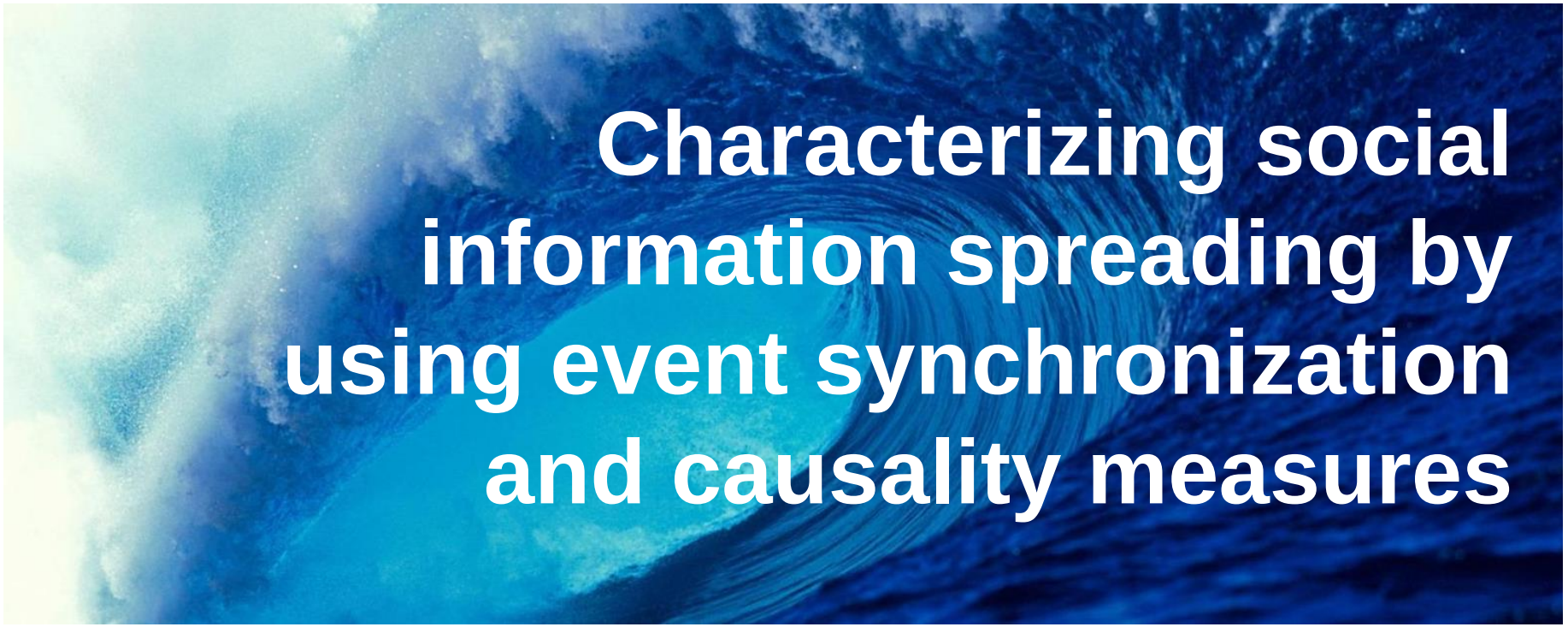


Characterizing social information spreading by using event synchronization and causality measures

Lucio Garcia(1), **Giulio Tirabassi**(2), **Cristina Masoller**(2), **Pablo Balenzuela**(1)
(1) Universidad de Buenos Aires
(2) Universitat Politecnica de Catalunya

X COMPLEXITAT DAY
June 21, 2023

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Campus d'Excel·lència Internacional

ICREA

AGAUR

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

Data

- 28000 news articles published in Argentina in six main cities.
- 26/05/2022 - 26/09/2022
- Use an unsupervised non-negative matrix factorization algorithm to classify the articles in 20 non-orthogonal topics.

