9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Real time study of the political opinion using massive data

Keywords: Social media, data based opinion dynamics, semantic networks, real time community tracking, memory effects in Twitter

Extended Abstract

Online social networks, particularly Twitter, have become a common arena for researching the spread of political views. Political parties and traditional media use the discussions on specialized platforms to gauge social opinion trends, despite concerns about the representativeness of online network data. Radio and television programs often discuss these online conversations, thus spreading their content beyond the restricted social groups that typically use online platforms.

Most empiric works of online social opinion are based on the study of the usages of a set of keywords chosen a priori, according to the subject under study. Following a completely automated approach [1, 2], we base our study on a weighted network of hashtags where the weighted link represents the number of single users that have used two hashtags in the same tweet. Then, we determine the topics of discussion in the platform by community detection on this semantic network. This procedure has been successfully applied [2] using an aggregated semantic network based on data collected over all the studied period. As a consequence, the determination of the topics discussed at a given time, includes some information from the *future*, which is not useful if one is interested in following the events in *real time*. In this work we present a method that allows to do so. We need to deal with evolving semantic networks which implies a compromise between the instantaneity of the collected information and a reasonable amount of data required to make it robust. Here we compare two different procedures of building the semantic network. In one case, we cumulate data for the first month and then we aggregate the new data of each week. In the other, we start with the same network as before, and we use a sliding window of a week to integrate new data and drop the oldest one. In the first case our data keeps the memory of all the discussions in Twitter since the beginning of the capture, in the second, the system loses memory every week. In both cases, for each new network we determine communities that constitute the topics of discussion of the week. Fig.1 shows that the two procedures give qualitatively the same behavior of the dynamic similarity curves among the supporters of different candidates. The exception appearing for the supporters of a very atypical one –Zemmour- in the second week of December. As the hashtags used by the supporters of the candidates at a given point in time are the same for both methods, the difference arises from a different characterization of the topics with and without memory.

Furthermore, since our approach leads us to update the semantic landscape, by reconstructing the cooccurrence graph on a weekly basis, we are able to characterize and follow the evolution of the topics themselves. We use a dynamical procedure to track the communities over time [3], and we observe how discussion topics grow, shrink (eventually disappear), split or merge with others. In Fig.2 we show that for controversial subjects like vaccination policies, as time passes, the topics integrate new highly used hashtags that tend to extremist positions.

References

[1] F. M. Cardoso, S. Meloni, A. Santanchè, and Y. Moreno, "Topical alignment in online social systems", Frontiers in Physics **7**, 58 (2019).

[2] T. Mussi Reyero, M. G. Beiró, J. I. Alvarez-Hamelin, L. Hernández, and D. Kotzinos, "Evolution of the political opinion landscape during electoral periods", EPJ Data Science **10**, 31 (2021).

[3] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks", in 2010 international conference on advances in social networks analysis and mining (Aug. 2010), pp. 176–183.

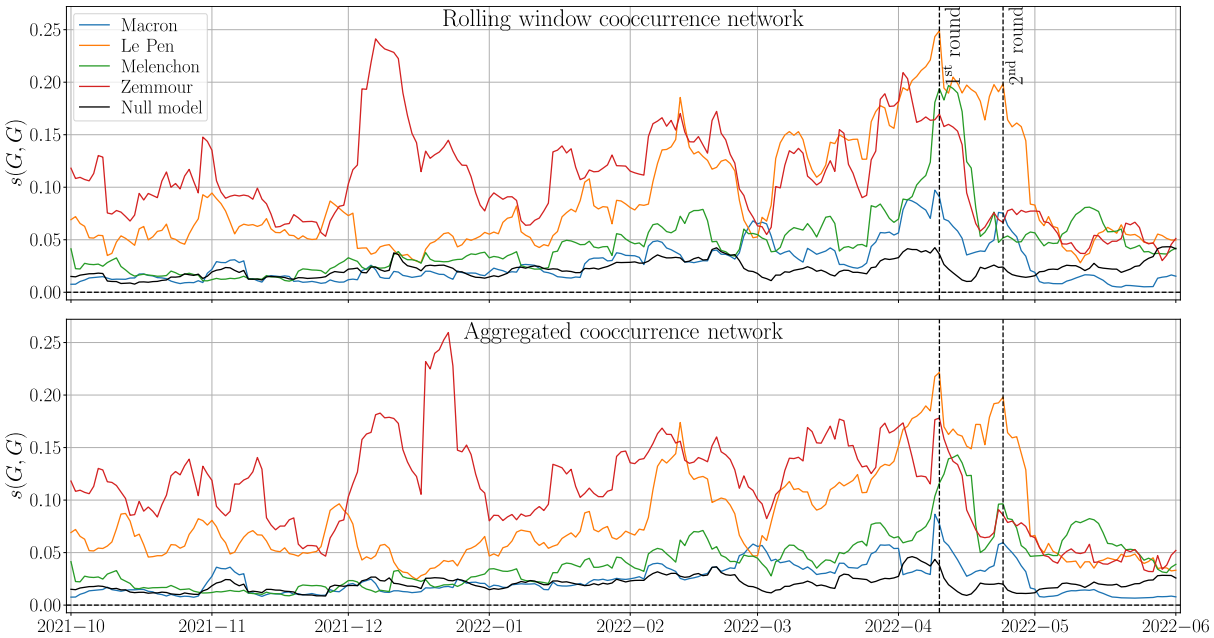

Figure 1: Dynamical self-similarities among the supporters of some typical candidates to the French presidential election in 2022. Same color represents the supporters of the same candidate in both panels. Top panel: Topics have been calculated using a semantic network computed over one month with a rolling window of a week (no memory). Bottom panel: Topics have been calculated using a semantic network that aggregates one week at each time (memory is kept)

Figure 2: Dynamic Topic Tracking, example of an anti-vax topic weekly evolution. The topic is labeled with the most used hashtag and represented by a node whose size is proportionnal to the number of hashtags constituting the topic. Topics have been calculated using a semantic network that aggregates one week at each time.