

# Dinámicas del discurso político en una red de Twitter en la Argentina

## Elecciones presidenciales del 2015

Tomás Mussi Reyero<sup>1</sup>

Director: Dr. Ing. Mariano Beiró<sup>1</sup>

Co-director: Dr. Ing. José Ignacio Alvarez-Hamelin<sup>1</sup>

<sup>1</sup>Grupo de Redes Complejas y Comunicación de Datos  
INTECIN (Universidad de Buenos Aires-CONICET), Argentina

26 Noviembre 2019



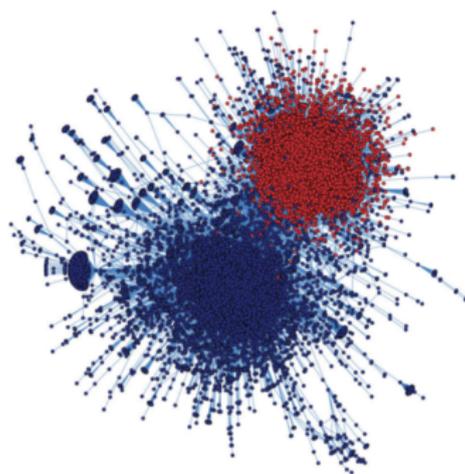
# Resumen

- 1 Estado del arte
- 2 Conjunto de datos y contexto político
- 3 Metodología
- 4 Resultados
- 5 Conclusiones

## Estado del arte

## Polarización política en Twitter (Conover *et al.*, 2011)

- Capturaron 250.000 *tweets* publicados por 45.000 usuarios durante las seis semanas previas a las elecciones legislativas de EE.UU. del 2010
- Optimización de modularidad de Newman para dividir la red de usuarios en dos comunidades.
- Algoritmo de detección de comunidades basado en propagación de etiquetas sobre red de *retweets* y menciones.
- **Red de *retweets* altamente polarizada entre republicanos y demócratas. Red de menciones con una única comunidad**



<sup>1</sup> Imagen extraída del trabajo

## Twitter : *Echo Chamber* o espacio público ? (Colleoni *et al.*, 2014)

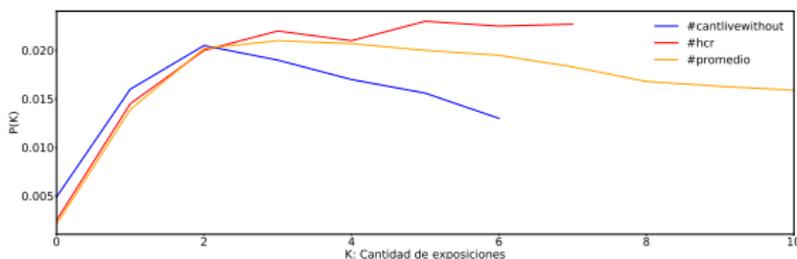
- Estudiaron si Twitter favorece el debate de ideas
- 467M de *tweets* publicados por 20M de usuarios de EE.UU. durante un período de siete meses en el año 2009.
- Análisis de sentimiento sobre los *tweets* para clasificar usuarios entre demócratas o republicanos
- Medición de homofilia para el grupo de demócratas y republicanos por separado.
- **Demócratas exhiben mayor homofilia entre ellos y republicanos exhiben mayor homofilia con las cuentas oficiales.**



<sup>2</sup>Imagen extraída del portal *The conversation*

<http://theconversation.com/the-problem-of-living-inside-echo-chambers-110486>

## Diferencias en los mecanismos de difusión de información en Twitter (Romero *et al.*, 2011)



- Estudiaron como se propaga la utilización de *hashtags* en la red.
- La cantidad de exposiciones y el tiempo transcurrido desde la última exposición son factores que afectan a la propagación.

- Capturaron más de 3.000M de *tweets* publicados por 60M de usuarios durante seis meses.
- **Adopción de posturas políticas controversiales requiere más exposiciones que la adopción de otras ideas. (Complex contagion)**

## Conjunto de datos y contexto político

## Contexto político : elecciones presidenciales del 2015

- Presidente saliente : Cristina Kirchner (dos mandatos)
- Candidatos capturados :



Mauricio  
Macri



Daniel  
Scioli



Sergio  
Massa



Margarita  
Stolbizer

Partido	PRO Cambiemos	PJ - FpV	Frente renovador	GEN Progresistas
---------	------------------	----------	---------------------	---------------------

## Contexto político : elecciones presidenciales del 2015

- Presidente saliente : Cristina Kirchner (dos mandatos)
- Candidatos capturados :



Mauricio  
Macri



Daniel  
Scioli



Sergio  
Massa



Margarita  
Stolbizer

Partido

PRO  
Cambiamos

PJ - FpV

Frente  
renovador

GEN  
Progresistas

1<sup>ra</sup> vuelta

**34%**

**37%**

**21%**

**3%**

## Contexto político : elecciones presidenciales del 2015

- Presidente saliente : Cristina Kirchner (dos mandatos)
- Candidatos capturados :



Mauricio  
Macri



Daniel  
Scioli



Sergio  
Massa

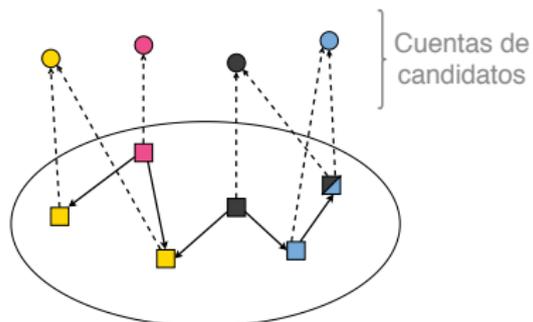


Margarita  
Stolbizer

Partido	PRO Cambiamos	PJ - FpV	Frente renovador	GEN Progresistas
1 <sup>ra</sup> vuelta	<b>34%</b>	<b>37%</b>	<b>21%</b>	<b>3%</b>
Ballotage	<b>51%</b>	<b>49%</b>	-	-

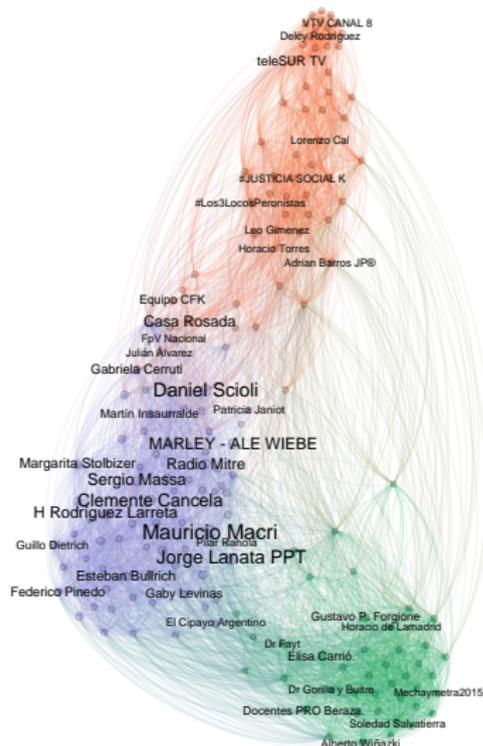
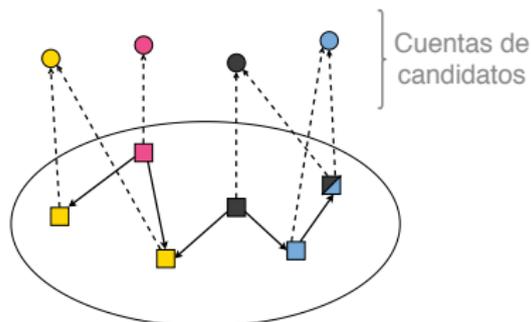
## Conjunto de datos

- 300,000 usuarios que siguen a un político
- 7M relaciones de seguidor-amigo entre ellos
- 50M tweets capturados
- Período : Julio 2015 - Marzo 2016



