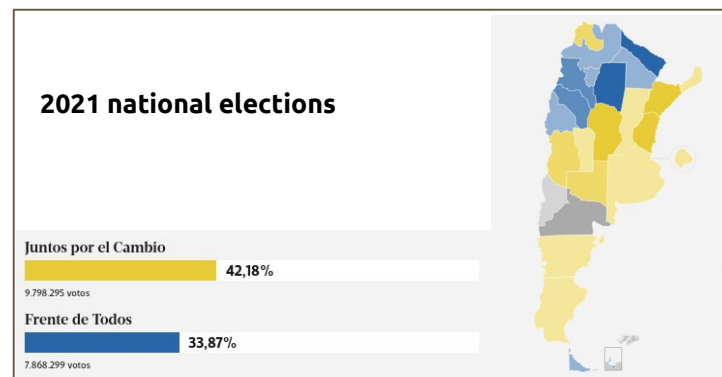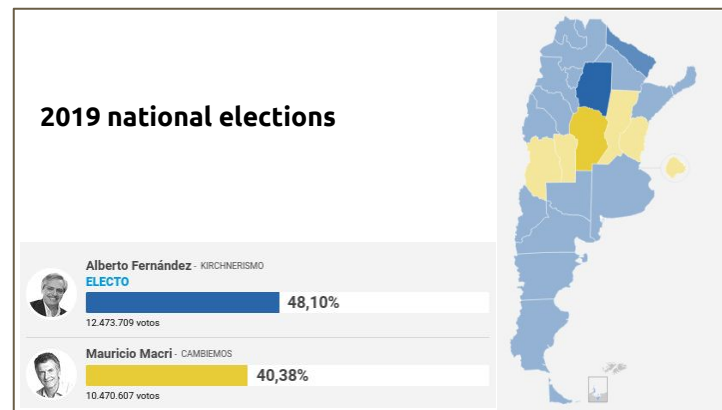# News sharing on Twitter

Emergent fragmentation of media agenda and persistent polarization

Tomás Cicchini, Sofía del Pozo, Enzo Tagliazucchi & Pablo Balenzuela

Political and Mediatic Argentinian Landscape

Political polarization

- There are two main political coalitions
  - FdT, kirchnerismo, leaded by CFK and AF, center-left
  - JxC, macrismo, leaded by MM, center right

2019 national elections

Alberto Fernández - KIRCHNERISMO
ELECTO
48,10%
12.473.709 votos

Mauricio Macri - CAMBIEMOS
40,38%
10.470.607 votos

2021 national elections

Juntos por el Cambio
42,18%
9.798.295 votos

Frente de Todos
33,87%
7.868.299 votos

Political and Mediatic Argentinian Landscape

Political polarization

- There are two main political coalitions
  - FdT, leaded by CFK, center-left
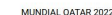 - JxC, leaded by MM, center right

Media fragmentation

- Studies accounts for the fragmentation of the media agenda. Mainly, there are two groups:
  - Clarin, La Nacion, Infobae: Center and Center-Right
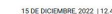 - Página 12, El Destape: Center, Center-Left

Twitter, users interactions and news consumption

On twitter, users interact between each others on multiple ways. Moreover, they share content, such as news.

Twitter, users interactions and news consumption

On twitter, users interact between each others on multiple ways. Moreover, they share content, such as news.
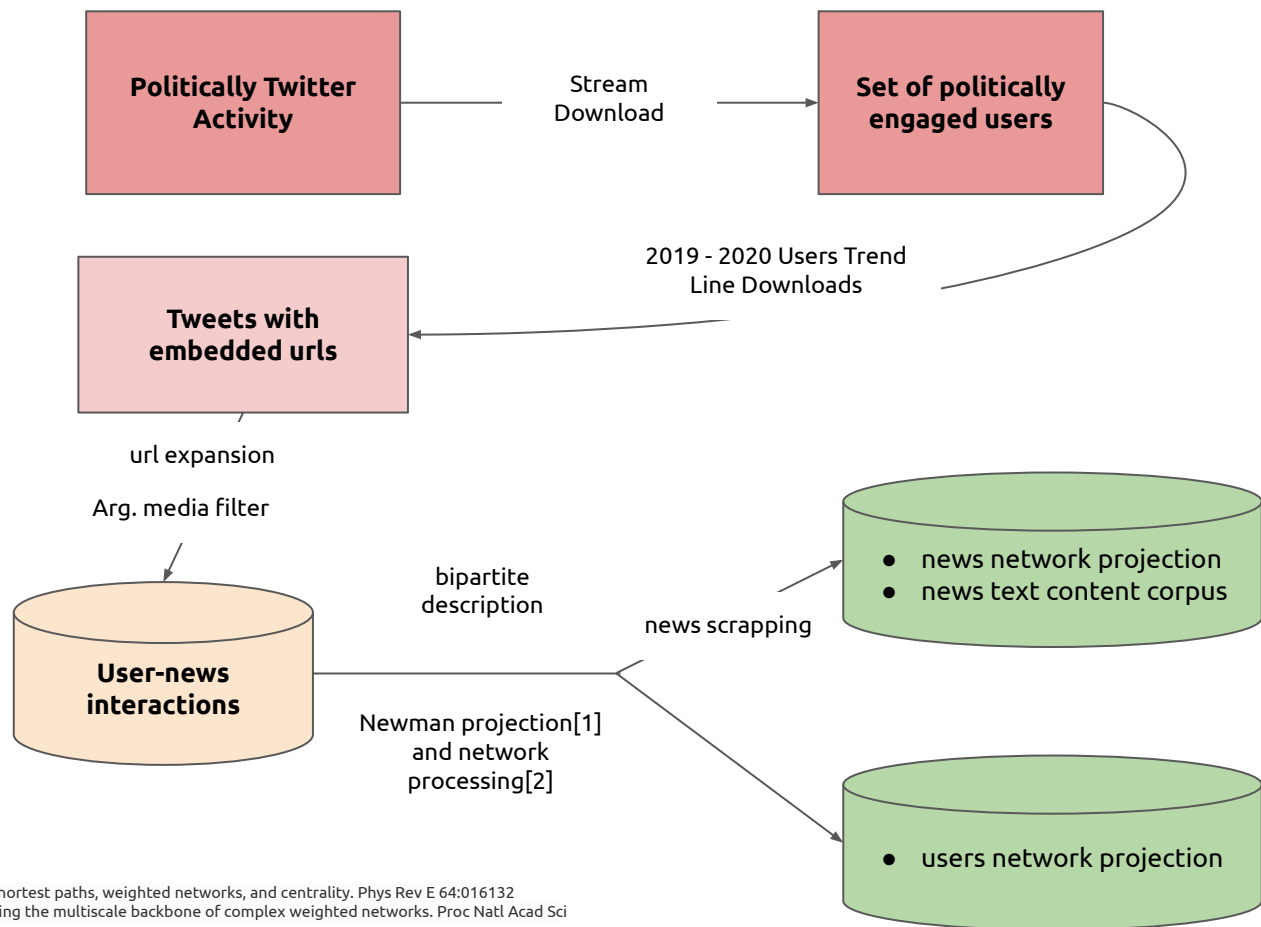
Main hypothesis

Individual news sharing behavior on Twitter is driven by personal preferences and ideological affinity, making it possible to detect emerging polarized groups as a consequence of social media interactions.

