

Controversy detection and automated characterization of polarized communities by compression distances

Extended Abstract

Luis Pérez Miguel & David Arroyo Guardado

Grupo de investigación en Ciberseguridad Y Protección de la Privacidad (GiCP); Consejo Superior de Investigaciones Científicas (CSIC)

Special thanks: XAI-DisInfodemics: eXplainable AI for disinformation and conspiracy detection during infodemics



Financiado por
la Unión Europea
NextGenerationEU



- 1 Background
- 2 Methodology
- 3 Preliminary results
- 4 Key findings
- 5 Future direction

Background: Previous work

Characterization of new actors as belonging to an in-group by Normalized Compression Distance [1]

$$NCD(x, y) = \frac{C(x :: y) - C(x)}{C(y)}$$

- Low computation
- Fast and simple
- State-of-the-art performance in the domain of Covid Deniers

[1] Following Negationists on Twitter and Telegram: Application of NCD to the Analysis of Multiplatform Misinformation Dynamics, A. de Paz et al

Can we automatically characterize two sides of a discourse?

How to define controversy? How we separate each side?

- By the common use of different hashtags [2]
- Using NCD to measure the closeness in speech between users

[2] Quantifying Controversy in Social Media, K. Garmiella et al

Close to submitting this work, researchers lose access to twitter data.

Currently repivoting same goals for Telegram. But data has other structure:

- Inside channels, the discourse is not divided
- Between channels, the discourse may not have points in common

Preliminary results in a database with spanish politicians twitter accounts:
296 politicians timelines.

- 1 Background
- 2 Methodology**
- 3 Preliminary results
- 4 Key findings
- 5 Future direction

- Topic building around a hashtag in a timeframe
Hashtag $\Rightarrow$ Closest 20 hashtags by NCD distance
- Graph building
 - By hashtag use: *Connect two users if they have used at least 3 hashtag in common*
 - By NCD neighbours: *Connect each user to the 5 closest users by NCD distance*
- Separation in two communities by greedy modularity
- Quantification of controversy as Random Walk with Restart

$$RWC = P_{XX+}P_{YY+} - P_{XY+}P_{YX+} \quad , \text{ where } P_{XX+} = Pr[\text{from } X \text{ to } X^+]$$

- 1 Background
- 2 Methodology
- 3 Preliminary results**
- 4 Key findings
- 5 Future direction

Preliminary results: Covid

Hashtag: Covid19

- Not controversial (in institutional political communications)
- Both centralized graphs
- Low controversy measure

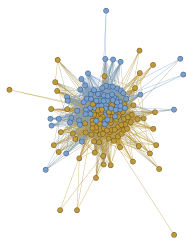


Figure: #Covid19 topic graph by hashtag use. $RWC = 0.116$



Figure: #Covid19 topic graph by NCD neighbours. $RWC = 0.272$

Preliminary results: Covid

The particulars of this dataset allows us to compare with political affiliation

- High overlap between sides of discourse and left/right political affiliation

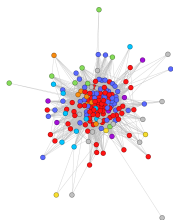


Figure: Political affiliation in #Covid19 by HT use. *Overlap = 56%*

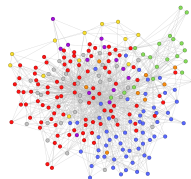


Figure: Political affiliation in #Covid19 by NCD. *Overlap = 78%*

Preliminary results: National budget distribution

Hashtag: pge2022

- Very controversial topic
- Several hashtags pushed by political parties
- Hashtag use shows a complete separation

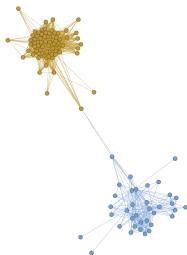


Figure: #pge2022 topic graph by hashtag use. $RWC = 0.982$

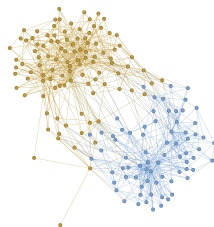


Figure: #pge2022 topic graph by NCD neighbours. $RWC = 0.508$

Preliminary results: National budget distribution

The separation by hashtag use easily generates an overview of each side:

- **Yellow:** #ingresomínimovital, #pgrecuperacionjusta, #avanzamos
- **Blue:** #noalainseguridadciudadana, #funcionariosdeprisiones, #pgedelaruina

Concurrent with the hashtags pushed by the left/right political parties

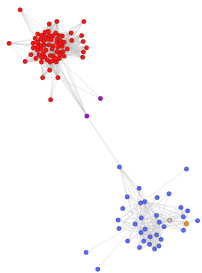


Figure: Political affiliation in #pge2022 by HT use. *Overlap* = 100%

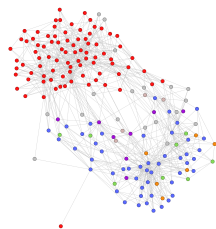


Figure: Political affiliation in #pge2022 by NCD. *Overlap* = 93%

Other remarks

Far-right party slogan: #soloquedavox

- Not divisive, used by a single community

	# use	NCD
<i>RWC</i>	0.163	0.242
<i>Overlap</i>	53%	60%

General hashtags: #ultimahora

- Not controversial. Higher overlap when using NCD

	# use	NCD
<i>RWC</i>	0.250	0.310
<i>Overlap</i>	59%	73%

Hashtags about political decisions: #sesiondecontrol

- Better performance by hashtag use.

	# use	NCD
<i>RWC</i>	0.602	0.480
<i>Overlap</i>	93%	91%

- 1 Background
- 2 Methodology
- 3 Preliminary results
- 4 Key findings**
- 5 Future direction

Key findings

- The difference in hashtags separates controversial topics in a more dramatic way. Might be due to observing users from a political organization, where slogan hashtags are being pushed.
- The separation by NCD overlaps with the political party affiliation. Not surprising, as we have seen compression distances are good in describing in-groups. A useful tool in finding repeated messages and arguments.

Content

- 1 Background
- 2 Methodology
- 3 Preliminary results
- 4 Key findings
- 5 Future direction**

Future direction

Monitoring of denier communities on telegram. Is there any use to check if a topic is being discussed differently between two channels?

- Instead of if $\Rightarrow$ How separated two sides are? Can we quantify radicalization of speech?
- Is the graphing by hashtag use applicable to a non-hashtag environment? Same results with keyword based topics?

Controversy by hashtag bears good performance in political party communications. Can we separate discourse in Congress Proceedings?

Thanks

Thanks for your attention

luis.perez@csic.es