# Conjunto de datos

- 300,000 usuarios que siguen a un político
- 7M relaciones de seguidor-amigo entre ellos
- 50M tweets capturados
- Período : Julio 2015 - Marzo 2016



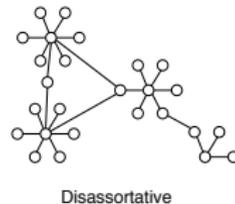
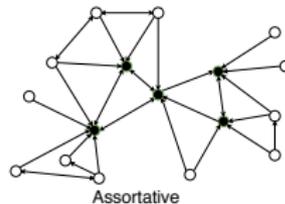
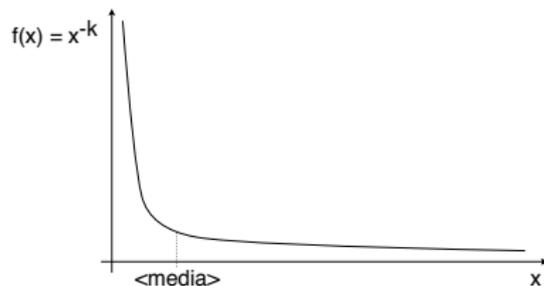
## Metodología

## Enfoque : Redes complejas

- Redes de gran tamaño con características y topología particulares
- Ej : redes sociales, interacción entre proteínas, redes de transporte de energía, Internet, la Web, entre otros.
- Distribución de grados de nodos en forma de cola pesada
- *Small world*
- Coeficiente de *clustering* elevado
- *Assortative mixing*

## Enfoque : Redes complejas

- Redes de gran tamaño con características y topología particulares
- Ej : redes sociales, interacción entre proteínas, redes de transporte de energía, Internet, la Web, entre otros.
- Distribución de grados de nodos en forma de cola pesada
- *Small world*
- Coeficiente de *clustering* elevado
- *Assortative mixing*



## Detección de tópicos

UsUARIO ←

Hashtag ←

Fecha y hora ←

**More**  
@moreira\_ngo

#Elecciones2015 #Argentina :Cerró la votación: final ajustado con posibilidad de #ballotage #Macri vs #Scioli  
<https://t.co/yGPB1mjeXM>

1:37 - 25 de octubre 2015

The image shows a tweet interface with three blue boxes highlighting specific elements. The first box encloses the user's profile picture, name 'More', and handle '@moreira\_ngo'. The second box encloses the hashtag '#Elecciones2015'. The third box encloses the timestamp '1:37 - 25 de octubre 2015'. Arrows point from the labels 'UsUARIO', 'Hashtag', and 'Fecha y hora' to these respective boxes. The tweet text includes '#Argentina :Cerró la votación: final ajustado con posibilidad de #ballotage #Macri vs #Scioli' and a URL.

## Detección de tópicos

Usuario ←

Hashtag ←

Fecha y hora ←

More  
@moreira\_enzo

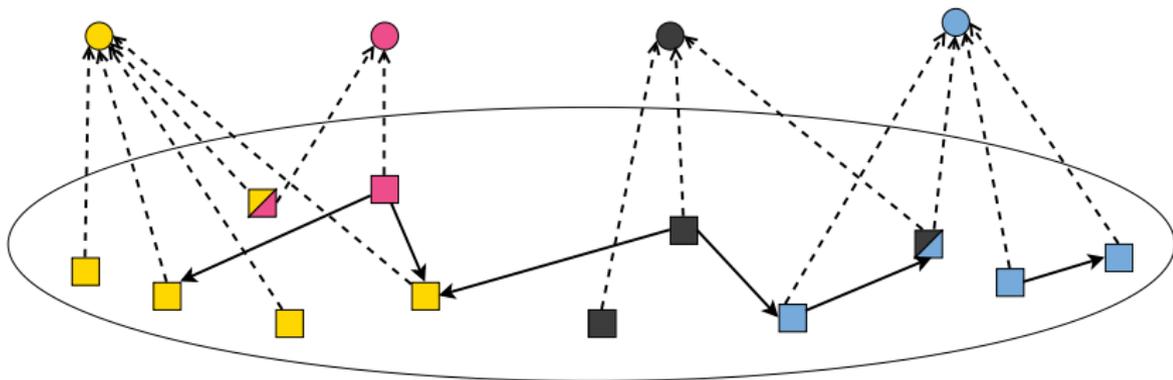
#Elecciones2015 #Argentina :Cerró la votación: final ajustado con posibilidad de #ballotage #Macri vs #Scioli  
<https://t.co/yGPB1mjeXM>

11:37 - 25 de octubre 2015

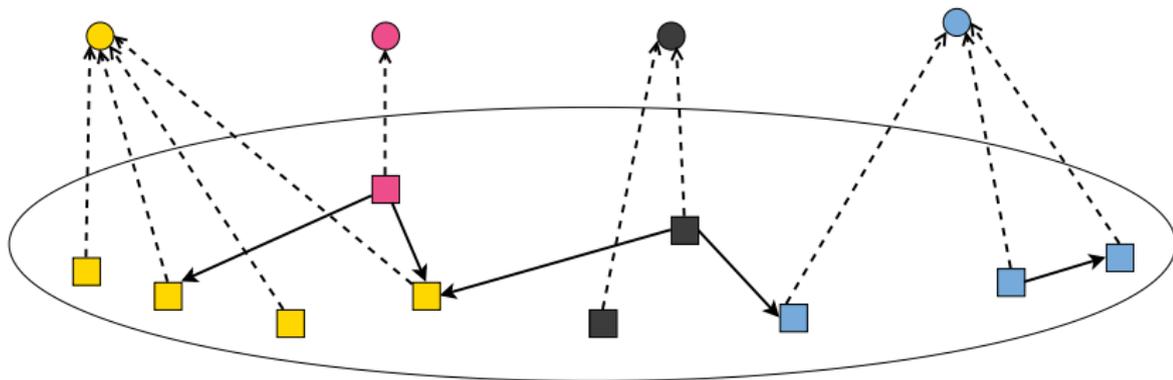
	#Boca	#MM15	#Macri	#DS	#VotaScioli
#Boca	-	7	102	0	2
#MM15	7	-	849	5	4
#Macri	102	849	-	3	6
#DS	0	5	3	-	796
#VotaScioli	2	4	6	796	-

Co-ocurrencia ←

# Detección de tópicos

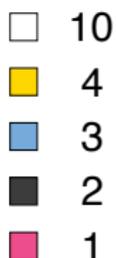
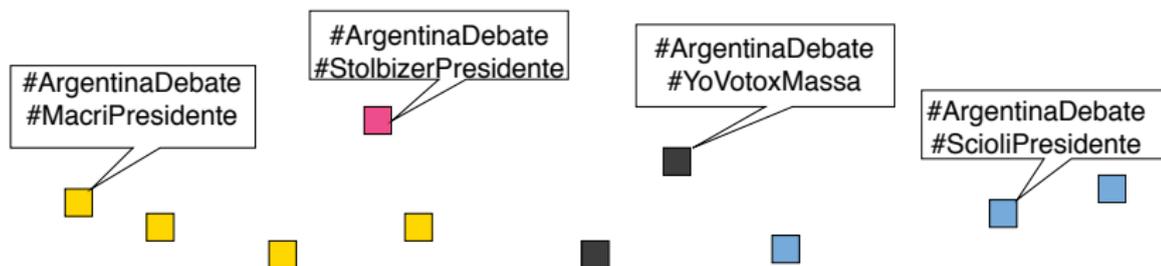