S. Pinto, F. Albanese, C. O. Dorso, and P. Balenzuela. *Quantifying time-dependent media agenda and public opinion by topic modeling*. Physica A, 524:614, 2019

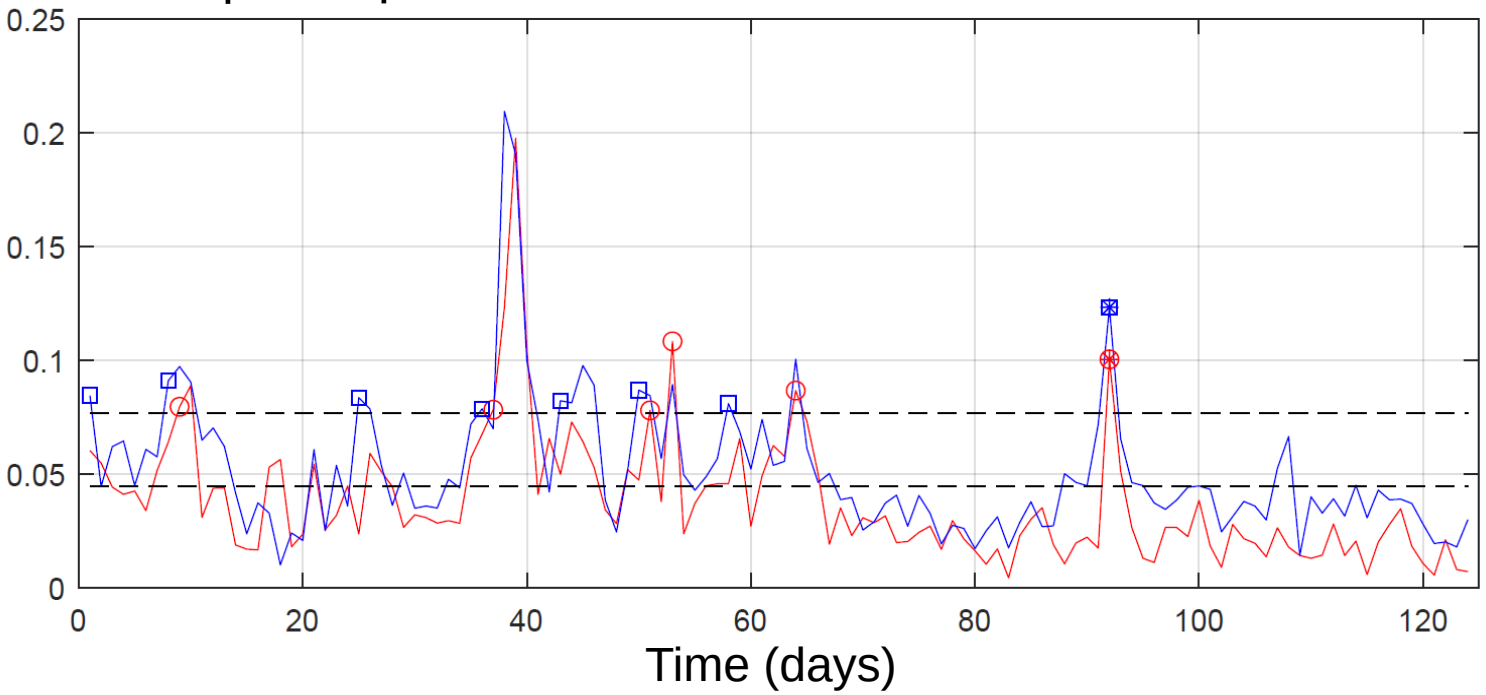
cristina.masoller@upc.edu @cristinamasoll1

Data

We obtain a time series for each topic in each town, by adding the number of articles per day.

We define “events” using two thresholds.

