POLITECNICO DI MILANO
Corso di Laurea Magistrale in Ingegneria Informatica
Scuola di Ingegneria Industriale e dell'Informazione

Assessing Italian disinformation spread on Twitter during the 2019 European elections

Relatore: Prof. Stefano Ceri
Correlatore: PhD cand. Francesco Pierri

Tesi di Laurea di:
Alessandro Artoni
Matricola 899343

Anno Accademico 2018-2019

Prometti solo di fare del tuo meglio...
  quel che puoi, come puoi, del tuo meglio.

Lezàrd

Ringraziamenti

Sono sempre stato negato con i ringraziamenti.. Di solito mi viene meglio chiedere scusa.

Vorrei prima di tutto ringraziare il Professor Stefano Ceri e il mio corelatore Francesco Pierri. Non solo senza di loro questo lavoro non sarebbe stato possibile ma se oggi mi sento un cittadino (informatico) migliore è grazie ai loro consigli e al loro supporto.

Un grazie enorme va ai miei genitori, per il loro sostegno in tutto e per tutto da sempre.

Ringrazio mia sorella Carlotta, per avermi insegnato che la determinazione è la chiave del successo.

Grazie a mio fratello gemello Alberto, per la complicità dimostrata in tutti questi anni.

Vorrei dire grazie anche alle persone che ho incontrato e che mi hanno supportato in collegio. Grazie Don Luca, Don Luciano e Don Franco. Grazie a Marco per l'ascolto incodizionato e per aver saputo stimolare la mia curiosità e grazie ai collegiali tutti. Tra questi, voglio dire un grazie *grosso* a Massimiliano e Stefano, per le loro opinioni, la loro passione e la loro voglia di emergere, insieme.

Grazie al gruppo di Poviglio, per la loro compagnia ed amicizia, in particolare a chi in questi anni è stato sempre presente, nel bene e nel male.

Ringrazio i miei compagni di Liceo, per ricordarmi continuamente cosa vogliono dire spensieratezza e allegria.

Ringrazio chi, come Marta e Lorenzo, ha allietato i miei viaggi in treno in questi anni, condividendo le tristi sorti dell'essere pendolare.

Vorrei ringraziare anche tutte le magnifiche persone che ho incontrato in questi anni di università: dagli amici del Laboratorio, ai miei compagni passati e presenti.

In particolare ringrazio Federico e Luca. Il primo per la sua ingenuità in certe cose, la voglia di fare e la sua determinazione; il secondo un po' per tutto. Ma soprattutto grazie, ad *entrambi*, perché ci siete sempre stati.

Ringrazio infine Francesca, per il loro sostegno in tutto e per tutto *per* sempre.

III

IV

Sommario

Dopo le elezioni Americane del 2016 è cresciuto esponenzialmente l'interesse dell'intera communità accademica nei confronti della disinformazione, e di come questa possa guidare o modificare l'opinione degli elettori al fine di influenzarne il voto.

Questo interesse ha portato alla pubblicazione di numerose analisi sui motivi, economici e politici, che portano alla creazione e alla condivisione di disinformazione. D'altra parte, si è cercato di descriverne l'entità, di verificare se effettivamente sia stata rilevante nella campagna americana o se ci siano dei profili più vulnerabili e perciò più sensibili alla sua condivisione, o infine se esistano evidenze di particolari attori che hanno coordinato e creato intere campagne disinformative, per esempio i cosiddetti "troll russi".

C'è ancora molto dibattito inoltre, su quali possano essere le migliori strategie per frenare la diffusione di notizie false. Infatti, se qualcuno intervenisse regolamentando troppo, esiste un rischio di limitazione alla libertà di parola e quindi di nascita di una vera e propria censura legalizzata. D'altra parte se non si regolamenta esiste il rischio che le notizie false possano giocare un ruolo nell'orientare l'elettorato.

Tenendo presente queste considerazioni, il nostro lavoro punta a capire e caratterizzare la disinformazione nel contesto italiano, concentrandoci sulle Elezioni Europee del 2019. Per farlo, indaghiamo la propagazione di contenuti su Twitter prodotti da un insieme di siti che tipicamente condividono notizie complottistiche, di propaganda o palesemente false. In primo luogo forniamo una descrizione dei volumi delle loro interazioni e descriviamo le narrative che hanno caratterizzato il dibattito dei portali selezionati.

Successivamente, tramite la creazione di una rete di diffusione, individuiamo quali sono i nodi principali fornendo suggerimenti per poter efficacemente distruggere la rete, frenando la diffusione di disinformazione.

Infine, mostriamo l'esistenza di collegamenti tra siti di disinformazione, portali di informazione tradizionali e siti di notizie spazzatura europei, russi e americani.

Abstract

Since 2016 American elections, the interest of academia on misinformation and how it could drive the opinion of voters has grown exponentially. Numerous analyses have been published on the reasons, both economical and political, that led to the creation and the share of disinformation. At the same time, researchers tried to address its entity, or to verify how relevant disinformation has been during the U.S. Presidential Elections, or also to identify if there are vulnerable targets, or if there are evidences of particular actors that created coordinated disinformation campaigns - such as the so called "Russian trolls".

There's still a lot of debate towards what could be the best strategies to curb fake news spread. If on one hand you regulate too much, there is the risk that limiting the freedom of speech may result in censorship. On the other hand, if you do not regulate, there is the possibility that some actors will sponsor their own agenda by producing false news.

Bearing this in mind, our work aims to understand and characterize disinformation in Italy, focusing on the 2019 European Elections. To do that, we investigate the propagation of content on Twitter produced by a set of website that typically share conspiracy theories, propaganda or completely invented fake news. First, we describe the volume of interactions, and the narratives that characterized the disinformation websites we monitored. Next, we build a diffusion network, and we identify who are the most important nodes, giving highlights on how to destroy the diffusion network and effectively curbing disinformation spread

We finally show the existence of links between disinformation websites, mainstream outlets and European, Russian and American junk news websites.

Contents

List of Figures

List of Tables

XVI

Chapter 1

Introduction

1.1 Context and Problem Statement

When taking decision, it is fundamental to have correct and accurate information. Imagine you want to buy a new house: you would like to know exactly where it is, how big it is, how much it will cost you. Then you would also like to have some opinions on how your neighbors are, how easily is to access that zone, if the mayor plans to do some investments, and so on. You would like to be fully-aware before taking your decision.

One of the most influential decision we take is voting and we ground our decision mostly on information we get. Having well documented and informed citizens is most relevant in order to obtain a better democracy. Democracy as we know it is at risk: the way we get information has dramatically changed with the growth and the widespread of social medias.

We want live, fast and short news that we can easily consult using a smartphone. As reported by AGCOM [1], the Italian authority for communications guarantees, even if online news are still perceived as less trusted then traditional outlets, the trend of getting information from the web has risen and interestingly 41.8% of Italians consume news **every day** from internet.

In this scenario, a lot of inaccuracies are often published. Some are inevitable as they appear also in online mainstream outlets, but some others are instead peculiar of the new media; they are produced by malicious actors with the objective of manipulating people's opinion. Misinformation has found a new way to reach people: and it's broader and deeper than ever.

Since the first suggestion that fake news might have played an important role in 2016 U.S. presidential elections [1], awareness of citizens and organization has risen a lot and a lot has been made so far. The European

---

[1]Report AGCOM 19/02/2018.

Union, for example, started a very large campaign against disinformation [2], with guidelines [3], events and platforms [4] where they verify statements (an example is showed in figure 1.1) made by European politicians.

Mostly true: Italy's employment rates peaking while youth unemployment at its lowest, claims Italian minister

False: "The Netherlands has the most flexible labour market in Europe"

Figure 1.1: Example of news fact checking from a eufactcheck.eu. The first one reports a statement from the Italian Minister of Infrastructure and Transport Danilo Toninelli which turned out to be mostly True. The second is a claim from Dutch MEP Agnes Jongerius which was completely False.

ONGs as well become active and reported malicious accounts to social media companies [5] who are starting to actively be part of the ecosystem of tackling disinformation online [6] closing some pages that spread hate [7]. Still, a lot more has be done to effectively curb fake news spread.

Misinformation actually, is not new at all in history. It has always existed in the form of **propaganda**, **satire** and also **hoaxes**. For example, in 1835, on the New York Sun a series of articles, known as the Great Moon Hoax [1], pitched the existence of life of the Moon.

In recent years, the term **fake news** has been used to indicate broader aspects of misinformation, and lately with a political purpose: whatever is used as a news against a politician can be labelled as a fake news. For this reason, a further focus on the terminology is needed.

---

[2]EU - Tackling online disinformation.

[3]EU - Code of Practice on disinformation.

[4]www.eufactcheck.eu.

[5]Avaaz report.

[6]Facebook statement on disinformation.

[7]Facebook closes 23 Italian pages sharing fake news and hate.

1.2 Terminology

Following Pierri [2] we provide a list of the definitions from the literature, which is by no means exhaustive. While there is common agreement that these terms indicate deceptive information, it appears that an agreed and precise definition is still missing.

1. **False news** are articles that are considered potentially or intentionally misleading for the readers, as they are deliberately false. They can represent fabricated information which mimics traditional news content in form, but not in the intent or the organizational process.

2. **Fake news** are false news with a political connotation.

3. **Misinformation** is all kind of information that is inaccurate or misleading, that is usually spread without harmful purpose.

4. **Disinformation** is instead, is false news consciously spread to promote a biased agenda, or to deceive people.

5. **Hoaxes**, which is a very common term in Italy attached improperly to fake-news, is a kind of disinformation, more *humorous* and *mischievous*.

6. **Satirical** news are written primary to entertain readers, but similarly to hoaxes, they can be shared harmfully and they can be misunderstood. They are easily recognized by their humor, their absurdity and irony.

7. **Propaganda** is defined as information that tries to influence emotions, opinions and actions of the target audiences by means of deceptive, selectively omitted and one-sided messages. The main purpose is to drive a particular political, ideological or religious agenda.

8. **Click-bait** is a very low kind of journalism which is intended to attract traffic and monetize via advertisement revenues.

9. **Junk news** is a more broader term that indicates propaganda, hyperpartisan or conspiratorial news and information referred more commonly to a publisher rather than a to a single article.

10. **Rumors** are claims that do not have origin from any news events.

The topic of disinformation therefore has many faces, and needs a deeper investigation.

1.3   Position of this work

In this work, we investigate the presence (and the influence) of the phenomenon of disinformation in Italy, focusing on the run-up 2019 European Parliament elections, monitoring the period between the 7th of January and 26th of May, the election day. In particular, we address the following research questions:

RQ1: What was the reach of disinformation which circulated on Twitter in the run-up to European Parliament elections? How active and strong was the community of users sharing disinformation?

RQ2: What were the most debated narratives of disinformation? How much were they influenced by national vs European-scale topics?

RQ3: Who were the most influential spreaders of disinformation? Did they exhibit precise political affiliations? How could we dismantle the disinformation network?

RQ4: Did disinformation outlets organize their deceptive strategies in a coordinated manner? Can we identify inter-connections across different countries?

Following these research questions, we show that a few outlets accounted for most of the deceptive information circulating on Twitter, which was driven by controversial and polarizing topics of debate such as immigration, national safety and (Italian) nationalism. Besides, we showed that few users account for most of the disinformation spread, with political affiliations towards the Italian conservative and far-right political environment. We finally conclude showing the existence of an intricate network of interconnections between different disinformation outlets across Europe, U.S. and Russia, which seemingly acted in a coordinated manner in the period before the elections. Results presented in this work are also presented in [3], and are currently under review for publication.

1.4   Outline

The thesis structure, after pitching the background and the related work, follows the research questions:

- Chapter *2* presents the background knowledge that has been used to support the analysis.

- Chapter *3* describes the related work this thesis is based on.

- Chapter *4* contains the description of the pipeline used to collect our data, a brief description of the data set itself, including a comparison with other social media.

- Chapter *5* starts the investigation on the most relevant disinformation outlets, trying to assess who were the main sources and what was their reach [**RQ1**].

- Chapter *6* shows what were the most debated topic of disinformation, and what was their impact [**RQ2**].

- Chapter *7* analyses the diffusion network of tweets, highlighting political affiliations, and showing how robust it is [**RQ3**].

- Chapter *8* investigate the possibility of disinformation outlets to have a sort of coordination [**RQ4**].

- Chapter *9* summarizes the work, discussing the results and proposes future enhancements.

Chapter 2

Background

This chapter contains all the techniques and methodologies that have been used to ground this thesis.

2.1 Natural Language Processing

NLP is the field of Artificial Intelligence that tries to get computers to perform useful tasks involving human language [4]. It includes fields such as speech and language processing, speech recognition and synthesis, computational linguistics and so on.

2.1.1 Regular expressions

Regular expressions are one of the most well-known tools of NLP and can be used to extract strings that follow specific patterns from documents. A regular expression (regexp) is a string r defined over the letters of the terminal alphabet Σ and that contains a few metasymbols, according to the following cases, where s and t are regexp:

- $r = \emptyset$

- $r = a, a \in \Sigma$

- $r = s \cup t$

- $r = (s.t) = (st)$

- $r = (s)^*$

2.1.2 Tf-Idf

A way to understand how much a word t is important in a document d having a set of documents, is **tf-idf**. Tf stands for **term frequency** and it is defined as follows:

$$tf_{t,d} = \begin{cases} 1 + \log_{10} \text{count}(t,d) & \text{if count}(t,d) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Idf instead means **inverse document frequency** and it is a way to assign to more rare words a more important weight. It is defined as follow:

$$idf_t = \log_{10}\left(\frac{N}{df_t}\right)$$

Tf-idf then considers the weight, or the "importance" of a word t in document d, $w_{t,d}$ combining thus the tf with idf:

$$w_{t,d} = tf_{t,d} \times idf_t$$

2.1.3 Part of Speech tagging

It is commonly attributed to Dionysius Thrax of Alexandria (c. 100 B. C.) the first description of modern linguistic vocabulary. It also included the description of **part of speech tagging:** *noun, verb, pronoun, preposition, adverb, conjunction, participle* and *article*. **POS** tagging is very useful, because it can reveal a lot about a word and its neighbors. There are different approaches mainly divided in **rule based techniques** and **statistical based techniques** [4].

Given a sentence, POS tagging tries to assign to each *word* its own *tag*, by maximizing the probability that a certain word has a particular tag. We then want to find $\hat{t}_1^n$, which is the sequence of tags $t_1...t_n$ that maximizes the probability $P(t_1...t_n|w_1...w_n)$. In other words:

$$\hat{t}_1^n = \arg\max_{t_1^n} P(t_1^n|w_1^n) \tag{2.1}$$

To make the formula operational, Bayes Rule can be used:

$$\hat{t}_1^n = \arg\max_{t_1^n} P(w_1^n|t_1^n)P(t_1^n) \tag{2.2}$$

The first term $P(w_1^n|t_1^n)$ is called *likelihood* and the second $P(t_1^n)$ *prior*. The formula can be further simplified by making Markov and Independence

assumptions. The first says that the probability of a state depends only on the k-th previous states. This creates a k-th order Markov model. The second says that the probability of a word w_n depends only on its POS tag t_n. The formula becomes:

$$\hat{t}_1^n = \arg\max_{t_1^n} \prod_{i=1}^{n} P(w_i|t_i)P(t_i|t_{i-k}...t_{i-i}) \qquad (2.3)$$

It's important to note that mainly first and second order Markov model are used.

Methods then, differ for how the prior - also called transition probability - is estimated. N-grams often uses the *Maximum likelihood estimation*(MLE) principle. If we consider a second order Markov model, given $F(t_{n-2}t_{n-1}t_n)$ the number of occurrences of the trigram $t_{n-2}t_{n-1}t_n$ in the corpus, and $F(t_{n-2}t_{n-1})$ the number of occurrences of the bigram $t_{n-2}t_{n-1}$, then the transition probability is estimated in this way:

$$P(t_n|t_{n-2}t_{n-1}) = \frac{F(t_{n-2}t_{n-1}t_n)}{F(t_{n-2}t_{n-1})} \qquad (2.4)$$

This approach has some limitations: it is hard to understand if a trigram with zero probability is simply incorrect, or rare. Therefore some smoothing techniques are applied to take care of this issue.

In our work, we use a probabilistic tree tagger that exploits decision trees [5]. This approach in contrast, estimates the transition probability using a binary decision tree performing overall better reaching a $\sim 96.3\%$ accuracy.

2.1.4 Sentiment Analysis

Sentiment Analysis is the computational study of people's opinions, appraisals, attitudes and emotions toward entities [6]. A typical application of this technique is to understand a customer opinion towards a product in an automated way. There are two main techniques to do sentiment analysis:

- **Classification based on Supervised Learning**. The problem could be formulated as a supervised learning problem with three classes (positive, neutral and negative).

- **Classification based on Unsupervised Learning**. Another approach is to identify a priori words and phrases that are dominating indicators for sentiment classification and exploit them for the classification task.

