EPJ Data Science 8, 35 (2019). DOI:10.1140/epjds/s13688-019-0213-9

Quantifying echo chamber effects in information spreading over political communication networks

Wesley Cota[1], Silvio C. Ferreira[1,2], Romualdo Pastor-Satorras[3], and Michele Starnini[4]

[1]Departamento de Física, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil
[2]National Institute of Science and Technology for Complex Systems (INCT-SC), Brazil
[3]Departament de Física, Universitat Politècnica de Catalunya, Campus Nord B4, Barcelona, Spain
[4]Data Science Laboratory, ISI Foundation, Torino, Italy

UFV Universidade Federal de Viçosa

wcota.me | wesley.cota@ufv.br | @wlcota

How does Twitter work?

- At a time t, a user can post a message (*tweet*) mentioning other users.
- There are two basic types of interactions:
  - **Reply or mention (@):** the message is directed to someone.
  - **Retweet (RT @):** a tweeted message is transmitted.
- A tweet can have *hashtags* (#).

Data collection and classification

- The data were collected from March 5th to December 31st of 2016 about the *impeachment* of Dilma Rousseff.
- From the total of 48.2 million tweets, 12.3 million contained at least one hashtag.

Table 1: Some events during the impeachment process.

Date	Event	Activity
Sun Mar 13	Biggest street manifestation against the government spread out in more than 250 cities.	−1
Wed Mar 16	Supreme court permits the constitution of a commission on the chamber of deputies	+1
Tue Mar 29	MDB (Brazilian political party "Movimento Democrático Brasileiro") left the government	+1
Sun Apr 17	Deputy chamber approves impeachment with 367 votes against 137	+1
Thu May 12	Rousseff leaves the presidency after Senate approval	−1
Mon May 23	Audio of Senator Romero Jucá saying "Estancar a sangria"	+1
Fri Jul 29	Rousseff delivers final arguments in the Deputy chamber	−1
Mon Aug 29	Rousseff's defense in Senate	+1
Wed Aug 31	Senate approves impeachment with 62 to 20 votes	+1

