Detection and characterization of misleading information spreading on online social media

Francesco Pierri
2nd year PhD Candidate in *"Data Analytics and Decision Sciences"*

Supervisors*:* Stefano Ceri, Fabio Pammolli

PUBLICATIONS

Francesco Pierri, Stefano Ceri **False News On Social Media: A Data-Driven Survey.** *ACM SIGMOD Record* Vol 48 issue 2 (2019)

Brena, G., Brambilla, M., Ceri, S., Di Giovanni, M., **Pierri, Francesco**, & Ramponi, G. "News Sharing User Behaviour on Twitter: A Comprehensive Data Collection of News Articles and Social Interactions." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 13. No. 01. 2019.

Francesco Pierri, Carlo Piccardi, Stefano Ceri **Topology comparison of Twitter diffusion networks effectively reveals misleading news.** (2020) *Scientific Reports*

Francesco Pierri, Alessandro Artoni, Stefano Ceri **Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections.** (2020) *Plos One*

Francesco Pierri The diffusion of mainstream and disinformation news on Twitter: the case of Italy and France. (2020) *Companion Proceedings of The 2020 World Wide Web Conference WWW '20*

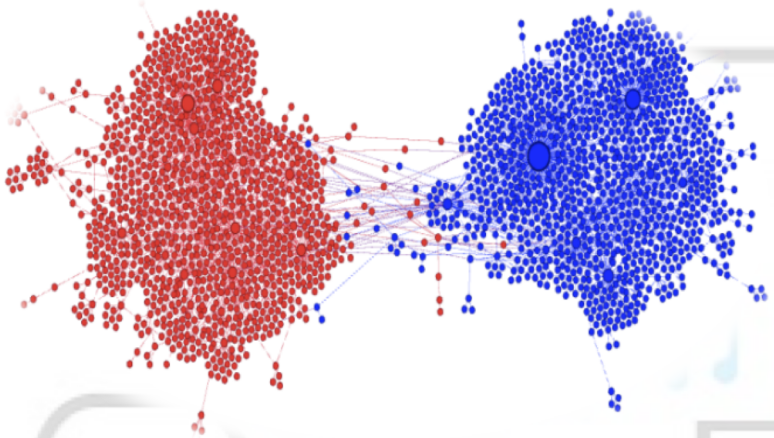
What is **misleading information**?

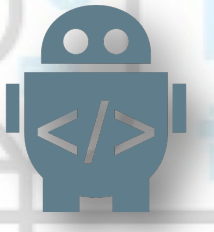
1. Disinformation
2. Misinformation
3. Hoaxes
4. Rumours
5. Satire
6. Clickbait
7. Junk news
8. Propaganda
9. Fake news
10. ...

No single definition nor general agreement on different meanings!

What happens on **Social Media?**

- No quality control
- Echo chambers
- Malicious agents: bots, cyborgs, trolls
- Human factors: confirmation bias, naïve realism, etc …





What's the **REAL** problem?

- **Politics**: UK 2016 Brexit, US 2016 elections, FR 2018 Gilets Jaunes, IT 2018 Elections, EU 2019 Elections
- **Finance**: "Obama injured" stocks wipe-out 2014, Starbucks free coffee 2017
- **Healthcare**: Ebola, Zika, vaccines in general

Challenges

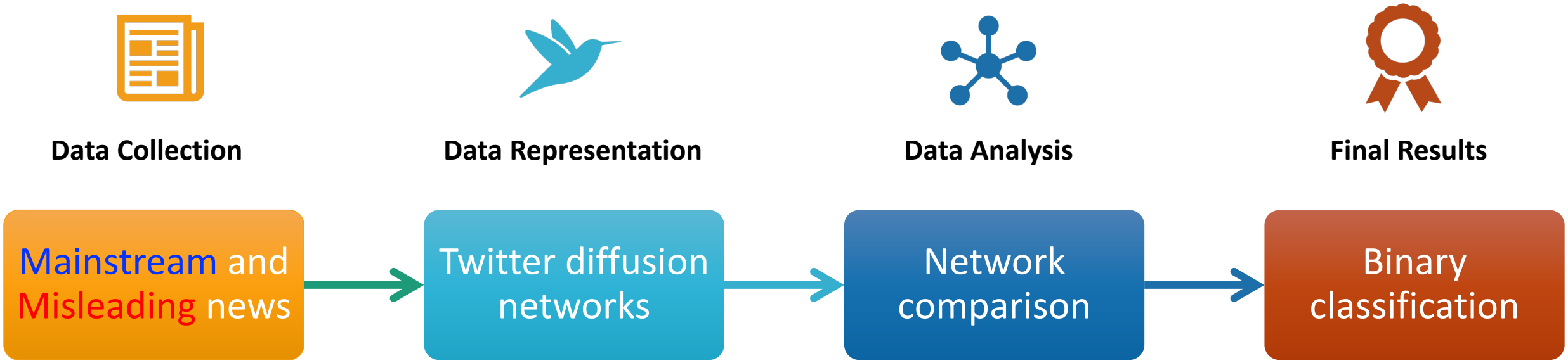
Misleading information is deliberately created to **deceive** people and **mimic** traditional news outlets (“adversarial setting”)