Both of the techniques could also exploits some typical NLP features such as *tf-idf*, *POS tagging*, *syntax dependency* and so on.

2.2   Network analysis

2.2.1   Network definition

A **network** or a **graph** (mathematically) is a collection of vertices (or nodes) connected by edges. Formally: given V the set of vertices and $E \subseteq \{\{x, y\}|(x, y) \in V^2 \wedge x \neq y\}$ the set of edges, $G = (V, E)$ is a graph. There are different ways to represent a network, one of the most used is using an **adjacency matrix**. The adjacency matrix $\mathbf{A}$ of a simple graph is the matrix with elements a_{ij} such that:

$$a_{ij} = \begin{cases} 1, & \text{if i and j are connected by an edge} \\ 0, & \text{otherwise} \end{cases}$$

Weighted Networks

In some networks, there might be some edges more important than others. It is possible to attach a label to the edge that indicates the weight of that edge.

Directed Networks

A **directed network** (also called DiGraph), is a network in which each edge has a direction, pointing from one node to another.

2.2.2   Multipartite Networks

In the case where vertices belong to some sort of group, the network is called **Multipartite network** or **k-partite graph**, where k is the number of partitions. To represent the graph it is used a rectangular **incidence matrix**, where given n the number of nodes in the network and g the number of groups, the incidence matrix $\mathbf{B}$ is a $g \times n$ matrix having elements b_{ij} such that:

$$b_{ij} = \begin{cases} 1, & \text{if vertex j belongs to group i} \\ 0, & \text{otherwise} \end{cases}$$

2.2.3 Centrality metrics

Most of the ideas of centrality in networks come from the study of social networks and in particular from social sciences. The focus is to find the most important nodes in the network, but **"what is importance?"**.

Degree centrality

One of the simplest metrics is **Degree Centrality**, which is the number of edges connected to a node. In Directed networks, nodes can have both **In-Degree** and **Out-Degree**, where the former is the number of incoming edges and the latter is the number of outgoing edges.

Taking as an example a social media context, this metric supposes that people with many connections (highest degree) are more influential than others. Another classical example from the academic world, takes into consideration a set of papers and builds a citation network from this set. The more a paper is important, the higher its in-degree will be. is considering the number of citation of a paper, which is the in-degree in a citation network.

Strength

Considering Weighted Directed graphs, nodes have both **In-Strength** and **Out-Strength**, where the former is the sum of the weights belonging to incoming edges, and the latter is the sum of the weights of all outgoing edges. The **Strength** of the node is then the sum between in-strength and out-strength.

PageRank

PageRank centrality become very famous since it was used in the first Google's search engine implementation [7]. It takes in consideration both the count and the quality of links in a page to estimate the importance of a website in a network, assuming that the most important websites will likely be the ones that receive more links from other websites. PageRank also tries to intuitively model random walk on graphs, in its simplest version it corresponds to the probability distribution of a random walk over a graph.

Betweenness

Betweenness is a different concept of centrality [8]. It exploits the idea that, considering all the shortest paths that links two node, a node importance increase considering all the shortest paths it belongs to. Formally,

betweenness is defined as the sum of the ratio between the number of minimum paths passing through a node, and all the minimum paths linking the two nodes, for all pairs of nodes.

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (2.5)$$

Where v,s,t are vertices, σ_{st} is the minimum path between the vertex v and t, and $g(v)$ is the betweenness of a vertex.

K-core

K-core is defined as the maximum subset of nodes such that each node is connected to at least k others in the subset [9]. It is very on of the most used centrality metrics since it is very easy to compute.

2.2.4 Network robustness

The process of taking a vertex and removing it from a graph is called percolation [10].

A network is robust if it can resist failures and perturbations. A well-known method to test the robustness of a network is percolation. By selecting some nodes and removing them from the network randomly you could test if the network still works, or how much it is still connected. In general, some metrics could also be used to understand what are the best nodes you can disconnect to eliminate the maximum number of edges.

2.2.5 Community detection

Another studied problem in networks is community detection. The goal is to partition a network into groups of highly connected vertices. This partition could be interesting as it could reveal more information on nodes themselves thanks to their connections. The *quality* of the partition is often measured using a scalar value called **modularity** [12] that has values between -1 and +1 and that measures the density of links inside a community, compared to the others. In particular, considering weighted networks, modularity is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \qquad (2.6)$$

In the equation, A_{ij} is the weight of an edge between vertex i to vertex j, $k_i = \sum_j A_{ij}$ is the sum of weights of all edges from i, c_i is the community

Figure 2.1: Community detection applied to a network, courtesy of Becatti et at al. [11]

.

to which vertex i is assigned, the $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise, while $m = \frac{1}{2} \sum_{ij} A_{ij}$. Sadly, using this kind of modularity only fails for very large networks. Blodel et al. [13] proposed a faster way to calculate communities for very large networks. In the first step, they assign communities to all nodes, having as many nodes as communities. Then, for each node i, they consider its neighbours j and evaluate the gain of modularity by eventual removal of node i from its community, finally placing it in the community where the gain is maximized. The process continues until all nodes cannot have further improvements, reaching a local maximum. Note that the output of the program depends on how the nodes are considered. The gain modularity is calculated in this way:

$$\Delta Q = \left[\frac{\sum_i n + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (2.7)$$

where $\sum_{in}$ is the sum of weights of the links inside C, $\sum_{tot}$ is the sum of weights of the links incident to nodes in C, k_i is the sum of the weights of the links incident to node i, $k_{i,in}$ is the sum of the weights of the links from

i to nodes in C and m is the sum of all weights in the network. A similar expression is used to compute the change of modularity when i is removed from the community.

In the second phase of the algorithm, a new network where nodes are the community, links between communities are the sum of weighted links from nodes belonging from one community to the others, considering also self-loops. The first phase of the algorithm is then reapplied to the new network, obtaining a lower number of communities. The steps are iterated until the number of communities is stable.

An example of how the output of the algorithm might be is showed in figure 2.1.

2.3 Topic modeling

The task of finding latent semantic structures in documents is called **topic modeling**. Those semantic structures are called **topics** while the statistical methods to find them are called **topic models**. A lot of different topic models have been proposed and many are constantly being improved [14][15]. Models in the area can be divided into two big categories: **parametric** and **non parametric**. In the former approach, the number of topics is a hyper parameter that needs to be tuned, while in the latter, no hyper parameter is needed because those methods can estimate the most probable number of topics. Currently the most widespread and used topic model is Latent Dirichlet Analysis (LDA) [16], and also some extensions such as probabilistic LDA (pLDA) are available. Another common parametric algorithm is Latent Semantic Analysis (LSA) [17][18]. The most promising non parametric topic model is Hierarchical Dirichlet Process (HDP) [19].

2.3.1 Latent Dirichlet Allocation (LDA)

The main idea of LDA is that each document is a probability distribution over latent topics, and that each topic is defined by a distribution over words [16]. LDA also makes the assumption that the similar topics will use similar words. To better understand how LDA works, let's look at figure 2.2.

The bigger rectangle, M, represents the documents, while the smaller rectangle N represents the words. The position of parameter is relevant: it means where the parameter applies whether at document level or at word level. Focusing on the parameters:

- α is the parameter of the Dirichlet prior on the per-document topic distribution. Tuning this parameter influences on the number of topics

Figure 2.2: Graphical representation of LDA model

to which a document refers to. Having a high α means having a mix of topics in the same document, while having a low α results in few topics in the same document.

- β is the parameter of the Dirichlet prior on the per-topic word distribution. High *beta* means that each topic will contain a high mixture of words, while low β indicates that each topic will contain a lower mixture of words.

- θ_m is the topic distribution for document m.

- $z_{m,n}$ is used to indicate a particular topic for the $n - th$ word in the $m - th$ document.

- $w_{m,n}$ is a particular word.

Given as input the number of topics K, the algorithm first assigns to each word in a document randomly a topic. Then for each document d, it assumes that all the topics, except the one taken in consideration, are true, calculating two proportions:

1. Proportion of words assigned to a topic t which is $p = (t|d)$

2. Proportion of assignments to topic t from the word w $p = (w|t)$

Proportions are then multiplied and the topic with the highest value is assigned to word w. The algorithm continues until no more assignments are made.

Since the first step is random, LDA is not deterministic.

2.3.2 Latent Semantic Analysis (LSA)

LSA [17] is another example of parametric topic model. It creates a word-document matrix which describes the occurrences of terms in a document. It then exploits SVD to reduce the number of rows (words) preserving the similarity between columns (documents). Columns are then compared using a cosine similarity, and very similar columns creates a topic.

2.3.3 Topic Coherence

It is quite difficult to evaluate topic models performance. **Topic coherence** are a set of metrics that try to assess the interpretability, from a human perspective, of generated topics [20].

The two most used coherence metrics [21] are the UCI measure [22] and the Umass measure [23]. Both measures are computed by summing the pairwise distributional similarity scores over words assigned to each selected topic t. The coherence of a topic T is then defined as follows:

$$coherence(t) = \sum_{(v_i,v_j) \in V} score(v_i, v_j, \epsilon) \tag{2.8}$$

Where V is the set of words assigned to topic t, ϵ is a smoothing factor and $v_i and v_j$ are words in the topic's vocabulary. Both metrics differ on how the score function is computed, in particular:

- The C_{UCI} metric defines a word pair's score to be the pointwise mutual information (PMI) between two words from an external corpus:

$$score(v_i, v_j, \epsilon) = \log \frac{p(v_i, v_j) + \epsilon}{p(v_i)p(v_j)} \tag{2.9}$$

- The C_{UMass} metric defines the score to be based on document co-occurrence:

$$score(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)} \tag{2.10}$$

  Where $D(x, y)$ counts the co-occurrences of x and y in a document, and $D(x)$ is simply the count of occurrences of a word in a document.

It is important to note that while the C_{UCI} is an **extrinsic measure**, which means it is based on co-occurrences from a different corpus such as for example Wikipedia, the C_{UMass} since it's evaluated on the co-occurrences over the original corpus it's an **intrinsic measure**.

2.4 Mann-Kendall statistical test

The Mann-Kendall statistical test [24] [25] it's a way to assent statistically if there is an upward or downward trend over time on some variable of interest. It is a non-parametric test, thus it does not make any assumptions on data distribution. The null hypothesis H_0 is that no monotonic trend is present and it is tested against the alternative hypothesis H_a that there is either an upward or downward monotonic trend. These trends can be also non linear.

2.5 Main technologies

Developing Framework

We used as developing framework Python 3.5.2 [1] due to its popularity and its richness in terms of libraries and documentation.

Beautiful Soup

We first tried to use some other tools that were reliable on scraping American news such as Newspaper3k [2] but we realized that the library failed crawling Italian news. Then, we decided to implement manually a crawler for the 20 most shared websites, using **urlrequest** to download the full html page and then parsing them using **Beautiful Soup**[3].

Postgresql

To store data collected using Hoaxy and to store the articles scraped, we relied on **postgresql**[4], an open source relational database.

Dandelion API

In order to perform sentiment analysis, we relied on Dandelion API's service[5]. They claim to perform semantic text analysis and thus they are able to provide accurate sentiment on short phrases such as reviews or tweets.

---

[1]Python's homepage.
[2]Documentation of newspaper3k.
[3]Beautiful Soup documentation.
[4]PostgreSQL.
[5]Dandelion homepage.

Wordcloud

Wordclouds is a visual representation of words, where most important words are bigger than less important ones. To make our owns, we relied on the python's module Wordcloud[6].

Gephi

To represent visually our networks, and to apply some algorithms (like k-core decomposition and community Leuvain detection), we used a tool called Gephi [7].

Networkx

The module Networkx [8], was used to build the network and to apply to it some analysis related to the centrality metrics and to the Network dismantling.

Matplotlib

Matplotlib [9] is a widely used Python 2D plotting library.

Seaborn

Searborn [10] is a Python data-visualization library built on top of Matplotlib, that has more plots types than matplotlib and that is able to make them look a bit prettier.

Botometer API

Botometer [11] is a joint project of the Network Science Institute (IUNI) and the Center for Complex Networks and Systems Research (CNetS) at Indiana University. Botometer's offer a service that tries to classify twitter accounts as automatic non-human entities (bots) with respect to real human users.

---

[6]Wordcloud documentation.
[7]Gephi's homepage.
[8]Networkx documentation.
[9]Matplotlib homepage.
[10]Seaborn homepage.
[11]Botometer homepage.

Gensim

A very useful tool for topic modelling in Python is Gensim [12], where there are already implemented and ready-to-use algoritmhs such as LDA,LSA and tf-idf.

pyLDAvis

A handy module for python's LDA exploration, is pyLDAvis [13]. It offers an intuitive and interactive interface, where the user can manually explore each topic seeing what words were more important than others and therefore better understanding what is the name of the topic.

Netvizz

In order to get data from Facebook, we exploited a popular application called Netvizz [14], which will stop working on the $4^t h$ of September, that was able to get permission from Facebook's API to automatically download content from Facebook public pages.

Mann-Kendall Test module

To perform the Mann-Kendall Test for trend, we relied on the homonim python module [15] which already implements the functions needed.

---

[12]Gensim homepage.
[13]pyLDAvis documentation.
[14]Netvizz installation page.
[15]mkt documentation.

Chapter 3

Related Work

Most of the research interest towards disinformation and fake news started in 2016 fist with the so called "Brexit" campaign and exploded with the 2016 U.S. presidential elections. In this chapter, we first give a definition of social media and what is their relation with disinformation spread. We then give a broader overview on research topics in the area of disinformation. We finally shift our focus to the 2016 U.S. Presidential election, reporting the most important contributions related to our work.

3.1   Social media and disinformation

In order to better understand what is the link between social media and disinformation, we first state what social media are. Kaplan and Haenlein [26] define them as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content (UCG)". In their work, they make a classification of social media with respect to how much they involve self-presentation or self-disclosure and on how much the media itself is rich or not. It turns out that the possibility of users to generate content is what makes social media so successful.

Types of UCG can be posts where they share opinions, images, videos and so on. One of the most important is news, either shared from some sources attaching a link or created from users themselves who report events often in real-time.

A recent report in fact, showed that nearly two Americans out of three consume news on social media [1]. In Italy the situation is similar, with 70%

---

[1]Report Pew Journalism Center.

of the population retrieving news from Internet [2] - with an increasing trend.

Getting information online purely by the feed of social media could be dangerous in terms of information quality.

Garrett [27] highlighted in fact, that if on one hand Internet exposes people to get different perspectives from their own, on the other, they are more likely to choose a news that is aligned with their opinion. Pariser [28] also argued that the two phenomena could be linked together. The diversity of point of views, in fact, would have stressed the possibility of people that share the same opinion to form **echo chambers** or **filter bubbles** where they would be isolated from contrary beliefs.

These are two of the three main factors that helped disinformation spread in social media.

Alongside these two factors (echo chambers and filter bubbles), social media allow the management of an account using bots. Humanity has been always look for the creation of an intelligent machine, able to understand and answer questions in a human natural language way. Considering one of the first example was ELIZA [29], systems such chat bots have nowadays made a lot of progress and they are able to interact with users, answering a lot of questions. Social bots (such as chat bots) can have a benign or useful purpose automating publication of contents, helping users in some tasks (augmented bots) or even providing information in real time. Some others instead, that do the same tasks, can have a malicious purpose and thus it is important to detect them [30]. They could for example, in the early moments of a post or an hashtag, share it and boost its ranking in any human social feed.

Even if the success of fake news can't be explained by social bots themselves, [31], it has been shown that they contribute with different strategies in the earlier life of disinformation articles making them reach a broader audience in a faster and deeper way [31][32].

3.2 Identifying fake news

In the literature, there are a lot of techniques to detect false news, and according to Pierri [2], the problem has been usually formulated as a binary classification task. The main research contributions can be divided in:

- **Content based.** Techniques that focus on detecting false news using the title or the body of an article.

---

[2]Report AGCOM 19/02/2018.

- **Context based.** Methods that exploit information obtained by users social interactions such as shares, likes and so on.

- **Content and context based.** Approaches that use both content and context features.

- **Network based.** This procedure is looked as the most promising one, and focuses on creating and studying diffusion cascades, assuming that false information in general spreads significantly faster, deeper, farther and broader than the truth [31].

In general, most of the datasets are built using a consolidated technique which labels disinformation based on their source [33][34][35].

3.3 Disinformation

In this section, most of the literature concerning the social impact of disinformation is presented.

We remind that **disinformation**, is any "false information that is spread deliberately to deceive people" [2].

US Politics

Since 2016 U.S. presidential elections, concern on circulation of "fake-news" has risen. Silverman [36] showed that in the days before the elections, top fake news stories outperformed top news from mainstream outlets by engagement (figure 3.1) and that most Americans believed them [37]. Some commentators even suggested that if it wasn't for disinformation, Trump wouldn't have won the elections [38]. Although a lot has been written in the news, the very first analysis that gave a theoretical and empirical background to properly frame this debate was from Alcott and Gentzkow [1]. After an important introduction on the economics of fake news, they collected false articles from fact checking portals [3] [4] and from BuzzFeed [5] and they showed that the articles collected tended to favor Donal Trump. Still, they concluded that it is hard to evaluate the effectiveness of these stories, and therefore it's impossible to determine if Trump wouldn't have won if it wasn't for fake news.