Example: Topic “*Alberto Fernandez*” in **BsAs** and in **Mendoza**



cristina.masoller@upc.edu @cristinamasoll1

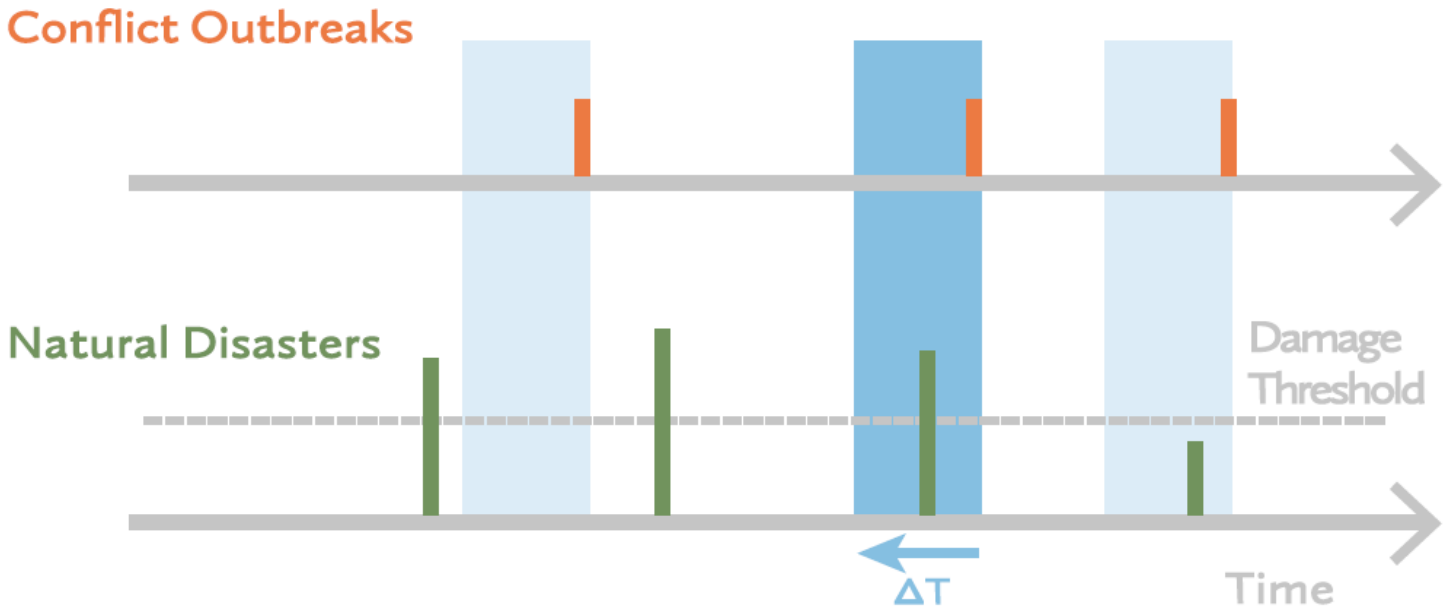
How to find “synchronized events” in two time series? Example

Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries

Carl-Friedrich Schleussner[a,b,c,1], Jonathan F. Donges[a,d], Reik V. Donner[a], and Hans Joachim Schellnhuber[a,e,1]

[a]Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany; [b]Climate Analytics, 10969 Berlin, Germany; [c]Integrative Research Institute on Transformations of Human–Environment Systems, Humboldt University, 10099 Berlin, Germany; [d]Stockholm Resilience Centre, Stockholm University, 114 19 Stockholm, Sweden; and [e]Santa Fe Institute, Santa Fe, NM 87501

9216–9221 | PNAS | August 16, 2016 | vol. 113 | no. 33



cristina.masoller@upc.edu @cristinamasoll1

Event synchronization measures

- Define “events” in each time series.
- Count $c^{\tau}(x|y)$ = number of times an event appears in x shortly after (within interval τ) an event appears in y. Idem for $c^{\tau}(y|x)$.
- Calculate:

$$Q_{\tau} = \frac{c^{\tau}(y|x) + c^{\tau}(x|y)}{\sqrt{m_x m_y}} \qquad q_{\tau} = \frac{c^{\tau}(y|x) - c^{\tau}(x|y)}{\sqrt{m_x m_y}}$$

m_x, m_y are the number of events in each time series.

- $Q_{\tau} = 1$: the events of the signals are fully synchronized.
- $q_{\tau} = 1$: the events in x always occur before those in y.
- $q_{\tau} = -1$: the events in x always occur after those in y.

No “causal” information

Quian Quiroga et al, PRE 66, 041904 (2002).

cristina.masoller@upc.edu @cristinamasoll1

Granger Causality

Hypothesis: X_1 and X_2 can be described by stationary autoregressive linear models.

$$X_1(t) = \sum_{j=1}^{p} A_{11,j}\ X_1(t\ -j) \quad \text{(past of } X_1\text{)} \qquad + E_1(t) \quad \text{(Residual error)}$$

$$X_1(t) = \sum_{j=1}^{p} A_{11,j}\ X_1(t\ -j) \quad \text{(past of } X_1\text{)} + \sum_{j=1}^{p} A_{12,j} X_2(t\ -j) \quad \text{(past of } X_2\text{)} + E'_1(t) \quad \text{(Residual error)}$$