They are produced and spread **massively** (can’t verify them all manually)

Low-availability of **data** (Twitter API limitations, FB privacy restriction)

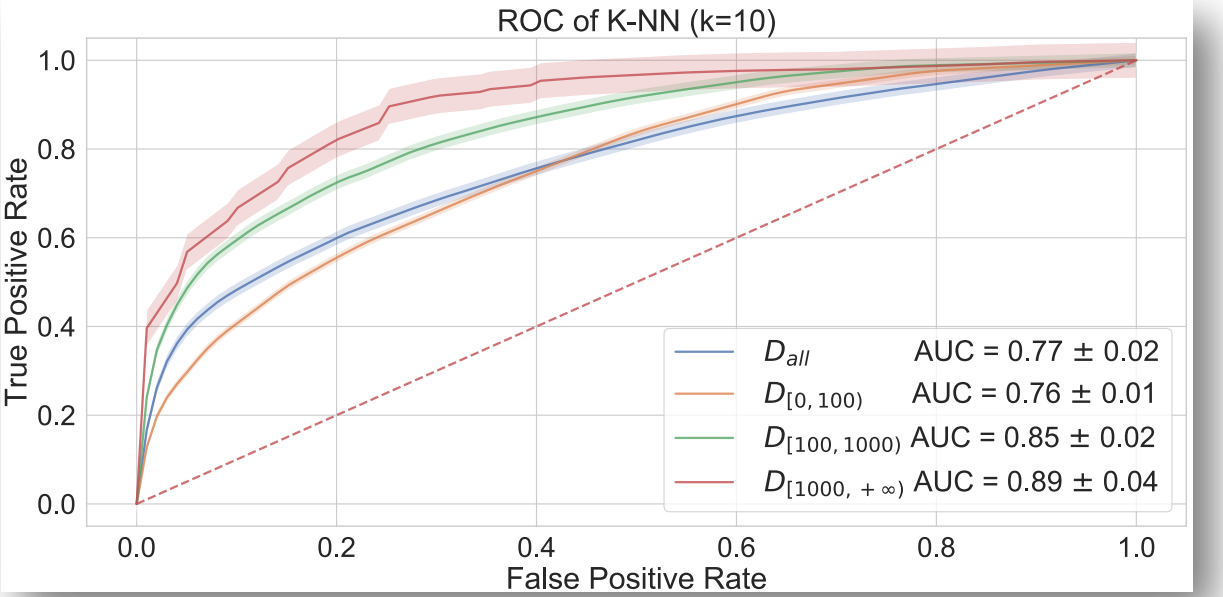
Censorship concerns about potential intervention by platforms

Topology comparison of Twitter diffusion networks effectively reveals misleading information *(Pierri, Piccardi, Ceri Scientific Reports 2020)*

