Characterizing social information spreading by using event-synchronization and causality measures

Lucio Garcia[1], Giulio Tirabassi[2], Cristina Masoller[2], and Pablo Balenzuela[1]
[1]Departamento de Fisica, Universidad de Buenos Aires, Argentina
[2]Departament de Fisica, Universitat Politecnica de Catalunya, Terrassa, Spain

Understanding the diffusion of information is a fundamental challenge of complexity science. Here we analyze 28000 news articles published in Argentina in the period 26/05/2022 - 26/09/2022 in six main cities, and classify them in 20 non-orthogonal topics (Fig. 1). Then, we obtain a time series for each topic, $n = 1, \ldots, 20$, in each city, $i = 1, \ldots, 6$, by adding the number of articles per day.

Next, we use two causality measures, Granger causality, GC, and pseudo Transfer Entropy, pTE [2] to study how the information about a topic that appears in the local press in one city, spreads to news articles published in other cities. GC and pTE, however, have the drawback that assume stationarity. Therefore, we also use event synchronization measures, Q_s and Q_a, as proposed in [3]. To calculate Q_s and Q_a we count the number of events, c_{ij}, that occur in one time series, j, after an event occurred in a time series i, allowing for a lag of up to 3 days. The events are detected by using two thresholds, one to detect when media attention grows above a certain value, and another, to detect when media attention decays, and then, a different event may occur latter. These thresholds are defined in terms of the relative importance of a topic.

Fig. 1. Unsupervised distribution of articles in 20 topics.

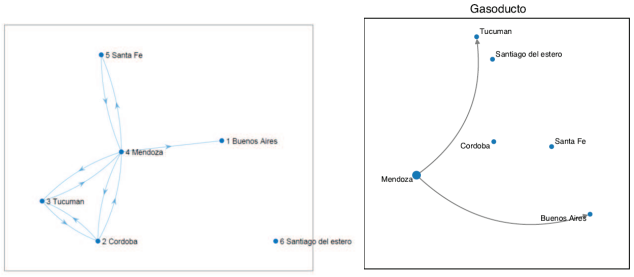

Fig. 2. Links obtained for topic *Gasoducto* from (left) event synchronization measures and (right) causality measures.

Finally, the process of information spreading is represented as a multiplex network, in which the different topics represent the layers, the six cities in Argentina represent the nodes, and the links are defined by thresholding the values of (GC and pTE) or (Q_s and Q_a). Figure 2 shows, as an example, the links obtained for topic *Gasoducto*, and Fig. 3 displays the corresponding time series.

Figure 4 displays the causal layers obtained from the 20 topics analyzed. In this talk/poster, the results obtained with different event-detection criteria and with different synchronization/causality measures will be compared and discussed.

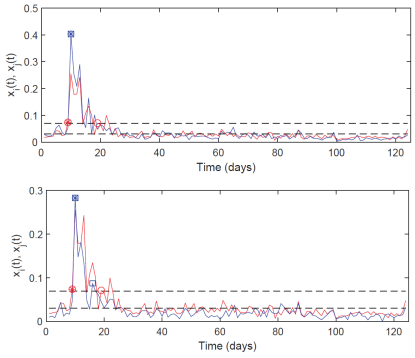

Fig. 3. Examples of time series obtained for topic *Gasoducto*. Top: Mendoza (red) and Buenos Aires (blue); Bottom: Mendoza (red) and Tucuman (blue). Empty (solid) symbols indicate not-synchronized (synchronized) events.

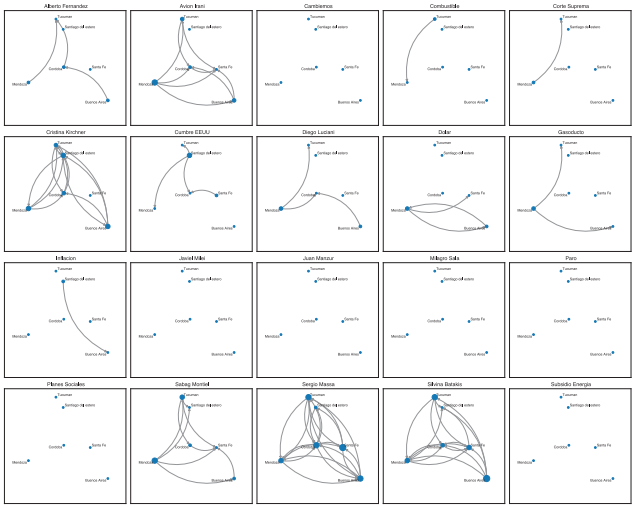

Fig. 4. Layers obtained by thresholding the values of GC and pTE. Some layers do not have links, which also happens when thresholding Q_s and Q_a.

[1] S. Pinto et al., *Quantifying time-dependent Media Agenda and public opinion by topic modeling*, Physica A **524**, 614 (2019).

[2] R. Silini and C. Masoller, *Fast and effective pseudo transfer entropy for bivariate data-driven causal inference*, Sci. Rep. **11**, 8423 (2021).

[3] R. Quian Quiroga, T. Kreuz, and P. Grassberger, *Event synchronization: A simple and fast method to measure synchronicity and time delay patterns*, Phys. Rev. E **66**, 041904 (2002).