Tw. Argentina: a mirror of Argentina's political actuality?

In order to probe our hypothesis, we propose the next questions as a guideline:

- Is news sharing **constrained by features** related to users or news?
- Do users tend to form **clusters** according to their **preferences** in news consumption?
- Can the news consumption profiles of users be used to define and identify **echo-chambers**?
- Does media consumption in social media reflect the **political polarization** in Argentina?
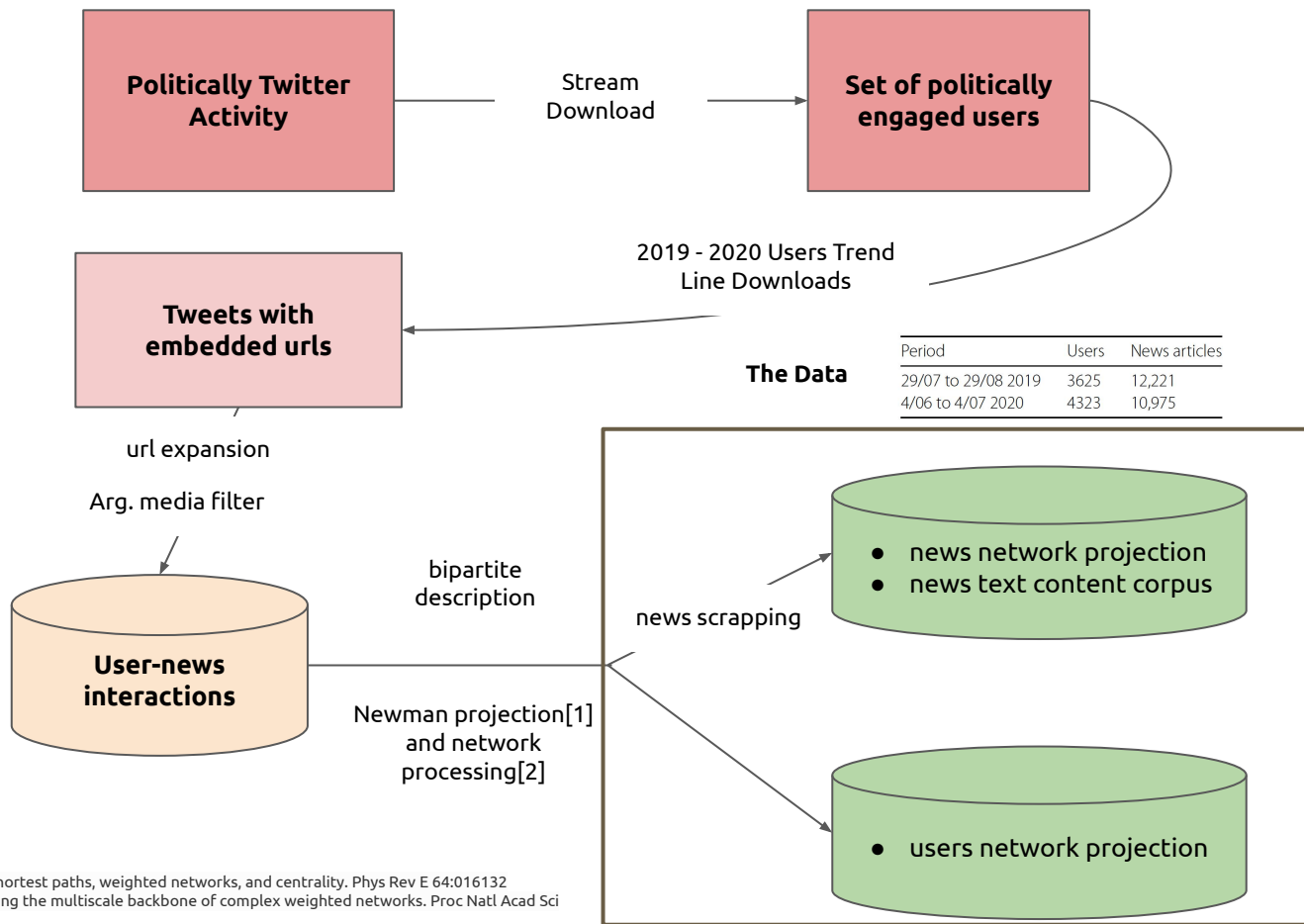
The workflow

Politically Twitter Activity → Stream Download → **Set of politically engaged users**

2019 - 2020 Users Trend Line Downloads → **Tweets with embedded urls**

url expansion

Arg. media filter

User-news interactions

bipartite description → news scrapping →
- news network projection
- news text content corpus

Newman projection[1] and network processing[2] →
- users network projection

[1] Newman MEJ (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E 64:016132
[2] Ángeles Serrano M, Boguñá M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. Proc Natl Acad Sci 106(16):6483–6488

The workflow

Politically Twitter Activity → Stream Download → **Set of politically engaged users**

2019 - 2020 Users Trend Line Downloads

Tweets with embedded urls

url expansion

Arg. media filter

The Data

Period	Users	News articles
29/07 to 29/08 2019	3625	12,221
4/06 to 4/07 2020	4323	10,975

User-news interactions

bipartite description

news scrapping

- news network projection
- news text content corpus

Newman projection[1] and network processing[2]

- users network projection

[1] Newman MEJ (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E 64:016132
[2] Ángeles Serrano M, Boguñá M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. Proc Natl Acad Sci 106(16):6483–6488

Methods: network, news and users metrics
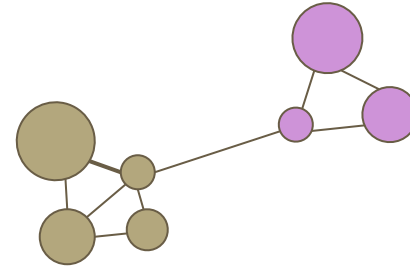
Network analysis:

- Communities detection [3][4]

