



PROGRAMA UNIVERSITARIO
DE ESTUDIOS SOBRE
DEMOCRACIA, JUSTICIA Y SOCIEDAD



DOCUMENTO DE TRABAJO 8

Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad | ENERO 2022

Estrategias metodológicas

Comunidades en redes socio-digitales

Identificación de narrativas y cuentas representativas

Martín Zumaya
Diego Espitia
Luis Ángel Escobar
Autores

ENERO 2022



Documento de trabajo 8

Comunidades en redes socio-digitales. Identificación de narrativas y cuentas representativas

El presente documento de trabajo fue elaborado por investigadores adscritos al Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad (PUEJJS) de la Universidad Nacional Autónoma de México

Primera edición, enero 2022

D.R. © Universidad Nacional Autónoma de México,
Programa Universitario de Estudios Sobre Democracia,
Justicia y Sociedad, Torre UNAM-Tlatelolco, Piso 13
Ricardo Flores Magón número 1,
Colonia Nonoalco Tlatelolco Alcaldía Cuauhtémoc,
Código Postal 06995, Ciudad de México
www.puedjs.unam.mx

Cómo citar:

Zumaya Martín, Diego Espitia, Luis Ángel Escobar (2022), "Comunidades en redes socio-digitales. Identificación de narrativas y cuentas representativas", Documento de Trabajo núm. 8. PUEJJS, UNAM, México, 30 páginas.



CONACYT

Consejo Nacional de Ciencia y Tecnología

Este documento de trabajo se realizó en el marco del Proyecto La disputa por la cultura política en el México actual: democracia, redes digitales y movimientos sociales", apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) en el año 2021 dentro del marco de los Programas Nacionales Estratégicos (PRONACE), el contenido y opiniones son responsabilidad exclusiva de los autores.

| **contacto:** puedjs@humanidades.unam.mx



ÍNDICE

Presentación	2
1. Introducción	3
1.1. Las cuentas: componentes elementales de las plataformas	3
1.2. Posibles acciones e interacciones entre cuentas	4
1.3. Redes de interacción entre cuentas	4
1.4. Caracterización de nodos dentro de la red	7
1.5. Estructura de comunidades. El algoritmo de Girvan-Newman	8
2. Metodología	10
2.1. Eventos y narrativas como objeto de estudio	10
2.2. Roles y relevancia de cuentas en eventos focales	10
2.3. Audiencia, Alcance directo e indirecto de cuentas	12
2.4. Caracterización del contenido de las publicaciones	13
Modelado de temas	13
Búsqueda Inversa	15
3. Aplicación de la metodología	18
3.1. Detección y caracterización de comunidades	18
3.2. Audiencia y Alcance de las cuentas	21
Modelado de temas y búsqueda inversa	24
4. Conclusiones	26



Presentación

Las herramientas digitales y la Internet se han convertido en una parte fundamental de nuestro día a día, permitiéndonos comunicarnos, trabajar, entretenernos y recibir, gracias a los dispositivos móviles, oleadas de información directamente en la palma de nuestras manos.

De las herramientas digitales, la que quizá ha cobrado mayor relevancia en nuestra vida cotidiana, dada su omnipresencia, son las redes socio-digitales, las cuales, por un lado, permiten extender la interacción social del espacio físico al virtual, haciendo posible compartir intereses sin la necesidad de compartir el mismo espacio físico o temporal. Por otro lado, las redes socio-digitales también se han convertido en un espacio en el cual las personas se informan, expresan sus opiniones y debaten sobre temas de interés general.

Estas y otras razones han hecho que las redes socio-digitales se conviertan en las nuevas plazas públicas en las que se genera información y se desarrolla una gran parte de la conversación de los temas actuales y de relevancia para las sociedades, siendo una de sus facetas la de servir como un medio de comunicación.

Sin embargo, a diferencia de los medios de comunicación tradicionales (radio, prensa escrita, televisión), la generación y difusión del contenido informativo o de entretenimiento no se realiza de manera unilateral en el que las personas usuarias juegan el rol pasivo de espectadores, sino que tienen un papel activo en la generación y difusión de contenidos a través de la interacción y acción colectiva entre los millones de usuarios de estas plataformas. Dichos usuarios, pueden ser personas comunes, periodistas, las versiones digitales de medios de comunicación tradicionales, actores políticos, líderes de opinión, cuentas automatizadas, entre otros.

La actividad colectiva e interacciones de todas estas cuentas durante eventos focales alrededor de temas relevantes, como por ejemplo, movilizaciones sociales, coyunturas sociales o políticas, etc, pueden dar lugar a distintas narrativas que son representaciones de las distintas interpretaciones, lecturas e intereses de los contenidos que las cuentas emiten y amplifican.

Estudiar la compleja dinámica de las cuentas en las redes socio-digitales, es importante tanto para identificar distintas narrativas alrededor eventos focales, como también para la determinar la existencia de mecanismos de manipulación que pueden sesgar contenidos o propagar información falsa, afectando a las narrativas que ocurren en éstas nuevas ágoras digitales.

En este documento presentamos un marco metodológico empírico, que hemos desarrollado desde Tlatelolco LAB, para la identificación y caracterización de comunidades, actores principales y narrativas en la plataforma Twitter, y que será aplicado a los distintos análisis de coyuntura y estudios de caso realizados por el laboratorio.

▶ 1. Introducción

En esta sección describimos las distintas etapas del marco metodológico empírico para el análisis de datos y la identificación de comunidades, narrativas y actores relevantes alrededor de coyunturas en redes sociales digitales. Aunque en principio sería posible desarrollar estrategias similares para distintas redes sociales digitales, en el resto del documento nos enfocaremos en la plataforma Twitter, debido principalmente a dos factores. El primero es la facilidad de la obtención de datos en comparación con otras plataformas digitales.¹

El segundo es debido a la relevancia que tiene Twitter. En lo referente al estudio de las diferentes redes socio digitales, Twitter se ha posicionado como la plataforma más importante para la comunicación política en la mayoría de los debates democráticos en todo el mundo [7], y es usada masivamente por distintos movimientos ciudadanos para propagar sus opiniones y movilizar a la ciudadanía [9].