If $\langle E'_1(t)\rangle < \langle E_1(t)\rangle$ ➡ $X_2 \rightarrow X_1$

C. W. J. Granger *Investigating causal relations by econometric models and cross-spectral methods*. *Econometrica* 37, 424–438 (1969) **(> 10000 citations)**

cristina.masoller@upc.edu @cristinamasoll1

Transfer Entropy (TE)

- TE: is the *Conditional* Mutual Information, given the "past" of one of the variables.

$$TE\ (x,y) = MI\ (x,\ y|x_{\tau})$$

$$TE\ (y,x) = MI\ (y,\ x|y_{\tau})$$

- $MI\ (x,y) = MI\ (y,x)$ but $TE\ (x,y) \neq TE(y,x)$

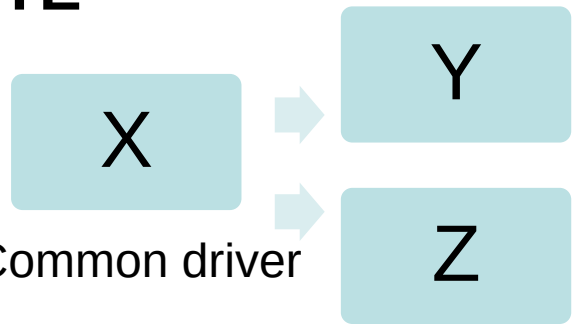
- TE and GC are equivalent for Gaussian processes.

T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85, 461 (2000).

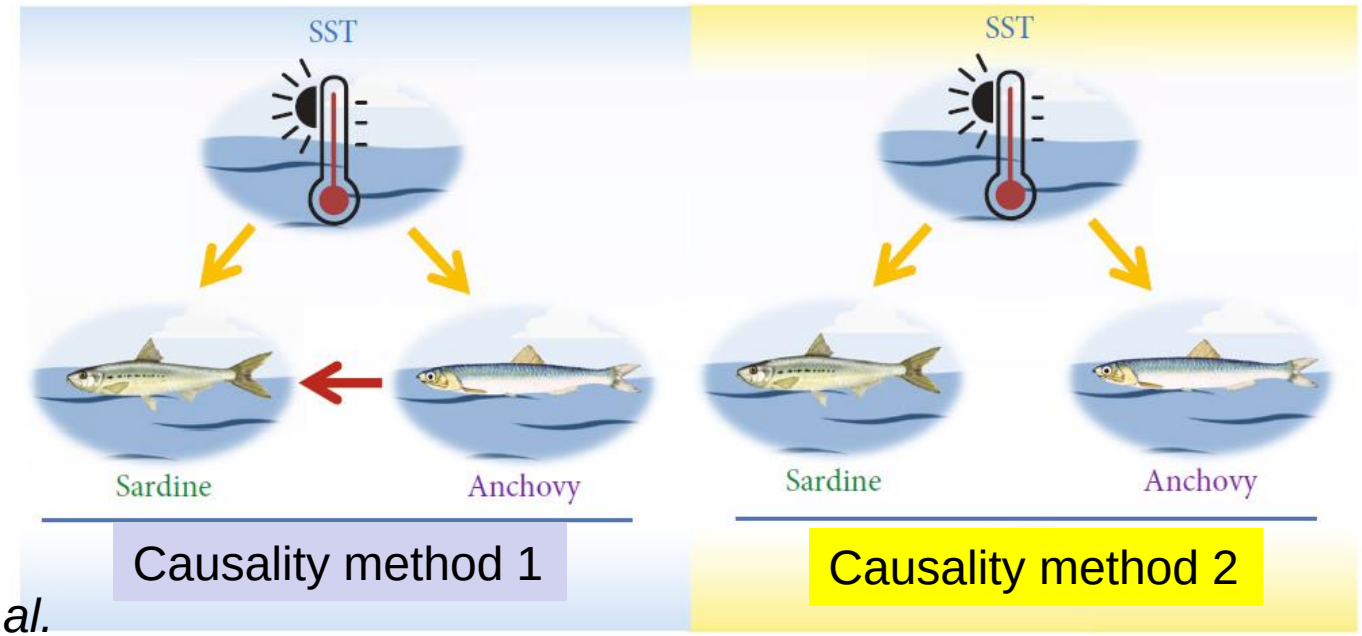
cristina.masoller@upc.edu @cristinamasoll1

Problems of Grange Causality and TE

Indirect link X→Z?



