

Definition 1.6.3 (Lognormal distribution)**.** A nonnegative random variable $\mathcal{X}$ is said to have a lognormal distribution if the random variable $Y = \ln(X)$ has a normal distribution. The resulting PDF of a lognormal random variable when $\ln(X)$ is normally distributed with parameters μ (mean) and σ (standard deviation) is

$$f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

It is very important to point out that the parameters μ and σ are not the mean and standard deviation of $\mathcal{X}$ but of $\ln(X)$.

Finally, for the sake of completeness, we state that the mean and variance of a lognormal random variable can be shown to be:

$$E(X) = e^{\mu+\frac{\sigma^2}{2}} \qquad Var(X) = e^{2\mu+\sigma^2} \cdot (\, e^{\sigma^2} - 1)$$

Various examples of lognormal distributions with different values of the mu and sigma parameters can be seen in **Figure 1.7**.

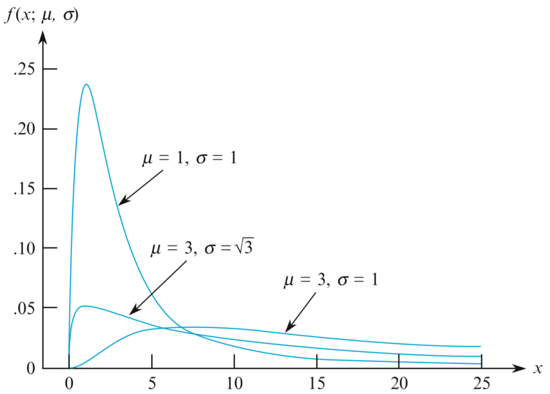

Figure 1.7: Various examples of lognormal distribution; we can easily see the right-skewness of the distribution as μ and σ tends to be equal to 1

Power law distribution

The power-law distribution, as stated in Chapter 1.5, is the distribution that the scientific literature has decreed as the best type of distribution that can approximate the degree distribution of a social network (or, in general, a scale free network).

It has become common practice to call a network scale free if the distribution of vertex degrees follows a power law[3].

Roughly speaking, this means that the probability that an arbitrary node has degree k is proportional to $(1/k)^{\alpha}$ for some number $\alpha > 1$ called the scaling exponent. In mathematical terms,

$$P[k] \propto k^{-\alpha} \tag{1.1}$$

For most real-world scale-free networks, it turns out that $2 < \alpha < 3$ [44].

For convenience, the power law distribution is reported in a log-log plot, thus resulting:

$$\log P[k] \propto -\alpha \log k$$

However, it is important to remember as written in [41] that in real systems we rarely observe a degree distribution that follows a pure power law. Instead, for most real systems p_k has two recurring features as displayed in **Figure 1.8**:

- **Low-degree saturation** is a common deviation from the power-law behavior. Its signature is a flattened p_k for $k < k_{sat}$. This indicates that we have fewer small degree nodes than expected for a pure power law.
- **High-degree cutoff** appears as a rapid drop in p_k for $k > k_{cut}$, indicating that we have fewer high-degree nodes than expected in a pure power law.

We can therefore reformulate our initial Formula (1.1) as follows:

$$P[k] \propto (k + k_{sat})^{-\alpha} e^{-\frac{k}{k_{cut}}} \tag{1.2}$$

1.6.2 Testing Goodness of Fit (GoF)

Once both distributions that best approximate the empirical distribution have been found, we can then proceed to compare them.

[3] Obviously, the rank of a node (in our case a tweet), is given by the number of interactions that generated that tweet.

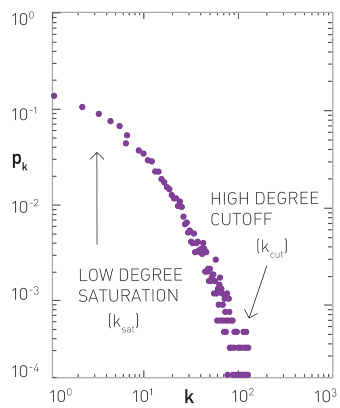

Figure 1.8: Example distribution that makes evident the two major problems in real distributions

Two-side p-value

Before introducing the concept of a two-sided p-value test, the definition of a test statistic and p-value must first be stated[4].

The test statistic T is a random variable constructed from two other random variables (mean $\bar{X}$ and sample standard deviation S) in the following way:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Its value is subject to uncertainty prior to obtaining the sample data.

The last important thing is that if the population distribution is normal, Gosset's Theorem implies that when the null hypothesis is true, T has a t distribution with $n-1$ degrees of freedom.

> **Definition 1.6.4** (P-value)**.** The **P-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as contradictory to H_0 as the value calculated from the available sample. The smaller the P-value, the more the data contradicts the null hypothesis, so H_0 should be rejected in favor of H_a if the P-value is sufficiently small.
> More specifically, select a number α reasonably close to 0; then reject the null hypothesis if P-value $\leq \alpha$ and do not reject the null hypothesis if P-value $> \alpha$. The selected α is called the **significance level** of the test.

Usually the values of the significance level α are 0.05, 0.01 and 0.001. In this thesis will be used always $\alpha = 0.05$.

[4]The theoretical part of this section is taken from [46].

Definition 1.6.5 (One sample T test)**.** Consider testing the null hypothesis $H_0 : \mu = \mu_0$ based on a random sample $X_1, X_2, \ldots, X_n$ from a normal population distribution (the plausibility of the normality assumption should be checked by examining a normal probability plot). The test statistic is

$$T = \frac{\hat{X} - \mu_0}{S/\sqrt{n}}$$

The calculated value of this test statistic is $t = \frac{\hat{x} - \mu_0}{s/\sqrt{n}}$. The determination of the P-value depends on the choice of H_a (see **Table 1.1**).

Alternative Hypotesis	**P-value**
$H_a : \mu > \mu_0$	Area under the t_{n-1} curve to the right of t
$H_a : \mu < \mu_0$	Area under the t_{n-1} curve to the left of t
$H_0 : \mu \neq \mu_0$	$2 \cdot$ (Area under the t_{n-1} curve to the right of $\|t\|$)

Table 1.1: Table summarising possible cases for the calculation of the p-value.

As far as our thesis is concerned, we only calculated the two-sided p-value as it is more suitable for our type of investigation, as we only need to find out whether or not to reject the null hypothesis.

In order to be more explanatory with regard to the case we are going to analyse (the third), in **Figure 1.9** it is possible to see graphically what is meant by p-value and all the elements involved.

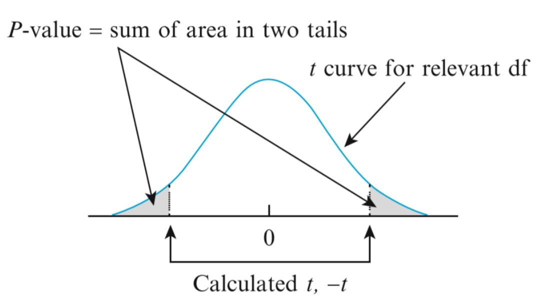

Figure 1.9: Two-tailed (sided) test where H_a contains the inequality $\neq$

Kolgomorov-Smirnov (K-S) Test Statistic

The first numerical measurement that was performed was the Kolgomorov-Smirnov Test, carried out to see how far the lognormal or power law distribution deviates from the empirical one. Note that this test will also be used in estimating the best power law distribution.