▶ 1.1. Las cuentas: componentes elementales de las plataformas

Tanto en Twitter como en otras plataformas digitales, los componentes elementales son las cuentas o perfiles, los cuales pueden ser considerados como representaciones digitales, en primer lugar de individuos, y en un sentido más amplio, de entidades como: marcas, empresas, productos y medios de comunicación.

En el caso particular de la plataforma Twitter, a cada cuenta se le asigna un número de identificación único, proporcionado por la plataforma misma, y cada cuenta puede elegir un nombre de usuario, el cual puede cambiar en cualquier momento, y un handle, el cual debe de ser único y no puede modificarse una vez que ha sido creada la cuenta².

Las cuentas tienen la posibilidad de generar tweets o publicaciones, en las que pueden incluir contenido como: texto, audio, imágenes o vídeos. En el texto de un tweet, una cuenta puede hacer mención de otra utilizando la sintaxis **@handle**, o utilizar hashtags, con la sintaxis **#hashtag** los cuales pueden representar, en un sentido amplio, sentimientos y temas relacionados con el contenido de la publicación.

¹La plataforma Twitter cuenta con una API, i.e. una interfaz de solicitud y obtención de información, por medio de la cual es posible, entre otras funciones, recuperar publicaciones realizadas alrededor de hashtags o palabras clave especificadas por el usuario. El acceso a esta API se obtiene mediante un proceso de solicitud y aprobación por parte de la plataforma, para mayor información recomendamos visitar la página Twitter Developers

²Específicamente, el handle es el que aparece al final del URL al perfil del usuario: <http://twitter.com/usuario>, mientras que el nombre de usuario es el que se muestra al resto de los usuarios de la plataforma cuando realiza publicaciones.

→1.2. Posibles acciones e interacciones entre cuentas

Cada usuario o cuenta en la plataforma puede elegir a que otras cuentas seguir, lo que significa que las publicaciones realizadas por las cuentas a las que se sigue, se muestran en la línea de tiempo de cada una de ellas. De tal forma que cada cuenta tiene la posibilidad de seleccionar el contenido que recibe de acuerdo a sus intereses y afinidades. Cabe mencionar que las publicaciones generadas por una cuenta llegan directamente a las líneas de tiempo de sus seguidores.

De manera adicional, al seguir otras cuentas, se pueden realizar diversas interacciones entre las personas usuarias; específicamente:

- **Retweet o Reposteo:** La publicación a la que se realiza el retweet o reposteo se muestra de manera íntegra a los seguidores de la cuenta que genera la acción. El propósito de esta acción puede considerarse como de amplificación o difusión del tweet original.
- **Quote o Cita:** Esta acción es similar al Retweet o al reposteo, en el sentido en que tiene la función de amplificar o difundir el tweet original, con la diferencia de que la cuenta que genera la acción incluye un comentario a la publicación original.
- **Reply o Respuesta:** Es una acción en respuesta a la publicación de otra cuenta, que no necesariamente se muestra a los seguidores de la cuenta que generó la acción.
- **Like:** Es una acción que puede representar el respaldo o aprobación de la cuenta que inicia la acción del contenido en la publicación original. No representa una amplificación o difusión del contenido original en la misma forma que el retweet o el quote, ya que no necesariamente llega al total de los seguidores de la cuenta que inicia la acción.
- **Menciones:** Se refiere a la acción en la que una cuenta hace mención de alguna otra mediante la sintaxis: @usuario, en el texto de su publicación. La intención de esta acción puede ser para llamar la atención de la cuenta a la que se hace mención, ya que esta última recibe una notificación, o expresar que el texto o el contenido dentro de la publicación está dirigido hacia la cuenta que se menciona.

De manera general, estas son los distintos tipos de acciones e interacciones que una cuenta puede generar dentro de la plataforma, y son las que tomaremos como punto de partida para la metodología y análisis que desarrollamos a continuación.

→1.3. Redes de interacción entre cuentas

De manera general podemos definir a una red ³ como un conjunto de elementos, que llamaremos nodos o vértices, con conexiones o relaciones entre ellos, a los que denominaremos enlaces o aristas

³Otra nomenclatura es la de grafo, la cual proviene de la literatura de las matemáticas, específicamente de la teoría de grafos, mientras que el término red es más común en el área de las ciencias de la complejidad. Aunque esencialmente estos términos son equivalentes, en este documento preferiremos utilizaremos el término red para referirnos a esta forma de representar las relaciones e interacciones entre nodos.

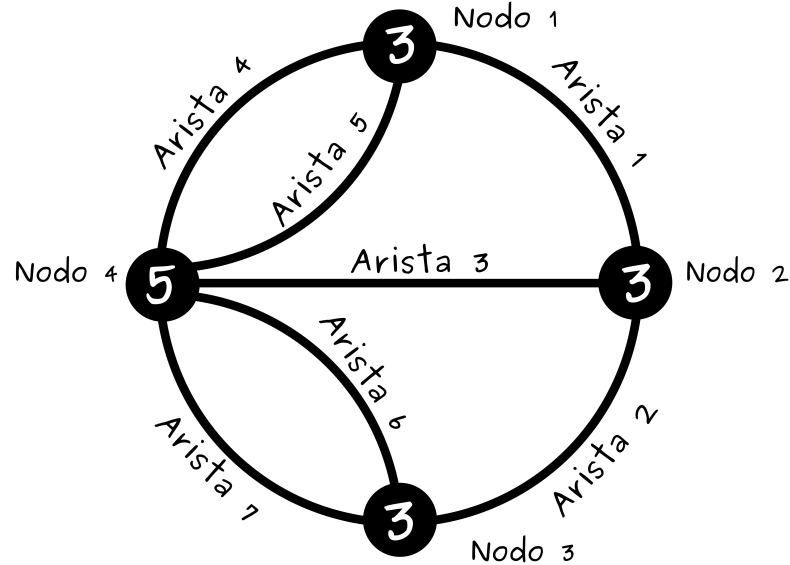


Figura 1.1: Ejemplo de una red con cuatro nodos y siete aristas. Los círculos de color negro representan a los nodos; las aristas son representadas por las líneas que conectan a los nodos. La etiqueta numérica dentro de los nodos (número de color blanco) hace referencia al grado del nodo, es decir, al número de aristas que un nodo tiene.