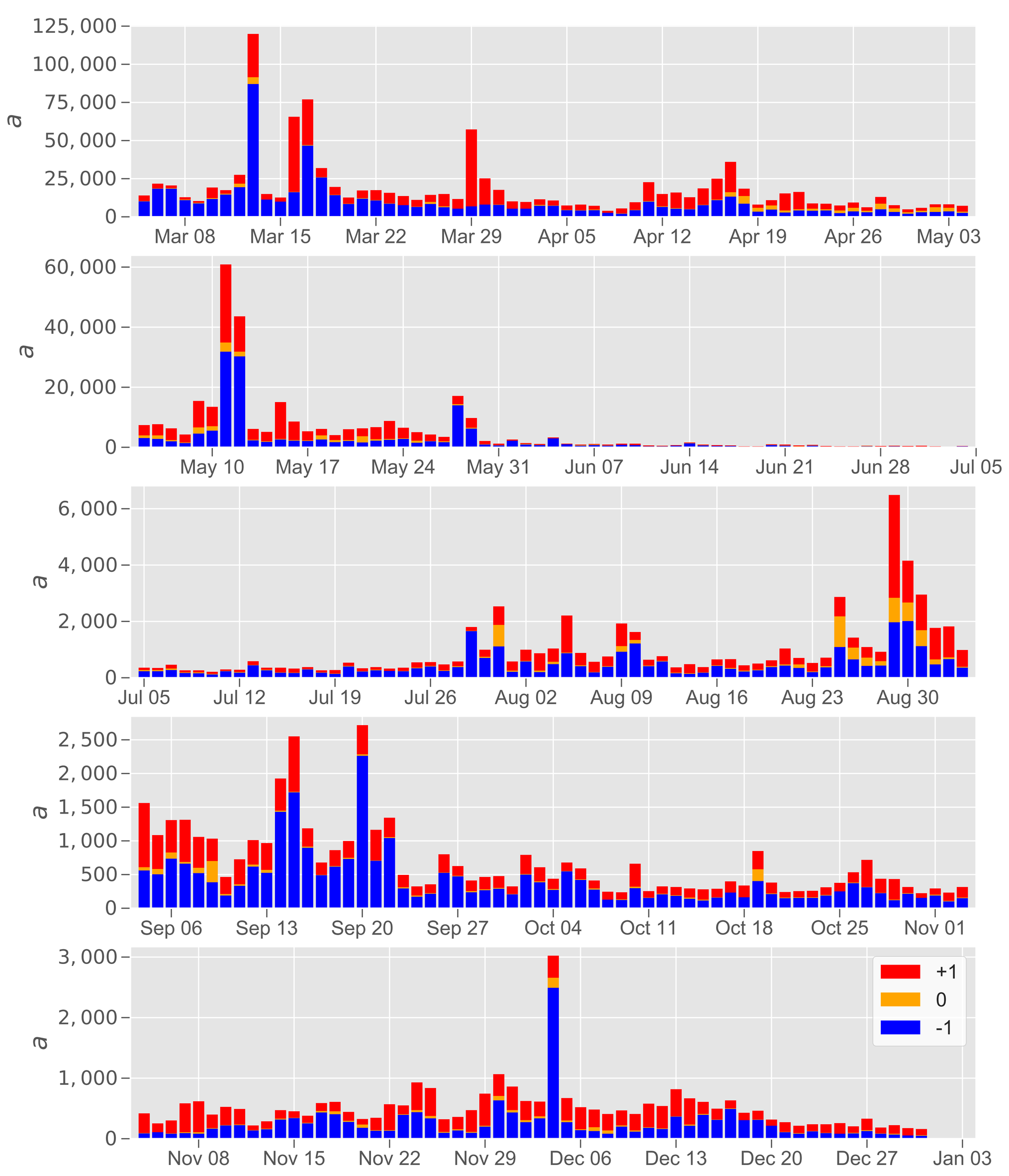


Figure 1: Daily activity of tweets with at least one hashtag.

Data classification

- Each tweet has a sentiment $s_t = \{-1, 0, +1\}$ given by the classified hashtags.
- From the 12.322.322 tweets, 7.486.459 (60.8%) had hashtags from the main classification.
- About 50% were retweets and discarded.
- The strongly connected component (SCC) has $N = 31.412$ users with $W = 1.552.389$ interactions.

Word clouds of hashtags

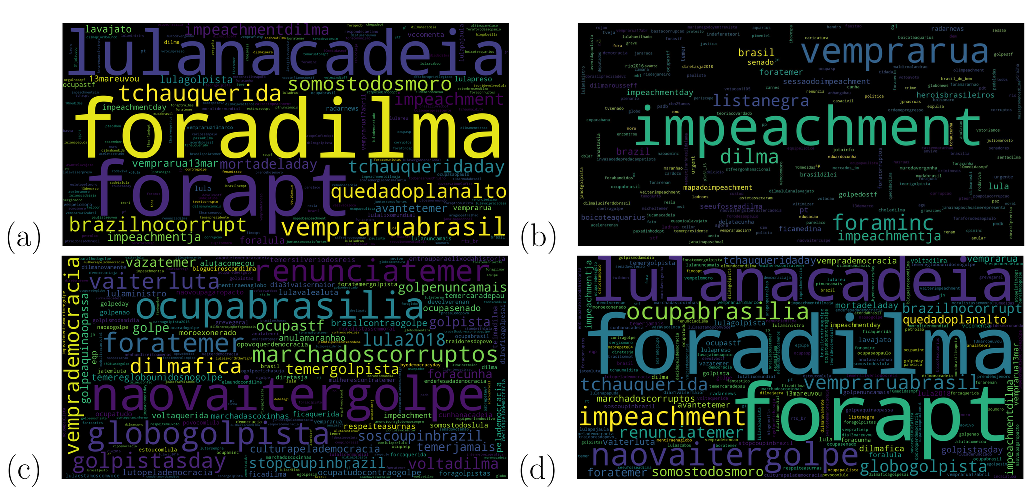


Figure 2: Word clouds for: (a) -1, (b) 0, (c) +1, (d) all sentiments.

