H2020-ICT-2018-2 /ICT-28-2018-CSA
SOMA: Social Observatory for Disinformation and Social Media Analysis

D3.3 Data Intelligence toolkit description

Project Reference No	SOMA [825469]
Deliverable	D3.3 Data Intelligence toolkit description
Workpackage	WP3 Observatory set-up and operation
Type	Report
Dissemination Level	PU = Public
Date	30/10/2019
Status	V1.0
Editor(s)	Stefano Guarino, LUISS
Contributor(s)	Stefano Guarino, LUISS; Noemi Trino, LUISS
Reviewer(s)	Marina Klitsi, ATC
Document description	The purpose of this report is to provide an overview of design and implementation choices for the upcoming data intelligence toolkit, to be embedded in the SOMA platform.

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
V0.1	28/01/2019	Deliverable concept and structure definition	LUISS
V0.2	21/04/2019	Initial Draft	LUISS
V0.3	27/10/2019	Final Draft	LUISS
V1.0	30/10/2019	Final Version	LUISS

Executive Summary

Operated by the H2020 SOMA Project, the recently established Social Observatory for Disinformation and Social Media Analysis relies on a web-based collaborative platform to support researchers, journalists and fact-checkers in their quest for quality information. This document presents DisInfoNet, one of the technological and algorithmic engines of the SOMA platform. DisInfoNet is designed for non-specialized users and combines elements of text mining and classification with modules for network analysis and visualization to help understand the dynamics of (fake) news dissemination in social media.

In this document, we explain why DisInfoNet fills a much-needed gap in the fact-checking community and we motivate and discuss the design and implementation choices. To provide a practical demonstration of DisInfoNet's potential, we also describe the prototype version of DisInfoNet and report on a use case scenario with Twitter data. Even if several extensions of the DisInfoNet are in the offing, our toolbox is already capable of: (i) tracking relevant news stories and reconstructing their prevalence over time and space; (ii) detecting central debating communities and capturing their distinctive polarization/narrative; (iii) identifying influencers both globally and in specific disinformation networks. When completed, we expect DisInfoNet to be an extremely valuable instrument for investigating the role of disinformation in the social debate.

Table of Contents

List of Figures

List of Tables

1 Introduction

The SOMA platform is aimed at supporting, coordinating and guiding the efforts of researchers, fact-checkers, journalists and society at large in the fight to online and social disinformation. Its main purpose is to provide a shared space that makes available tools and fosters cooperation for the verification of digital (user-generated) content and the analysis of its prevalence in the social debate. To this end, the SOMA platform will follow a collaborative approach to debunking, by making available a special instance of Truly Media[1] and, through that, providing access to a set of softwares (e.g., TruthNest[2]) and tools for the (semi-)automatic identification of information sources and factual claims. The establishment of such a platform requires the systematization – into a user-friendly web-based application – of techniques and tools that are instrumental to monitor the social debate and to track deceptive information.

In this document, we focus on the algorithmic and technological engine of the platform: a data-driven social disinformation toolbox named DisInfoNet that supports the users of the SOMA platform in collecting and processing social media data with the goal of tracking popular and controversial contents, understanding the dynamics of (fake) news dissemination in social media, and identifying the origin and the broadcasters of false information. We present the first prototype of DisInfoNet, we overview its current features and preview its future extensions, and we demonstrate how DisInfoNet can be used in practice for the analysis of a Twitter dataset.

As a general principle, the toolbox is conceived so as to make use of data collected on social media in at least three ways:

1. For identifying evidence of disinformation disorders and assessing the role of temporal and spatial covariates. To this end, the toolbox will allow using text-mining techniques (e.g., keyword-based expressions or text-similarity algorithms) for tracking specific news pieces in the data and visualizing their prevalence over time/space.
2. For analyzing the semantics of shared content and understanding the relevance of disinformation in the social debate. For this purpose, the toolbox will include a tweet/comment/post classifier based on extracting clusters in a (hash) tag co-occurrence graph to be used for training.
3. For understanding the structure of social ties and their impact on (dis)information flows. The toolbox will in fact be endowed with modules for extracting, analyzing and visualizing social interaction graphs, supporting community-detection and user classification.

On a practical side, DisInfoNet is implemented on top of well-known python libraries, and all source code will be made publicly available by the end of the SOMA project. Currently, DisInfoNet can be used by adjusting a configuration file, but the development of a practical GUI is in progress. Additional modules are on the horizon, such as a user-friendly interface for topic modeling, built on top of the powerful and flexible Structural Topic Modeling (Roberts et al. 2014), supporting both sentiment analysis – performed either globally or at the topic level – and n-grams (Guarino and Santoro 2018).

[1] https://www.truly.media/
[2] https://www.truthnest.com/

When DisInfoNet will be fully developed, we expect it to be a fundamental instrument for examining the message strategies used on social media by disinformation players, as a means to share information, build community, and promote action. It will make it possible to:

- measure the impact that these strategies had on audience engagement (e.g., in terms of replies and retweets/repost), on both a temporal and geographical scale;
- understand to what extent the production and sharing of information, community and action posts are driven by renowned authoritative accounts;
- identify elements of agreement/membership or conflict/discussion, and to describe patterns of interaction;
- measure (dis)continuity and cross-ideological exposure, thus assessing the impact of echo-chambers in the success of disinformation campaigns.

In particular, by combining topic modeling, text classification and manual inspection with the structural analysis of the interaction graph, we expect DisInfoNet to bring to the surface the degree of polarization of disinformation communities as well as their internal configuration, their main influencers and leaders.

To demonstrate the usability and the potential of DisInfoNet, in this document we also present a case study: we analyzed a dataset of over 1.3M Italian tweets dated back to November 2016 and hinged on the main subject of debate in Italy at the time: the Constitutional Referendum that was to be held on December 4 2016. The choice of such a dataset relies on three main observations:

- the significant diffusion of fake news in the phase of political campaign before the vote;
- the dichotomic structure of referendums, that fosters user polarization thus making it possible to test the impact of our classifier;
- finally, the relative distance in time of the event, which allows treating disinformation related to a critical political event and prevents the risk of recentism in analyzing social phenomena.

We found evidence of a few relevant false stories in our dataset and, by relating polarization and network analysis, we were able to gain a better understanding of their patterns of production/propagation and contrast and of the role of renowned authoritative accounts and outsiders in driving the production and sharing of information. From a purely quantitative point of view, it is worth noting that our findings diverge significantly from the outcomes of the analysis realised by SOMA partner Pagella Politica at the time[3], underlining once more that Twitter and Facebook provide a very different snapshot of society and that further support of social media platforms is needed for boosting the study of information disorders in social media.

[3] https://pagellapolitica.it/blog/show/148/la-notizia-pi%C3%B9-condivisa-sul-referendum-%C3%A8-una-bufala

2 Background and Desiderata

To start, let us briefly review the state-of-the-art of research in the area, in order to identify desiderata and especially urgent requirements for the pursued social media analysis toolbox. We underline that the purpose of this section was the main goal of Deliverable 3.2: Algorithms of Data Intelligence, Complex Network Analysis, and Artificial Intelligence for the Observatory AI Driven. Some superimpositions between the two documents are therefore unavoidable. However, Deliverable 3.2 contains a much more detailed literature review, whereas in the following we focus on those aspects that are especially relevant for the design and implementation of DisInfoNet.

2.1 Literature Review

As reported by a recent Science Policy Forum article (Lazer et al. 2018), stemming the viral diffusion of fake news and characterizing disinformation networks largely remain open problems. Besides the technical setbacks, the existence of the so-called “continued influence effect of misinformation” is widely acknowledged among socio-political scholars (Skurnik et al. 2005). This socio-psychological phenomenon, consisting in people frequently using inaccurate information in their reasoning even after a credible retraction has been presented, questions the intrinsic potential of debunking in contrasting the proliferation of fake news. Yet, the body of research work on fake news detection and (semi-)automatic debunking is vast and heterogeneous. Linguistics-based techniques, relying on word-frequency patterns (Markowitz and Hancock 2014), deep syntax analysis (S. Feng, Banerjee, and Choi 2012) or semantic analysis (V. W. Feng and Hirst 2013), coexist with network-based techniques, using knowledge networks (e.g., DBpedia) (Ciampaglia et al. 2015), or network-based text analysis such as CRA (Papacharissi and Fatima Oliveira 2012). Attempts at designing an end-to-end fact-checking system exist (Hassan et al. 2017), but are mostly limited to detecting and evaluating strictly factual claims. An alternative approach, followed by the 2016 Fake News Challenge (FNC)[4], is to support professional fact-checkers with tools for automatic stance detection, that is, determining an article’s attitude towards a topic or headline. Unfortunately, even the best submissions to the FNC were later found to mostly capture relatedness rather than agreement/disagreement (Hanselowski et al. 2018). Approaches specifically conceived for measuring the credibility of social media rumours appear to benefit from the combined effectiveness of analyzing textual features, classifying users’ posting and re-posting behaviors, examining external citations patterns, and comparing same-topic messages (Boididou et al. 2018; Castillo, Mendoza, and Poblete 2011; Zubiaga et al. 2018). However, this is well beyond what social media analytics and editorial fact-checking tools on the market permit.