La figura 1.1 corresponde a la representación de una red con cuatro nodos y siete aristas. Cada uno de los nodos (círculos de color negro) simboliza al objeto de estudio que conforma la red (en nuestro caso cuentas de Twitter). Las aristas (líneas que conectan a los nodos), son las representaciones de las interacciones entre cuentas (tweets, retweets, etc). El número de interacciones que cada nodo tiene, se conoce como grado del nodo. Este número se obtiene al contar las aristas que tiene un nodo (etiqueta numérica de color blanco en la figura); y es el punto de partida para el análisis matemático y estadístico de una red [2].

El estudio matemático de una red, se conoce como teoría de redes. Esta teoría es fundamental para la comprensión de diversos fenómenos en campos tan dispares como pueden ser la física, la lingüística, la sociología, la biología y un largo etcétera. El gran poder que tienen las redes es que son representaciones abstractas y a la vez simples, de una enorme variedad de sistemas de distinta naturaleza, como por ejemplo: la Internet, las relaciones sociales entre individuos, las redes de regulación genética o metabólicas, las redes tróficas, las redes de suministro de energía, rutas de distribución, redes de citas entre artículos científicos, entre muchos otros. Muchos de los resultados que se obtienen del análisis matemático de una red son universales, es decir, comunes a muchos sistemas, constituyendo así la piedra angular para los estudios multidisciplinarios.

Ahora bien, desde el punto de vista de la teoría de redes, los enlaces entre nodos pueden ser de dos tipos: directos o indirectos. Decimos que un enlace es directo, si relaciona a dos nodos de manera bidireccional; por ejemplo como sucede en Facebook: sí yo soy **amigo** de Juan, él a su vez es mi **amigo**. Por otro lado, cuando hay una dirección en el enlace que

relaciona a los nodos, entonces se dice que la arista es indirecta; este es el caso de Twitter, en donde yo puedo ser **seguidor** de Juan (y por lo tanto sus publicaciones aparecen en mi línea de tiempo), pero Juan no necesariamente es mi **seguidor** (mis publicaciones no aparecen en su línea temporal).

Esta diferencia entre los enlaces, permite clasificar a las redes en dos tipos: **redes directas o redes dirigidas**, cuando todos los enlaces son directos; y **redes indirectas o no dirigidas**, que es cuando todos los enlaces son indirectos [2].

Una manera conveniente de representar a una red es a través de su **matriz de adyacencia**. Por ejemplo, en la figura 1.2 a), tenemos una red no dirigida, compuesta por cuatro nodos n_1, n_2, n_3 y n_4 . Es fácil ver que para este ejemplo, el nodo 1 está conectado con el nodo 2 y el nodo 3; a su vez, el nodo 2 está en conexión con los nodos 1, 3 y 4, y así con los demás nodos. La matriz de adyacencia de una red, es una matriz de N filas y N columnas, cuyos elementos satisfacen las siguientes condiciones:

- $A_{i,j} = 1$ si hay un enlace entre el nodo i y el nodo j .
- $A_{i,j} = 0$ si los nodos i y j no están conectados.

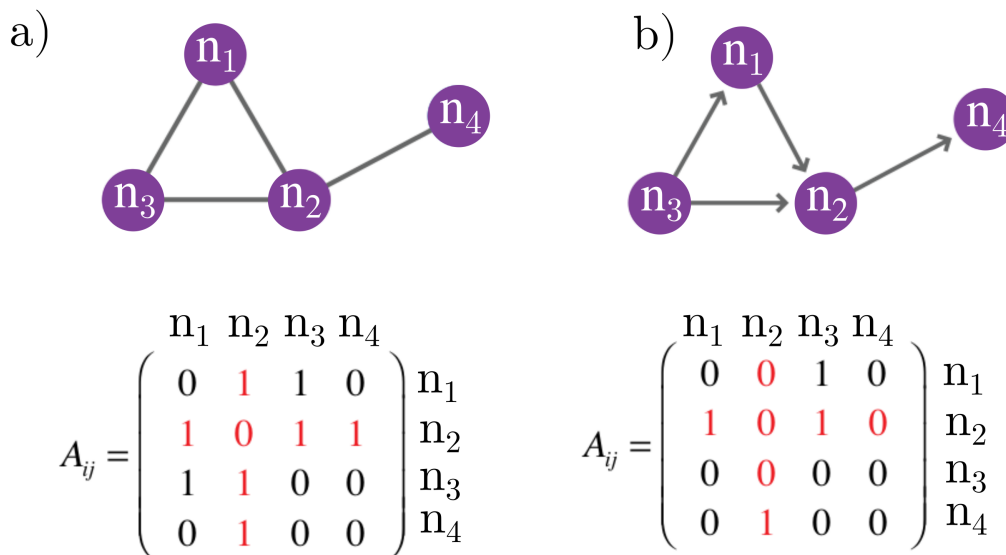


Figura 1.2: Redes y matrices de adyacencia. La figura a) es la representación de una red no dirigida. Para este caso, la matriz de adyacencia toma valores de 1 (en rojo) cuando dos nodos están conectados. En la figura b) tenemos una red dirigida. Ahora los elementos de la matriz de adyacencia toman valores distintos de 0 solo cuando el enlace es dirigido. Por ejemplo, el nodo 2 está conectado con el nodo 4, pero el nodo 4 no se encuentra en conexión con el nodo 2. (Figura tomada de [2]).

