

Detección de comunidades políticas en Twitter y posibles aplicaciones

Palabras clave: Comunidades políticas; Twitter; igraph; rtweet; Social Network Analysis

Introducción

Los estudios académicos sobre el uso político de redes sociales han proliferado en tiempos recientes. Pese a las obvias bondades de trabajar con datos de redes sociales (principalmente, estudiar comportamiento real, en tiempo real), existe una importante limitación en el acceso a información relevante a nivel individual. Por ejemplo, en el caso de Twitter, no es posible obtener información relevante a nivel individual sobre los usuarios que participan en una discusión política dada. Sin embargo, existen ciertas estrategias para inferir actitudes políticas de los usuarios de Twitter. Una de ellas es la detección de comunidades políticas mediante la aplicación de técnicas de teoría de redes, en particular, la detección de comunidades. En teoría de redes se utiliza la detección de comunidades para identificar subgrafos cuyos nodos están altamente conectados entre sí pero con poco conectados con otros grupos de nodos. La idea subyacente es identificar subgrupos cohesivos (Borgatti, 2013).

Detección de comunidades en Twitter

El paquete *rtweet v.0.7.0* (Kearney, 2020) permite una sencilla interacción con la API oficial de Twitter para la descarga de datos. En orden de ilustrar cómo detectar comunidades políticas en Twitter utilizaremos datos del debate presidencial uruguayo de noviembre de 2019. Para descargar la discusión política referida al debate descargamos aquellos *tweets* que contenían los *hashtags* *#DebatePresidencialUY* o *#DebatePresidencial2019* usando en la noche del debate el siguiente comando del paquete *rtweet*:

```
search_tweets("#debatepresidencialUY OR #DebatePresidencial2019", n = 100000,  
retryonratelimit=TRUE)
```

Esto nos devuelve un *dataframe* con todos los *tweets* que utilizaron al menos uno de los dos *hashtags*. Sin embargo, si lo que queremos es comparar patrones de uso de Twitter de los partidarios del candidato Luis Lacalle Pou *vs* los partidarios de Daniel Martínez, a priori, necesitamos algún criterio de caracterización de los usuarios. Aquí es donde radica la utilidad de la detección de comunidades. A no ser de contar con datos de comportamiento previo de los usuarios, tales como índices de afiliación partidaria basados en el historial de *retweets* de los usuarios como en (Bogliaccini et al., 2019), nos resultaría muy dificultoso identificar cuáles usuarios pertenecen a cuál bando de la discusión. De esta forma, la detección de comunidades nos permite un atajo rápido y efectivo -aunque aproximado- para lograr ubicar a los usuarios en función de sus preferencias políticas.

El primer paso para identificar comunidades es transformar el *dataframe* que descargamos en una red. En esta red todos los usuarios participantes son nodos y están unidos -de forma dirigida- si uno de ellos retuiteó a otro. Luego, nos valemos del paquete *igraph v.1.2.4.1* (Csárdi, 2019) para crear la red y extraer el componente conectado más grande (*Largest Connected Compound*) y remover los nodos que no están conectados a ningún otro nodo. Una vez extraído el componente conectado más grande buscamos comunidades utilizando un modelo spin-glass con un tope de dos comunidades (la decisión de la cantidad de comunidades depende de la estructura empírica de los datos y de la teoría sobre la división de la discusión política en cuestión) de la siguiente forma:

```
net %>% cluster_spinglass(spins = 2)
```

La Figura 1 grafica la red de retweets relacionados al debate a la luz de las comunidades identificadas (cada una pintada con colores distintos). La red parece estructurarse en dos comunidades bien definidas: a la izquierda de la figura vemos un subgrafo muy denso entre sí pero poco conectado con el otro subgrafo. Un análisis de las autoridades cada comunidad (usuarios que recibieron mayor cantidad de *retweets*) revela que estas están claramente identificadas con los dos bloques políticos que representaba cada candidato en el debate.