This test is formally defined as follows[5]: The K-S one-sample test compares the CDF of empirical distribution against the chosen distribution; basically it finds the point at which these two distributions show the largest divergence. Then, the test uses the largest divergence to identify a two-tailed probability estimate p to determine if the samples are statistically similar or different.

To perform a K-S test, we determine the relative empirical frequency distribution based on the observed sample. We then calculate its midpoint M and its standard deviation s as follows:

$$M = \frac{x_{max} + x_{min}}{2} \qquad s = \sqrt{\frac{\sum(f_i x_i^2) - \frac{(\sum f_i x_i)^2}{n}}{n-1}}$$

where x_{max} and x_{min} are the largest and smallest values of the sample respectively, x_i is the i^{th} value of the sample, f_i is the frequency associated with the i^{th} value and n is the number of observed samples. We need these two elements to compute the so-called z-scores to determine the probability associated with each sample value. These p-values are the relative frequencies of the empirical CDF $\hat{f}_r$.

Finally, we find the relative values of the obseverd frequency distribution f_r. Z-scores and f_r are computed with the following formula:

$$z = \left| \frac{x_i - M}{s} \right| \qquad f_r = \frac{f_i}{n}$$

Since the Kolmogorov–Smirnov test uses CDF, both the relative empirical frequency distribution and relative observed frequency distribution must be converted into cumulative frequency distributions $\hat{F}_{x_i}$ and S_{xi}, respectively. We then calculate the absolute value divergence $\tilde{D}$ and D between the CDF:

$$\tilde{D} = \left|\hat{F}_{x_i} - S_{x_i}\right| \qquad D = \left|\hat{F}_{x_i} - S_{x_{i-1}}\right|$$

Then we use the largest divergence to calculate the K-S test statistic Z:

$$Z = \sqrt{n} \max(|D|, |\tilde{D}|)$$

Vuong's Test

Without going into specific details, as the calculation method is very complex and not functional for our purpose, the Vuong's test is a likelihood ratio test for model

[5] The theoretical part of this section is taken from [46].

selection using the Kullback-Leibler criteria. Importantly, the two models must be non-nested[6].

The test statistic, R, is the ratio of the log-likelihoods of the data between the two competing models. **The sign of R** indicates which model is better [48].

More formally, this statistic test this two hypotheses:

- H_0: The two models are equally closed to the true data.
- H_1: Model 1 is closer than model 2.

1.6.3 Correlation

After having thoroughly analysed the distributions of our network, we went on to analyse whether there is a linear correlation between the different types of interaction (like, retweet and reply). Of the many types of correlation, we chose the Pearson Correlation[7]:

> **Definition 1.6.6** (Pearson Correlation)**.** Let a and b be two zero-mean real-valued random variables. The Pearson correlation coefficient (PCC) is defined as
>
> $$\rho(a, b) = \frac{E(ab)}{\sigma_a \sigma_b}$$
>
> where $E(ab)$ is the cross-correlation between a and b, and $\sigma_b^2 = E(b^2)$ are the variances of the signals a and b, respectively.

[6] Two models (or hypotheses) are said to be "non-nested" if neither can be obtained from the other by the imposition of appropriate parametric restrictions or as a limit of a suitable approximation. This is our case since power law distribution and lognormal distribution are non-nested models [47].

[7] Definition taken from [49].

2 Measuring polarization on Twitter: A case study

This chapter will deal with the elaborated plots and numerical results of the project; the first part announces the results obtained from downloading the data and what elements it consists of and in proportion, while the second part begins with the results of fitting the distributions, leading to the results of the polarity and finally the significant results we obtained in quantitatively studying the users who interacted with the debunk news.

2.1 Polarity Index

The method used to identify polarised users is as simple as it is very effective: once we have downloaded the additional information for each tweet (the second dataset mentioned in Section 2.2.1), we identify a polarised user of a faction if of all their interactions in our possession more than 95% are attributable to a particular community.

More formally, let us assume that user u has performed x and y likes in science and conspiracy posts, then

$$\rho(u) = \frac{y - x}{y + x}$$

Thus, user u for whom $\rho(u) = -1$ is biased towards the scientific community, whereas if $\rho(u) = 1$ he is biased towards the conspiracy community.

We define the polarisation of a user $\rho \in [-1, 1]$ as the ratio of the difference in interactions between conspiracy and proscience posts, divided by the total number of interactions.

2.2 Datasets

The tweets of the accounts we decided to download are essentially the most relevant accounts of each community and were carefully hand-selected; therefore, being highly polarised and popular accounts, we assumed that all posts made by these accounts belonged to one of the three categories (Science, Conspiracy, Debunk).

The second clarification that needs to be made concerns the size of these three communities on Twitter Italy; the community of pro-science people is clearly larger than the conspiracy community (this is due to the fact that even in real life it is possible to see that conspiracy people are just a small minority in society).

More in detail, we have chosen:

- 13 pro-science accounts (1.8 mln total followers)
- 13 conspiracy accounts (320 k total followers)
- 6 debunk news accounts (570 k total followers)

The account usernames for each category are reported in **Table 2.1**.

Science	**Conspiracy**	**Debunk**
RobertoBurioni	MartinKeufaver	Bufalenet
esaspaceflight	Autismo_Vaccini	Labbufala
valigiablu	Pianetablunews	iovaccino
le_scienze	MinervaMCGrani1	disinformatico
oggiscienza	a_meluzzi	butacit
Focus_it	MarcelloLyotard	DavidPuente
dariobressanini	PotereVerita	
cicap	dessere88	
esa_italia	piersar62	
mediainaf	byoblu	
ASI_spazio	sapereeundovere	
wireditalia	barbarab1974	
gravitazeroeu	Silvana_demari	

Table 2.1: Table showing all the usernames of the accounts taken into account and their category.

In conclusion, when analysing the results, the balance of importance between the two communities must be taken into account.

2.2.1 Tools and software

The software that was used in the information retrieval phase was mainly RStudio and Python (the latter programming language through the interface of Anaconda and Jupyter Notebook).

In fact, both tools contain external libraries that make it easy to interface with Twitter's API; in fact, the possibility of creating automatic scripts made it possible to overcome the fact that Twitter only allows you to download 10 million posts per month for each Academic account[1], but most importantly the limit of 750 requests every 15 minutes to retrieve information concerning the likes and retweets of the downloaded posts (which corresponds to approximately 10,000 interactions every 15 minutes).

To give a time reference, in order to get all the tweets of 33 Twitter Account of the last 4 years (gen 18 - gen 22) has been completed within about three months, using two accounts with Academic Access and four other accounts without special privileges (which were dedicated to retrieving tweet information such as likes and retweets).