---

[3] www.snopes.com .
[4] www.politifact.com .
[5] buzzfeednews.com

Total Facebook Engagements for Top 20 Election Stories

MAINSTREAM NEWS

FAKE NEWS

8.7 million

7.3 million

15 million · 12 million · 9 million · 6 million · 3 millon

Feb.–April · May–July · Aug—Election Day

ENGAGEMENT REFERS TO THE TOTAL NUMBER OF SHARES, REACTIONS, AND COMMENTS FOR A PIECE OF CONTENT ON FACEBOOK SOURCE: FACEBOOK DATA VIA BUZZSUMO

Figure 3.1: Picture by BuzzFeedNews

The spread of fake news online

Del Vicario et al. [39] focus on how Italian **conspiratorial theories** and **scientific theories** cascades compete. By collecting data using Facebook Graph API [6] - which at that time had not the restrictions it has today - they were able to build a dataset of conspiracy theories and scientific theories spread on Facebook, relying on debunking groups. They showed that although both consumers of science news and conspiratorial theories have similar consumption patterns, their cascades - in terms of lifetime - differ greatly suggesting that science news is usually assimilated quickly and that after a while they become less and less interesting; conspiracy theories instead have a longer assimilation time but as time passes, their size increases as shown in figure 3.2.

Vosoughi et al. [31], instead of considering conspiratorial theories, focused on **rumor cascades** and showed how they compete compared to mainstream news. After characterizing their depth (from the first tweet to the last retweet), their size (the number of unique users involved in the

---

[6]Facebook Graph API documentation.

Figure 3.2: How science and conspiratorial cascades compete by DelVicario et al. [39]

rumor cascade), and their breadth (the maximum number of unique users involved in a single hop) they showed that "false news spreads farther, faster, deeper and more broadly than the truth". They also firstly assessed the impact of social bots in disinformation spread, showing that the veracity of disinformation is mainly due to humans.

Fake news on Twitter during the 2016 U.S. presidential election

More recently, Grinberg et al. [33], associated public voter registration records to Twitter accounts, and labelling sources constructed by fact checkers, journalist and academia were able to describe both the relevance in terms of volume of disinformation compared to mainstream outlets and typical political affiliation of users sharing misinformation. They estimated the composition of each person's news feed by randomly sampling tweets posted by their followees, calling them "exposures". They showed that 5.0% of exposures to political URLs were from fake news sources. Still, 5.0% of disinformation outlets was responsible to 50% of the volume of fake content exposures and that 1.0% of individuals accounted for 80% of the false content exposures. They concluded that a conservative, older person is more likely to be engaged in fake political news.

Hoaxy

Shao et al. introduced **Hoaxy**[35], a platform to collect online disinformation and its related fact-checking efforts [7]. The platform uses different technologies (such as social media api, crawlers and RSS) to monitor news from social medias and from chosen websites which usually spread misinformation and well known fact-checking websites such as snopes.com, politifact.com, factcheck.org and so on. Their main analysis was conducted on Twitter, collecting tweets that contained URLs belonging to the early mentioned lists. From the analysis they showed that the volume of fake news is 10 times bigger than the volume of fact checkers, and that exists a lag of 10-20 hours from fake news to fact checking sharing.

Another work from Shao et al.[32], using data collected from Hoaxy from mid-May 2016 to March 2017, tried to infer the role of bots in low-credibility content spread. Using Botometer[40], they assigned a score to users: high scores are users that are more likely to be bots, low scores are users that are less likely to be bots. They identified different strategies bot used to spread news:

1. Being very active in the first seconds after the publishing of an article. They made the conjecture that this helps exposing more users, increasing the chances that an article will become viral.

2. Mentioning very popular accounts and attaching a fake content in the tweet. They thought this is done to help the credibility of the news shared.

They finally concluded that few accounts were responsible for a large share of misinformation. These accounts were shown to be likely bots and that their contribution was relevant in the early moments of the false news life. Their analysis complemented the work done by [31], arguing that surely bots alone didn't entirely explained success of fake news, but also humans were engaged in the sharing of fake news.

Hui et al. [41] introduced a new way to analyze data coming from Hoaxy, considering the diffusion Network. In the network, two nodes are linked if one node retweeted the other, the edge is directed and if follows the flow of information (from retweeted to retweeting) and is weighted by the number of retweets. Using a majority rule, they then categorized "claim" edges (those edges that are linked where sharing news from a misinformation source), and "facts" edges (those edges coming from fact-checking websites).

---

[7]Hoaxy's platform

Figure 3.3: Main core of hoaxy

In another work Shao et al. [42], using data collected from their plat-
form Hoaxy, compared misinformation spread with fact-checking, highlight-
ing what were the characteristics of the core of their diffusion network and
that it is possible to reduce the overall number of misinformation. They
showed that the fact checkers accounted for only 5.8% of the total tweet
volume and that when analysing the main k-core fact checkers almost dis-
appear. Then, they used different centrality metrics to identify the top-10
users for each centrality and manually inferred their partisanship, if they
were humans or bot, if the account was a personal account or belonged to
an organization. Lastly, they tried to curb disinformation spread, evaluat-
ing the robustness of the network. They showed that the efficiency of the
network is greatly impaired by disconnecting only 10 accounts.

3.4   How to curb disinformation spread

In their conclusion, Shao et al. [42], using a network dismantling technique
showed the possibility of effectively disrupt the spread of false-news. Accord-
ing to the authors, social media shouldn't fully automate this process, but
using it as a tool to review more efficiently users, in order to find accounts
that does not follow platforms' terms of service. In another labour, Lazer
et al. [43], points out the necessity of a multidisciplinary effort to address
fake news. Looking at the past, U.S. news media failure of the 20th century
allowed the rise of journalistic practices and norms that "served well" creat-

ing an overall credible information system. They pointed two main paths to stem disinformation. Acknowledging that sometimes fact checking can be counterproductive, the first is then empower people through education by injecting into primary and secondary schools critical-information skills. The second path that has to be followed according to them, is to involve social media companies into actively take care of malicious accounts or content. They suggested for example to promote quality information, or to provide customers with signals of the quality of the information they consume.

3.5   European context

Europe itself is not spared from disinformation at all. There is in fact a lot of material coming from the COMPROP group at Oxford's university in the form of data memos [8]. Their methodology is in general to select a set of hashtags related to the current election (such as main candidates, the name of the elections, or main parties), download tweets containing the selected hashtags and evaluate how much each one of them was discussed. They then consider tweets containing URLs, and they manually labelled most of the sources. Label used are categorizied under "big" sections such as "Professional News Content", "Professional Political Content", "Divisive & Cospiratorial Content", "Other Political News & Information", "Other" - which are social medias or websites where you can buy items and any sources not belonging to the other labels. In their last work [44], they provided an overview of their findings of an analysis conducted from 5 April to 20 April 2019 (the methodology where they describe hashtags and source's categories can be foundhere), for the countries: Italy, Germany, France, Poland, Spain and Sweden. They showed that less than 4% of sources circulating on Twitter were "Junk News" or "Russia" (labels inside the category "Divisive & Cospiratorial Content), with users sharing overall far more links to mainstream outlets (34%) and that on Facebook, a single junk news story "can hugely outperform even the best, most important professionally produced stories, drawing" four times its share. They also noted that the most shared junk news stories are about **populist themes** such as anti-immigration and Islamophobic sentiment, with a few of them also promoting Eurosceptiscism or mentioning directly European leaders. In Italy, as showed in figure 3.4, it is interesting to note that the ratio between "Junk News"(JN) and "Professional News" (PN) is very high - second only to Poland - namely 38.9% of sources shared were "PN" and 8.7% "JN".

---

[8]COMPROP homepage.

IN EACH LANGUAGE SPHERE

◯ Junk News
◯ Professional News

Figure 3.4: Ratio between Junk news and professional news in Europe, courtesy of [44]

3.6 Italian disinformation

So far to our knowledge, there has been little work in the literature that specifically address the Italian disinformation context. A recent work from EU Disinformation lab [45], monitors 171 stories that have been considered misleading, satirical, misleading or conspiratorial and evaluates how they propagate on twitter.

In the context of Italian Political Elections in 2018 instead, and in particular in the new media political coverage [46], the disinformation is treated "collaterally" as a matter of fact, highlighting its presence, but not investigating its sources. Anyway, it has been pointed out that "explicitly partisan and hyper-partisan sources catalyzed a significant share of the social media interactions performed by the online audience in the lead up of the election", polarizing the discussion towards controversial topics such as immigration and the lack of an effective welfare.

3.7 Italian political context

After the Political elections held the 4th of March 2018, a new landscape for Italian politics was born: the two populist forces ("Movimento 5 Stelle" and "Lega Nord"), boosted their presence in the Italian Parliament and after a couple of months they decided to create a new government based on a contract, naming Giuseppe Conte as Prime Minister, "lawyers of italians". They shared a lot of ideas in common. For example, they presented themselves as new, anti-establishment forces, they claimed the necessity of a "rediscussion" of Treaties with the European Union, and together they called for a less stricter regulations on vaccines. Where they disagreed in some positions, they were able to make a deal. The strength of the government reflected on social media: both leaders were very active during and after the elections. In particular Matteo Salvini, more than any other politician in the world, has been testing the relationship between clicks and consensus using social media [9] to bring his supporters closer to him, making them feeling as if *he is one of them*, taking selfies with them, posting on socials photos of his meals or of him at the beach. His language has humanized far-right rhetoric and made public discourse rougher and more aggressive, increasing also the hate speech via social [10] After 15 months as the Vice Prime Minister of the government, Salvini's relentless campaigning — while Interior Minister — helped his party double its standing in opinion polls, clearly outpacing any other party. The "Lega" in fact, placed first at the Italian European Election in May, with 34% of votes. In parallel, "Movimento 5 Stelle" struggled to maintain their consensus and some of their voters became supporters of "Lega". Even if their network of interactions is overall bigger than the one of "Lega", the popularity of Matteo Salvini's social pages overtakes by a lot the pages of the most influential leaders of "M5S". The oppositions, in particular "Partito Democratico" saw primaries in March, electing a new Secretary with the hard task to reorganize the party against the populist. "Forza Italia", a center-right party, is in crisis. After losing to Lega in 2018, it lost even more consensus after the European Election and a lot of deputies moved to "Lega". Who is gaining more and more consensus is "Giorgia Meloni", leader of "Fratelli D'Italia", a right-party, also thanks to her usage of social media. We observe that, even if there are some similarities between Trump and Matteo Salvini, the Italian political context is very different from the U.S.

---

[9] The New Populist Playbook

[10] Hate speech: si confermano rom e immigrazione i temi che generano maggiori polemiche.

3.8 Position of this work

The aim of this work is to replicate most of the analysis made in U.S. and across other countries, in Italy during the European Electoral campaign and, given the two different political context, compare the two results. In particular we first address the disinformation reach on Twitter, and how active the community of users that shared disinformation was. We next focus on the narratives of the debate, looking if there was any major topic. We then focus on the disinformation network: trying to look if there were some particular users responsible for most of the disinformation, and describing who these users are. We finally conclude investigating the presence of a coordinated network across different countries.

Chapter 4

Data Collection

In this chapter we present how the data has been collected and a brief description of the dataset. The approach used to gather the data is similar to the one used in Hoaxy [35]. Leveraging on Twitter Streaming API, we collected tweets containing an explicit Uniform Resource Locator(URL) associated with news articles shared on a set of Italian disinformation websites list we created. We considered the period before the Italian European Election, starting our analysis on the 7th of January and using data until the 26th of May, the day for the European Elections in Italy. In the chapter we first describe in more details what were the steps to create our data collection and in particular the list of fact checking and claim website monitored and how the data has been collected. We next report the entities collected, focusing on tweets, users and articles. We finally conclude with a comparison on how two sources are active on twitter and on Facebook.

4.1 Dataset creation

In order to create the dataset we followed these steps:

1. We built a list of fact checking and claim websites.

2. We gathered tweets with URLs belonging to these lists.

3. We identified unique articles.

In the following sections, each step will be described.

4.1.1 Claim and fact checking websites

We first created a list of websites that typically share deceptive content, relying on famous Italian debunking websites, namely **"pagellapolitica.it"**,

"**butac.it**", "**bufale.net**" and "**lavoce.info**", which have blacklists where they collect unreliable domains of portals that typically share hoaxes or fake news. We call the list of unreliable domains **claim websites** because they usually share **claims**, or disinformation; the second list of the aforementioned **debunking** websites is called **fact-checking**. Table 4.1 reports all the websites we monitored. Label "C" stands for claim websites, "F" for fact checking websites.

Figure 4.1: Weekly time series of the data collected with highlights of main events covered by disinformation.

Since some claim websites during the observation period became inactive, and new ones appeared, we had to keep the lists updated and check every now and then the activity of the websites. Table 4.1 shows also when a certain website was added. We remark that - compared to fact-checkers (that are 4) - claim websites are more numerous (62). This is probably due to the fact that is very hard to verify a news completely, grounding and sourcing every statement. Instead, it's very easy to create a website out of the blue and write totally fake articles as reported from Allcott [1].

4.1.2   Twitter API

To collect tweets, we relied on the Twitter API, and in particular on the module **tweepy** for python. As suggested by the Twitter API[1] itself, in order to find useful tweets for our analysis, we used only the domains as a query of

---

[1]Twitter Developer Guide.

Website domain	Type	Date	Website	Type	Date
accademiadellaliberta.blogspot.com	C	15/3	altrarealta.blogspot.com	C	15/3
aurorasito.wordpress.com	C	15/3	compressamente.blogspot.com	C	15/3
freeondarevolution.wordpress.com	C	15/3	ilsapereepotere2.blogspot.com	C	15/3
laveritadininconaco.altervista.org	C	15/3	madreterra.myblog.it	C	15/3
neovitruvian.wordpress.com	C	15/3	olivieromannucci.blogspot.com	C	15/3
pianetax.wordpress.com	C	15/3	terrarealtime.blogspot.com	C	15/3
corrieredelcorsaro.it	C	15/3	siamonapoletani.org	C	15/3
ilfattoquotidaino.it	C	15/3	conoscenzealconfine.it	C	15/3
5stellenews.com	C	7/1	breaknotizie.com	C	7/1
bufale.net	**F**	7/1	butac.it	**F**	7/1
disinformazione.it	C	15/3	ecplanet.org	C	15/3
filosofiaelogos.it	C	15/3	hackthematrix.it	C	15/3
ilpuntosulmistero.it	C	15/3	libreidee.org	C	15/3
liberamenteservo.com	C	15/3	nibiru2012.it	C	15/3
pandoratv.it	C	15/3	tmcrew.org	C	15/3
tankerenemy.com	C	15/3	ununiverso.it	C	15/3
ilprimatonazionale.it	C	7/5	skytg24news.it	C	7/5
daily-screen.com	C	9/1	adessobasta.org	C	9/1
byoblu.com	C	7/1	catenaumana.it	C	9/1
comedonchisciotte.org	C	7/1	direttanews24.com	C	7/1
essere-informati.it	C	7/1	il-giornale.info	C	7/1
ilmessangero.it	C	9/1	il-quotidiano.info	C	7/1
ilvostropensiero.it	C	9/1	informarexresistere.fr	C	7/1
informazionelibera.eu	C	7/1	interagisco.net	C	7/1
italianosveglia.com	C	7/1	jedanews.it	C	7/1
lavoce.info	**F**	7/1	lettoquotidiano.it	C	9/1
lonesto.it	C	7/1	mag24.es	C	9/1
mondodiannunci.com	C	9/1	notiziarioromacapitale.blogspot.com	C	9/1
pagellapolitica.it	**F**	7/1	saper-link-news.com	C	17/1
silenziefalsita.it	C	17/1	skynew.it	C	9/1
sostenitori.info	C	17/1	tg24-ore.com	C	17/1
tuttiicriminidegliimmigrati.com	C	17/1	voxnews.info	C	17/1
webtg24.com	C	17/1	zapping2017.myblog.it	C	9/1

Table 4.1: List of 'C' domains and 'F' domains, namely "claim" and "fact checkers". "Date" indicates when the domain was added in our analysis.

the service, dropping any "www", "https" and so on. The final query looked like: "`skynew24 com OR byoblu com OR voxnews info OR ...`". Knowing that some twitter accounts have the same of the website we were monitoring, for example byoblu[2], we also collected some tweets that mentioned those accounts, that were sharing a URL (not necessarily coming from the sources we were observing). We called the sources of those tweets "No_site" and we later decided to remove them from our analysis. In the next sections, we will consider the dataset as a whole first and then we'll neglect debunking and "No_site" sources, focusing on the data set containing only claim tweets. We report in figure 4.1 the time series corresponding to the weekly volume of tweets we collected and the main events corresponding to the peaks we observed in our time-series. There are also some negative peaks, where the machine we used to collect tweets crashed.

