



**Politecnico
di Torino**

Politecnico di Torino

Physics of Complex Systems

A.y. 2025/2026

Graduation Session March 2026

User-Source interaction patterns in misinformation spread:

**analyzing source consistency and source-topic relationships in
a fact-checked Twitter dataset**

Supervisors:

Luca Dall'Asta
Joana Gonçalves de Sá
Ana Vranić

Candidate:

Carlotta Burzio

A Tom

Abstract

The spread of misinformation online is a growing phenomenon that remains difficult to tackle due to its scale and complexity. Social media platforms play a major role in this spread because users are simultaneously consumers and publishers of information, serving as vectors for the dissemination of content that often escapes the veracity checks applied to traditional information channels. However, although misinformation in online environments and users' news consumption habits have been extensively studied, relatively little attention has been paid to the relationship between users and (dis)information sources. Understanding this link is important not only because the sources users choose to share from shape the pathways through which misinformation spreads, but they also reveal patterns of trust, exposure, and information diversity. For example, if users systematically rely on a limited set of (mostly social media) sources, this may reinforce information silos and facilitate the persistence and amplification of misleading content. This thesis investigates this relationship on Twitter/X by analyzing source preference (social media and conventional news) in a large dataset of shared fact-checked claims. In particular, it addresses two main questions: 1) How consistent are users in the selection of information sources they share from?, and 2) How does this behavior vary across user characteristics and external events?

For each user identified as having shared at least one fact-checked claim (from our database), sequences of posted tweets were constructed using the domain of the shared URL as a proxy for the information source (e.g., Twitter, Wikipedia). These sequences were then aggregated based on user characteristics (e.g., popularity, activity, gender, reliability) and transformed into networks, with domains as nodes and sharing as edges. Using descriptive network science metrics and entropy measures, employed to quantify the variety of sources to which users have access, behavioral patterns were investigated across categories. Two null models were used as benchmarks: one controlling for user-specific choice bias and the other for source popularity bias.

The results indicate that users tend to be consistent but limited in their source selection while being influenced by source popularity (i.e., source diversity is found to increase with activity and number of followers), but, contrary to previous results, not by gender. The persistence in users' source selection was closely linked to the topic of the shared content, revealing a strong source-topic association. Moreover, distinguishing social platform sources from publisher websites showed that users tend to be less segregated when sharing from a more diverse set of sources and that they return more to social media. Topic analysis over time shows that the COVID-19 pandemic led not only to an increase in health-related tweets, but also to an increase in entropy after 2020, reflecting a wider distribution of topics, while the sharing regularity remained stable.

Overall, these findings suggest that exposure to misinformation is influenced less by demographic factors and more by how people consume information, especially how concentrated their sources are and how much they rely on social media platforms.

Table of Contents

1	Introduction	1
2	Literature review	5
2.1	Disinformation and user behavior	5
2.2	Topics, languages, and the impact of COVID-19	6
2.3	Network approaches to information flow	7
2.4	Information flow metrics and source diversity	7
3	Background knowledge	9
3.1	Complex networks	9
3.1.1	Network properties	11
3.2	Entropy in information theory	13
4	Methods	15
4.1	Datasets	15
4.1.1	Global Claims dataset	16
4.1.2	Twitter dataset	18
4.1.3	Users dataset	18
4.2	Building users networks	19
4.2.1	Identifying the sources	20

4.2.2	Building sharing sequences	20
4.2.3	Building sharing networks	21
4.3	Consecutive topics analysis	22
4.4	Entropy analysis	23
4.4.1	Entropy measures	23
4.4.2	Null models	24
4.5	Favorability and Return scores	25
4.5.1	Favorability Score	25
4.5.2	Return Score	26
4.6	Entropy with propensity score matching	27
5	Results	29
5.1	Users characteristics and sampling	30
5.2	Networks Characteristics	32
5.3	Correlation between sources and topics	44
5.4	Restricted source consumption driven by popularity	46
5.5	Segregation around source types	53
5.5.1	Favorability score	53
5.5.2	Return score	56
5.6	Increase in user entropy after COVID-19	60
6	Conclusions	63
A	Networks Characteristics	67
B	Entropy Analysis	76
	Bibliography	82

Chapter 1

Introduction

Social media platforms have increasingly displaced traditional media as primary channels for news consumption [1, 2], fundamentally altering both the production and the distribution of information. Unlike traditional content diffusion models, online platforms enable any user to simultaneously become a potential broadcaster of information [3], effectively collapsing the distinction between audience and publisher. This has created an environment particularly favorable to the proliferation of misinformation, defined as false or misleading content that can spread through social networks so fast that it outpaces traditional verification mechanisms [4], and that represents one of the most significant and enduring challenges of the contemporary information landscape.

Understanding the phenomenon of misinformation requires attention to the behavioral patterns of those who spread it. In fact, exposure to (mis)information is far from neutral; selective exposure, intended as the tendency of individuals to prefer information that confirms pre-existing beliefs, has been extensively documented and explained both by personal interests and political leaning and by algorithmic and social amplification [5, 6, 7]. Studies have further suggested that misinformation spreads faster and more broadly than accurate information on social media [8], a finding that points to the importance of characterizing the mechanisms governing this phenomenon. The scale of this problem has been documented across different platforms, such as Facebook [5, 9] and Twitter [10, 11, 12]. The latter is the reference social platform used in this thesis.

What has received far less scrutiny, however, is the relationship between users and the specific sources from which they choose to share. Understanding this link matters because source selection reflects and reinforces the structure of individual

information consumption, with direct consequences for the spread and persistence of misinformation. Bertani et al. [10] provided a notable exception by analyzing the news media diet of disinformation spreaders through entropy-based measures, showing that such users tend to rely on a narrow and consistent set of sources. Extending this line of research working on the large-scale multilingual dataset of fact-checked claims compiled by Vranić et al. [13], this thesis explores how Twitter users interact with information sources across different user categories, determined according to online profile characteristics, content reliability, and source type preferences.

We first ask whether users’ sharing behavior exhibits patterns and whether these are uniform across the user population or depend on user characteristics. For example, do more popular users share from more diverse sources? This matters because heterogeneity in sharing behavior across user types would suggest that misinformation exposure is not uniform, and that different user profiles may require different interventions. Second, we test whether there is a correlation between source choice and the topic of the shared content. A tight coupling between source and topic would imply that users’ exposure to specific topics is not independent of where they get their information from. Third, we analyze whether users show consistency in their source choices and how much the popularity of certain sources impact individual selection. We hypothesize that more diverse source selection may indicate broader exposure to information and, potentially, less misinformation spreading (as in [10]). Finally, given that information ecosystems are not static, we ask whether large-scale external events, such as the COVID-19 pandemic outbreak, can perturb the observed behavioral patterns. It is natural to anticipate that the pandemic led to an increase in novel health-related claims and, therefore, to a shift in the topic distribution of our dataset, with health-related content gaining prominence, but did this topical change also leave a measurable imprint on users’ source sharing behavior?

To address these questions, a research pipeline was used, as illustrated in Figure 1.1. It relies on the Twitter dataset, containing 4.6 million tweets that shared fact-checked URLs, whose veracity was identified according to a collection of fact-checked claims: the Global Claims dataset [13]. In parallel, a Users dataset provides additional profile-based metrics describing user characteristics.

Using the Twitter dataset, temporal sharing sequences are reconstructed for each user, where the domain of each shared URL serves as a proxy for the information source. These sequences are then transformed into weighted, undirected networks that capture the structural relationships among the sources shared by the users.

Further, users are aggregated into categories, enabling comparisons of sharing behavior across groups. Several methods were then applied. First, network analysis examines the structural properties of individual sharing networks. Second, transition matrices are reconstructed to investigate topic-source continuity, assessing how changes in sources relate to changes in topics. Third, entropy-based measures quantify the diversity and consistency of sources that users share. These were compared with two null models: one addressing possible biases over personal choice (e.g., if a user shares ten times from one source and once from another), and another controlling for source popularity (e.g., unsurprisingly, Twitter posts - re-shares - were much more common than Facebook shares and this should be accounted for: the results would likely be different if we were studying another platform). Together, these benchmarks allow us to disentangle personal preference from popularity-driven behavior. Fourth, *favorability and return scores* capture patterns of segregation and loyalty in users' choices between different source types, such as social media and publisher websites (global news, commentary, institutional, and alternative media websites).

Finally, the pipeline concludes with a temporal analysis centered on the COVID-19 outbreak, using propensity score matching to evaluate whether the pandemic induced measurable changes in the relationship between users and the information sources they share, focusing on their entropy.

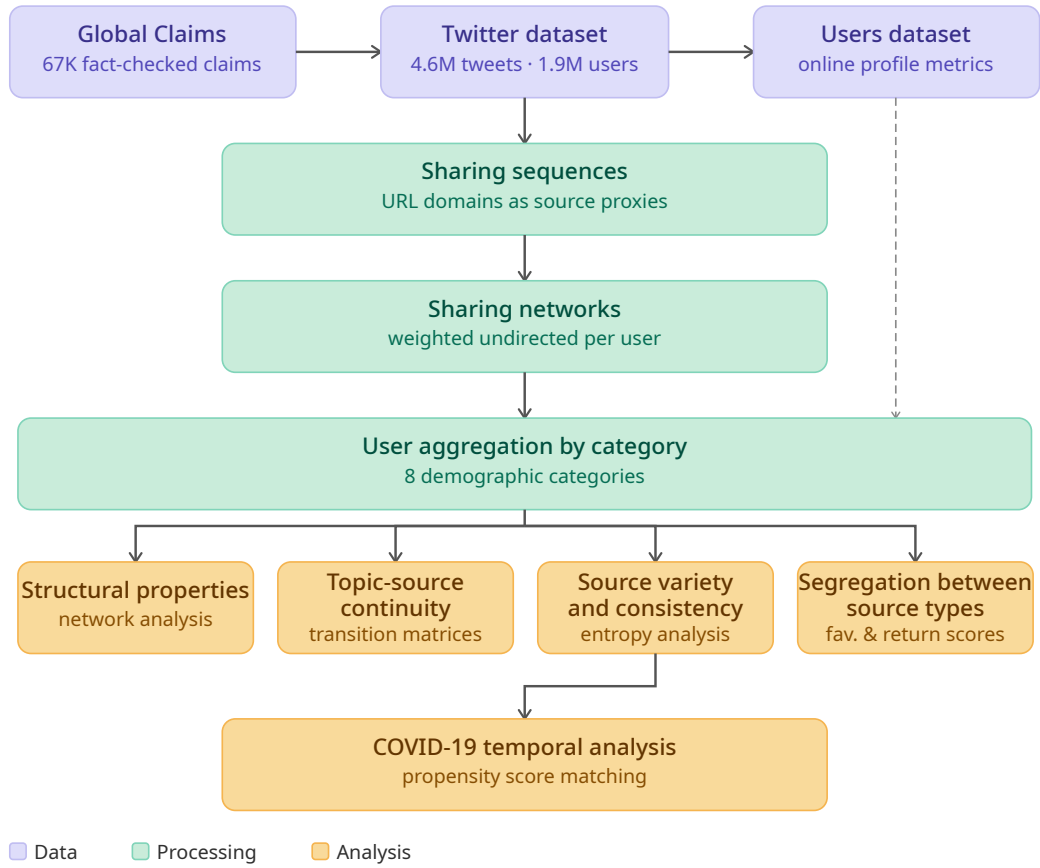


Figure 1.1: Research pipeline for analyzing user-source interaction patterns in misinformation spread.

Chapter 2

Literature review

The main goal of this thesis is to study how specific user preferences influence the spread of disinformation and, in particular, the extent to which users diversify news sources and how these preferences vary across different characteristics and over time. In this chapter, we summarize recent research on disinformation and discuss established methods to study (mis)information spread on social networks.

2.1 Disinformation and user behavior

The large scale diffusion of false or misleading information through digital platforms has attracted growing scientific and public attention over the past decade, driven by increasing evidence of its real-world consequences, including influence on public opinion [14, 15, 16], electoral outcomes [17], and deterioration of institutional trust, affecting societal values and redefining beliefs [18, 19, 20]. A landmark contribution by Vosoughi et al. [8] suggested that false news spreads faster, deeper, and more broadly than accurate information, and this difference is attributed at least partially to human users.

These dynamics are amplified by the shift in the way people access news. Social media (e.g., Facebook, Twitter) have become a primary news source in many countries [1, 2]. On these platforms, any user can act as a potential broadcaster of information [3], bypassing the editorial and verification processes of traditional journalism. Algorithmic personalization further deepens the problem by confining users to ideologically homogeneous information environments [21, 22, 5, 23], while

individual consumption habits introduce additional biases, such as the tendency to draw from a narrow set of sources focusing on one or two topics of interest [24], or to favor the most popular ones [25], creating *social bubbles*.

A growing body of research has focused on characterizing users who spread misinformation, rather than the content itself. A recurring finding is that sharing is highly concentrated: a small minority of users accounts for the majority of false content diffusion [26, 27, 28, 12, 29], with politically polarized individuals and leading accounts playing a disproportionate role [30, 31, 32, 33]. However, findings show that not all political leanings lead to similar sharing behavior; for instance, during Argentina’s 2019 election on Twitter, center-right users shared predominantly from center-right media, while the media consumption of center-left individuals was more diverse [34]. In parallel, demographic factors further shape both exposure and sharing behavior [35, 36]; for example, research showed that male and white users tend to be more active in sharing news, biasing the audience exposure to their interests [37]. Moreover, users’ inclinations towards sharing misinformation tend to persist across topics and over time [38].

2.2 Topics, languages, and the impact of COVID-19

As discussed above, most existing research on misinformation has focused on politically charged content, predominantly in English and primarily in the US context [4, 12, 17, 26, 15, 16]. The COVID-19 pandemic represents a notable exception to the political focus of the literature, generating a substantial body of research on health misinformation. It largely confirmed the patterns described above: the concentration of spreading among a minority of users, the community segregation, and the persistence of individual sharing habits over time [14, 39, 11, 36, 38, 40]. However, interestingly, some authors found that users turned primarily to trustworthy outlets during this period [41], and that accurate content drove more engagement than misinformation [42], suggesting that the relationship between external shock and misinformation dynamics is not straightforward.

While this work has provided important insights, existing misinformation datasets are often restricted to specific topics [40], single fact-checking organizations [43], or single countries [44, 45, 46], limiting the generalizability of the findings across different cultural and informational environments. Furthermore, little is known about whether source preferences vary across user demographics, topics, and languages. It also remains unclear whether the behavioral patterns

identified in the literature [47] are stable individual traits or whether they can be disrupted by large-scale external events. The construction of the *Global Claims* dataset [13], used in this thesis, directly addresses these gaps by providing a large-scale multilingual and multi topical collection of fact-checked claims and associated tweets spanning an extended period of time.

2.3 Network approaches to information flow

Network-based methods have proven particularly productive for studying information diffusion and user behavior on social media. Early theoretical work established foundations for understanding how information cascades propagate through social networks [48], later exploited to characterize the structural differences between misinformation and accurate information diffusion [8, 49].

The majority of this work models users as nodes and their interactions (retweets, follows or co-exposure) as edges, revealing for instance that retweet networks can capture significant signals of news consumption patterns [50]. Bipartite user-source networks have also been employed, showing that accounts sharing content from similar sources tend to have similar credibility profiles [51] and that news sharing behavior reflects persistent ideological polarization driven by media outlet preferences [52]. A complementary perspective places sources rather than users as nodes of the networks. Connecting sources by audience overlap has revealed that source credibility is associated with structural differences in network topology [53], and that source-centered networks can effectively characterize the structure of the information landscape independently of individual user behavior [54].

This thesis adopts a related source-centered perspective, constructing networks representing consecutive sharing events for each user, allowing for a direct characterization of source diversity and consistency at the individual level.

2.4 Information flow metrics and source diversity

Beyond network structure, quantifying the diversity and regularity of information consumption requires dedicated metrics. Information theory has been adopted as a useful framework for this purpose, with Shannon entropy [55] providing a measure of uncertainty and heterogeneity in probability distributions. In the context of online information consumption, entropy has been applied to detect echo-chambers

and characterize the concentration of users' information exposure [56], and to quantify uncertainty in the spread of misinformation on social media [57, 58]. A particularly relevant precedent is the work of Bertani et al. [10], which applied entropy-based measures to characterize the news media diet of disinformation spreaders on Twitter, finding that such users tend to concentrate their sharing on a narrow set of sources and exhibit high regularity in their source selection. This work represents the most direct methodological predecessor of the present thesis, which extends the entropy framework to a broader range of user categories and adds network-level analysis.

While entropy captures the diversity and regularity of users' source choices, it does not account for the temporal order in which sources are encountered and shared. To capture the sequential structure of sharing behavior, Markov chains have also been employed in related contexts, from modeling information diffusion of social networks [59] to detecting media polarization [60] and studying user browsing patterns [61], suggesting that the order in which users encounter and share content carries meaningful information about their online behavior.

Overall, while the dynamics of misinformation spread, the structure of news media diet, and the role of user behavior have been each studied in depth, the specific question of how users relate to their information sources in terms of consistency, diversity, and topical coupling and how these patterns vary across user characteristics and in response to external events has not been systematically addressed. The combination of network analysis and entropy measures applied to a large multilingual fact-checked dataset employed in this thesis represents a step towards filling that gap, offering a user-centered characterization of source interaction patterns that extends beyond a single topic or cultural context.

Chapter 3

Background knowledge

This chapter offers an overview of what constitutes a *complex network* and outlines the *networks properties* that will be analyzed throughout the thesis. Moreover, it introduces the concept of *entropy* as it is used in *information theory*.

3.1 Complex networks

In general, a *network* is a set of *vertices*, also called *nodes*, connected by *edges* (or *links*) (Figure 3.1) studied in the form of mathematical graph theory [62]. There are many systems in the real world that can be represented as networks. Examples include the World Wide Web, where nodes are webpages and link connections between them are edges, metabolic networks, where substrates and products are the nodes, and edges represent reactions, postal delivery routes, with locations as nodes and paths between them as edges, and citation networks, where nodes are the papers and edges are the references. In fact, interest in network science exploded during the first years of the 21st century as a tool to encode the interactions between the components of a *complex system*, its collective behavior being difficult to derive otherwise [63].

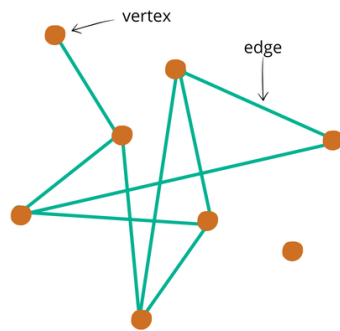


Figure 3.1: Example of a small network with eight vertices (nodes) and nine edges.

3.1.1 Network properties

Networks can also be weighted or unweighted. In the latter, edges are binary in weight, which means that they can be present (weight equal to one) or absent (weight equal to zero) [64]. In the first, each edge (i, j) acquires a unique weight w_{ij} , which classifies the connection. Examples of weighted networks can be social networks where weights indicate the strength of the relation between two people (vertices) or power grids where the weight is the amount of current flowing through a transmission line. In addition, networks (or graphs) are called undirected when there is no preferred direction in the flow from one vertex to another (for example, just indicating that two people are connected), or directed when all edges point to a defined direction (for example, when A shares content from B, the direction of the edge would be from A to B, but if B shares content from A, the edge and its weight would remain the same but the direction would reverse). In this case, not all nodes can be reached from a given node, and different network properties, such as the appearance of communities, depend on the starting point. Figure 3.2 illustrates the graph representation of these different configurations.

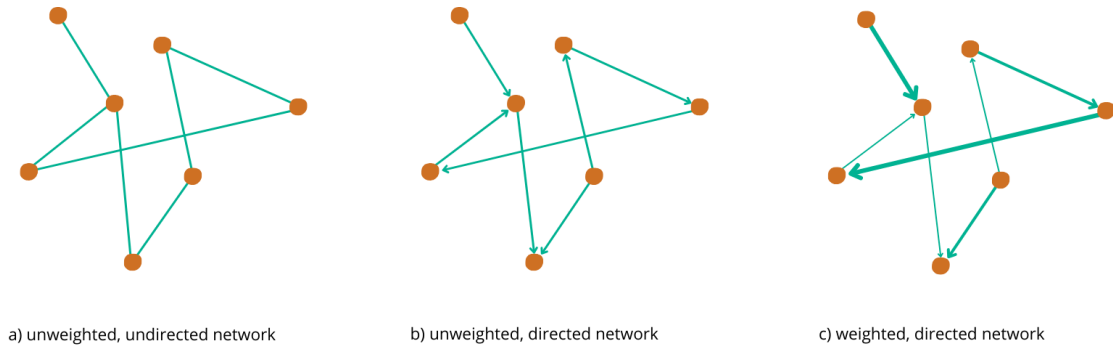


Figure 3.2: Examples of different network types: a) unweighted and undirected, b) unweighted and directed, c) weighted and directed.

When dealing with a network, different properties can be analyzed. The following definitions are based on the *Network Science* book by Albert-László Barabási [63]. Many of these properties can be computed using the Julia packages `Graphs` and `MetaGraphsNext`. The latter allows for the addition of metadata to vertices and edges, facilitating the insertion of weights and complex labels.

- The **size** of the network is the number of nodes N in a network and represents the number of components of the system. It allows one to have preliminary

information about the extension of the network, still without knowing anything about the number of links and the wiring between vertices.

- The **density** is defined as the total number of links over the maximum number of links. It can be interpreted as a measure of connectivity, since it describes the proportion of possible links in a network that are actually present. Low density indicates sparse connections, while high density indicates tight connections. It is usually represented as follows:

$$d = \frac{L}{L_{max}} = \frac{2L}{N(N-1)}$$

where L is the number of edges in the network, L_{max} is the maximum number of edges, and N is again the number of nodes. However, in the networks used in this work, we will often consider self-loops, representing self-interactions of nodes with themselves (for example, if user A shared consecutively from the same source more than once). Therefore, to include self-loops in the density computation and obtain a meaningful result between 0 and 1, the definition of the maximum number of edges becomes:

$$L_{max, self-loops} = \frac{N(N+1)}{2}$$

- The **average degree** is defined as the mean number of connections (edges) per node. In undirected networks, which are the ones that will be used in this work, it is represented by:

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

where k_i is the degree of the node i , which is the number of links it has to other nodes. To account for weights and self-loops, one must carefully compute k_i : each self-loop counts as two edges, while supposing the network has integer weights, one can interpret each edge as the number of edges corresponding to its weight and count the degree consequently. This can be easily generalized to cases for which weights are not integers.

- The **clustering coefficient** captures the degree to which neighbors of a given node link to each other. It is a measure of the network's local link density, and it is defined as follows:

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

where L_i is the number of edges between the k_i neighbors of vertex i . Its values range from 0 to 1. When $C_i = 0$, none of the neighbors of vertex i is connected to each other, while when $C_i = 1$ the neighbors of vertex i form a *complete graph*, where every neighbor links to each other.

- The **average clustering coefficient** captures the degree of clustering of the whole network. From the probabilistic point of view, it is the probability that two neighbors of a randomly selected node link to each other. It is defined as follows:

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

which is the average of the clustering coefficient of all nodes.

- The **diameter** is defined as the maximum shortest path in the network, i.e., the maximum distance between any pair of its nodes [64]. Often, it is computed by exploiting the *Breadth-First Search* (BFS) algorithm. In case of a weighted network, the internal function **diameter** of the Julia package **Graphs** can account for weights when computing the shortest path.
- The **betweenness centrality** offers a metric to identify the best connected nodes in the network. The betweenness centrality of a vertex i is the number of shortest paths between other vertices that run through i . It can be interpreted as a measure of network resilience: it encodes the number of shortest paths that would become longer if that vertex were to be removed from the network [62, 65, 66]. It is defined as follows:

$$bc(i) = \frac{1}{Z} \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

where σ_{jk} is the total number of shortest paths from node j to node k , $\sigma_{jk}(i)$ is the number of those paths that pass through i (excluding where i is an end point) and Z is the number of pairs of nodes not including i .

3.2 Entropy in information theory

In *information theory*, where information can be treated as a physical quantity [67], entropy measures how uncertain or unpredictable a random variable is. Essentially, it captures how much information is required, on average, to describe the outcome of a variable given how its probabilities are distributed across all possible states. The concept was proposed by Shannon in 1948 [55], who introduced a formula to calculate the entropy of information analogous to the mathematics of statistical thermodynamics. In fact, Shannon entropy is defined after the Boltzmann and Gibbs entropy theory [68] as follows:

$$H(X) = \mathbb{E}[I(X)] = \mathbb{E}[-\log(p(x))] = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x)$$

where X is a discrete random variable taking values in the set χ and distributed according to probability p , $\mathbb{E}$ is the average operator and I is the content of the information of X . Often, the logarithm is in base 2, because information theory was developed around binary communication systems.

$H(X)$ can be seen as the average information per emitted symbol and indicates the degree of non-determination of the system from an informational point of view [69].

This formula has been developed by assuming that the source is a *DMS* (Discrete Memoryless Source), so an independent identically-distributed random variable. Mathematically, this implies relying on a complete system of events in which a final equilibrium of states can be reached. Although these conditions can usually be satisfied in technical systems (used as models by Shannon himself), this is not always true in biological, economic, or social systems [69]. Nevertheless, Shannon entropy has often been used as a diversity index or uncertainty measure, a practical proxy of heterogeneity in systems where the strict assumptions of information theory may not hold [10, 56, 57].