Dataset cleaning

The amount of data downloaded is very substantial, both in terms of the number and the information that can be obtained for each tweet/user; I have only retained the information necessary for the purposes of my research, eliminating superfluous information.
The following information has been retained for each tweet:

Information Name	Description
ID	A numerical value that uniquely identifies the individual tweet.
Conversation ID	A numeric value that uniquely identifies the tweet conversation.
Referenced_tweet_replyTo_ID	A numeric value identifying the ID of the tweet being replied to.

[1] Academic accounts are verified Twitter accounts that have researcher privileges: without them, it is only possible to download tweets created within the last 30 days, and this type of account also has higher rate limits and tweet caps.

Information Name	Description
Referenced_tweet_Retweeted_ID	Numerical value identifying the ID of the tweet that was retweeted.
Referenced_tweet_Quoted_ID	Numerical value identifying the ID of the tweet that was mentioned.
Author_ID	Unique numeric value identifying the author of the tweet.
Created_At	Date of creation of the tweet.
Like_Count	Number of likes the tweet received when it was downloaded.
Quote_Count	Number of quotes the tweet received when it was downloaded.
Reply_Count	Number of replies the tweet received when it was downloaded.
Category	It identifies at what community the tweet belong (Science, Conspiracy, Debunk).

Table 2.2: Table showing all the information retrieved for each tweet together with its description

The second dataset that was created concerns the interactions with each tweet (likes and retweets); again, it was decided to keep only a few which are summarised in the next table:

Information Name	Description
ID_User_Action	A numerical value that uniquely identifies the individual user.
Follower_Count	Number of followers the user has.
Following_Count	Number of person who follows the user.
Tweet_Count	Number of tweet the user has.
Author_Tweet_ID	ID of the tweet that the interaction is referred.
Category	It identifies at what community of users the interaction belong.

Information Name	Description
Like_retw	Variable that say if the interaction is a retweet or like action.

Table 2.3: Table summarising all the information retrieved for each interaction (like, retweet) together with its own description.

2.3 Dataset Overview

A total of 165.000 tweets and most of its iterations were downloaded, resulting in approximately 24.2 million. All these iterations were generated by 928.000 distinct users. In detail, the following were downloaded:

	No. tweets	No. interactions	No. distinct users
Science	45,000	10 milions	550,000
Conspiracy	60,000	10,6 milions	163,000
Debunk	63,000	3,6 milions	215,000

Table 2.4: Table showing for each type of community, the number of tweets, the number of interactions and the number of distinct users that were retrieved.

It is necessary to point out that the number of distinct users is given by the users who have interacted with the tweets of the 33 reference accounts, while in the our analysis we consider the community made of polarized users that will be ranked.

In this regard, not of all this users were classified into one of the three categories; in fact, about 40 K users were not given a category and were discarded from the subsequent analyses as irrelevant for our purposes.

2.4 Elaboration

Once the datasets had been reorganised, the resulting volume of data was analysed quantitatively.

The first type of analysis, that you could see in **Figure 2.1**, is usually performed to observe the behaviour of the attention patterns of our networks through the CCDF (Complementary, Cumulative Distribution Function, see Chapter 1.6 for detailed explanations about method and theory) of the number of total interactions, likes, retweets and replies.

This distribution is very useful to understand the general community behaviour in a social network; the cascading behaviour in log-log plot confirms that our network possesses the same characteristics as Scale-Free networks (see Chapter 1.5 for more details).

It is very clear that both communities exhibit similar behaviour, i.e. all distributions have heavy tails.

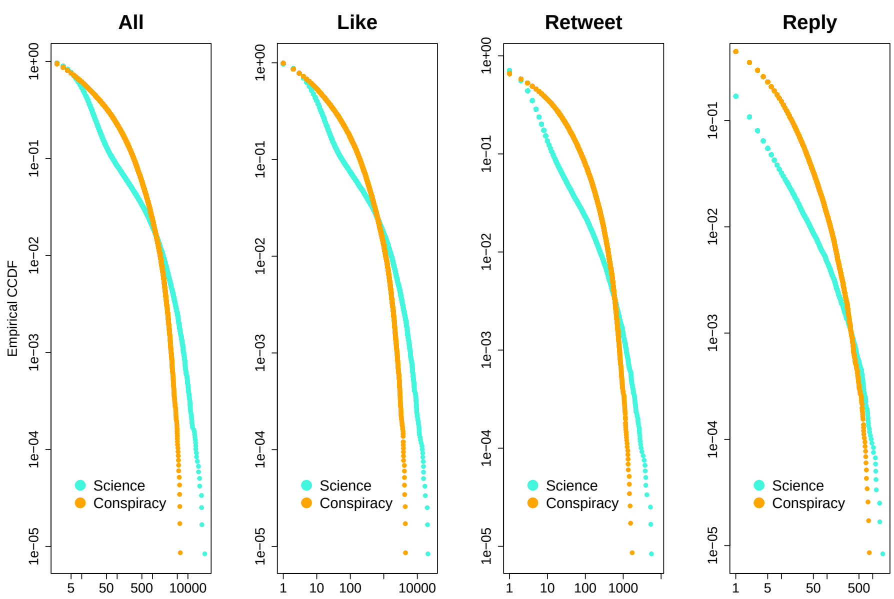

Figure 2.1: Empirical CCDF of users' activity (all activity, like, retweet, reply) for tweets grouped by the categories of science (light blue) and conspiracy (orange). The distributions are indicating heavy-tailed consumption patterns.

2.4.1 Goodness Of Fit (GoF) of the models

It was decided to further investigate the best distribution that can approximate the behaviour of the two communities.

We know from the literature that the frequency distributions of interactions of people can be approximated (in the general case) by a power-law distribution because the distribution is very heavy-tailed. To further confirm this view, I tried to see if another very popular type of distributions, such as the lognormal distribution, could compete with a power-law distribution.

We therefore proceeded to check the fitting of a lognormal distribution, checking the goodness of fit graphically (see **Figure 2.2**)

Fitting Lognormal

As explained in Section 1.6.1, the lognormal distribution has two parameters, which were estimated by means of a library function fitting this distribution to the empirical data provided. The parameters of the two distributions and the K-S test between the model and the empirical data are shown in **Table 2.5**, while in **Figure 2.2**, one can see how the two distributions fit the data quite well; the results of the K-S test are very low, probably too low, which makes us think it **might be an overfitting problem**.

	Mean	**St. deviation**	**K-S test**
Science	6.079	1.095	0.01003
Conspiracy	7.245	0.490	0.01100

Table 2.5: Summary table showing estimated parameters for the two lognormal distributions.