En el caso de una red dirigida, figura 1.2 b), la matriz de adyacencia se construye de la misma manera, solo que la primera condición debe satisfacer que $A_{i,j} = 1$, cuando hay un enlace dirigido entre el nodo i y el nodo j .

Una de las ventajas de tener esta representación matricial de una red, es que el grado

de un nodo se obtiene fácilmente a través de la siguiente fórmula:

$$k_i = \sum_{j=1}^N A_{i,j} = \sum_{i=1}^N A_{j,i}, \text{ para una red no dirigida.} \quad (1.1a)$$

$$k_i^{in} = \sum_{j=1}^N A_{i,j}, \quad k_i^{out} = \sum_{i=1}^N A_{j,i}, \text{ para una red dirigida.} \quad (1.1b)$$

Donde k_i^{in} es el número de enlaces que se dirigen hacia el i -ésimo nodo, y k_i^{out} es el número de enlaces que salen del nodo i .

Las redes también pueden clasificarse en aleatorias, cuando la distribución de probabilidad de los grados, sigue una distribución de Poisson; y en redes libres de escala, que es cuando la distribución de probabilidad del grado sigue una ley de potencias [2].

→1.4. Caracterización de nodos dentro de la red

Como hemos mencionado, la representación de la actividad e interacciones entre cuentas en la plataforma Twitter por medio de redes de interacción, nos permite hacer uso de conceptos y técnicas de la ciencia de redes para caracterizar, tanto la topología de la red en su conjunto, como a los nodos individuales con respecto a su posición y conexiones en la red.

Una manera de caracterizar a los nodos en la red es por medio de **medidas de centralidad**, las cuales cuantifican la importancia de los nodos en la red con respecto a distintas métricas, algunas de las cuales describimos a continuación:

- **Grado de entrada y salida:** El grado de un nodo, puede usarse como una medida de centralidad, y tiene por objetivo determinar a los nodos más importantes según su número de enlaces [10]. Si la red es dirigida, entonces el grado del i -ésimo nodo se obtiene al contar todos los enlaces de entrada y de salida del nodo

$$k_i = k_i^{in} + k_i^{out}, \quad (1.2)$$

En el caso que la red sea no dirigida, no hay distinción entre grado de entrada y salida, y decimos que k_i , dada por la ecuación 1.1a es el número de enlaces del nodo.

- **Eigenvectores:** La importancia de un nodo se obtiene al calcular el eigenvector correspondiente al eigenvalor más grande de la matriz de adyacencia de la red [6].
- **Pagerank:** La medida de centralidad dada por el **pagerank** se obtiene al realizar una caminata aleatoria por la red; esto es, dado el nodo i , se elige con una cierta probabilidad, alguno de los nodos vecinos a i . Este proceso se repite hasta que se han visitado todos los nodos de la red. La puntuación de centralidad se obtiene al calcular el tiempo promedio que se pasa en cada nodo durante la caminata aleatoria [8].

- **Betweenness para nodos:** La centralidad tipo *betweenness* mide la frecuencia con la que cada nodo aparece en la ruta más corta entre dos nodos cualesquiera de la red. Dado que pueden haber varios de estos caminos más cortos entre dos vértices i y j , la centralidad del nodo u esta dada por:

$$c(u) = \sum_{i,j \neq u} \frac{n_{ij}(u)}{N_{ij}}, \quad (1.3)$$

donde, $n_{ij}(u)$ es el número de caminos más cortos entre i y j que pasan a través de u , y N_{ij} de caminos más cortos entre i y j [10].

→1.5. Estructura de comunidades. El algoritmo de Girvan-Newman

Los seres humanos hemos evolucionado para tener un comportamiento gregario y cooperativo. Esto hace que tengamos la tendencia a formar vínculos con otros humanos con los que compartimos similitudes, dando lugar a la aparición de estructuras comunitarias en las sociedades. Las redes socio-digitales, al ser el reflejo de dichas sociedades, también se organizan en comunidades. Develar a ésta estructura es una de las tareas que puede realizarse con la ciencia de redes.

En teoría de redes, se dice que hay una estructura comunitaria en una red, si en ésta existen nodos densamente conectados entre sí formando grupos. A su vez dichos grupos se conectan unos con otros a través de conexiones muy específicas. Existen diferentes métodos para detectar a éstos enlaces; y en particular en este trabajo, usaremos el algoritmo de **Girvan-Newman** para la detección de comunidades [2].

En la figura 1.3 se muestra de manera esquemática cómo funciona el algoritmo de Girvan-Newman. El ingrediente clave aquí es el **betweenness de enlaces**⁴; que se define como el número de caminos más cortos que atraviesan a un enlace en una red [4]. Cada arista puede ser asociada con un valor de betweenness de enlace. Entonces, una conexión con un alto puntaje de éste betweenness representa a un conector (que hace las veces de **punte**) entre dos grupos de una red, y su eliminación puede separar a la red en comunidades. El algoritmo consta de los siguientes pasos:

1. Se calcula el betweenness para todos los enlaces de la red.
2. Se remueve de la red el enlace con mayor valor de betweenness.
3. Se vuelve a calcular el betweenness para todas las conexiones de la red afectadas por la remoción del anterior enlace.
4. Se repite desde el paso 2 hasta que la Modularidad sea máxima.

⁴Aunque similar, el **betweenness de enlaces** no debe confundirse con el betweenness de nodos. Existen diferentes algoritmos para calcular el **betweenness de enlaces**, ver por ejemplo [4] y referencias allí mencionadas.