Chapter 4

Methods

This chapter describes the data and the methodological framework that underlies the thesis analysis. It introduces the three datasets used, outlines the process of constructing user sharing networks from URL domains, and presents the analytical tools employed: transition matrices for the source-topic analysis, entropy measures and their null models for quantifying source diversity, favorability and return scores for investigating users orientation towards source types, and a propensity score matching approach for the COVID-19 temporal analysis.

4.1 Datasets

The focus of this work is to analyze the relationship between online users and the information sources they spread when sharing content on social media platforms. Compared to previous datasets, which focused only on specific events or topics [40, 46] or are country-specific [44, 45], the datasets used in this research cover a wide range of topics and multiple languages and countries over an expanded period of time, allowing for a comprehensive analysis of user behavior. From the *Global Claims* [13] dataset, a collection of URLs of fact-checked claims classified by veracity and topic, the *Twitter dataset* and the *Users dataset* were constructed by collecting the Twitter posts containing those URLs and the profiles of the users who shared them.

4.1.1 Global Claims dataset

The *Global Claims* dataset [13] consists of a large annotated collection of 67,000 fact-checked claims gathered from more than 200 fact-checking organizations and covering 50 languages. It combines the *Google Fact Check Tool*, a Google search engine for fact-checked results [70], with direct data extraction from fact-checking websites. Each claim is linked to its original URL and was classified according to its veracity and topic.

The veracity labeling of the different fact-checking sources was standardized into a unique *veracity scale*:

- **true**: verified factual statements;
- **false**: misleading or demonstrably inaccurate claims;
- **other**: claims with mixed accuracy;
- **satire**: false claims, but with a humorous or satirical purpose and, therefore, separated from misinformation [71]

Most claims were labeled false, in line with the idea that implausible statements are more likely to attract the attention of fact-checkers [72], while the relatively small number of true claims likely seemed doubtful enough to justify verification.

The classification of topics was conducted using a hybrid annotation approach that combined expert, non-expert, and ChatGPT-based annotations. Firstly, experts created a *codebook* that defined 5 main topics and 23 subtopics according to the Comparative Agenda Project [73]. The list of topics is illustrated in Table 4.1. These topics served as a baseline that was later refined after several rounds of classification (see [13] for details). Mturk and ChatGPT annotators were allowed to assign up to two (sub)topics to each claim. This approach enabled the construction of a probability distribution across topics, allowing analysis that capture topic overlap (e.g., identifying politicians discussing COVID-19) or, alternatively, more conservative analysis that relies only on the most probable main topic.

Claims were also classified by *salience* according to their presence on Twitter (high salience indicates more shares); in fact, the URLs associated to the claims were queried on the platform, identifying approximately 7,000 distinct URLs shared by Twitter users.

Table 4.1: List of topics and subtopics used to classify claims.

Topic	Subtopic
Politics	Political Events and Institutions
	Politicians
	Foreign Affairs and Defense
	Social Groups and Identity
	Immigrations
	Welfare and Economy
Health	Other
	Nutrition
	Vaccination
	Covid-19
	Health Problems
Environment	Other
	Activists
	Environmental Problems
Science	Other
	Scientists
	Scientific Discoveries
Society	Other
	Celebrities
	Events and Places
	Crime
Other	Other

4.1.2 Twitter dataset

As mentioned above, all claims URLs were searched on Twitter using the *Twitter Academic API* (now discontinued), allowing the identification of approximately 7K claims shared on the platform. These were shared by 1,932,773 unique users, totaling 4,660,483 tweets posted between 2013 and 2022 (around 9 years and 2 months). These correspond to over 800,000 original tweets and more than 3 million retweets (80.11% of the dataset).

Because these tweets reference claims from the *Global Claims* dataset, they can be directly linked to their associated annotations. Therefore, every tweet that shares a claim’s URL inherits the corresponding veracity and topic labels.

As shown in Figure 4.1, most tweets are labeled false, following the trend of the *Global Claims* dataset. The most frequent topic is politics, followed by health, science, environment, and other.

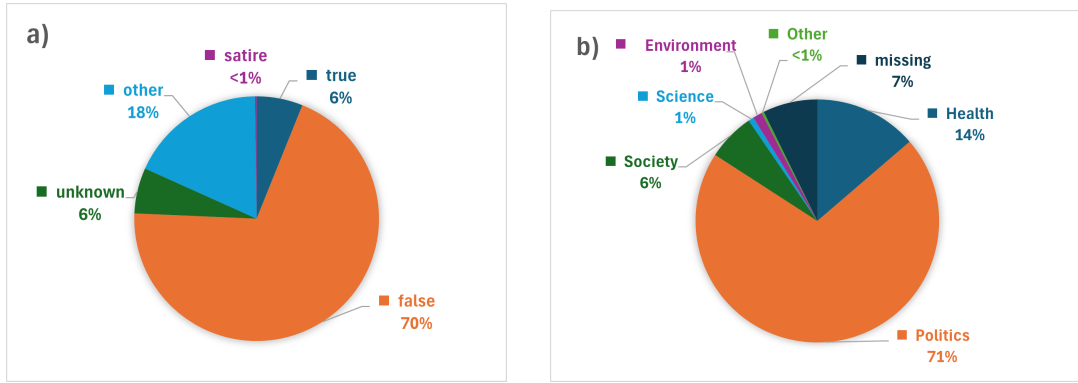


Figure 4.1: a) Pie chart showing the distribution of veracity labels in the Twitter dataset. b) Pie chart showing the distribution of topics in the Twitter dataset.

4.1.3 Users dataset

Once the tweets were collected, the publicly available information for the users who shared or re-shared at least one of the fact-checked claims was retrieved. In particular, this dataset includes the profile name, username, verification status, the number of followers and following, the number of tweets posted at the time of the collection, and information from self-reported fields such as biography and location.

Additionally, gender and location of users were inferred. Gender was identified, when possible, using a multistep heuristic algorithm [74] from name, username, biography, and location. The **gender-guesser** Python library was applied to the first token of the name and, for unidentified cases, gender-related keywords were searched in the users' bio and location fields. Gendered word lists were compiled in different languages to allow identification in a multilingual context. Then, a manual validation process was performed. Instead, a geoparsing tool (**geoparser** Python library) was used to infer the location of users to the country level, because the location field on Twitter is self-reported and expressed as free text. Country was eventually inferred for 859,669 users (44.5%), while gender was assigned to 1,052,400 users (54.4%).

Figure 4.2 shows the structures of the datasets and how they are linked to each other.

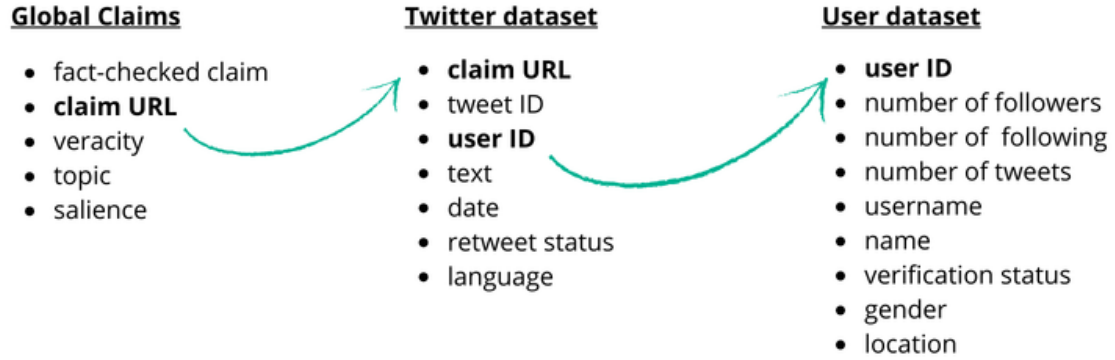


Figure 4.2: Diagram showing how the three datasets (Global Claims, Twitter and Users) are linked to each other and their structure.

4.2 Building users networks

In the *Twitter dataset*, each tweet shared by each user contains a URL. In fact, for each claim, as explained in 4.1, the embedded URL was collected and its domain identified. The latter was used as a proxy for the information source. This allowed us to analyze how users share and propagate content subject to fact-checking from specific URL domains across the Twitter platform. Sources were used to create a temporal sharing sequence and a weighted undirected network for each user, and

users were aggregated into different categories.

4.2.1 *Identifying the sources*

In the dataset, the sources of information are represented by the domains of the URLs of the fact-checked claims. Following the *DNS* (Domain Name System) hierarchy [75], for each URL the first- and second-level domains were filtered. For example, if the URL is `https://en.wikipedia.org/wiki/Domain_Name_System`, the domain is `wikipedia.org`, where `.org` is the first-level domain, and `wikipedia` is the second-level domain.

An exception was `twitter.com`; in fact, 81.1% of the collected tweets have `twitter.com` as a domain. This appears consistent with the notion that Twitter users tend to favor sharing content that originates within the platform itself [13]. However, the disproportionate usage of this domain would have led to a too rough analysis. Thus, to allow for a wider variety of sources, the username was included in the domain: `twitter.com/username`. In this way, each Twitter profile is considered as a unique information source (e.g., `twitter.com/therealdonaldtrump` and `twitter.com/nickiminaj` are considered as two different sources, instead of aggregating them under the same source `twitter.com`.) To be coherent, the structure of the URLs coming from other Social Networks, namely *Instagram*, *YouTube*, and *Facebook*, was checked to see if the same kind of domain structure used for managing the *Twitter* URLs should have been applied. However, in those platforms, the username is not included in the URL of the posted content, therefore the identification of the user profile proved too challenging. Nevertheless, their presence represents only 2.4% of the collected tweets, thus the original domain definition was applied, keeping only the first and second levels (e.g., for URLs of the kind `https://www.youtube.com/watch?v=TRL7o2kPqw0`, only `youtube.com` was kept as source).

4.2.2 *Building sharing sequences*

Once the sources were defined, for each user, the domains of the URLs were collected, constructing a temporal sequence of information sources. Obviously, one should keep in mind that this temporal sequence does not include the whole set of tweets shared by each user, but only those present in the dataset, meaning those including a link which has been fact-checked. Ultimately, the sequence of collected tweets corresponds to the set of misinformation the user shared.

4.2.3 Building sharing networks

Lastly, each sequence was converted to an undirected weighted network. Nodes represented unique domains and edges represented consecutive sharing (e.g., if user A shared a URL from `twitter.com/username` and then from `bbc.uk`, there is an edge connecting these two domains). If the same pair of sources has been shared more than once, the weight of the edge increases by 1 for each repetition. Self-loops represented the same source being shared more than once consecutively (e.g., if user B shared `youtube.com` and then `youtube.com`, there will be a loop at that source node). Figure 4.3 illustrates the process.

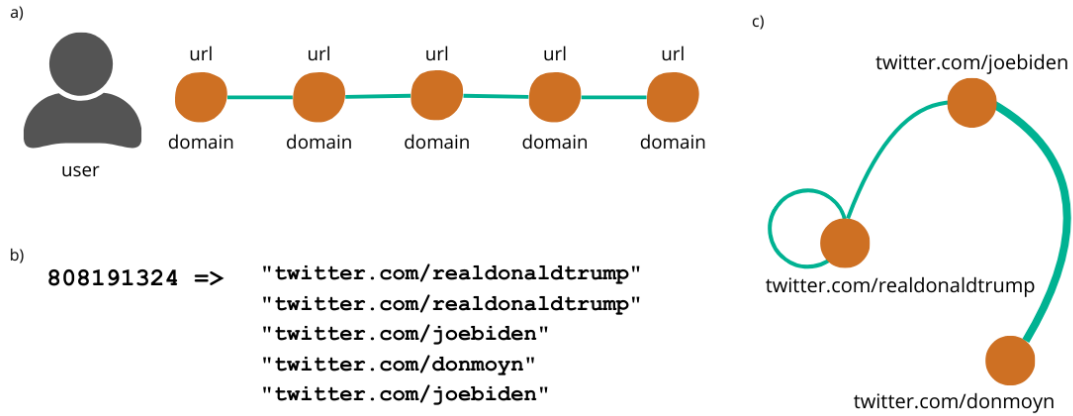


Figure 4.3: Example of a network based on a user's tweets sequence. a) Illustration of the temporal sequence of users. b) Example of the sequence of tweets shared by a specific user. c) Network constructed from the previous sequence. The thicker edge indicates greater weight, indicating that the user returned to the pair of domains. The loop represents the consecutive share of the same domain.

In order to retain only meaningful sequences and networks, two different user filters were applied, both based on their sharing sequences. The first one retains user sequences that have a minimum length of two, allowing to define a minimal network structure. Thus, it focuses on users sharing at least two fact-checked tweets. The second filter yields sequences containing at least four different sources, limiting the analysis to users with a moderate-to-high source variety.

To evaluate the properties of the sharing sequences and the relative networks, users were then aggregated into categories, according to their online profile characteristics.

The evaluated categories were:

- **verification status**
- **gender**
- **country** of origin
- **popularity**: number of followers at the time of the data collection
- **activity**: number of tweets posted at the time of data collection
- **influenceability**: number of following at the time of data collection
- **reliability**: percentage of tweets labeled false over the total number of tweets containing a fact-checked URL
- **source type** of tweets containing a fact-checked URL

The latter category distinguished social platforms, namely `twitter.com`, `youtube.com`, `facebook.com`, and `instagram.com`, from all other types of source present in the dataset, which can be global news, commentary, institutional, and alternative media websites, and were collectively called *publisher websites*. This name identifies websites where content is produced, edited and distributed under the control of the site’s owning organization rather than contributed primarily by users, contrary to what happens on social media.

Once the categories have been determined, an analysis of the user networks aggregated by category was performed to compare different network features. Size, density, average degree, average clustering, diameter and betweenness centrality were computed using the definitions in section 3.1.1.

For the most unbalanced categories, a sampling procedure was performed keeping as a reference the least populated subgroup, and the analysis was repeated applying both user filters previously mentioned.

4.3 Consecutive topic analysis

As mentioned before, for each tweet also the topic was retained. Claims were classified into six different main topics, choosing the most probable among the annotated topics (details in section 4.1.1): politics, health, society, science, environment, and other.

To investigate possible relations between sources and topics over time, Markov transition matrices were computed. Markov chains have been used to model information diffusion on social networks [59], detecting media polarization [60] or classifying fake versus genuine information flow through probabilistic transition paths [76]. In this context, transition matrices were used to investigate topic probabilities in sequentially posted sources. For each user, the sequence of sources of the posted tweets was considered. Each element of the transition square matrix was the count of how often, for each pair in the sequences, the next tweet is about one topic, given the topic of the previous one. Matrices were then normalized by row.

The goal was to inspect the frequency with which topics tend to change in consecutively posted tweets and infer whether there is a relation with the change of source. Firstly, all pairs of consecutively posted tweets were observed. Then, the same analysis was performed considering only pairs where the source was the same. Finally, the analysis was repeated keeping only pairs in which the two considered tweets had different sources, thus removing all consecutive repetitions of source.

4.4 Entropy analysis

One of the most interesting patterns to investigate regarding misinformation is source diversification.

Inspired by a paper by Bertani et al. [10], entropy measures were exploited to explore source diversification among different types of users. In general, entropy has been widely used in research related to misinformation spread, from detecting echo-chambers [56] to investigate credibility and uncertainty in communication ecosystems ([58], [57]). In this user-focused context, entropy is an uncertainty measure that quantifies the variety of sources to which each user has access. Two different measures of entropy were used to better explore whether some pattern can be recognized between users.

4.4.1 Entropy measures

Firstly, *Random Entropy* was measured, representing the most natural method to infer source diversification. It is defined as follows:

$$S_{rand,i} = \log_2(N_i)$$

The mathematical definition is borrowed from text information theory (details can be found in section 3.2), explaining thus the logarithm in base 2. Here, N_i is the length of the set of sources shared by the user i . It can be interpreted as a predictability measure of the media diet, assuming uniform usage since all domains are considered equally likely.

Then, the temporal uncorrelated entropy was computed as the *Shannon entropy*. The mathematical equation is the following:

$$S_{unc,i} = - \sum_{j=1}^{N_i} p(x_j) \cdot \log_2(p(x_j))$$

Where $p(x_j)$ is the observed frequency of the source x_j in the sequence shared by the user i . Thus, it is the historical probability that a web domain x_j was shared by the user i . This kind of entropy is a finer measure of heterogeneity of the news media diet of each user because it captures how unevenly domains are shared, even though the specific order of the sequence is ignored.

These two entropy measures become more informative when taken together. That is why *Shannon Entropy* was normalized over *Random Entropy*. This ratio measures consistency in domain usage. In fact, the random entropy value acts as a baseline entropy that captures uniform usage, so the ratio captures how much user choice deviates from the uniform choice. The closer it is to one, the more uniformly the user chose from the batch of sources she shared.

$$S_{norm,i} = \frac{S_{unc,i}}{S_{rand,i}}$$

where i is the user.

These two measures of entropy and their ratio were computed over users aggregated by category, applying separately both filters mentioned previously (users sharing at least twice and users sharing from at least 4 different domains). In particular, special attention was given to the average entropy of each category subgroup to highlight possible different behaviors.

4.4.2 Null models

To compensate for some of the biases in the dataset, two null models were proposed to compare with the findings.

The first null model aimed to remove the bias over user's choice of sources and therefore compensate possible individual personal preferences of users toward

certain sources. Thus, the distribution of the sources was kept fixed, but each sequence was randomly recreated for each user, sampling without replacement. This operation was repeated 100 times. Then, for each user, the entropy of each new sequence was computed and averaged over the number of repetitions to obtain an average measure of the entropy for each user in the null model.

The second null model, instead, aimed to remove the bias over source popularity and therefore compensate for the unbalanced distribution of sources in the dataset. This time, the length of the sequence of each user was kept fixed and filled randomly assuming a uniform distribution over the set of unique sources, making them equally probable. The procedure was repeated 100 times. After that, for each user, the entropy of the generated sequences was computed and averaged over the number of repetitions to estimate the average user entropy in the null model.

4.5 Favorability and return scores

The web domains present in the dataset under analysis can be clearly separated into two subgroups, namely *social platforms* and *publisher websites*, as suggested during the division into categories mentioned previously (4.2.3). Approximately, the publisher websites represent 65% of the sources, dropping to 60% when filters are applied to users.

To gain a better insight on users relation with the type of source, two indexes were implemented: the *favorability score* and the *return score*.

4.5.1 Favorability Score

The so-called *favorability score* is a measure from which the composition of an audience according to a binary characteristic can be inferred. In previous work, this score has often been associated with the composition of political ideology [5, 6, 7], but in this context it has been linked to the type of source shared by users (social platforms or publisher websites). The favorability score was defined as:

$$fav_i = \frac{S_i - N_i}{S_i + N_i}$$

where S_i is the number of social platforms sources shared by the user i , while N_i is the number of publisher websites shared by the same user. Normalization constrains the score in $[-1, +1]$. Thus, a favorability score was obtained for each

user. Positive values indicated a higher proportion of social media sources and negative values indicated a higher proportion of publisher websites. The metric therefore provided a normalized behavioral indicator of source orientation in the sharing of misinformation.

As already done with previous analysis methods, scores were aggregated by user category to compare differences in behavior, if any, between subgroups. The usual two different filters over users were applied separately.

4.5.2 *Return Score*

Due to the unbalanced distribution of sources across the collected tweets, in order to measure the relative tendency to share again previously shared sources, a second metric was introduced: the *return score*, defined as follows:

$$ret_i = \frac{S_i(R) - N_i(R)}{S_i + N_i}$$

The logic is similar to the favorability score, but here $S_i(R)$ denotes the number of shares of the same social platform, while $N_i(R)$ denotes the number of shares of the same publisher website. The denominator normalizes the score by the total length of the sequence considered and restricts the score values in $[-1, +1]$. The return score was intended to capture the loyalty to the source type in the re-sharing behavior. Positive values were expected when a user shared previously encountered social media sources more frequently than previously encountered publisher websites, whereas negative values indicated the opposite trend. The return score exactly equal to zero presented an ambiguity because by construction it could indicate both an equal number of re-shares from both types of sources and no re-shares at all. Thus, to avoid interpretation issues, users who never re-shared any source were excluded from the analysis.

As was already done for favorability scores, the scores were aggregated by user category to compare differences in behavior, if any, between subgroups. The usual two different filters over users were applied separately.

To observe the same issue from a coarse-grained point of view, a focus on the behavior of mixed users (those who shared from both social platforms and publisher websites) towards source type was proposed by constructing a Markov transition table. The underlying question was the probability that a user who shared a URL

from a social platform (publisher website) would have shared again from the same source type in her next tweet. Consecutive pairs of tweets in the users' sequences were analyzed, and results were presented as a heatmap normalized by rows.

4.6 Entropy with propensity score matching

Since the *Twitter dataset* covers a time period that includes the outbreak of the COVID-19 pandemic, an initial analysis was conducted to determine whether this event modified the distribution of health-related content in the dataset by considering how the different topics percentages varied between years. Subsequently, to assess whether this possible change in the distribution of topics affected the variety of sources shared by users, the average entropy of users sequences was computed before and after 2020, that is, before and after the outbreak of the COVID-19 pandemic.

To obtain a non-zero entropy, only users posting from at least two different sources before and/or after 1 January 2020 were considered, obtaining a set of 438,955 users. The users were then separated into two groups, one containing those who posted before the COVID-19 outbreak and one containing those who posted afterward. Users in the two groups were considered independent, even when they are present in both.

Then, users in the two groups were matched by propensity score. *Propensity score matching* is a statistical method developed to address the lack of randomization inherent in real-world treatment, policy, or intervention allocation, which is influenced by characteristics which in turn influence the outcome [77]. Propensity score methods emulate idealized *randomized controlled trials* (RCTs), minimizing baseline differences that are systematically present in observational studies. It serves as a balancing score: it collapses multivariate baseline data into a scalar probability. When subjects are matched or compared based on their propensity score, those who received the treatment and those who did not will tend to have similar background characteristics at baseline [77, 78]. Mathematically, the propensity score is defined [78] as follows:

$$e_i = Pr(Z_i = 1|X_i)$$

where $Z_i = 1$ indicates the treatment assignment to subject i and X_i represents the vector of observed baseline covariates. In this context, the outbreak of COVID-19 was considered the treatment; therefore, users who posted from 2020 represented the

group receiving treatment ($Z = 1$), while users who posted before 2020 represented the control group ($Z = 0$).

It is mandatory to include in the computation the *confounders*, variables that affect both exposure and outcome; their selection should be guided by subject-matter expertise, because it affects the quality of the analysis. Here, the online profile information, namely the number of followers, the number of following, and the number of tweets, were chosen as confounding parameters. The set of selected users was thus cleaned of those for which these parameters were not available, with 422,664 users remaining for analysis. The length of the sequence of each user was added as an additional confounding parameter because it has a fundamental role in computing entropy over sources.

The standard model used to estimate the propensity score is *logistic regression*, in which treatment status is regressed on the observed baseline characteristics [77]. Nevertheless, other methods have also been employed, such as bagging or boosting [79, 80] or neural networks [81].

The logit model can be formulated as follows:

$$\text{logit}(e(x)) = \log \frac{e(x)}{1 - e(x)} = \log \frac{\text{pr}(z = 1)}{1 - \text{pr}(z = 1)} + Q(x) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$$

where $e(x)$ is the propensity score, x_i is the vector of covariates of the individual i , p is the total number of covariates and β_j are the regression coefficients.

The matching is then performed by pairing treated and untreated subjects with similar scores [78]. The most common implementation is one-to-one matching [77], which can be carried out with or without replacement [82]. To select untreated subjects with a similar propensity score to that of a treated subject, the most common approaches are nearest neighbor matching, which identifies the closest available match, and nearest neighbor matching with a caliper distance, which restricts matches to those falling within a predefined distance threshold. In the following thesis, nearest neighbor matching with replacement will be employed, exploiting the `MatchIt` Julia package.

After matching the users in the two groups by propensity score, random entropy, Shannon entropy, and normalized Shannon entropy were computed. The results were then compared with a null model obtained by randomizing over the popularity of the sources, which means that the length of the sequences was kept as the original ones, but they were randomly populated with replacement from the set of unique sources (as it was done in the general computation of entropy, section 4.4.2).

Chapter 5

Results

We investigated how users relate to information sources when sharing misinformation on Twitter, aiming to explore how consistent they are towards source selection and how this behavior varies across different demographic and online profile factors, source types, and external events.

First, we analyzed the properties of the sharing networks of users aggregated by categories to explore possible patterns across user characteristics. Then, we inspected the variation of tweet topics in consecutive sources through Markov transition matrices and found a close correlation between topic and source choice. To gain a clearer insight into the consistency and regularity of source sharing, we analyzed the entropy of the user sequences, comparing it with two null models concerning users individual choices and source popularity. After separating sources in two main groups, social platforms and publisher websites, we employed two distinct measures, *favorability score* and *return score*, to explore the segregation of users between these two source types. Finally, noticing that the dataset temporally expands throughout the COVID-19 pandemic outbreak, we linked the expected increase in attention to health-related content to an increase in the variety of sources, computed by matching users before and after 2020 through propensity score.

5.1 Users characteristics and sampling

As mentioned in 4.1.2, the dataset contains almost 2 million unique users, but, for the majority, only a single fact-checked tweet was collected. Therefore, as outlined in 4.2.3, we applied two different filters to the users set to retain only the most meaningful ones. The first filter selected users who shared at least two tweets, excluding those who posted misinformation only once. This allowed the creation of a minimal network composed of at least one edge (which may be a loop if the same source was shared twice). After applying this filter, 641,729 users remained for the analysis (33% of the total). The second filter was more restrictive, focusing on users who shared from at least four different sources, exhibiting greater diversity in sources selection. The application of this filter drastically reduced the number of users to 122,390 (6% of the dataset). This stricter filter allowed for a precise behavioral analysis of users with a relatively diverse media diet, excluding users with too extreme behavior. This choice was also motivated by the will to compare our results with those by Bertani et al. [10], who applied a similar filtering in the analysis of entropy (discussed later).