[3] Blondel VD et al. (2008) Fast unfolding of communities in large networks. J Stat Mech 2008:P10008
[4] Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. Sci Rep 2(1):336
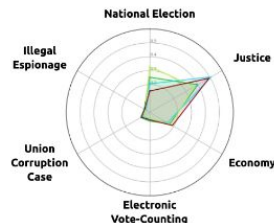
Methods: network, news and users metrics

Network analysis:

- Communities detection [3][4]

News metrics:

- Media outlet label
- Semantic content & topic decomposition (NMF) [6]
- Sentiment Bias (SA)

infobae

El impacto de la condena: CFK hundió su discurso en la grieta, transitó la ilegalidad y sacudió la interna

Mié, 7 Dic 2022

[3] Blondel VD et al. (2008) Fast unfolding of communities in large networks. J Stat Mech 2008:P10008
[4] Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. Sci Rep 2(1):336
[5] Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791

Methods: network, news and users metrics

Network analysis:

- Communities detection [3][4]

News metrics:

- Media outlet label
- Semantic content & topic decomposition (NMF) [6]
- Sentiment Bias (SA)

Users metrics:

- Media outlet consumption vector

$$m^i = (1, \dots, j, \dots, 0)$$

i: i-th user
j: number of shared news from j-th media outlet

[3] Blondel VD et al. (2008) Fast unfolding of communities in large networks. J Stat Mech 2008:P10008
[4] Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. Sci Rep 2(1):336
[5] Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791

The News Projection - 2019 and 2020 Analysis

The News Projection

2019 activity

12K news

2020 activity
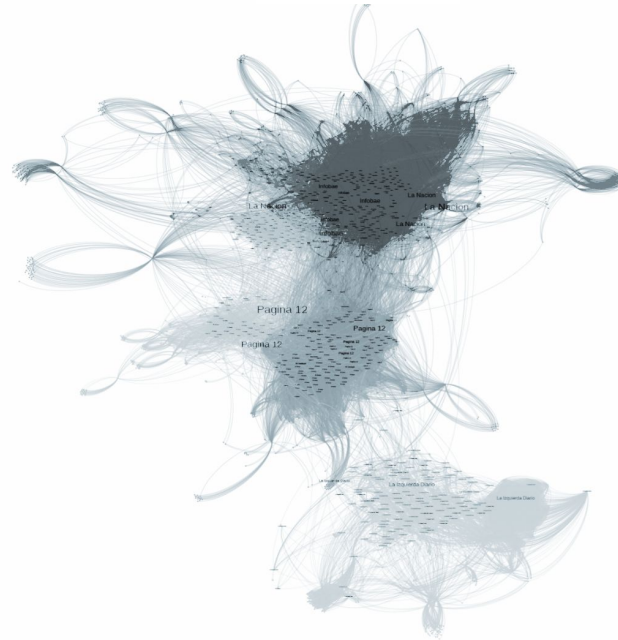
11K news

The News Projection

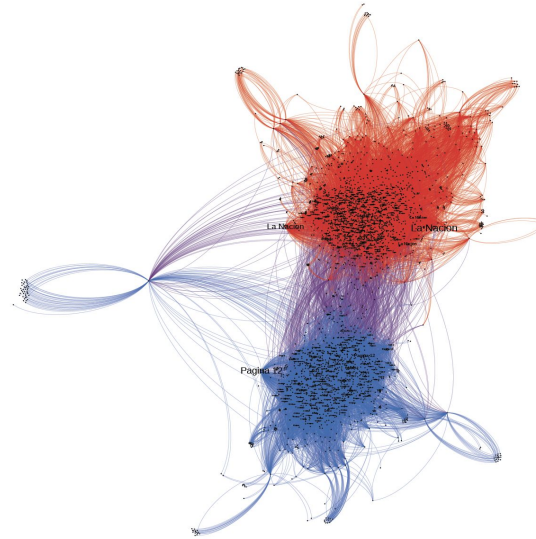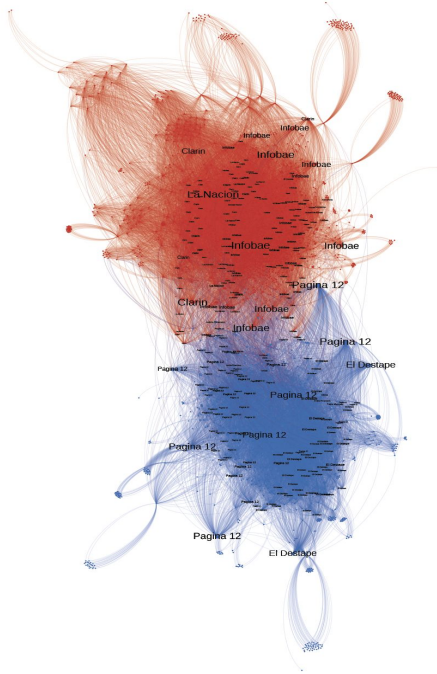Two main communities ----> 35% of the total news

2019 activity
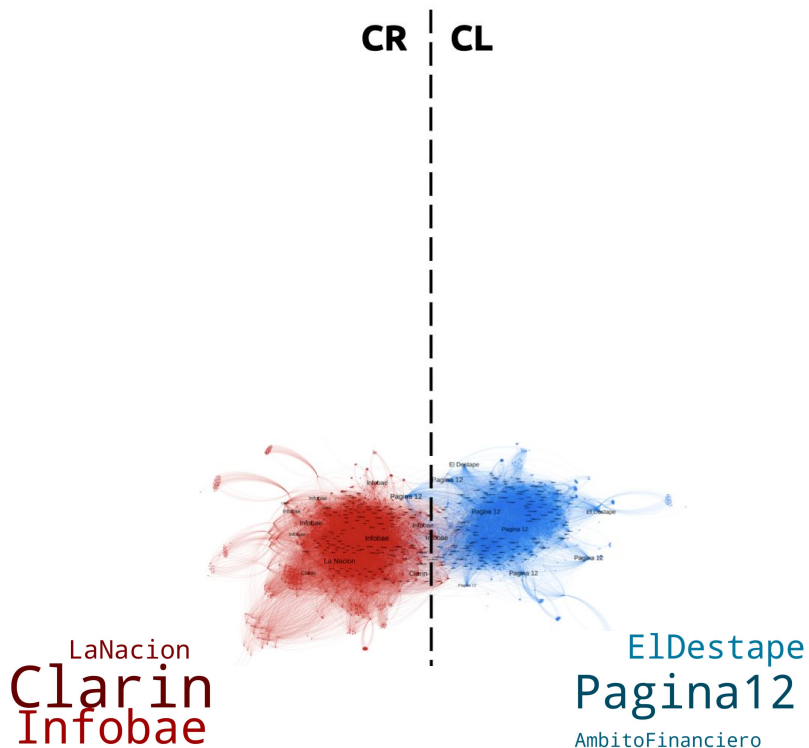
2020 activity

The News Projection: Main Two Communities

What are the media to which the news belong?