Other studies focused on characterizing the prevalence, the impact and the reasons behind the success of disinformation in social media. Recent work confirmed the general perception that, on average, fake news get diffused farther, faster, deeper and more broadly than true news (Allcott and Gentzkow 2017; Vosoughi, Roy, and Aral 2018). The prevalence of false information is often deemed to be caused by the presence of “fake” and automated profiles, usually called bots (Boshmaf et al. 2013). The role of bots in disinformation campaigns is however far from being sorted out: albeit bots seem to be the main responsible for fake news production and are used to boost the perceived authority of successful (human) sources of disinformation (Bessi and Ferrara 2016), they have been found to accelerate the

[4] http://www.fakenewschallenge.org/

spread of true and false news at the same rate (Vosoughi, Roy, and Aral 2018). Many approaches have been proposed for bot detection, mostly classifiable as either graph-based (e.g., (Paradise, Puzis, and Shabtai 2014)) or machine-learning based (e.g. (Cresci, Di Pietro, et al. 2015)), but none can be considered a definitive solution. Models for explaining the success of false information without a direct reference to bots have also been recently proposed, either based on information overload vs. limited attention (Qiu et al. 2017), or on information theory and (adversarial) noise decoding (Brody and Meier 2018). Finally, investigating the relation between polarization and information spreading has been shown to be instrumental for both uncovering the role of disinformation in a country's political life (Bovet and Makse 2019) and predicting potential targets for hoaxes and fake news (Vicario et al. 2019).

2.2 Desiderata

Based on the above review of research work in the area, we believe that the following features would be highly desirable for enhancing the reach of the SOMA platform.

Data collection. Currently, systematically collecting social and online data is a possibility mostly restricted to users with some non-negligible programming skills. The platform should overcome this limitation by embedding dashboards for collecting data from selected social media and news agencies. This can be done in two ways:

- allowing users to launch a fully new research through a GUI for social media APIs and web crawling;
- by having a data collection routine running in the background to construct and continuously update a SOMA data lake that the user can query.

Debunking. Despite the many technical and socio-psychological issues, debunking remains a fundamental element of a healthy information system. The SOMA platform, whose main purpose is making debunking easier and more effective, could benefit from a wide range of functions capable of supporting debunking in a broad sense. The following seem to be especially useful for the SOMA platform:

- A parser for identifying check-worthy factual claims in a text document.
- A stance detection tool for automatically determining a document's attitude towards a topic or headline.
- A tool that combines some of the aforementioned algorithms, including the data extractor, to search a repository of fact-checks and documents for evidence that supports or refutes a specific claim.

Social graph analysis. Network-oriented analysis is instrumental for understanding the prevalence of misinformation in social media. In particular, the SOMA platform may benefit from both general-purpose tools (that can be used to gain insights into data characterized by a common theme or collected in a precise time frame) and instruments that are particularly relevant for uncovering disinformation campaigns. We identify the following main desiderata:

- A tool for extracting a graph representation from social and news media data. This should be flexible in supporting both hashtags/keywords and users/accounts/authors networks.

The former model the co-occurrence patterns of words, connected based on whether they appeared in the same post/tweet/message/document. The latter model the interplay among subjects, possibly considering several types of interactions, such as friendship, retweeting/reposting, replying, mentioning, writing about the same topic, etc.

- A community-detection tool that returns the community structure of an input graph (possibly, the output graph of the previous tool). Related community-based metrics may also be desirable, such as the well-known Guimerà-Amaral cartography (Guimera and Amaral 2005) that describes the role of the members of each community based on their intra- and inter-community connections.
- Instruments for identifying the key-players of an input graph, based on well-known centrality metrics (degree, PageRank, Betweenness, Closeness) or on measures of influence in a more dynamic sense, for instance through the so-called Linear Threshold (LTM) and Independent Cascade (ICM) models (Hernández and Van Mieghem 2011).
- A social bot meter, potentially combining graph-based and feature-based approaches.
- A tool for tracking and visualizing the path traveled by a specific news piece through sharing, mentioning, retweeting, and other forms of information forwarding in social media.
- A combination of the latter two tools, able to estimate the proportion of bots vs. humans that contributed to the diffusion of a news piece.

Text analysis. Besides all aforementioned purpose-specific algorithms, a few more generic text analysis tools would allow gaining a better understanding of the discussion going on on social media about a theme of interest, and of the polarization of users involved in disinformation campaigns. We especially envisage:

- A tool for user-friendly topic modeling of a text corpus (Blei 2012). Topic modeling returns a thematic organization of the corpus that highlights the topics discussed in the selected documents, the keywords characterizing each topic, and the main topics treated in each document.
- A tweet/comment/post polarization/sentiment classification tool.

3 The Designing of DisInfoNet

As emerged in Section 3, preventing disinformation to enter the news stream is not only difficult, but possibly useless, unethical and maybe even illegal. Along this line, DisInfoNet is not designed for filtering information, but rather for empowering individuals to defend from a storm of heterogeneous and often unreliable news pieces.

The importance of data collected on social media in understanding disinformation disorders (Bovet and Makse 2019) is self-evident, and especially paramount in the following tasks:

- elaborating quantitative analysis of the diffusion of unreliable news stories (Allcott and Gentzkow 2017);
- comprehending the relevance of disinformation in the social debate, possibly incorporating thematic, polarity or sentiment classification (Vosoughi, Roy, and Aral 2018);
- unveiling the structure of social ties and their impact on (dis)information flows (Bessi and Ferrara 2016).

Among all desiderata listed in Section 3.2 we therefore believe that the following features are especially urgent for our DisInfoNet to help researchers, journalists and fact-checkers characterizing the prevalence and dynamics of disinformation on social media:

- tracking specific news pieces in the data and visualizing their prevalence over time/space;
- classifying content in a semi-automatic fashion (relying on clustering a keyword/hashtag co-occurrence graph)
- extracting, analyzing and visualizing social interaction graphs, embedding community-detection and user classification.

That being said, in the following paragraphs we consider all requirements and report on the extent to which DisInfoNet will implement them. For those that will not be supported, we motivate our choice and provide pointers to further resources the SOMA platform may alternatively rely upon.

Data collection. DisInfoNet will surely include a dashboard for collecting data from selected social media and news agencies. Both options suggested in Section 3.2 have pros and cons. Allowing on-demand searches would potentially permit highly customized queries and guarantee richer and more up-to-date results, but at the cost of a less trivial interface and of longer processing time. On the other hand, implementing a data collection routine would make querying the obtained datalake easier and faster, but the results would be limited to the available data and the data gathering/scraping process would require a careful tuning (e.g., accounting for research criteria that vary based on recent events) to produce useful and manageable datasets. To date, we tested a tweepy-based Python scraper for Twitter data that accepts a list of keywords and returns a variable-length list of tweets containing at least one of such keywords. We are currently working on an extension of the scraper that considers at least Reddit and Facebook data (as much as possible through Facebook's Graph API), as well as popular news/media websites.

Debunking. The three elements we suggested as our debunking desiderata are basically the main modules of the ClaimBuster (Hassan et al. 2017) system. However, the current status of the ClaimBuster initiative makes clear that only specific sub-tasks can be reliably addressed at the moment, with all others being mostly open problems in the research community. In this context, developing a fully-working system from scratch seems well beyond the scope and the reach of the SOMA project. On the contrary, for the tasks that appear to be feasible, relying on APIs or other open-source code is a much more affordable and practical solution. Specifically, we believe that the SOMA platform would greatly benefit from embedding the following:

- ClaimBuster's claim spotter module;
- the PyTorch implementation of the FEVER pipeline baseline for fact extraction and verification[5];
- the FEVER system for document retrieval, sentence retrieval, natural language inference and aggregation[6].

That being said, previous work shows that both identifying factual claims and detecting a document's stance towards a claim can be modeled as classification tasks and addressed using supervised learning. In case the SOMA project decides to develop a prototype or to test/extend existing implementations, we collected a few sources of training data for claim extraction/verification. These datasets include:

- Claimbuster's debates datasets[7];
- FEVER's datasets[8];
- BuzzFeed data[9] related to a report on partisan Facebook pages[10];
- the FNC dataset (headlines and articles annotated with one of four classes: "unrelated", "agree", "disagree", "discuss")[11];
- the Emergent dataset[12];
- the Stanford Natural Language Inference (SNLI) Corpus[13].