We then classified the users into eight different categories (4.2.3), each of which contained between two and four subgroups. For some categories, specifically popularity (number of followers), activity (number of tweets), and influenceability (number of following), the boundaries of the subgroups were arbitrarily chosen, with the aim of being both socially meaningful and as balanced in size as possible. Another source of arbitrariness was in the choice of countries, for which we kept only the top three by the number of users, but also in the definition of reliability, which, differently from [10], we defined as a function of the proportion of fake news among the total number of shared fact-checked tweets. We kept the extremes, so users who always share false content and those who never do, and split the remaining users into two groups. Additionally, also the source type category relies on the arbitrary choice of separating users sharing exclusively from publisher websites from those sharing exclusively from social platforms and those sharing from both kinds of source. Table 5.1 illustrates the different subgroups and their population.

We observed that, despite the efforts, some categories remained heavily imbalanced. To address this, we performed a sampling procedure on the four most imbalanced ones prior to any analysis: *verification*, *country*, *source type* and *reliability*. In each case, the smallest subgroup was used as a size reference and the larger subgroups were sampled 100 times, with results averaged across samples, in order to obtain comparable measures.

Table 5.1: Number of users in each category for each filter applied. In the first column, users sharing less than two tweets were neglected. In the second column, users sharing less than 4 different sources were neglected.

Category	Subgroups	Number of users sharing at least 2 tweets	Number of users sharing from at least 4 different domains
verification	verified	15,819	3,866
	unverified	602,999	113,428
gender	male	164,691	20,262
	female	118,686	18,385
country	USA	204,899	44,763
	India	16,022	2,237
	UK	15,780	2,099
popularity	<100 followers	111,924	101,86
	(100-1K) followers	302,579	50,329
	(1K-10K) followers	178,442	49,123
	>10K followers	25,873	7,656
activity	<10K tweets	162,692	4,881
	(10K-100K) tweets	368,640	72,416
	>100K tweets	87,486	39,997
influenceability	<100 following	40,782	4,558
	(100-1K) following	305,739	41,466
	>1K following	272,297	71,270
reliability	0% fake news	59,265	639
	<50% fake news	15,6011	23,454
	>50% fake news	178,100	85,647
	100% fake news	248,353	12,650
source type	publisher websites	49,401	4,057
	social platforms	427,953	40,248
	mixed	164,375	78,085
total		641,729	122,390

5.2 Networks Characteristics

To investigate if possible patterns in source sharing are related to specific demographic factors or online profile metrics or instead there are common behaviors across users who generally share misinformation, we created a network for each user, in which the nodes are the sources and the edges are consecutive shares and started by analyzing different characteristics of the networks aggregated by user categories (e.g., gender, verified status, etc).

For each category, we computed the size, density, average degree, average clustering coefficient, and diameter distributions. The two previously discussed filters on users were applied: one excluded users who posted only once in the dataset and the other excluded users who shared from fewer than four different sources. In the following analysis, we will consider the latter, but the results obtained applying the first filter can be found in Appendix A and follow the same trends. Moreover, as mentioned above, four categories (verification, reliability, country and source type) were balanced for size through sampling, but we did not control for the number of posts.

The top left panel in Figures 5.2 - 5.9 shows the probability mass function of *size*, plotted in logarithmic scale, for each category grouping. The number of sources shared by each user varies from 4 (our lower-cutoff) to 43 and, as expected, follows a power law distribution across all categories. This means that the vast majority of users have small networks, with less than 10 nodes (i.e., shares from only a few sources), while a small number of users share from more than 10 different sources. Network size gives us a preliminary idea about the variety of sources each user shares. Focusing first on the characteristics of the online profile, namely popularity (number of followers), activity (number of tweets posted also outside of the dataset), influenceability (number of following) and verification status (Figures 5.2, 5.3, 5.4, 5.5), we see that users who interact with a smaller social bubble (few followers and/or few following and/or relatively low activity) tend to have smaller sharing networks. This is also true for verified users, which in general have an higher number of followers with respect to unverified users (the mean number of followers of verified users is 700,534 when considering the broader filter and 497,820 when considering the stricter one. Unverified users, instead, have a mean number of followers, respectively, of 1,841 and 2,809). Regarding the percentage of false news shared (Figure 5.7), reliable users (i.e., those who did not share any false URL that we could identify), have extremely small networks, the largest having only eight nodes.

The top right panel in Figures 5.2 - 5.9 illustrates the probability density function of *density* as an histogram grouped in six equal bins. Density, which can be interpreted as a measure of connectivity and provides information on how structured sequential sharing is, is generally low, between 0.2 and 0.5 for most networks in each category. This indicates that sparse connections prevail over tight ones. It should be noted that when applying this filter to users, we do not obtain any complete graph (graph where each node is connected to every other node, and therefore the density is equal to one). Complete graphs are only obtained when a small number of sources (below 4) is considered, as can be seen in Appendix A.

The middle left panel in Figures 5.2 - 5.9 shows the *average degree* distribution plotted in logarithmic scale. It represents the average number of co-occurrences each source has with the other sources, which can be interpreted as the overall connectivity of networks. The results show that average degree follows approximately a power law distribution, showing a concave curve in log-scale. Large values are thus less probable than a pure power law would predict, and nodes in most networks have an average degree below 10. This is consistent with the low density found previously.

To further investigate connectivity, we computed the *average clustering*, which is a measure of the average local cohesion of nodes in networks. Its probability density distributions are showed in the middle right panel in Figures 5.2 - 5.9. We can observe a distinct peak in zero common to all user categories, which means that networks often do not contain triangles and are organized in tree-like structures (Figure 5.1) and/or contain many self-loops. The preference of tree-like structures may suggest that users, even when repeating sources, tend to have a few favorite sources to which they come back often rather than randomly jumping from one source to the other in the pool of the sources they share. Focusing on the average clustering distribution after the peak, we see that users with few followers (Figure 5.2) tend to have more clustered networks than users with more followers. A similar consideration can be done for users sharing fewer tweets (Figure 5.3), for users sharing only from social platforms (Figure 5.8) and also for highly unreliable users (Figure 5.7). Indian users networks clustering distribution is massively peaked around zero, while USA users tend to have more structured networks (Figure 5.9).

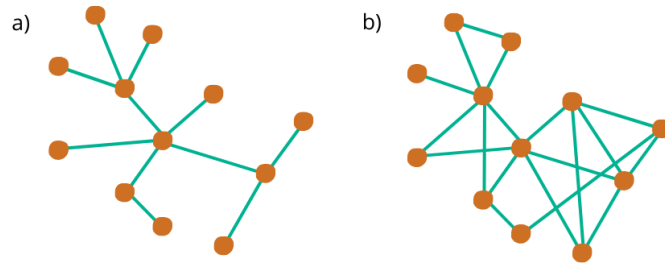


Figure 5.1: a) Example of a network with a tree-like structure, where neighbors of nodes are not connected and therefore the average clustering coefficient is zero. b) Example of a network with many triangles and therefore an high average clustering coefficient.

After the considerations made on the average clustering distribution, we would expect a quite large diameter since, in general, tree-like structures imply longer shortest paths between nodes with respect to clustered networks. However, the *diameter* distribution of the networks (bottom left panel in Figures 5.2 - 5.9) always peaks between 3 and 4. This can be explained by the abundance of networks with small size. Focusing on activity (Figure 5.3), it is noteworthy that the diameter distribution becomes broader with the number of tweets; this suggests that users who post more share in a less structured way with more elongated networks. The same trend, even if less emphasized, can be seen with the increase in the number of followers (Figure 5.2) and partially with the percentage of fake news (Figure 5.7). Moreover, users sharing from mixed sources tend to have less structured networks than users sharing from only one type of source (Figure 5.8).

Figure 5.6 shows how there are not noticeable differences in the sharing behavior dictated by gender; the properties of the networks of males and females are surprisingly similar to each other, in contrast with previous research that found male users more active in sharing [37]. This is also true for users distinguished by verification status (Figure 5.5), with the exception of size.

It is interesting to see which are the nodes with the highest *betweenness centrality*, so the best connected nodes across the different users networks groups. Taking into account the entire dataset, the five most recurring sources with the highest betweenness centrality are `twitter.com/realdonaldtrump`, `washingtonpost.com`, `twitter.com/drpaulgosar`, `twitter.com/danscavino`, and `rt.com`. When aggregating users by categories, these five sources often return among those with the highest betweenness centrality, with some remarkable exceptions. When considering users grouped by location, results show that, despite having networks with similar properties (Figure 5.9), users sharing from India have a completely different set of nodes with highest betweenness centrality with respect to users from the USA and UK. This is not surprising, as it reflects the popularity of different public figures: `twitter.com/yuvadesh`, `twitter.com/rssurjewala`, `twitter.com/msisodia`, `twitter.com/koenamitra`. On the other hand, the only other user group that shows a quite different set of sources among the nodes with highest betweenness centrality is the group of users who do not share any false URL; here, the nodes with the highest centrality are `twitter.com/donmoyn`, `twitter.com/kenidrarwoods_`, `youtube.com`, `twitter.com/realdonaldtrump`, and `twitter.com/manufacturingni`.

Users networks distributions: popularity

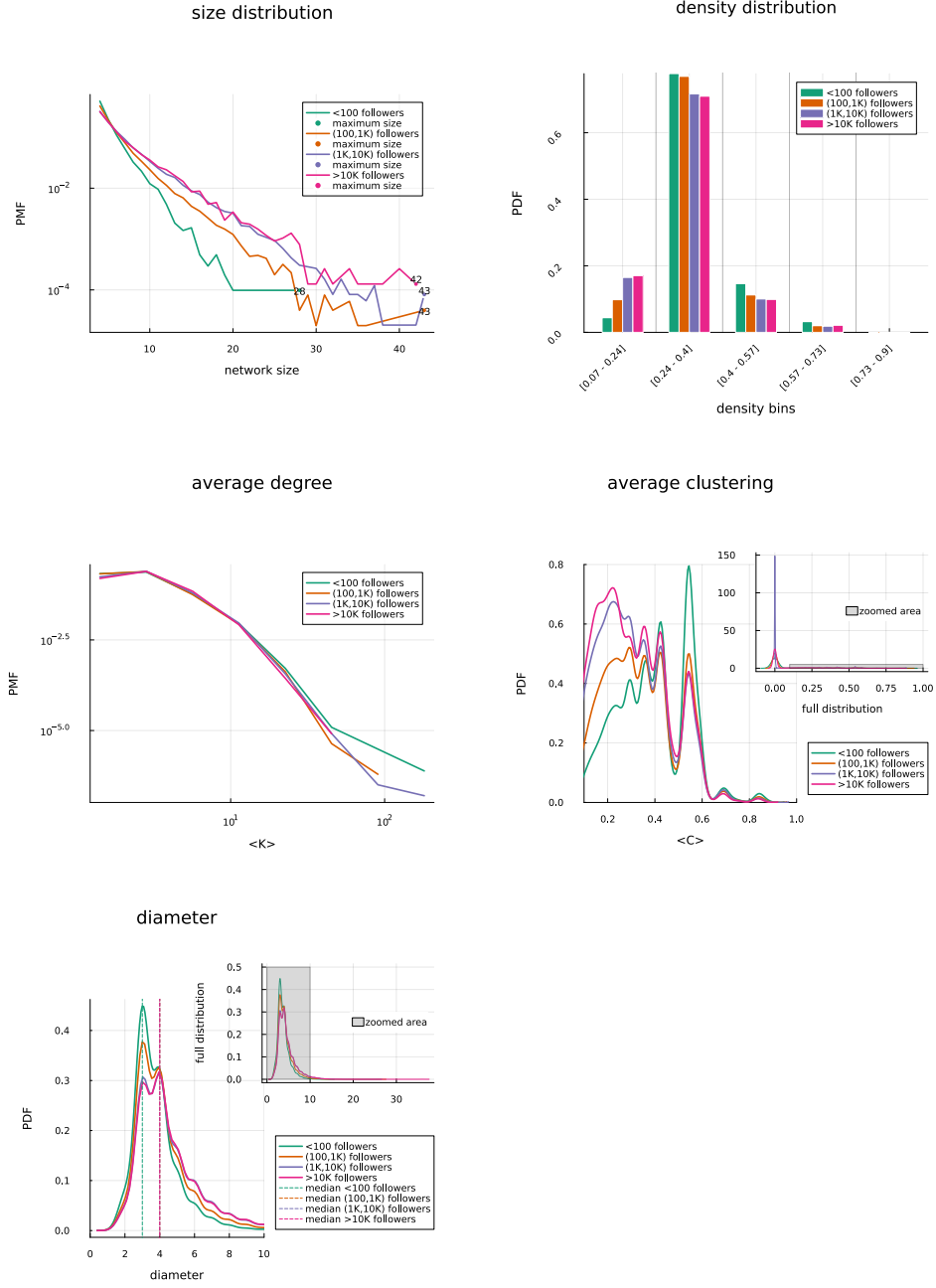


Figure 5.2: Network characteristics of users sharing from at least 4 different domains aggregated by popularity (number of followers).

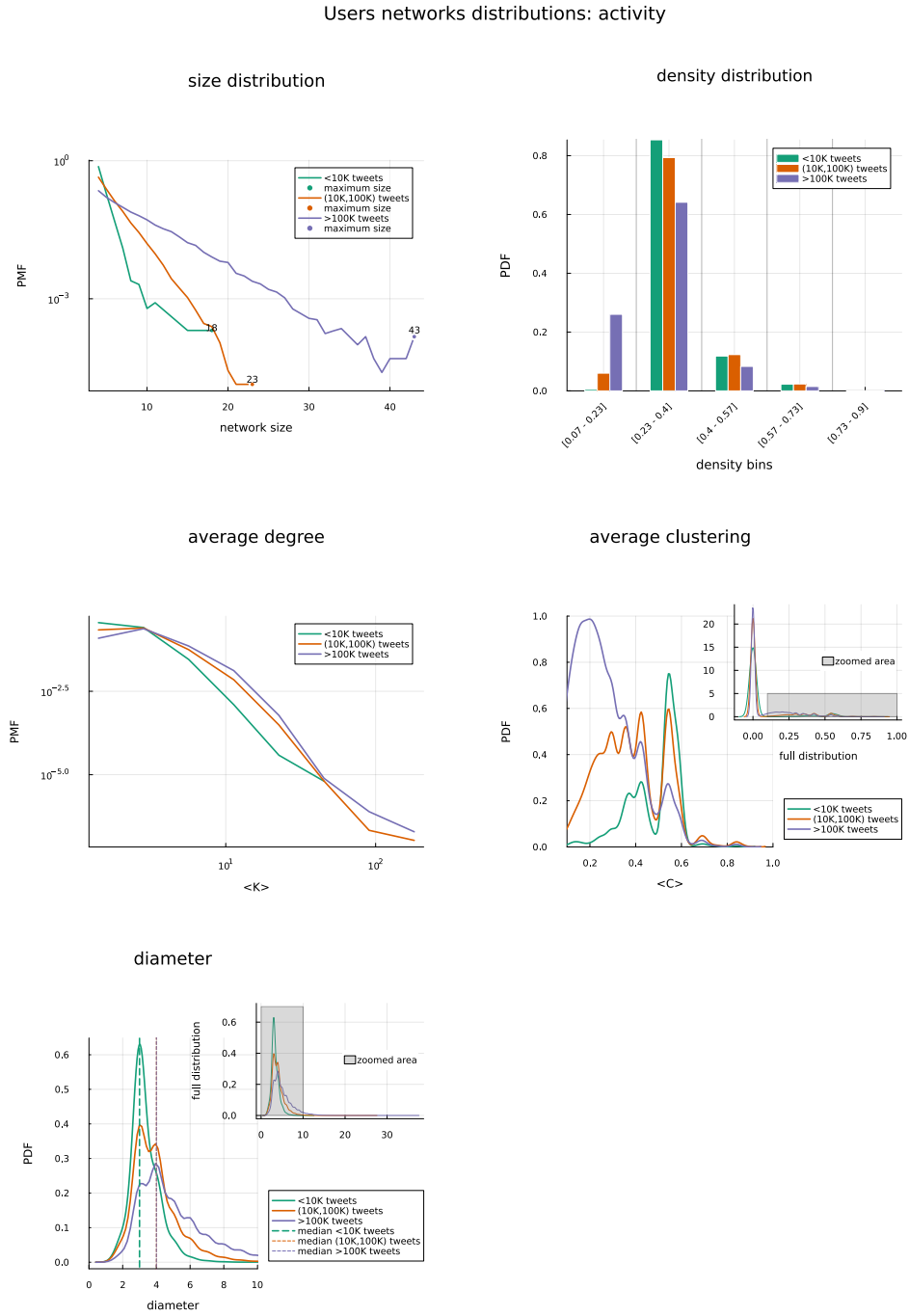


Figure 5.3: Network characteristics of users sharing from at least 4 different domains aggregated by activity (number of tweets).

Users networks distributions: influenceability

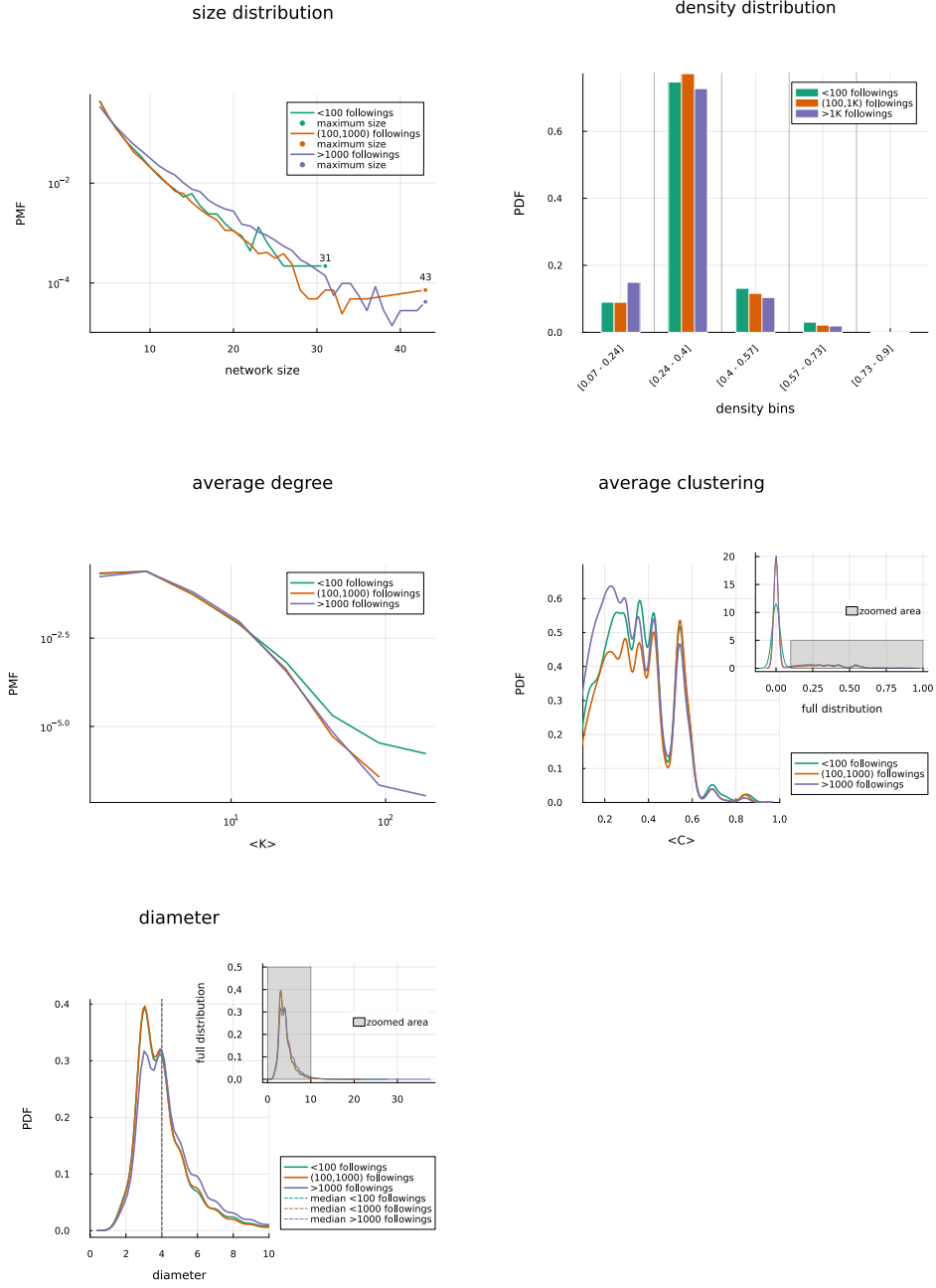


Figure 5.4: Network characteristics of users sharing from at least 4 different domains aggregated by influenceability (number of following).

Users networks distributions: verification (with sampling)

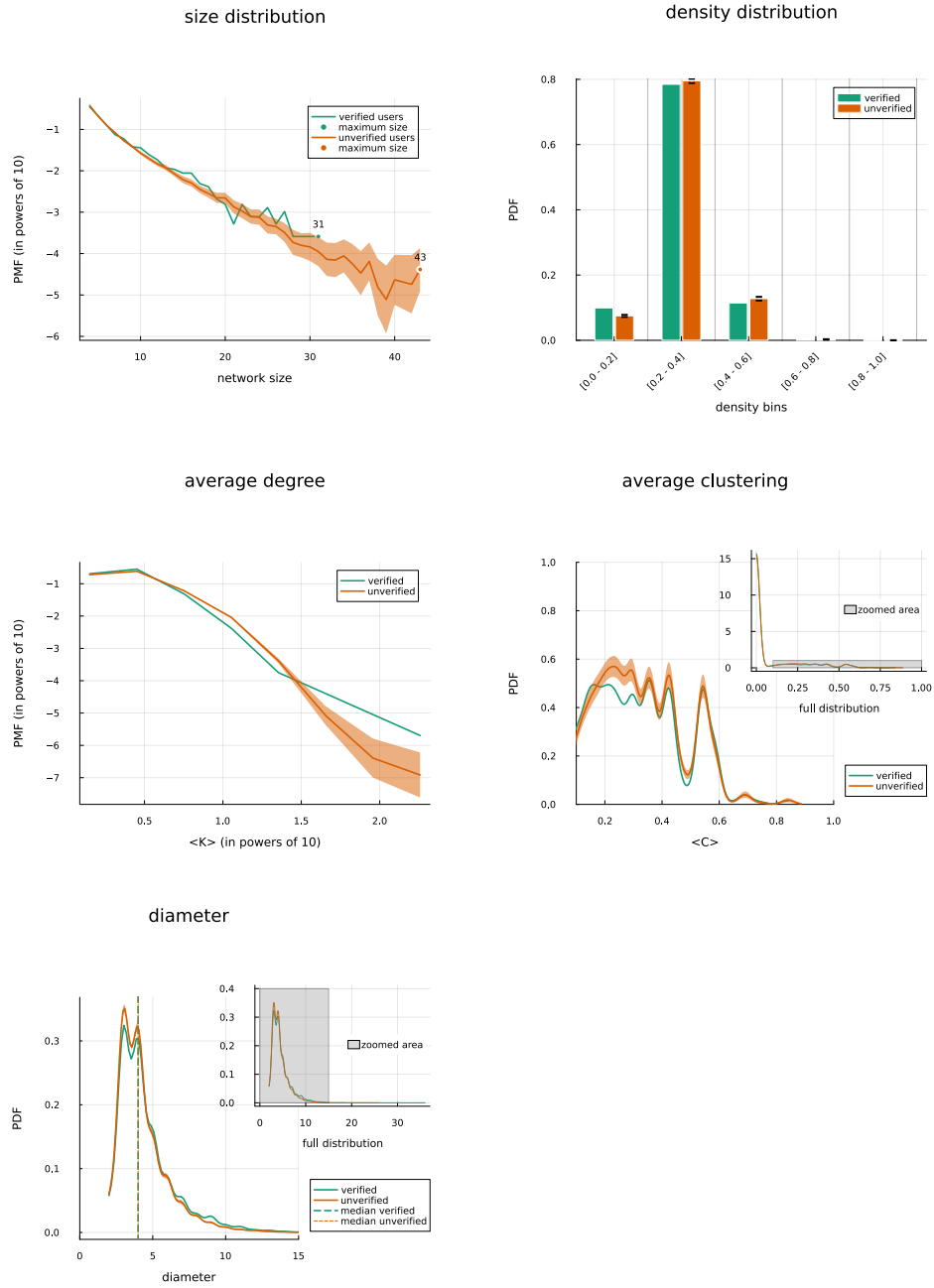


Figure 5.5: Network characteristics of users sharing from at least 4 different domains aggregated by verification status, controlling for group size by sampling.