What are the news talking about?

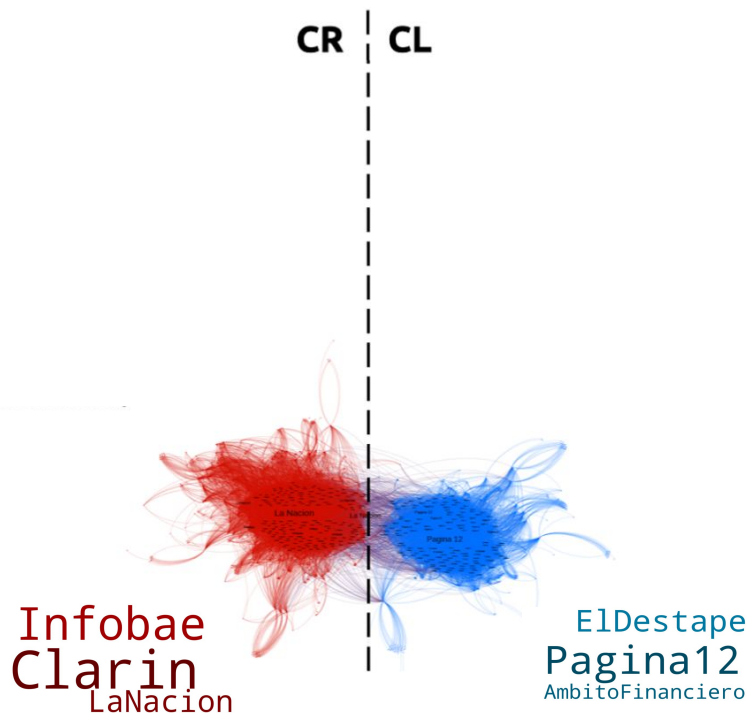# Media outlets and topics

La Nación, Clarín e Infobae are the main medias the first community; while Página 12 y El Destape Web are the main medias of the second community.
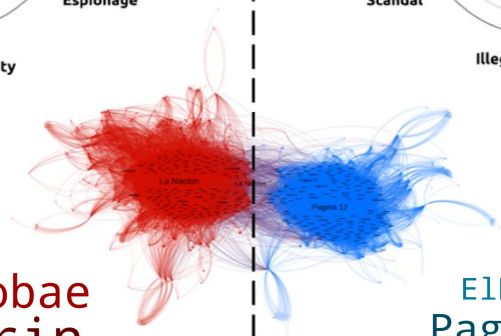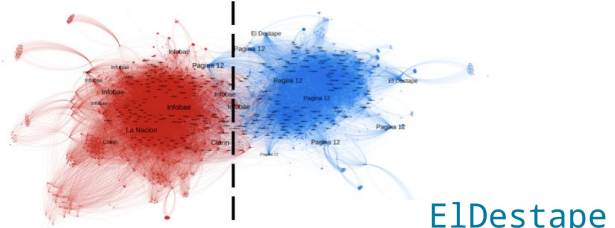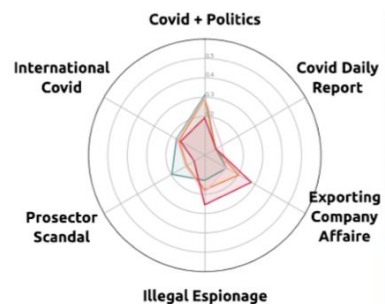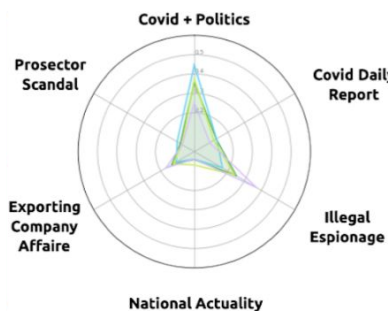
2019 activity

CR | CL

LaNacion
Clarin
Infobae

ElDestape
Pagina12
AmbitoFinanciero

2020 activity

CR | CL

Infobae
Clarin
LaNacion

ElDestape
Pagina12
AmbitoFinanciero

Media outlets and topics

2019 activity

CR | CL

National Election
Justice
Illegal Espionage
Union Corruption Case
Economy
Electronic Vote-Counting

National Actuality
Economy
National Election
Electronic Vote Counting
Illegal Espionage
Drug Traffic Affaire

LaNacion
Clarin
Infobae

ElDestape
Pagina12
AmbitoFinanciero

2020 activity

CR | CL

Covid + Politics
Covid Daily Report
Prosector Scandal
Illegal Espionage
Exporting Company Affaire
National Actuality

Covid + Politics
Covid Daily Report
International Covid
Exporting Company Affaire
Prosector Scandal
Illegal Espionage

