



Disinfonet Luiss Data Lab

Disinfonet
User Manual



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825469.



Disinfonet User Manual



Go to

<https://soma-disinfonet.luiss.it/>



Disinfonet

Summary

- Disinfonet
- Dataset Management Homepage
- Dataset List
- Dataset Upload
- Dataset Overview
- Filter Section
- Create New Filter
- Classifier Section
- Create New Classifier
- Analysis Result: Histogram Prevalence
- Analysis Result: Histogram Polarization
- Analysis Result: Time Distribution
- Analysis Result: Map
- About
- Privacy and Policy

Disinfonet

The platform **DisInfoNet** provides the ability to upload datasets taken from Twitter, in the [JSON Tweet Format](#), and carry out analyzes on them.

It is possible to create **Filters** to isolate, manage and analyze a fake news or a set of related fake news through dynamic statistics and visualizations.

You can create one or more binary classifiers to cluster a dataset. **Classifier** is based on a semi-automatic “self-training” process, in which a list of hashtags associated with two classes of interest are used to automatically extract a training set.

For each dataset, several **interactive data visualizations** are automatically generated, describing the quantitative, temporal and spatial distribution of the data.

Disinfonet - Dataset Management Homepage

A brief introduction to the DisInfoNet prototype is available on the dataset management homepage and in the main page of all single dataset overviews.

Disinfonet

Operated by the H2020 SOMA Project, the recently established Social Observatory for Disinformation and Social Media Analysis supports researchers, journalists and fact-checkers in their quest for quality information. At the core of the Observatory lies the DisInfoNet Toolbox, designed to help a wide spectrum of users understand the dynamics of (fake) news dissemination in social networks. DisInfoNet combines text mining and classification with graph analysis and visualization to offer a comprehensive and user-friendly suite that allows users to:

1. Track relevant news stories and reconstruct their prevalence over time and space;
2. Detect central debating communities and capture their distinctive polarization/narrative;
3. Identify influencers both globally and in specific “disinformation networks”.

Dataset list

Dataset Name	Status
--------------	--------

Disinfonet - Dataset list

In the **Dataset list** section you can view the list of datasets currently available in the system. For each dataset it is possible to view its name and current status.

Each dataset is pre-processed to optimize performance in terms of space and time used; when a dataset is in the "*complete*" state it is ready for analysis.

Dataset list

Dataset Name	Status
Referendum November	complete
amazzonias_twitter	complete
7244_Cambiamento climatico - hashtag	complete
Qanon Eng	complete

Disinfonet - Dataset upload

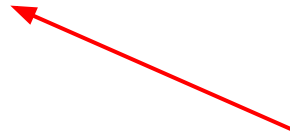
Under the Dataset List section, there is the Import New Dataset section, through which you can insert a new dataset inside the system. To do this, follow the next instructions:

1. click on the button “*browse...*” at the end of the page to select a dataset from your own filesystem; the dataset must be a JSONEachRow in the [JSON Tweet Format](#), compressed through gzip.