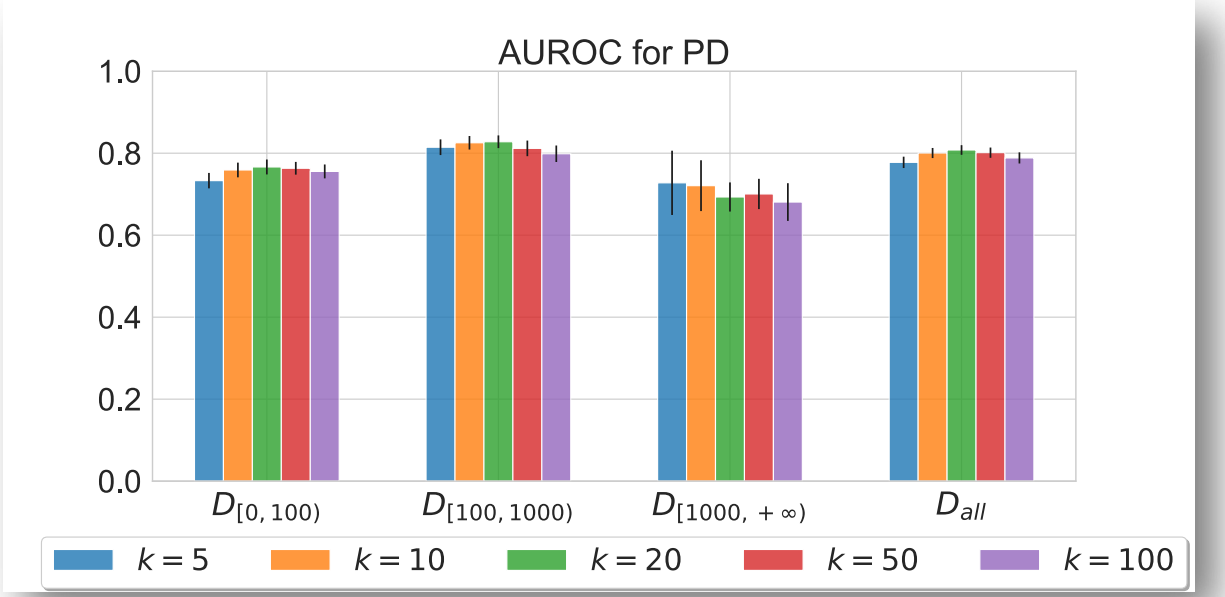


Classification results

AUROC up to 94% with a simple Logistic Regression classifier!



Receiver operating characteristic curve (ROC) for K-NN (k=10) classifier trained on global network properties, evaluated on different partition sizes.



AUROC value for K-NN classifier (for different number of neighbors) trained on the Portrait Divergence similarity matrix, evaluated on different partition sizes.

Qualitative results (in accordance with literature findings!)

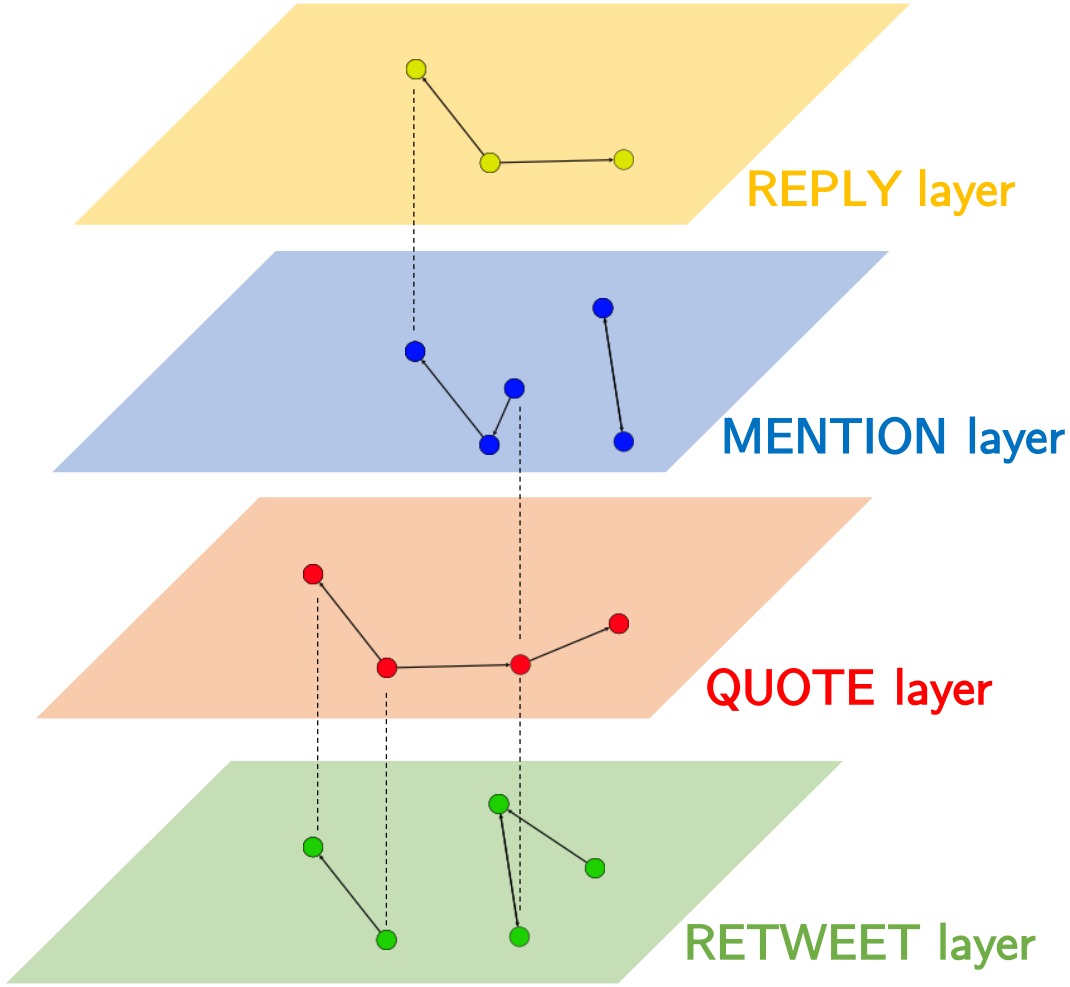
Global network properties can be **interpreted** in terms of social dimensions, e.g. WCC is the number of cascades, LWCC is the largest cascade, CC indicates the degree of connectedness between users, etc.