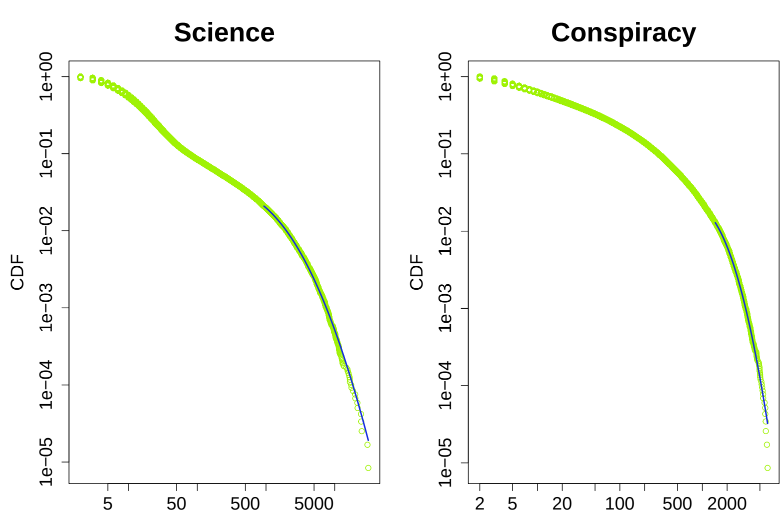

Figure 2.2: Fitting of lognormal distributions (blue line) in the two scientific and conspiracy communities.

Fitting Power law

The fitting of the power law distribution was much more complex; initially, the values of K_min[2] and *alpha* were estimated such that they fit the basic power law (Formula 1.1). We want to find the best tuple which minimised the *KS distance* D; so we proceed calculating this distance for every possible tuple. In **Figure 2.3** for **Science community** and in **Figure 2.4** for **Conspiracy community**, it can be seen graphically in **graph A** that the value of K_min was in the global minimum of D, while in the second **graph B** we can see all the possible values of *alpha* associated with each K_min; finally in the third **graph C** we can see how the best distribution with the best parameters approximates the empirical distribution in a log-log plot.

The values of both models are summarised in **Table 2.6**.

Fitting power law distribution for Science Community

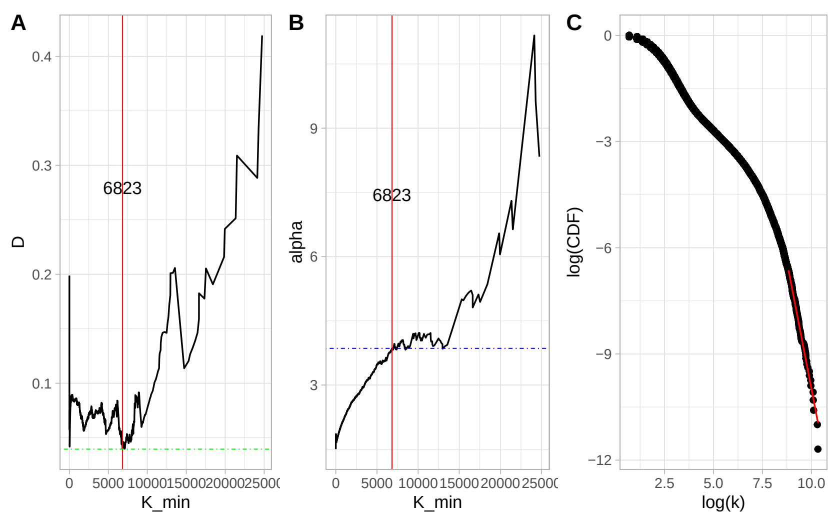

Figure 2.3: Finding the best *K-min* and *alpha* parameters (A and B) and fitting the best distribution to the empirical data (C)

	D	α	**K_min**
Science	0.0346	3.856	6823
Conspiracy	0.0490	6.095	3102

Table 2.6: Summary table of calculated parameters for the first type of power law.

[2] since for any positive value of *alpha* the distribution diverges when k tends to zero, it is normal to impose a minimum value K_min

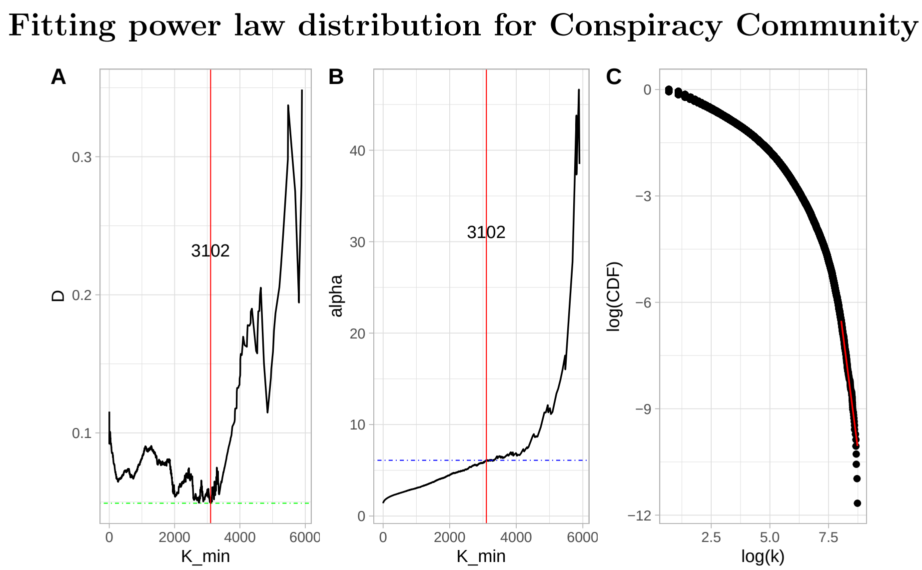

Figure 2.4: Finding the best *K-min* and *alpha* parameters (A and B) and fitting the best distribution to the empirical data (C)

At this point, I moved on to find a fitting for the second power law distribution Formula 1.2 stated in Section 1.6.1, in order to obtain the most generic possible model for the empirical distribution.

The values of k_sat and k_cut were estimated as follows:

1. All possible tuples of k_min and k_max were computed (with the first k_min computed as the lower bound);

2. At this point, for each pair, the KS distance D was computed and we select the tuple which minimize D.

The final values of k_sat and k_cut thus correspond respectively to the values of k_min and k_max that minimise D.

The values obtained for both communities are shown in **Table 2.7**, while in **Figure 2.5** you can see the fitting of the new distribution in comparison with the first type of power law and the lognormal distributions.

	D	α	**K_sat**	**K_cut**	**p-value**
Science	8.5676	3.950	7129	7146	0.505
Conspiracy	0.01346	4.4423	1973	1974	0.620

Table 2.7: Summary table of calculated parameters for the second type of power law.

As can be seen from Table 2.7, the results of these new power laws are peculiar: the values of the K-S distance D are rather low compared to the level of generalisation we wanted to achieve, however α values are more in line with the type of network we expect to analyse. The values of k_sat and k_cut are far too close to each other, probably in order to obtain a result that best approximates the model, we need to take into account not only the tuple with the lowest D, but also a minimum distance between k_sat and k_cut. Finally, the p-value shows that the null hypothesis cannot be rejected since $p > 1\%$, and consequently our distribution could be power law.