Import new dataset

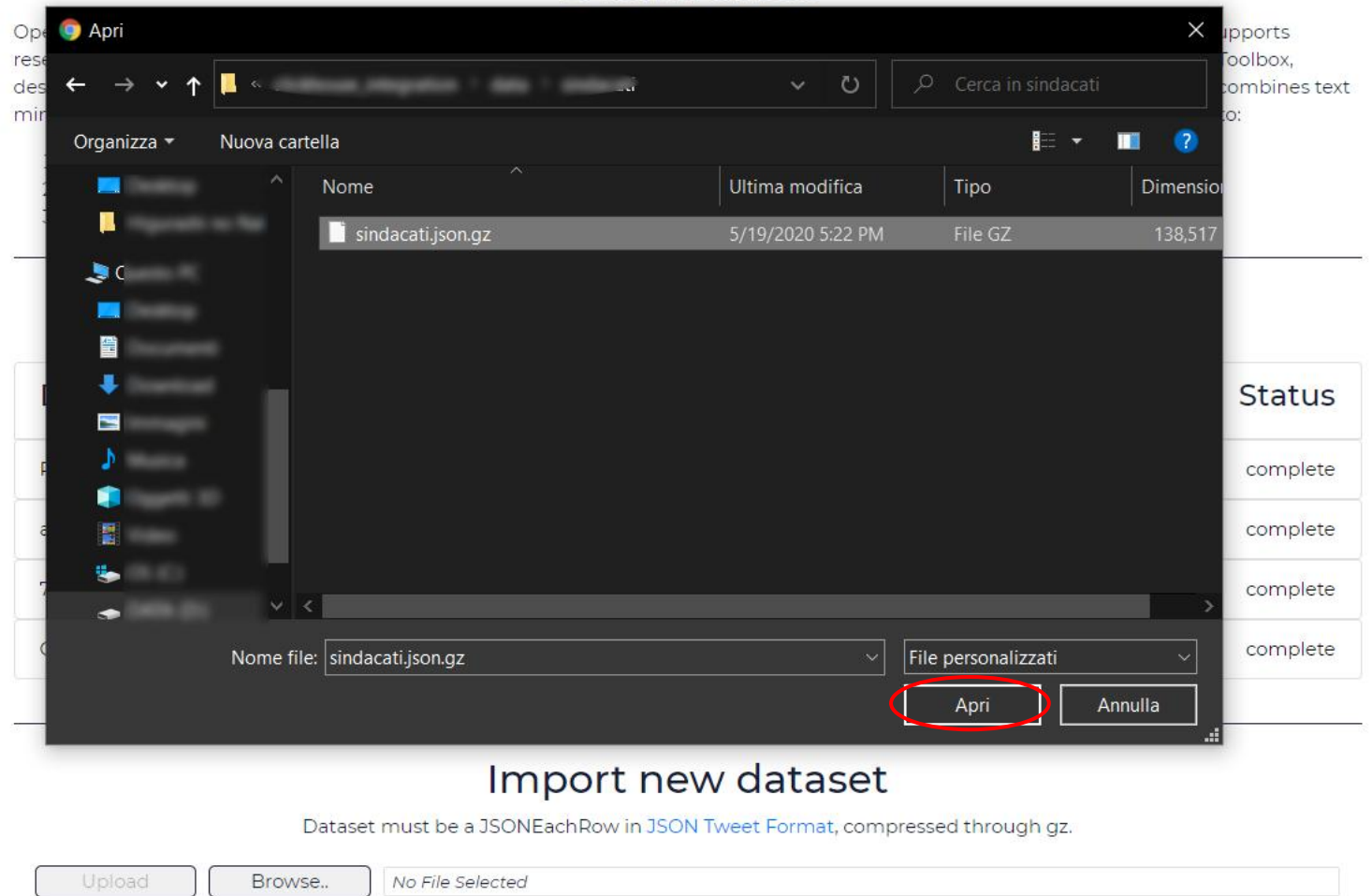
Dataset must be a JSONEachRow in [JSON Tweet Format](#), compressed through gz.

No File Selected



Disinfonet - Dataset upload

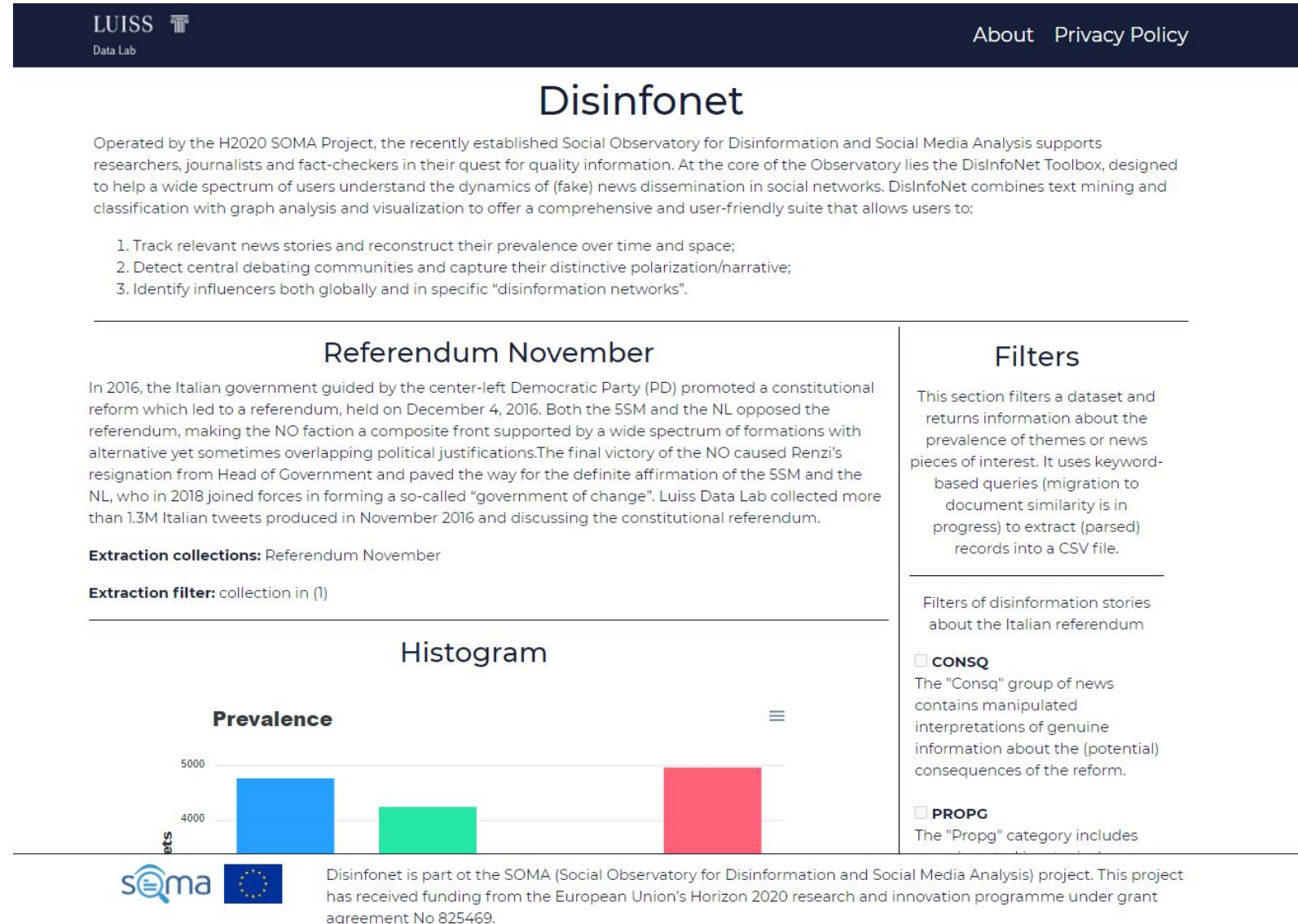
2. After choosing a file, click on the download button and wait until the dataset is transferred to the server. The new dataset will appear in the list of dataset with the status *"loading"*.
3. After a few minutes (depends on the dataset dimensions), the status of the dataset will go to complete and it will be ready to perform new analysis on it.