# Detección de tópicos



- 10
- 4
- 3
- 2
- 1

# Detección de tópicos



#ArgentinaDebate

#MacriPresidente

#ScioliPresidente

#YoVotoxMassa

#StolbizerPresidente

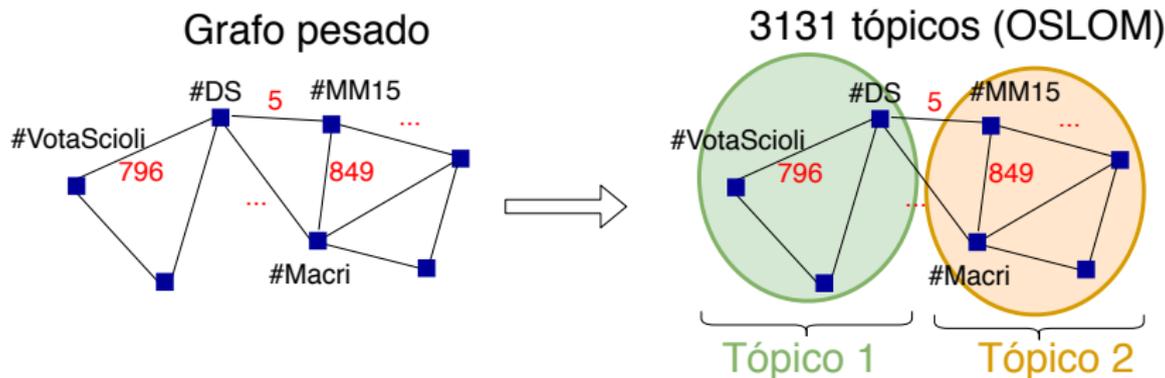
	Yellow	Blue	Black	Pink	White
#ArgentinaDebate	40	30	20	10	100
#MacriPresidente	67	2	23	54	146
#ScioliPresidente	4	60	1	1	66
#YoVotoxMassa	0	0	40	0	40
#StolbizerPresidente	0	0	3	80	83

## Detección de tópicos

	40%	30%	20%	10%	Divergencia Kullback-Leibler
#ArgentinaDebate	40	30	20	10	0.000
#MacriPresidente	67	2	23	54	0.467
#ScioliPresidente	4	60	1	1	0.826
#StolbizerPresidente	0	0	3	80	2.122
#YoVotoxMassa	0	0	40	0	1.609
#FutbolPorSiempre	5	5	3	2	0.012

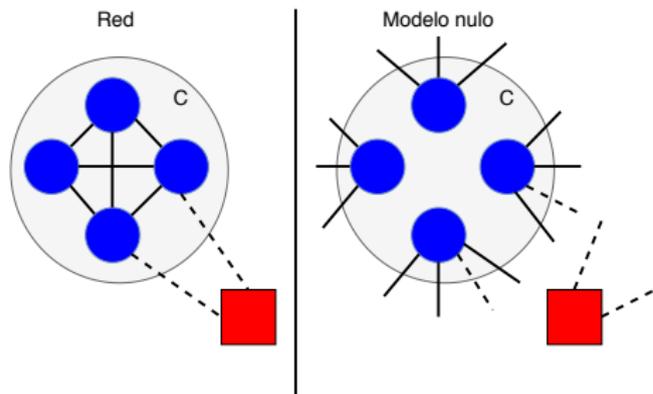
# Detección de tópicos

	#Boca	#MM15	#Macri	#DS	#VotaScioli
#Boca	-	7	102	0	2
#MM15	7	-	849	5	4
#Macri	102	849	-	3	6
#DS	0	5	3	-	796
#VotaScioli	2	4	6	796	-



## OSLOM (Lancichinetti et al., 2011)

- Método de optimización local
- Aplicable a redes de gran escala
- Genera un modelo nulo manteniendo la distribución de grados de los vértices del grafo original
- Mientras menor sea la probabilidad de encontrar la comunidad  $C$  en el modelo nulo, mayor es la probabilidad que  $C$  sea una comunidad en la red









## Interés de los usuarios

	Tópico 1		Tópico 2		Tópico 3	
	#River	#Boca	#DS	#VotaScioli	#MM15	#Macri
	2	5	1	0	9	13
	0	1	10	15	1	0
	10	13	1	3	0	2

## Interés de los usuarios

	Tópico 1		Tópico 2		Tópico 3	
	#River	#Boca	#DS	#VotaScioli	#MM15	#Macri
	2	5	1	0	9	13
	0	1	10	15	1	0
	10	13	1	3	0	2

Sumamos el uso de hashtags dentro del tópico...

## Interés de los usuarios

	Tópico 1		Tópico 2		Tópico 3	
	#River	#Boca	#DS	#VotaScioli	#MM15	#Macri
	7			1		22
	1			25		1
	23			4		2

## Interés de los usuarios

	Tópico 1		Tópico 2		Tópico 3	
	#River	#Boca	#DS	#VotaScioli	#MM15	#Macri
	7			1		22
	1			25		1
	23			4		2

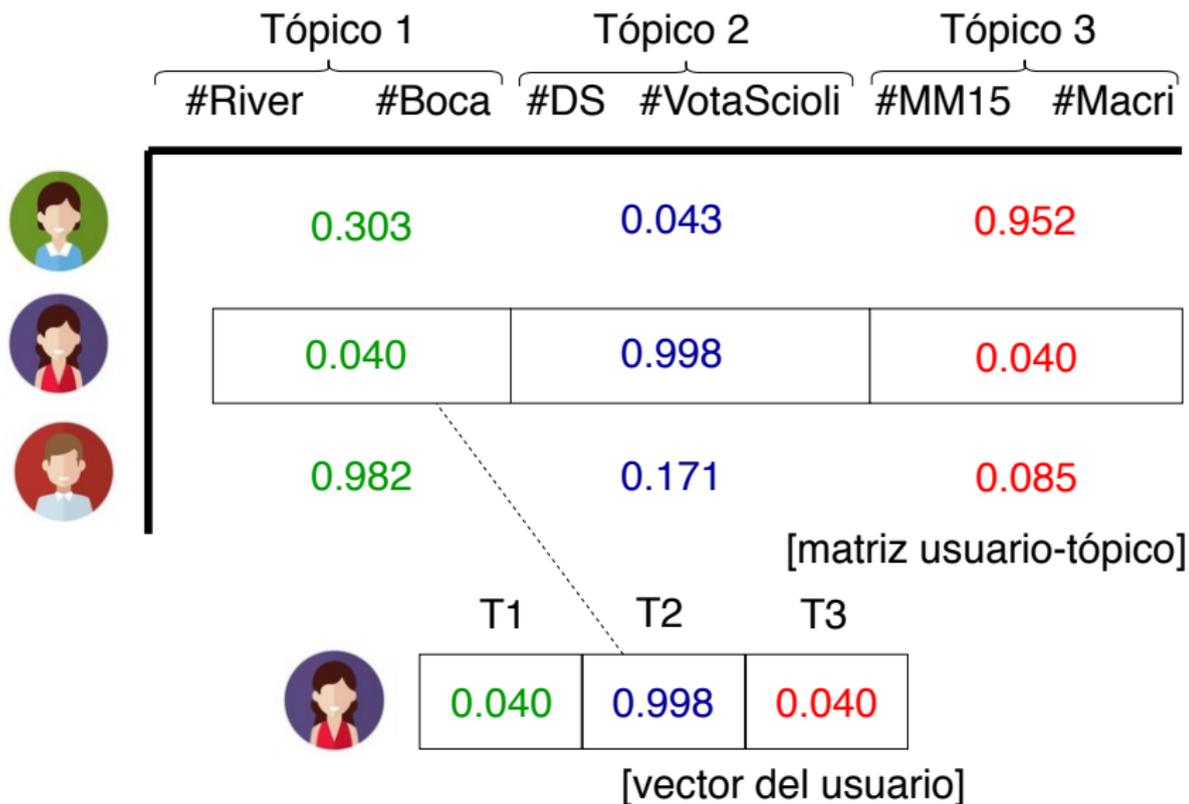
Aplicamos TF-IDF y normalizamos vectores ( $L^2$ )...