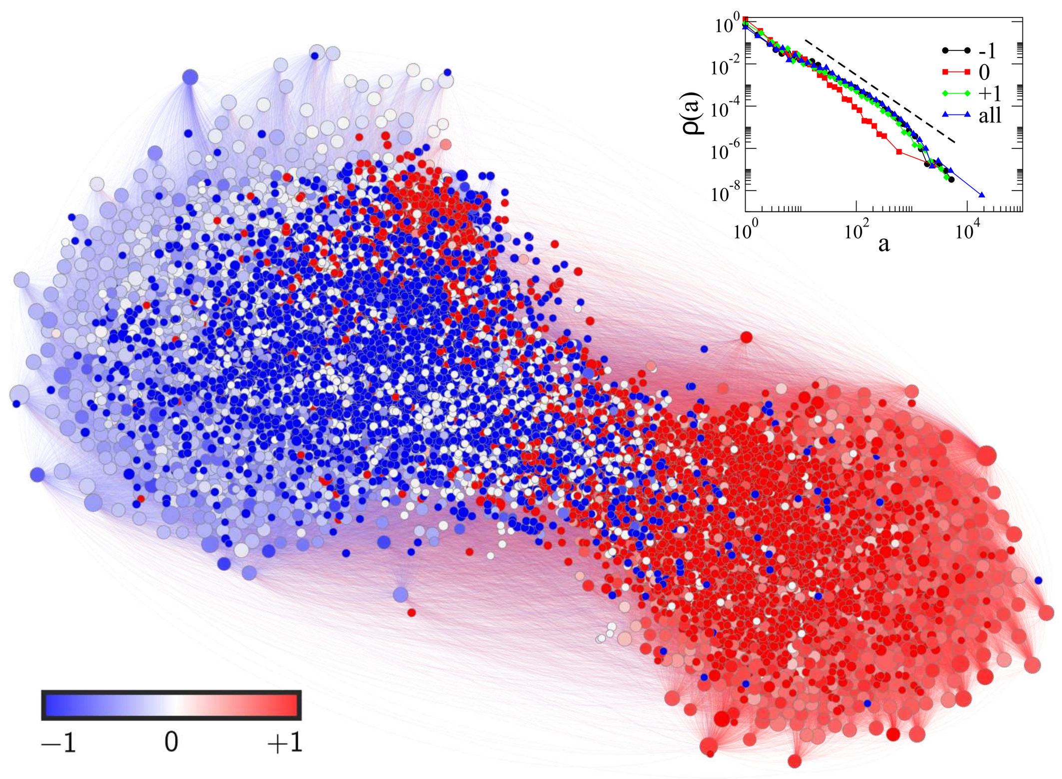


Figure 3: Visualization of the time-aggregated network.

Temporal and static networks

- **Users:** $\mathcal{N} = \{1, 2, \ldots, N\}$.
- Each interaction from i to j ($i, j \in \mathcal{N}$) occurs at time t with sentiment s_t: $e = (i, j, t, s_t)$.
- **Interactions:** $\mathcal{E} = \{e_1, e_2, \ldots, e_W\}$.
- Static representation by the **adjacency matrix** $\mathcal{A} = \{A_{ij}\}$ and the **weight matrix** $\mathcal{W} = \{W_{ij}\}$, counting all the interactions.
- Total number of links and interactions:

$$L = \sum_{ij} A_{ij} \quad \text{and} \quad W = \sum_{ij} W_{ij}. \tag{1}$$

- Activity of a sender a_i or receiver a_i^{IN} are defined as:

$$a_i = \sum_j W_{ij} \quad \text{and} \quad a_i^{\mathrm{IN}} = \sum_j W_{ji}. \tag{2}$$

- **Political position** and **content diversity**:

$$P_i \equiv \sum_{t=1}^{a_i} \frac{s_t}{a_i} \quad \text{and} \quad \varpi_i \equiv \sum_{t=1}^{a_i} \frac{(s_t - P_i)^2}{a_i} \tag{3}$$

- **Neighbors** and **static diversity**:

$$P_i^{\mathrm{NN}} \equiv \sum_j \frac{A_{ij} P_j}{k_{\mathrm{out},i}} \quad \text{and} \quad \varsigma_i \equiv \sum_j \frac{A_{ij}(P_j - P_i^{\mathrm{NN}})^2}{k_{\mathrm{out},i}} \tag{4}$$