Disinfonet - Dataset Overview

To see the analysis carried out on a dataset, search on the main page the dataset of interest and click on it (if its state is “*complete*”).

You will be redirected to the **Dataset Overview** page with all the analyzes and results made for this specific dataset. The top of the page contains a description of the SOMA project and the DisInfoNet platform. The first column contains all the data visualization, while the second one contains the filters and classifier management interface.



Disinfonet - Filter Section

In the right-side of the results web page of a dataset, there is a **Filter** section where you can manage the created filters related to this specific dataset.

A **Filter** is a set of queries that is used to isolate tweets about a disinformation topic. Each query is a set of words connected through a logic of AND/OR/NOT, that represent a fake news.

From here, you can see a description of each existing filter or delete/create new ones.

The screenshot shows the 'Filters' section of the Disinfonet interface. It contains a list of filters with their descriptions and actions. A blue curved arrow points from the 'Filters' title to the 'PROPQ' filter description. Another blue curved arrow points from the 'CONSQ' filter description to the 'Delete Filters' button.

Filters

This section filters a dataset and returns information about the prevalence of themes or news pieces of interest. It uses keyword-based queries (migration to document similarity is in progress) to extract (parsed) records into a CSV file.

Filters of disinformation stories about the Italian referendum

- ☐ **PROPQ**
The "Propq" category includes news inserted in a typical populist frame, opposing people vs the elite.
- ☐ **QUOTE**
The "Quote" category includes entirely fabricated quotes of public figures endorsing one or the other faction or defaming voters of the other side.
- ☐ **FRAUD**
The "Fraud" category involves the integrity of the electoral process, gaining unauthorized access to voting machines and altering voting results.
- ☐ **CONSQ**
The "Consq" group of news contains manipulated interpretations of genuine information about the (potential) consequences of the reform.

Delete FiltersCreate New Filters

Disinfonet - Create New Filter

To create a new **Filter**, select a dataset in the main page and then click the button “New Filters”, in the section Filters of the results page. You will be redirected to a new page with a form with this fields:

- **tag**: an identifier for the new filter;
- **description**: a short description for the new filter;
- **title**: a short description for the following query;
- **query**: AND/OR/NOT logic query; the query follows Google’s logic research (e.g. space between two words is meaning AND, etc..)

Create new Filter

Tag (the main topic of the analysis, composed by one or more filters):	Description (a brief description of the main topic of the analysis):
<input type="text" value="Tag"/>	<input type="text" value="Description"/>
Tag is required	
Title (the filter of the analysis):	Query (the exact word or group of words to filter the dataset):
<input type="text" value="Title"/>	<input type="text" value="Query"/>
Title is required	Query is required
	Queries use the Google's logic research
<input type="button" value="Add Query"/>	<input type="button" value="Save Filter"/>
	<input type="button" value="Back Results"/>

Disinfonet - Create New Filter

For each word in the query, it is possible to add a list of synonyms, separate by a comma, that will be used during the research.

Through the button “Add Query”, is possible to add a new title and query for the current filter.

The button “Save Filter” stores the just created filter, with all associated queries, on db.

With the button “Back Result” you can go back to result without saving the new filter.

Create new Filter

Tag (the main topic of the analysis, composed by one or more filters):

Covid 5g

Description (a brief description of the main topic of the analysis):

Fakes about Coronavirus and 5g

Title (the filter of the analysis):

5g causes coronavirus

Query (the exact word or group of words to filter the dataset):

5g coronavirus

[Queries use the Google's logic research](#)

Synonyms

5g:

Synonyms

coronavirus:

covid19, covid

Add Query

Save Filter

Back Results

Disinfonet - Classifier Section

A **Classifier** is an algorithms that allows you to assign a label to each tweet, based on a semi-automatic binary polarization process, clustering the dataset into belonging groups.