Comparison of all distributions processed so far

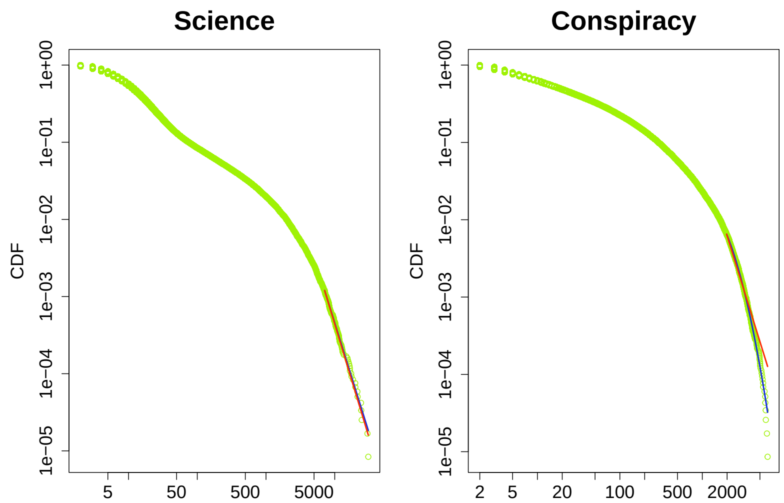

Figure 2.5: Comparison of the lognormal distribution (in blue) and the powerlaw distribution taking into account the cutoff and saturation (in red).

As can be seen from **Figure 2.5**, all distributions approximate our empirical distribution very well, however in the case of the lognormal distribution there is an **overfitting problem** that does not allow us to obtain a generally valid model for every distribution of this kind.

Furthermore, we can observe how from the first power law the α value is very high (even for the conspiracy community it reaches 6!), when theory tells us that a Scale Free network value should settle $2 \leq \alpha \leq 4$ (see Chapter 1.5 for further details) value.

Finally, another factor that makes us say that the simple power law suffers from **the problem of overfitting** are the values of the K-S distance D which are far too low.

In conclusion, it is safe to say that the best model is the second power law with k_sat and k_cut.

However, in order to obtain yet another numerical verification, we performed a *Vuong's test* and calculated the *two-sided p-value*, obtaining results which confirm the superiority of the latter over the lognormal model. The values obtained are the following:

- Science Community (Power law vs Lognormal):
  - Vuong's Test: 0.2023
  - Two sided p-value: 0.043
- Conspiracy Community (Power law vs Lognormal):
  - Vuong's Test: 0.3217
  - Two sided p-value: 0.035

The results of the Vuong's test together with the two sided p-value show that in both cases the best model is the power law distribution; in fact, the numerical value of the Vuong's test is positive, therefore preferring the first model, while the two sided p-value is $0.01 <$ p-value ≤ 0.05 so we can reject the null hypothesis and state that the test is significant (for a better understanding of the results see Section 1.6.2 and 1.6.2).

These attention patterns and mathematical confirmation reveal how both echo-chambers behave similarly and can be well approximated by a power-law distribution (but taking only the specific case it could be approximated by a lognormal distribution).

2.4.2 Correlation between different type of interaction

The next step was to check whether there is a correlation between the three different types of interaction; to do this, the most suitable Pearson Correlation was used.

The graphs in **Figures 2.6** and **2.7** show the correlation situation between the following pairs of interactions: like-retweet, like-reply, retweet-reply.

Looking at the correlation value, we can see that there is no difference in correlation between the two communities for the like-retweet pair, while the value is higher in the conspiracy community in the other two pairs.

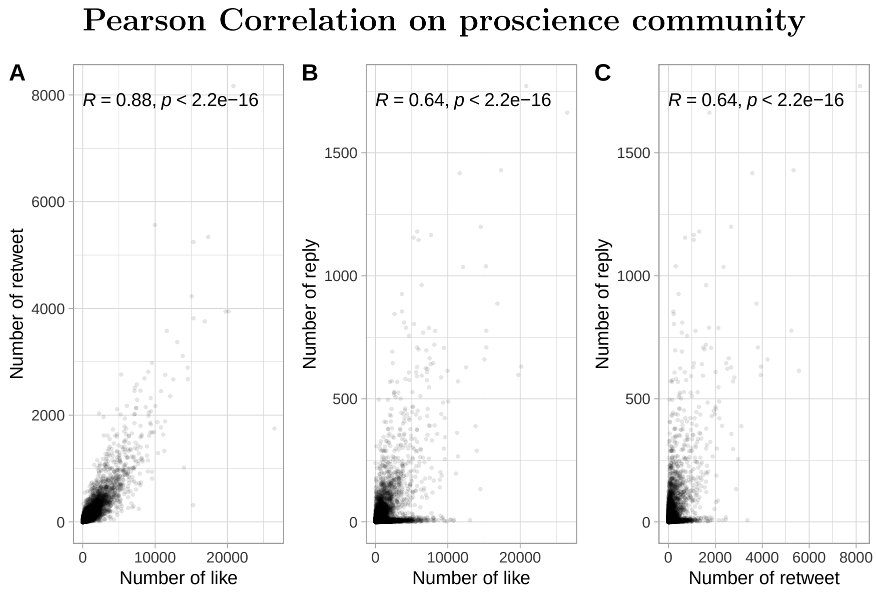

Figure 2.6: The three graphs showing the correlation between pairs of interactions (like, retweet, reply) for the proscience community

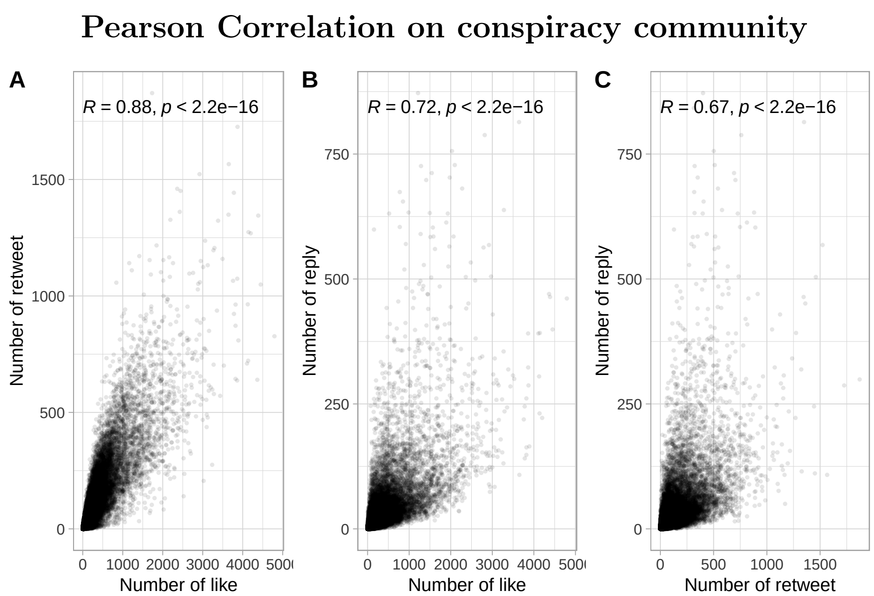

Figure 2.7: The three graphs showing the correlation between pairs of interactions (like, retweet, reply) for the conspiracy community

This gives us another very useful piece of information, namely that the conspiracy community is more likely to retweet a piece of content by commenting on it (and thus potentially create other subtopics from the same content).