Una vez identificadas las comunidades y su asociación política el paquete `igraph` ofrece una función que permite transformar la red -junto a sus atributos de los nodos como la comunidad a la que pertenecen- a un *dataframe* de nuevo:

```
as_long_data_frame(net)
```

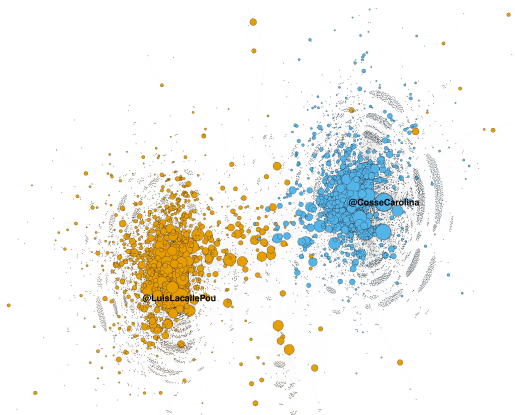


Figura 1: Red de retweets del debate con dos comunidades detectadas

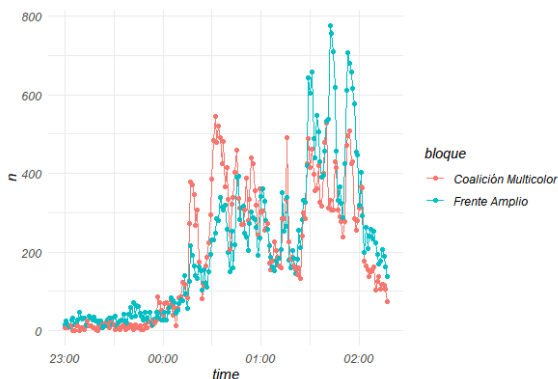


Figura 2: Retweets por minuto de los usuarios identificados con cada comunidad

Posibles aplicaciones

El procedimiento descrito anteriormente produce un *dataframe* que contiene los tweets descargados junto a una aproximación binaria de la afiliación partidaria del autor del *tweet*. Esto permite estudiar la diferencia en la discusión política entre los usuarios próximos a cada partido. Por ejemplo, una de las posibles aplicaciones es realizar un análisis de sentimiento de los *tweets* de los distintos grupos partidarios para identificar el tono de los mensajes de ambos bloques o estudiar la estructura (densidad, distribución de medidas de centralidad) dentro de cada subgrafo.

Otra posible aplicación de la detección de comunidades políticas en Twitter es medir la frecuencia con la que los usuarios de distintos partidos participan durante una discusión dada. Existe evidencia empírica de que el entusiasmo de las personas hacia un evento político aumenta las acciones positivas mientras que el enojo genera que las personas eviten el tema (MacKuen et al., 2010). En base a esto, Calvo et al. (2019) proponen un modelo de adjudicación de eventos utilizando como variable dependiente el tiempo para retuitear (*time-to-retweet*). De forma similar, es posible aproximarse al nivel de entusiasmo que un evento genera en los grupos partidarios según la evolución de retweets por minuto que envía cada grupo, como muestra la Figura 2.

Bibliografía

- Bogliaccini, Juan, et al. *Twitterquía. La política de las redes en Uruguay*. (2019). *Tunel Ediciones y Universidad Católica del Uruguay*. Montevideo, Uruguay
- Borgatti, Stephen P., Martin G. Everett, and Jeffrey C. Johnson. *Analyzing social networks*. Sage, 2018.
- Calvo, Ernesto, et al. "Winning! Adjudication and Dialogue in Social Media." (2019).
- Csardi G, Nepusz T (2006). "The igraph software package for complex network research." *InterJournal, Complex Systems*, 1695. <http://igraph.org>.
- Kearney MW (2019). "rtweet: Collecting and analyzing Twitter data." *Journal of Open Source Software*, 4(42), 1829. doi: 10.21105/joss.01829, R package version 0.7.0, <https://joss.theoj.org/papers/10.21105/joss.01829>.
- MacKuen, Michael, et al. Civic engagements: Resolute partisanship or reflective deliberation." *American Journal of Political Science* 54.2 (2010): 440-458.