## Interés de los usuarios

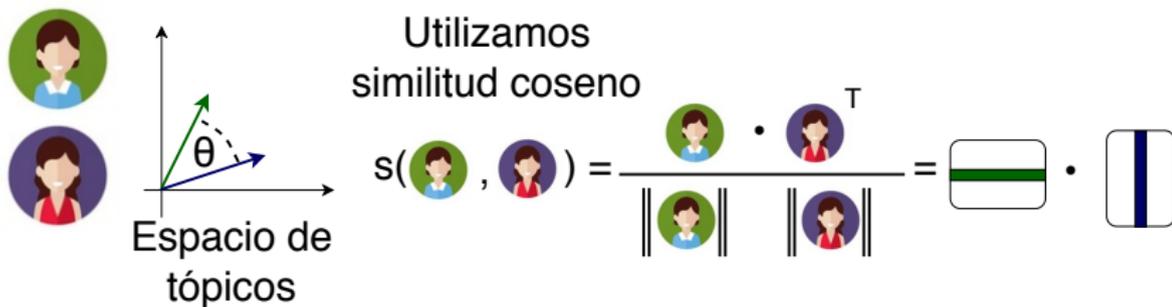
	Tópico 1		Tópico 2		Tópico 3	
	#River	#Boca	#DS	#VotaScioli	#MM15	#Macri
	0.303			0.043		0.952
	0.040			0.998		0.040
	0.982			0.171		0.085

[matriz usuario-tópico]

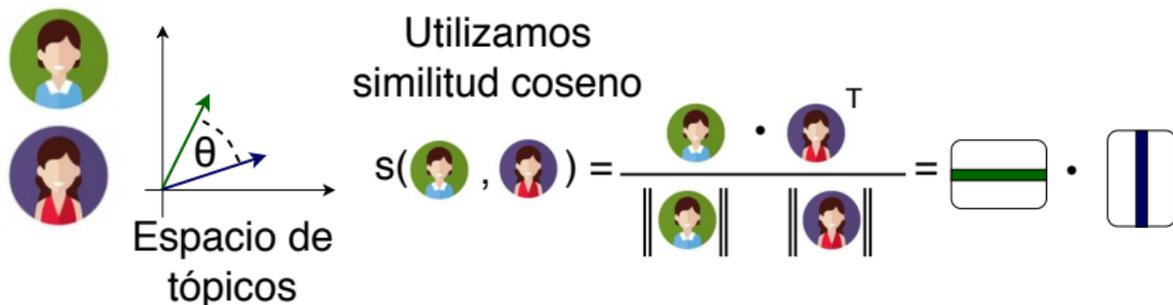
## Interés de los usuarios



## Computando similitud entre usuarios



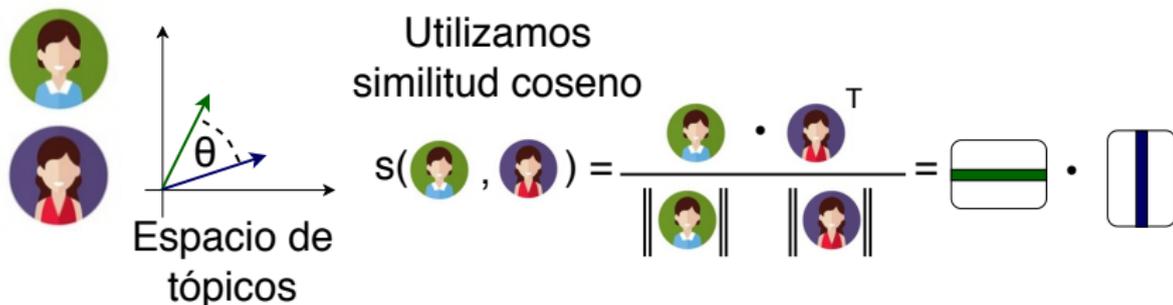
## Computando similitud entre usuarios



Y entre grupos de usuarios...

$$s(G_1, G_2) =$$

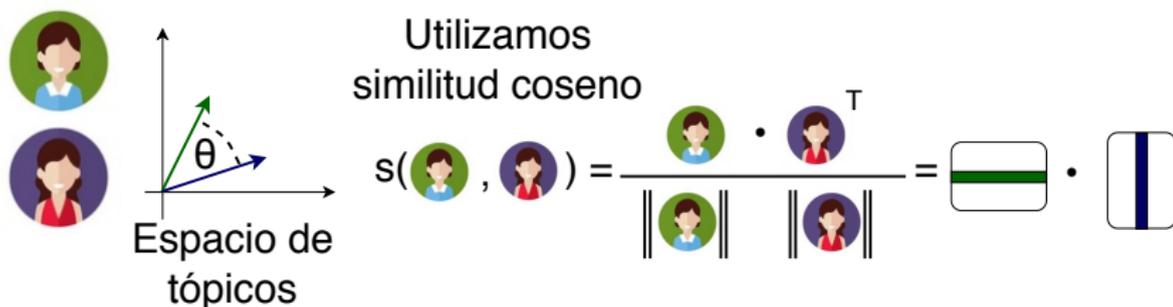
## Computando similitud entre usuarios



Y entre grupos de usuarios...

$$s(G_1, G_2) = \text{green\_bar} \cdot \text{blue\_bar}$$

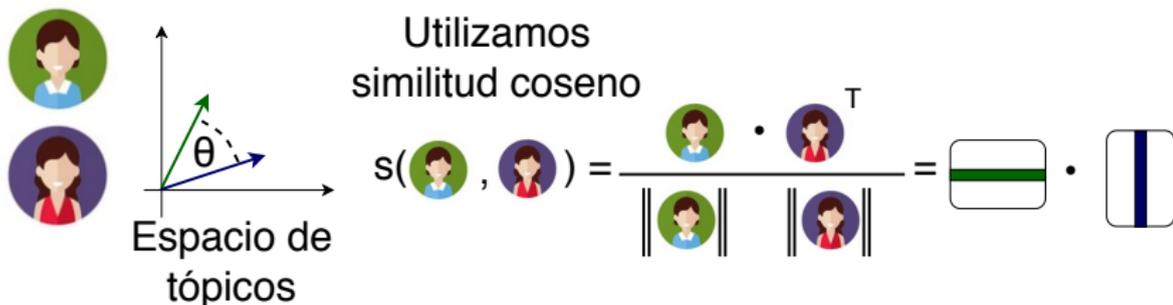
## Computando similitud entre usuarios



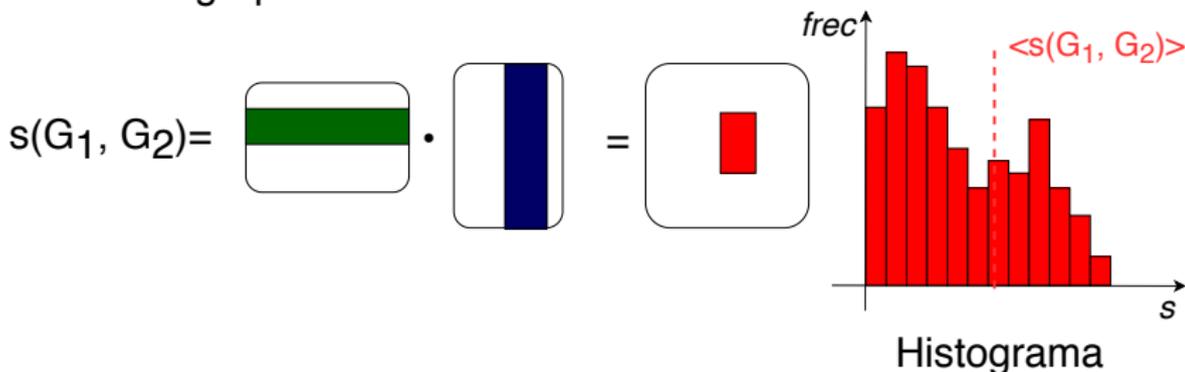
Y entre grupos de usuarios...