Social graph analysis. This is the part of DisInfoNet that we consider both urgent and feasible, and for this reason these will be the founding modules of the toolbox. We already motivated on the importance of network analysis tools. While many tools exist for network analysis, however, only commercial tools are conceived for the average user with no or limited programming skills, but these tools are not though for tracking disinformation and offer no customization options – other than being very expensive. We will therefore implement a thorough library for graph extraction, analysis and visualization, mostly based on the igraph Python library. This will support all most known and used instruments of network analysis, including community detection, centralities, and vertex classification. The results will be returned with an interactive interface thought for non-expert users. For instance, buttons will be used for switching from influencers (high closeness and/or pagerank) to bridges (high betweenness centrality) when key-players are visualized. In all cases that involve working with a graph, several

[5] https://github.com/sheffieldnlp/fever-naacl-2018
[6] https://github.com/uclmr/fever
[7] https://idir.uta.edu/claimbuster/debates
[8] http://fever.ai/resources.html
[9] https://www.kaggle.com/mrisdal/fact-checking-facebook-politics-pages
[10] https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis#.ia1QB2KJl
[11] https://github.com/FakeNewsChallenge/fnc-1
[12] https://drive.google.com/drive/folders/0BwPdBcatuO0vYTAxSnA1d09qdGM
[13] https://nlp.stanford.edu/projects/snli/

formats will be supported. For instance, edge-lists, csv files, or raw json data as provided by most web and social media APIs.

Importantly, the library will support the inclusion, in both analysis and visualization tools, of information coming from other modules on DisInfoNet. This includes, for instance, making the polarization of social media profiles with respect to a topic of interest visible in the plots of an interaction graph. Additionally, by specifying one or more news pieces of interest, the user will be allowed to only focus on the subgraph induced by the messages that discuss that topic, or to track and visualize, on a temporal scale, the path traveled by these news pieces through sharing, mentioning, retweeting, etc. To this end, it must be stressed that these functionalities can be seen as a special case of the database querying tool discussed in Section 3.2, thus the quality and comprehensiveness of the results will be strongly dependent on the quality of the available datalake.

For the same reasons delineated above with respect to debunking, implementing a social bot meter from scratch makes little sense. For prototyping, supervised learning would be the chosen paradigm, with the dataset[14] and the benchmark provided in (Cresci, Di Pietro, et al. 2017) available for training, testing and evaluation. However, a probably more viable alternative to devising a novel algorithm is to hook the SOMA platform to the APIs provided by Botometer, the most performing machine-learning based bot detector to date (Davis et al. 2016; Ferrara 2017). Once a reliable bot-meter will be embedded in the platform, bot statistics will also be available in the network analysis tool.

Text analysis. Finally, text analysis tools will be included in DisInfoNet as indicated in the desiderata. In particular, topic modeling and sentiment analysis will be available through a module for extended Structural Topic Modeling (STM), implemented in R on top of the STM[15] library. This will include:

- An automated process for selecting the most suitable number of topics based on well-known metrics such as coherence and perplexity.
- A sentiment analysis module that can be run independently to establish the sentiment of single documents and infer the sentiment distribution of the corpus, or together with STM to also establish a sentiment distribution for each topic.
- Several options for specifying vocabulary, stopwords, synonyms, and covariates to be used for content and semantic analysis.

Let us underline that, since topic modeling is computationally intensive, real-time results cannot realistically be expected. On the other hand, the module will be completely automatic, with the user being only expected to provide or select the data of interest and, possibly, the number of topics to identify. For what concerns polarization/sentiment analysis, it is worth mentioning that this is a strongly application dependent issue and designing a tool that works with heterogeneous data and provides reliable results is substantially an open problem in the research literature. However, many middle-ground solutions can be considered, such as unsupervised clustering based on document distance (Kusner et al. 2015). A possible semi-automatic approach is already included in the current prototype of DisInfoNet and is described in detail below.

[14] mib.projects.iit.cnr.it/dataset.html

[15] https://github.com/bstewart/stm

4 The DisInfoNet Prototype

At a high level, DisInfoNet is a Python library built on top of well-known packages (e.g., igraph, scikit-learn, NumPy, Gensim). It provides modules for managing archives, elaborating and classifying text, building and analyzing graphs, and more. It is memory-efficient to support large datasets and, albeit a few functions are optimized for Twitter data, generally flexible.

DisInfoNet also implements a pipeline designed to enable journalists and fact-checkers with no coding expertise assessing the prevalence of disinformation in social media data. This pipeline, depicted in Figure 1, consists of three main tools, described in the following. It can be fed a CSV file or the (possibly GZ-compressed) JSON file returned by the Twitter search or streaming API. One of DisInfoNet's main features is the ability to extract and examine both keyword co-occurrence graphs and user interaction graphs induced by a specific set of themes of interest, thus providing valuable insights into the contents and the actors of the social debate around disinformation stories.

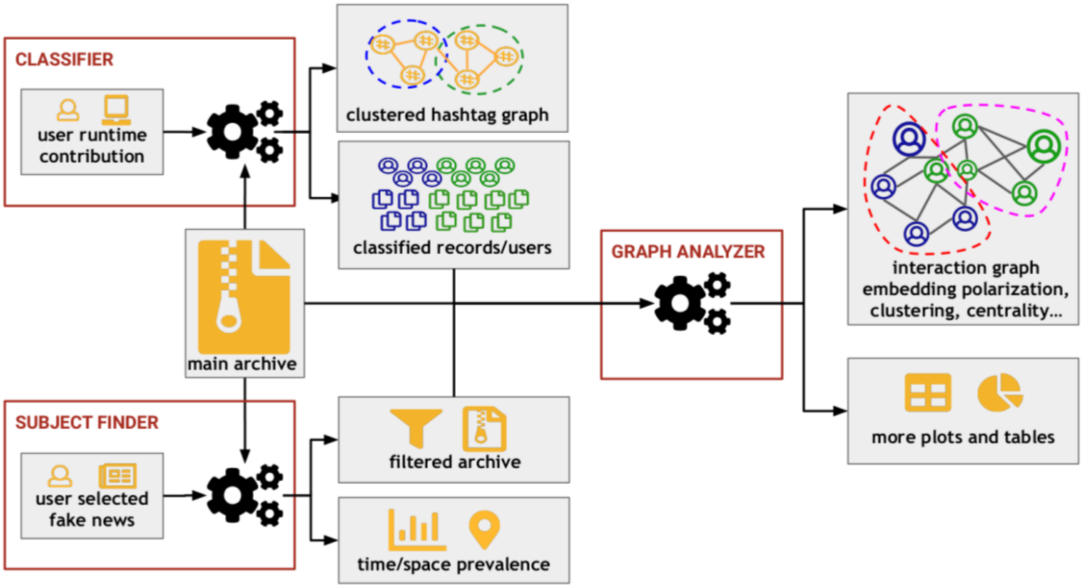
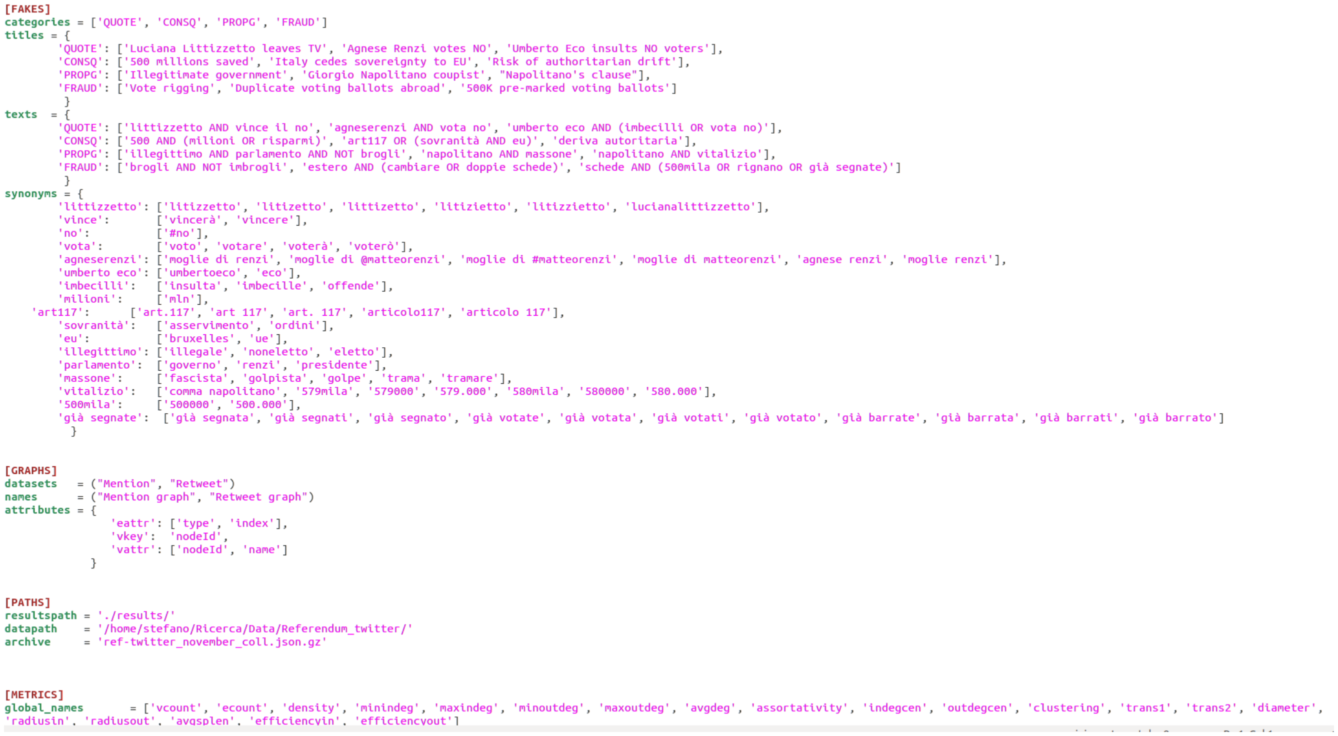