4.1.3   Url canonicalization

In the previous work [35], they used url canonicalization to handle duplicate articles. We preferred to take a different approach. Using crawlers we built for most of the websites ($\sim$20), we crawled article's title and content and used those two elements and the site name as primary keys to find different urls that were mentioning the same articles.

The difference between their approach and ours is that they [35] take into account modifications of the same articles, continuously crawling the disinformation outlets and thus they might also discard some articles that have been deleted. We decided to discard url canonicalization since we considered our method more reliable. In fact, we first downloaded the full html page of the url that appears in the tweet using **urlopen**, in order to obtain a page of the article if that page is removed (which happens very often). When we finally finished to collect all our tweets (after the election day), we crawled again the title and the content of the articles, discarding articles that had 0-length content and keeping the old ones. In this way we ensured the collection of consistent articles, updated in case of modifications and present in case of deletion, with still the possibility to discriminate them using titles and content as keys.

The crawlers were built using first the module **urlopen** to download the html page and then **BeautifulSoap** to parse the html. Most of the websites were different and in order to retrieve the correct title and content, ad-hoc functions were implemented.

---

[2]Byoblu's twitter account.

Tweet type	Number of elements
Tweet	103401
Retweet	265088
Quote	61996
Reply	7198
Mention	317718

Table 4.2: Different types of tweets for the whole dataset.

4.2 Description of the dataset

In this section, we describe the entities of our dataset, giving an idea of the size of our problem.

The total number of tweets we collected was **437683**, ∼**23000** related to fact check and ∼**414000** related to claim websites.

4.2.1 Tweets

Up until now we used the term *tweet* considering all the types of tweets. Actually a *tweet* can indeed be a tweet, but also a reply, a mention or a retweet:

- A **retweet** it's a repost or a forward of a post of another user called **tweet**.

- A **quote** is a retweet with additional text from the user.

- A **reply** it's when a user directly comment a tweet of another user.

- A **mention** it's a particular tweet where the owner of the tweet mentions another user by adding a @ before the user's screen name.

In table 4.2 we show how the number of tweets, mentions, retweets and replies are distributed in our dataset.

Similar to other findings [42], we can see that fact-checking tweets accounts for only the 5.5% of our data. We decided to consider in our analysis only the "claim" websites, dropping the "fact" tweets and to keep only tweets with url pointing to our disinformation websites, therefore dropping the "No_site" sources.

4.2.2   Users

A total of $\sim$**35000** users were actively involved in our analysis. Dropping
"No_site" tweets, $\sim$**23000** user shared news actively from claim websites,
$\sim$**10000** from fact checking websites and $\sim$**2000** from both claim and fact
checking websites.

Using Twitter API, and in particular the **tweepy** module for python,
we retrieved the "Twitter Profile" object and we analyzed the creation date,
more details are available in section 5.3.

Figure 4.2: User features description

Figure 4.3: Boxplot of users followers

Figure 4.4: Boxplot of users friends

Figure 4.5: Boxplot of users statuses

Figure 4.6: Boxplot of profiles lifetime

From figure 4.2 we can notice that the average user has created it's profile
more than 5 years ago, did around 800 statuses, follows ~ 500 users and is
followed by ~ 200 users.

We overall report **21124** active accounts, **20** of which are also verified,
800 deleted, **124** protected and **112** suspended accounts. Verified accounts
("accounts that are of public interests" [3] such vips,artists, politicians...)
were altogether involved in **5761** tweets, only **18** of which in an "active"
way, i.e. a verified account actually authored the tweet.

---

[3]Twitter verified accounts description.

4.2.3 Articles

We collected a total of **27740** articles, ∼**24000** of claims and less then ∼**4000** of fact checking. As mentioned before, one problem we had to handle was article duplicates. One user could share an article from the sources itself having then a so called "canonical url", or use for example some other redirect links having different urls referring to the same articles. For these reasons, Shao [42] used url canonization. Since we believe url canonization was unreliable for cases in which the url itself could be packed using url shortener tools, we decided to follow a different approach and to crawl only once articles with the same url (so if the article was edited we don't have any different text in our database) and to use as key of our database articles' text, title and website. In this way we were able to find different urls that pointed to the same article having a total of ∼17000 unique claim articles. In some cases, user mentioned only the website (so instead of an article they were mentioning the homepage, ex. "byoblu.com"). We are not considering these on our articles descriptive statistics, but we kept the tweets for other analysis. In chapter 6 we will further characterize the topic and the sources of these articles.

Figure 4.7: Shares comparison between Twitter(red) and Facebook(blue) for two disinformation outlets, namely "byoblu"(left) and "silenziefalsita.it"(right).

4.3   Comparison with other social media

According to Reuters [4], twitter is the 6th social media used for sharing news in Italy. For this reason, we tried to evaluate the different reach of disinformation for both Twitter and Facebook, collecting data regarding two disinformation outlets , *"byoblu.com" and "silenziefalsita.it"*, which have an associated page, and that belong to the top 3 most shared and engaged misinformation outlets.

We relied on **netvizz** to gather statistics on Facebook's daily shares of posts of the two aforementioned outlets. We then compared it with the traffic we observed from Twitter. We found that there is a huge discrepancy: figure 4.7 shows that the number of shares of Twitter(red) and Facebook(blue) for two disinformation outlets, namely "byoblu"(left) and "silenziefalsita.it"(right), between 07/01/2019 and 27/05/201, is hugely in favor of Facebook. These results are consistent with the Reuters report and with other findings, suggesting that most of the Italian disinformation is spread mainly on Facebook.

---

[4]Reuters Institute Digital News Report 2019.

Chapter 5

Disinformation reach and users activity

In this chapter we answer to **RQ1**: What was the reach of disinformation which circulated on Twitter in the run-up to European Parliament elections? How active and strong was the community of users sharing disinformation? To do that, we first give a description of the activity of the content published from our disinformation sources, in terms of engagement and in terms of number of shared tweets. We then look to see if we can find a trend of news consumption towards the European election and we finally evaluate the activity of the users community sharing disinformation.

5.1 Sources of disinformation

To better understand the reach of disinformation, we first computed the distribution of the number of articles and tweets per sources.

From figure 5.1 we see that the most active websites, in terms of production of articles shared, are **voxnews.info**, **tuttiicriminidegliimmigrati.com** and **skynew.it**. But some articles have been shared more than others, and if we look at figure 5.2 we can see that **voxnews.info** still remains the most shared, but it is followed by **ilprimatonazionale.it** and **byoblu.com**.

Voxnews.info, with $\sim 200k$ tweets (over 50% than the total volume) and $6k$ unique articles (1/3 of the total number) is the most active disinformation website in our database. It spreads disinformation on different topics: from immigration to health-care and conspiratorial theories. We were not able to to understand who is behind the website. They also have a fictitious *fact checking* service, to deceive the unaware reader, quoting

Figure 5.1: Distribution of the total number of shared articles

Figure 5.2: Distribution of the total number of shared tweets

very often misinformation sources and mainstream outlets. In the latter case, they usually manipulate the mainstream content by just taking a little portion of what is written decontextualizing it.

Skynew.it was very active at the beginning of our monitoring, but then stopped in the middle of March. Nevertheless, new websites with similar names and literally the same webpages (layout, style, font) appeared and also stopped publishing articles after a while - namely skytg24news.it,

Title	Shares
1 Blitz dei Ros: certificati falsi agli stranieri per farli votare PD	**2558**
2 Matteo Salvini, frena tutti: "Non cambio idea sui migranti, i porti restano chiusi"	**1845**
3 Più della metà degli account Twitter che seguono Boldrini e Saviano sono falsi	**1740**
4 Modena, nordafricano dà fuoco a sede vigili: due morti e tre feriti gravi	**1487**
5 CI STANNO DISARTICOLANDO, SARÀ UNA GUERRA LUNGA E DURA	**1303**
6 "Ha favorito le banche": Finalmente indagato Mario Monti, il vero traditore...	**1213**
7 Il Piano B per un'Italia fuori dall'Euro, in diretta streaming su Byoblu. Diffondi il contagio.	**1198**
8 ORA E' UFFICIALE, IL NAUFRAGIO DEI 117 MORTI ERA UNA BUFALA	**1143**
9 Macron: "L' Italia si merita tutti gli immigrati, perchè era un Paese fascista"	**1108**
10 Adam confessa: "Non è stato Ramy a telefonare, Niccolò il vero eroe"	**981**

Table 5.1: Top ten most shared articles in our dataset.

tg24.skynew.it and catenaumana.it. These websites were linked to a twitter account called "CatenaUmana".

Tuttiicriminidegliimmigrati.com instead claims to be a news aggregator, which, quoting from their website, collects 40% of the news automatically using a software, 30% using volunteers and the remaining from readers and from data shared from the police. Of course, most of the news are inaccurate or completely made up so we think this is a "nice" (actually terrible) example of propaganda anti-immigration.

Ilprimatonazionale.it is a websites that call itself *quotidiano sovranista* ("a sovranist newspaper"), close to "Casa pound", a (former) far-right italian party.

Byoblu.com is Claudio Messora's website, an "independent journalist that spread independent information" who has also been the communication advisor for "Movimento 5 Stelle" ("5 stars movement").

Most shared articles

We in particular digged into the ten most shared articles, shown in Table 5.1, revealing that some of them are completely false (6, 8, 9, 10), some are misleading (1, 3), propaganda (2), and one is even true (4).

The most shared article (1) for example, it's taken from an old news published by Salerno today [1] in 2015, and it's meant to mislead the reader of the tweet that does not open the article into believing that either PD let immigrants vote in their last primary elections - held the 3rd of March - or into the EU elections.

The second most shared article (2) is actually a "True" article. As stated in sections 9.2 in fact , one of the main limits of this analysis is the

---

[1]Eboli, falsi certificati agli stranieri per farli votare al Pd: blitz dei Ros.

assumption that all claim articles are false. We also report that this article was shared by **matteosalvinimi** - the official account of the Leader of Lega party - and this is probably the reason it had so many shares. We remark anyway that, the outlet still typically share disinformation, as reported by **pagellapolitica**, and that inside the body there are grammatical errors (for example repetitions of "merito" or the incorrect use of a verb "invitare" .. "di" instead "invitare".. "a") and that a particular polarized language is used ("buonista"), showing that the article still belongs **propaganda**, as defined in the introduction 1.2.

For what concerns the third (article 3), it is partially true: in fact the system of **Audit**[2] used to evaluate the users is considered reliable, but it points out the lack of activity of the followers and not them being a bot; so the news would be that most of the accounts that follow Boldrini and Saviano are not very active. We also point out that the article has a partisanship in the language and its political targets.

As pointed out in some debunking articles instead, news number 6 is totally fake [3], as it is fake news number 8 [4], number 9 [5] and number 10 [6].

The last two articles are from byoblu (5, 7) and they long videos sharing euro-sceptical content. We already stated that byoblu is a well-known disinformation spreader in the previous section.

Tweets per week

In figure 5.3 we show the most active websites per week. From the figure we see again the huge activity of the already mentioned voxnews.info, ilprimatonazionale.it and byoblu.com. We also highlight the even if we started collecting tweets from ilprimatonazionale.it during the second week of May, it is very active and has a comparable volume of tweets to voxnews.

5.2   Daily engagement

We also checked the **engagement** of each websites per week, defining engagement as the ratio between the number of shares and the number of articles published per day:

---

[2]Audit's homepage.

[3]No, Mario Monti non è indagato per aver favorito le banche.

[4]≪Il naufragio dei 117 morti è una bufala≫. Macché, è questa la bufala!

[5]Notizia falsa: Macron: "L'Italia merita tutti i migranti perché era un Paese fascista".

[6]La disinformazione e la verità sulla telefonata al 112 di Adam e Rami.

Figure 5.3: Sources share of tweets per week

$$engagement = \frac{number\,of\,shares}{number\,of\,articles} \qquad (5.1)$$

From the figure 5.4 we can see that "byoblu.com" has remarkable spikes of engagement, with respect to a small number of total tweets compared to the other outlets, whereas **mag24.es** increased its activity during May. This might be for two different reasons:

1. With the approach of the elections, they increased their activity successfully trying to share propaganda.

2. Their Facebook pages were closed thanks to Avaaz Report [7].

With the approach of the election, we were interested to understand if there were some particular trends. For this analysis we didn't consider "ilprimatonazionale.it" as it was added in the last may of data collection. Focusing on the other top 10 sources, we performed a Mann-Kendall test to identify the presence or absence of a monotonic upward or downward trend in the time-series of daily shared tweets and of daily engagement. Taking in consideration the Bonferroni's correction, the test was rejected at $\alpha = 0.005$. For both daily shared tweet and daily engagement, we found an upward trend only for "byoblu.com", whereas the remaining disinformation

---

[7]The report where Avaaz successfully tracked a lot of different disinformation sources and managed to make Facebook close those pages can be found here.

Figure 5.4: Daily engagement per top-10 websites.

sources are either stationary or monotonically decreasing. Since this outlet strongly supports euro-skeptical positions (both towards Euro and EU), we suggest that in the run-up to the European elections its agenda became slightly more captivating for its social audience.

5.3   User investigation

As already mentioned in section 4.2.2, we leveraged on Twitter API, and in particular on the **tweepy** module for python, to retrieve the "Twitter Profile" object. We used as query the *"screen_name"* of each user that appeared in our collection. We then analyzed their activity, their types and some features related to their profile.

User analysis

In order to model the activity of the users, we first computed the distribution of number of shared tweets per number of users, observing that a restricted amount of users were responsible for the spreading of most of the disinformation we collected.

In fact, 20% ($\sim 4k$) of our users accounted for less than 90% ($\sim 330k$) of the total tweets as shown in figure 5.5. This is similar to other findings [33][42][47]. Among these users, by manually looking into their profile's description and screen name, we identified accounts officially associated to

Figure 5.5: Users tweets volume cumulative distribution

18 different outlets that overall shared alone 8310 tweets.

We also decided to distinguish users into five different classes of activity i.e. the number of actively shared tweets containing a URL to claim websites:

- **Rare**. Users that shared only one tweet. ($\sim 9.5k$ users)

- **Low**. Users that shared more than 1 tweet, but less than 10 ($\sim 8k$ users).

- **Medium**. Users that shared more than 10 tweets, but less than 100 ($\sim 3k$ users).

- **High**. Users that shared more than 100 tweets, but less than 1000 (~ 500 users).

- **Extreme**. Users that shared more than 1000 tweets (20 users).

From these classes, we can notice that one users over five shared more than 10 claim articles in five months. In figure 5.6, we can see the total volume of tweets according to activity of the users. In the inner cake, lighter colors show users that were created within 6 months from the election day. These accounts respectively for the 0.18% (Rare), 0.6% (Low), 2.04% (Medium) and 2.98% (High) of total tweets. We can note that "High" and "Extreme" users account for half of the misinformation tweets.

Account types

As already stated in section 4.2.2, we overall report **21124** active (20 of which are also verified), **800** deleted, **124** protected and **112** suspended

Medium (34.93%)

Low (10.97%)

Rare (3.44%)

Extreme (7.56%)

High (43.11%)

Figure 5.6: Total volume of tweets according to activity of the users.

accounts. Verified accounts were altogether involved in 5761 tweets, only 18 of which in an "active" way, i.e. a verified account actually authored the tweet. We observed that they were mostly called in with the intent to mislead their followers, adding deceptive content on top of quoted statuses or replies.

Retweeting rate

We follow our analysis looking at the distribution of the numbers of users with respect to their re-tweeting activity defined as:

$$Retweeting(user) = \frac{NumRetweets(user)}{NumInteractions(user)} \tag{5.2}$$

where the NumRetweets(user) is the number of retweets of a certain user in our collection, and the NumInteractions(user) is the number of all shares of the same user in our dataset.

We noticed from figure 5.7 that the Retweeting rate is strongly bimodal, and users that typically share disinformation are mostly "retweeters". More than 60% of the accounts show a re-tweeting activity larger than 0.95% while less than 30% have a re-tweeting activity smaller than 0.05%. This is similar to other findings and suggests that few accounts are responsible for conveying in the first place disinformation to Twitter, and then the rest of the community propagates it.

Figure 5.7: Retweeting rate

User profile features distribution

Next, we focused on the feature distribution obtained using the Twitter API, namely the number of followers, the number of friends (aka the users an account follows), the length of their timeline (aka the number of tweets, retweets, quotes or replies a user did in its lifetime) and the creation date as shown in figure 4.2. We also computed, from the creation date, the age of the account in months with respect to the European election day, as shown in figure 5.8.