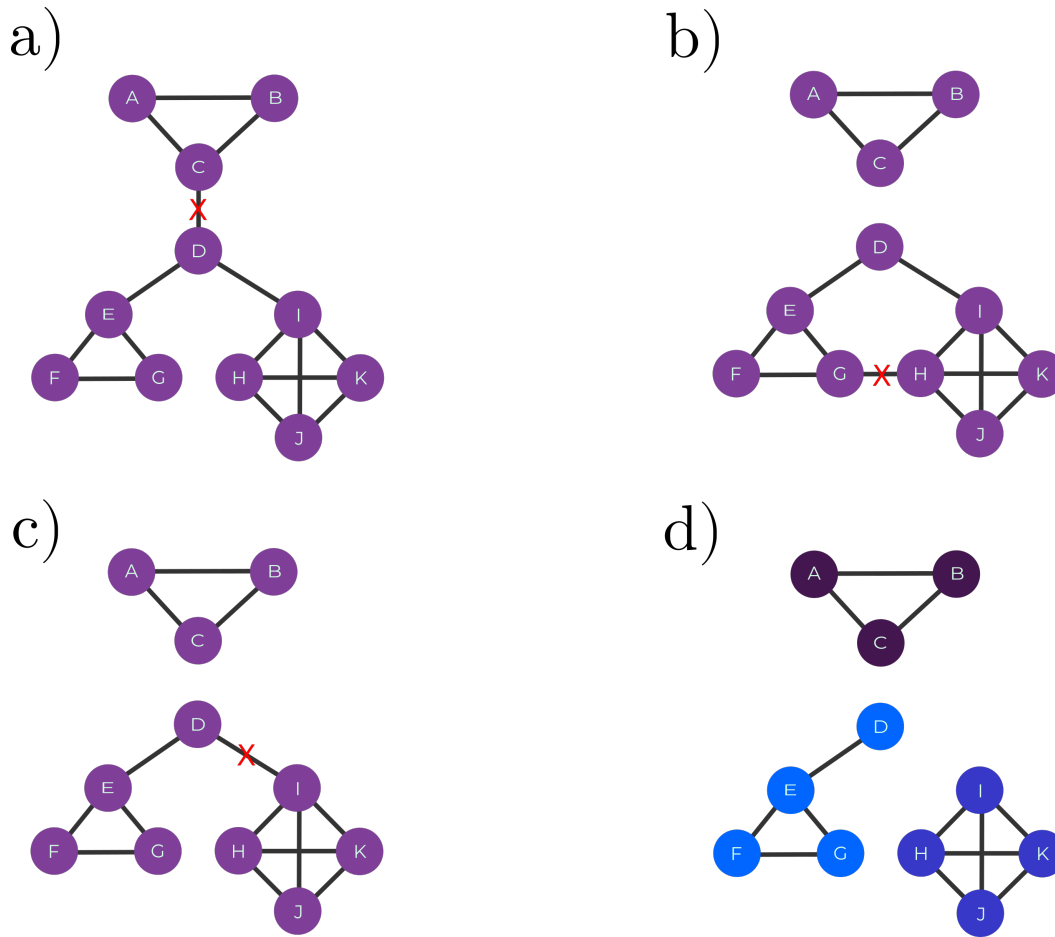


Figura 1.3: Representación gráfica del Algoritmo de Girvan-Newman. a) Dada una red, se calcula el betweenness de los enlaces y se remueve aquel con mayor valor (x de color rojo en la figura). b) Se repite el cálculo de betweenness de enlaces para determinar el nuevo enlace con mayor valor. c) El procedimiento se repite removiendo siempre el enlace con el valor más alto de betweenness. d) Este procedimiento permite encontrar la estructura comunitaria de la red.

La modularidad, se define como

$$M = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \right], \quad (1.4)$$

donde L_c es el número de enlaces en una red alternativa que se obtiene al aleatorizar las aristas la red original, mientras el grado de cada nodo se mantiene sin cambios. De igual manera k_c es el grado total de los nodos en una comunidad (se calcula recursivamente) y L es el número de enlaces en la red original. Finalmente n_c es un número que se fija desde el principio como parámetro inicial del algoritmo [2].

▶ 2. Metodología

Una vez que hemos hecho un recorrido por los conceptos fundamentales que nos permiten describir y caracterizar la actividad y patrones de interacción entre cuentas en la plataforma, procedemos ahora a la descripción de la propuesta metodológica que hemos desarrollado en Tlatelolco LAB, para el análisis de la actividad en Twitter alrededor de eventos coyunturales, con los principales objetivos de identificar las cuentas principales y comunidades de cuentas que participan en ellas, así como las narrativas asociadas a ellas.

→ 2.1. Eventos y narrativas como objeto de estudio

Aunque el contenido en las plataformas de las redes sociales digitales se genera de manera continua, hay momentos en los que la actividad en dichas plataformas es especialmente intensa, usualmente desencadenada por **eventos o coyunturas** que ocurren en la capa física o mediática [5], como por ejemplo eventos políticos; como las jornadas electorales, las declaraciones de algún actor político, o movilizaciones sociales, en las que las redes son utilizadas tanto para convocar, documentar y reaccionar a las mismas, o durante conmemoraciones que incluyen movilizaciones y marchas en las calles.

Es durante estos momentos de alta intensidad, que denominamos de manera general **eventos focales**, en donde se construyen distintas **narrativas** desde las múltiples visiones de los usuarios de las redes socio-digitales.

Dado que en éste documento, nos referiremos principalmente a la plataforma Twitter; nos enfocaremos en la resonancia y narrativas alrededor de los eventos focales que se manifiestan en ésta plataforma. Como se discutió en la sección 1.3; las cuentas, sus publicaciones e interacciones, nos permiten construir redes, que son nuestro principal objeto de estudio.

La construcción de redes, pone a nuestra disposición un diverso número de herramientas matemáticas y estadísticas, que constituyen el núcleo de nuestra metodología. Herramientas matemáticas tales como la detección de comunidades, medidas de centralidad, y el modelado de temas, nos permitirán estudiar y describir a las diferentes manifestaciones que las culturas políticas del México contemporáneo, en la capa digital.