Inspecting feature **distributions**:

Communities of users sharing disinformation are more **clustered** and **connected**.	Disinformation cascades are **broader** and **deeper**, and they are shared in a less **broadcast** flavour compared to mainstream.

Multilayer diffusion networks

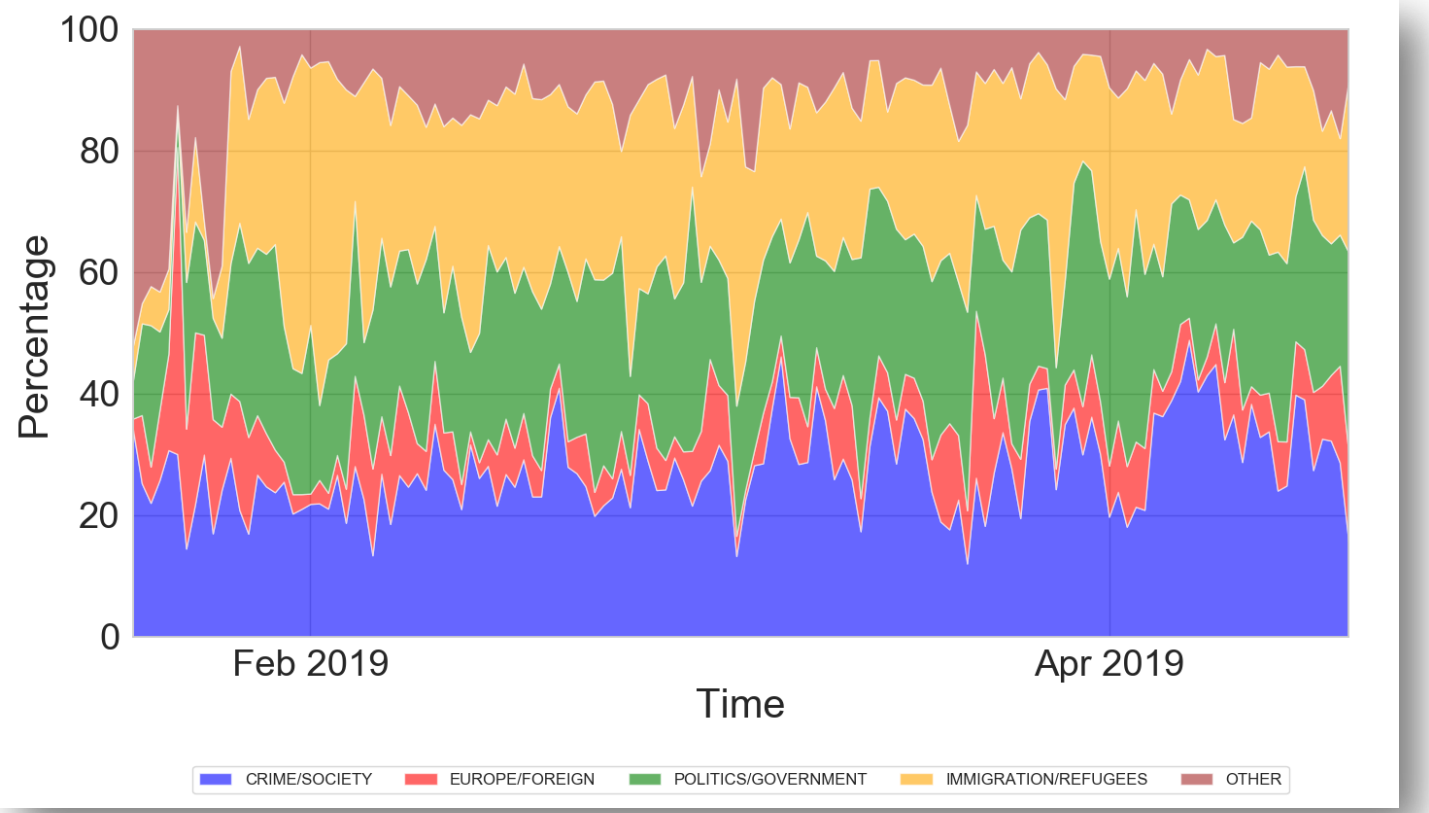
- A layer for each Twitter **action** (concatenated features)
- New **features**: density and structural virality
- New **Italian** dataset (2019 mainstream and disinformation news)
- **Better** results (AUROC > 80%)