In the right-side of the results web page of a dataset, under the Filter section, there is a **Classifier** section, where you can manage the created classifiers. From here, you can see a description of each label of each existing classifier and check the current status of it. If the status is “*complete*”, you can see the results of the analyzes on that classifier; If the status is “*computing*”, this means that the classification process is still ongoing.

Through the “*create new classifier*” button you can access the page for creating a new classifier for the current dataset.

Classifier

This section partitions records into classes based on a semiautomatic “self-training” process. By building and clustering a keyword cooccurrence graph, it presents the user with an excerpt of the keywords associated with the obtained classes.

☒ Classifier 1

Status: complete

yes: Fakes tweets about referendum of november 2018 that are pro yes.

no: Fakes tweets about referendum of november 2018 that are pro no.

New
Classifier

Disinfonet - Create New Classifier

To create a new **Classifier**, select a dataset in the main page and then click the button “New Classifier” in the section Classifier of the results page. You will be redirected to a new page with all the instructions for a new classifier. Through the button “Back Result” you can go back to result without saving the new classifier.

Create new Classifier

Classifier is based on a semi-automatic “self-training” process”, in which a list of hashtags associated with two classes of interest are used to automatically extract a training set. The classifier works according to a three-step process:

1. It builds and cluster a keyword co-occurrence graph, that presents the user with an excerpt of the keywords associated with the obtained classes.
2. It provides the user with the possibility to prune of central (high pagerank) yet generic and/or out-of-context keywords, detrimental to clustering.
3. After performing community detection on the graph, 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.

Back Results

Disinfonet - Create New Classifier

The classification process and the creation of a new classifier is composed by the following steps:

1. An hashtag graph of the dataset is constructed and the pageranks are computed to order the nodes; Now it is possible to select and delete, among the 30 top hashtags, the misleading ones to clean the dataset.

Top Hashtag

These are the top 30 hashtag. Now you can clean the dataset deleting any hashtag that not interest you. Click on "Finish" when you're done.

- | | | | |
|--|---|---------------------------------------|--|
| <input type="checkbox"/> #iovotono | <input type="checkbox"/> #referendum | <input type="checkbox"/> #renzi | <input type="checkbox"/> #referendumcostituzionale |
| <input type="checkbox"/> #bastaunsi | <input type="checkbox"/> #iodicono | <input type="checkbox"/> #4dicembre | <input type="checkbox"/> #iovotosi |
| <input type="checkbox"/> #no | <input type="checkbox"/> #riformacostituzionale | <input type="checkbox"/> #ottoemezzo | <input type="checkbox"/> #costituzione |
| <input type="checkbox"/> #pd | <input type="checkbox"/> #m5s | <input type="checkbox"/> #portaaporta | <input type="checkbox"/> #salvini |
| <input type="checkbox"/> #iostoconmarino | <input type="checkbox"/> #si | <input type="checkbox"/> #noino | <input type="checkbox"/> #deluca |
| <input type="checkbox"/> #boschi | <input type="checkbox"/> #movimentonesti | <input type="checkbox"/> #bastaunsi | <input type="checkbox"/> #renziacasa |
| <input type="checkbox"/> #piazzapulita | <input type="checkbox"/> #lariachetira | <input type="checkbox"/> #si | <input type="checkbox"/> #nonrubo |
| <input checked="" type="checkbox"/> #trump | <input type="checkbox"/> #italia | | |

Finish

Disinfonet - Create New Classifier

2. The hashtag graph is rebuilt without the deleting ones selected in the previous phase, and a clustering algorithm is applied. A new user interface is shown and you can select 2 clusters from those computed, to polarize all tweets between this. For each cluster you must specify a label and a description.
3. Now the training/prediction process can start. The new defined classifier can be seen in the Classifier section on the result page, with state computing.

After that the training/predict process is terminated (it can last several hours depending on the size of the dataset), the classifier passes in the state complete and the related analysis can be viewed.

Cluster selection

These are the top cluster of dataset. Choose 2 cluster for classification and specify labels for them. Click on "Finish" when you're done.

☐

["referendum","no","costituzione","brexit","si"]

☒

["referendumcostituzionale","bastaunsi","iovotesi","riformacostituzionale","italia"]

Yes

All tweets pro yes.

☒

["iovetono","iodicono","4dicembre","m5s","renziacasa"]

No

All tweet pro no.