→ 2.2. Roles y relevancia de cuentas en eventos focales

A grandes rasgos, podemos distinguir la actividad de las cuentas en las redes socio-digitales en dos grandes categorías: **las generadoras y las difusoras** de contenido. A su vez, a estas dos grandes categorías de actividad podemos asociarles dos principales roles:

Cuentas generadoras de contenido

La actividad de este tipo de cuentas está dominada por la generación de publicaciones propias, por medio de las cuales se establecen las distintas narrativas presentes en los eventos focales a los que hemos hecho referencia anteriormente. A este tipo de cuentas que establecen líneas discursivas o de sentido, usualmente se les denomina *influencers*.

Cuentas difusoras o amplificadoras de contenido

La actividad de este tipo de cuentas está dominada principalmente por *retweets* y *quotes*, acciones que redirigen el contenido de las publicaciones originales a los seguidores de las cuentas que realizan la acción (véase la sección 1.2), cuyo objetivo principal es el de **difundir o amplificar** las narrativas presentes en los eventos focales.

Es en este tipo de cuentas, en las que usualmente se ha observado la acción de mecanismos de manipulación de tráfico, y de amplificación coordinada de tendencias o hashtags. Si bien, este es un tema de suma importancia para el análisis de la conversación en las redes sociales digitales, no es tratado en el presente documento, ya que el objetivo de este trabajo es la identificación de las narrativas principales y las cuentas que las impulsan durante los eventos focales.

Características de las cuentas tipo *influencer*

Las cuentas generadoras de contenido que asociamos con las narrativas principales en las coyunturas, poseen ciertas características que las hacen relevantes. Dichas características les otorgan la posibilidad de influir en las decisiones, opiniones o interacciones de las cuentas con las cuales se relacionan. A dichas cuentas las denominamos como **influencers**. En general se consideran como aspectos principales de los influencers características tales como:

Popularidad:

- Número de seguidores.
- Recepción de interacciones de otras cuentas (Retweets, Respuestas, Likes).
- Difusión de otras cuentas del contenido que genera la autoridad, la actividad y la autoridad.

Actividad:

- Número de publicaciones.
- Participación de la cuenta en un evento.
- Generación de interacciones con otras cuentas.

Autoridad:

- Relaciones con otras cuentas “importantes”.
- Otras consideraciones cualitativas, como su relevancia en los temas que se discuten durante un evento focal, fuera de la capa digital.

→2.3. Audiencia, Alcance directo e indirecto de cuentas

Para caracterizar a las cuentas de Twitter, haremos las siguientes definiciones, que capturan a las características recién mencionadas

Audiencia: la definimos como la fracción de cuentas a las que un usuario puede llegar en la comunidad. Para ello definimos la audiencia total como:

$$F_T^C = \sum_{i \in C} f_i^C, \quad (2.1)$$

donde f_i^C son todos los seguidores, de todos los miembros de la comunidad C . Entonces, la audiencia para una cuenta i , se puede definir como

$$\alpha_{u_i}^j = \frac{f_i^j}{F_T^j}, \quad (2.2)$$

donde f_i^j es el número de seguidores de la i -ésima cuenta en la j -ésima comunidad.

Alcance directo: Si elegimos una comunidad, entonces el alcance directo es el número de usuarios al que una cuenta llega, multiplicado por el número de publicaciones que realiza en un intervalo de tiempo

$$A_{u_i}^D = \alpha_{u_i} (T_i + R_i + Q_i), \quad (2.3)$$

donde α_{u_i} es la audiencia, T_i es el número de tweets, R_i el número de Retweets y Q_i , el número de quotes.

Alcance indirecto: es el número de usuarios al que una cuenta llega, debido a las interacciones que realizan sus seguidores

$$A_{ij}^I = \alpha_{u_j} (R_{ij} + Q_{ij}) \quad (2.4)$$

Alcance indirecto total

$$A_{u_i}^I = \sum_{j \in \Omega_i} A_{ij}^I \quad (2.5)$$

Alcance total es la combinación de las ecuaciones anteriores

$$A_{u_i}^T = \omega A_{u_i}^D + (1 - \omega) A_{u_i}^I, \quad (2.6)$$

donde el parámetro ω nos permite seleccionar un alcance u otro

- $\omega = 0 \rightarrow$ Solo alcance indirecto
- $\omega = 1 \rightarrow$ Solo alcance directo
- $\omega = 1/2 \rightarrow$ Mismo peso para ambos

→2.4. Caracterización del contenido de las publicaciones

Hasta el momento nos hemos enfocado en la caracterización de las cuentas y las posibles redes de interacción entre ellas, solamente tomando en consideración la actividad individual y colectiva de ellas en la plataforma, sin prestar atención al contenido que publican y comparten.

Si bien, el tipo de descripción que hemos realizado hasta el momento nos permite identificar

Modelado de temas

Parte importante de la presente metodología, es hacer análisis estadístico de textos (tweets) a través del modelado de temas; es decir, usando algoritmos que agrupan palabras similares, y que al estudiarlas dentro de un marco matemático, son relevantes para examinar, organizar conjuntos de documentos e inferir los temas y el balance de los mismos en un corpus determinado [1].

Entre los diferentes métodos que existen para identificar los temas en un corpus dado, el **Latent Dirichlet Allocation** (LDA por sus siglas en Inglés), es uno de los métodos más usados en el modelado de temas, y es el que usaremos en esta metodología.