Users networks distributions: gender

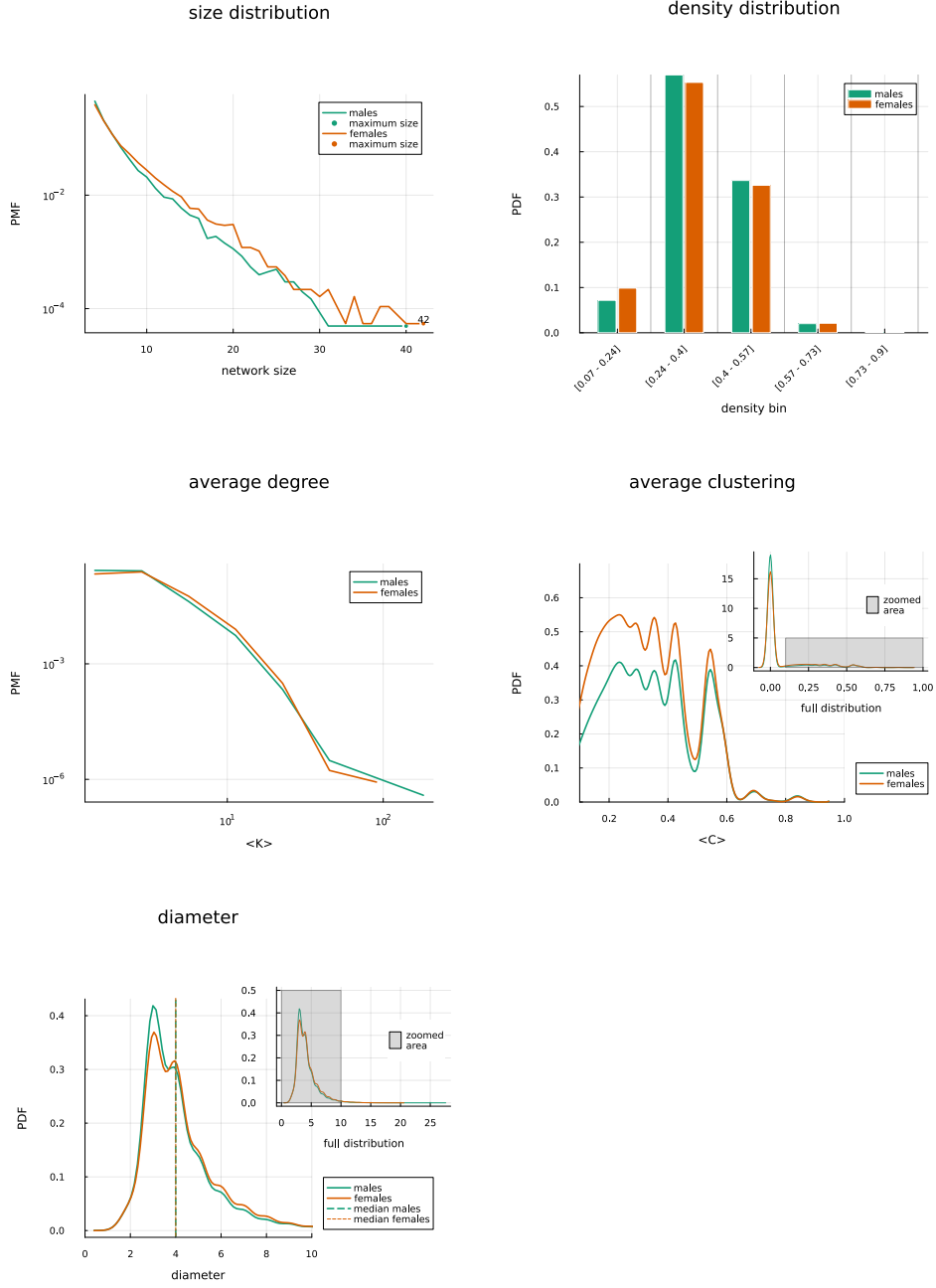


Figure 5.6: Network characteristics of users sharing from at least 4 different domains aggregated by gender.

Users networks distributions: reliability (sampled)

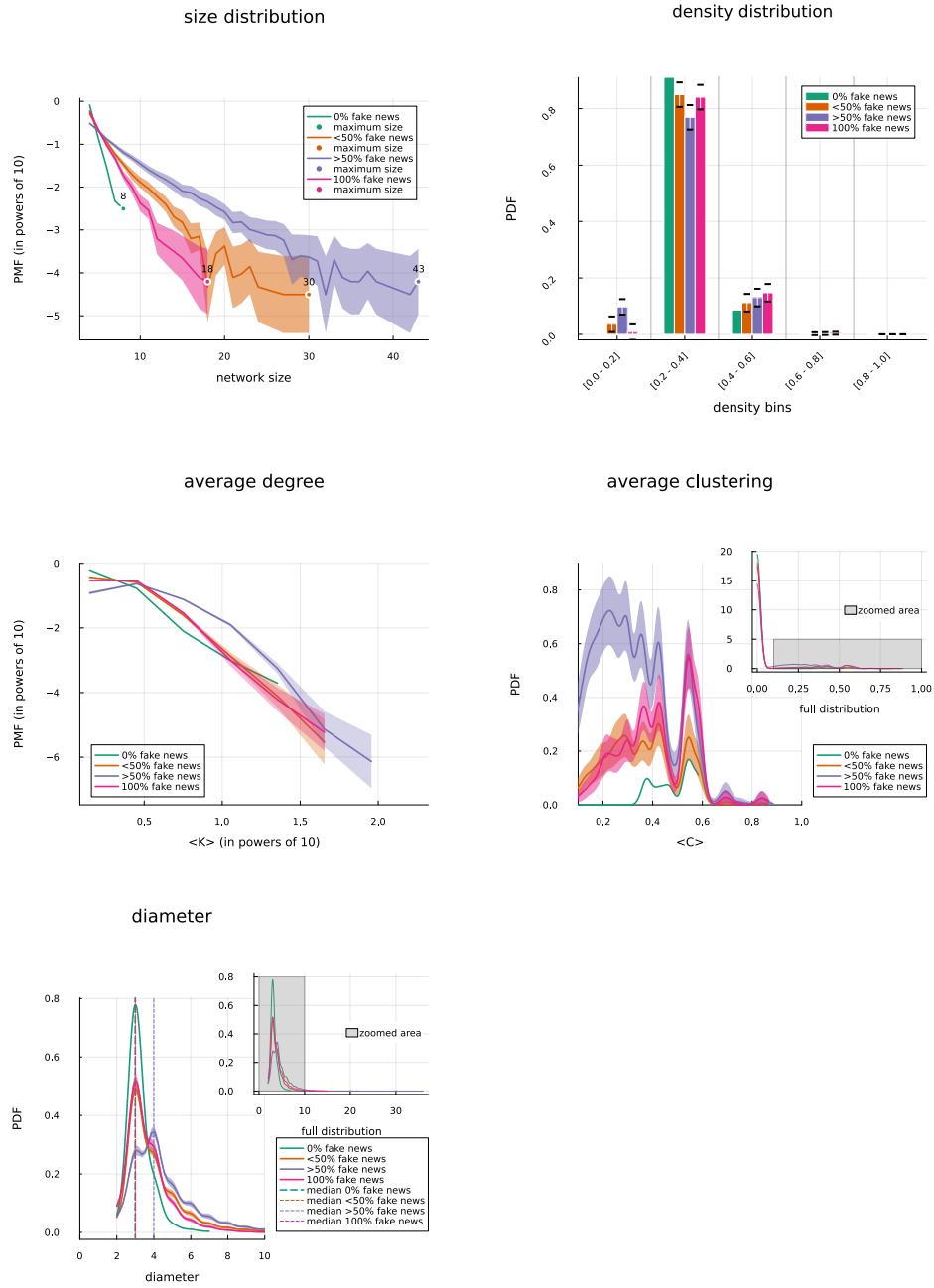


Figure 5.7: Network characteristics of users sharing from at least 4 different domains aggregated by reliability (percentage of fake news shared), controlling for group size by sampling.

Users networks distributions: source type (sampled)

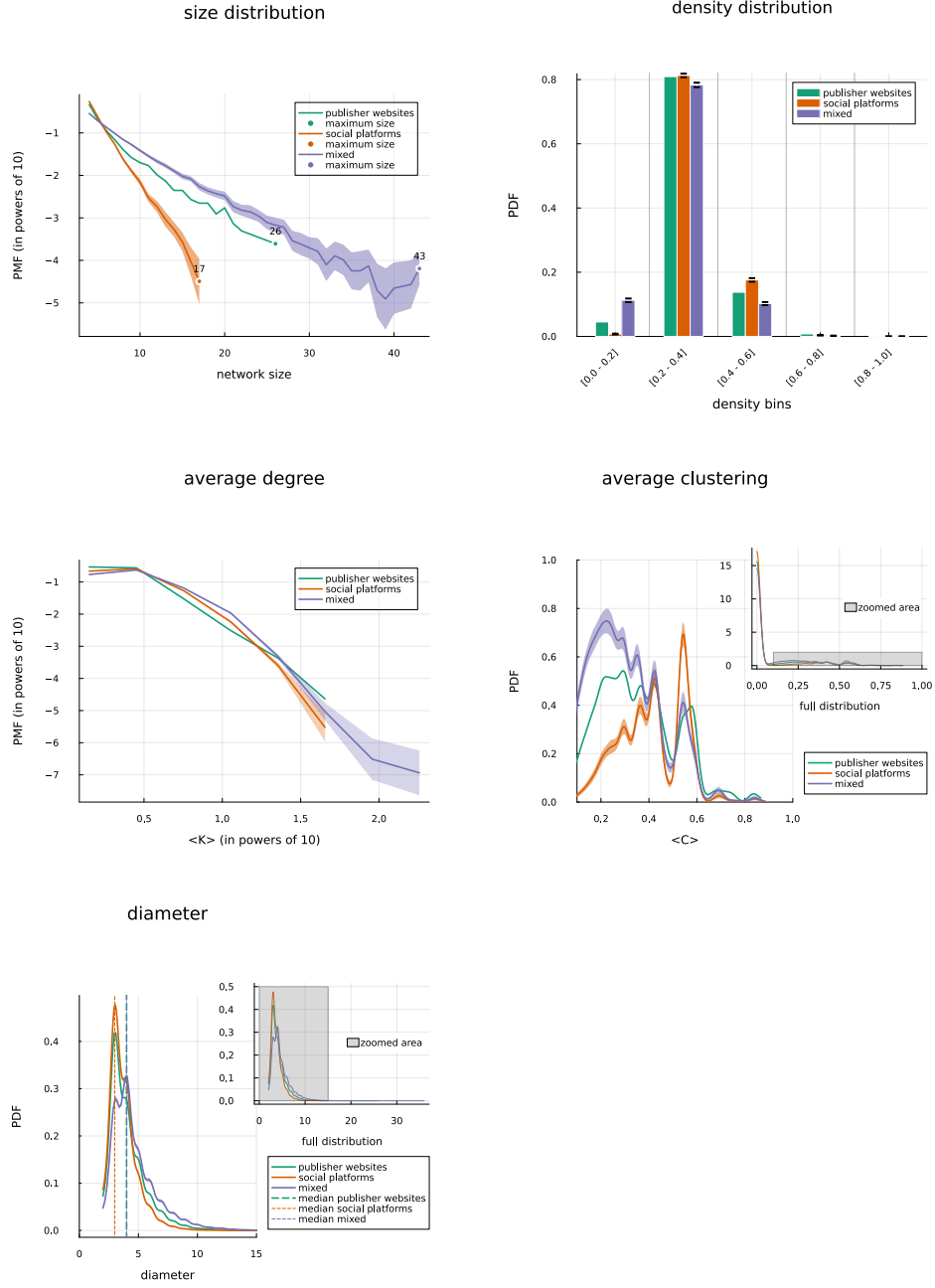


Figure 5.8: Network characteristics of users sharing from at least 4 different domains aggregated by the type of source shared, controlling for group size by sampling.

Users networks distributions: countries (sampled)

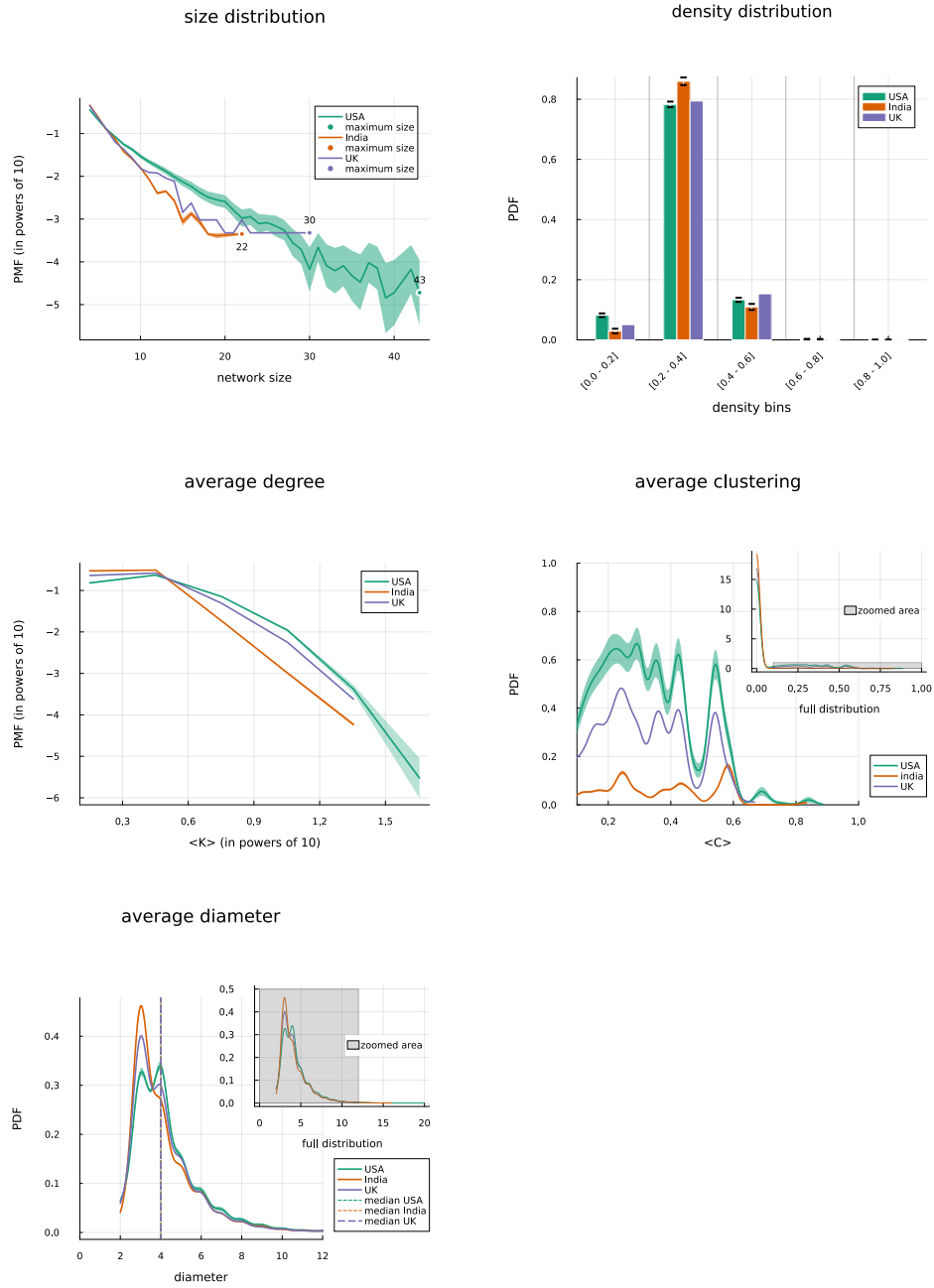


Figure 5.9: Network characteristics of users sharing from at least 4 different domains aggregated by the three top countries by number of users, controlling for group size by sampling.

5.3 Correlation between sources and topics

Before moving on to the analysis of source consistency, since the dataset [13] allowed to classify each tweet also by topic, we investigated whether a correlation could be found between source and topic choice. The analysis of consecutive topic transitions in user sharing sequences reveals a strong correlation between the source of a tweet and its topic, where switching sources substantially disrupts the tendency to repeat the same topic. This suggests that source choice shapes, at least partially, which topics users are exposed to.

To show this, we constructed three different Markov transition matrices that describe the probability of sharing a specific topic in the next tweet conditioned on the current topic. Results showed in Figure 5.10 are obtained retaining users sharing from at least four different sources, but are completely superimposable to the ones obtained filtering users sharing at least two tweets.

In the first plot (a), all consecutive pairs in the user sequences were considered. The heatmap highlights the predominance of political tweets, followed by the health-related ones (expected by the composition of the dataset, illustrated in Figure 4.1); however, the high saturation on the diagonal suggests that in consecutive tweets, the probability of repeating the same topic is higher than the probability of changing it. To understand whether this trend is related to the repetition of the same source, plot (b) focuses on the pairs of consecutive tweets with unchanged source. In this case, the diagonal is even more pronounced, indicating that when the source is unchanged, the probability of repeating the topic also increases. This already represents a strong hint of correlation between source and topic. To exclude that the tendency of repeating the same topic is a generally valid behavior independent of the source, plot (c) examines the variation of topics when considering consecutive pairs of tweets with different sources. Interestingly, the predominance of the diagonal vanishes; therefore, changing the source substantially reduces the likelihood of topic repetition.



Figure 5.10: a) Markov transition matrix visualized as an heatmap showing topic transition probabilities when considering all consecutive sources in users sequences. b) Markov transition matrix visualized as an heatmap showing topic transition probabilities when considering only pairs of repeated consecutive sources in users sequences. c) Markov transition matrix visualized as an heatmap showing topic transition probabilities when considering only different pairs of consecutive sources in users sequences.

5.4 Restricted source consumption driven by popularity

As proposed by Bertani et al. [10], the entropy of user source sequences was used as a measure of the variety of sources and the regularity of sharing. Results show that users' media consumption is concentrated on a limited pool of sources, narrower than expected by chance, and, for users sharing from a moderate-to-high set of sources, driven primarily by source popularity rather than personal choice. This suggests that source concentration is, at least partially, a passive consequence of the structure of the information environment. In agreement with the findings of [10], users sharing a higher proportion of false URLs tend to exhibit a more regular and constrained source selection with respect to more reliable users.

These results emerge from the following analysis, in which we computed three different measures of entropy over the sequences of tweets shared by each user (as explained in section 4.4) and aggregated the results by user category. As for the previous analysis, four categories (verification, reliability, country and source type) were balanced for size through sampling, but we did not control for the number of posts.

In the dataset, some domains are much more shared than others, affecting the diversity of the users' media diet (Figure 5.11). That is why we proposed the two null models as benchmarks to understand how much user sharing behavior is conditioned by the source distribution in the dataset (4.4.2. The first null model removes the bias over user choices (showed as a dashed black line in average and as a black dot for each category in Figures 5.12 - 5.19), the second one removes the bias over source popularity (showed as a dashed red line in average and as a red dot for each category).

In Figures 5.12 - 5.19, we focus on users sharing from at least four different sources; results with the second filter retrieving users sharing at least two tweets can be consulted in Appendix B.

The findings indicate that the actual entropy of users' sequences is consistently lower than the one we would have if all sources were equally popular (all users groups average entropies are below the red dashed line in Figures 5.12 - 5.19). This indicates that users media consumption is concentrated on a smaller subset of sources than the one theoretically available; they rely on a limited set of preferred domains, reducing the overall diversity of their media diet. However, when randomizing over users choice while preserving the dataset source distribution, the results over users

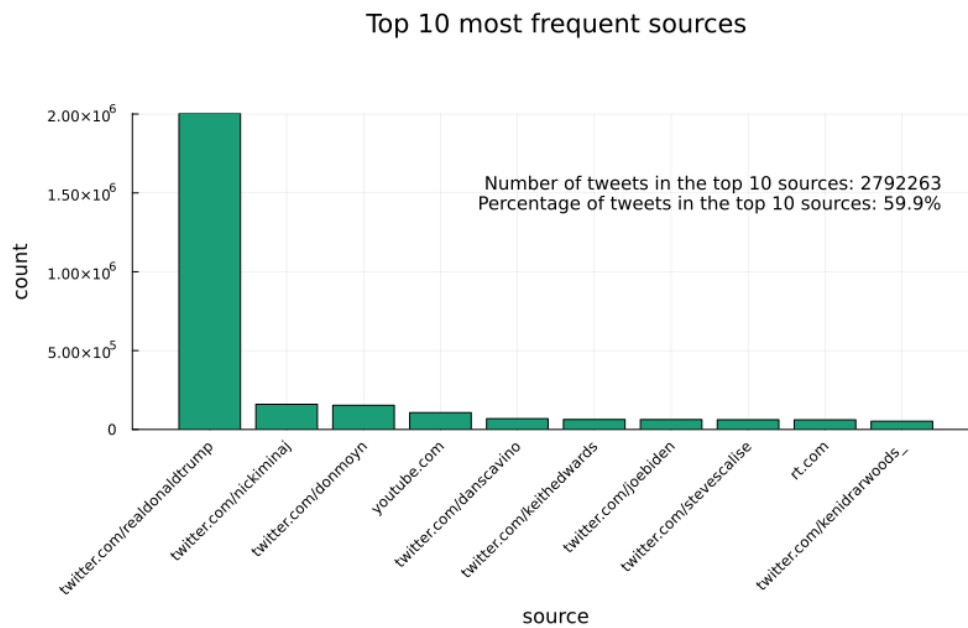


Figure 5.11: Bar chart showing the top 10 most frequent sources in the dataset. Each bar represents a domain, with the height indicating its total count. As highlighted in the annotation, 10 sources represent 59.9% of the dataset, with the single most shared source representing 43% of the dataset alone. This clearly shows a widely unbalanced distribution of sources.

with a relatively high source diversity (the ones showed in Figures 5.12 - 5.19) indicate that the effective entropy is often similar to the one obtained with the null model (black dots in Figures 5.12 - 5.19). This suggests that users' source selection is largely passively shaped by the existing popularity landscape. In other words, users do not appear to actively seek out a diverse range of sources when sharing misinformation, at least when we focus on users with a moderate-to-high media diversity.

There are, however, some exceptions: the actual entropy of users sharing a low percentage of false URLs is higher than the one we would have removing personal choice (Figure 5.17); the same happens with users sharing only from publisher websites (Figure 5.19) and Indian users (Figure 5.18).

Since random entropy encodes the variety of the media diet and Shannon entropy characterizes its heterogeneity by considering the frequency of sharing of each source, it is interesting to focus on the third measure, which shows the normalization of Shannon entropy through random entropy and encodes regularity in sharing [10]. As illustrated in the third plot in Figures 5.12 - 5.19, regularity is similar when comparing popularity (Figure 5.12), influenceability (Figure 5.14), verification status (Figure 5.15), and gender (Figure 5.16); however, less active users are less regular in sharing than the more active ones (Figure 5.13), as well as users sharing only from publisher websites (Figure 5.19) in comparison to users sharing also from social platforms. Notably, users sharing a large percentage of false URLs appear to be more regular in their source choice than users sharing less false URLs (Figure 5.17). This finding is in line with the results of Bertani et al. [10], that showed that users disseminating unreliable content exhibit a more varied but more regular choice of web-domains. It is however important to notice that, in our case, the same conclusion can't be reached for users sharing exclusively fake news among the fact-checked tweets they shared. When focusing on location, in comparison with USA and UK users, Indian ones have lower regularity in sharing (Figure 5.18); this contrast may be linked to the quite different pool of sources available to the subgroups, which are subject to different popularity, as highlighted by the betweenness centrality analysis in section 5.2.

Interestingly, users aggregated by gender do not show any noticeable differences in entropy, just as they did not exhibit substantial differences in the network analysis.

In this analysis, users were filtered for those sharing from at least four different sources. As mentioned above, in Appendix B, one can find the entropy results obtained when selecting users posting at least twice. The most noticeable difference lies in the comparison with the randomization over user choice; when users with

very low source variety are included, their average entropy falls below both null models, thus below what we would expect if sources were chosen purely based on their popularity. This suggests that users relying on a very limited set of sources exhibit stronger individual-level source loyalty, returning to a specific set of sources beyond what popularity-driven selection would predict. An example of the detailed distribution of entropy used to compute the averages analyzed above can also be found in the same appendix.

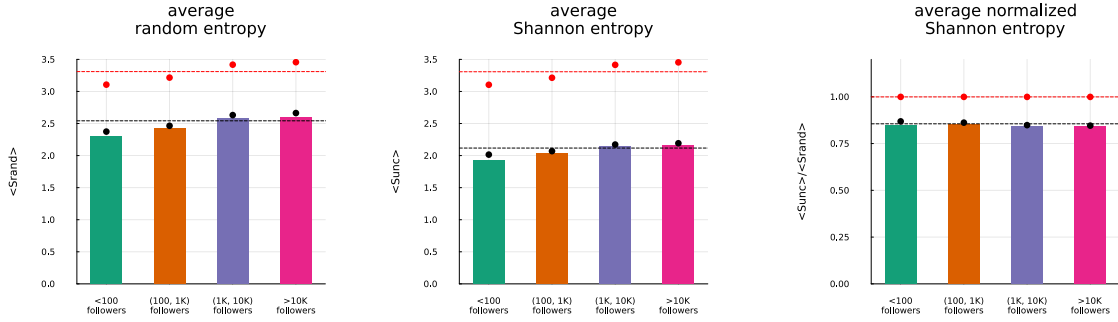


Figure 5.12: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by popularity (number of followers). The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources were considered.

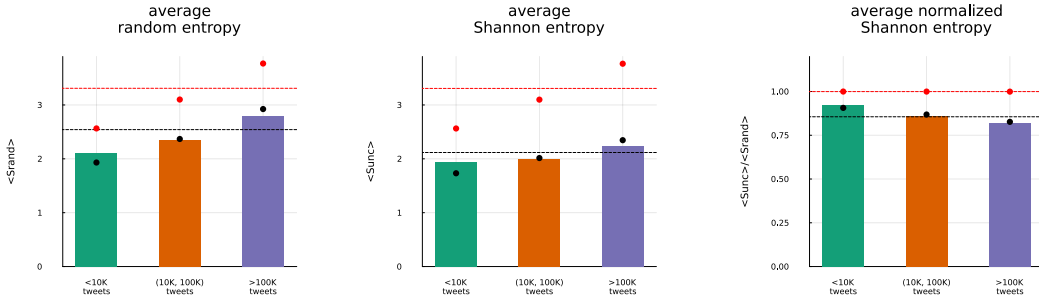


Figure 5.13: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by activity (number of tweets posted). The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources were considered.