$$s(G_1, G_2) = \text{Green Bar} \cdot \text{Blue Bar} = \text{Red Bar}$$

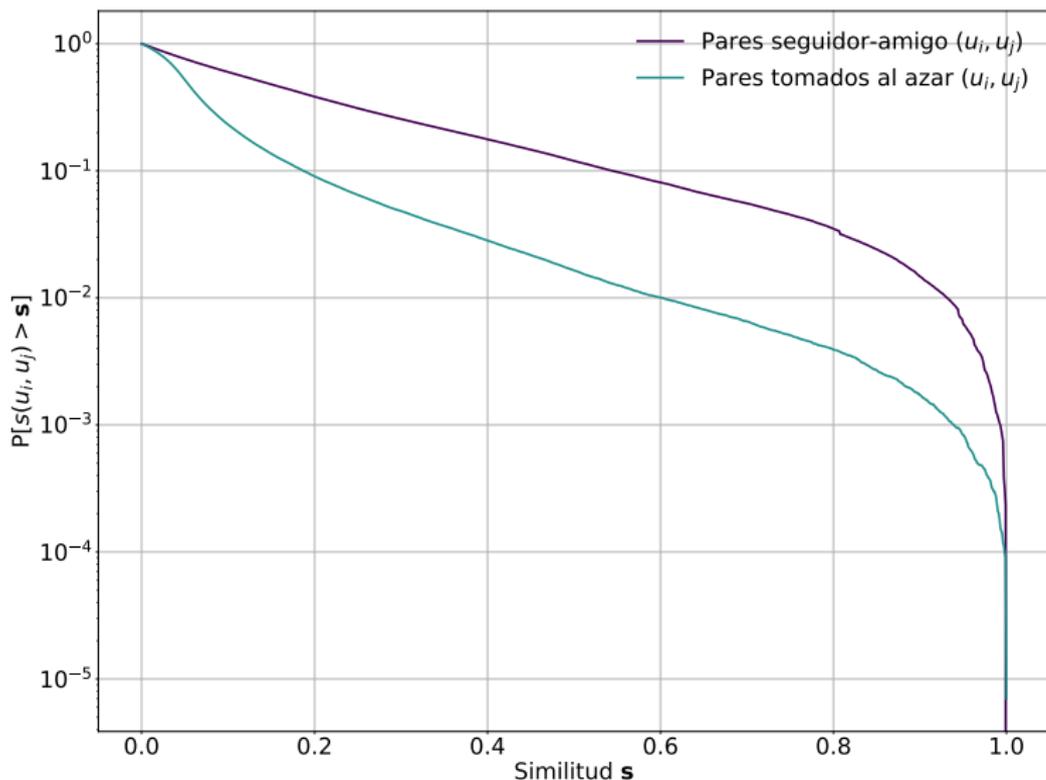
## Computando similitud entre usuarios



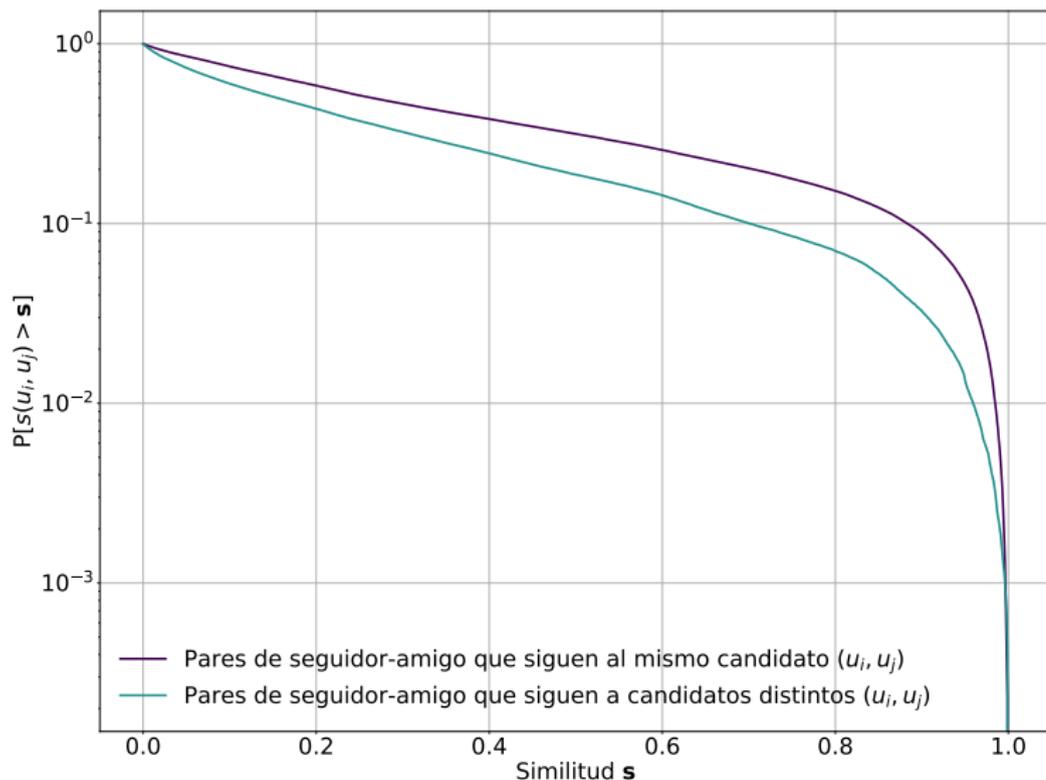
Y entre grupos de usuarios...