An example of Twitter multilayer diffusion network.

Investigating Italian disinformation on Twitter in the run-up to 2019 European Parliament elections *(Pierri, Artoni, Ceri PLOS One 2020)*

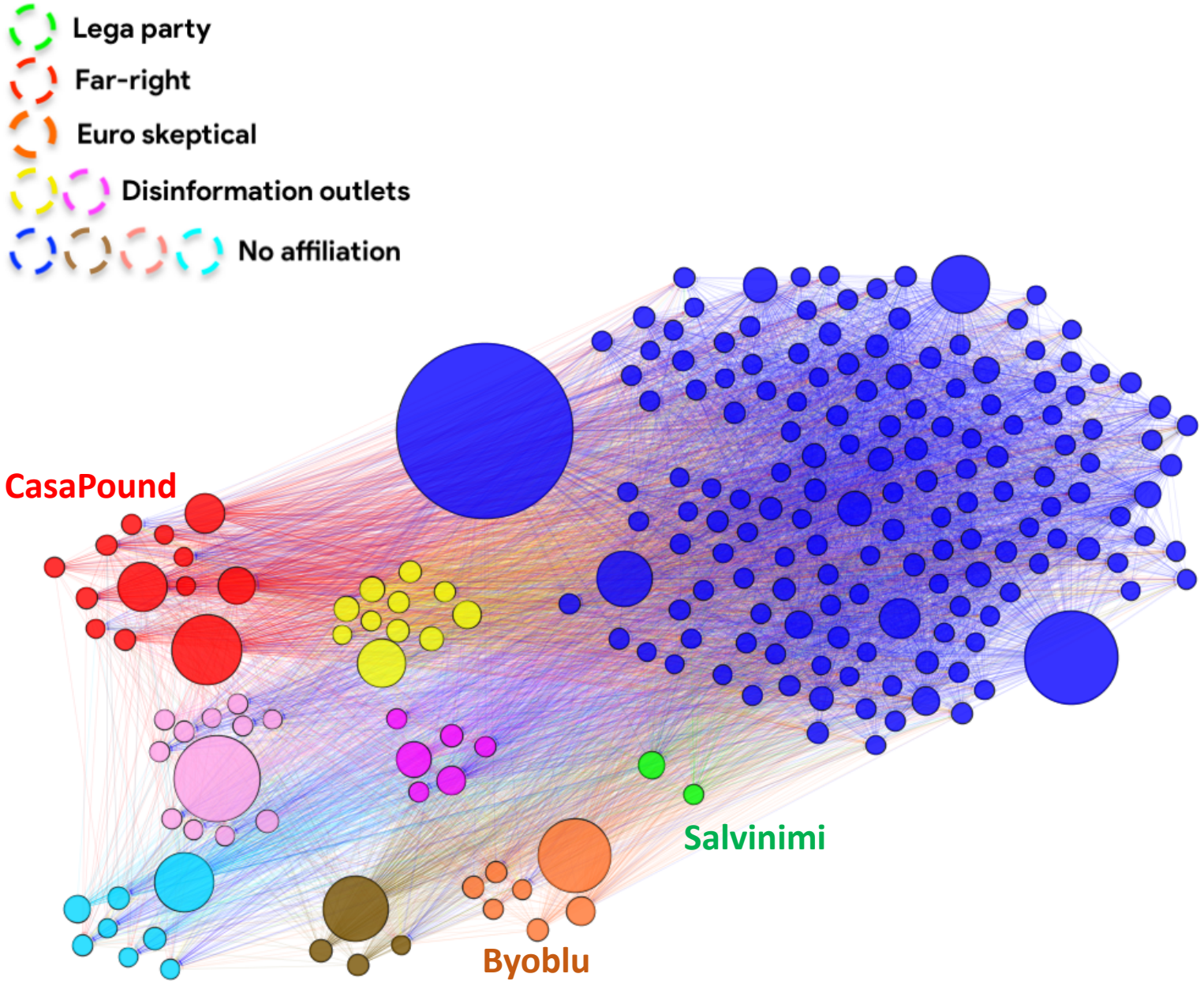
- Twitter Streaming API (07th Jan -27th May 2019)
- 60+ Italian disinformation websites
- 360k Tweets containing URLs
- Focus on controversial topics (e.g. immigration, national safety)