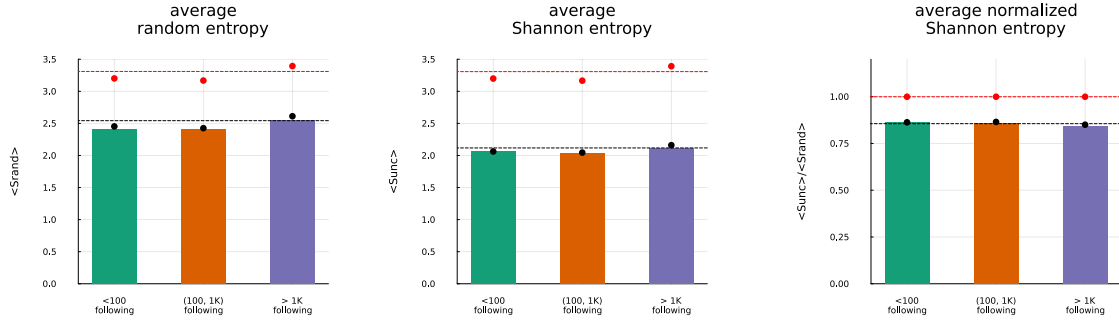


Figure 5.14: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by influenceability (number of following). The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources where considered.

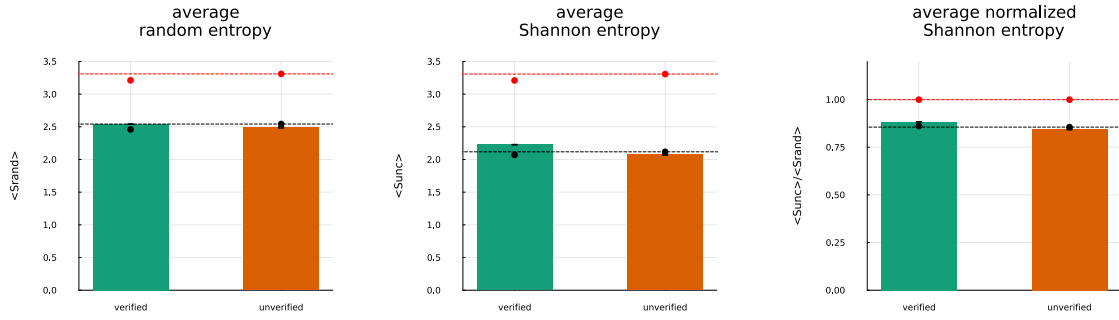


Figure 5.15: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by verification status, controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources where considered.

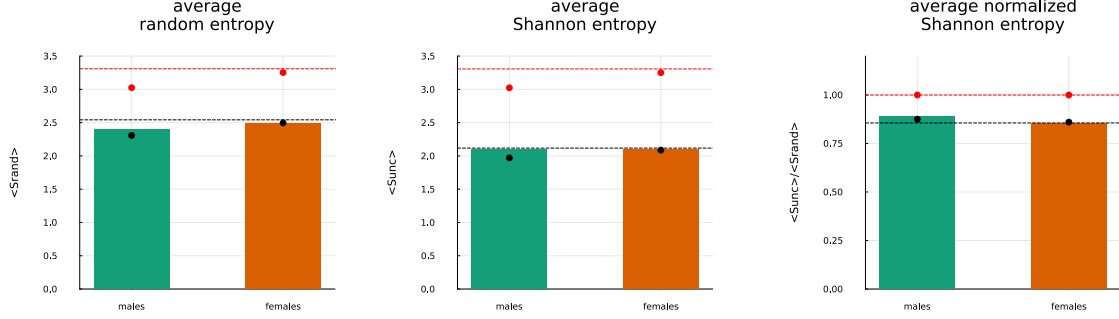


Figure 5.16: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by gender. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources were considered.

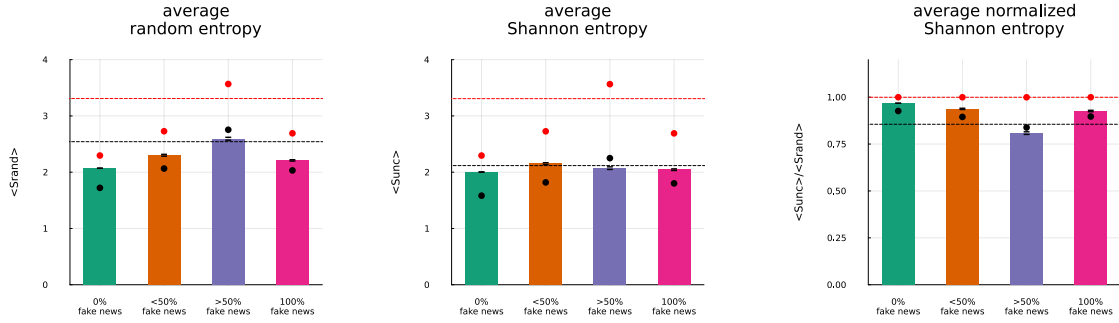


Figure 5.17: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by reliability (percentage of fake news shared), controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources were considered.

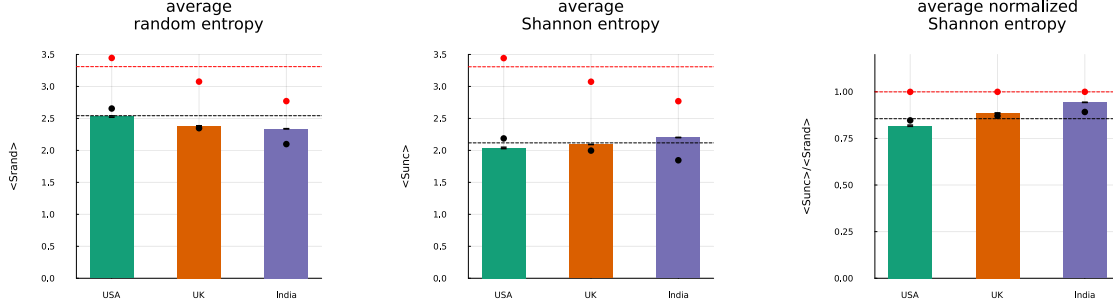


Figure 5.18: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by the three top countries by users number, controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources were considered.

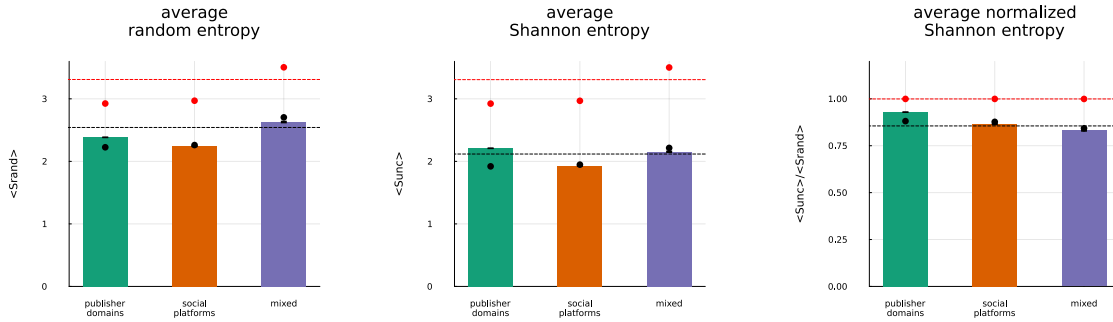


Figure 5.19: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by the type of source shared, controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing from at least 4 different sources were considered.

5.5 Segregation around source types

Once having gained a better picture of user consistency in sharing through entropy, we aimed to understand how users behave towards publisher websites sources and social platform sources. As mentioned in section 4.5, in the Global Claims dataset publisher websites are more than social platforms, but the latter are shared much more frequently in the Twitter dataset. We measured the *favorability score* and the *return score* and found a consistent tendency among users to concentrate their sharing behavior around a single source type, where social platforms emerge as the dominant vehicle for disinformation sharing, especially when considering re-sharing behavior. This pattern holds across both user filters, though users with higher source diversity show a less polarized tendency. Interestingly, users following a low number of accounts tend to share more from publisher websites than more susceptible users, and those not sharing fake news return more often to publisher websites than users who do share fake news. Moreover, consistent with previous results, we do not see any noticeable differences due to gender.

5.5.1 *Favorability score*

We computed the favorability score to investigate the segregation of users between source types in the sharing of misinformation [5], as explained in section 4.5. Values towards -1 indicate a tendency to share more from publisher website sources, while values towards +1 indicate a tendency to share more from social platforms. In the following, we compare the results obtained by filtering users sharing at least two tweets and users sharing from at least four different sources. In both cases, we controlled for group size in the three most unbalanced categories (verification status, reliability, and country) by a sampling procedure. With both user filters, three main peaks in the distributions are recognizable: in -1, identifying users sharing only from publisher domains, in 0, encoding users sharing equally from both source types, and in +1, identifying users sharing only from social platforms. Applying the first filter, shown in Figure 5.20, the segregation appears stronger, because we consider users with highly concentrated behavior; the distribution between the peaks is almost flat, except for the highly active users (activity panel, top left of figure 5.20), who are more broadly distributed in the positive part of the plot, and users sharing a large number of fake URLs (reliability panel, bottom left), who do not show any peak in 0. Interestingly, users who do not share any false URL (reliability panel, bottom left) and users with few followings (influenceability panel, top right) have a higher peak in -1 than the other users in their same category,

which means that they tend to share more from publisher websites than the others. The results obtained filtering users sharing at least 4 different sources, thus focusing on users engaging with a wider set of sources, are illustrated in Figure 5.21 and show in general lower peaks and broader distributions towards the positive part of the plots, suggesting a reduced segregation but oriented towards social platforms, highlighting the central role of social platforms in the spread of misinformation.

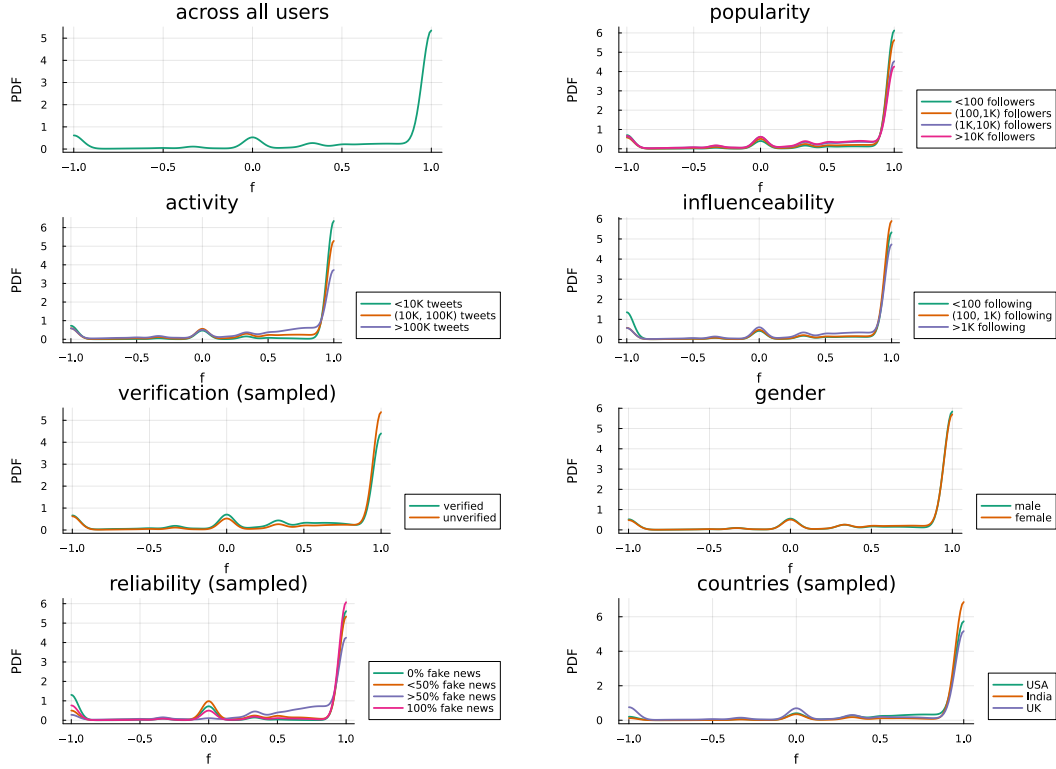


Figure 5.20: Favorability score computed across users categories filtering users sharing at least 2 tweets.

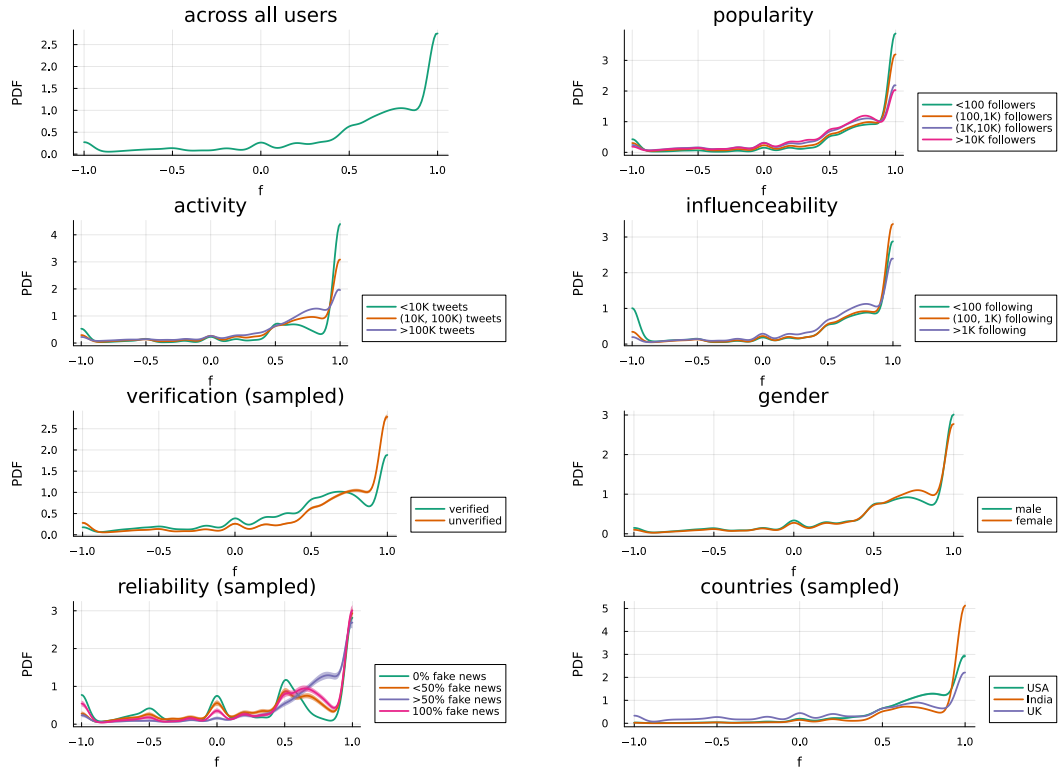


Figure 5.21: Favorability score computed across users categories filtering users sharing at least from 4 different sources.

5.5.2 Return score

Since we noticed that the distribution of tweets is highly unbalanced (Figure 5.25), we adapted the favorability score to count the repeated sources in users sharing sequences. Therefore, the *return score* captures the loyalty to source type in the re-sharing behavior (Section 4.5). As before, we compare the results obtained applying both filters on users and controlling for group size in the three most unbalanced categories (verification status, reliability, and country) by a sampling procedure. When filtering on users sharing at least 2 tweets (figure 5.22), results show broader distributions with respect to the favorability score ones, but still maintaining visible peaks in the extremes and in the positive part of the plots. This suggests that re-sharing behavior tends to favor social platforms, but still there are users who are highly loyal to previously encountered publisher websites and re-share only those. This is especially true for users with few followers (influenceability panel, top right) and users not sharing any false URL (reliability panel, bottom left). The situation is very different when considering users sharing from at least 4 different sources (Figure 5.23); here, the peaks at the extremes of the plots disappear and the distributions broaden especially in the positive part of the plots, indicating a wide range of intermediate re-sharing behaviors. Therefore, an increased source diversity may imply a more flexible re-sharing habit, however centered around social platforms.

The above analysis suggests that there is a tendency to repeat the same type of source along the sharing sequences. We can also observe this trend from a more coarse-grained point of view by looking at the transition matrices constructed by considering consecutive sources in each user sequence. Figure 5.24 illustrates the Markov transition table that describes the probability of the source type of the next tweet conditioned on the source type of the current tweet, differentiated by the applied user filter. The results are similar in the two cases, but more nuanced when selecting users sharing from at least 4 different sources. In both cases, it is more probable to share again from the same source type, even if the disparity in frequency between tweets with a social platform source and tweets with a publisher website source implies that the probability of going from a publisher website source to a social platform one is much higher than the contrary (around 40% in the first case and less than 10% in the second). This result emphasizes once more the consistent behavior of users towards source choice, also taking into account the biased dataset on which analysis is performed.

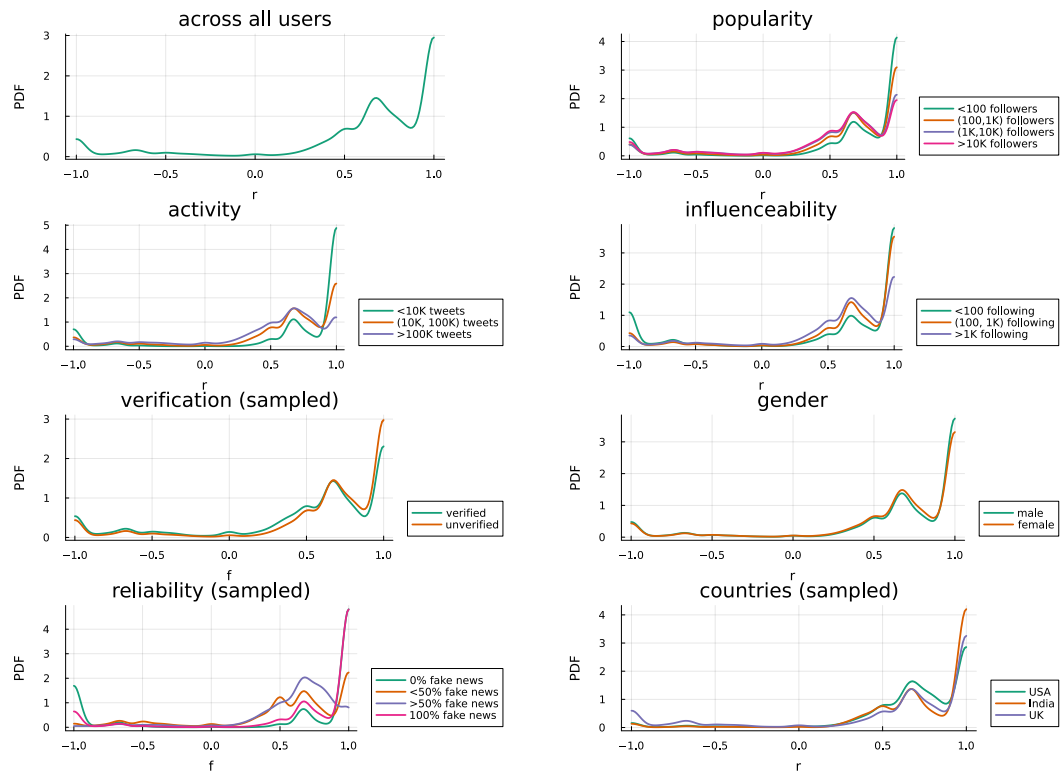


Figure 5.22: Return score computed across users categories filtering users sharing at least 2 tweets.

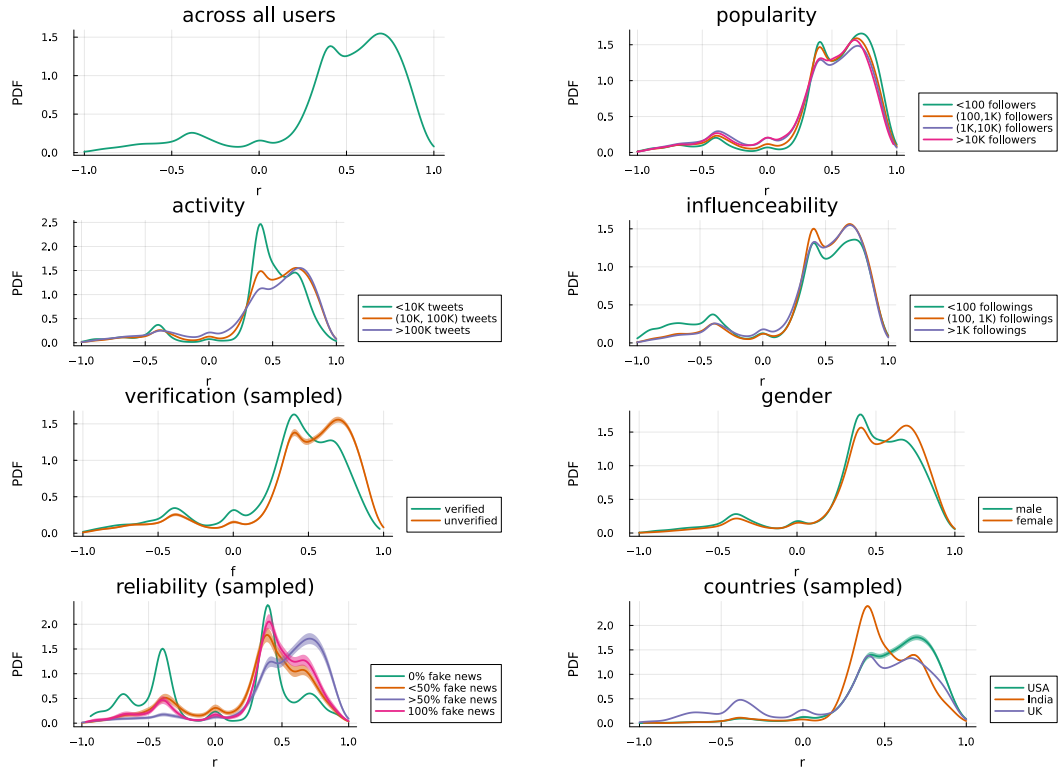


Figure 5.23: Return score computed across users categories filtering users sharing at least from 4 different sources.

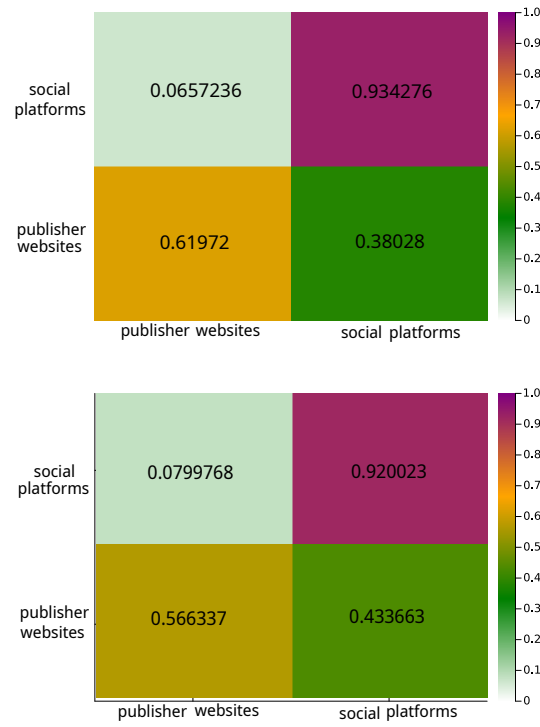


Figure 5.24: a) Markov transition matrix showing the probability of the source type of the next tweet knowing the source type of the present tweet. Here, users sharing at least 2 tweets are considered. b) Markov transition matrix showing the probability of the source type of the next tweet knowing the source type of the present tweet. Here, users sharing from at least 4 different domains are considered.

5.6 Increase in user entropy after COVID-19

In the Twitter dataset, the years with the most tweets collected are 2019, 2020, and 2021. Interestingly, these years span across a major world-wide event: the COVID-19 pandemic outbreak at the beginning of 2020. Focusing on the topics shared during those three years, a shift in topic preference was noticeable. Although politics remained the most shared topic, health related tweets gained a lot of attention in a short time period, allowing for a distribution of topics less peaked around politics. Figure 5.25 illustrates the shift in topic proportion from 2019 to 2021.