Infobae
Clarin
LaNacion

ElDestape
Pagina12
AmbitoFinanciero

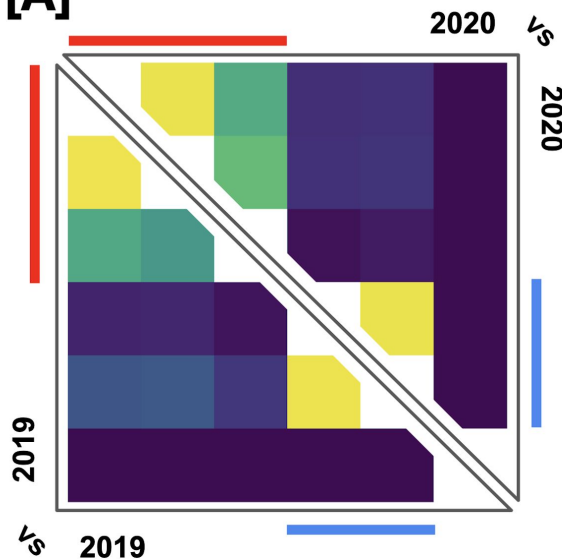# Quantifying Media Outlet Distributions
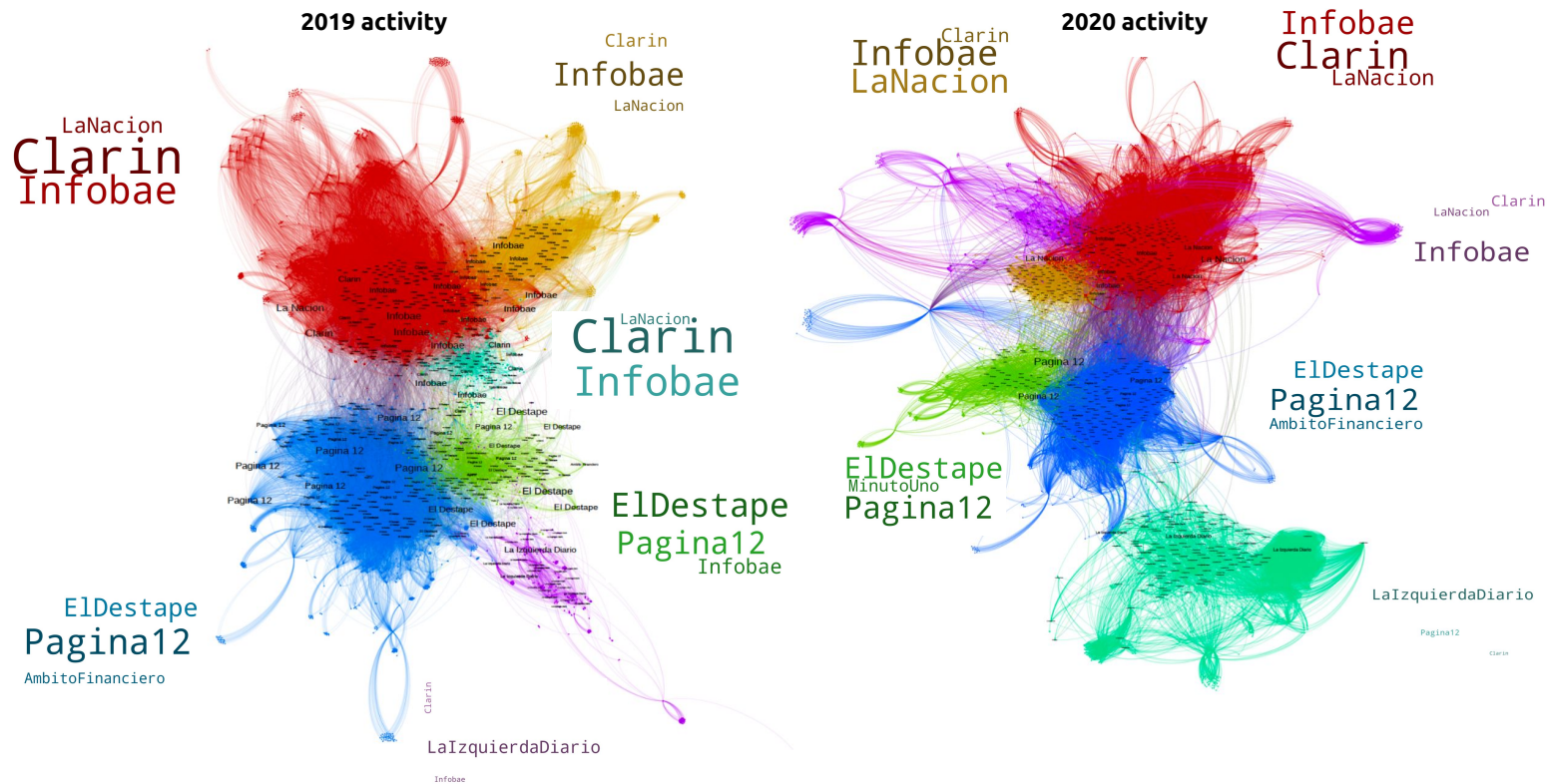
Media Outlet Distribution Similarities
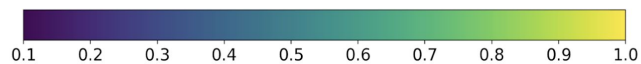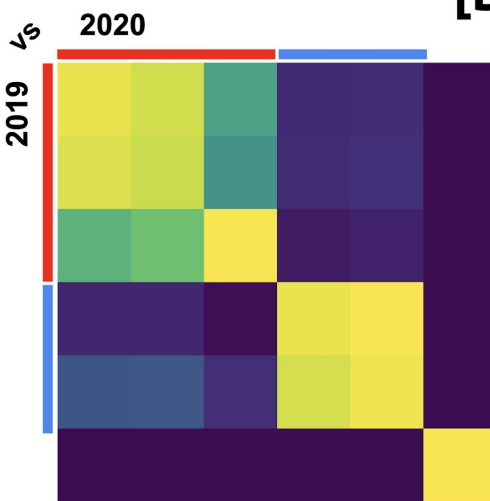
We assigned to each community a media outlet vector:: $C_i = [\#N^i_{M1}, \#N^i_{M2}, \ldots, \#N^i_{MN}]$
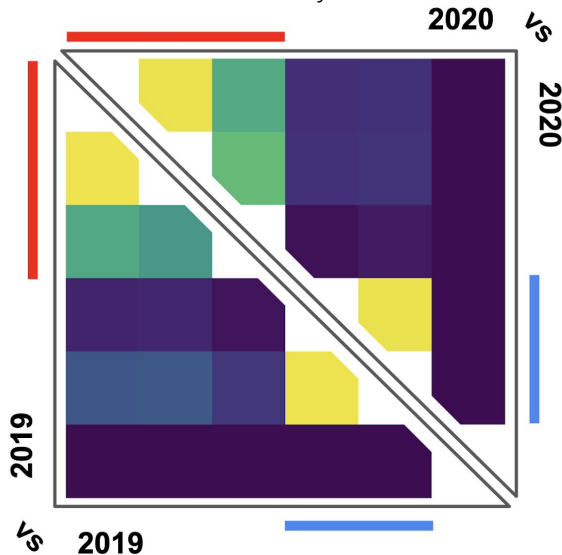

[B] We computed the similarity between communities media outlet vectors

Media Outlet Distribution Similarities

We assigned to each community a media outlet vector:: $C_i = [\#N^i_{M1}, \#N^i_{M2}, ..., \#N^i_{MN}]$

Between communities of the same year

Between different years communities
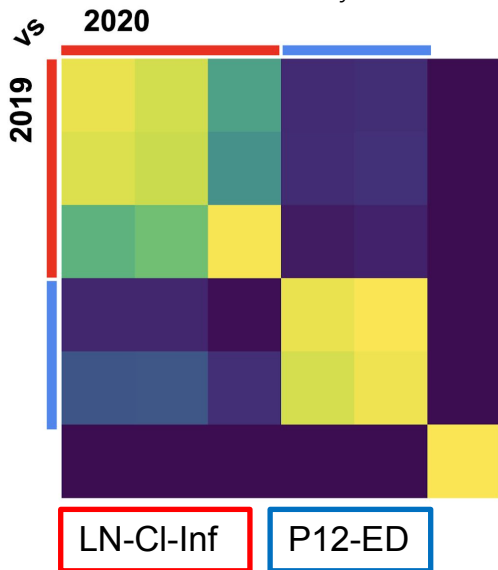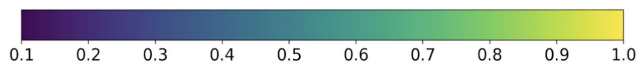
We computed the similarity between communities media outlet vectors

There are similars from one year to another, but different from the rest

The Izquierda Diario community remains present from one year to another

LN-Cl-Inf   P12-ED

Does the communities structure reflects the underlying political polarization?