☐

["renzi","bastaunsi","pd","riforma","si"]

☐

["ottoemezzo","lariachetira","portaaporta","piazzapulita","tagadala7"]

☐

["lavoro","jobsact","scuola","labuonascuola","buonascuola"]

☐

["leopolda7","leopolda2016","firenze","leopolda","facciaafaccia"]

☐

["beatricedimaio","renzie","brunetta","laprocuraindaga","iacoboni"]

☐

["trendingtopic","idoneifantasma","abilitatitfa","odproroga2018","8000esiliatifaseb"]

☐

["mentana","basta","stato","rai3","calderoli"]

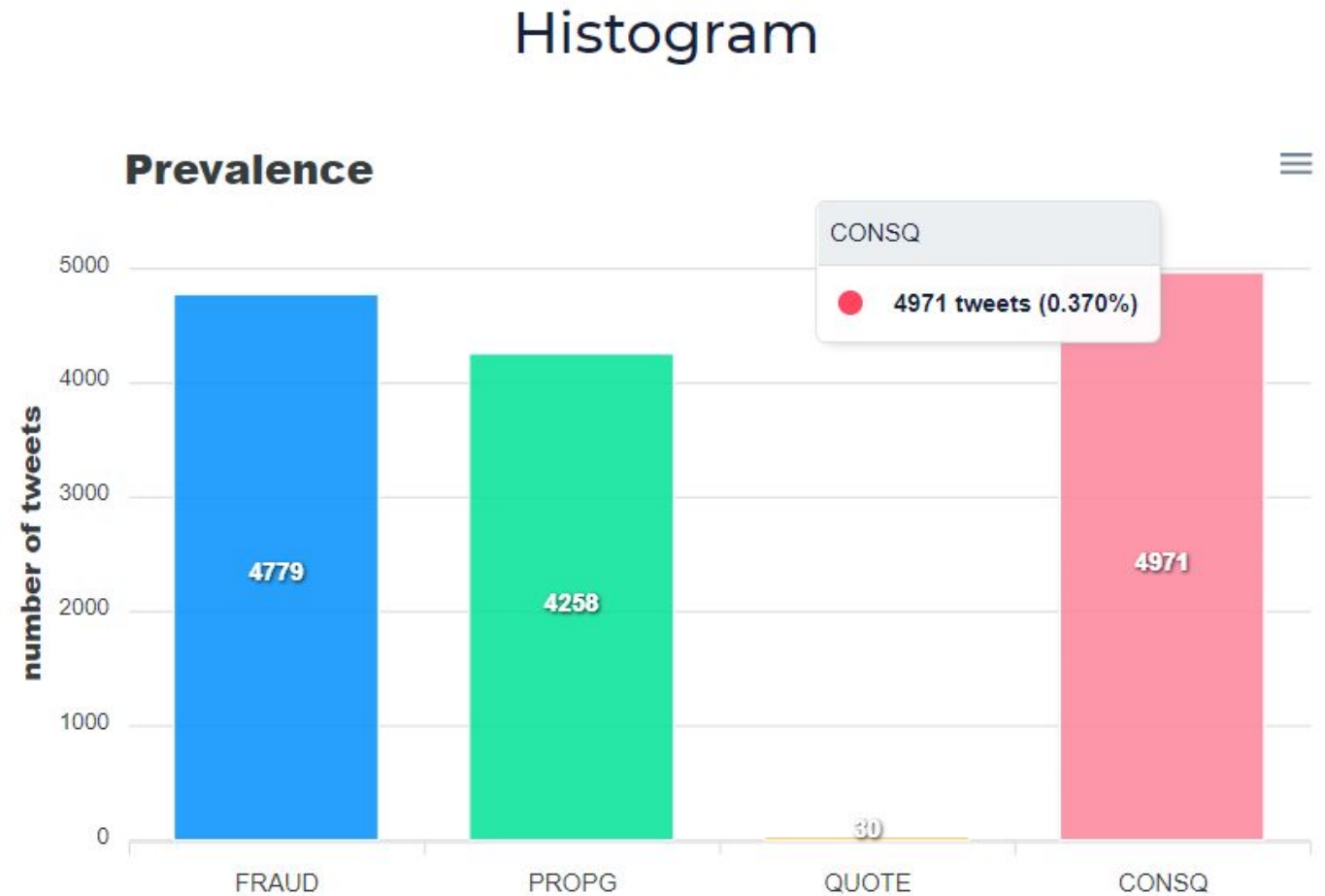
Finish

Disinfonet - Analysis Result: Histogram Prevalence

An interactive histogram that highlights the prevalence of each group of filtered tweets.

For each filter it is possible to view the quantity and percentage of the total prevalence.

Using the drop-down menu at the top right of the graph, you can download a screenshot of the graph in .svg or .png format, or the .csv of the result data.



The figure provides the prevalence of all available filters in the dataset.