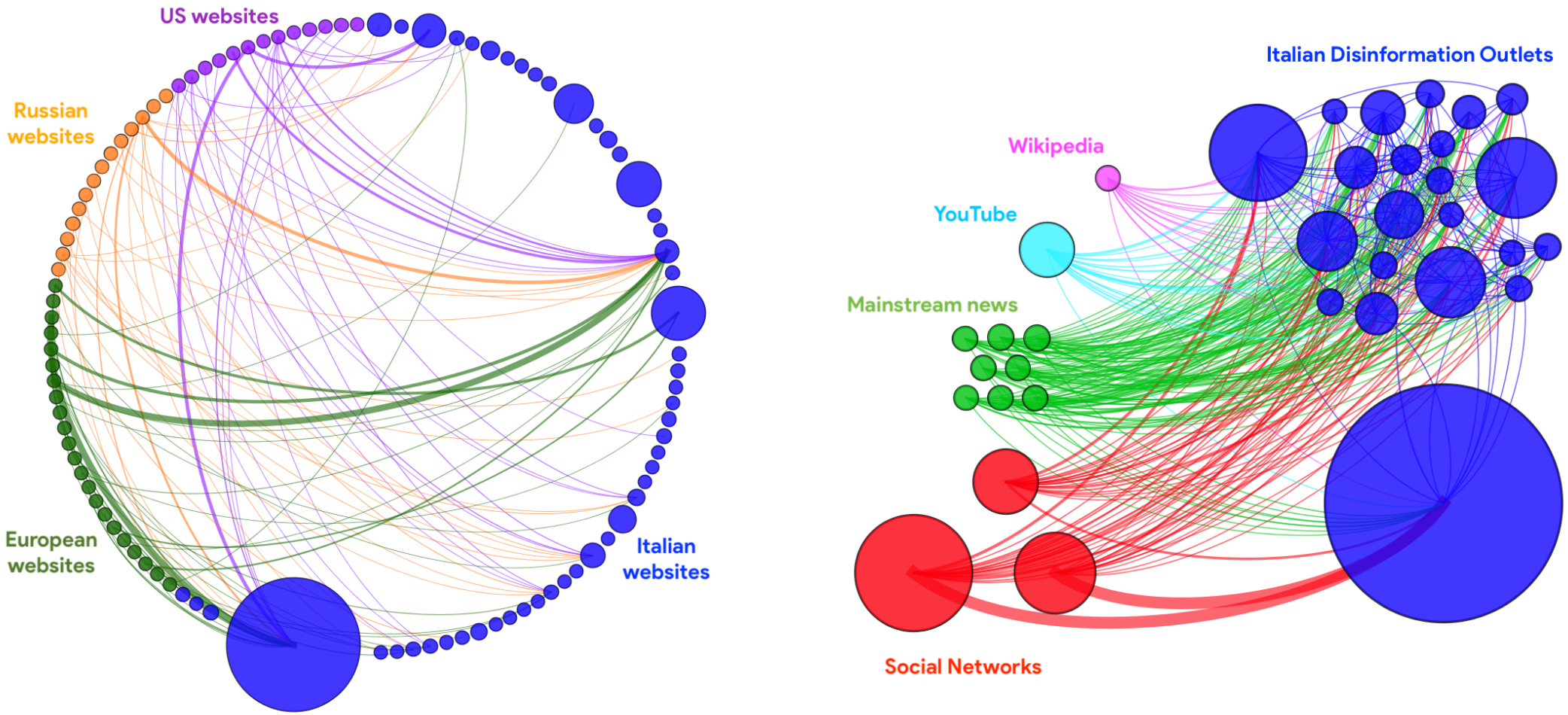
A stacked-area chart showing the distribution of different **topics** over the collection period. The daily coverage on themes related to Immigration/Refugees and Europe/Foreign is stationary, whereas focus on subjects related to Crime/Society and Politics/Government is monotonically increasing towards the elections (end of May 2019).