Figure 1: DisInfoNet's main pipeline

As DisInfoNet is thought for non-developers, it can be used by simply adjusting a configuration file, where the user may specify several parameters, including: paths to data; which of the modules she/he wants to run; what are the news pieces she/he wants to look for in the data; what type of interaction graph she/he wants to build and analyze; which centrality measure and community detection algorithm she/he wants to use and visualize in the plots. An excerpt of the configuration file is shown in Figure 2. A user-friendly dashboard will be later made available on the SOMA project's website to overcome the need for manually modifying the configuration file. When sufficiently mature, all source-code, documentation and code snippets needed to access the data will be hosted on a gitlab repository and released under the GPL.



```
[FAKES]
categories = ['QUOTE', 'CONSQ', 'PROPG', 'FRAUD']
titles = {
        'QUOTE': ['Luciana Littizzetto leaves TV', 'Agnese Renzi votes NO', 'Umberto Eco insults NO voters'],
        'CONSQ': ['500 millions saved', 'Italy cedes sovereignty to EU', 'Risk of authoritarian drift'],
        'PROPG': ['Illegitimate government', 'Giorgio Napolitano coupist', "Napolitano's clause"],
        'FRAUD': ['Vote rigging', 'Duplicate voting ballots abroad', '500K pre-marked voting ballots']
         }
texts  = {
        'QUOTE': ['littizzetto AND vince il no', 'agneserenzi AND vota no', 'umberto eco AND (imbecilli OR vota no)'],
        'CONSQ': ['500 AND (milioni OR risparmi)', 'art117 OR (sovranità AND eu)', 'deriva autoritaria'],
        'PROPG': ['illegittimo AND parlamento AND NOT brogli', 'napolitano AND massone', 'napolitano AND vitalizio'],
        'FRAUD': ['brogli AND NOT imbrogli', 'estero AND (cambiare OR doppie schede)', 'schede AND (500mila OR rignano OR già segnate)']
         }
synonyms = {
        'littizzetto': ['litizzetto', 'litizetto', 'littizetto', 'litizietto', 'litizzietto', 'lucianalittizzetto'],
        'vince':       ['vincerà', 'vincere'],
        'no':          ['#no'],
        'vota':        ['voto', 'votare', 'voterà', 'voterò'],
        'agneserenzi': ['moglie di renzi', 'moglie di @matteorenzi', 'moglie di #matteorenzi', 'moglie di matteorenzi', 'agnese renzi', 'moglie renzi'],
        'umberto eco': ['umbertoeco', 'eco'],
        'imbecilli':   ['insulta', 'imbecille', 'offende'],
        'milioni':     ['mln'],
    'art117':      ['art.117', 'art 117', 'art. 117', 'articolo117', 'articolo 117'],
        'sovranità':   ['asservimento', 'ordini'],
        'eu':          ['bruxelles', 'ue'],
        'illegittimo': ['illegale', 'noneletto', 'eletto'],
        'parlamento':  ['governo', 'renzi', 'presidente'],
        'massone':     ['fascista', 'golpista', 'golpe', 'trama', 'tramare'],
        'vitalizio':   ['comma napolitano', '579mila', '579000', '579.000', '580mila', '580000', '580.000'],
        '500mila':     ['500000', '500.000'],
        'già segnate':  ['già segnata', 'già segnati', 'già segnato', 'già votate', 'già votata', 'già votati', 'già votato', 'già barrate', 'già barrata', 'già barrati', 'già barrato']
          }


[GRAPHS]
datasets   = ("Mention", "Retweet")
names      = ("Mention graph", "Retweet graph")
attributes = {
                'eattr': ['type', 'index'],
                'vkey':  'nodeId',
                'vattr': ['nodeId', 'name']
             }


[PATHS]
resultspath = './results/'
datapath    = '/home/stefano/Ricerca/Data/Referendum_twitter/'
archive     = 'ref-twitter_november_coll.json.gz'


[METRICS]
global_names       = ['vcount', 'ecount', 'density', 'minindeg', 'maxindeg', 'minoutdeg', 'maxoutdeg', 'avgdeg', 'assortativity', 'indegcen', 'outdegcen', 'clustering', 'trans1', 'trans2', 'diameter',
'radiusin', 'radiusout', 'avgsplen', 'efficiencyin', 'efficiencyout']
```

Figure 2: An excerpt of the configuration file used to control the DisInfoNet prototype.

Subject Finder The Subject Finder filters a Twitter dataset and returns information about the prevalence of a theme or news piece of interest in the dataset. Currently, it relies on an user-provided keyword-based boolean expression, but extensions based on word embeddings and document similarity measures are on top of the list of upcoming features. The tool can be used to obtain a list of relevant tweets together with a number of covariates, such as author, timestamp, geo-localization, retweet count, hashtags, mentions. This information can also be aggregated at the user level in order to directly obtain a list of users that are especially active on the theme or just mentioned when others talk about it. The module outputs a number of data structures, including:

- a CSV file containing all texts that satisfied the query with an arbitrary set of covariates (e.g., author, timestamp, retweet count);
- the set of all timestamps of texts that satisfied the query;
- the set of all available gps positions for texts that satisfied the query together with statistics on the proportion of available locations.

The distribution of all tweets and disinformation-related tweets over time may as well be plotted. Similarly, the Subject Finder allows creating an interactive html map where geo-localized tweets may be visualized and explored.

Classifier The Classifier is based on a “self-training” process in which a list of hashtags associated with two (or more) classes of interest are used to automatically extract a training set. The Classifier then trains both a Logistic Regression and a Gradient Boosting Classifier, it uses 10-fold cross-validation and it returns the overall best performing model. Significantly, the list of hashtags is derived in a semi-automatic fashion: a hashtag graph is extracted from the available dataset by connecting any two hashtags that co-occur at least once in the same tweet and weighing the edge by the total count of such co-occurrences; after the user is given the chance to prune a few central (high pagerank) but generic and/or out-of-context hashtags (that could be detrimental to identifying clear and meaningful clusters), community detection is performed on the graph; finally, the user is presented with an excerpt of these clusters for manual inspection and she/he can pick and label any of these communities as representative of a specific class. As a result, the classifier can use as many as 10/100/100× more hashtags for classification than with any realistic fully manual approach, without sacrificing accuracy and possibly bringing to light previously unknown highly discriminative hashtags. For tweet representation the Classifier may use tf-idf vectors, doc2vec (Le and Mikolov 2014) or a combination of both. When only two classes are used (which is often enough: Republican vs. Democratic or right vs. left wing in a political debate; pro vs. against in a referendum debate; discussing political views vs. environmental consequences in the recent Amazon fires debate, etc.), the obtained classification can be extended to users by averaging over the classification of all tweets produced by a specific user.

Graph Analyzer The Graph Analyzer extracts an interaction graph from a Twitter dataset, i.e., a directed graph in which accounts are connected based on their patterns of interaction. Any combination of the following are admitted: (i) $user_a$ → $user_b$ if $user_a$ mentions $user_b$ in any of his/her tweets; (ii) $user_a$ → $user_b$ if $user_a$ replies to any of $user_b$'s tweets; (iii) $user_a$ → $user_b$ if $user_a$ retweets any of $user_b$'s tweets. Significantly, the Graph Analyzer may also build the subgraphs induced by the specified themes or stories fed to the Subject Finder, i.e., the graphs obtained by only considering the tweets that satisfy the corresponding query. For each extracted graph, the Graph Analyzer allows:

- Extracting a global description of the graph through a set of metrics including radius, diameter, average distance, assortativity, degree distribution, clustering coefficient, directed transitivity, eccentricity (M. Newman, Barabasi, and Watts 2006).
- Computing the importance of all vertices of the graph through centrality metrics such as degree, PageRank, Betweenness, Closeness (Hernández and Van Mieghem 2011).
- Partitioning the graph into communities relying on the well-known Louvain (Blondel et al. 2008) or Leading Eigenvector (M. E. Newman 2006) algorithms.
- Applying the Guimerà-Amaral cartography (Guimerà and Nunes Amaral 2005) to establish the role of each node within its community based on the volume of inter- and intra-community connections.
- Producing a number of plots, including the (possibly fitted) distributions of degrees and other metrics, an interactive plot of any portion of the graph, the community graph, the Guimerà-Amaral cartography.