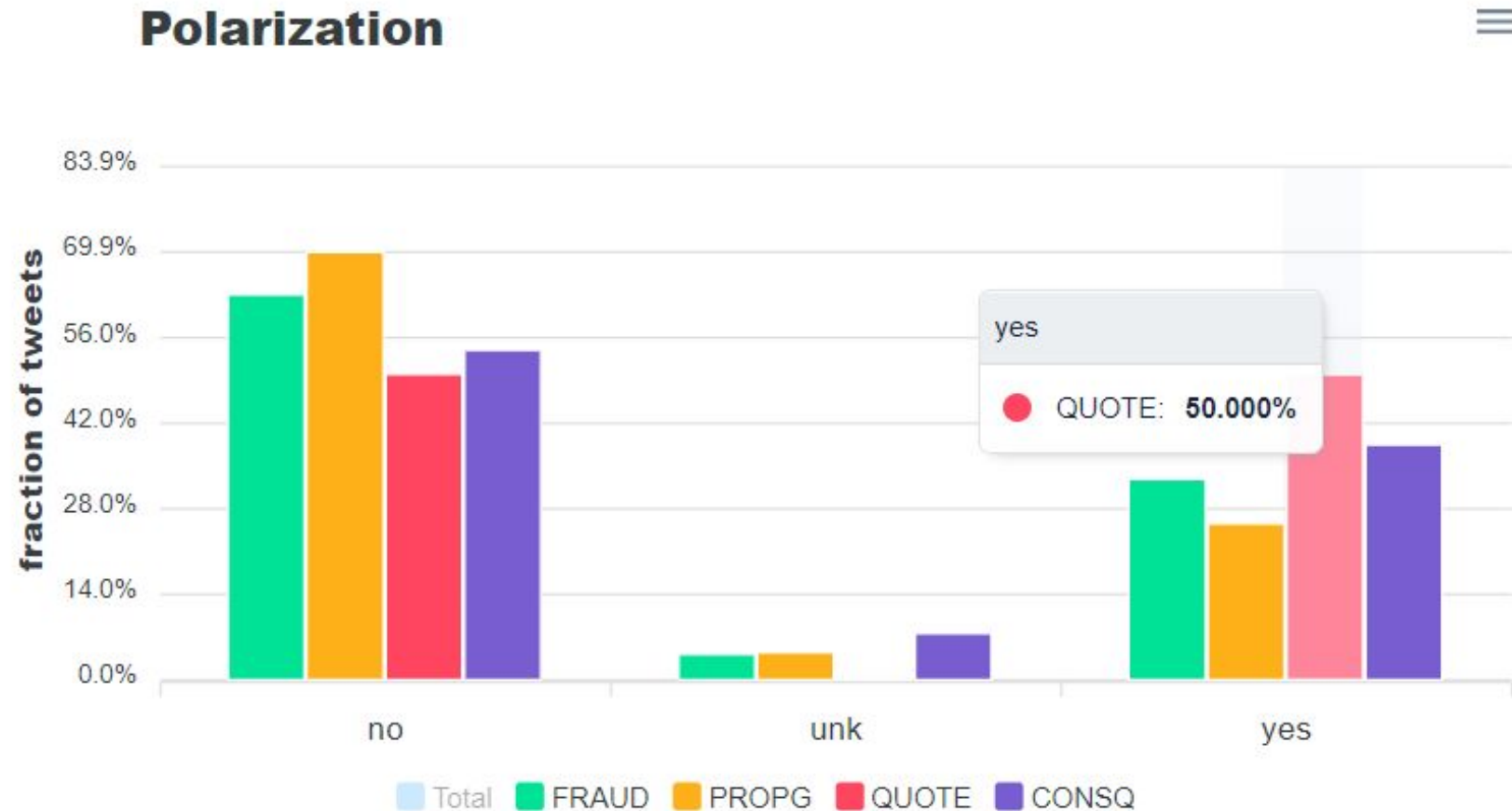
Disinfonet - Analysis Result: Histogram Polarization

An interactive histogram that highlights the polarization of each group of filtered tweets with respect to the selected classifier.

For each group of filtered tweets, it is possible to see the percentage subdivision derived from the classification, compared to the total number for that filter.

Through the legend below the graph, it is possible to select/deselect a group of filtered tweets to highlight/hide them in the graph.

Using the drop-down menu at the top right of the graph, you can download a screenshot of the graph in .svg or .png format, or the .csv of the data.



The figure provides the polarization of tweets for each available filter.