Core of the disinformation diffusion network

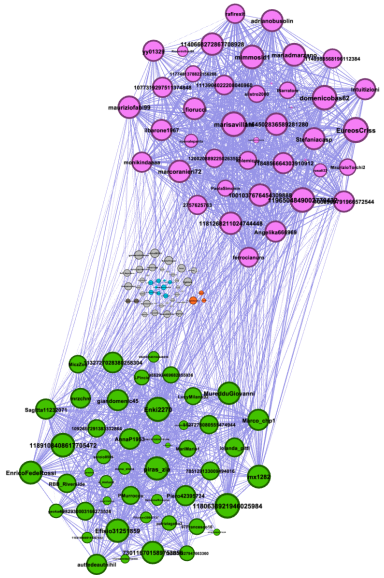
- Main K-core (k = 47)
- Links with Italian **far-right** and **Lega** parties
- No **bot** prevalence (using Botometer)
- Easy to **dismantle**

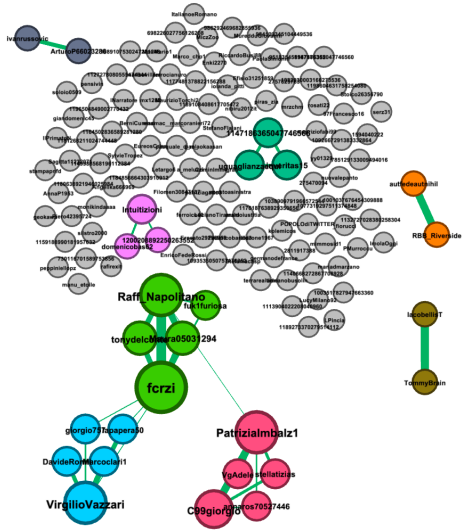


Network of websites



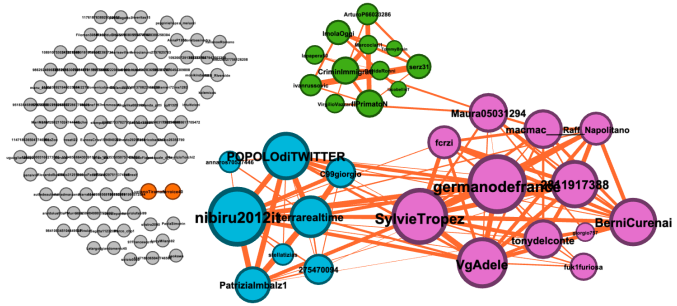
Two different views of the network of websites. The size of each node is adjusted w.r.t to the Out-strength, the color of edges is determined by the target node and the thickness depends on the weight (i.e. the number of shared tweets containing an article with that hyperlink).
Left. The main core of the network (k = 14); blue nodes are Italian disinformation websites, green ones are Italian traditional news outlets, red nodes are social networks, the sky-blue node is a video sharing website and the pink one is an online encyclopedia.
Right. The sub-graph of Russian (orange), EU (olive green), US (violet) and Italian (blue) disinformation outlets.