Network properties

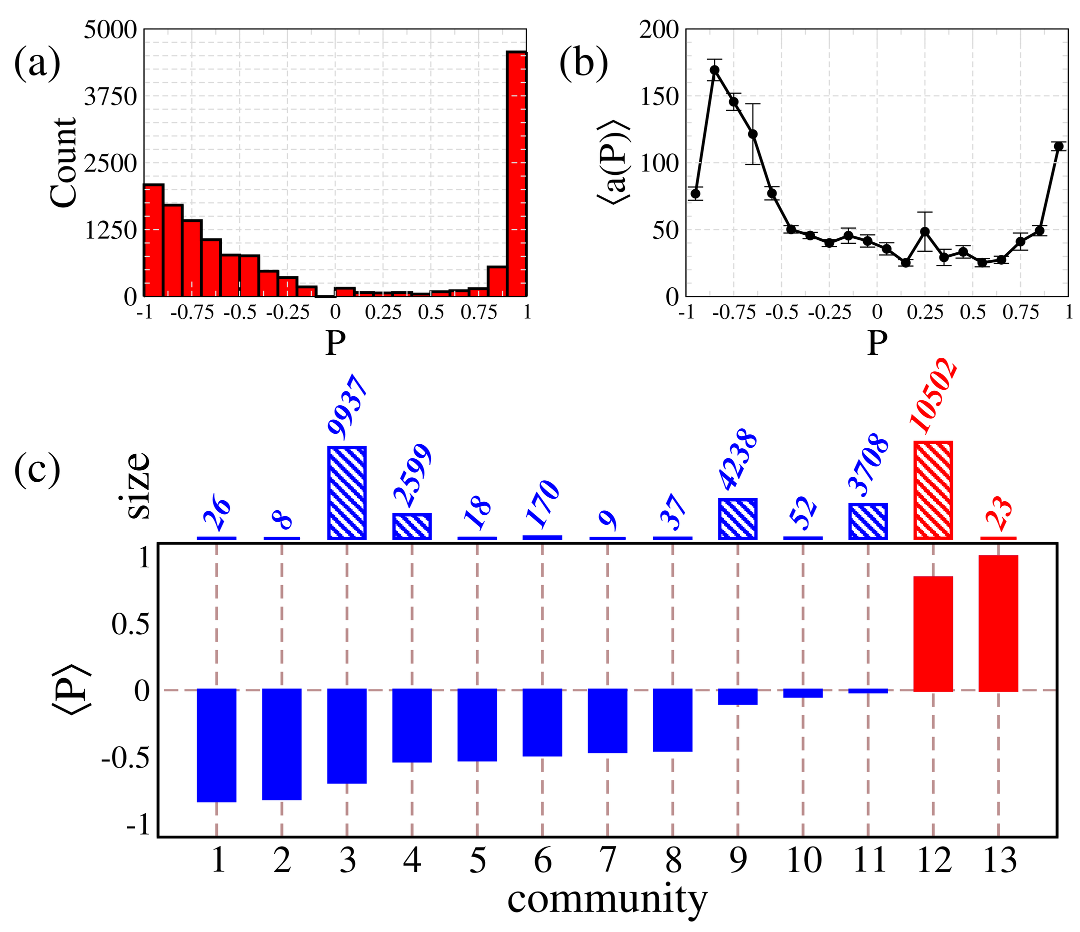


Figure 4: (a) Number of users and (b) average activity as function of position; (c) community size and average position.