The interaction of sharing (which is known to be more important for the spread of tweets), is supported by a comment that reinforces the shared content and thus the affiliation of the person in the community.

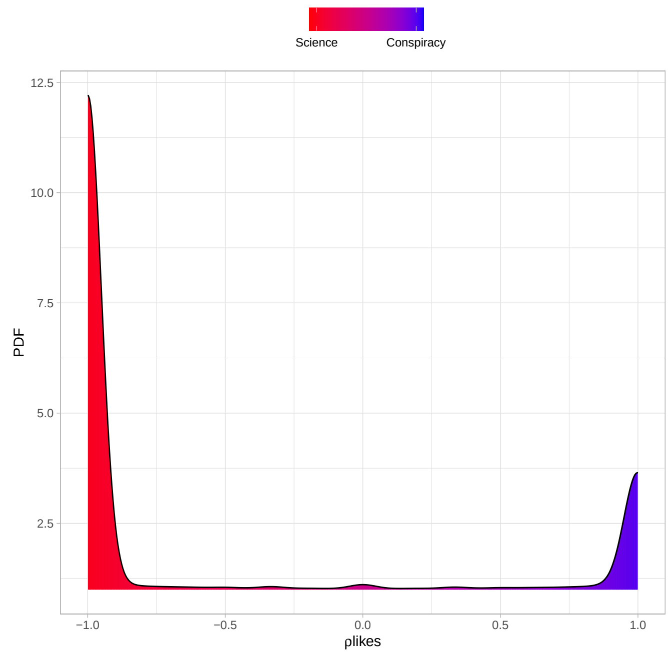

Figure 2.8: Probability density function (PDF) of the bias of all users counted against the total number of downloaded interactions.

2.4.3 Polarized Users

At this point, after analysing the tweets of the most influential users of each community (which we selected by hand), and after analysing what type of interaction users have with them, we move on to the phase of identifying the polarised users for each community.

The following graph in **Figure 2.8** shows the probability density function (PDF) for the polarisation of all the users in the dataset; as can be seen, most of the users show a very clear polarisation towards the two extremes, while only a very small proportion are not polarised to one side or the other.

Finally, this graph gives us yet another confirmation of how the community of pro-science users is much larger in number than that of conspiracists.

Once the polarised users were identified, the first analysis I thought it was necessary to see in percentage terms how the two types of iterations were distributed between the two communities.

The first graph in **Figure 2.9** shows how in absolute terms pro-science users performed the most interactions compared to conspiracy users.

It is possible, however, to notice a consistent percentage difference between the likes and retweets made by conspiracists; against 17% of likes, the percentage of retweets is considerably higher (21%). This could be a first clue as to why the conspiracy phenomenon sometimes appears larger than it is: conspiracy users tend to interact with information more by sharing than by liking.

This practice not only reinforces the dissemination of information in general in the social network, but also strengthens the loyalty of the conspiracy user.

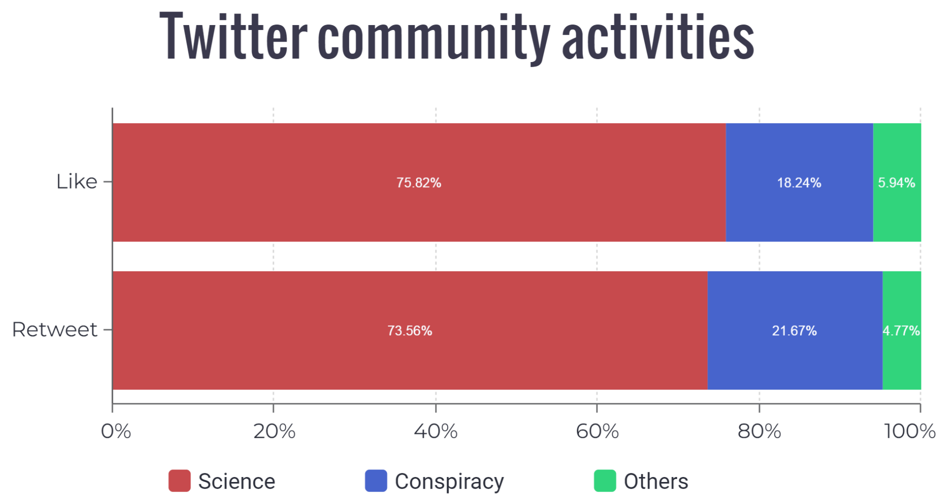

Figure 2.9: Graph comparing the two types of interaction (like and retweet) and the percentage of use in the two communities.

A very interesting fact concerns the number of users of the two communities who interacted with a tweet from a debunk profile: we can see that of 500,000 distinct users polarised in the scientific community, 117,000 interacted (corresponding to 23.4%), while of 119,000 users polarised in the conspiracy community, only 10,313 users interacted (corresponding to 8.7%). We can therefore state, as repeatedly confirmed by other publications in the past, that the scientific community is the largest consumer of debunk posts.

Interaction with debunking

This subsection is dedicated to investigating the behaviour of users of both analysed communities when interacting in the common field of debunk pages; the purpose of debunk pages should be to debunk fake news and rise above the debate to say what is objectively true from what is not.

We therefore took the users who had interacted with these posts from our dataset and calculated:

- the number of interactions they had with these posts;
- how much time (calculated in months) passed from the first interaction with a post on a debunk page to the last;
- calculate the polarity index before/after the first/last interaction with a debunk post.

At this point, we calculate the frequency of interaction respect to the the number of months and the difference in polarisation between before and after the interactions; these information came in handy for the representations in Figure 2.10 and 2.11.

Before looking at the results, it is good to remember that the difference in the polarisation index goes from -2 to +2, so when there are points in the graph with these values of coordinate, it will mean that the users have shifted their polarisation (they go from conspiracists to pro-science and vice versa).