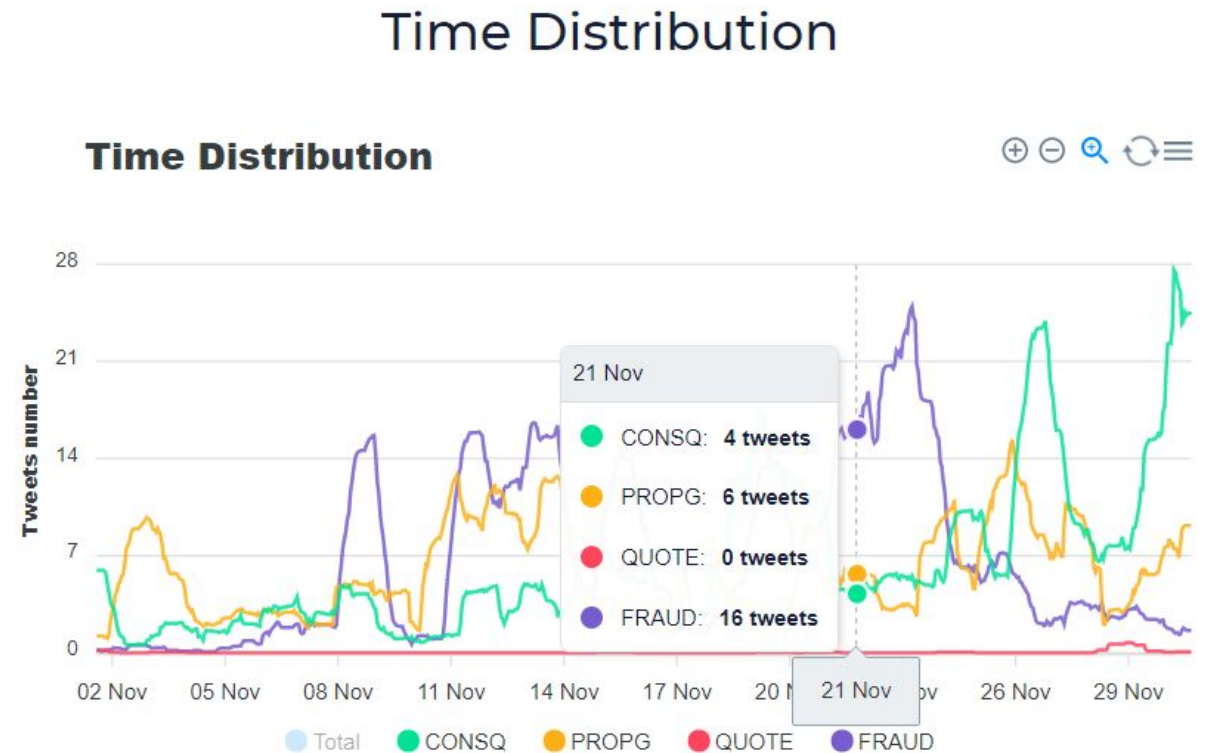
Disinfonet - Analysis Result: Time Distribution

An interactive graphic that shows the time distribution of each group of tweets by filter created.

By placing the mouse on a point on the graph, the number of tweets made that day will be displayed for each filter.

It is possible to zoom the figure by selecting a time window inside the graph, or by using the two buttons “zoom in” and “zoom out” in the menu bar at the top right. To reset the zoom just select the appropriate button in the menu bar. Through the legend below the graph, you can select/deselect a group of filtered tweets to highlight/hide them in the graph.

Using the drop-down menu, at the top right of the graph, you can download a screenshot of the graph in .svg or .png format.



The figure provides the one-day rolling mean of each obtained class in the considered period of time, compared with the overall trend.