Figure 5.8: Violin plot with the average lifetime of accounts in months

We observed two things: first that in general profiles that share misin-

formation are quite **"old"** with an average of 5 years (figure 4.6) with more than one thousand authored statuses; second that in the last 20 months a lot of **new accounts** were created (figure 5.8), in particular almost a thousand registered during the collection period. Compared to the others, those recently created users show similar features distributions.

Looking again at figure 5.6, we can see that most of the recent account are either medium or highly active. Interestingly, some of them show abnormal activities having produced more than $10k$ tweets (in general, not coming from our collection) in their small lifetime.

Focusing on these users, we performed a Mann-Kendall test to the time series of daily tweets shared by such users (see 5.9), assessing the presence of a monotonically increasing trend (at a significance level $\alpha = 0.05$). The main referenced source of disinformation is "voxnews.info" with more than 60% ($\sim 12k$ tweets) of the total number of shared stories.

Figure 5.9: Daily time series of tweets created by recent users

Such activity is quite suspicious and suggests the presence of "cyber-troops" (bots, cyborgs or trolls) that either attempted to drive public opinion in light of up-coming elections via so-called "astroturfing" campaigns [48], or simply redirected traffic as to generate online revenue through advertising [1] and should be further investigated.

Chapter 6

Disinformation Narratives

This chapter gives an answer to the second research question, **RQ2: What were the most debated narratives of disinformation? How much were they influenced by national vs European-scale topics?**

Before answering the research question, we performed a data exploration of our articles in order to get a more precise idea of what we should find from our analysis. We already showed what were the most popular articles in Table 5.1, our focus is now on topics and words. We tackled the problem of identifying the most debated narratives using firstly LDA and LSA, which are well known topic modelling techniques. Unsatisfied by the results, we opted for a keyword approach, selecting them among the most used words in articles. We also considered first tweet's hashtags, looking for particular ones that supported a specific parties, and then profile hashtags, retrieved from profiles of users belonging to our dataset, to finally identify any political affiliation of the users in our dataset. We conclude the chapter with a section that assesses the hate speech of the debate.

6.1   Description of the data

If the chapter will mostly perform topic analysis, in this section we first filtered the stopwords (like articles, conjunctions..) as they don't add particular meaning to a sentence. We next looked at what were the most used words in our dataset's article, by weighting each word for the number of times the article was shared, and finally we created the word clouds shown in figure 6.1. On the left (figure 6.1a), it is showed the word cloud of the titles. It is interesting to observe news titles more than news bodies, because they are what a user reads first when he surfs on the web, and they are able to catch its attention or even deceive its perceptions. In this case, since

(a) Wordcloud of article titles   (b) Wordcloud of article bodies

Figure 6.1: Wordcloud of articles

they're twitter cascades, the article is even shared and debated. As we can already see the most frequent keyword was "video", proving that a remarkable fraction of disinformation was shared as multimedia content. The next one is "Salvini" and then "immigrati" (which means immigrants) showing that the politics of the government and the immigrant crisis were the most used words to catch users attention. Looking then at the word cloud of the article bodies 6.1b we can see that the focus shifts more on italy and Italians (so to "sovranism"). In the next sections we will give a more grounded base to our statements, showing first the preprocessing techniques used, and then showing the results for each method chosen.

6.2 Preprocessing

In order to execute both LDA and LSA algorithms, we cleaned our data doing some preprocessing: we first removed all the punctuation (",",".",...), we next lowered all the words and finally used TreeTagger to do POS-tagging and lemmatization of each word. We also tested each model with and without *tf-idf*, or using a certain parts of speech with respect to others. Results will be discussed in the next sessions.

6.3 LDA

We used LDA [16] on articles, trying to find what were the topics of the discussions for disinformation. Since LDA needs in input the number of topics, we used the Coherence metric C_v discussed in chapter 2.3.3 to evaluate what could be the best number of topics.

We tested the model using different parameters: either the text or the

titles of the articles, either the "nouns" or the "verbs" or both or all words, either with tf-idf or without.

In general, according to the coherence metric considered, using only "nouns" was the best choice because "nouns" embed most of the meaning; words appearing in titles resulted in a better coherence score and *tf-idf* did not improve the understandability nor the coherence score of the topics relevantly.

Figure 6.2: Plot of the C_v measure respect to the number of topics for LDA(blue) and LSA(red).

From figure 6.2 we can see that the best number of topic is number 10: a bigger number would result in a lot of words in different topics (therefore hard to interpret) and a smaller would have a lower topic coherence. Still, exploiting the **pyVis** module [49], we were able to visualize each topic's word and to tune the relevance σ of each word with respect to the topic. We realized that the "topics" didn't make sense at all: most of the words were in fact synonyms put in different topics: for example immigration-related words (such as "profugo", "immigrato", "clandestino", "immigrazione" ...) appeared most of the topics found.

Instead of using the topic coherence, we decided to tune manually the number of topics, exploiting the visualization tool and tested all the possible number of topics (from 2 to 15) until we were satisfied.Exploring all the models for different number of topics, we still were not satisfied at all by the topics we found because they either contained synonyms, overlapping each other, or they were too specific, reporting news instead of a topic. We overall had difficulties to understand properly what kind of topic those words were

referring to.

We then decided to try other approaches.

6.4   LSA

The next algorithm used was LSA [17]. We performed the same preprocessing we did with LDA and explained in section 6.2. We then evaluated again our model using the coherence score. We found similar results compared to LDA (as shown in figure 6.2), with LSA having a better coherence with few topics, and LDA with a lot of topics. Since topics did not make sense again, we decided to move to a keyword-based technique.

6.5   Keyword approach

We then decided to change again our approach using a keyword one.

By considering the most 500 frequent words (some are shown in figure 6.1) that appeared in article titles and taking into account the most relevant events occurred in the observation period, we employed two volunteers to manually compile a list of keywords associated to manually compile a list of keywords associated with the following four topics: Politics/Government (PG), Immigration/Refugees (IR), Crime/Society (CS), Europe/Foreign (EF). The list is presented in table 6.1.

In particular, PG refers to main political parties and state government (e.g. "salvini", "pd", "m5s", "conte") as well as the main political themes of debate (e.g. "reddito", "banche", "decreto"). IR includes references to immigration, refugees and hospitality (e.g. "immigrati", "profughi", "ong", "porti") whereas CS includes terms mostly referring to crime, minorities and national security (e.g. "rom", "mafia", "polizia"). Finally EF contains direct references to European elections and foreign countries (e.g. "euro", "macron", "onu"). It is worth mentioning that the most frequent keyword was again "video" (which does not belong to any specific topics), proving that a remarkable fraction of disinformation was shared as multimedia content. Any articles that did not contain any keywords from the categories just described, was considered belonging to "OTHER".

We then computed the relative presence of each topic in each article, using a plurality voting rule, and accordingly assessed their distribution across tweets over the sampling period.

In figure 6.3 it is showed the distribution of the different topic over the collection period. We can observe that the debate was quite stable

Governo/politica	Immigrazione	Europa/Esteri	Crimini/Società	Altro
salvini	immigrati	euro	rom	video
italia	profughi	europa	milano	anni
pd	clandestini	ue	casa	contro
italiani	profugo	fusaro	bergoglio	foto
m5s	ong	diego	morti	vuole
italiana	porti	meluzzi	mafia	può
italiano	migranti	libia	bambini	vogliono
milioni	africani	macron	roma	parla
lega	immigrato	soros	donne	byoblu
sinistra	islamici	francia	bruciato	via
casapound	imam	francesi	confessa	niccolò
maio	seawatch	gilet	falsi	casal
soldi	nigeriani	gialli	bus	vero
guerra	nigeriana	europee	choc	ufficiale
cittadinanza	nigeriano	germania	figli	bufala
prima	islamica	tedesca	case	anti
raggi	africano	mondo	chiesa	sta
governo	stranieri	notre	famiglia	grazie
renzi	chiusi	dame	magistrato	casarini
zingaretti	sea	francese	polizia	farli

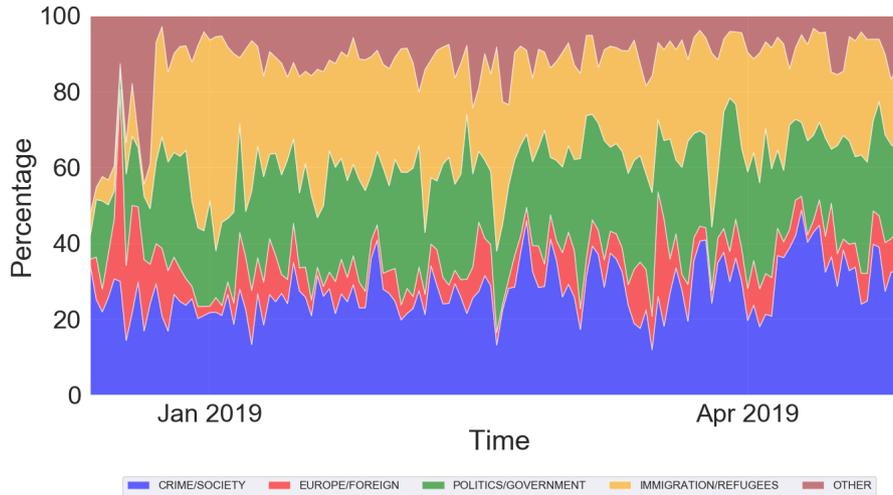*Table 6.1: Top twenty keywords associated with each topic.*

Figure 6.3: Distribution of the different topic over the collection period.

on controversial topics such immigration, refugees, crime and government whereas interest on European election was quite low in the period, with a spike in January, corresponding to the quarrel between France and Italy, in particular on news invented about Macron and CFA franc. We then performed Mann-Kendall test to find the presence of any monotonic trends in the daily distribution of different topics; we rejected the test for $alpha = 0.05/5 = 0.01$ for IR and EF whereas we accepted it for the remaining topics, detecting the presence of an upward monotonic trend in CS and PG, and a downward monotonic trend in OTHER.

In the observation period, the disinformation agenda was well settled on main topics supported by Italian leading parties, namely "Lega" and "Movimento 5 Stelle", since 2018 general elections; this suggests that they might have profited from and directly exploited hoaxes and misleading reports as to support their populist and sovranist views (whereas "Partito Democratico" appeared among main targets of misinformation campaigns); this phenomena has empirical evidence widely reported elsewhere [46].

However, after the election day, the electoral outcome confirmed the decreasing trend of "Movimento 5 Stelle" (17.1%) consensus in favor of "Lega" (34.3%). Differently from Italian 2018 politics [46] we see that the main cited leader is Matteo Salvini, leader of "Lega" party. This is consistent with a recent report on online hate speech [1], which has shown that his activity (and reception) on Twitter and Facebook is 5 times higher than Luigi Di Maio (leader of "Movimento 5 Stelle"); not surprisingly, his main agenda focuses (negatively) on immigration, refugees and Islam (which generated most of online interactions in 2018 [46]), which are also the main objects of hate speech and debate in online conversations of Italian political representatives overall.

It appears that in Italy, mainstream news didn't cover too much the European topics before in the months before the elections, focusing on the national debate [2]. This fact was observed also in other European countries according to FactCheckEU, observing that disinformation struggled to dominate online conversations mainly because European elections are extremely un-polarised and not very interesting compared to national elections.

We believe in conclusion that this might have affected the agenda-setting of disinformation outlets, which are in general susceptible to traditional media coverage [50], thus explaining the focus on different targets in their deceptive strategies.

---

[1] Report Amnesty International on hate speech during Italian European Elections.
[2] "Elezioni Europee ma poca europa", Repubblica.

6.6   Hashtag usage

We also analysed both the hashtags that appeared in the tweets and hashtags that appeared in the user profile description.

Figure 6.4: Top 10 hashtags for unique users (orange) and Tweets(blue)

Tweets hashtags

Among the most relevant hashtags shared along with tweets - in terms of number of tweets and unique users who used them (see figure 6.4) - a few indicate main political parties (cf. "m5s", "pd", "lega") and others convey supporting messages for precise factions, mostly "Lega" (cf. "salvininon-mollare", "26maggiovotolega"). Interestingly, we manually checked if there were some political affiliations with the usage of "m5s", "pd" and "lega", finding that generally "m5s", "lega" hashtags, were used to support the government, while "pd" was generally used to denigrate the "Partito Demo-cratico" party. In particular, a lot of the tweets were about the possibility to "cheat" during the primary elections of PD held the 3rd of March, by voting more than one time.

Other hashtags instead, manifested active engagement in public debates which ignited on polarizing and controversial topics (such as immigrants hospitality, vaccines, the Romani community and George Soros). We also found explicit references to (former) far-right party "CasaPound" and the associated "Altaforte" publishing house, as well as some disinformation websites

(with a remarkable polarization on "criminiimmigrati" which was shared more than 5000 times by only a few hundred accounts).

Profile hashtags

Figure 6.5: Most recurrent hashtags in profiles

By leveraging again on Twitter's API, we also extracted hashtags directly embedded in the profile description of users collected in our data. After collecting each user profile description, in order to retrieve each hashtag, we used the following regular expression:

```
(?:^|\s)[#]{1}(\w+)
```

We provide a wordcloud of the hashtags found in figure 6.5. The majority of them expresses extreme positions in matter of Europe and immigration: beside explicit references to "Lega" and "Movimento 5 Stelle", we primarily notice euro-skeptical (cf. "italexit", "noue"), anti-Islam (cf. "noislam") and anti-immigration positions (cf. "noiussoli", "chiudiamo i porti") and, surprisingly enough, also a few (alleged) Trump followers (cf. "maga" - which means Make America Great Again - and "kag" - "Keep America Great"). The latter finding is odd but somehow reflects the vicinity of Matteo Salvini and Donald Trump on several political matters (such as refugees and national security). On the other hand, we also notice "facciamorete", which started as a Twitter grassroots anti-fascist and anti-racist movement that

was born on December 2018 [3], as a reaction to the recent policies in matter of immigration and national security of actual Italian establishment, but as reported from an article of Open [4], the risk of the manipulation the hashtag only to create new fake accounts or trolls is very high.
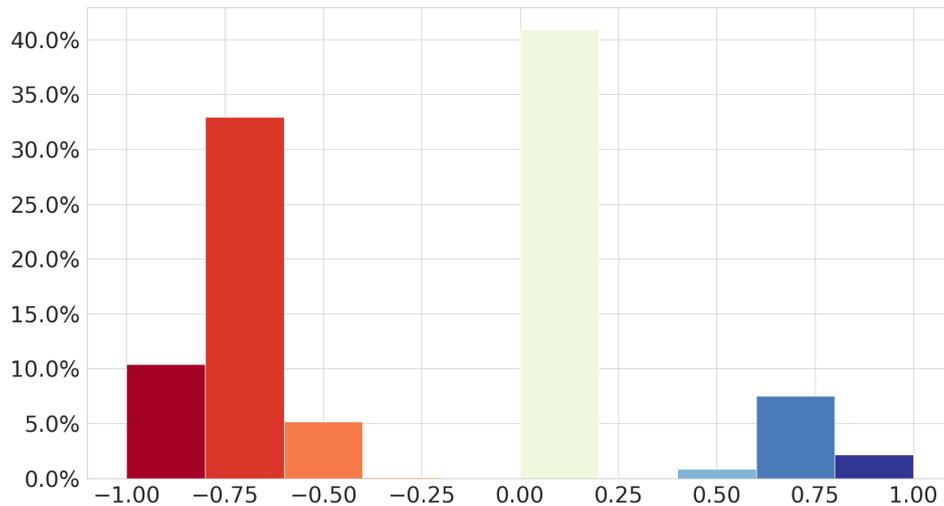
6.7 Assessing the hate speech

Figure 6.6: Sentiment analysis on article titles

According to Amnesty International [5] Italian European election campaign saw nearly 11.5% of comments on Facebook or Twitter related to politician's posts, on hate speech specifically used to insult and discriminate immigrants, religious minorities, women and ONGs. Using Dandelion API, [6], we performed semantic sentiment analysis on our articles titles - we could not perform it on tweets due to the platforms' rate limit and we could not perform it on the content due to the fact that the platform limits the length of the "sentence" to analyze. We noted from figure 6.6 that in general, most of the article titles tends to have a negative sentiment ($\sim 48\%$) as shown in the figure in red. Examples of these titles are: "Salvini non va da Fazio: "Sei un comunista con il Rolex". E lui gli risponde così", "Inghilterra, la FA Cup bandisce lo champagne: per non urtare i giocatori musulmani", "Di Maio infuriato con la Raggi: 'Aiuti i romani non i rom'". A lot of articles

---

[3]La rete Twitter più grande d'Italia: i dati dell'hashtag #FacciamoRete.
[4]Gli errori del #facciamorete.
[5]Report on Italian's hate speech.
[6]Dandelion homepage.

still have a neutral sentiment ($\sim 41\%$). Example of these articles are: "A caccia di bellezza: il teatro greco di Taormina", "A che punto siamo col federalismo differenziato", "In 2 mila a Milano per ricordare Sergio Ramelli – FOTO". Very few articles are instead considered "positive" by Dandelion API ($\sim 11\%$) and are shown in blue. Articles that are considered positive are for example: "Briatore loda Salvini: 'Ottimo lavoro su immigrazione'", "Canarie, paradiso europeo: benzina ad 1 euro, sigarette a 2 e IVA al 7%", "Ecco cosa succede quando si massaggia questo punto dell'orecchio.Fantastico!".