Common driver ↕ ?



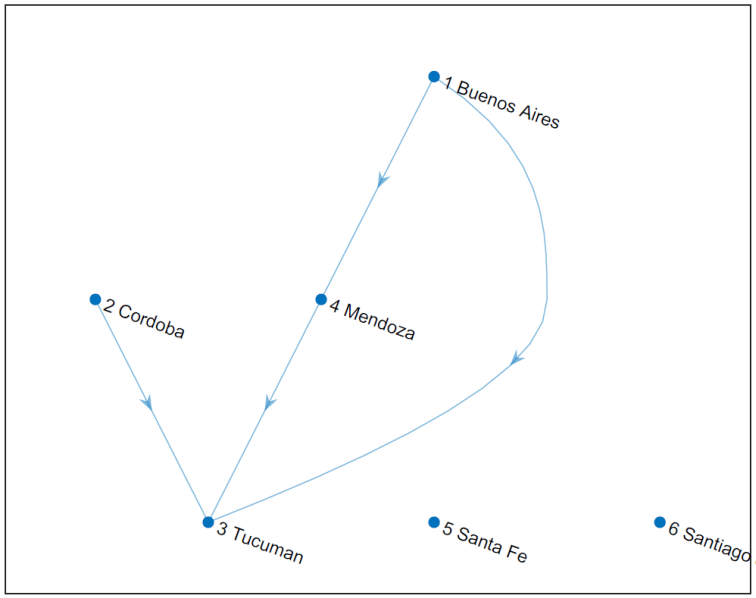
X. Ying et al.
AAAS Research 2022

Many alternative approaches to try to “solve” these problems.

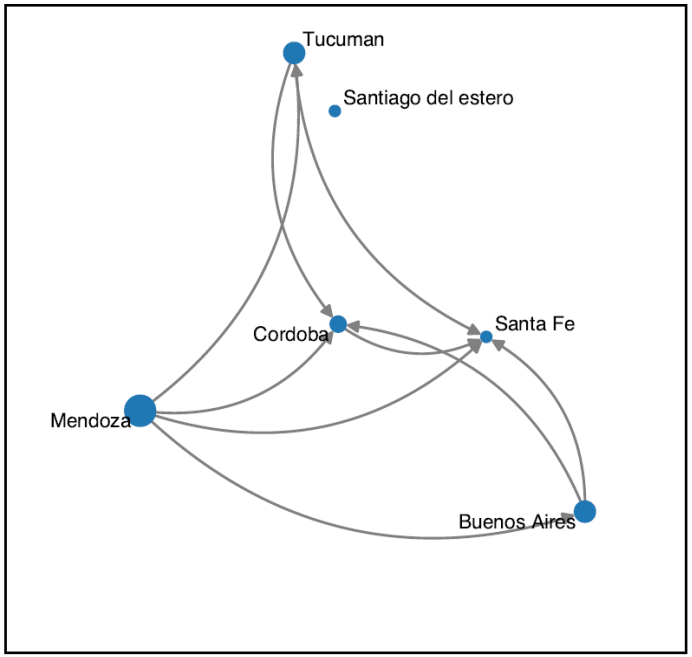
cristina.masoller@upc.edu @cristinamasoll1

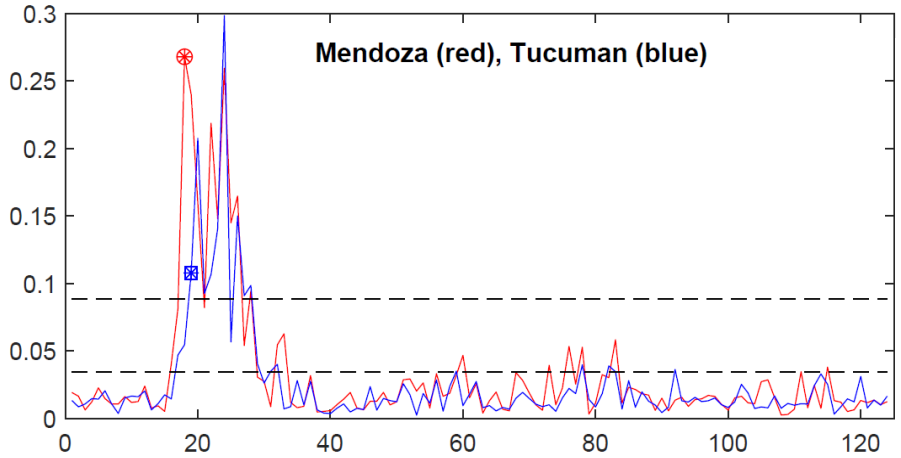
Results. Example topic "*Avion Irani*"