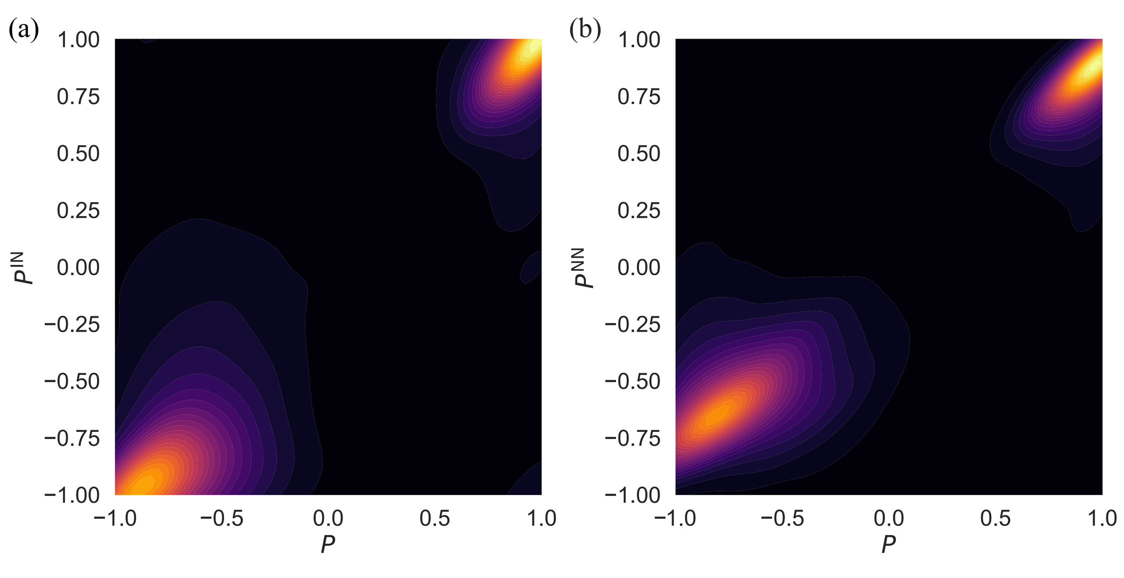
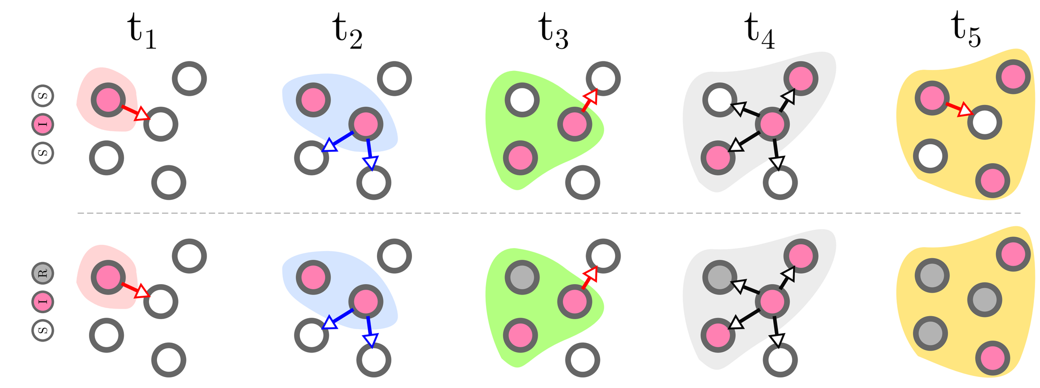


Figure 5: Contour maps for the position of the (a) tweets received P^{IN} and (b) nearest-neighbors P^{NN} against the position P.

Information spreading



- Simple models: Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered (SIR).
- Users become infected/informed with probability λ and susceptible or recovered after fixed time τ.
- Being i the source of information, the users that were infected form the *set of influence* $\mathcal{I}_i$.
- *Spreadability:*

$$S_i(\lambda, \tau) \equiv \frac{|\mathcal{I}_i(\lambda, \tau)|}{N}. \tag{5}$$

- Average (μ_i) and variance (σ_i) of the position of $\mathcal{I}_i$:

$$\mu_i \equiv \sum_{j \in \mathcal{I}_i} \frac{P_j}{|\mathcal{I}_i|} \quad \text{and} \quad \sigma_i \equiv \sum_{j \in \mathcal{I}_i} \frac{(P_j - \mu_i)^2}{|\mathcal{I}_i|} \tag{6}$$