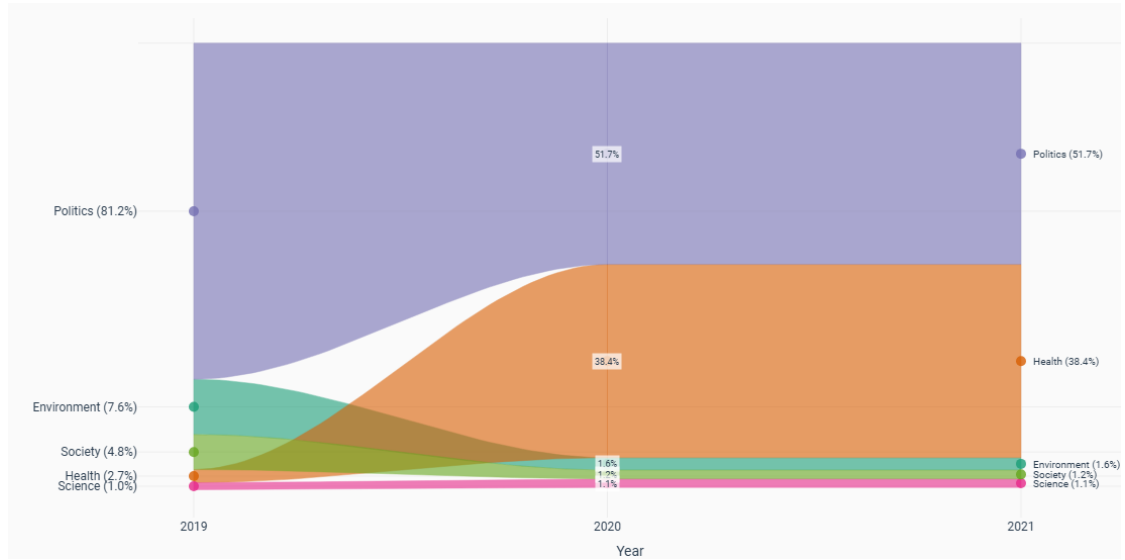


Figure 5.25: Percentage of tweets for each main topic in year 2019, 2020, and 2021. Topics are ordered from the most frequent to the less one, highlighting the increase of attention given to health from 2020. In 2019, the order was politics (81.2%), environment (7.6%), society (4.8%), health (2.7%) and science (1.0%), while in 2021 it became politics (51.7%), health (38.4%), environment (1.6%), society (1.2%) and science (1.1%).

Since we previously found a strong correlation between sources and topics, we investigated whether the less biased towards politics topic distribution after 2020 could be linked to a higher variety in the users source choice. We found that users entropy increased after 2020, and this result was even more remarkable when considering that the null model randomizing over source popularity showed instead a decrease after the outbreak of the pandemic.

To establish this, we compared entropy before and after 2020, matching users by propensity score (the detailed process is explained in section 4.6), selecting as confounding parameters the number of followers, of tweets and of following and the length of the shared sequence. We compared results against a randomization over source popularity, similar to the null model previously exploited in the entropy computation (section 4.4.2), but again matching users by propensity score. Figure 5.26 shows that the entropy increased after the COVID-19 outbreak, while the entropy obtained by removing the bias over the popularity of sources (and therefore considering all sources equally probable) instead decreased after 2020. The regularity in sharing, encoded in the normalized Shannon entropy, remained constant before and after 2020 (third plot), suggesting that users maintained similar behavioral patterns in terms of posting dynamics. Thus, the change in the topical landscape was accompanied by a broader diversification in the sources shared by users, while the structural regularity of user sharing sequences was not substantially altered by the COVID-19 outbreak.

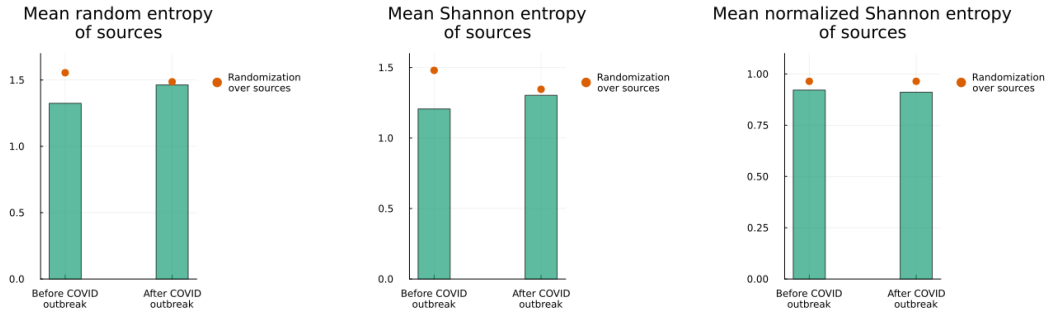


Figure 5.26: Random entropy, Shannon entropy and normalized Shannon entropy compared before and after the COVID-19 outbreak, against a null model randomizing over source popularity.

Chapter 6

Conclusions

This thesis investigated the relationship between Twitter (now X) users and the sources they utilize when sharing fact-checked content, aiming to characterize behavioral patterns across a range of user categories. Using a dataset of over 4.6 million tweets, each linked to a fact-checked URL [13], each user’s sharing behavior was modeled as a temporal sequence of web domains, serving as a proxy for information sources. Sequences were represented as weighted undirected networks and analyzed through different network metrics, transition matrices, entropy measures, and favorability and return scores.

First, network analysis revealed that most users share from a relatively small set of sources, with tree-like network structures suggesting habitual return to a limited number of preferred sources rather than an exploratory attitude across the available source pool. Second, source choice and topical exposure were found to be strongly coupled: sharing from the same source largely increases the probability of sharing the same topic, while switching sources disrupts this continuity. Third, entropy analysis confirmed a consistently concentrated source diet across all user categories, where the source selection of users with a moderate-to-high source variety is largely shaped by source popularity rather than personal preference. Additionally, we found that users with low source variety are more segregated between source types than users with higher diversity, who revealed a consistent orientation towards social platform sources. Finally, the COVID-19 pandemic outbreak provided a natural framework for testing previously found entropy patterns; entropy was found to increase after the onset of the pandemic, reflecting the broadening of topic distribution caused by the sudden increased attention to health-related content. Notably, however, the normalized entropy, which captures regularity in sharing,

remained stable before and after 2020, suggesting that users expanded the range of sources they shared without altering their habitual posting dynamics.

The concentration of sources across all user categories suggests that restricted source consumption is a general feature of how users interact with (mis)information on social media rather than a characteristic exclusive to some particular group. This is consistent with documented tendencies towards homogeneity in online information exposure [25, 24], and mirrors the filter bubble dynamics observed on major platforms [5, 21, 22]. This tendency is reinforced by the fact that retweets account for more than 80% of the *Twitter dataset*; their dominance suggests that the sharing behavior analyzed here is predominantly passive, meaning that users amplify existing content rather than actively selecting it. Thus, the nature of source loyalty is called into question: the concentration of sources may be a consequence of low effort participation in platform-driven content circulation more than of deliberate information choices.

The source-topic coupling identified through transition matrices suggests a re-framing of the echo-chamber concept [56]: beyond ideological and user communities, echo-chambers can be thought in terms of sources, where repeated exposure to the same source also constrains the topical diversity of one’s information diet. However, in this context, this source-level selective exposure is complicated by the fact that the majority of the shared content in the dataset originates from Twitter posts themselves, blurring the line between information sources and individual users (whose profiles are considered sources). In this environment, a concentrated source diet also involves repeatedly amplifying the same voices, regardless of their reliability.

In line with Bertani et al. [10], users sharing a higher proportion of unreliable content exhibit more varied yet more regular source selection than users sharing more reliable URLs. However, our results suggest a more nuanced picture: users sharing exclusively false content among the fact-checked tweets do not always show the same pattern, as well as users never sharing fake news. The discrepancy may reflect differences in dataset composition, in our case focused only on fact-checked content, or the inclusion of retweets. Nevertheless, the regularity of source selection among higher disinformation sharers could indicate systematic reliance on a limited set of sources associated to disinformation, and be exploited as a behavioral signature to complement content or network-based approaches to misinformation detection [51] and inform the training of AI models to identify likely misinformation spreaders. These detection approaches would need to account for geographic and cultural context, as highlighted in our analysis by the contrasting sharing regularity of Indian users compared to USA and UK users.

From our comparisons, verified users do not exhibit substantially different sharing patterns from unverified ones, consistently with previous work suggesting that verification status is not a reliable indicator of content quality [31], questioning the effectiveness of platform verification mechanisms as an indirect indication of trustworthiness. Another group not showing any appreciable difference across all the analyses is gender, suggesting that exposure to misinformation is shaped more by behavioral patterns than by demographic characteristics. This is interesting because previous research highlighted that male users were more active in terms of sharing news [37]. Moreover, user sharing exclusively from publisher websites showed lower regularity compared to those sharing exclusively from social platforms, suggesting that reliance on social media as primary information source is a driver of concentrated and regular behavior, pointing to the architecture of social platforms as a key factor in shaping misinformation dynamics. From a platform design and policy perspective, the finding that source diversity is driven more by source popularity than by personal choice, at least for users with a moderate-to-high source variety, suggests that interventions diversifying algorithmic recommendations may be more effective than those directly targeting individual user behavior [83].

Focusing on the temporal analysis around COVID-19 outbreak, the regularity of sharing remaining stable before and after 2020 points to a remarkable stickiness of misinformation-sharing behavior, apparently in contrast with the evidence that during the pandemic users broadly turned to trustworthy outlets [41] and that factual content drove more engagement than misinformation [42]. However, our result focuses primarily on misinformation spreaders, therefore we do not have a proper picture of users behavior linked to trustworthy outlets, and is instead consistent with findings that users' historical inclination toward sharing misinformation persists across topics and over time [38]. This suggests that interventions targeting misinformation spreaders may need to address long-term behavioral patterns.

Several limitations to this work must be accounted for. First, the dataset is restricted to fact-checked URLs, representing a biased sample in which false claims are over-represented, as fact-checkers tend to prioritize implausible or contested statements. Therefore, the observed user behavior reflects sharing patterns limited to misinformation contexts and may not generalize to broader news consumption. Moreover, analyzed source sequences of each user represent only the subset of their tweets that contains fact-checked links, without capturing the full range of their information diet. The used dataset implies also an interpretation issue on entropy results. The finding that observed entropy tends to match the null model obtained removing users choice may be biased by the fact that the selected sources comes from fact-checked URLs, which often are widely shared and already have high

visibility. Moreover, the dataset does not allow us to distinguish human users from bots, and it is likely that a non-negligible fraction of the accounts analyzed are automated [33], which could inflate the apparent regularity and concentration of source selection. Second, the Twitter Academic API used to collect the data is now discontinued, which limits replications and extensions to more recent periods. A further methodological concern regards the arbitrary choice of the two user filters applied. The stricter filter, which retains only users sharing from at least four different sources, is motivated by the choice of focusing on users with moderate source diversity and the need for networks with sufficient structure to compute meaningful properties, but it excludes the majority of users, who shares from very few sources and have a highly concentrated behavior. Therefore, both filters are considered throughout the analysis and results can be consulted in the appendix. Finally, the confounding parameters chosen for the propensity score approach may not fully capture all relevant baseline differences between pre- and post- pandemic users, therefore the causal attribution of entropy changes to the pandemic should be interpreted with caution.

To overcome some of the mentioned limitations, the proposed analysis could be repeated by considering broader posting sequences, allowing a more complete picture of users information diet and a comparison between misinformation sharing behavior and everyday news consumption. Moreover, the observed consistency in source selection could be exploited for link prediction [84] in bipartite user-source graphs [51]; given the regularity with which users return to a small set of domains, predicting future sharing behavior may be feasible and potentially useful for anticipating misinformation spread. More broadly, extending this framework to other platforms would help to assess the generalizability of the patterns identified here, and incorporating temporal dynamics beyond the COVID-19 case study would allow for a more systematic investigation of how external events perturb information ecosystems.

Appendix A

Networks Characteristics

Results of networks characteristics obtained by aggregating users by category, retaining users posting at least two tweets. As for the previous analysis, four categories (verification, reliability, country and source type) were balanced for size through sampling, but we did not control for the number of posts.

Users networks distributions: verification (with sampling)

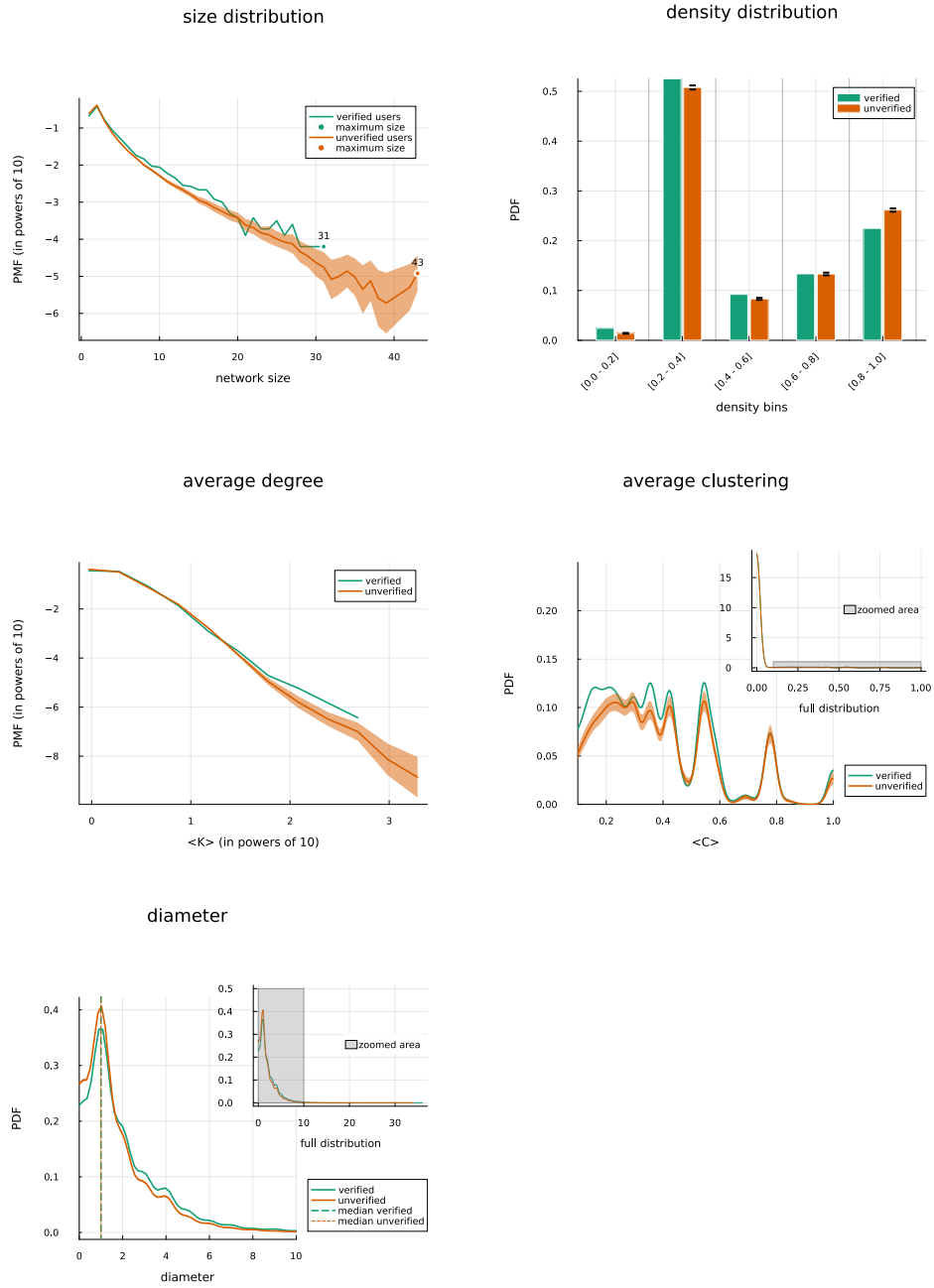


Figure A.1: Network characteristics of users sharing at least 2 tweets aggregated by verification status, controlling for size by sampling.

Users networks distributions: popularity

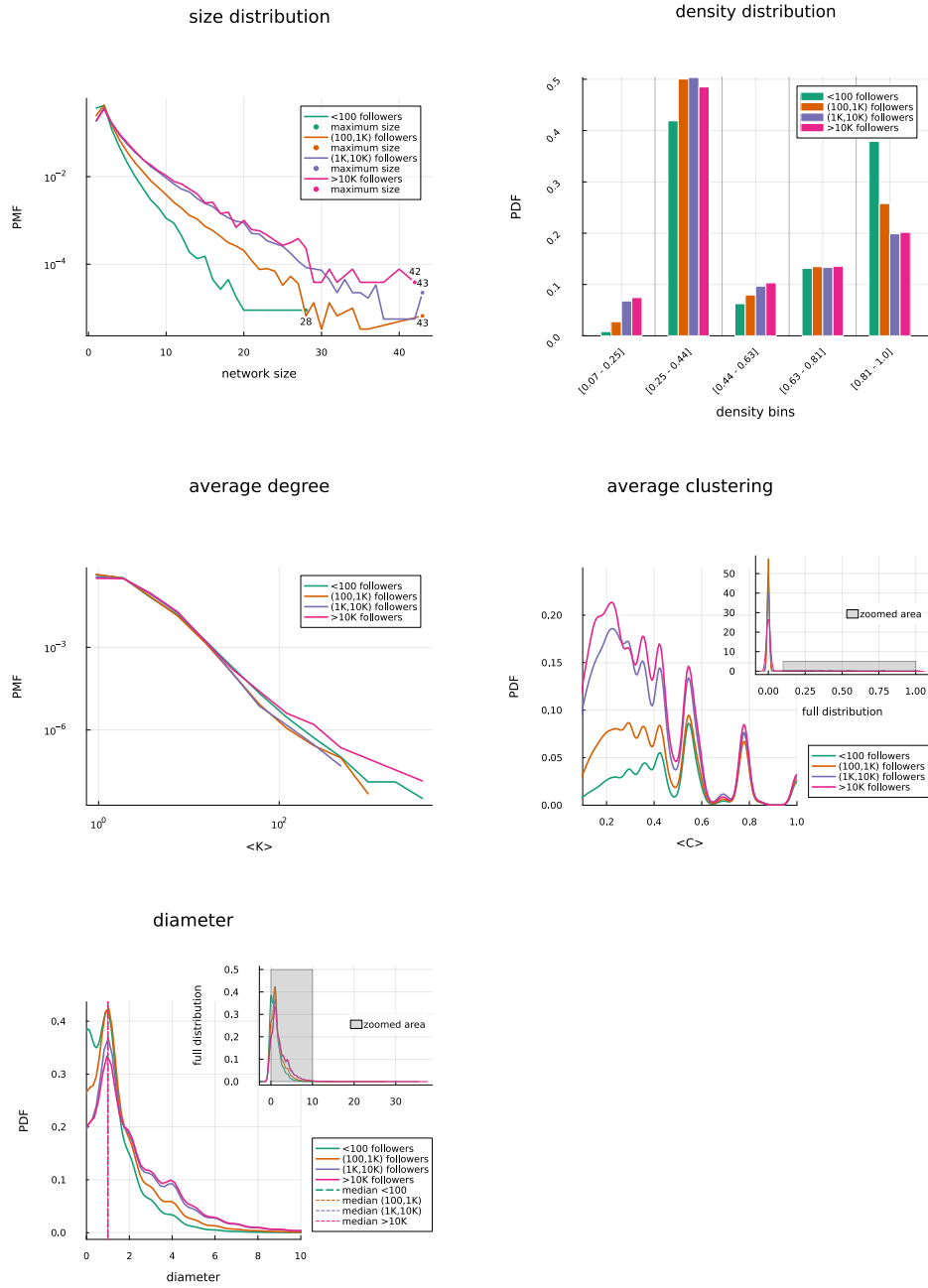


Figure A.2: Network characteristics of users sharing at least 2 tweets aggregated by popularity (number of followers).

Users networks distributions: activity

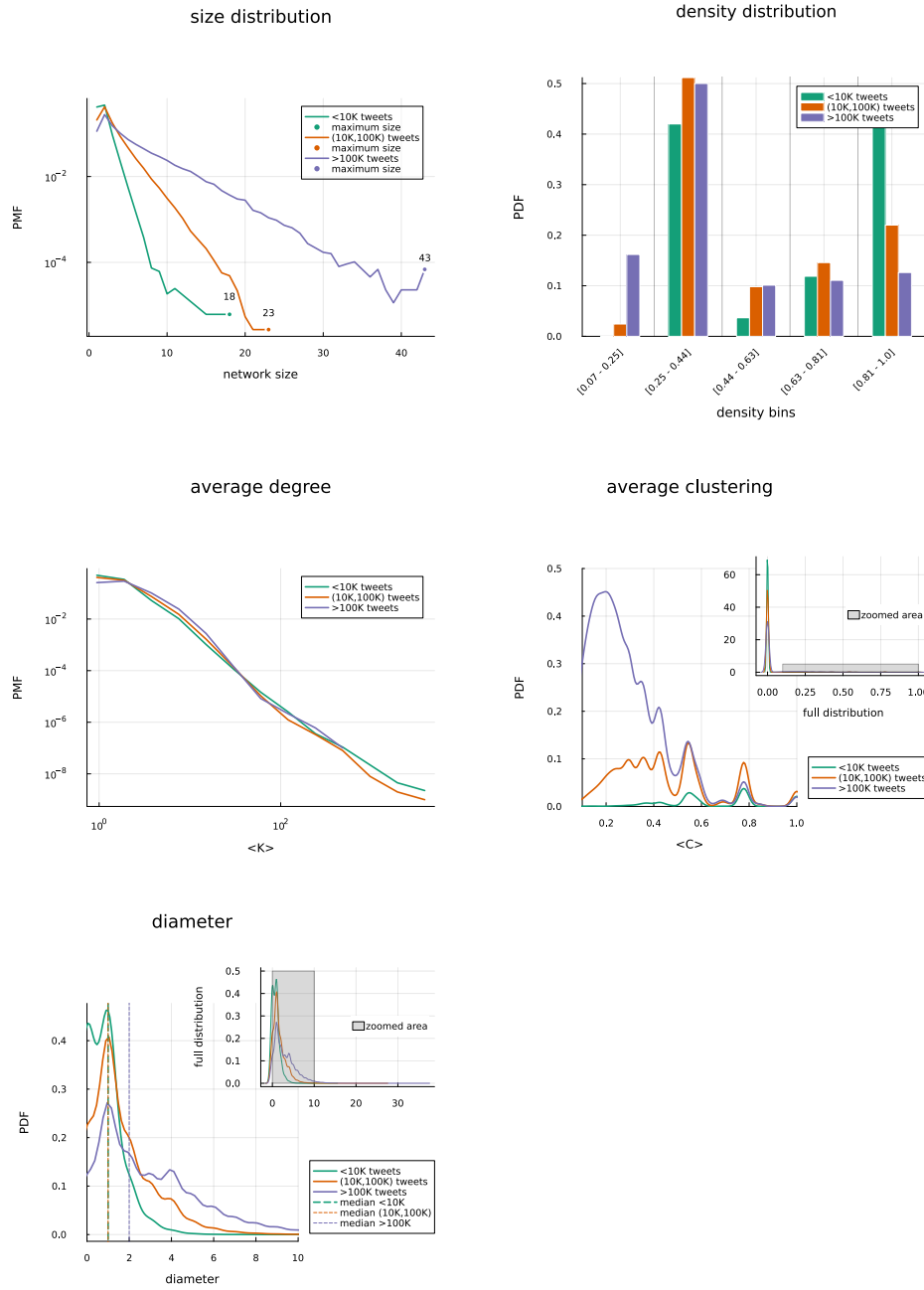


Figure A.3: Network characteristics of users sharing at least 2 tweets aggregated by activity (number of posted tweets).

Users networks distributions: influenceability

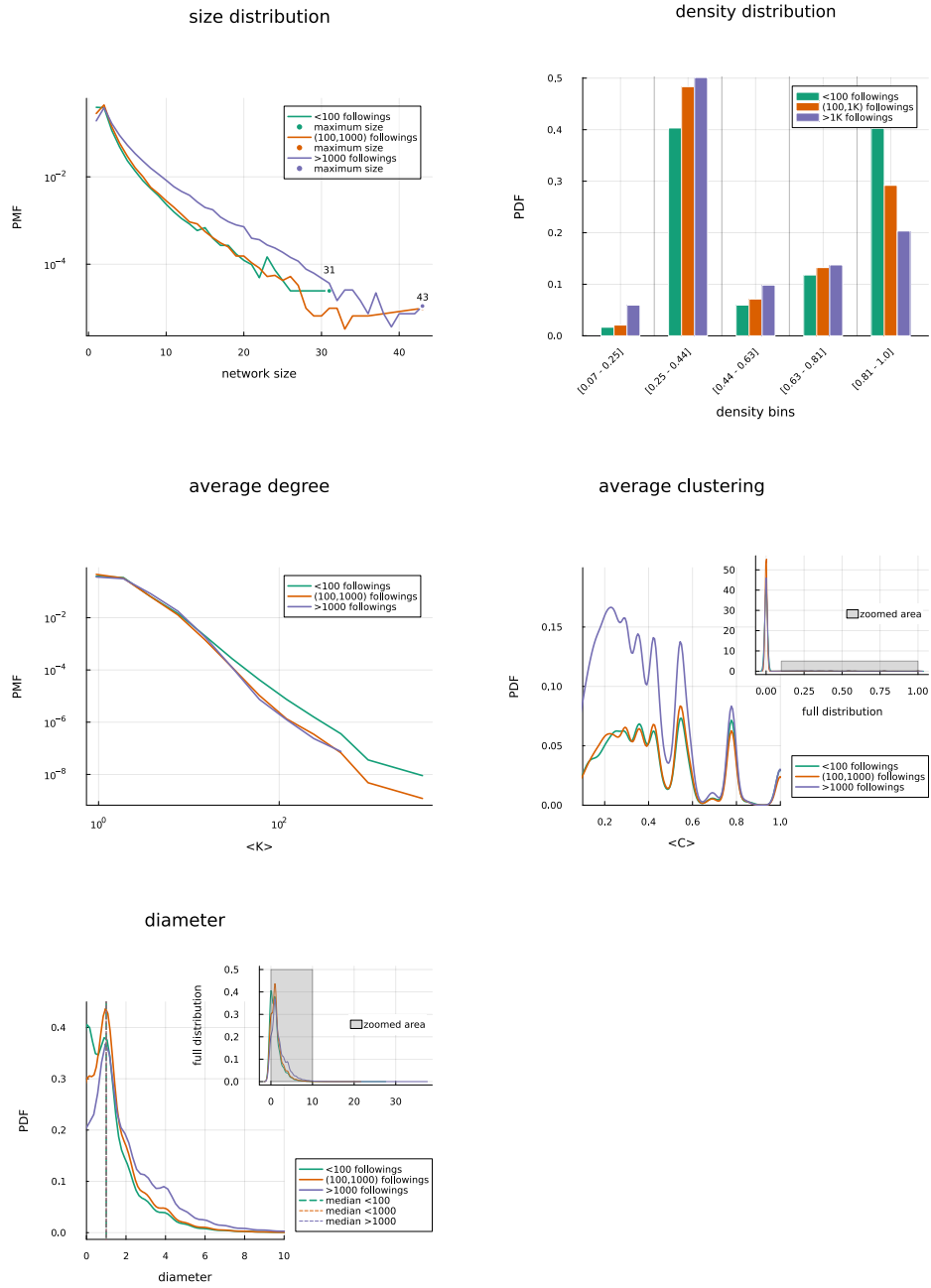


Figure A.4: Network characteristics of users sharing at least 2 tweets aggregated by influenceability (number of following).