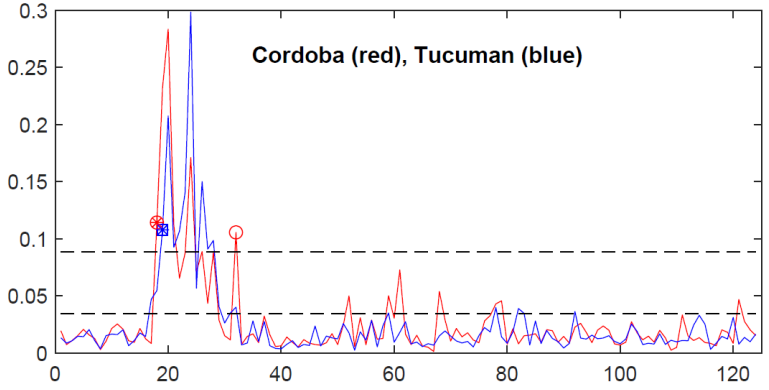
Qs or Qa significant



GC and TE significant



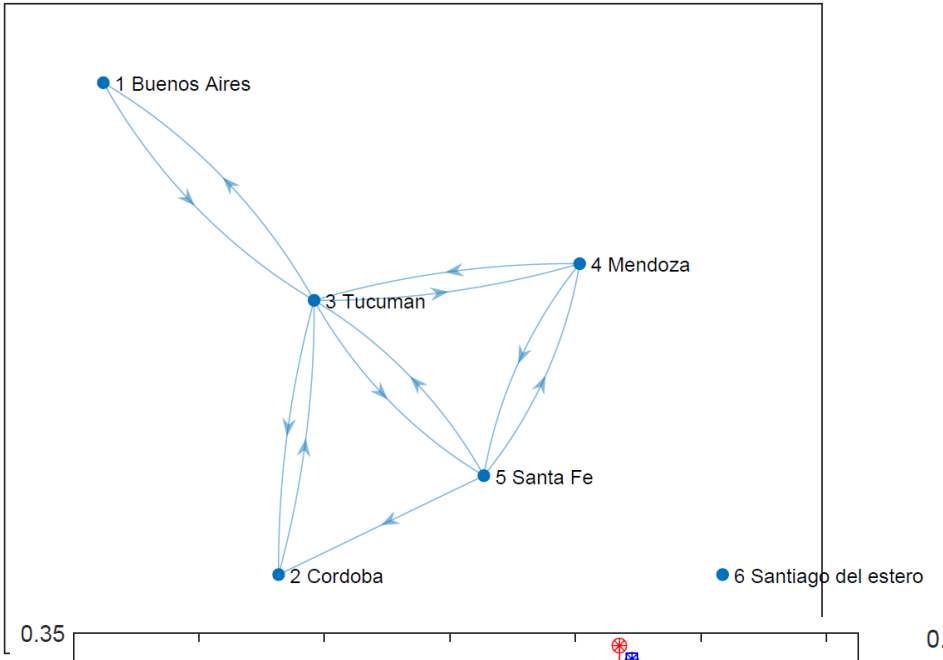




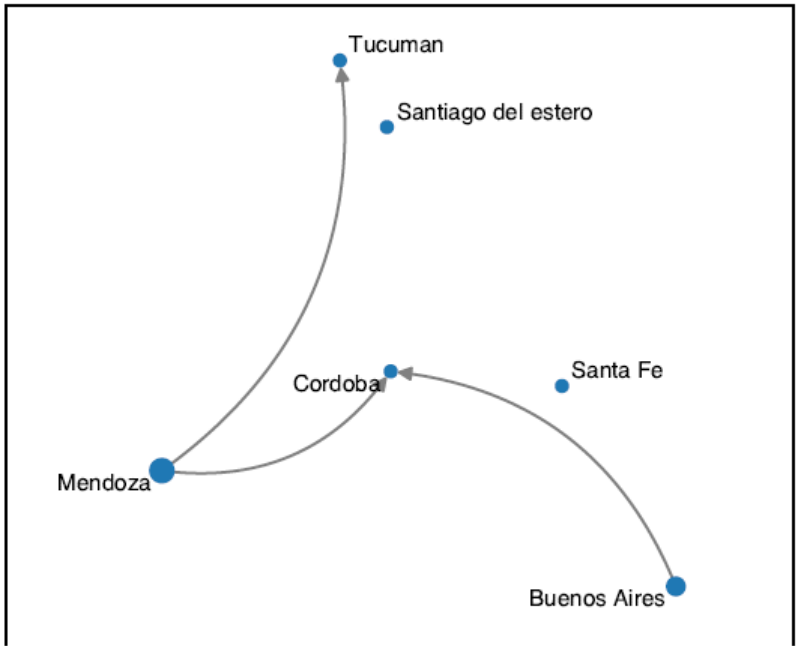
cristina.masoller@upc.edu @cristinamasoll1

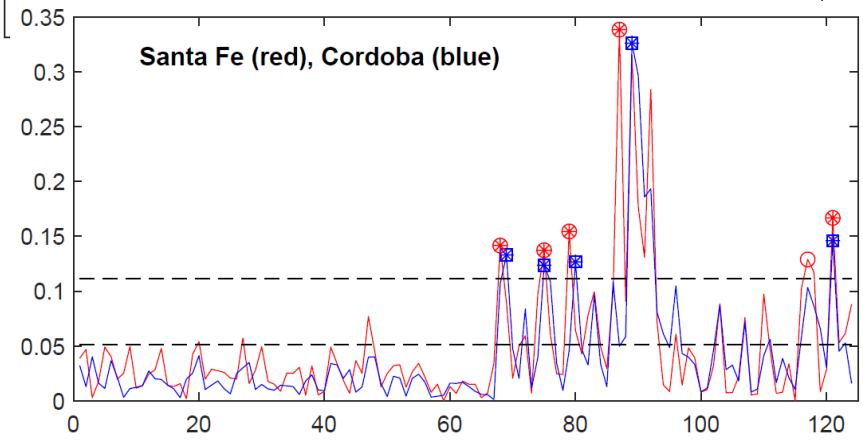
Results. Example topic "*Diego Luciani*"