From these results, we note that compared to the Amnesty findings, disinformation articles tend to be very negative, raising indignation and hate speech. We still find that the targets of these "campaign of hate" are mostly immigrants.

Chapter 7

Diffusion Network analysis

In this chapter we look for an answer to **RQ3:** Who were the most influential spreaders of disinformation? Did they exhibit precise political affiliations? How could we dismantle the disinformation network?

In particular, the chapter starts d escribing the typologies of tweets and how we they are considered in order to create a diffusion network. Next section digs deeper in the analysis of the network, describing who were the most influential and most important actors in the spread of disinformation and looking if it's possible to associate them to a precise political party. The last section of the chapter evaluates the robustness of the network: if a network is very robust, it's hard to avoid disinformation spread. On the other hand, a very weak network would offer an interesting solution to curb disinformation spread.

7.1 Data description

From the data gathered using the methodology described in chapter 4, during the period from 7th of January to 26th of May which, removing the "fact-checking" websites, we obtained the dataset described in table 7.1.

Number of tweets collected	Type of tweets
All types	414468
Tweet	96897
Retweet	253076
Quote	59481
Reply	5014

Table 7.1: Dataset description considering only claims

From the dataset, we built a **diffusion network** according to Hoaxy's methodology [42] , which basically creates a network that follows the information flow.

Network creation

As already said, there are different types of *tweets*:

- **Tweet**: a tweet is a post from an author. If it is alone, the author is considered as a single node of the network.

- **Retweet**: a retweet is a new post that shares the content of another post and can add some content. When a user b retweets a tweet from user a, we built an edge from a to b.

- **Reply**: a reply is a new tweet starting with `screen_name`. When a user b replies to a tweet from user a, we built an edge from a to b.

- **Quote**: a quote is retweet that includes new text.If a user b quotes a tweet from user a, we built an edge from a to b.

- **Mention**: when a user b mentions another user a, we built an edge from b to a.

From these, we were able to create a new graph $G = <N, E>$ where N is the set of users, E is the set of edges before defined.

7.2 Analysis

The graph just defined, is called **diffusion network**. Our diffusion network has more than 20k nodes and more than 100k edges.

Most important users

In order to analyze the diffusion network we first find the main k-core using the k-core decomposition algorithm, a recursive algorithm that, given a k-core, extracts the $k + 1$ core by removing all the nodes with degree k. The last non empty graph it's called indeed the main k-core.

The number of nodes in the main k-core were 218, while the number of edges 8538 and $k = 48$. The main k-core it is shown in figure 7.1, and the colors corresponds to the output of the Louvain modularity's algorithm able to find communities and they will be described deeper in the next sections.
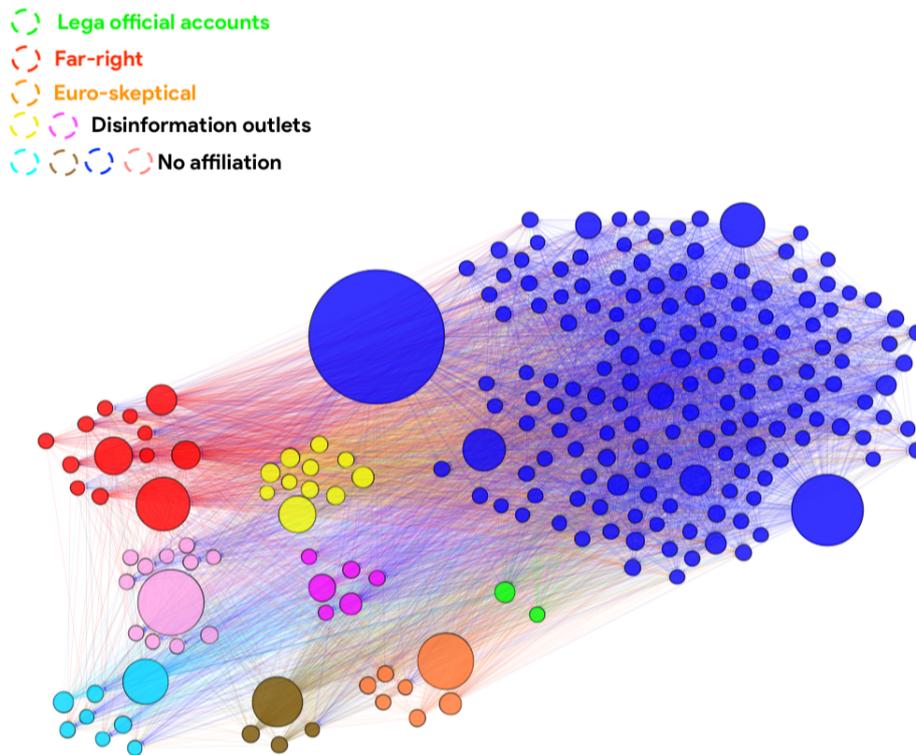
Figure 7.1: Main k-core of the disinformation network.

We next exploited other centrality measures such as **Degree**, **Strength**, **Betweenness** and **PageRank**. These metrics embed different meanings.

In particular, the **In-Degree** of a node a is the number of users that a has quoted, replied to, or retweeted plus the number of other nodes that have mentioned a. We could say that if you have a high in-degree you're usually interacting a lot with other people's tweets.

In parallel, the **Out-Degree** of a node a is the number of users that have replied, retweeted or quoted node a plus the number of other nodes that have been mentioned by a. We could say that the higher your out-degree the more your posts are being shared over twitter, which means that to higher out-degrees usually correspond more famous users.

The **Degree** of a node is the sum of the in-degree and out-degree. It basically embeds both of the features explained earlier.

Rank	In-Strength	Out-Strength	Betweenness	PageRank
1	napolinordsud ×	Filomen30847137	IlPrimatoN	IlPrimatoN
2	RobertoPer1964	POPOLOdiTWlTTER	matteosalvinimi	matteosalvinimi
3	razorblack66	laperlaneranera	Filomen30847137	Sostenitori1 ×
4	polizianuovanaz ×	byoblu	byoblu	armidmar
5	Giulia46489464	IlPrimatoN	a_meluzzi	Conox_it ×
6	geokawa	|petra_romano	AdryWebber	lauraboldrini ×
7	Gianmar26145917	araldoiustitia	claudioerpiu	pdnetwork ×
8	pasqualedimaria ×	max_ronchi	razorblack66	libreidee ×
9	il_brigante07	Fabio38437290	armidmar	byoblu
10	AngelaAnpoche	claudioerpiu	Sostenitori1 ×	Pontifex_it ×

Table 7.2: Top ten users according to different centrality measures. The cross indicates nodes that do not belong to the main k-core (k=48) of the network.

Strengths (in and out) of a node considers the number of interactions of a node rather than the number of users to which the node has been interacting with and in particular **In-Strength** of a node a represents the number of interactions a did (in terms of quotes, replies retweets) plus the number of times it has been mentioned. Similarly to in-degree to higher in-strengths correspond higher "active" interactions to users. Similarly, the **Out-Strength** of a node a is the number of interactions with node a plus the number of other nodes that have been mentioned by a. Again, analogously to out-degree, we could say that the higher your out-strength the more you are famous. Even though it's harder to see in social media applications the meaning of **PageRank** and **Betweenness** they are still considered as a very strong centrality metrics for networks: the first used originally to rank web pages in search engine queries while the second quantifies the probability of a node to act as a bridge (so as a must-go path) between two other nodes.

Table 7.2 shows the top-10 users according to different the different centrality measures; the cross indicates nodes that do not belong to the main k-core of the network but are found by the corresponding centrality metric. We note that, despite their different concepts, the majority of nodes with highest values of In-Strength, Out-Strength and Betweenness belongs also to the main k-core of the network; the same does not hold for users which have a large PageRank centrality value, in fact six nodes out of ten does not belong to the k-core.

A few users strike the eye:

1. `matteosalvinimi` is Matteo Salvini, actual leader of the far-right wing "Lega" party and vice-prime minister of the Italian Parliament; he is not an active spreader of disinformation, being responsible for just one

(true) story coming from disinformation outlet "lettoquotidiano.com" (available here), which was shared over 1800 times. He is generally **passively** involved in deceptive strategies of malicious users who attempt to "lure" his followers by attaching disinformation links in replies/retweets/mentions to his account. This technique - reach followers exploiting famous accounts - is very common in the social media environment.

2. `a_meluzzi` is Alessandro Meluzzi, a former representative of centre-right wing "Forza Italia" party (affiliated to Silvio Berlusconi); he is a well-known supporter of conspiracy theories [45] and a very active user in the disinformation network, with approximately 400 deceptive stories shared overall.

3. Accounts associated to disinformation outlets, namely `IlPrimatoN` with "ilprimatonazionale.it", `byoblu` with "byoblu.com", `libreidee` with "libreidee.org", `Sostenitori1` with "sostenitori.info" and `Conox_it` with "conoscenzealconfine.it".

A manual inspection revealed that most of the influential users are indeed actively involved in the spread of disinformation, with the only exception of `matteosalvinimi` who is rather manipulated by other users, via mentions/retweets/replies, as to mislead his huge community of followers (more than 2 millions). The story shared by Matteo Salvini underlines a major issue with disinformation outlets identified in this analysis: they often publish simple true and factual news as to bait users and expose them to other harmful and misleading content present on the same website.

Besides, we identified a few influential users who represented the target of several disinformation campaigns:

1. `lauraboldrini` is Laura Boldrini, representative of left-wing "Liberi e Uguali" party and actual member of the Italian Parliament; in the last few years she has been repetitively a target of fake news.

2. `pdnetwork` is the account of the centre-left "Partito Democratico" party; as the former ruling party it has been severely attacked in the propaganda of both actual ruling parties ("Lega" and "Movimento 5 Stelle").

3. `Pontifex_it` is the account of Papa Francesco; due to his recent statements in favor of immigration he has become another target of Italian far-right online hateful speech.

We also report a suspended account (`polizianuovanaz`), a protected one (`Giulia46489464`) and a deleted user (`pasqualedimaria`).

    We then identified different communities using **Louvain modularity-based** algorithm [13], on the whole diffusion network and we investigated these communities of users in the main K-core (see figure 7.1), and we noticed systematic interactions between distinct accounts. We manually inspected usernames, most frequent hashtags and referenced sources, deriving the following qualitative characterizations:

1. the **Green** community corresponds to "Lega" party official accounts: `matteosalvinimi` and `legasalvini`. We also noticed the presence of a third account, `noipersalvini`, belonging to the same community but not appearing in the core.

2. the **Red** community represents Italian far-right supporters, with several representatives of CasaPound (former) party (including his secretary `distefanoTW` who does not appear in the core), who obviously refer to "ilprimatonazionale.it" news outlet. Most frequent hashtags are "#M5S" (18), "#CasaPound" (14) and "#rom" (11). Most referenced sources are "ilprimatonazionale.it" (19649), "voxnews.info" (2539) and "skynew.it" (839).

3. the **Yellow** community is strongly associated with two disinformation outlets, namely "silenziefalsita.it" (`SilenzieFalsita`) and "jedanews.it" (`jedasupport`); the latter was one of the pages identified in Avaaz report[1] and deleted by Facebook. Most frequent hashtags are "#ottoemezzo" (7), "#M5S" (6) and "#DiMaio" (6). Most referenced sources are "silenziefalsita.it" (7371), "voxnews.info" (2322) and "jedanews.it" (1824).

4. the **Orange** community is associated to the euro-skeptical and conspiratory outlet "byoblu.com" (`byoblu`), and it also features Antonio Maria Rinaldi (`a_rinaldi`), a well-known euro-skeptic economist who has just been elected with "Lega" in the European Parliament. Most frequent hashtags are "#Soros" (8), "#SalonedelLibro" (8) and "#vaccini" (7). Most referenced sources are "byoblu.com" (13857), "ilprimatonazionale.it" (4695) and "comedonchisciotte.org" (2547).

5. the **Purple** community corresponds to the community associated to "tuttiicriminidegliimmigrati.com" (`TuttICrimin`) disinformation outlet. Most frequent hashtags are "#CriminiImmigrati" (2923), "#26mag-

---

[1]Avaaz report

giovotoLega" (9), "#PD" (8) and "#portichiusi" (8). Most referenced sources are "tuttiicriminidegliimmigrati.com" (3609), "skynew.it" (1200) and "voxnews.info" (381).

6. the remaining **Blue** (`Filomen30847137`), **Light-blue** (`araldoiustitia`) and **Brown** communities (`petra_romano`) represent different groups of very active "loose cannons" who do not exhibit a clear affiliation. Most frequent hashtags are altogether "#CriminiImmigrati" (368), "#portichiusi" (319), "#SalviniNonMollare" (292). Most referenced sources are altogether "voxnews.info" (125127), "ilprimatonazionale.it" (4701) and "skynew.it" (4114).

For what concerns hashtags, we notice that they are not largely employed when directly sharing disinformation news, with a few exceptions being "#CriminiImmigrati", "#portichiusi" and "#SalviniNonMollare": the first one is practically associated to the corresponding outlet ("tuttiicriminidegliimmigrati.com") where hoaxes and factual news about immigrants committing crimes are published on a daily basis; the other two ("#portichiusi" stands for "closed docks", whereas the other means "Salvini do not give up") are two traditional slogans associated with Matteo Salvini and his fierce position against immigration; we also find some direct references to main political parties. As previously highlighted, hashtags mostly relate to controversial topics and matters of debate which have inflamed the political discussion during the run up to European elections.

For what concerns sources of disinformation, we observe that the global trend prevails also in the core, with "voxnews.info" and "ilprimatonazionale.it" holding the highest number of tweets among several communities; nonetheless, we observe a polarization of other communities w.r.t specific sources, e.g. Purple mainly engage with "silenziefalsita.it" and "jedanews.it", Pink with "tuttiicriminidegliimmigrati.com" and Sky-blue with "byoblu.com".

Investigating bot activity

Eventually, we employed Botometer algorithm [40] as to detect the presence of social bots among users in the main core of the network. The algorithm uses more than 1k features, and group them in 6 different classes: **Network**, **User**, **Friends**, **Temporal**, **Content**, **Sentiment**. We set a threshold of 50% on the Complete Automation Probability (CAP) - i.e. the probability of an account to be completely automated - which, according to the authors, is a more conservative measures that takes into account an estimate of the overall presence of bots on the network; besides, we computed the CAP value

based on the language independent features only (which are grouped in the classes called network, users, friends and temporal), as the model considered also some features conceived for English-language users. We only detected two bot-like accounts, namely `simonemassetti` and `jedanews`, respectively with probabilities 58% and 64%, that belong to the same Purple community. A manual check confirmed that the former habitually shares random news content (also mainstream news) in an automatic flavour whereas the latter is the official spammer account of "jedanews.it" disinformation outlet. We noticed also that after some time, twitter suspended `simonemassetti` and that `jedanews` account was deleted.

7.3   Network robustness

In the previous section, we used some centrality to find out who were the most important nodes in our network. In addition to these (s_{in}, s_{out}, $betweenness$ and $PageRank$) we also considered the s_{innotw} which is the **in** strength of a node not taking in consideration the weight of each edge, which would be the s_{in} in an unweighted network, and $s_{outnotw}$ which is the **out** strength of a node considering only the number of outgoing edges. Then, similar to Shao et al. [42], we exploit these metrics in an automatic fashion, to evaluate the robustness of the network eliminating the most important nodes, and finally measuring the impact on tweet's volume and on number of interactions i.e. the number of edges.

Dismantling the disinformation network

We first rank nodes in decreasing order with respect to each metric and k-core. Then, we remove each node one by one while tracking the resulting amount of edges and tweets in the network. In figure 7.4 we show our results.

We can observe that with respect to edge dismantling (shown in figure 7.2), betwenness, out, out not weight initially perform similarly, but at some point removing nodes considering their out strength non weighted degree turns out to be the best strategy, allowing to remove 80% of the edges, by just removing 100 users (0.4% of users). Considering the tweet volume (as shown in figure 7.3), we see no relevant distinction between the metrics, but it looks like the in strength is the most effective with less users. By removing 100 users, we can eliminate 20% of tweets.