The Graph Analyzer may also read and/or dump two files representing the graph, a node list and an edge list, based on the options specified in the configuration file.

5 DisInfoNet in Action

Finally, let us demonstrate the potential of DisInfoNet by using the currently available prototype version to analyze a dataset of more than 1.3M Italian tweets produced in November 2016 and discussing the upcoming Constitutional Referendum. A complete contextualization of this case-study is provided in a recently accepted conference paper (Guarino, Trino, et al. 2019). In the following, we will show that DisInfoNet allows shedding light on the dynamics of social disinformation as Italy approached the Referendum. Broadly speaking, we aim at unveiling patterns in the spreading of (dis)information by understanding how specific stories reach their (intended?) target and identifying the actors of disinformation within and across communities.

Following the literature, in order to identify the main topics and themes of disinformation of the political campaigning we relied on the activity of fact-checking and news agencies who reported lists of fake news that went viral during the referendum campaign. Mostly based on the work by fact-checking web portal Bufale.net (Mastinu 2016), online newspaper Il Post (Post 2016), and SOMA partner and political fact-checking agency Pagella Politica (Politica 2016), we were able to collect a number of potentially interesting pieces of (dis)information related to the referendum. We ended up considering twelve stories, which can be broadly classified into four categories, reported below. These stories include both general theories and very specific news pieces, allowing us to study disinformation at different levels of granularity. Significantly, this type of classification-based approach is fully supported by DisInfoNet and easily available through the configuration file.

The first category includes entirely fabricated content, involving made-up quotes of famous people public endorsing one or the other faction or defaming voters of the other side:

QUOTE – made-up quotes of popular public figures

1. showgirl/comedian Luciana Littizzetto saying she will leave television if the NO wins;
2. Matteo Renzi’s wife Agnese Renzi saying she will vote NO despite her husband being the main promoter of the Constitutional Reform ;
3. deceased novelist/critic/semiotician Umberto Eco calling NO voters idiots. It has to be noted that Umberto Eco had died earlier that same year, long before the referendum was announced.

The second group of news can be labelled as manipulated content, as it contains manipulated interpretation of genuine information (about the reform):

CONSQ – presumed consequences of the YES/NO victory

1. the claim that a victory of the YES would guarantee over €500M of public saving every year;
2. the claim that a victory of the YES would make Italy yield national sovereignty to EU institutions (especially referring to a hidden clause in art.117);
3. the claim that a victory of the YES would cause a shift towards authoritarianism.

The third category includes news inserted in a typical populist frame, opposing people vs the élite:

PROPG – anti-establishment propaganda

1. the claim that the parliament and/or the government and/or the premier have been illegitimately elected;
2. the claim that YES-supporter former president Giorgio Napolitano is a Freemason orchestrating a coup d'état;
3. the claim that the reform contains a "comma Napolitano" guaranteeing a rich life annuity to former presidents;

A final category involves the integrity of the electoral process and gaining unauthorized access to voting machines and altering voting results:

FRAUD – alleged reports of frauds

1. the fear that the government (and, thus, the YES side) was organizing vote rigging;
2. the claim that duplicate voting cards were prepared for abroad voters in order to manipulate their vote;
3. the news story of 500K YES pre-marked voting card having been found;

To widen the scope of the analysis, we considered all the news, theories and topics of discussion that could be associated with information disorders in its broader sense. This includes deliberate and organised disinformation and propaganda, unintentionally propagated misinformation, but also rumors, hearsays, clickbaits items and conspiracy theories, often used by the two sides to accuse one another. However, we are not interested here in discussing the political implications of our findings. Providing the reader with details on the set of fake-news considered is only instrumental for discussing the results of applying DisInfoNet.

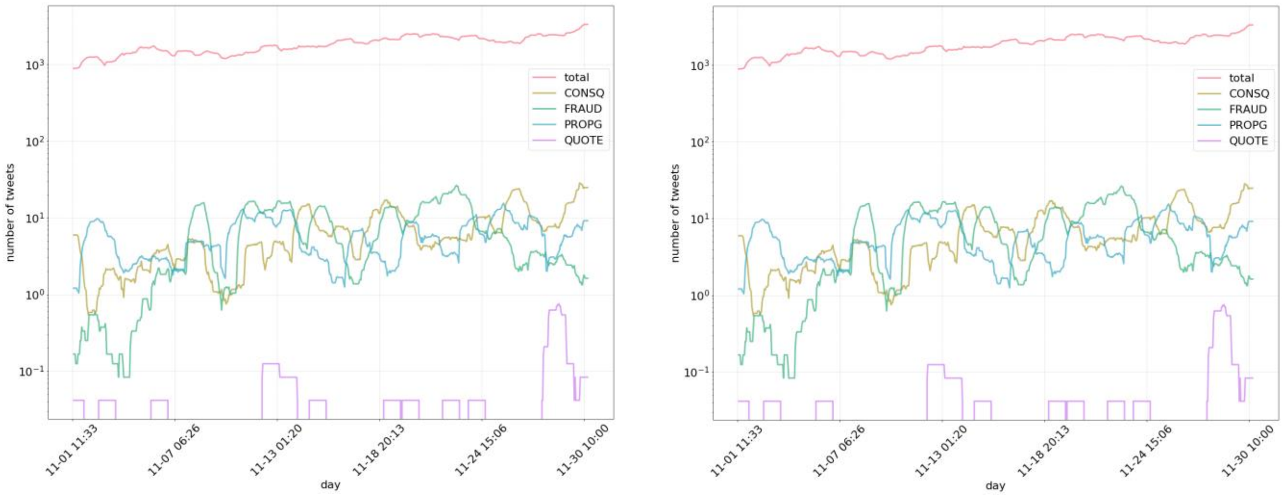
5.1 The Subject Finder: Disinformation Prevalence

With the set of news stories defined above represented by a suitable set of keyword-based queries, we ran the Subject Finder to get tweets related to each of these twelve stories (an example query is reported in Table 1). We also labelled these tweets with the four classes. We extracted timestamps and locations to study the time and space distribution of tweets in our dataset.

Table 1: The condition for PROPG (1) news story.

('illegittimo' OR 'illeggittimo' OR 'illegal' OR 'non eletto') AND ('parlamento' OR 'governo' OR 'renzi' OR 'presidente')

In figures 3a and 3b we report, respectively, the distribution (one-day rolling mean) of the four aforementioned classes and of the four most relevant news pieces across November 2016. In both cases, these distributions are compared with the overall trend. In general, we see a limited presence of disinformation in the dataset. However, while the share of QUOTE tweets is almost negligible, each of the other three classes accounts for ≈ 5% of the total. The volume of discussion about fake/conspiracy news stories does not seem to simply increase at the approach of the referendum as for the general discussion. Different stories have different surges, that an informed journalist may relate with events (e.g., a politician giving an interview) or with the activity of some influencer.



(a) By class.

(b) By story (top 4).

Figure 3: The time distribution of disinformation tweets compared with the total.

Regarding the geography of the debate, we first measured the proportion of geo-localized tweets in the dataset, reported in Table 2. We observe that overall only 29716 tweets – that is, 2.21% of the whole dataset – are geo-tagged, and that this percentage is significantly lower among disinformation tweets, suggesting that users involved in this type of discussions are generally more concerned about privacy than the average user. The map, reported in Figure 4, shows some activity in Great Britain and the Benelux area, but disinformation topics appear to be substantially absent outside Italy.

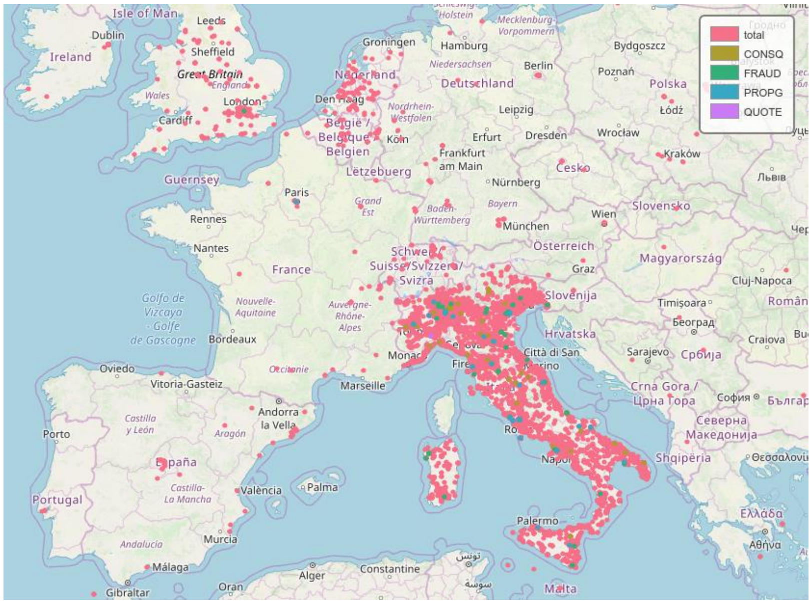


Figure 4: The geographic distribution of tweets for the four classes (compared with the total).