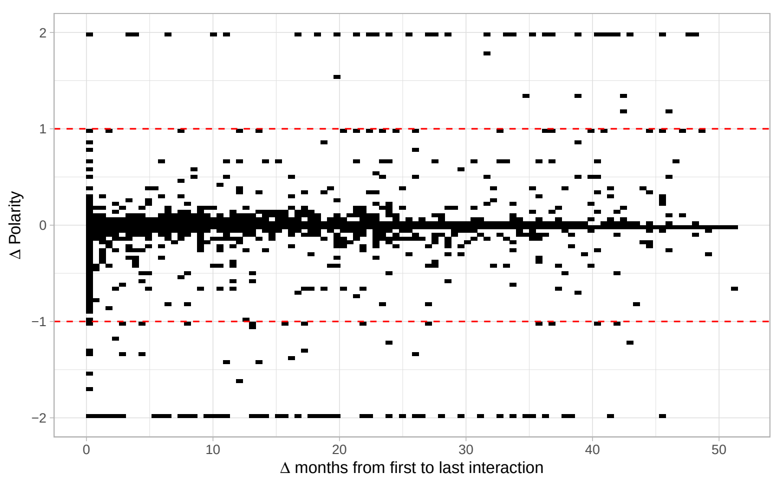

Figure 2.10: Binned scatter plot showing the distribution of users polarity among the months that has been passed from the first interaction on debunk news posts to the last. Above and below the red dashed lines there are the users who shift its polaritazion.

The first graph in **Figure 2.10** aims to highlight a possible correlation between the users' shift of polarisation and how long they have been interacting with debunk news; it can be seen that the interaction that lasted the longest is more performed by people who shift their polarisation towards the conspiracy community, while those who veered towards the scientific community had an interaction that lasted much less.

This result is very interesting as it could give us information on the exposure of debunk posts in the long run, or it could confirm once again how conspiracists are

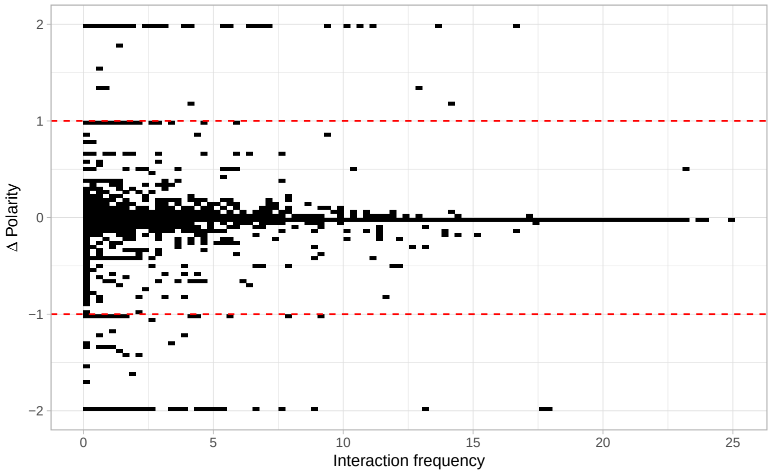

Figure 2.11: Binned scatter plot showing the distribution of the users taking into account the difference in bias and the frequency with which they interacted with debunking posts over time. Above and below the red dashed lines there are the users who shift its polaritazion.

more determined to defend their ideas by interacting in spaces that do not belong to their community.

Observing this result, I then proceeded to check whether this trend also existed in relation to the frequency with debunk posts; indeed, **Figure 2.11** reveals another important piece of information: there are few users who comment on debunk posts over a long period of time, yet those who do so during the time recorded a bias towards the conspiracy community.

As a final analysis, I wanted to analyse how users behaved in quantitative terms, i.e. to see if there was a shift in polarisation from one side to the other throughout this observation period (which would justify the role that debunk pages play, i.e. to depolarise the debate and downsize the community of conspiracy-mongers who use fake news). The results are not encouraging as we cannot see a consistent movement of people.

More in detail (people with neutral polarisation are not considered as they are perfectly split between the two communities):

- of the 37936 people polarised towards the scientific community, only in 110 we observe a shift in their polarisation (the **0.2% of people**);
- of the 5031 people polarised towards the conspiracy community, in 60 users we observe a shift of their polarisation (the **1.2% of people**).

These users are all contained above and below the red dotted horizontal lines positioned in the graphs at $y = 1$ and $y = -1$ respectively.

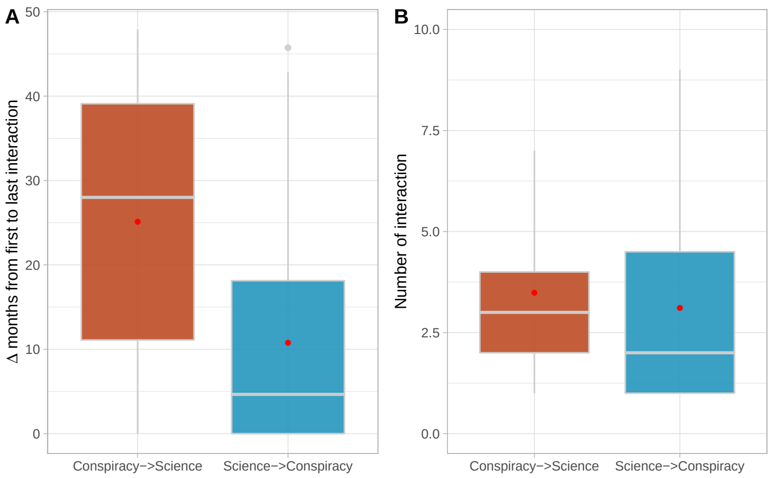

Figure 2.12: Box plot showing the behaviour of users who changed bias in terms of time spent interacting with debunk pages and number of interactions.

The last result that was processed concerns the users where we observe a shift in their polarisation; we wanted to see if they had any behaviour (on a quantitative level) that differentiated them. We noted that users who switched from conspiracy community to science community had a higher interaction time (about 25 months on average) than those who made the opposite switch (10 months on average). Finally, the average number of interactions is very similar, leading us to exclude this factor.

Pausing to think about the final results, we could see that a larger proportion of the conspiracy community changed their bias than the other side; however, to make this change on average takes more than twice as long to interact with these debunk pages.

Conclusions

It was recently demonstrated in [36] and [37] that there is a link between the polarization of users in social networks and the spread of disinformation (through the sharing of fake news). Furthermore, the existence and main characteristics of echo chambers have been studied in depth.

In this study, we investigated various models that could approximate the interactions between users of two communities (scientific and conspiracy) in order to apply them in other contexts. Also in relation to the two communities, a study was carried out on the correlations between the different types of interaction possible on Twitter, showing that the conspiracy community uses the sharing tool much more than the others (in fact amplifying the capacity for spreading the fake news that is the basis of the community itself).

Finally, the polarization index of users was studied, first showing the general distribution, and then going on to see how the polarization index changes over time when users interact in a common field that is the debunk pages. Quantitative parameters were calculated that could motivate this change in polarization such as the number of interactions or the frequency of them over time, showing how conspiracy users interact much more frequently and for longer periods with such debunk posts. Finally, the flows of people who moved from one community to another were analysed, finding that 1.2% of the conspiracy community moved to the scientific community, while only 0.2% made the opposite shift.

However, this shift did not occur with the same timing: while the average time of interaction with the debunk pages before the change of polarization was 25 months for conspiracists, it took only 10 months for those in the scientific community, with the same average number of interactions.