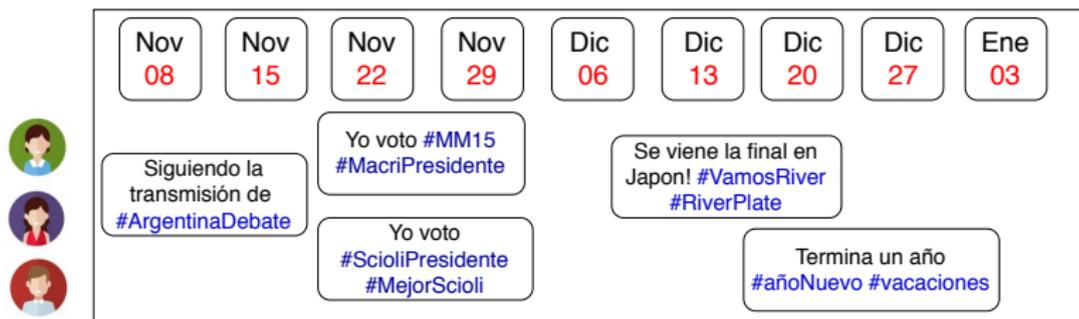
# Homofilia en la red de Twitter



## Homofilia entre seguidores de candidatos



## Evolución temporal de los distintos grupos políticos



## Evolución temporal de los distintos grupos políticos



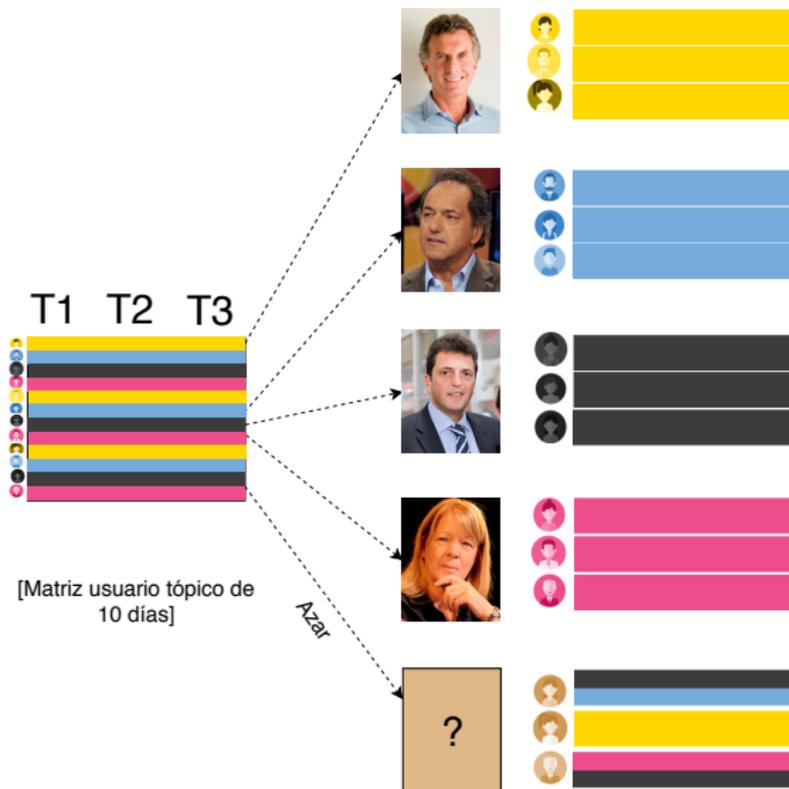


## Evolución temporal de los distintos grupos políticos

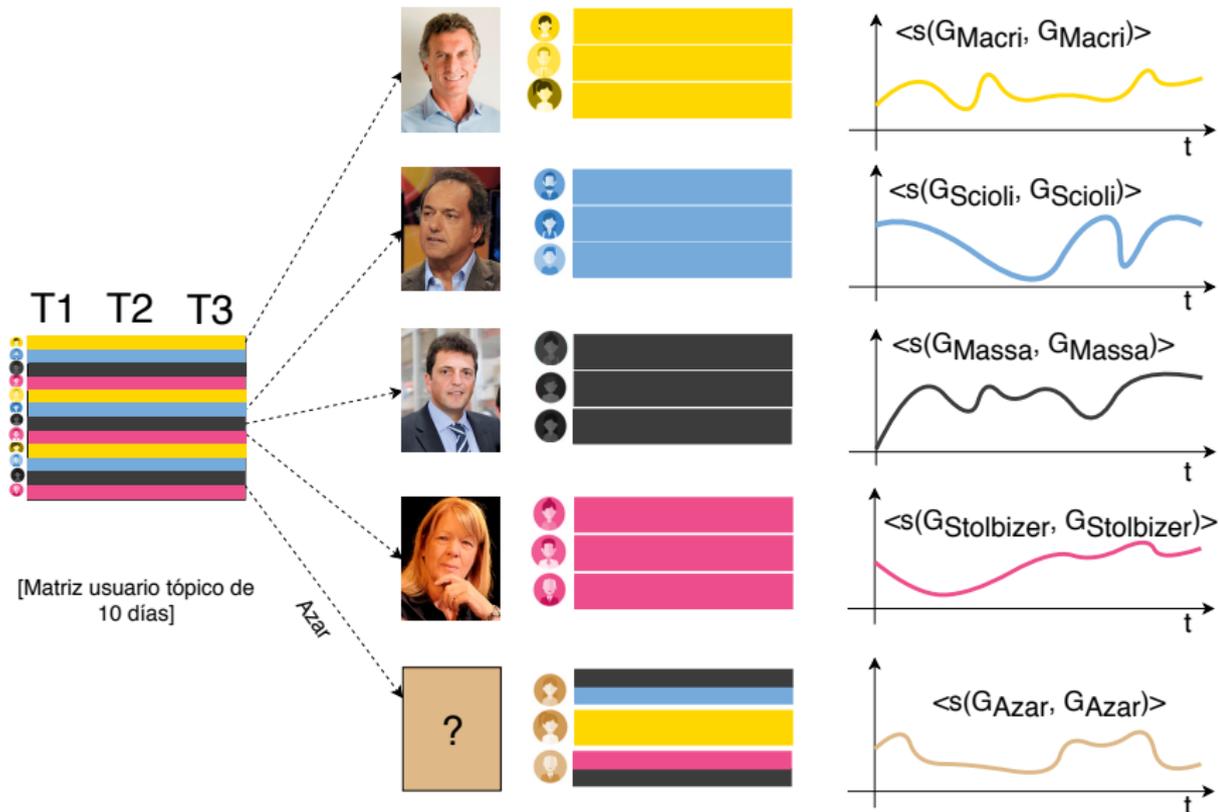


[Matriz usuario tópico de  
10 días]