Table 2: Number and percentage of geo-localized tweets.

class	tweets	geo-tags (%)
total	1344216	29716 (2.21%)
CONSQ	7909	71 (0.90%)
PROPG	4345	47 (1.08%)
FRAUD	5362	69 (1.29%)
QUOTE	57	1 (1.75%)

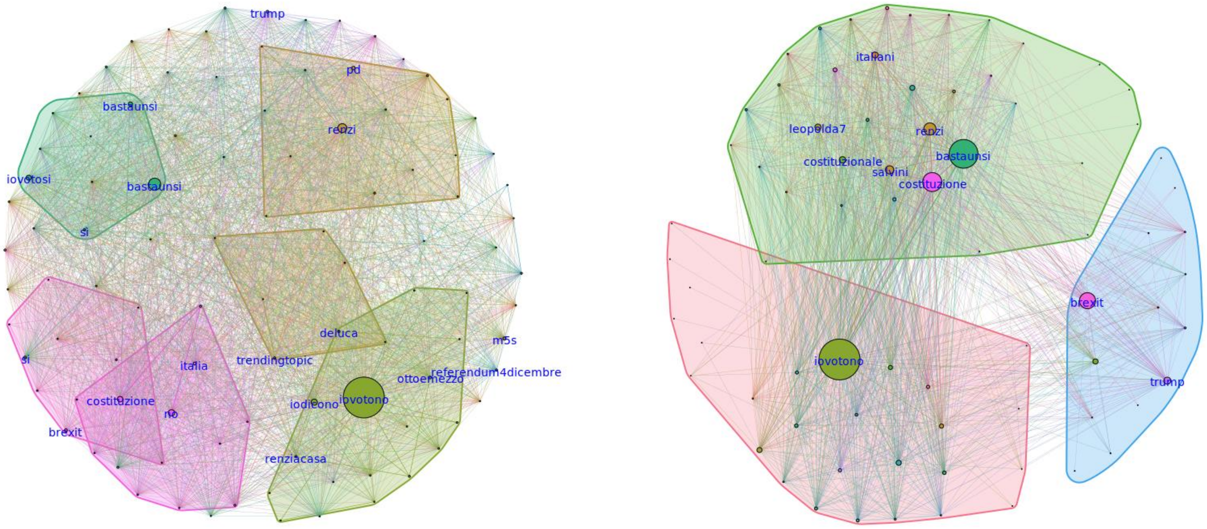
5.2 The Classifier: Polarization and Disinformation

The Classifier can now be used to gain a better understanding of users in our dataset and the relation between polarization and disinformation. Specifically, we used the semi-automatic self-training process as follows. First, we let the Classifier build a hashtag graph and return the top 30 hashtags by weighted degree (i.e., with the greatest number of co-occurrences with other hashtags). We pruned a few irrelevant or uninformative hashtags (“referendum”, “referendumcostituzionale”, “photo”, “riformacostituzionale”, “costituzione”, “4dicembre”, “trendingtopic”, “1w1l”) and let the Classifier run Louvain’s algorithm and plot both the graph and the cluster-graph that shows connections among communities. These two graph are reported in Figure 5 and show that: (i) hashtags used by the NO and YES supporters are strongly clustered; (ii) a third cluster-of-clusters emerges composed of hashtags probably used by international reporters; (iii) some hashtags are surprisingly central in these communities, such as “ottoemezzo” which is the name of a popular and supposedly impartial tv show discussing current political events. In particular, the two largest clusters of hashtags clearly characterize the two sides: the NO cluster is dominated by the hashtags "#bastaunsì" ("a yes is enough") and "#iovotosi" ("I vote yes"), whereas the YES cluster by "#iovotono" ("I vote no"), "#iodicono" ("I say no") and "#renziacasa" ("Renzi go home").

By interacting with the Classifier, we selected the two main clusters as the base sets of hashtags to be used for extracting a training set composed of tweets labelled as follows:

-1 / **NO** if the tweet only contains hashtags from the NO cluster;
+1 / **YES** if the tweet only contains hashtags from the YES cluster;
0 / **UNK** if the tweet contains a mix of hashtags from the NO and the YES cluster.

We used doc2vec features only and a Gradient Boosting Classifier was automatically selected as the best performing classifier. Significantly, we also obtained a continuous score in [-1,1] for users, since a user is classified with the average score of his/her tweets.



(a) The hashtag graph, with clusters highlighted. Vertex size is by PageRank.

(b) The cluster graph, with clusters-of-clusters highlighted. Vertex size is by cluster size.

Figure 5: The hashtag graph obtained while training the classifier.

The Classifier can be also set to plot a histogram that helps relating polarization and disinformation. In our case, this plot, reported in Figure 6, shows that:

- UNK tweets are substantially negligible – although this may be due to limitations of the classifier (indeed, cross-validation shows that the UNK class is the hardest to recognize);
- on the whole, NO tweets are almost 1.5x more frequent than YES tweets, supporting the diffused belief that the NO front was significantly more active than its counterpart in the social debate;
- disinformation news stories mostly follow the general trend, but: (i) topics of the QUOTE and PROPG class, which gathers attack vectors frequently used by the populist parties, are

especially popular among NO supporters (hence, debunking efforts are invisible); (ii) on the other hand, YES supporters are more active than the average in the CONSQ topics, probably due to the concurrent attempts at promoting the referendum and at tackling the fears of potential NO voters.

5.3 The Graph Analyzer: Interaction Graphs and Disinformation

Among the three supported types of interactions, we decided to only focus on retweets. Retweeting is in fact very easy – a lot easier than writing a new post (Kantrowitz 2019) – and is commonly interpreted as a form of endorsement, thus being a popular tool for promoting ideas and campaigns, and for community building, even relying on semi-automatic accounts.

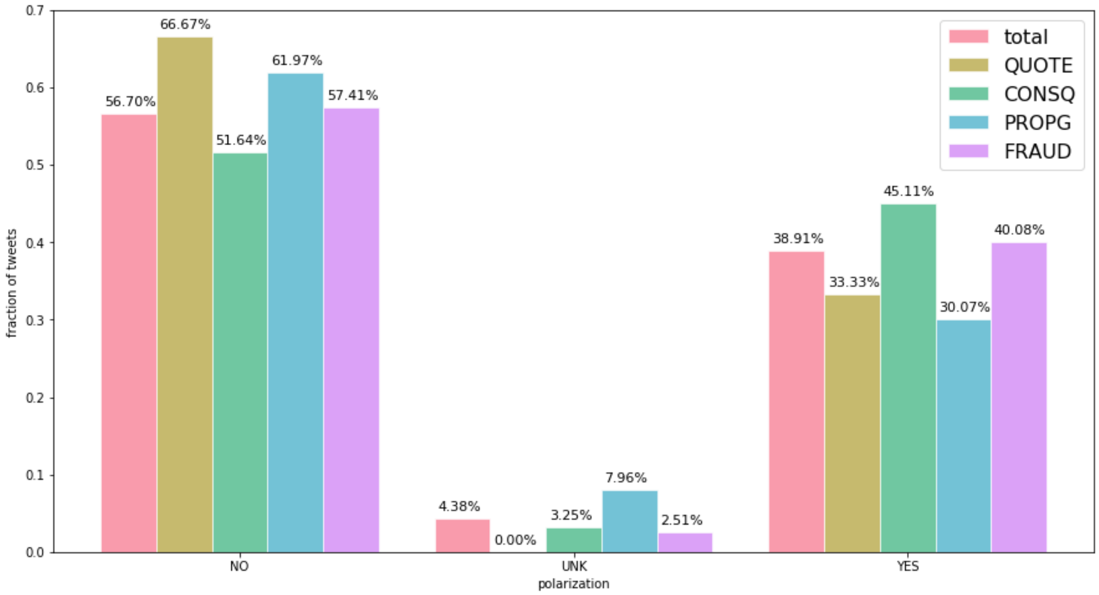
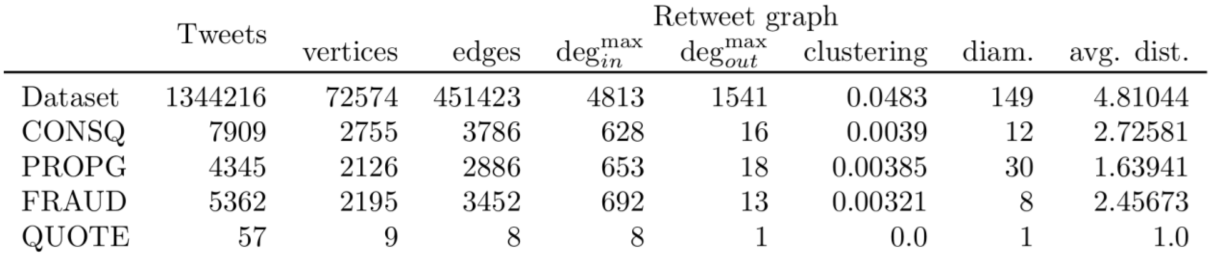


Figure 6: The polarization of tweets, in total and for the four disinformation classes.