We define the Sentiment Bias[6] as a balance between positive and negative [7] mentions of the different candidates of the 2019 National Election: Fernández-Fernández (FF) vs Macri-Pichetto(MP)

$$SB = (\#KF_+ - \#KF_-) - (\#MP_+ - \#MP_-)$$

[6] Albanese F, Pinto S, Semeshenko V, Balenzuela P (2020) Analyzing mass media influence using natural language processing and time series analysis. J Phys Complexity 1(2):025005
[7] Pérez J, Giudici J, Luque F (2021) Pysentimiento: a python toolkit for sentiment analysis and SocialNLP tasks

Does the communities structure reflects the underlying political polarization?

We define the Sentiment Bias[6] as a balance between positive and negative [7] mentions of the different candidates of the 2019 National Election: Fernández-Fernández (FF) vs Macri-Pichetto(MP)

$$SB = (\#KF_+ - \#KF_-) - (\#MP_+ - \#MP_-)$$

SB > 0 (Favorable to FF, unfavorable to MP)

[6] Albanese F, Pinto S, Semeshenko V, Balenzuela P (2020) Analyzing mass media influence using natural language processing and time series analysis. J Phys Complexity 1(2):025005

[7] Pérez J, Giudici J, Luque F (2021) Pysentimiento: a python toolkit for sentiment analysis and SocialNLP tasks

Does the communities structure reflects the underlying political polarization?

We define the Sentiment Bias[6] as a balance between positive and negative [7] mentions of the different candidates of the 2019 National Election: Fernández-Fernández (FF) vs Macri-Pichetto(MP)

$$SB = (\#KF_+ - \#KF_-) - (\#MP_+ - \#MP_-)$$

SB > 0 (Favorable to FF, unfavorable to MP)

SB < 0 (Favorable to MP, unfavorable to FF)

[6] Albanese F, Pinto S, Semeshenko V, Balenzuela P (2020) Analyzing mass media influence using natural language processing and time series analysis. J Phys Complexity 1(2):025005
[7] Pérez J, Giudici J, Luque F (2021) Pysentimiento: a python toolkit for sentiment analysis and SocialNLP tasks

Does the communities structure reflect the underlying political polarization?

We define the Sentiment Bias[6] as a balance between positive and negative [7] mentions of the different candidates of the 2019 National Election: Fernández-Fernández (FF) vs Macri-Pichetto(MP)

So that

SB > 0 (Favorable to FF, unfavorable to MP)
SB < 0 (Favorable to MP, unfavorable to FF)

| Sentiment Bias		
Year / Communities	2019	2020
Center-left group	0.17 ± 0.05 *	0.02 ± 0.05
Center-right group	-0.04 ± 0.05	-0.13 ± 0.05 *

[6] Albanese F, Pinto S, Semeshenko V, Balenzuela P (2020) Analyzing mass media influence using natural language processing and time series analysis. J Phys Complexity 1(2):025005
[7] Pérez J, Giudici J, Luque F (2021) Pysentimiento: a python toolkit for sentiment analysis and SocialNLP tasks

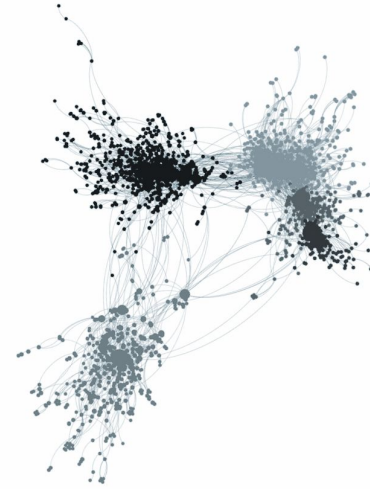# The Users Projection - 2019 and 2020 Analysis

Users Projection

3.6K users

4.3K users

Users Projection

Communities detection by Louvain
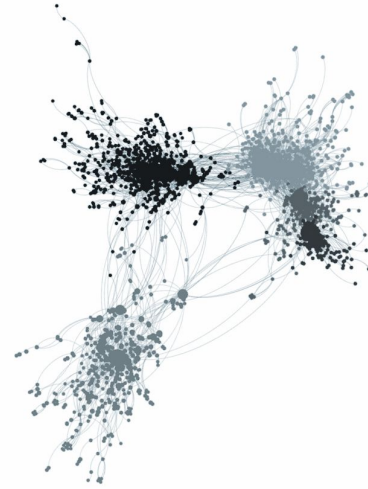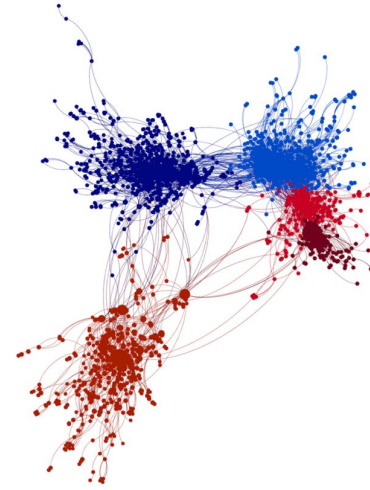
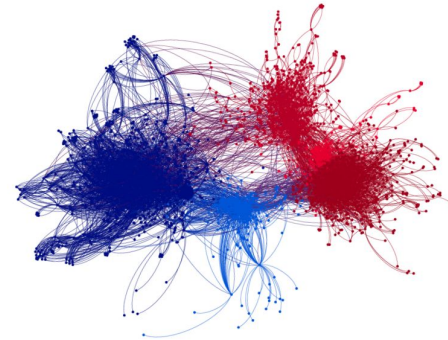**2019**