# Evolución temporal de los distintos grupos políticos

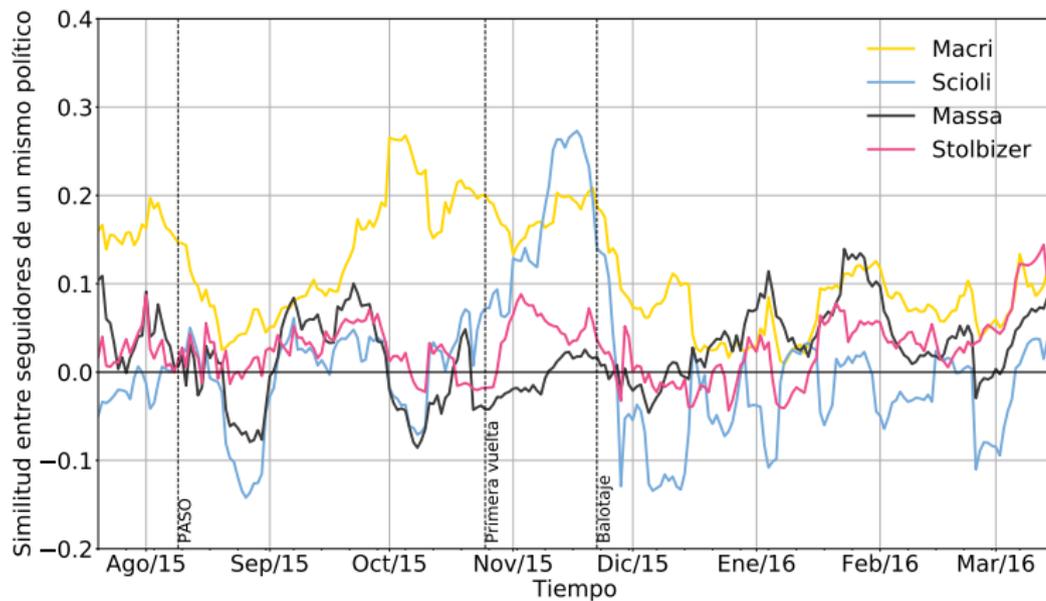


# Evolución temporal de los distintos grupos políticos

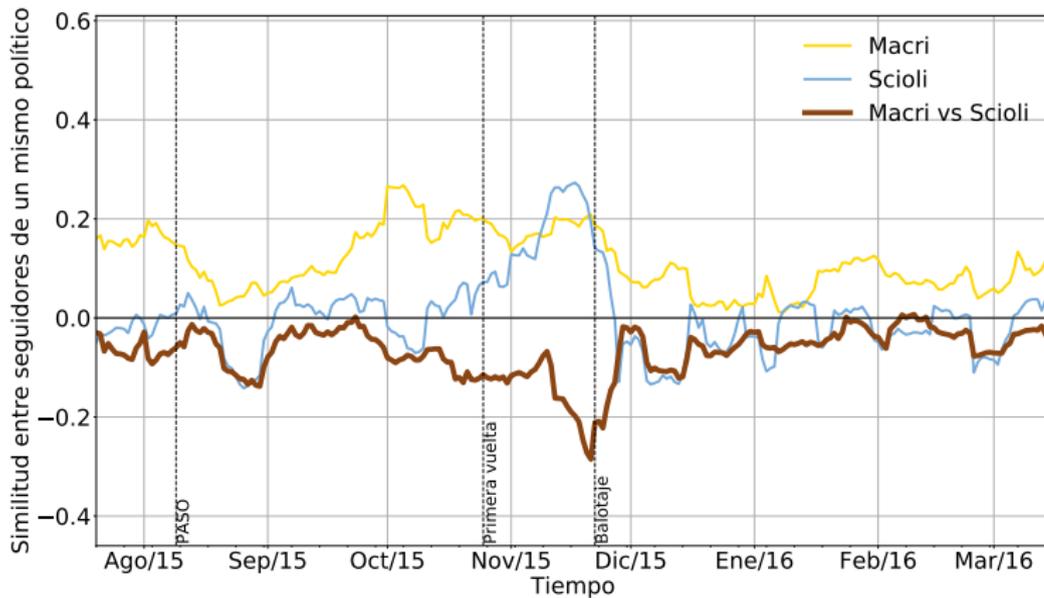


## Resultados

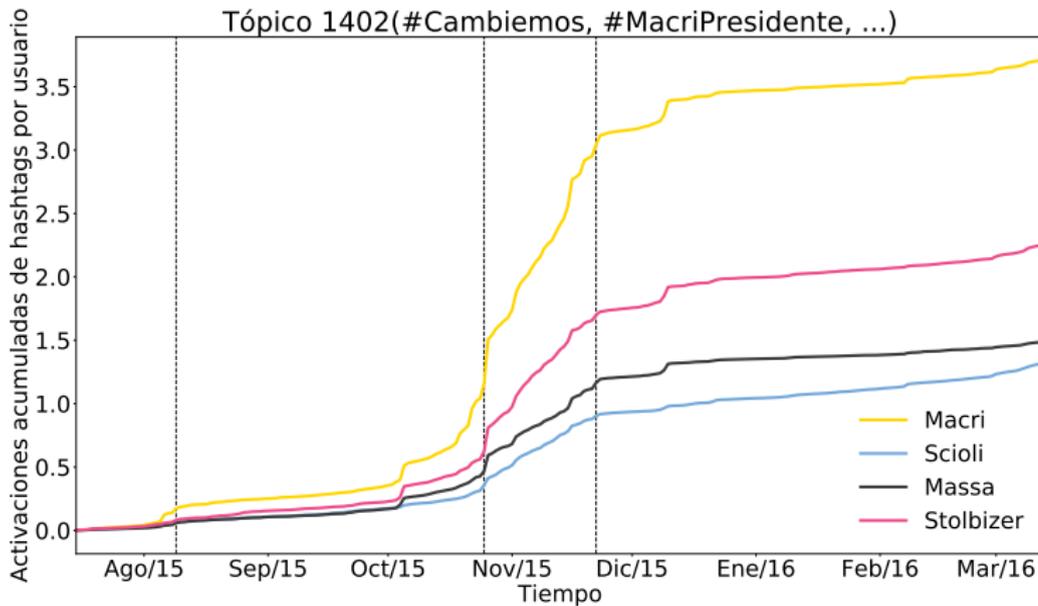
## Evolución de la similitud de cada grupo



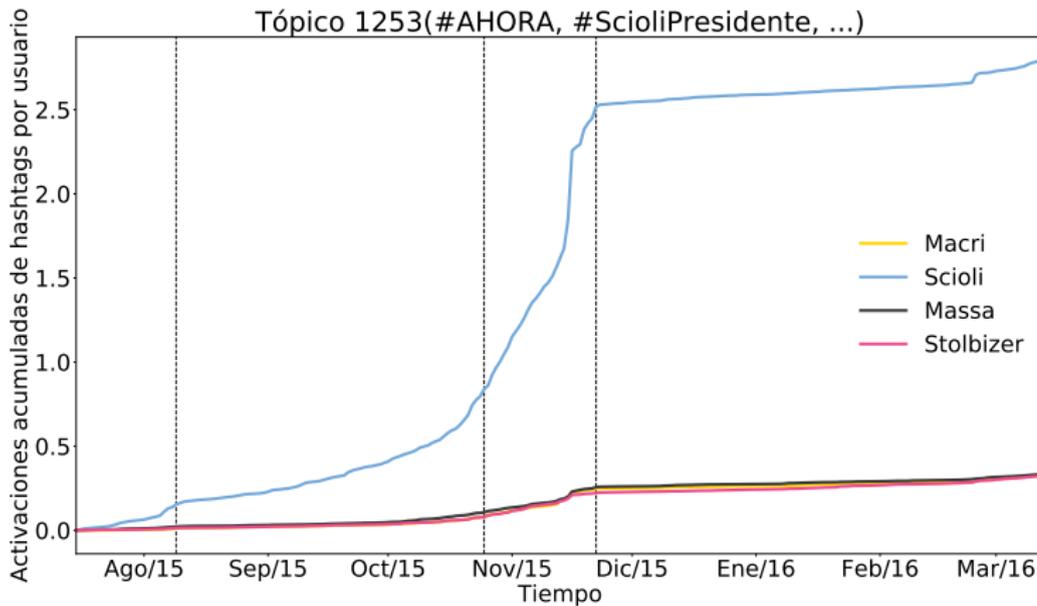
## Evolución de la similitud entre grupos



## Evolución del uso del tópic #1402 (Macri)



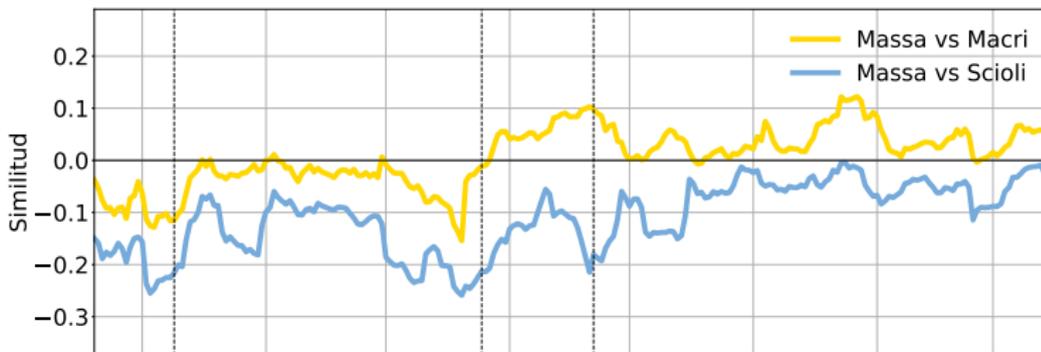
## Evolución del uso del tópico #1253 (Scioli)



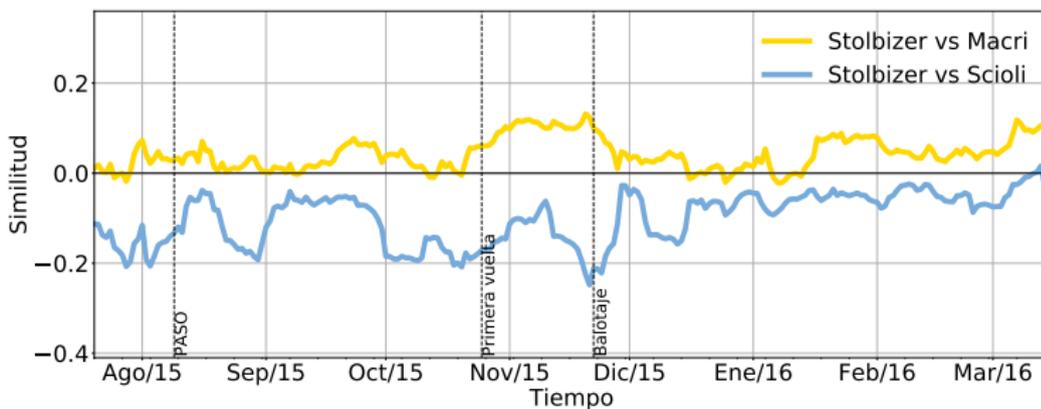
## Evolución de la similitud de los candidatos alternativos