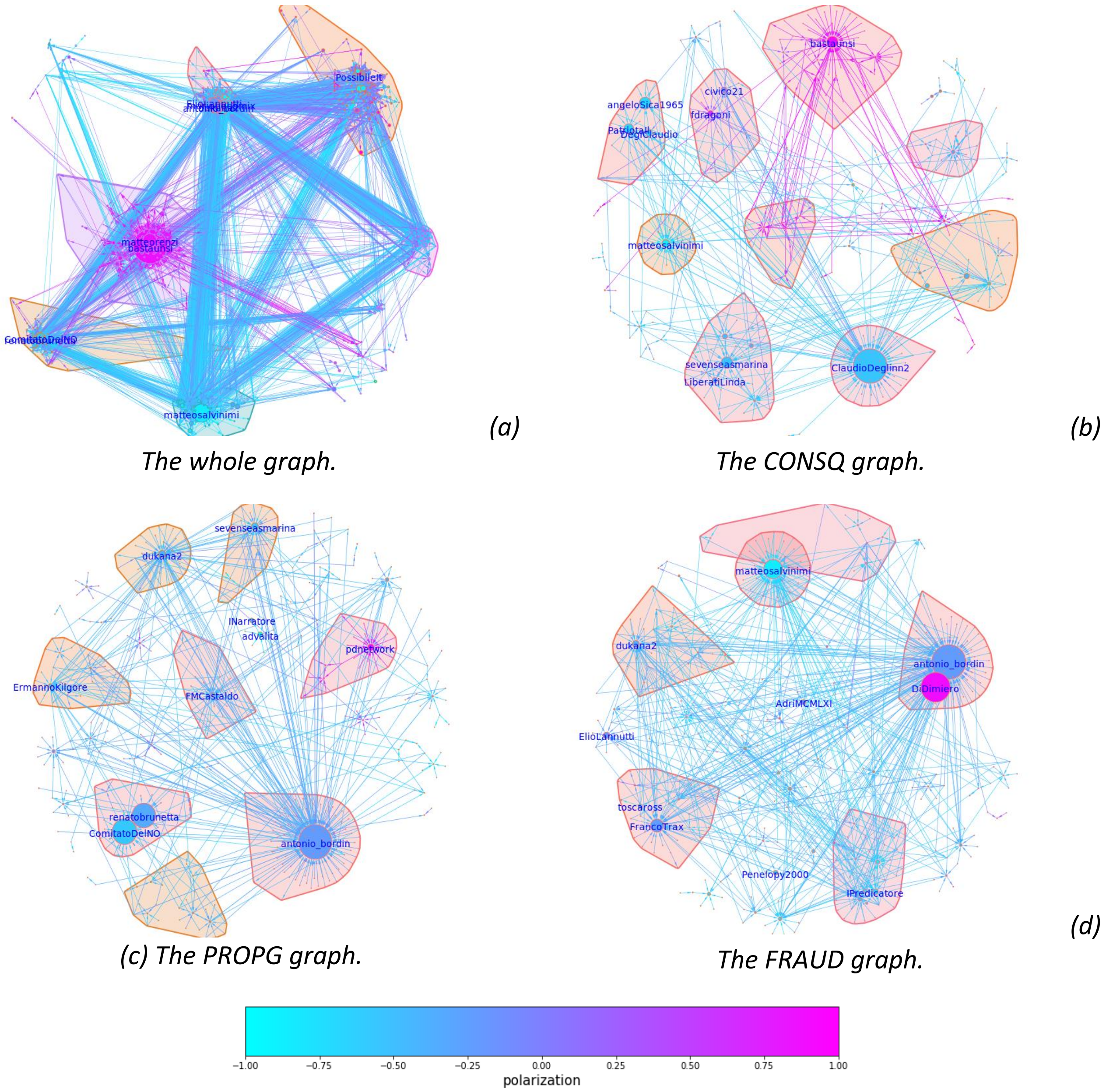
In Table 3 we report on the global information returned by the Graph Analyzer on the studied graphs. These macroscopic descriptors yield insights into the structural similarities and differences of the four graphs. For instance, the CONSQ and PROPG are similar in size (2755 vertices and 3786 edges vs. 2126 and 2886) and have similarly sized in- and out-hubs (628 and 16 vs. 653 and 18), but the diameter of the CONSQ graph is significantly smaller (12 vs. 30) despite it having a larger average distance (2.73 vs. 1.64). These numbers suggest that PROPG disinformation stories travelled less on average, but were sporadically able to reach very peripheral users. Additionally, we see that the clustering coefficient of the two graphs is almost identical and rather small (≈ 0.004), more than one order of magnitude smaller that the clustering coefficient of the whole graph. This suggests that these disinformation networks may not be “self-organizing” and their structure might be governed by artificial diffusion patterns.

Table 3: Dataset overview with interaction graph metrics.



	Tweets	Retweet graph						
		vertices	edges	$\deg_{in}^{max}$	$\deg_{out}^{max}$	clustering	diam.	avg. dist.
Dataset	1344216	72574	451423	4813	1541	0.0483	149	4.81044
CONSQ	7909	2755	3786	628	16	0.0039	12	2.72581
PROPG	4345	2126	2886	653	18	0.00385	30	1.63941
FRAUD	5362	2195	3452	692	13	0.00321	8	2.45673
QUOTE	57	9	8	8	1	0.0	1	1.0

In Figure 7 we instead show the 500 top users by pagerank for the whole graph and the subgraphs induced by the CONSQ, PROPG and FRAUD disinformation classes – the QUOTE class being barely present in the dataset and therefore scarcely informative. In these plots, users are colored by their polarity, obtained from the Classifier, and edges take the average color of the connected vertices. The size of a vertex is proportional to its pagerank, whereas the width of an edge to its weight (i.e., number of interactions). These plots highlight a number of interesting aspects. First of all, the NO front appears to be generally dominant. This is especially true in the PROPG and FRAUD graphs, whereas in the CONSQ graph YES supporters come back into action. Also, there seems to be limited interaction between YES and NO supporters, as can be noted by the fact that edges almost always link vertices of similar or even identical color. Additionally, when considering the whole graph the main vertices are well known politics or political parties, whereas in the disinformation graphs other users emerge. These users vary at least partially from graph to graph and include accounts that cannot be immediately linked to any known public figure/influencer.



(a) *The whole graph.*

(b) *The CONSQ graph.*

(c) The PROPG graph.

(d) *The FRAUD graph.*

Figure 7: 500 top users by PageRank. Color is by polarity, size by PageRank

5.4 Comments

Using the DisInfoNet prototype on our use case dataset was fairly easy but yielded a number of valuable insights. First of all, we gained an overview of the prevalence of different disinformation topics in the social debate on the Referendum. We then learned that the disinformation items that had higher and broader diffusion on our Twitter dataset are related to conspiracy themes and anti-élite sentiment. We were also able to correlate disinformation and polarization, in such a way to map the considered topics to the levels of activities of different communities, and to measure the volume of cross-ideological interactions.

Focusing on single profiles, among the leaders of the NO front we found well-known public figures (e.g., politicians Renato Brunetta and Fabio Massimo Castaldo in the PROPG graph) along with accounts not associated with any publicly known individual. To test the interplay of DisInfoNet with other tools available through the SOMA platform, we used TruthNest to obtain reports and analytics on the usage patterns of specific accounts, summarized into a bot-likelihood score. In most cases, we learned that these apparently common users are indeed militants of the NO front, sometimes having multiple aliases, and whose activity is characterized by a high number of retweets and mentions of well-known actors belonging to the same community (e.g., Antonio Bordin, Claudio Degl'Innocenti, Angelo Sisca, Liberati Linda). In particular, one of the most influential nodes of the PROPG graph, @INarratore, came out having a suspiciously high 60% bot-score, other than only 1% of original tweets and a considerable number of "suspicious followers". In the same graph, @dukana2 has a 50% bot-score, while the account @advalita has been suspended from Twitter. In the CONSQ graph, the most central user is @ClaudioDeglinn2, characterized by a relatively low 10% bot-score, but apparently in control of at least other 7 aliases and strongly connected with other amplification accounts. Two of these "amplifiers" are especially noteworthy: @IPredicatore, having a 40% bot-score, and @PatriotaII, having a 30% bot-score, mentioning @ClaudioDeglinn2 in more than 20% of his tweets, and producing only 3% original tweets. Altogether, we seem to have found indicators of coordinated efforts to avoid bot detection tools while reaching peripheral users and expanding the network.