Users networks distributions: gender

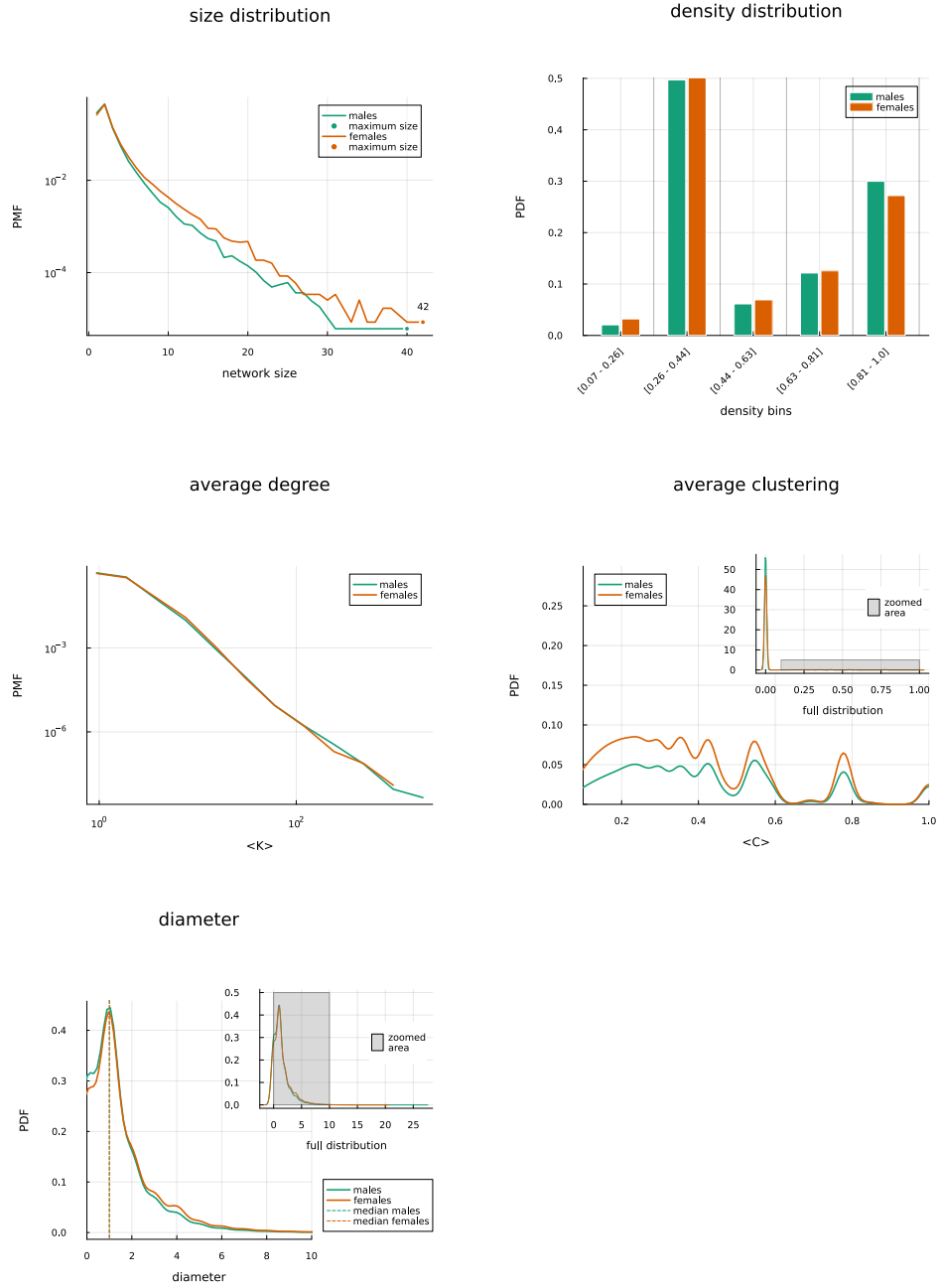


Figure A.5: Network characteristics of users sharing at least 2 tweets aggregated by gender.

Users networks distributions: reliability (with sampling)

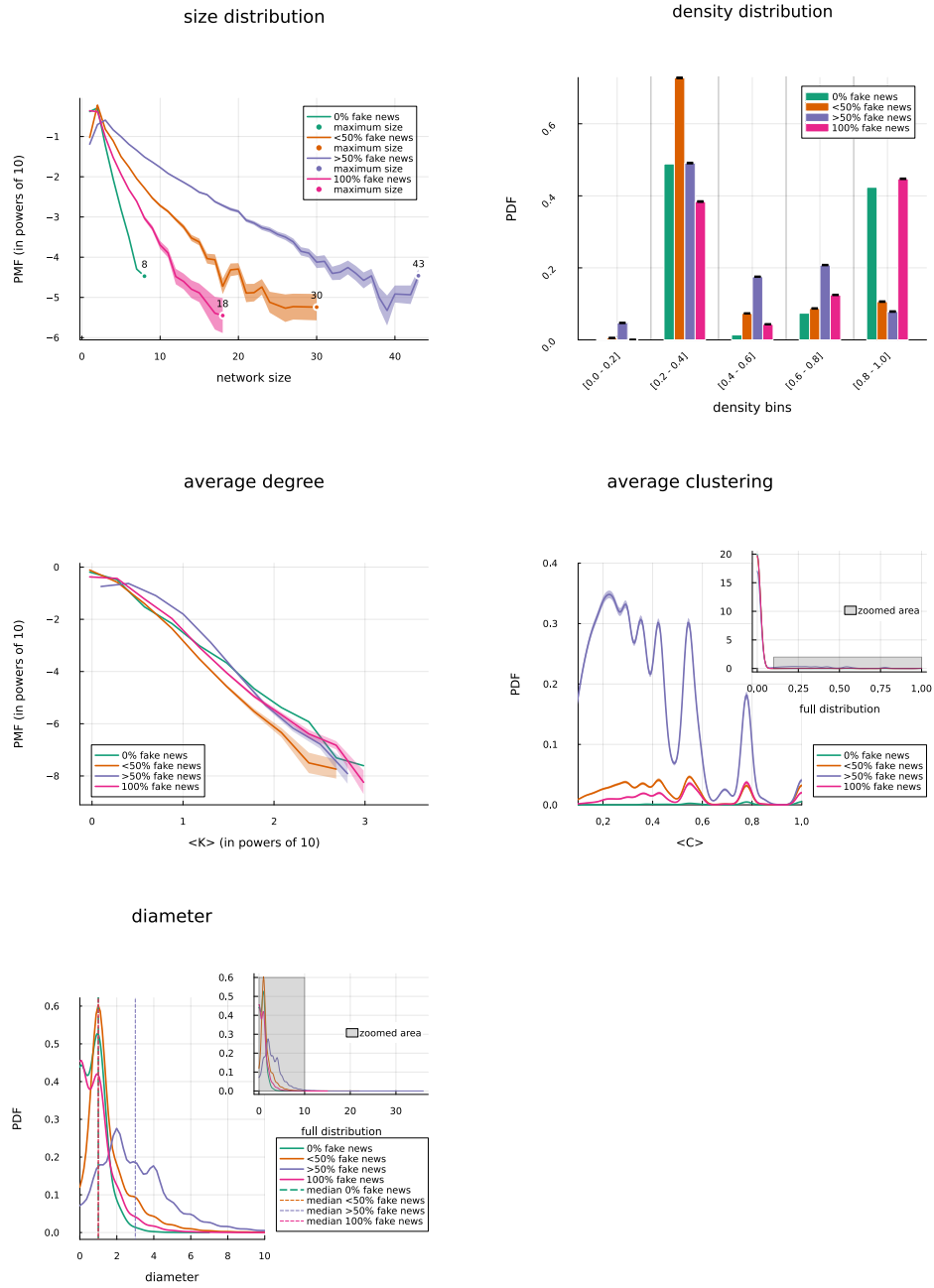


Figure A.6: Network characteristics of users sharing at least 2 tweets aggregated by reliability (percentage of fake news shared), controlling for group size by sampling.

Users networks distributions: source type (with sampling)

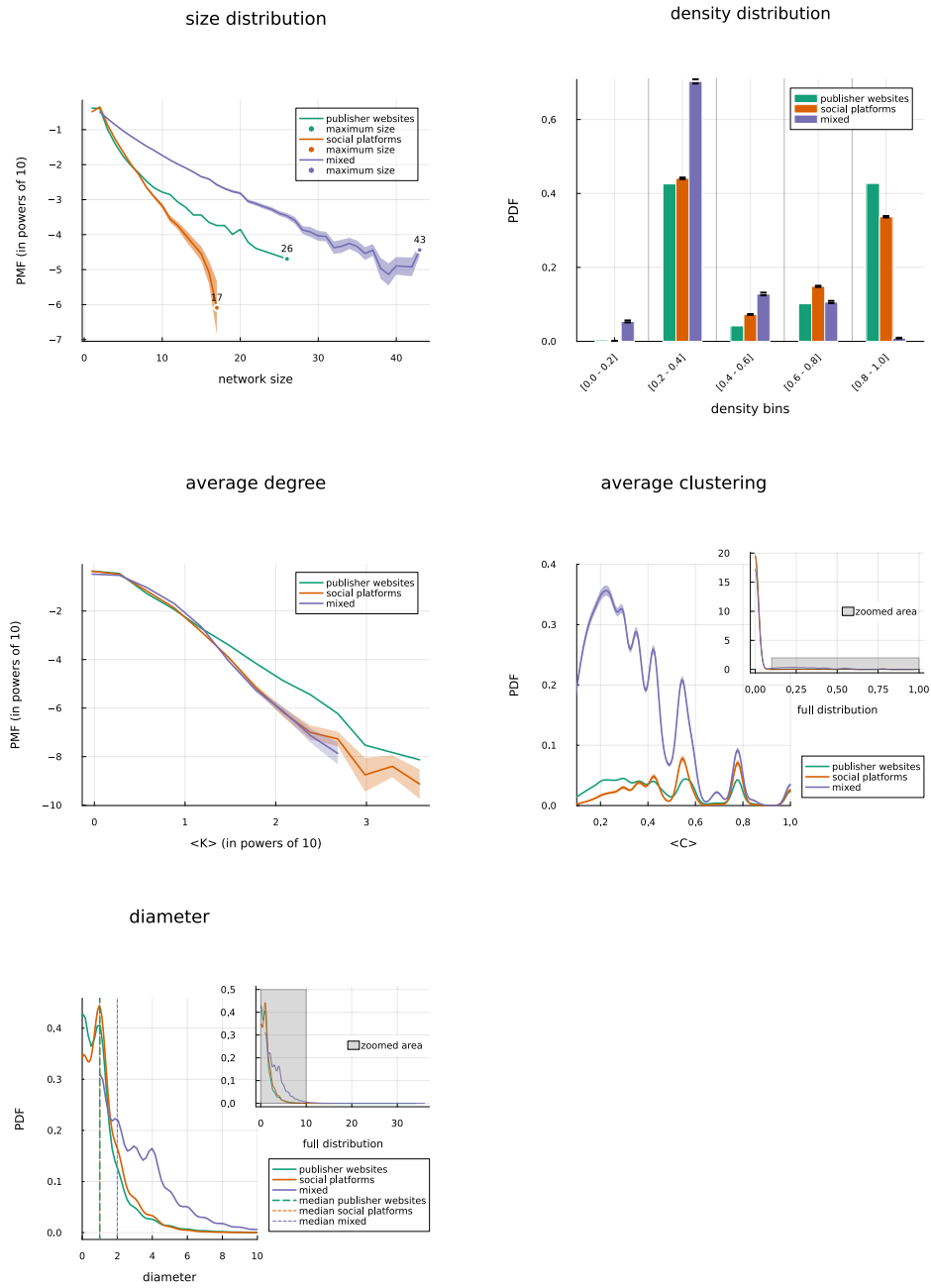


Figure A.7: Network characteristics of users sharing at least 2 tweets aggregated by the type of the source shared, controlling for group size by sampling.

Users networks distributions: country (with sampling)

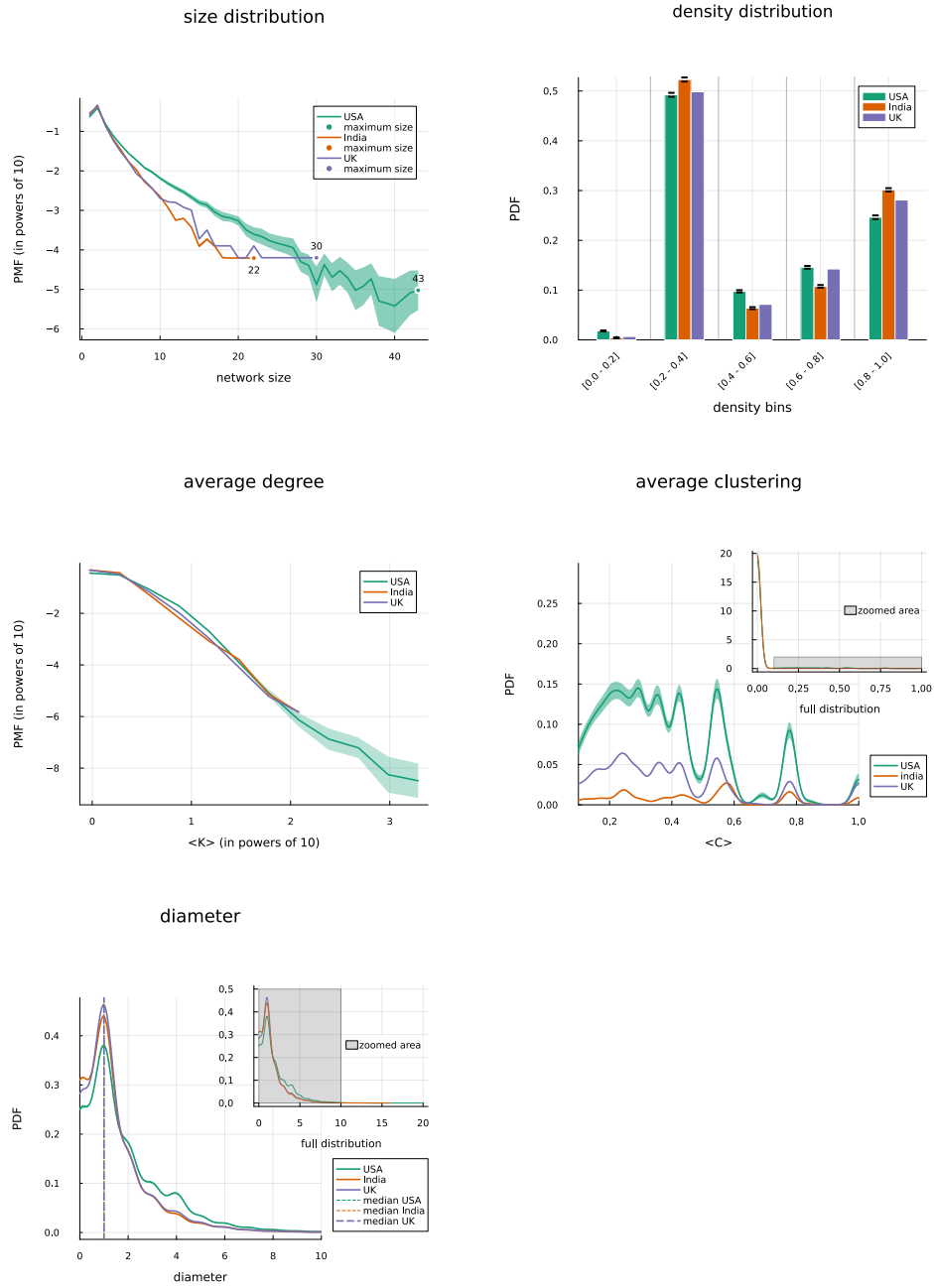


Figure A.8: Network characteristics of users sharing at least 2 tweets aggregated by the three most popular countries, controlling for group size by sampling.

Appendix B

Entropy Analysis

Results of entropy measures obtained by aggregating users by category, filtering users posting at least two tweets. As for the previous analysis, four categories (verification, reliability, country and source type) were balanced for size through sampling, but we did not control for the number of posts. When considering also users with very low source variety, the effective average entropy is always smaller not only than the entropy expected if all sources were equally probable but also than the one expected if users chose sources solely based on their popularity. This suggests that users relying on a very limited set of sources exhibit a more selective and constrained behavior, less subject to source popularity.

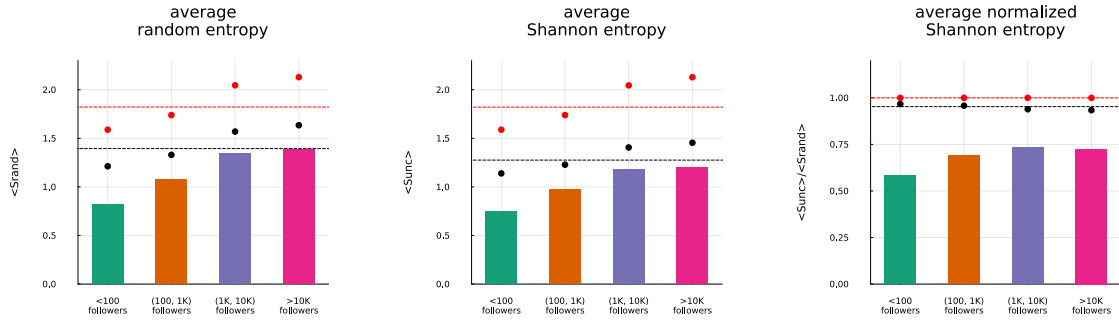


Figure B.1: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by popularity (number of followers). The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets were considered.

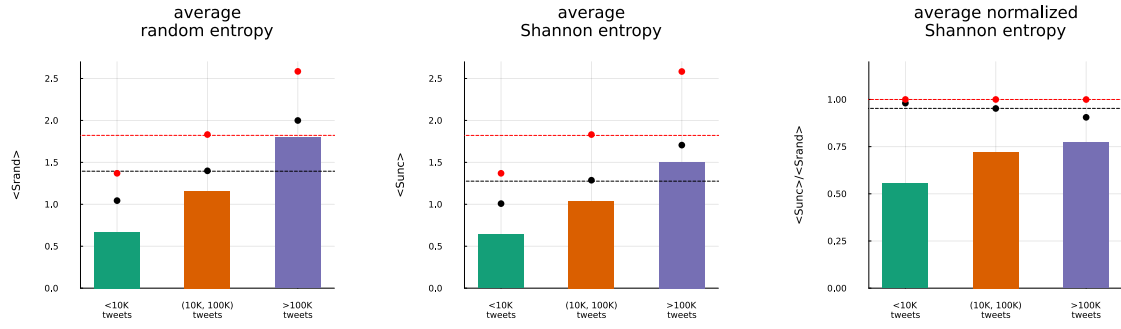


Figure B.2: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by activity (number of tweets posted). The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets where considered.

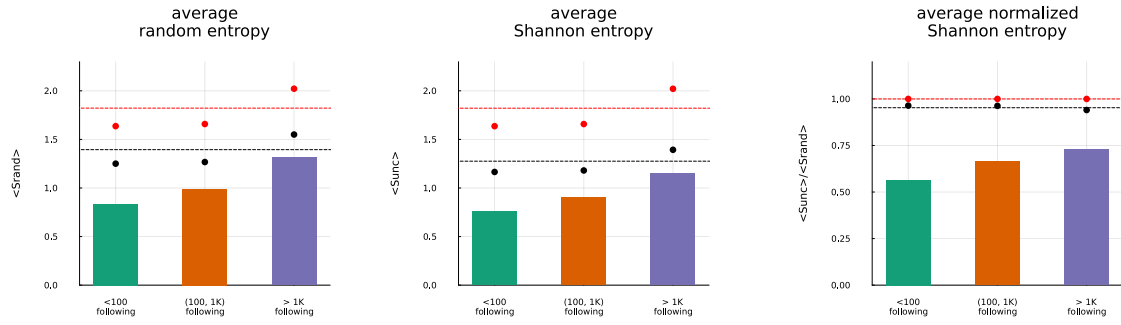


Figure B.3: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by influenceability (number of following). The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets where considered.

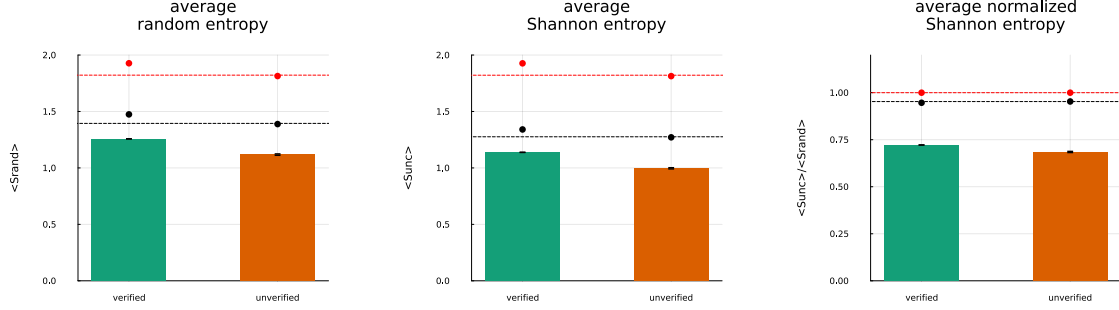


Figure B.4: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by verification status, controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets where considered.

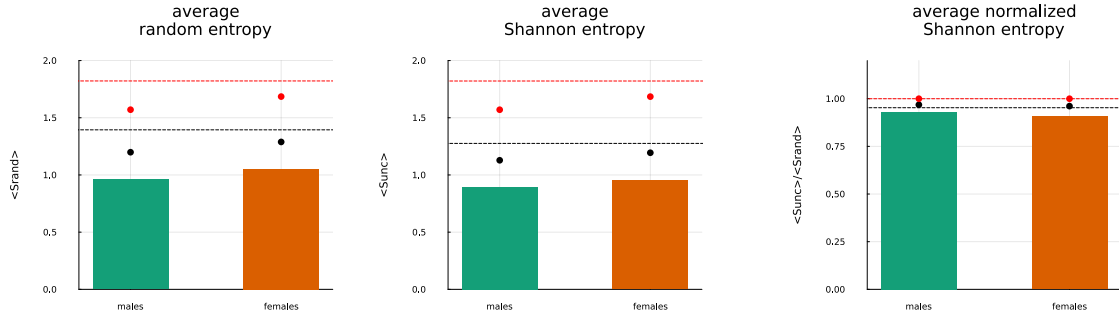


Figure B.5: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by gender. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets where considered.

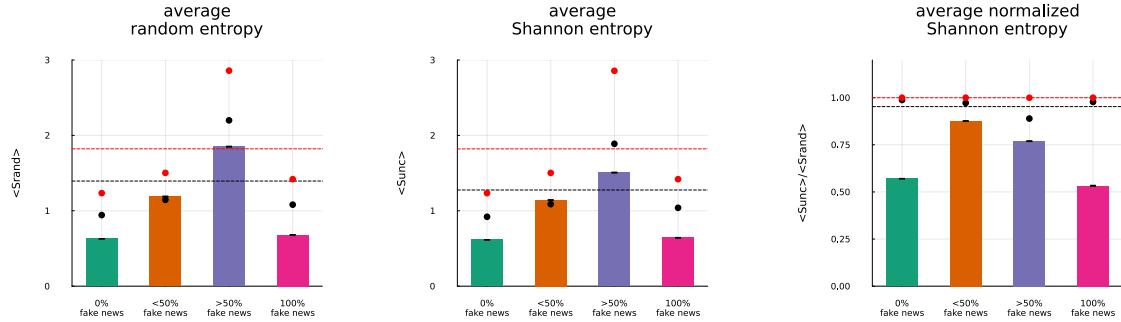


Figure B.6: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by reliability (percentage of fake news shared), controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets were considered.