3.6K users

2020
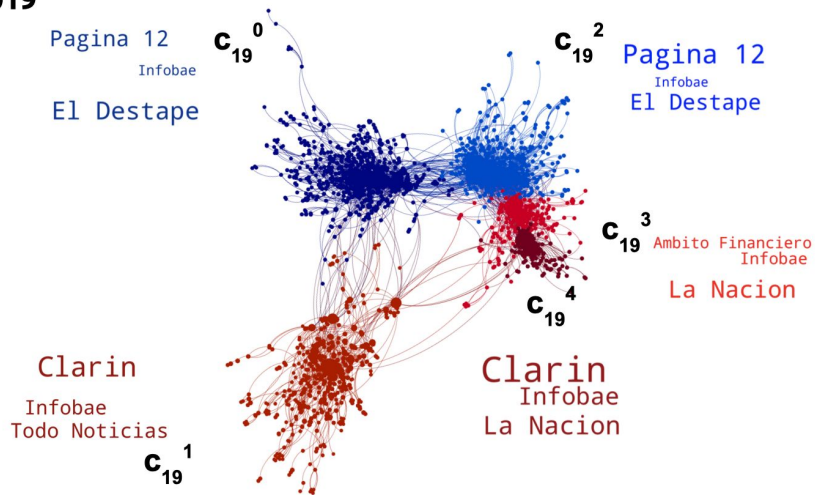

4.3K users

Users Projection

2019

2020

Users Projection

Describing users as a vector that accounts to the number of consumed news of each media outlet, we compute the similarities between users of the same communities and between different communities.
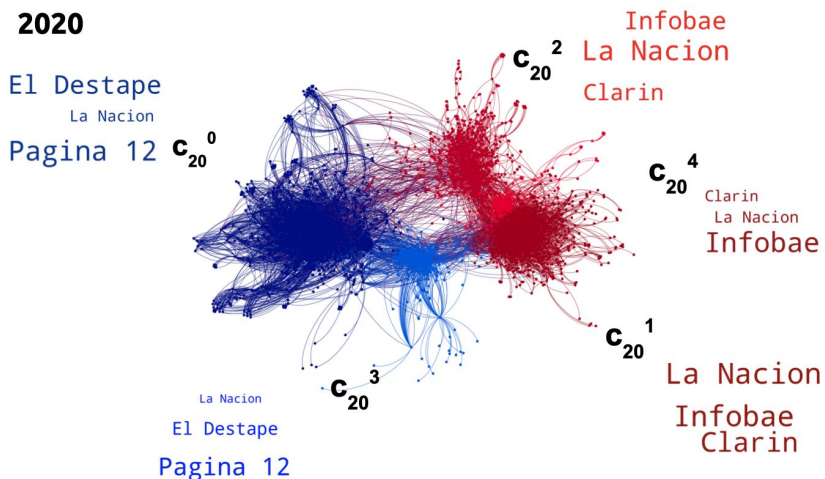
$$m^i = (1, \ldots, j, \ldots, 0)$$

i: i-th user
j: number of shared news from j-th media outlet

User Projection: Communities Similarities

Similarities between users media vector and the communities averages

Mean of the cosine similarities between the average media vector of the i-th community and all the users media vectors of the j-th community

Nodes (%)	Main Consumed Media Outlets
2019	
15.86	El Destape, Pagina 12
12.61	Clarin, Todo Noticias, Infobae, La Nacion
11.09	Pagina 12, El Destape
5.58	La Nacion, Infobae, Ambito Financiero, Pagina 12
5.31	Clarin, La Nacion, Infobae
2020	
16	Pagina 12, El Destape
11.17	La Nacion, Infobae, Clarin
8.11	La Nacion, Infobae, Clarin
7.35	Pagina 12, El Destape
7.12	Infobae

Discussion and Next Steps

Discussion

- Two large news groups emerge, identified by their belonging to the media where they are published: **Página 12 y El Destape on the one hand, Clarín-Infobae-La Nación on the other.**

- This distribution corresponds to a statistically **significant sentiment bias** in the mentions towards the candidates.

- Homogeneous groups of users are formed that can be identified by their pattern of media consumption.

- The results are consistent for different groups of users and for different years (with and without elections).

Next Steps

- Study the dependence of the structure in communities with the time window with which the original bipartite network was built (for more and less than a month).

- Replicate the analysis for different moments of interest.

- Extend the work to a temporal analysis, defining time-dependent networks in such a way as to be able to study the evolution of, for example, the sentiment associated with certain topics, or the behavior of users in relating to those who consume the same media.
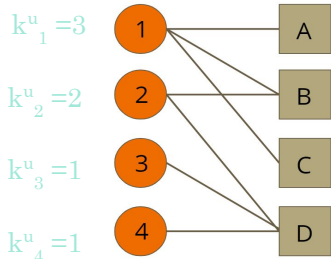
☀ **Thanks!** ☀
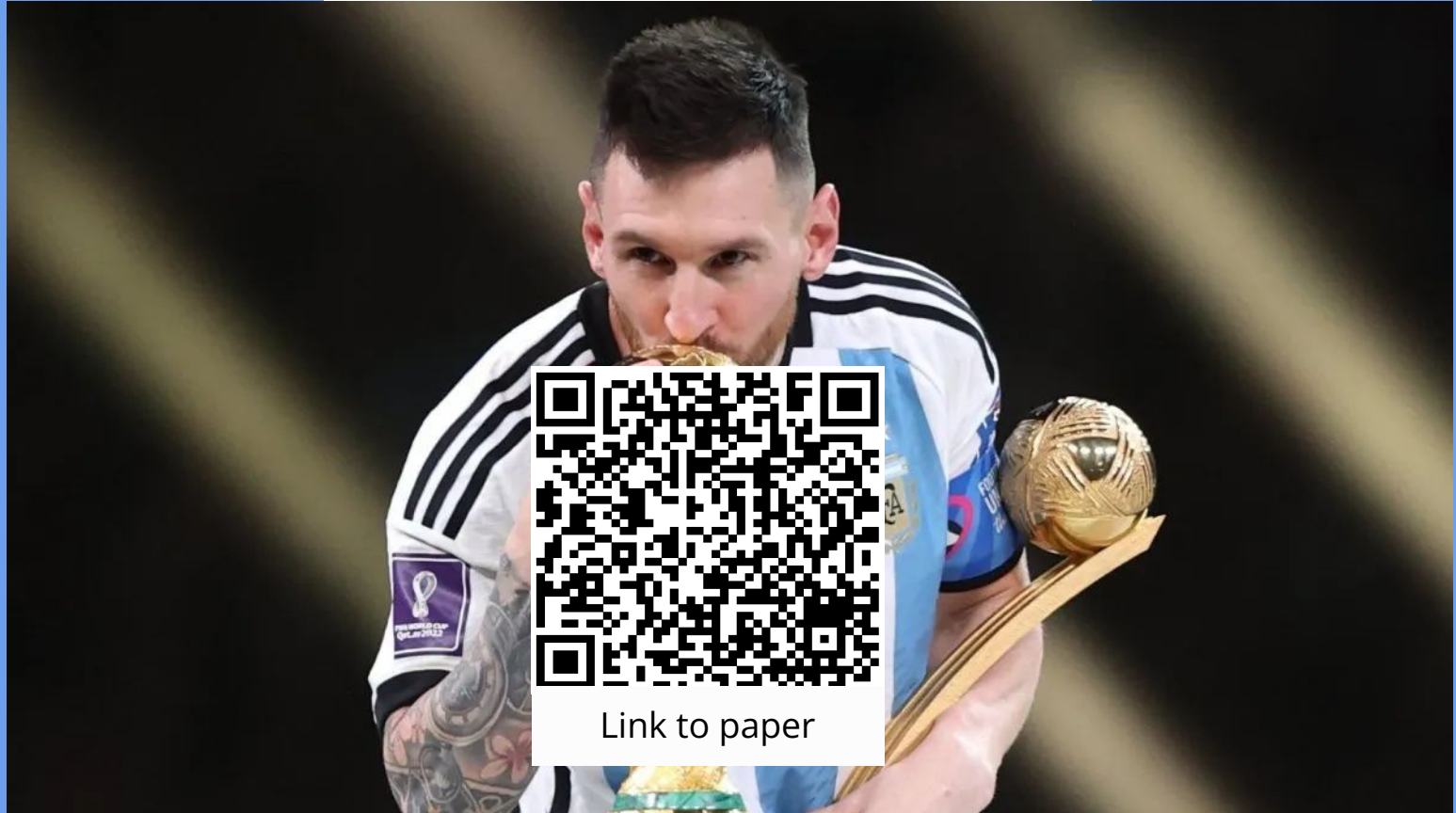
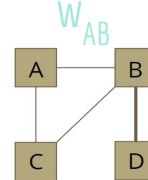Link to paper