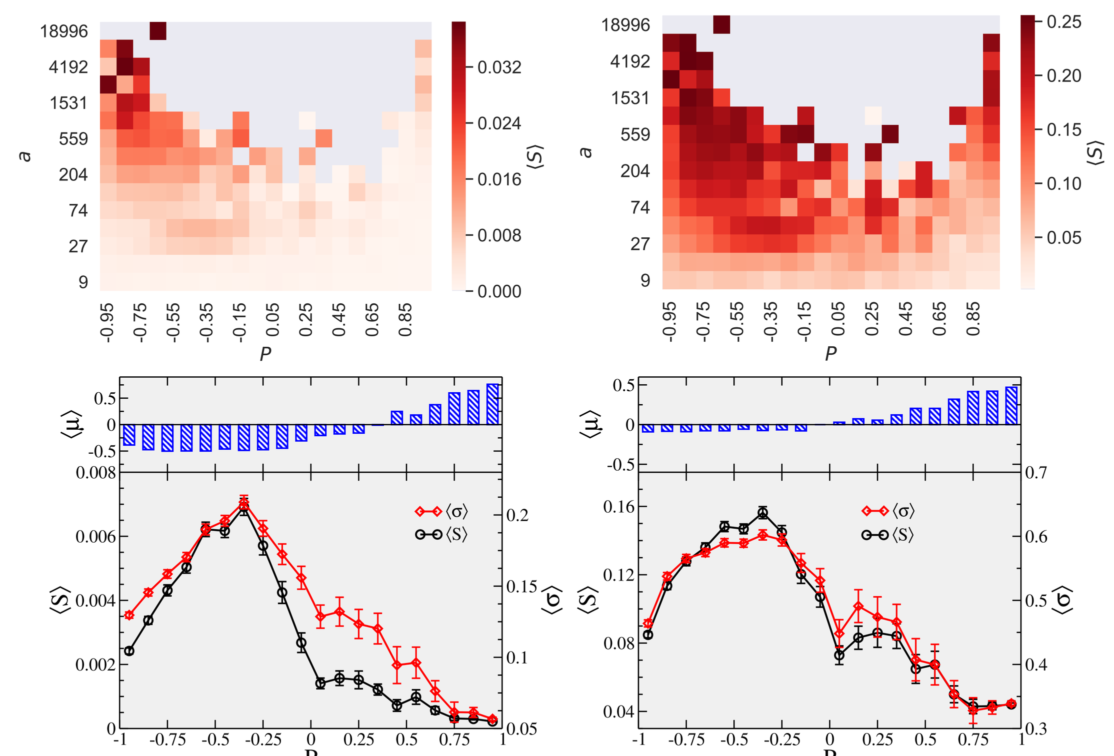


Figure 6: SIS, $\tau = 7$ days, $\lambda = 0.05$ (left) and 0.20 (right). Top: Heat maps of $\langle S \rangle$ versus P and a. Bottom: average values of S, μ and σ.

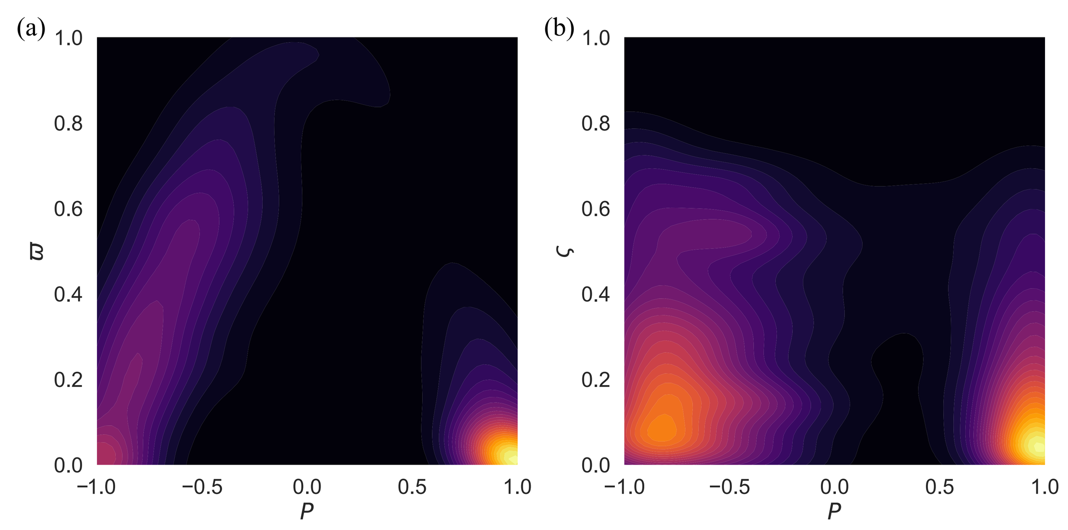


Figure 7: Contour maps of the (a) content diversity ϖ and (b) static diversity ς.

Final considerations

- The efficiency of users in propagating information strongly depends on their position.
- Users with pro-impeachment sentiments had a larger audience than those expressing anti-impeachment sentiments.
- Breaking the echo chambers is important for the information spreading.

Some references

1. W. Cota *et al.*: EPJ Data Science **8**, 35 (2019)
2. M. D. Conover *et al.*: EPJ Data Science **1**:6 (2012)
3. K. Garimella *et al.*: Proc. WWW '19, 913–922 (2018)
4. V. D. Blondel *et al.*: JSTAT **2008**(10), P10008 (2008)

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

CAPES CNPq FAPEMIG JSMF MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD ICREA