This result is certainly positive from a numerical point of view, as we can say that the conspiracy community has decreased by 1% in these 4 years, however, the shift of polarized users to the scientific community should be monitored, as it took them less than half the time to switch to the other side.

Future Work

This work can certainly be further developed by conducting this type of analysis on other type of social networks, opposing types of communities, or other nationalities (our Italian case in fact may not be representative of the European case or even compared to the world/general case), comparing the results with this one. It is certainly possible to associate this type of quantitative study with a qualitative one concerning the texts produced by the users and the quality of the interactions made (e.g. to understand which users may be bots, and which are the hubs and authorities of each community). With regard to the development of the part concerning the effectiveness of fake news, the network of pages could be better analysed, understanding which articles in particular have been most effective in depolarizing users and which have only exacerbated the debate and created ill-feelings on both sides.

Bibliography

[1] Amanda Lenhart. Teens, technology and friendships. 2015.

[2] K. E. Shearer, E. Matsa. News use across social media platforms 2018., 2018.

[3] Ital Communications. La buona comunicazione dell'emergenza quotidiana. 2022.

[4] K Garimella, G De Francisci Morales, A Gionis, and M Mathioudakis. Proceedings of the 2018 world wide web conference, 2018.

[5] Elizabeth Dubois and Grant Blank. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, communication & society*, 21(5):729–745, 2018.

[6] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports*, 6(1):1–12, 2016.

[7] Tali Sharot and Cass R Sunstein. How people decide what they want to know. *Nature Human Behaviour*, 4(1):14–19, 2020.

[8] Claire Wardle and Hossein Derakhshan. Information disorder: Toward an interdisciplinary framework for research and policymaking, 2017.

[9] Lee Howell et al. Digital wildfires in a hyperconnected world. *WEF report*, 3(2013):15–94, 2013.

[10] Giacomo Villa, Gabriella Pasi, and Marco Viviani. Echo chamber detection and analysis. *Social Network Analysis and Mining*, 11(1):1–17, 2021.

[11] Russell Torres, Natalie Gerhart, and Arash Negahban. Combating fake news: An investigation of information verification behaviors on social networking sites. In *Proceedings of the 51st Hawaii international conference on system sciences*, 2018.

[12] Costel-Sergiu Atodiresei, Alexandru Tănăselea, and Adrian Iftene. Identifying fake news and fake users on twitter. *Procedia Computer Science*, 126:451–461, 2018.

[13] Alexandre Bovet and Hernán A Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 10(1):1–14, 2019.

[14] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, 96:104, 2017.

[15] Abeer Aldayel and Walid Magdy. Characterizing the role of bots' in polarized stance on social media. *Social Network Analysis and Mining*, 12(1):1–24, 2022.

[16] Kurt Wagner and Sarah Frier. Elon musk backs out of $44b twitter deal on 'misleading representations' (twtr), Jul 2022.

[17] Jack Bandy and Nicholas Diakopoulos. Curating quality? how twitter's timeline algorithm treats different types of news. *Social Media+ Society*, 7(3):20563051211041648, 2021.

[18] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one*, 10(2):e0118093, 2015.

[19] Alessandro Bessi, Antonio Scala, Luca Rossi, Qian Zhang, and Walter Quattrociocchi. The economy of attention in the age of (mis) information. *Journal of Trust Management*, 1(1):1–13, 2014.

[20] Dave Lewis and Joss Moorkens. A rights-based approach to trustworthy ai in social media. *Social Media + Society*, 6(3):2056305120954672, 2020.

[21] Andrea Baronchelli. The emergence of consensus: a primer. *Royal Society open science*, 5(2):172189, 2018.

[22] Ana Lucía Schmidt, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. Anatomy of news consumption on facebook. *Proceedings of the National Academy of Sciences*, 114(12):3035–3039, 2017.

[23] Ashish Goel, Pankaj Gupta, John Sirois, Dong Wang, Aneesh Sharma, and Siva Gurumurthy. The who-to-follow system at twitter: strategy, algorithms, and revenue impact. *Interfaces*, 45(1):98–107, 2015.

[24] We're launching new engagement features, ways to discover groups and more tools for admins, Oct 2020.

[25] Emanuele Brugnoli, Matteo Cinelli, Walter Quattrociocchi, and Antonio Scala. Recursive patterns in online echo chambers. *Scientific Reports*, 9(1):1–18, 2019.

[26] Fabiana Zollo, Alessandro Bessi, Michela Del Vicario, Antonio Scala, Guido Caldarelli, Louis Shekhtman, Shlomo Havlin, and Walter Quattrociocchi. Debunking in a world of tribes. *PloS one*, 12(7):e0181821, 2017.

[27] Pedro Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. A measure of polarization on social media networks based on community boundaries. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages 215–224, 2013.

[28] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118, 2021.

[29] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118, 2021.

[30] R. Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among Internet news users1. *Journal of Computer-Mediated Communication*, 14(2):265–285, 01 2009.

[31] Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10):1531–1542, 2015.

[32] Axel Bruns. Echo chamber? what echo chamber? reviewing the evidence. In *6th Biennial Future of Journalism Conference (FOJ17)*, 2017.

[33] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.

[34] Alexander Hanna, Chris Wells, Peter Maurer, Lew Friedland, Dhavan Shah, and Jörg Matthes. Partisan alignments and political polarization online: A computational approach to understanding the french and us presidential elections. In *Proceedings of the 2nd Workshop on Politics, Elections and Data*, pages 15–22, 2013.

[35] Wesley Cota, Silvio C. Ferreira, Romualdo Pastor-Satorras, and Michele Starnini. Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, 8(1), dec 2019.

[36] Cheuk Hang Au, Kevin KW Ho, and Dickson KW Chiu. The role of online misinformation and fake news in ideological polarization: barriers, catalysts, and implications. *Information Systems Frontiers*, pages 1–24, 2021.

[37] Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, and Fabiana Zollo. Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, 13(2):1–22, 2019.

[38] Katrin Weigmann. The genesis of a conspiracy theory: Why do people believe in scientific conspiracy theories and how do they spread? *EMBO reports*, 19(4):e45935, 2018.

[39] Swati Goswami, C.A. Murthy, and Asit K. Das. Sparsity measure of a network graph: Gini index. *Information Sciences*, 462:16–39, 2018.

[40] File:economics gini coefficient2.svg.

[41] Albert-László Barabási & Márton Pósfai. *Network Science*. Cambridge University Press, 2016.

[42] Duncan J Watts. *Six degrees: The science of a connected age*. WW Norton & Company, 2004.

[43] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[44] Maarten Van Steen. Graph theory and complex networks. *An introduction*, 144, 2010.

[45] Matthew A Carlton and Jay L Devore. *Probability with applications in engineering, science, and technology*. Springer, 2017.

[46] Gregory W Corder and Dale I Foreman. *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons, 2014.

[47] M Hashem Pesaran. Non-nested hypotheses. In *Econometrics*, pages 167–173. Springer, 1990.

[48] Colin S. Gillespie. Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software*, 64(2):1–16, 2015.

[49] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.