The results are impressive, in particular if we consider the edges dismantling, and shows how few users are responsible for most of the disinformation, and in particular of the network connection. By just removing these users
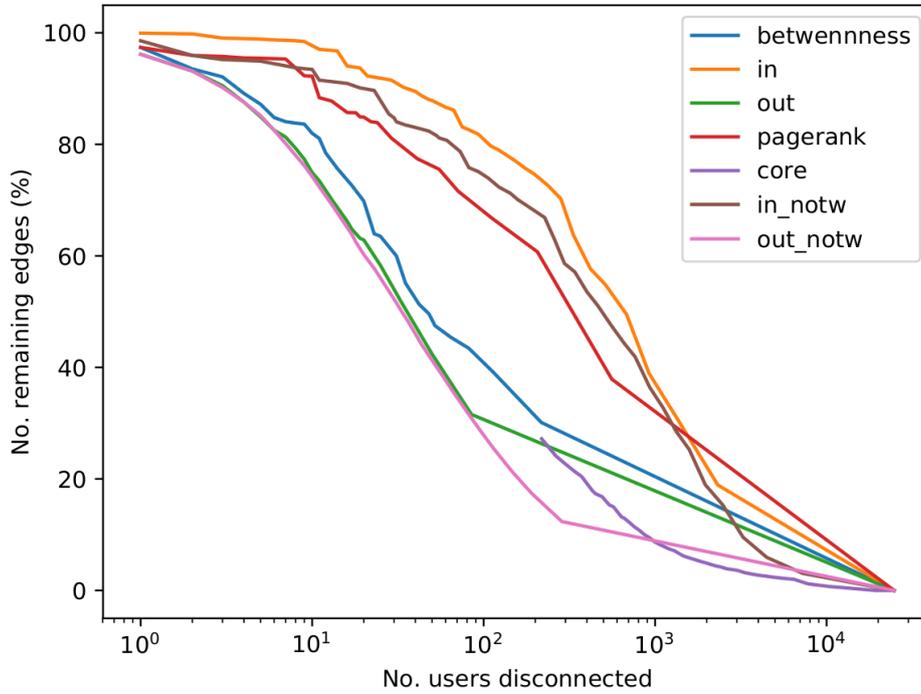
Figure 7.2: Dismantling analysis for edges

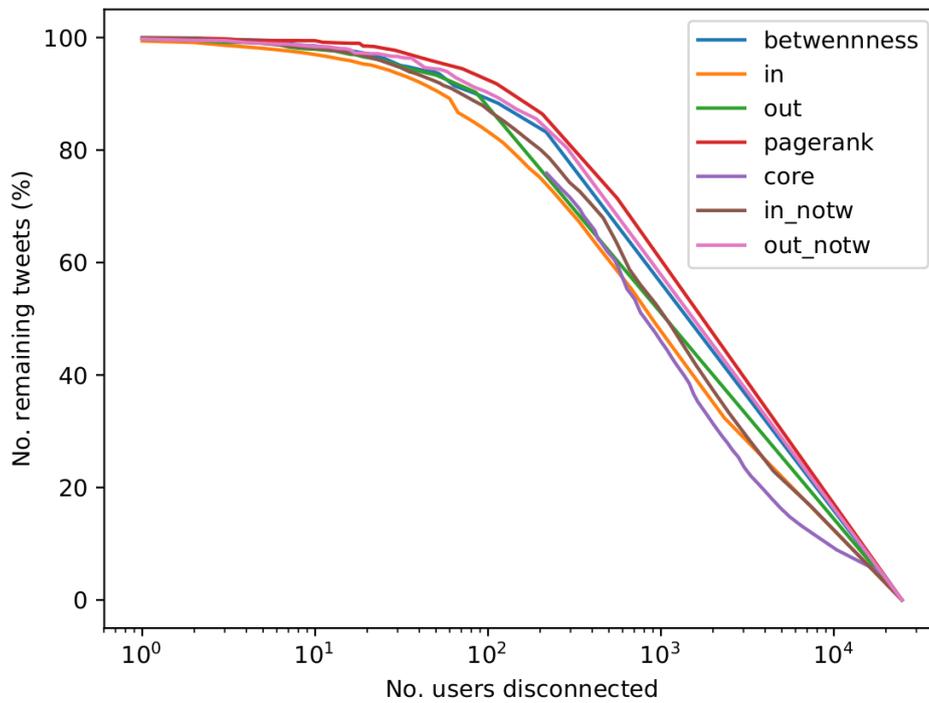*Figure 7.3: Dismantling analysis for tweets*

Figure 7.4: Results of different network dismantling strategies w.r.t to remaining edges in the network (top) and shared tweets (bottom). The x-axis (log scale) indicates the number of disconnected nodes and the y-axis the fraction of remaining edges (tweets) in the network.

we could curb the spread of disinformation by a lot.

Concerns on censorship

We finally raise a fundamental issue: is the social media (in this case Twitter, but it could be any) motivated enough to limit disinformation spread? Indeed, as Berinsky pointed out [51], platforms business model rely on monetizing our attention. The conflict of interests is evident: if platforms curb disinformation, they could lose revenues. They also claim that acting in in a stricter way would cause (motivated) complains on both freedom of speech and users privacy. For these reasons, should a third party agents (such as governments or European Union) intervene making laws to protect citizens? How much should they intervene? If you regulate too little, you're not effectively curbing the disinformation spread. But if you regulate too much, risks of censorship are very high.

Chapter 8

Coordinated disinformation strategy

This chapter presents the methodology used and the results found answering the **RQ4:** Did disinformation outlets organize their deceptive strategies in a coordinated manner? Can we identify inter-connections across different countries?

In the next sections, we first describe the method used to find links between disinformation outlets, and then we describe our findings. We finally check if our results show a trend towards American or Russian outlets.

8.1   Network creation

In our data, we didn't only scrape information from the article's web pages such titles and content, but we also downloaded their Hypertext Markup Language (HTML). From the HTML we considered all the Hypertext references containing a URL that did not point to the same domain. In order to get them, we relied on the **BeautifulSoup** module, identifying first all the `<a>` tags, extracting their `url` attribute, and verifying then that it pointed to a different domain.

Interestingly, we noted that almost all articles had a hyperlink in their body pointing to external sources(99%).

We then built a network, where nodes were website domain, and edges were urls that pointed from one node to another. In particular, considering a URL *https:/b/article* found in an article page with url *https:/a/article*, the nodes are *a* and *b* and there's and edge from *a* to *b*. The result is a directed and weighted graph where the weight of an edge is the number of twitter URL inside our data collection that have hyperlinks from the URL's

domain to any other domain.

We then used data memos from COMPROP [1] research group [44][52][53][54] and by Hoaxy [42] (mostly American websites), to identify classes of nodes belonging to "Mainstream news", "Social Networks", and "JN" (Junk-News) dividing them into EU ("opensources.co", "décodex.fr"), US ("Breitbart"), RU ("RussiaToday", "Sputnik") disinformation outlets and our "Italian Disinformation Outlets".

We finally obtained a directed weighted graph with $\sim 5k$ nodes and $\sim 8k$ edges, with a total of more than 75k external-redirections (some articles might quote different).
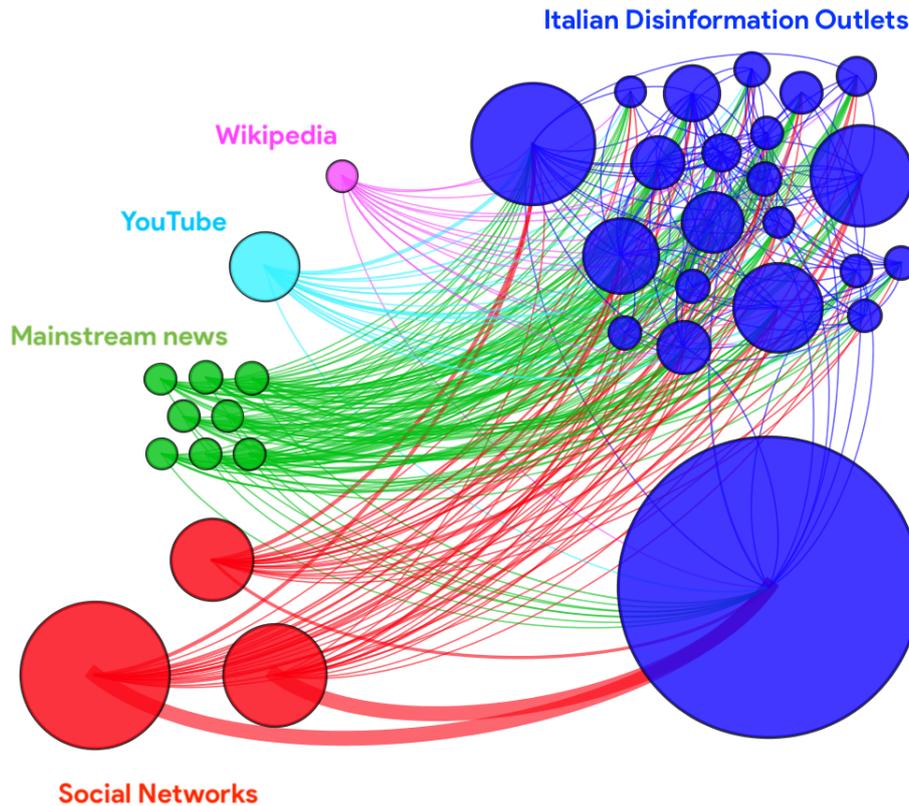

Figure 8.1: Main k-core disinformation outlets network

---

[1]COMPROP homepage

8.2   Network analysis

Main k-core

In this section we consider all the nodes and their classes providing a description of the core of the network, and discussing our findings.

Using the k-core decomposition algorithm, we found the main k-core of the network: a sub-graph containing 35 nodes and with $k = 14$. In the core (showed in figure 8.1), we first noticed frequent connections between distinct disinformation outlets, that suggested the presence of shared agendas and presumably coordinated deceptive tactics, as well as frequent mentions to reputable news websites as shown in particular in figure 8.2; among them we distinguished *IlFattoQuotidiano"*, which is a historical supporter of "Movimento 5 Stelle", and conservative outlets such as *"IlGiornale"* and *"LiberoQuotidiano"* which lean instead towards "Lega". There are also present news outlets that traditionally have moderate positions - for example *"Corriere della Sera"* which tends to have a center-right opinion or *"Repubblica"* which usually sponsors a center-left view - and some newspapers with a leaning towards left parties such *"HuffingtonPost"* and *"La Stampa"*. We also interestingly report the presence of *"Ansa"*, which is a news agency that publishes the news and it's considered very objective and, as reported from Reuters [2], it's also the most trusted outlet in Italy. From these findings, we investigated the connection between mainstream outlets and disinformation outlets, observing that most of the time the former were cited in a deceptive way, promoting the biased agenda of the disinformation outlet. We finally report that the more quoted newspaper are "ilgiornale.it" and "ilfattoquotidiano", as showed in figure 8.1, manifesting evidences of ideological proximity between some disinformation outlets and mainstream ones.

Considering again the whole main k-core, we also noted that most of the external re-directions point to social networks (Facebook and Twitter) and video sharing websites (Youtube); this is no wonder given that disinformation is often shared on social networks as multimedia content [2][43]. In addition, we inspected nodes with the largest number of incoming edges (In-degree) in the original network, discovering among uppermost 20 nodes a few misleading reports originated on dubious websites (such as "neoingegneria.com"), flagged by fact-checkers but that were not included in any blacklist. We believe that a more detailed network analysis could reveal additional relevant connections and we leave it for future research.

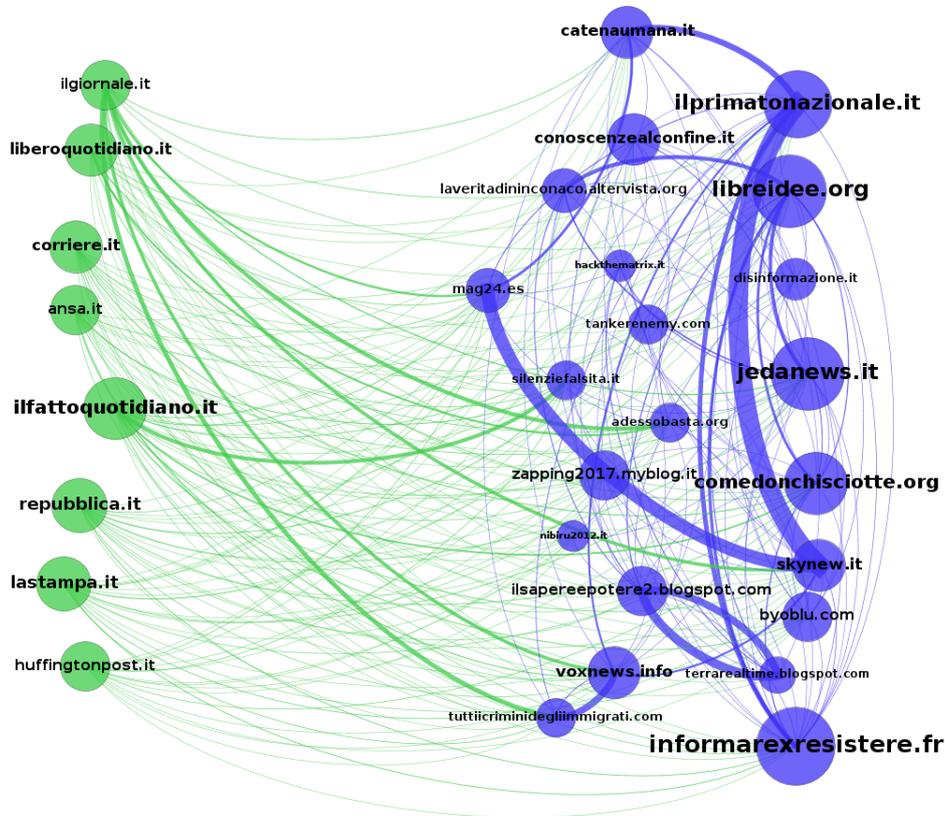[2] Reuters Institute Digital News Report 2019.

Figure 8.2: Italian mainstream outlets quoted in k-core

Coordinated disinformation

We then focused on the sub-graph composed of three particular classes of nodes, namely Russian (RU) sources, EU/US disinformation websites and our list of Italian (IT) outlets.

The resulting bipartite network - we filtered out intra edges between IT sources to better visualize connections with the "outside" world - contained over 60 foreign websites (RU, US and EU) and it is shown in Fig 8.4.

We observed a considerable number of external connections (over 500 distinct hyperlinks present in articles shared more than 5 thousand times) with other countries sources, which were primarily included within "voxnews.info", "ilprimatonazionale.it" and "jedanews.it".

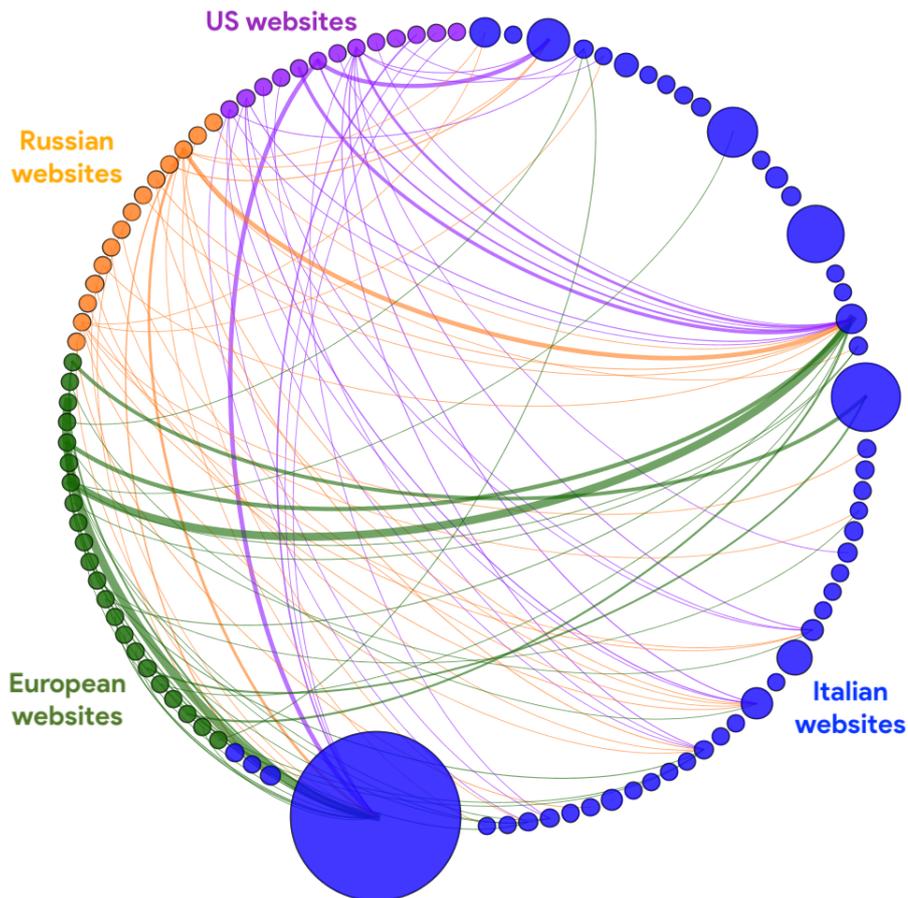Among foreign sources we encountered several well-known US sources

Figure 8.3: Sub-graph of Russian (orange), EU (olive green), US (violet) and Italian (blue) disinformation outlets.