Latent Dirichlet Allocation

A continuación haremos una breve reseña de cómo el método LDA funciona. La información presentada aquí está tomada de [3]. Formalmente, definimos los siguientes términos:

- Una **palabra** es en este caso definida como la mínima unidad (discreta) de dato. Las palabras son tomadas de un conjunto ordenado $\{1, \dots, V\}$ llamado **vocabulario**.
- Un **documento** es una secuencia de N palabras denotadas por $\mathbf{w} = \{w_1, w_2, \dots, w_N\}$, donde w_N es la n -ésima palabra en la secuencia.
- Un **corpus** es una colección de M documentos denotado por $\mathcal{D} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M\}$.

La idea detrás del LDA es la de encontrar un modelo probabilístico a partir de un corpus que permita asignar probabilidades a conjuntos de palabras que se encuentran relacionadas temáticamente. Dicho modelo se puede calcular usando siguiente ecuación:

$$p(\mathcal{D}|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d. \quad (2.7)$$

Esta ecuación describe a un modelo probabilístico, en donde los documentos son representados por una mezcla de temas (z_{dn}) latentes¹, y en donde cada tema está caracterizado por una distribución de probabilidad (θ_d) sobre las palabras. Esquemáticamente la ecuación (2.7) puede ser representada por la siguiente figura:

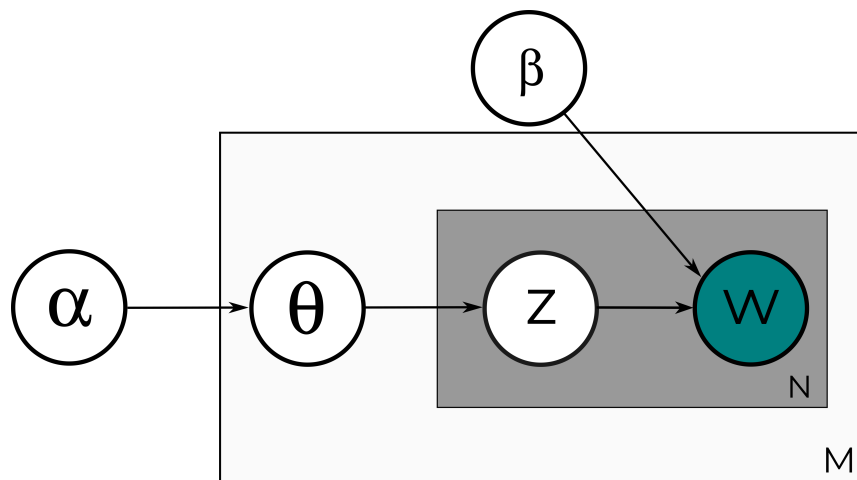


Figura 2.1: Representación gráfica del modelo de LDA. Las cajas son "placas" que representan réplicas. La placa exterior representa al conjunto de documentos (corpus de tweets en nuestro caso); mientras que la placa interior representa a cada documento (tweet). w representa al conjunto de palabras que componen al documento, y el hecho que esté en color turquesa significa que esta es la única variable observable en el modelo. Las demás variables (θ y z) son latentes. Finalmente α y β son parámetros de la distribuciones que deben ser calculados.

En la figura (2.1), z representa a los diferentes temas que hay en el conjunto de documentos, θ es la distribución de probabilidad de dichos temas. Adicionalmente, w hace referencia al conjunto de palabras, que son los únicos observables del modelo.

El método LDA consiste en determinar los parámetros α y β . El parámetro α está asociado a la distribución de probabilidad θ , es decir a la probabilidad que un tema aparezca en el conjunto de documentos; y el parámetro β se asocia a la probabilidad que un conjunto de palabras pertenezcan a un tema determinado.

La distribución de probabilidad $p(\theta_d|\alpha)$ se escoge de una distribución de Dirichlet (De ahí el nombre **Latent Dirichlet**); mientras que la distribución de probabilidad $p(w_{dn}|z_{dn}, \beta)$ se toma de una distribución multinomial.

Matemáticamente hablando, una distribución de probabilidad está identificada por sus

¹Latente viene del latín latens, que quiere decir oculto

parámetros. En nuestro caso, α y β son números reales y en principio cualquier número real permitiría resolver la integral (2.7). Sin embargo, dicha integral debe ser resuelta para determinar el **mejor** conjunto de parámetros que describan al corpus de documentos bajo estudio.

El proceso de encontrar al mejor conjunto de parámetros para describir el conjunto de documentos se conoce como **ajuste del modelo estadístico**. Dicho ajuste consiste en encontrar (usando técnicas matemáticas tales como la inferencia bayesiana, la simulación Monte-Carlo, o muestreo de Gibbs, entre otros) los parámetros que mejor describen a los datos u observables. Para ello, se debe **entrenar al modelo**. Esto se hace seleccionando de manera aleatoria un 10 % de los documentos, y con ellos se hace una validando al calculando la perplexidad:

$$\text{perplexity}(\mathcal{D}) = \exp - \frac{\sum_{d=1}^M \log p(\mathbf{w})}{\sum_{d=1}^M N_d}, \quad (2.8)$$

donde $p(\mathbf{w}) = \sum_{z=1}^k p(z) \prod_{n=1}^n p(w_n|z)$ y $\log p(\mathbf{w})$ es el **log-likelihood** de la probabilidad que una palabra aparezca en el el corpus de documentos de entrenamiento; N_d es el número de documentos y k el número de temas. En términos técnicos, entrenar al modelo significa hacer el ajuste para distintos valores del número de temas. En la siguiente figura podemos observar como se comporta la perplexidad como función del número de temas.

El objetivo de este entrenamiento es el de elegir un valor para k que minimice a la perplexidad. De la figura (2.2) es claro que para el ejemplo presentado, el valor de 15 representa un valor apropiado para el número de temas.

Una vez hallado el valor k , es decir, una vez encontrado el número de temas que posiblemente hay en el corpus, se procede a repetir el cálculo, pero esta vez se usa el 90 % de los documentos restantes. Esto permite asignar la probabilidad que una palabra del corpus pertenezca a cada uno de los temas determinados en el proceso de entrenamiento. Con dichas probabilidades se pueden hacer nubes de palabras relacionadas temáticamente; (ver capítulo 4 para más detalles).