Net projection

Bipartite Graph

Usuario   -   Noticia

$k^u_1 = 3$ 1 ——— A

$k^u_2 = 2$ 2 ——— B

$k^u_3 = 1$ 3 ——— C

$k^u_4 = 1$ 4 ——— D

$$A_{ij} = \begin{bmatrix} a_{11} & \dots & \dots & \dots & \dots & a_{1N} \\ & & & & & \\ & & & & & \\ a_{N1} & \dots & \dots & \dots & \dots & a_{NM} \end{bmatrix}$$

$u_1 \dots \quad \dots \quad \dots \quad \dots u_N$

n_1
$\dots$
$\dots$
$\dots$
n_M

$a_{ij} = 1$ if j-th user shares i-th news

Edge weight accounts how many times a pair of news was co-shared

$$w_{ij} = \sum_{\forall u} \frac{\delta_{iu} \delta_{ju}}{k_u - 1},$$

Monopartite graph (for example, news network)

W_{AB}

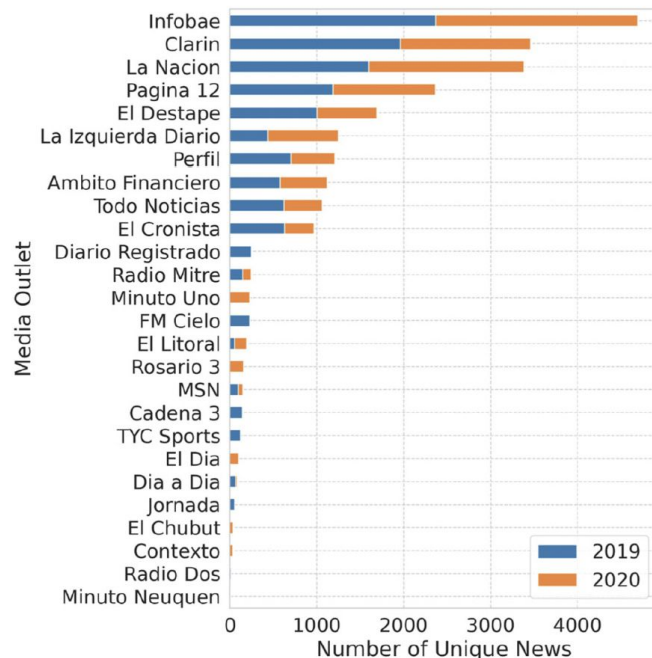A ——— B
| |
C   D

- Edge filtering with the disparity filter
- Community detection with Louvain
- Consensus network

Table 3 Main communities description for 2019 and 2020 news networks

Community Alias	Nodes (%)	Main Media Outlets	Main Topics
2019			
Center-Right I	18.68	Clarin, Infobae, La Nacion	Justice, National Elections, Economy
Center-Left I	14.75	Pagina 12, El Destape	National Actuality, Economy, National Elections
Center-Left II	8.34	El Destape, Pagina 12	National Actuality, National Election
International	5.5	Infobae	International Actuality
Center-Right II	4.4	Clarin, Infobae, La Nacion, Todo Noticias	Justice, National Election
Radical Left	3.4	Izquierda Diario	Politics, Public Health, Economy
2020			
Center-Right I	16.07	Clarin, Infobae, La Nacion	Covid + Politics, Illegal Espionage
Center-Left I	11.33	El Destape, Pagina 12	Covid + Politics, Exporting Company Affaire, Illegal Espionage
Center-Left II	8.61	Pagina 12, El Destape	National Actuality, Economy, Covid Daily Report
Center-Right II	7.43	Infobae, La Nacion, Clarin	International Covid, Covid Daily Report, Justice
Radical Left	7.1	La Izquierda Diario	Ilegal Espionage, Globally Known Racial Issue
International	4.5	Infobae	International Actuality

Table 5 Main communities features for 2019 and 2020 users networks

Nodes (%)	Main Consumed Media Outlets
2019	
15.86	El Destape, Pagina 12
12.61	Clarin, Todo Noticias, Infobae, La Nacion
11.09	Pagina 12, El Destape
5.58	La Nacion, Infobae, Ambito Financiero, Pagina 12
5.31	Clarin, La Nacion, Infobae
2020	
16	Pagina 12, El Destape
11.17	La Nacion, Infobae, Clarin
8.11	La Nacion, Infobae, Clarin
7.35	Pagina 12, El Destape
7.12	Infobae

Figure 1 Distribution of news articles according to media outlet. Here we plot the amount of articles of each media outlet corresponding to both analyzed years in the set of politically active users

Topic Word Clouds

CR 2019 | CR 2020 | CL 2019 | CL 2020 topic word clouds

- Topics differ between the same community on different years

- The covid+politics topic appeared in both communities of 2020

- In 2019, the national election topic highlighted the candidate names. We will use this fact while analysing the sentimental bias of the news