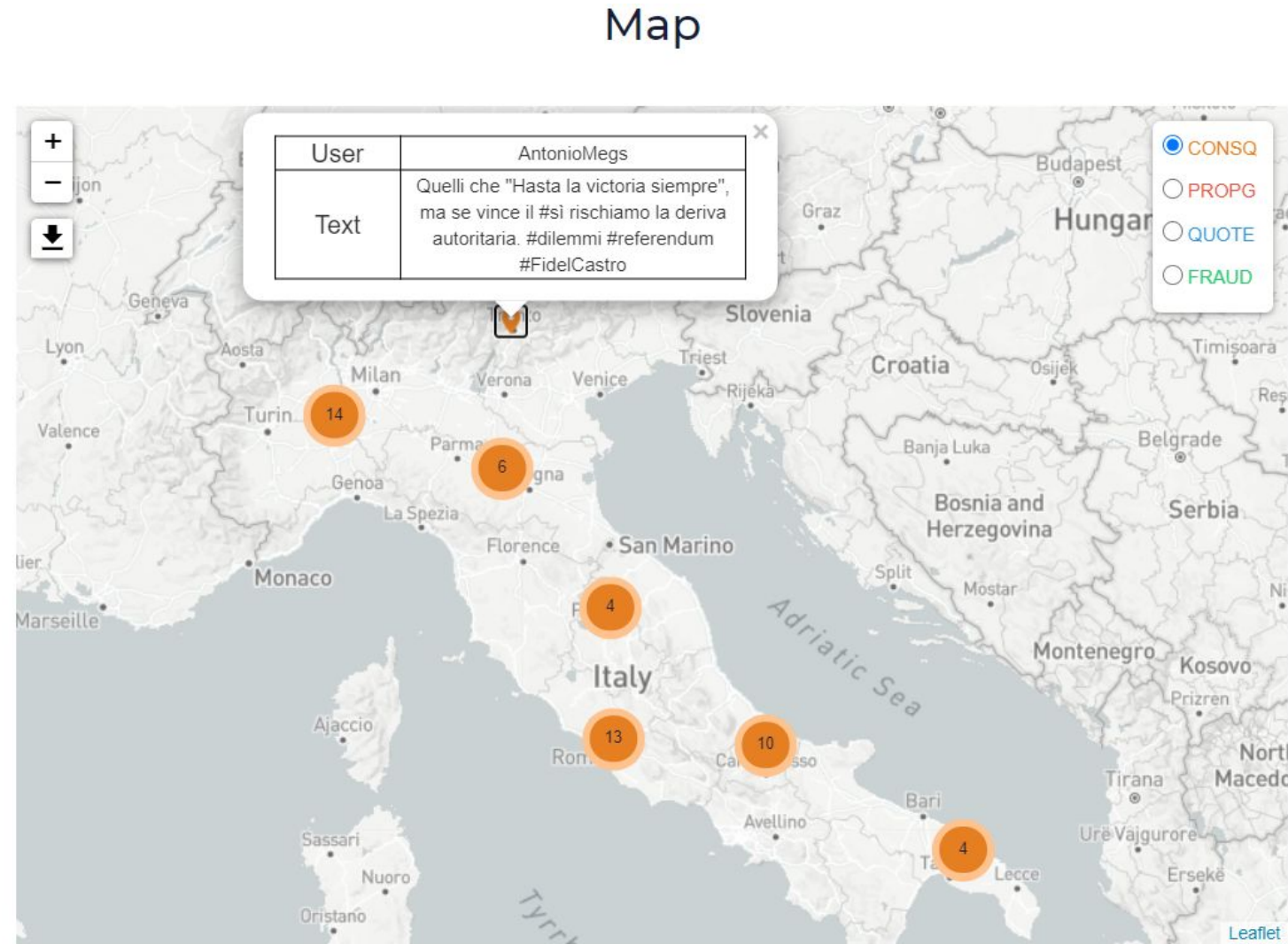
Disinfonet - Analysis Result: Map

The map visualization shows the geographical distribution, where possible, of the filtered tweets.

Through the menu at the top right, it is possible to select a specific filter to be displayed. It is possible to zoom the map by scrolling with the mouse or by using the + and - buttons located at the top left of the graph.

By clicking on a specific tweet shown on the map, the user who created it and the text of the tweet will be shown.

It is also possible to download a screenshot of the map using the button with the download symbol located at the top left.



The figure provides the geographic distribution of tweets, based on the percentage of geotagged tweets.

Disinfonet - About

The About page, accessible from the header, provides useful information on the Luiss Data Lab and SOMA project.

About us

Disinfonet is the main algorithmic engine of the SOMA verification platform, designed to support its users in understanding the dynamics of (fake) news dissemination in social media and tracking down the origin and the broadcasters of false information.

Disinfonet Toolbox combines several features: data collection, data analysis and visualization for exploring conversations from social networks.

DisInfoNet Toolbox is developed by [LUISS Data Lab](#), a Research Center based at the Department of Political Science of LUISS Guido Carli. The center carries out basic and applied research in the field of Big Data and digital transformation, relying on a multidisciplinary perspective.

Disinfonet is powered by [SOMA](#) "Social Observatory for Disinformation and Social Media Analysis", a H2020 Project aimed at supporting, coordinating and guiding the efforts of researchers, fact-checkers and journalists contrasting online and social disinformation.

Disinfonet - Privacy and Policy

The Policy and Privacy page, accessible from the header, provides useful information on the use and treatment of personal data and information of the user that uses the application.

Privacy Policy and Treatment of Personal Data

Premise

The information describes the characteristics of the treatments carried out by Luiss in relation to the navigation data of the website soma-disinfonet.luiss.it.

What personal data do we collect?

Luiss Datalab, as Data Controller, collects the following data:

- user identification data, such as, for example, the IP addresses and domain names of the computer;
- data relating to the use of the website by the user, through the use of technical, analytical third-party and profiling cookies;
- data relating to user preferences, expressed while browsing the website;
- data provided voluntarily by the user - identification information - for access to certain services offered through the website, for which reference is made to the specific information present in the relevant reference sections.

As regards the use of cookies, these are made up of text files that are automatically generated in the user's computer after visiting some pages of the site. Some of these files (session cookies) are automatically removed when the browser is closed. Another type of cookies, on the other hand, is recorded and stored on the User's computer (for example, to automate the procedure for accessing the reserved area, the User can choose that his User-ID and password identification data are stored in one of these files). A further typology is made up of analytical cookies, used to collect information, in aggregate form, on the number of users and on the methods by which they visit the website. The aforementioned cookies can be installed directly by the site operator or from a different site that installs them through the first (so-called "third party" analytical cookies). Finally, the last type of cookies is made up of profiling cookies, used for the purpose of sending advertising messages, for example through personalized banners, in line with the preferences expressed by the User while surfing the net. If the User prefers not to receive cookies, he can prevent their transmission by the website by properly configuring his Internet browsing browser. In some cases, however, the use of some parts of the site may be conditioned by the storage of cookies on the User's computer.

For what purposes do we use your personal data?

Luiss processes personal data for the following purposes:

- manage the correct use of the website;
- evaluate, in statistical form, the use of the website by users;

 LUISS