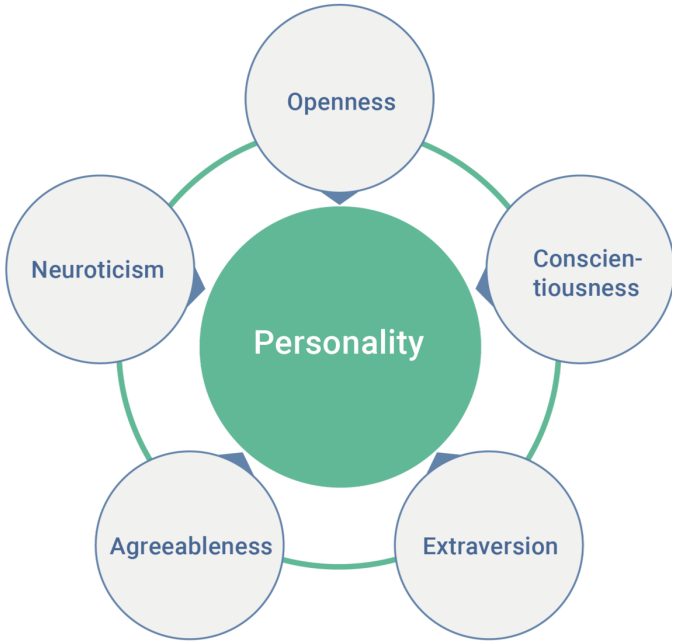
$$\mathbb{B}^3_{content} = \left\{ \begin{array}{ll} \texttt{N} & \longleftarrow \text{tweet contains no entities (plain text),} \\ \texttt{E} & \longleftarrow \text{tweet contains entities of one type,} \\ \texttt{X} & \longleftarrow \text{tweet contains entities of mixed types} \end{array} \right\} = \{\texttt{N}, \texttt{E}, \texttt{X}\}$$

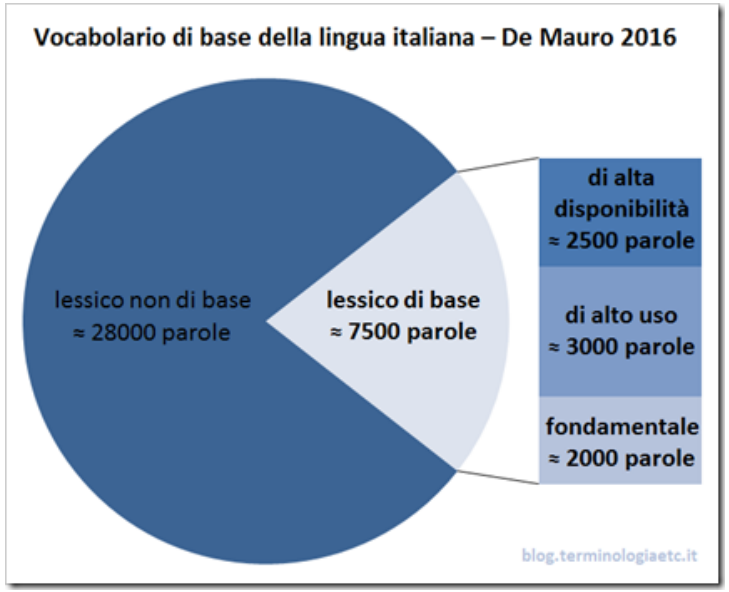
$$\mathbb{B}^6_{content} = \left\{ \begin{array}{ll} \texttt{N} & \longleftarrow \text{tweet contains no entities (plain text),} \\ \texttt{U} & \longleftarrow \text{tweet contains one or more URLs,} \\ \texttt{H} & \longleftarrow \text{tweet contains one or more hashtags,} \\ \texttt{M} & \longleftarrow \text{tweet contains one or more mentions,} \\ \texttt{D} & \longleftarrow \text{tweet contains one or more medias,} \\ \texttt{X} & \longleftarrow \text{tweet contains entities of mixed types} \end{array} \right\} = \{\texttt{N}, \texttt{U}, \texttt{H}, \texttt{M}, \texttt{D}, \texttt{X}\}.$$



Work in progress (1): Mining user roles in (dis)information diffusion

- Can we identify functional roles of accounts sharing news on Twitter?
- CS + NETSCI approach: profile/cluster users based on their interactions (digital DNA) and the resulting networks (centrality measures)





Work in progress (2) Psycho-socio-linguistic analysis

- Differences in users sharing mainstream vs disinformation news on Twitter?
- Personality/Demographics prediction
- Computational linguistics: LIWC, syntax, grammar, language richness, etc

Thank you for your attention!

Questions?