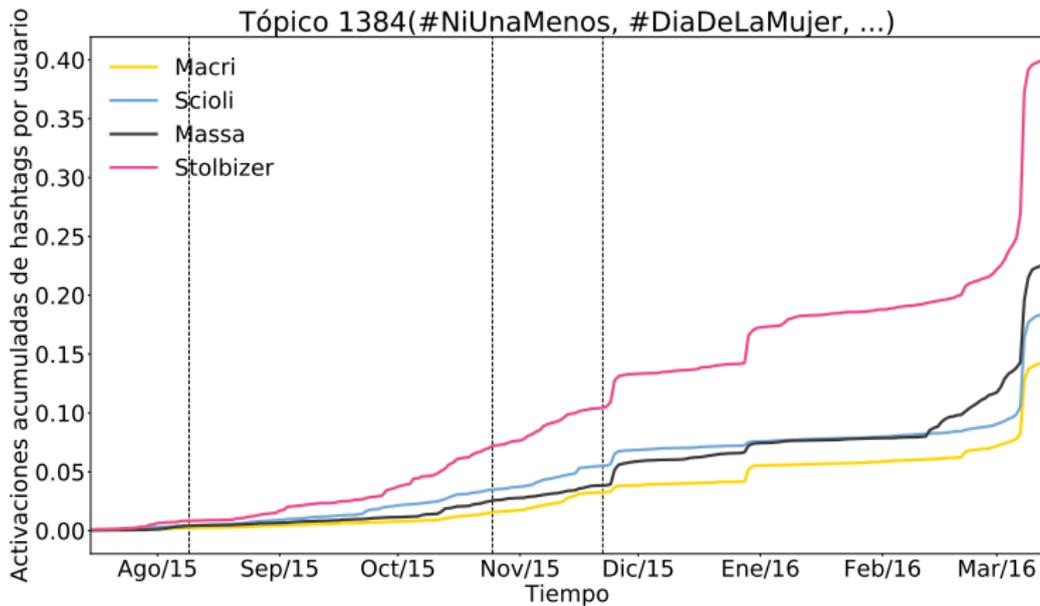
21%



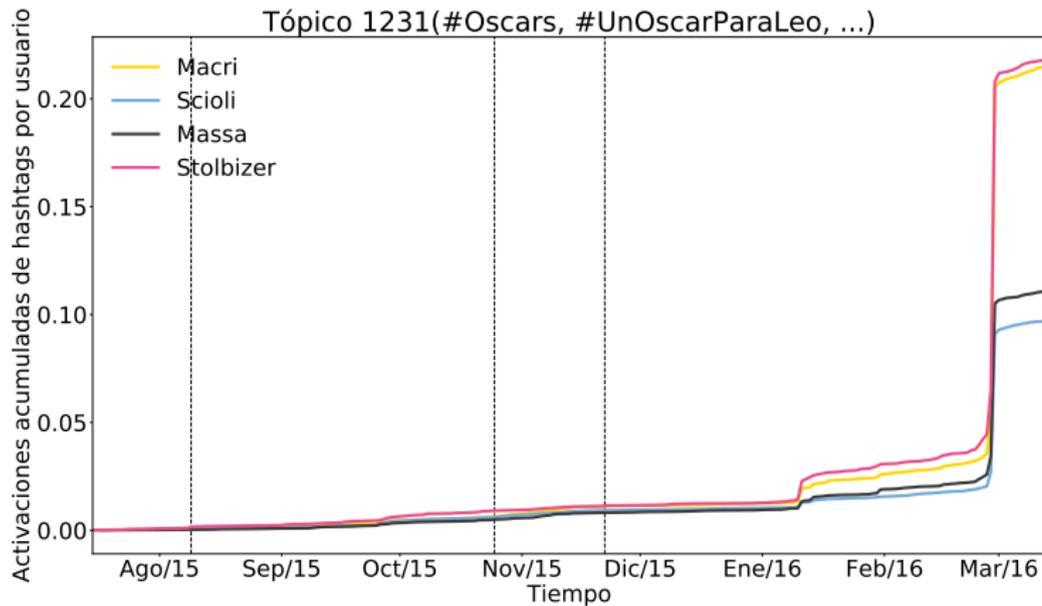
3%



## Evolución del uso del tópico #1384 (Feminismo)



## Evolución del uso del tópico #1231 (Oscars)



## Predicción de afiliación política

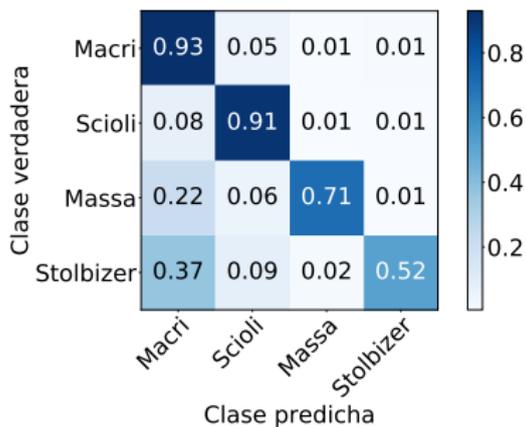
- Verificar la correctitud de los tópicos y determinar si representan la orientación política de los usuarios
- Usuarios que *retweetearon* contenido de sólo un candidato
- Clasificador *Random Forest*
- *Cross validation* utilizando *Stratified K-fold*
- *Grid search* probando combinaciones de seis parámetros
- Métrica de evaluación : Exactitud ponderada

$$\hat{w}_i = \frac{w_i}{\sum_j \mathbb{1}(y_j = y_i) w_j}$$

$$ba(y_i, \hat{y}_i, w) = \frac{1}{\sum \hat{w}_i} \sum_i \mathbb{1}(\hat{y}_i = y_i) \hat{w}_i$$

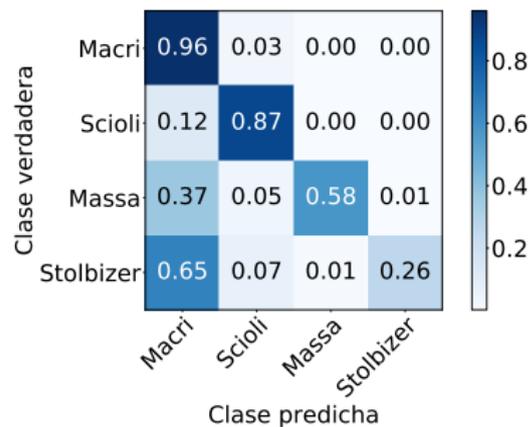
# Predicción de afiliación política

## Hasta primera vuelta



Exactitud ponderada : 77%

## Todo el período



Exactitud ponderada : 67%

## Conclusiones

## Conclusiones

- Detectamos **tópicos de debate** como comunidades dentro de una red pesada de *hashtags*.
- Descubrimos que estos tópicos de debate son **representativos de orientación política**.
- Verificamos que existe **homofilia temática** entre los usuarios de la red.
- Pudimos explicar los **resultados de las elecciones** en base al cambio de orientación del voto de los candidatos alternativos.
- A partir de un modelo que utiliza los tópicos para predecir la orientación política, **verificamos que muchos usuarios cambiaron su orientación entre la primera vuelta y el ballottage** y que esto mismo se manifestó en el contenido publicado en Twitter.

Gracias !

## Bibliografía



Colleoni, E., Rozza, A., Arvidsson, A. (2014).

Echo chamber or public sphere ? Predicting political orientation and measuring political homophily in Twitter using big data.

*Journal of communication*, 64(2), 317-332.



Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A. (2011).

Political polarization on twitter.

*In Fifth international AAAI conference on weblogs and social media.*



Romero, D. M., Meeder, B., Kleinberg, J. (2011).

Differences in the mechanics of information diffusion across topics : idioms, political hashtags, and complex contagion on twitter.

*In Proceedings of the 20th international conference on World wide web (pp. 695-704). ACM.*



Watts, D. J., Strogatz, S. H. (1998).

Collective dynamics of 'small-world' networks.

*Nature*, 393(6684), 440.



Lancichinetti, A., Radicchi, F., Ramasco, J. J., Fortunato, S. (2011).

Finding statistically significant communities in networks.

*PloS one*, 6(4), e18961.