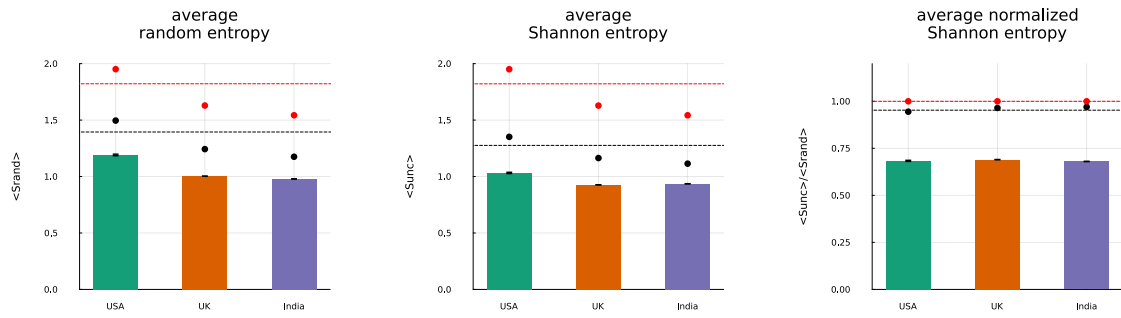


Figure B.7: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by the three top countries by number of users, controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets were considered.

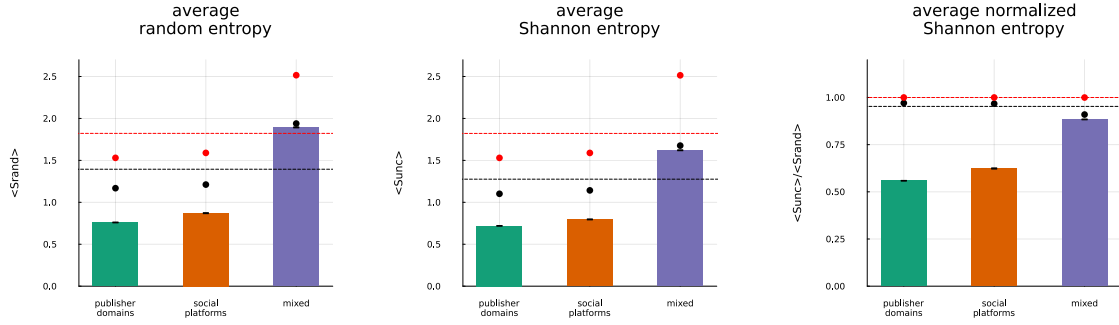


Figure B.8: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by the type of source shared, controlling for group size by sampling. The black dashed line is the average entropy and the black dots are the entropies per category obtained by randomizing over users' choice. The red dashed line is the average entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity. Here, users sharing at least 2 tweets were considered.

Figure B.9 shows an example of density distribution of entropy used to compute the averages analyzed in the Results section 5.4.

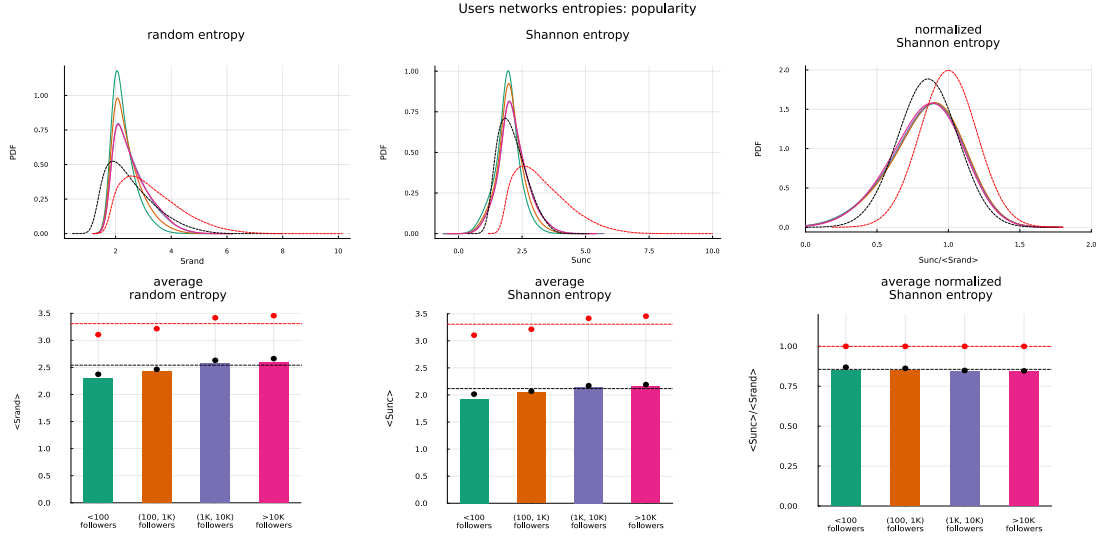


Figure B.9: Average random entropy, average Shannon entropy and normalized Shannon entropy of users networks aggregated by the number of followers. Here, users sharing from at least 4 different sources were considered. The first panel shows the density distributions of entropy of the different users subgroups, while the second panel shows the resulting averages. The black dashed line is the entropy and the black dots are the average entropies per category obtained by randomizing over users' choice. The red dashed line is the entropy and the red dots are the entropies per category obtained by randomizing over sources' popularity.

Bibliography

- [1] Rasmus Kleis Nielsen and Kim Christian Schrøder. «The relative importance of social media for accessing, finding, and engaging with news: An eight-country cross-media comparison». In: *Digital journalism* 2.4 (2014), pp. 472–489 (cit. on pp. 1, 5).
- [2] Oscar Westlund and Mathias A Färdigh. «Accessing the news in an age of mobile media: Tracing displacing and complementary effects of mobile news on newspapers and online news». In: *Mobile media & communication* 3.1 (2015), pp. 53–74 (cit. on pp. 1, 5).
- [3] Babajide Osatuyi. «Information sharing on social media sites». In: *Computers in human behavior* 29.6 (2013), pp. 2622–2631 (cit. on pp. 1, 5).
- [4] David MJ Lazer et al. «The science of fake news». In: *Science* 359.6380 (2018), pp. 1094–1096 (cit. on pp. 1, 6).
- [5] Sandra González-Bailón et al. «Asymmetric ideological segregation in exposure to political news on Facebook». In: *Science* 381.6656 (2023), pp. 392–398 (cit. on pp. 1, 5, 25, 53, 64).
- [6] Tian Yang, Sílvia Majó-Vázquez, Rasmus K Nielsen, and Sandra González-Bailón. «Exposure to news grows less fragmented with an increase in mobile access». In: *Proceedings of the National Academy of Sciences* 117.46 (2020), pp. 28678–28683 (cit. on pp. 1, 25).
- [7] Matthew Tyler, Justin Grimmer, and Shanto Iyengar. «Partisan enclaves and information bazaars: Mapping selective exposure to online news». In: *The Journal of Politics* 84.2 (2022), pp. 1057–1073 (cit. on pp. 1, 25).
- [8] Soroush Vosoughi, Deb Roy, and Sinan Aral. «The spread of true and false news online». In: *science* 359.6380 (2018), pp. 1146–1151 (cit. on pp. 1, 5, 7).
- [9] Jennifer Allen, Duncan J Watts, and David G Rand. «Quantifying the impact of misinformation and vaccine-skeptical content on Facebook». In: *Science* 384.6699 (2024), eadk3451 (cit. on p. 1).
- [10] Anna Bertani, Valeria Mazzeo, and Riccardo Gallotti. «Decoding the news media diet of disinformation spreaders». In: *Entropy* 26.3 (2024), p. 270 (cit. on pp. 1, 2, 8, 14, 23, 30, 46, 48, 64).

- [11] Francesco Pierri, Matthew R DeVerna, Kai-Cheng Yang, David Axelrod, John Bryden, and Filippo Menczer. «One year of COVID-19 vaccine misinformation on Twitter: longitudinal study». In: *Journal of Medical Internet Research* 25 (2023), e42227 (cit. on pp. 1, 6).
- [12] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. «Fake news on Twitter during the 2016 US presidential election». In: *Science* 363.6425 (2019), pp. 374–378 (cit. on pp. 1, 6).
- [13] Ana Vranić, José Reis, Íris Damião, Paulo Almeida, and Joana Gonçalves-Sá. «Global Claims: A Multilingual Dataset of Fact-Checked Claims with Veracity, Topic, and Salience Annotations». In: *Proceedings of the 2nd International Workshop on Diffusion of Harmful Content on Online Web*. 2025, pp. 85–94 (cit. on pp. 2, 7, 15, 16, 20, 44, 63).
- [14] Riccardo Gallotti, Francesco Valle, Nicola Castaldo, Pierluigi Sacco, and Manlio De Domenico. «Assessing the risks of ‘infodemics’ in response to COVID-19 epidemics». In: *Nature human behaviour* 4.12 (2020), pp. 1285–1293 (cit. on pp. 5, 6).
- [15] James Flamino et al. «Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections». In: *Nature Human Behaviour* 7.6 (2023), pp. 904–916 (cit. on pp. 5, 6).
- [16] Alexandre Bovet and Hernán A Makse. «Influence of fake news in Twitter during the 2016 US presidential election». In: *Nature communications* 10.1 (2019), p. 7 (cit. on pp. 5, 6).
- [17] Hunt Allcott and Matthew Gentzkow. «Social media and fake news in the 2016 election». In: *Journal of economic perspectives* 31.2 (2017), pp. 211–236 (cit. on pp. 5, 6).
- [18] Mulugeta Deribe Damota and MW Uninversity. «The effect of social media on society». In: *New Media and Mass Communication* 78.9 (2019), pp. 1–9 (cit. on p. 5).
- [19] Femi Olan, Uchitha Jayawickrama, Emmanuel Ogiemwonyi Arakpogun, Jana Suklan, and Shaofeng Liu. «Fake news on social media: the impact on society». In: *Information Systems Frontiers* 26.2 (2024), pp. 443–458 (cit. on p. 5).
- [20] Kevin Aslett, Zeve Sanderson, William Godel, Nathaniel Persily, Jonathan Nagler, and Joshua A Tucker. «Online searches to evaluate misinformation can increase its perceived veracity». In: *Nature* 625.7995 (2024), pp. 548–556 (cit. on p. 5).
- [21] Seth Flaxman, Sharad Goel, and Justin M Rao. «Filter bubbles, echo chambers, and online news consumption». In: *Public opinion quarterly* 80.S1 (2016), pp. 298–320 (cit. on pp. 5, 64).
- [22] Eytan Bakshy, Solomon Messing, and Lada A Adamic. «Exposure to ideologically diverse news and opinion on Facebook». In: *Science* 348.6239 (2015), pp. 1130–1132 (cit. on pp. 5, 64).

- [23] Marilena Hohmann, Karel Devriendt, and Michele Coscia. «Quantifying ideological polarization on a network using generalized Euclidean distance». In: *Science Advances* 9.9 (2023), eabq2044 (cit. on p. 5).
- [24] Juhi Kulshrestha, Muhammad Zafar, Lisette Noboa, Krishna Gummadi, and Saptarshi Ghosh. «Characterizing information diets of social media users». In: *Proceedings of the international AAAI conference on web and social media*. Vol. 9. 1. 2015, pp. 218–227 (cit. on pp. 6, 64).
- [25] Dimitar Nikolov, Mounia Lalmas, Alessandro Flammini, and Filippo Menczer. «Quantifying biases in online information exposure». In: *Journal of the Association for Information Science and Technology* 70.3 (2019), pp. 218–229 (cit. on pp. 6, 64).
- [26] Andrew Guess, Brendan Nyhan, and Jason Reifler. «Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign». In: *European Research Council* 9.3 (2018), p. 4 (cit. on p. 6).
- [27] Matteo Cinelli, Stefano Cresci, Alessandro Galeazzi, Walter Quattrociocchi, and Maurizio Tesconi. «The limited reach of fake news on Twitter during 2019 European elections». In: *PloS one* 15.6 (2020), e0234689 (cit. on p. 6).
- [28] Andrew Guess, Jonathan Nagler, and Joshua Tucker. «Less than you think: Prevalence and predictors of fake news dissemination on Facebook». In: *Science advances* 5.1 (2019), eaau4586 (cit. on p. 6).
- [29] Piergiorgio Castioni, Giulia Andrighetto, Riccardo Gallotti, Eugenia Polizzi, and Manlio De Domenico. «The voice of few, the opinions of many: evidence of social biases in Twitter COVID-19 fake news sharing». In: *Royal Society Open Science* 9.10 (2022) (cit. on p. 6).
- [30] Mathias Osmundsen, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann, and Michael Bang Petersen. «Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter». In: *American political science review* 115.3 (2021), pp. 999–1015 (cit. on p. 6).
- [31] Jana Lasser, Segun Taofeek Aroyehun, Almog Simchon, Fabio Carrella, David Garcia, and Stephan Lewandowsky. «Social media sharing of low-quality news sources by political elites». In: *PNAS nexus* 1.4 (2022), pgac186 (cit. on pp. 6, 65).
- [32] Sahar Baribi-Bartov, Briony Swire-Thompson, and Nir Grinberg. «Super-sharers of fake news on Twitter». In: *Science* 384.6699 (2024), pp. 979–982 (cit. on p. 6).
- [33] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. «The spread of low-credibility content by social bots». In: *Nature communications* 9.1 (2018), p. 4787 (cit. on pp. 6, 66).
- [34] Sofía M Del Pozo, Sebastián Pinto, Matteo Serafino, Federico Moss, Tomás Cicchini, Hernán A Makse, and Pablo Balenzuela. «Evaluating the relationship between news source sharing and political beliefs». In: *Scientific Reports* 15.1 (2025), p. 516 (cit. on p. 6).

- [35] Sandra González-Bailón et al. «The diffusion and reach of (mis) information on Facebook during the US 2020 election». In: *Sociological Science* 11 (2024), pp. 1124–1146 (cit. on p. 6).
- [36] Guido Caldarelli, Rocco De Nicola, Marinella Petrocchi, Manuel Pratelli, and Fabio Saracco. «Flow of online misinformation during the peak of the COVID-19 pandemic in Italy». In: *EPJ data science* 10.1 (2021), p. 34 (cit. on p. 6).
- [37] Julio CS Reis, Haewoon Kwak, Jisun An, Johnnatan Messias, and Fabricio Benevenuto. «Demographics of news sharing in the US Twittersphere». In: *Proceedings of the 28th ACM conference on hypertext and social media*. 2017, pp. 195–204 (cit. on pp. 6, 35, 65).
- [38] Mohit Chandra, Anush Mattapalli, and Munmun De Choudhury. «Understanding the Humans Behind Online Misinformation: An Observational Study Through the Lens of the COVID-19 Pandemic». In: *arXiv preprint arXiv:2310.08483* (2023) (cit. on pp. 6, 65).
- [39] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. «The COVID-19 social media infodemic». In: *Scientific reports* 10.1 (2020), p. 16598 (cit. on p. 6).
- [40] Bonka Kotseva et al. «Trend analysis of COVID-19 mis/disinformation narratives—A 3-year study». In: *Plos one* 18.11 (2023), e0291423 (cit. on pp. 6, 15).
- [41] Sacha Altay, Rasmus Kleis Nielsen, and Richard Fletcher. «Quantifying the “infodemic”: People turned to trustworthy news outlets during the 2020 coronavirus pandemic». In: *Journal of Quantitative Description: Digital Media* 2 (2022) (cit. on pp. 6, 65).
- [42] Luiz Giovanini et al. «People still care about facts: Twitter users engage more with factual discourse than misinformation». In: *International Symposium on Security and Privacy in Social Networks and Big Data*. Springer. 2023, pp. 3–22 (cit. on pp. 6, 65).
- [43] William Yang Wang. «“liar, liar pants on fire”: A new benchmark dataset for fake news detection». In: *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: short papers)*. 2017, pp. 422–426 (cit. on p. 6).
- [44] Shaina Ashraf, Isabel Bezzaoui, Ionut Andone, Alexander Markowetz, Jonas Fegert, and Lucie Flek. «Defakts: A german dataset for fine-grained disinformation detection through social media framing». In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. 2024, pp. 4580–4591 (cit. on pp. 6, 15).
- [45] Dilip Kumar Sharma and Sonal Garg. «IFND: a benchmark dataset for fake news detection». In: *Complex & intelligent systems* 9.3 (2023), pp. 2843–2863 (cit. on pp. 6, 15).

- [46] João A Leite, Olesya Razuvayevskaya, Kalina Bontcheva, and Carolina Scarton. «EUvsDisinfo: A dataset for multilingual detection of pro-Kremlin disinformation in news articles». In: *Proceedings of the 33rd ACM international conference on information and knowledge management*. 2024, pp. 5380–5384 (cit. on pp. 6, 15).
- [47] Anees Baqir, Alessandro Galeazzi, and Fabiana Zollo. «News and misinformation consumption: A temporal comparison across European countries». In: *Plos one* 19.5 (2024), e0302473 (cit. on p. 7).
- [48] Duncan J Watts and Peter Sheridan Dodds. «Influentials, networks, and public opinion formation». In: *Journal of consumer research* 34.4 (2007), pp. 441–458 (cit. on p. 7).
- [49] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. «The spreading of misinformation online». In: *Proceedings of the national academy of Sciences* 113.3 (2016), pp. 554–559 (cit. on p. 7).
- [50] Carolina Becatti, Guido Caldarelli, Renaud Lambiotte, and Fabio Saracco. «Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections». In: *Palgrave Communications* 5.1 (2019) (cit. on p. 7).
- [51] Bao Tran Truong, Oliver Melbourne Allen, and Filippo Menczer. «Account credibility inference based on news-sharing networks». In: *EPJ Data Science* 13.1 (2024), p. 10 (cit. on pp. 7, 64, 66).
- [52] Tomas Cicchini, Sofia Morena Del Pozo, Enzo Tagliazucchi, and Pablo Balenzuela. «News sharing on Twitter reveals emergent fragmentation of media agenda and persistent polarization». In: *EPJ Data Science* 11.1 (2022), p. 48 (cit. on p. 7).
- [53] Valeria Mazzeo and Andrea Rapisarda. «Investigating fake and reliable news sources using complex networks analysis». In: *Frontiers in Physics* 10 (2022), p. 886544 (cit. on p. 7).
- [54] Subhayan Mukerjee, Tian Yang, Georg Stadler, and Sandra González-Bailón. «What counts as a weak tie? A comparison of filtering techniques to analyze co-exposure networks». In: *Social Networks* 68 (2022), pp. 386–393 (cit. on p. 7).
- [55] Claude Elwood Shannon. «A mathematical theory of communication». In: *The Bell system technical journal* 27.3 (1948), pp. 379–423 (cit. on pp. 7, 13).
- [56] Manuel Pratelli, Fabio Saracco, and Marinella Petrocchi. «Entropy-based detection of Twitter echo chambers». In: *PNAS nexus* 3.5 (2024), pgae177 (cit. on pp. 8, 14, 23, 64).
- [57] Manar Alsaid, Siva Parvathi Panguluri, and Suliman Hawamdeh. «Combating misinformation on social media using social noise and social entropy as a measure of uncertainty». In: *Proceedings of the Association for Information Science and Technology* 61.1 (2024), pp. 25–35 (cit. on pp. 8, 14, 23).

- [58] Lokesh Jain. «An entropy-based method to control COVID-19 rumors in online social networks using opinion leaders». In: *Technology in Society* 70 (2022), p. 102048 (cit. on pp. 8, 23).
- [59] Louise A Dennis, Yu Fu, and Marija Slavkovik. «Markov chain model representation of information diffusion in social networks». In: *Journal of Logic and Computation* 32.6 (2022), pp. 1195–1211 (cit. on pp. 8, 23).
- [60] Roberto Casarin, Antonio Peruzzi, and Mark FJ Steel. «Media bias and polarization through the lens of a Markov switching latent space network model». In: *The Annals of Applied Statistics* 19.4 (2025), pp. 3416–3437 (cit. on pp. 8, 23).
- [61] Flavio Chierichetti, Ravi Kumar, Prabhakar Raghavan, and Tamas Sarlos. «Are web users really markovian?» In: *Proceedings of the 21st international conference on World Wide Web*. 2012, pp. 609–618 (cit. on p. 8).
- [62] Mark EJ Newman. «The structure and function of complex networks». In: *SIAM review* 45.2 (2003), pp. 167–256 (cit. on pp. 9, 13).
- [63] Albert-László Barabási₂₀₂₆. *Network Science*. 2026. URL: <https://networksciencebook.com/> (cit. on pp. 9, 11).
- [64] Réka Albert and Albert-László Barabási. «Statistical mechanics of complex networks». In: *Reviews of modern physics* 74.1 (2002), p. 47 (cit. on pp. 11, 13).
- [65] Mark EJ Newman. «Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality». In: *Physical review E* 64.1 (2001), p. 016132 (cit. on p. 13).
- [66] Petter Holme, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. «Attack vulnerability of complex networks». In: *Physical review E* 65.5 (2002), p. 056109 (cit. on p. 13).
- [67] James V Stone. «Information theory: a tutorial introduction to the principles and applications of information theory». In: (2024) (cit. on p. 13).
- [68] John R Waldram. «The theory of thermodynamics». In: *The Theory of Thermodynamics* (1985) (cit. on p. 13).
- [69] Monica Borda. *Fundamentals in information theory and coding*. Springer Science & Business Media, 2011 (cit. on p. 14).
- [70] Qiangeng Yang, Tess Christensen, Shlok Gilda, Juliana Fernandes, Daniela Oliveira, Ronald Wilson, and Damon Woodard. «Are fact-checking tools helpful? an exploration of the usability of google fact check». In: *International conference on data and information in online*. Springer. 2024, pp. 82–98 (cit. on p. 16).
- [71] Jennifer Golbeck et al. «Fake news vs satire: A dataset and analysis». In: *Proceedings of the 10th ACM conference on web science*. 2018, pp. 17–21 (cit. on p. 16).

- [72] Muhammad Bilal Zafar, Parantapa Bhattacharya, Niloy Ganguly, Saptarshi Ghosh, and Krishna P Gummadi. «On the wisdom of experts vs. crowds: Discovering trustworthy topical news in microblogs». In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 2016, pp. 438–451 (cit. on p. 16).
- [73] Bryan D Jones, Frank R Baumgartner, Sean M Theriault, Derek A Epp, Cheyenne Lee, Miranda E Sullivan, and Chris Cassella. «Policy agendas project: codebook». In: *Comparative Agendas Project*. URL: <https://www.comparativeagendas.net/us> (2023) (cit. on p. 16).
- [74] Marco Vicente, Fernando Batista, and Joao Paulo Carvalho. «Twitter gender classification using user unstructured information». In: *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE. 2015, pp. 1–7 (cit. on p. 19).
- [75] 2001. URL: https://en.wikipedia.org/wiki/Domain_Name_System (cit. on p. 20).
- [76] Mor Oren-Loberman, Vered Azar, and Wasim Huleihel. «Online auditing of information flow». In: *IEEE Transactions on Signal and Information Processing over Networks* 10 (2024), pp. 487–499 (cit. on p. 23).
- [77] Peter C Austin. «An introduction to propensity score methods for reducing the effects of confounding in observational studies». In: *Multivariate behavioral research* 46.3 (2011), pp. 399–424 (cit. on pp. 27, 28).
- [78] Paul R Rosenbaum and Donald B Rubin. «The central role of the propensity score in observational studies for causal effects». In: *Biometrika* 70.1 (1983), pp. 41–55 (cit. on pp. 27, 28).
- [79] Brian K Lee, Justin Lessler, and Elizabeth A Stuart. «Improving propensity score weighting using machine learning». In: *Statistics in medicine* 29.3 (2010), pp. 337–346 (cit. on p. 28).
- [80] Daniel F McCaffrey, Greg Ridgeway, and Andrew R Morral. «Propensity score estimation with boosted regression for evaluating causal effects in observational studies.» In: *Psychological methods* 9.4 (2004), p. 403 (cit. on p. 28).
- [81] Soko Setoguchi, Sebastian Schneeweiss, M Alan Brookhart, Robert J Glynn, and E Francis Cook. «Evaluating uses of data mining techniques in propensity score estimation: a simulation study». In: *Pharmacoepidemiology and drug safety* 17.6 (2008), pp. 546–555 (cit. on p. 28).
- [82] Paul R Rosenbaum. «Observational studies». In: *Observational studies*. Springer, 2002, pp. 1–17 (cit. on p. 28).
- [83] Jon Roozenbeek and Sander Van der Linden. «Fake news game confers psychological resistance against online misinformation». In: *Palgrave Communications* 5.1 (2019), p. 65 (cit. on p. 65).
- [84] David Liben-Nowell and Jon Kleinberg. «The link prediction problem for social networks». In: *Proceedings of the twelfth international conference on Information and knowledge management*. 2003, pp. 556–559 (cit. on p. 66).

Acknowledgements

My sincere thanks go to the Social Physics and Complexity Group (SPAC, LIP) in Lisbon for hosting me during my Master's research internship and for making it such a memorable experience. From the first day, I felt really welcome, not just as a visiting student, but as a real part of the group.

I am deeply grateful to Prof. Joana Gonçalves de Sá for her supervision and guidance throughout this experience. Her mentorship helped me grow as a researcher and pushed me to think more carefully about what it means to do science that actually matters.

A very special thank you goes to my tutor, Dr. Ana Vranić, for her patience, her support, and her availability that extended well beyond my permanence in Lisbon. Working closely with her gave me a real understanding of what the research process looks and feels like from the inside.

To the whole SPAC group: thank you for the conversations, the group meetings, and the general sense that curiosity is always worth following. You showed me, in practice, what it looks like when people from different scientific backgrounds come together around a shared question and how valuable different perspectives are.

Alla mia famiglia e i miei amici, grazie per il supporto costante, per avermi sempre aiutata a relativizzare le difficoltà quando sembravano insuperabili e per avermi ricordato che il valore di una persona non si misura con i numeri.