("breitbart.com", "naturalnews.com" and "infowars.com" to mention a few) as well as RU ("rt.com", "sputniknews.com" and associated networks in several countries), but we also find interesting connections with important disinformation outlets from France ("fdesouche.com" and "breizh-info.com"), Germany ("tagesstimme.com"), Spain ("latribunadeespana.com") and Sweden ("nyheteridag.se" and "samnytt.se"). Besides, a manual inspection of a few articles revealed that stories often originated in one country were immediately translated and promoted from outlets in different countries (see Fig 8.4). Such findings suggest the existence of coordinated deceptive strategies which span across several countries, consistently with claims in the latest

report by Avaaz [3] which revealed the existence of a network of far-right and anti-EU websites, leading to the shutdown of hundreds of Facebook pages with more than 500 million views just ahead of the elections. Far-right disinformation tactics comprised the massive usage of fake and duplicate accounts, recycling followers and bait and switch of pages covering topics of popular interest (e.g. sport, fitness, beauty).
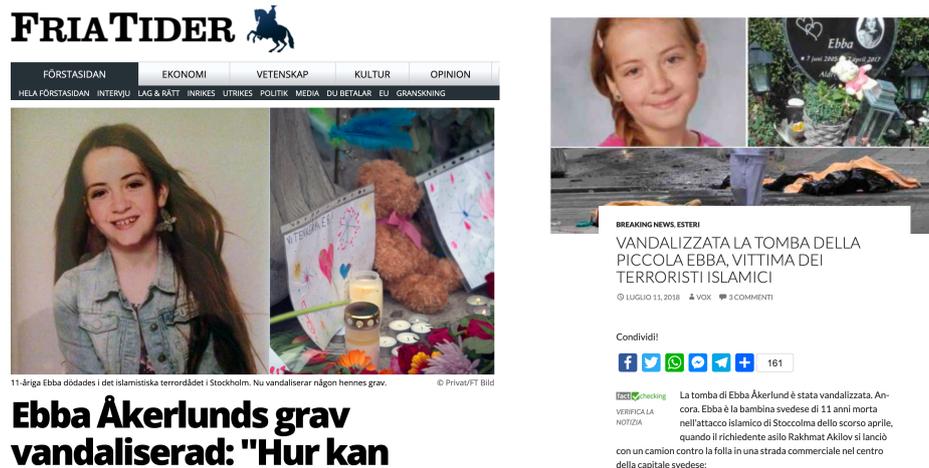

Figure 8.4: Fake news diffusion

It is interesting that Facebook decided on the basis of external insights to shutdown pages delivering misleading content and hate speech; differently from the recent past [32][43] it might signal that social media are more willing to take action against the spread of deceptive information in coordination with findings from third-party researchers. Nevertheless, we argue that closing malicious pages is not sufficient and more proactive strategies should be followed [43].

Trend detection towards RU, US

Finally, we performed a Mann-Kendall test to see whether there was an increasing trend towards the elections, in the number of external connections with US and RU disinformation websites; we rejected it at $\alpha = 0.05/2 = 0.0025$.

---

[3]Avaaz report.

Chapter 9

Conclusion

The main purpose of this thesis was to investigate Italian disinformation; we focused on a period of nearly six months before the European Election day (from 07/01/2019 to 26/05/2019). In particular, we structured our work following some specific research questions and we provided an answer to each of them.

9.1   Contributions

We found out that a small number of outlets accounted for most of the deceptive information circulating on Twitter; some of them belonged to a group of Facebook pages that were recently closed by the company itself, after violating the platform's term of use, as reported by Avaaz. Among these sources only one, **byoblu**, increased its activity and was also able to increase its shares, probably due to the fact that it sponsors mainly euro-skeptical positions. We also recognized a heterogeneous yet limited community of thousands of users who were responsible for sharing disinformation. Few of the accounts - 20% - were responsible of around 90% of the tweets volume, with a few hundreds accounts posting alone half of the tweets in our dataset [**RQ1**].

Next, we studied what were the most debated narratives of disinformation, observing that they mostly concern polarizing and controversial arguments of the local Italian political debate such as immigration, crime and national safety, reporting them in a negative way. Unfortunately, although it was a European election, we noted that discussion around the topics of Europe global management had a negligible presence throughout the collection period as mainly reported in other mainstream media. We also report affiliations with the American's republican party, in particular with *Trump*,

with hashtags such as #Qanon, #Maga (Make America Great Again) and #kag (Keep America Great) [**RQ2**].

We then conducted an analysis on the diffusion network resulting from users sharing disinformation articles on Twitter, to find who were the most influential spreaders of disinformation. We detected the presence of active groups with strong ties with the Italian far-right and conservative community, in particular with *"Lega"* party, as most of the users manifested explicitly support to the party agenda through the use of keywords and hashtags. Besides, we also noted that a common deceptive strategy was to passively involve *"Lega"*'s leader *Matteo Salvini* via mentions, quotes and replies attaching disinformation content, as to potentially mislead his audience of million of followers. While looking for a way to curb fake news' spread, we observed that the diffusion network is not robust at all: in fact disabling a limited number of central users in the network would effectively restrict the network size, but it would immediately raise concerns on censorship [**RQ3**].

We finally investigated inter-connections between different transnational deceptive agents, uncovering an intricate network of malicious portals, that seemingly acted in a coordinated manner during the period we analysed. We in fact discovered many cases where the same (or very similar) stories were shared in different languages not only across different European countries but also in U.S. and Russia [**RQ4**].

We conclude stating that despite Twitter is not really commonly used in Italy - it is the 6th social media according to Reuters - disinformation is present in the social media and that its spreaders are close to the right wing ideas in particular on topics such as immigration, domestic policy, sovereignism, no-EU feelings and some explicit support for Trump. This is somehow different from the U.S. findings where even if the **majority** of disinformation winks to the republican party, **some** disinformation that is close to democratic party's view still exists. We think that the lack of left-oriented disinformation in Italy is due to the fact that probably these topics are overall less appealing or to the lack of an active presence of left parties and supporters on social media.

9.2   Limitations

Although the results show an interesting picture of what is disinformation in Italy, we are aware of some limitations. The first one is that we can track only a limited amount of disinformation sources. This is due both to the fact that we can't know all the possible disinformation sources, and to data restriction's on Twitter's API volume. For what concerns Facebook,

obtaining information is very difficult due to the company's policy; to bypass them we used netvizz but this service will soon be closed, making this task even harder. The second limit is that we don't track falsities in mainstream outlets, which themselves sometimes have been found sharing literally fake news. The third limit, is our methodological assumption that all news shared by websites are false. As this assumption is sometimes false (as also showed in different sections). The fourth limit is about the new ways to spread disinformation. False news articles nowadays account only for **some** of the disinformation. Videos, photos also play an important role, and they are usually spread via chats which are not public and therefore is very hard to infer their entity.

9.3 Future work

Despite the limitations, we think that the study of disinformation through labelling websites is very promising for understanding how disinformation behaves and how to curb it.

During our work, we noticed that compared to mainstream outlets, disinformation websites were less "readable", they had a lot of ads (as Alcott [1] highlighted, one of the reasons of disinformation websites proliferation is their revenues) and did not contain certain information in the footer such as the business name, or a phone number. Some of them had the same layout, style and font with different names (skytg24news.it, tg24.skynew.it and catenaumana.it) although publishing different articles.

In a recent work, Castelo [55] explored a manual method to find disinformation sites from their look and feel, studying the differences with respect to mainstream outlets. We suggest instead to create a model able to embed these features using deep learning. An idea could be build a dataset of screenshots of each disinformation websites and mainstream outlets, and use them as unstructured data to predict if a new screenshot belongs to a disinformation website or to a mainstream one using some models that are able to understand "readability". Another idea could make use of the raw html (without it's content) instead of the images.

We found really promising also the investigation of connections between websites via hyperlinks. A simple augmentation to the model presented, could be the addition of "hops" to the analysis. So, instead of monitoring only the newspaper quoted in the disinformation portal, we would build a chain (or a network) of websites to find the source of the news. Following this line, one could detect how much the articles are different between hops and identify if some content was added and at what point of the "chain".

A new interesting field, as pointed out in our fourth limitation, is the study of images and videos that are manipulated or present fake content, and how to curb their spread on social media such Telegram, WhatsApp and Instagram.
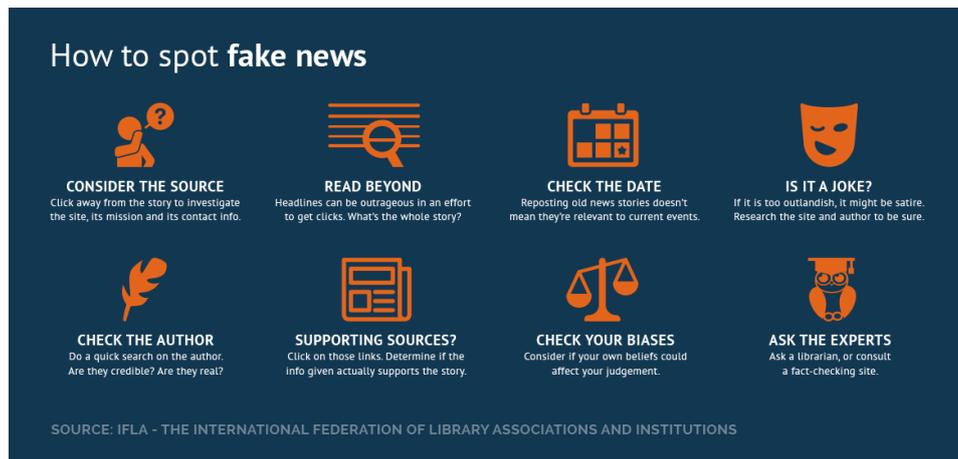

Figure 9.1: How to spot a fake news, courtesy of ifla.org

Finally, we stress the necessity of unmasking fake news and helping citiziens to recognize misinformation. It's very hard to convince someone that shared a falsity that it was indeed a false news. Even if debunking tries to address this problem, they have a huge limitation on resources and less reach (as showed in this work was around $\sim 5.5\%$), therefore they are not effective. One way to help citizens in spotting junk news is educating them to use **critical thinking** either at schools or with some targetted campaings. For example, figure 9.1 shows some tips to spot a fake news, shared by ifla [1], that are commonly used by debunkers to find false news.

---

[1]International Federation of Library Associations and Institutions

Bibliography

[1] Hunt Allcott and Matthew Gentzkow. Social Media and Fake News in the 2016 Election. 31(2):211–236, 2017.

[2] Francesco Pierri, Politecnico Milano, Dipartimento Elettronica, Stefano Ceri, Politecnico Milano, and Dipartimento Elettronica. False News On Social Media : A Data-Driven Survey.

[3] Francesco Pierri, Alessandro Artoni, and Stefano Ceri. Investigating italian disinformation spreading on twitter in the context of 2019 european elections. *arXiv preprint arXiv:1907.08170*, 2019.

[4] Daniel Jurafsky and James H Martin. Speech and Language Processing. 2018.

[5] Helmut Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, Manchester, UK, 1994.

[6] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012.

[7] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[8] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[9] Vladimir Batagelj and Matjaz Zaversnik. An o (m) algorithm for cores decomposition of networks. *arXiv preprint cs/0310049*, 2003.

[10] Dietrich Stauffer and Ammon Aharony. *Introduction to percolation theory*. Taylor & Francis, 2018.

[11] 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. *arXiv preprint arXiv:1901.07933*, 2019.

[12] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[13] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[14] Rubayyi Alghamdi and Khalid Alfalqi. A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6(1), 2015.

[15] Lin Liu, Lin Tang, Wen Dong, Shaowen Yao, and Wei Zhou. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, 5(1):1608, 2016.

[16] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[17] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.

[18] Thomas Hofmann. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc., 1999.

[19] Yee W Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems*, pages 1385–1392, 2005.

[20] Michael Röder, Andreas Both, and Alexander Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408. ACM, 2015.

[21] Keith Stevens, Philip Kegelmeyer, David Andrzejewski, and David Buttler. Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 952–961. Association for Computational Linguistics, 2012.

[22] David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 100–108. Association for Computational Linguistics, 2010.

[23] David Mimno, Hanna M Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. Optimizing semantic coherence in topic models. In *Proceedings of the conference on empirical methods in natural language processing*, pages 262–272. Association for Computational Linguistics, 2011.

[24] Henry B Mann. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, pages 245–259, 1945.

[25] Maurice George Kendall. Rank correlation methods. 1948.

[26] Andreas M Kaplan and Michael Haenlein. Users of the world , unite ! The challenges and opportunities of Social Media. (December), 2017.

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

[28] Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin, 2011.

[29] Joseph Weizenbaum et al. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[30] B Y Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The Rise of Social Bots.

[31] Soroush Vosoughi, Deb Roy, and Sinan Aral. news online. 1151(March):1146–1151, 2018.

[32] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature Communications*, (2018), 2017.

[33] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-thompson, and David Lazer. Fake news on Twitter during the 2016. 378(January):374–378, 2019.

[34] Francesco Pierri, Carlo Piccardi, and Stefano Ceri. Topology comparison of twitter diffusion networks reliably reveals disinformation news. *arXiv preprint arXiv:1905.03043*, 2019.

[35] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. Hoaxy : A Platform for Tracking Online Misinformation. pages 745–750.

[36] Craig Silverman. This analysis shows how viral fake election news stories outperformed real news on facebook. *BuzzFeed News*, 16, 2016.

[37] Craig Silverman and Jeremy Singer-Vine. Most americans who see fake news believe it, new survey says. *BuzzFeed News*, 6, 2016.

[38] Hannah Jane Parkinson. Click and elect: how fake news helped donald trump win a real election. *The Guardian*, 14, 2016.

[39] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.

[40] Onur Varol, Emilio Ferrara, Clayton A Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. In *Eleventh international AAAI conference on web and social media*, 2017.

[41] Pik-mai Hui, Chengcheng Shao, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. The Hoaxy Misinformation and Fact-Checking Diffusion Network. (Icwsm):528–530, 2018.

[42] Chengcheng Shao, Pik-mai Hui, Lei Wang, Xinwen Jiang, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. Anatomy of an online misinformation network. pages 1–23, 2018.

[43] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Ny- han, Gordon Pennycook, David Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[44] Nahema Marchal, Bence Kollanyi, Lisa-Maria Neudert, and Philip N Howard. Junk news during the eu parliamentary elections: Lessons from a seven-language study of twitter and facebook.

[45] Alexandre Alaphilippe, Chiara Ceccarelli, Léa Charlet, and Martin My- cielski. Disinformation detection system: 2018 italian elections. *Brus- sels: EU Disinfo Lab, June*, 1, 2018.

[46] Fabio Giglietto, Laura Iannelli, Luca Rossi, Augusto Valeriani, Nicola Righetti, Francesca Carabini, Giada Marino, Stefano Usai, and Elisa- betta Zurovac. Mapping italian news media political coverage in the lead-up to 2018 general election. *Available at SSRN 3179930*, 2018.

[47] Alexandre Bovet and Hernán A Makse. Influence of fake news in twit- ter during the 2016 us presidential election. *Nature communications*, 10(1):7, 2019.

[48] Jacob Ratkiewicz, Michael D Conover, Mark Meiss, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer Menczer. Detecting and tracking political abuse in social media. In *Fifth international AAAI conference on weblogs and social media*, 2011.

[49] Jason Chuang, Spence Green, Marti Hearst, Jeffrey Heer, and Philipp Koehn. Proceedings of the workshop on interactive language learn- ing, visualization, and interfaces. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 2014.

[50] Maxwell McCombs. *Setting the agenda: Mass media and public opinion.* John Wiley & Sons, 2018.

[51] Adam Berinsky, Kelly Greenhill, Filippo Menczer, Miriam Metzger, Brendan Nyhan, Gordon Pennycook, Michael Schudson, Cass Sunstein, Emily Thorson, Duncan Watts, and Jonathan Zittrain. The science of fake news. (March), 2018.

[52] Philip N Howard, Samantha Bradshaw, Bence Kollanyi, and Gillian Bolsolver. Junk news and bots during the french presidential election: What are french voters sharing over twitter in round two? Technical

report, Technical report, Data Memo 2017.4. Project on Computational Propaganda . . . , 2017.

[53] Freja Hedman, Fabian Sivnert, Bence Kollanyi, Vidya Narayanan, Lisa-Maria Neudert, and Philip N Howard. News and political information consumption in sweden: Mapping the 2018 swedish general election on twitter, 2018.

[54] Bence Kollanyi and Philip N Howard. Junk news and bots during the german parliamentary election: What are german voters sharing over twitter, 2017.

[55] Sonia Castelo, Thais Almeida, Anas Elghafari, Aécio Santos, Kien Pham, Eduardo Nakamura, and Juliana Freire. A topic-agnostic approach for identifying fake news pages. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 975–980. ACM, 2019.