6 Conclusion

In this document, we presented the plan for an integrated toolbox for monitoring social disinformation, conceived as part of the H2020 Social Observatory for Disinformation and Social Media Analysis. While the desirable features of a similar platform are numerous, we identified those that are both urgent and feasible – based on recent research advancements – and on which most efforts should hence be focused. For all desiderata that came out being hardly achievable within the scope of the SOMA project, we instead proposed alternative approaches based on open source libraries and existing tools with publicly available APIs. The resulting design of the DisInfoNet toolbox delineates a framework capable of providing a wide spectrum of users with instruments for quantifying the prevalence of disinformation and understanding its dynamics of diffusion on social media, relying on well-established techniques for text and graph mining. The toolbox will soon be available online and extended in the next future.

To make the description of DisInfoNet and its usage clearer, we also presented a prototype version of the toolbox, reporting both a high-level description of its operational pipeline and a detailed analysis of the three main constituent modules. Finally, we presented a case study analysis focused on the 2016 Italian constitutional referendum. We found evidence of a correlation between users' polarization and participation to different disinformation campaigns, and by highlighting the primary actors of disinformation production and propagation we could manually tell apart public figures, activists and potential bots. This shows the immediate relevance that our DisInfoNet may have on the daily work of researchers, journalists and fact-checkers.

7 References

1. Allcott, H. and M. Gentzkow (2017). “Social media and fake news in the 2016 election”. In: Journal of Economic Perspectives 31.2, pp. 211–36.
2. Bessi, A. and E. Ferrara (2016). “Social bots distort the 2016 US Presidential election online discussion”. In: First Monday 21.11-7.
3. Blei, D. M. (Apr. 2012). “Probabilistic Topic Models”. In: Commun. ACM 55.4, pp. 77–84. issn: 0001-0782. doi: 10.1145/2133806.2133826. url: http://doi.acm.org/10.1145/2133806.2133826.
4. Blondel, V. D. et al. (2008). “Fast unfolding of communities in large networks”. In: Journal of Statistical Mechanics: Theory and Experiment 2008.10, P10008.
5. Boididou, C. et al. (2018). “Verifying information with multimedia content on twitter”. In: Multimedia Tools and Applications 77.12, pp. 15545–15571.
6. Boshmaf, Y. et al. (2013). “Design and analysis of a social botnet”. In: Computer Networks 57.2, pp. 556–578.
7. Bovet, A. and H. A. Makse (2019). “Influence of fake news in Twitter during the 2016 US presidential election”. In: Nature communications 10.1, p. 7.
8. Brody, D. C. and D. M. Meier (2018). “How to model fake news”. In: arXiv preprint arXiv:1809.00964.
9. Castillo, C., M. Mendoza, and B. Poblete (2011). “Information credibility on twitter”. In: Proceedings of the 20th international conference on World wide web. ACM, pp. 675–684.
10. Ciampaglia, G. L. et al. (2015). “Computational fact checking from knowledge networks”. In: PloS one 10.6, e0128193.
11. Cresci, S., R. Di Pietro, et al. (Dec. 2015). “Fame for sale: efficient detection of fake Twitter followers”. In: Decision Support Systems 80, pp. 56–71. issn: 0167-9236. doi: http://dx.doi.org/10.1016/j.dss.2015.09.003.
12. Cresci, S., R. Di Pietro, et al. (2017). “The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race”. In: Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, pp. 963–972.
13. Davis, C. A. et al. (2016). “Botornot: A system to evaluate social bots”. In: Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, pp. 273–274.
14. Feng, S., R. Banerjee, and Y. Choi (2012). “Syntactic stylometry for deception detection”. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, pp. 171–175.
15. Feng, V. W. and G. Hirst (2013). “Detecting deceptive opinions with profile compatibility”. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 338–346.
16. Ferrara, E. (2017). “Disinformation and social bot operations in the run up to the 2017 French presidential election”. In: First Monday 22.8.
17. Guarino, S. and M. Santoro (2018). “Multi-word Structural Topic Modelling of ToR Drug Marketplaces”. In: 2018 IEEE 12th International Conference on Semantic Computing (ICSC). IEEE, pp. 269–273.
18. Guarino, S., N. Trino, et al. (2019). “Beyond Fact-Checking: Network Analysis Tools for

Monitoring Disinformation in Social Media”. In: International Conference on Complex Networks and their Applications, to appear.

19. Guimera, R. and L. A. N. Amaral (2005). “Cartography of complex networks: modules and universal roles”. In: Journal of Statistical Mechanics: Theory and Experiment 2005.02, P02001.
20. Guimerà, R. and L. Nunes Amaral (2005). “Functional cartography of complex metabolic networks”. In: Nature 433.895.
21. Hanselowski, A. et al. (2018). “A Retrospective Analysis of the Fake News Challenge Stance Detection Task”. In: arXiv:1806.05180.
22. Hassan, N. et al. (2017). “ClaimBuster: the first-ever end-to-end fact-checking system”. In: Proceedings of the VLDB Endowment 10.12, pp. 1945–1948.
23. Hernández, J. M. and P. Van Mieghem (2011). “Classification of graph metrics”. In: Delft University of Technology: Mekelweg, The Netherlands.
24. Kantrowitz, A. (2019). The Man Who Built The Retweet: “We Handed A Loaded Weapon To 4-Year-Olds”.www.buzzfeednews.com/article/alexkantrowitz/how-the-retweet-ruined-the-internet. [accessed 05-Aug-2019].
25. Kusner, M. et al. (2015). “From word embeddings to document distances”. In: International conference on machine learning, pp. 957–966.
26. Lazer, D. M. et al. (2018). “The science of fake news”. In: Science 359.6380, pp. 1094–1096.
27. Le, Q. and T. Mikolov (2014). “Distributed representations of sentences and documents”. In: International conference on machine learning, pp. 1188–1196.
28. Markowitz, D. M. and J. T. Hancock (2014). “Linguistic traces of a scientific fraud: The case of Diederik Stapel”. In: PloS one 9.8, e105937.
29. Mastinu, L. (2016). TOP 10 Bufale e disinformazione sul Referendum. www.bufale.net/top-10-bufale-e-disinformazione-sul-referendum/. [accessed 05-Jul-2019].
30. Newman, M. E. (2006). “Finding community structure in networks using the eigenvectors of matrices”. In: Physical review E 74.3, p. 036104.
31. Newman, M., A.-L. Barabasi, and D. J. Watts, eds. (2006). The Structure and Dynamics of Networks. Princeton University Press.
32. Papacharissi, Z. and M. de Fatima Oliveira (2012). “Affective news and networked publics: The rhythms of news storytelling on# Egypt”. In: Journal of Communication 62.2, pp. 266–282.
33. Paradise, A., R. Puzis, and A. Shabtai (2014). “Anti-reconnaissance tools: Detecting targeted socialbots”. In: IEEE Internet Computing 18.5, pp. 11–19.
34. Politica, R. P. (2016). La notizia più condivisa sul referendum? È una bufala. https://pagellapolitica.it/blog/show/148/la-notizia-pi\%C3\%B9-condivisa-sul-referendum-\%C3\%A8-una-bufala. [accessed 05-Jul-2019].
35. Post, R. I. (2016). Nove bufale sul referendum. www.ilpost.it/2016/12/02/bufale-referendum/. [accessed 05-Jul-2019].
36. Qiu, X. et al. (2017). “Limited individual attention and online virality of low-quality information”. In: Nature Human Behaviour 1.7, p. 0132.
37. Roberts, M. E. et al. (2014). “Structural topic models for open-ended survey responses”. In: American Journal of Political Science 58.4, pp. 1064–1082.
38. Skurnik, I. et al. (2005). “How warnings about false claims become recommendations”. In: Journal of Consumer Research 31.4, pp. 713–724.
39. Vicario, M. D. et al. (2019). “Polarization and fake news: Early warning of Potential misinformation targets”. In: ACM Transactions on the Web (TWEB) 13.2, p. 10.

40. Vosoughi, S., D. Roy, and S. Aral (2018). “The spread of true and false news online”. In: Science 359.6380, pp. 1146–1151.
41. Zubiaga, A. et al. (2018). “Detection and resolution of rumours in social media: A survey”. In: ACM Computing Surveys (CSUR) 51.2, p. 32.