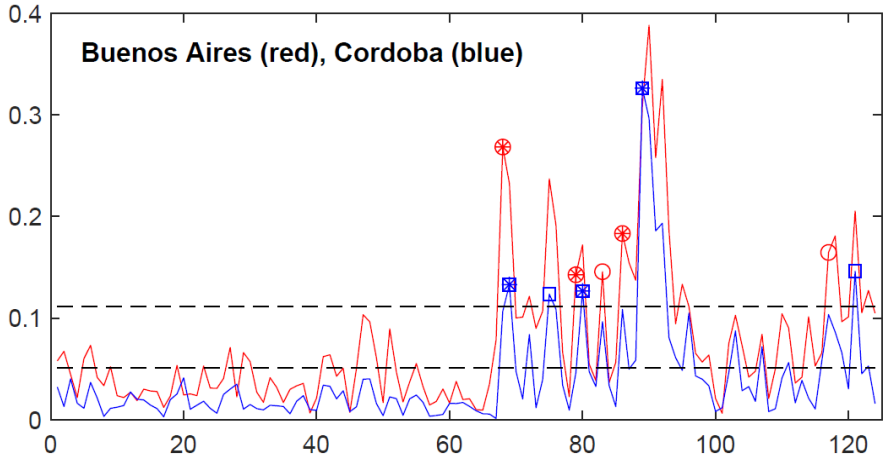
Qs or Qa significant



GC and TE significant

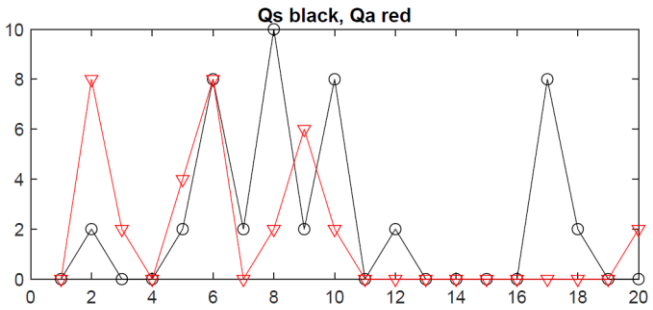


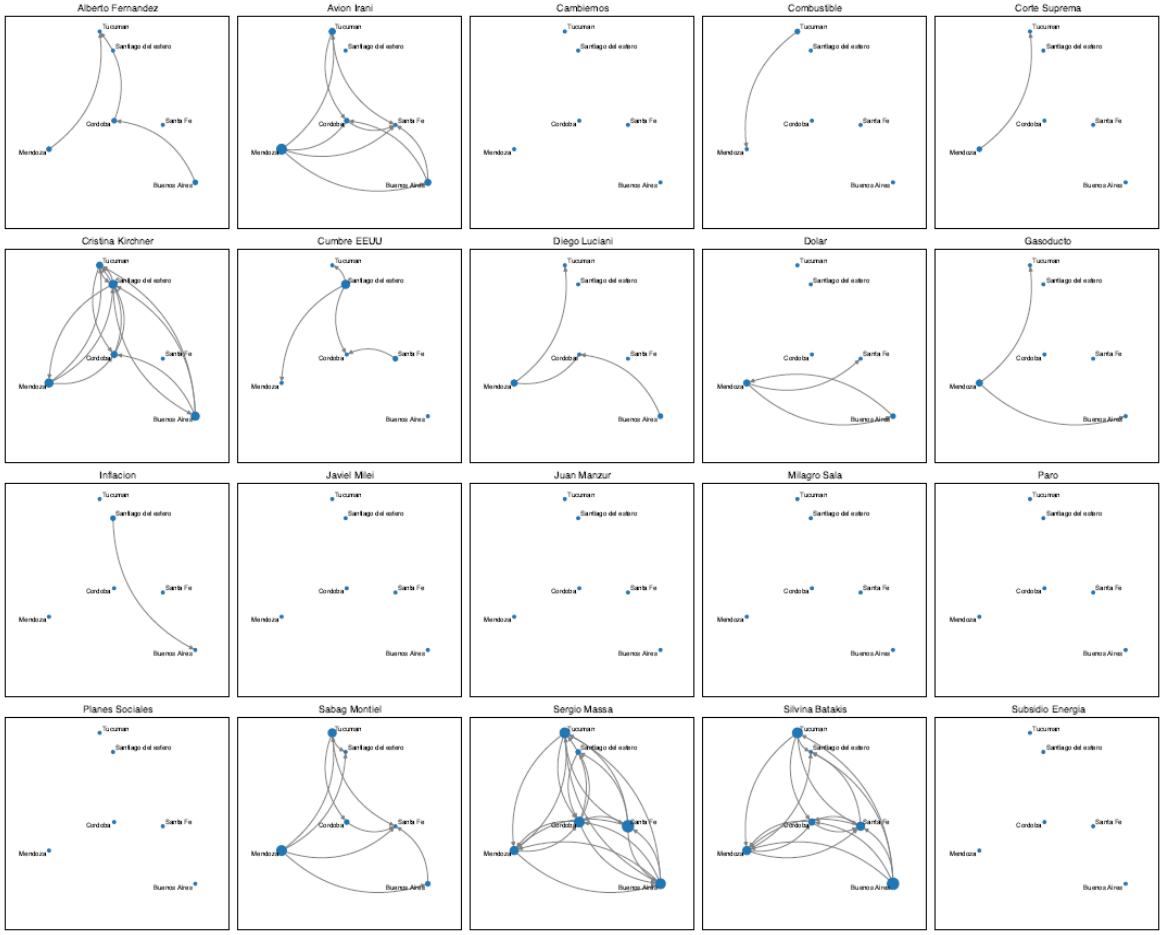




cristina.masoller@upc.edu @cristinamasoll1

Results





cristina.masoller@upc.edu @cristinamasoll1

Preliminary conclusions

- The networks obtained with event synchronization measures strongly depend on the criteria used to define the events and on the thresholds to obtain adjacency matrices.
- There are coincidences and differences with the networks obtained by combining GC and TE.
- It could be interesting to analyze networks obtained by counting only the number of synchronized events, without "penalizing" the occurrence of unsynchronized events.
- Temporal info can be obtained by counting events in time segments.

Thanks to my collaborators and
Thank you for your attention!



cristina.masoller@upc.edu @cristinamasoll1