Es importante señalar que dependiendo del corpus, habrá que tomar más o menos documentos para entrenar el modelo. Esa decisión depende de la cantidad de datos disponibles y del acceso a recursos computacionales para ajustar al modelo estadístico.

Búsqueda Inversa

La búsqueda inversa se basa en la idea que, una vez determinado el número de temas que hay en un corpus, se puede calcular el grado de pertenencia que tiene un documento a un tema determinado. Esto se hace determinando cuantas palabras en el documento i -ésimo pertenecen al tema j -ésimo

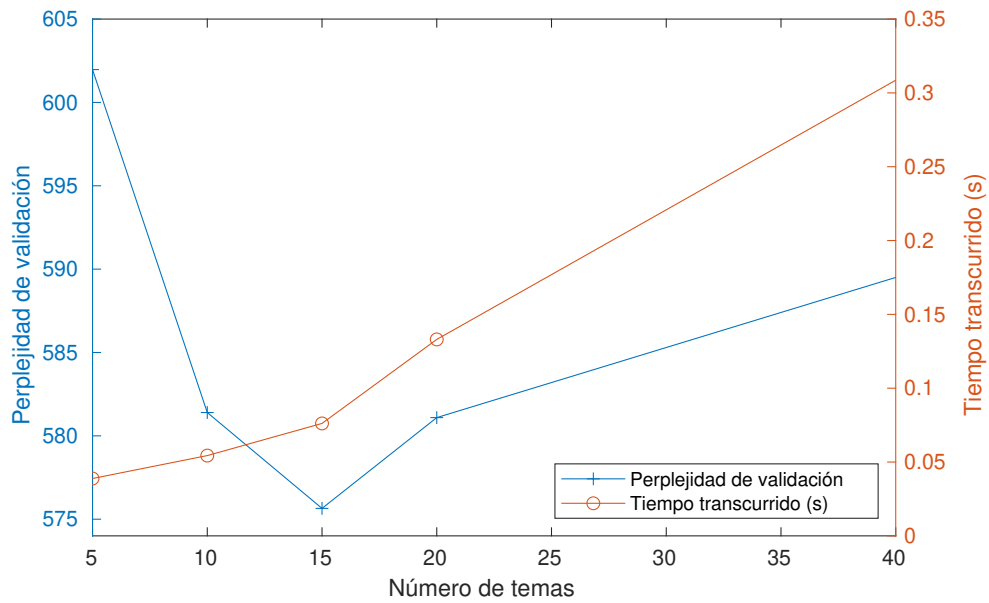


Figura 2.2: Perplejidad como función del número de temas. La línea roja representa el tiempo que se tarda el programa para determinar a la perplejidad. Este tiempo computacional es importante a tener en cuenta, dado que la cantidad de documentos en el corpus puede tener cientos de miles de elementos.

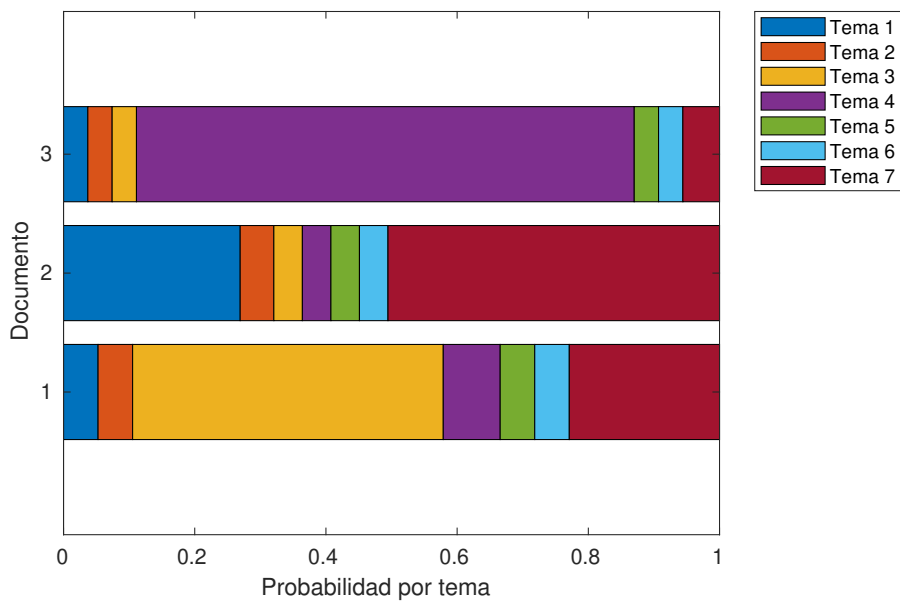


Figura 2.3: Mezcla de temas. En esta figura podemos ver cual es el grado de pertenencia, medido en número de palabras que pertenecen a cada uno de los temas determinado por el método, para cada uno de los documentos (3 en este caso).

En la figura (2.3) se visualiza la mezcla de temas para tres documentos de un corpus. De la figura es claro que cada documento tiene más pertenencia a un tema que a los otros, lo cuál permite asignar una etiqueta a cada documento. En éste ejemplo el documento 1 pertenece principalmente al tema 5, al igual que el documento 2; el documento 3 pertenece más al tema 2 que a los otros temas encontrados por el LDA.

La búsqueda inversa, junto con las otras herramientas que hemos presentado en este capítulo, aunadas a la información que se obtiene de la API de twitter (ID de usuario, fecha y hora de publicación, número de retweets, quotes, etc) constituyen la base metodológica para estudiar a las diversas expresiones que las culturas políticas manifiestan en Twitter, y potencialmente otros tipos de redes socio-digitales.

▶ 3. Aplicación de la metodología

Para ejemplificar la propuesta metodológica descrita en los capítulos anteriores, usaremos como evento focal, las manifestaciones realizadas en el marco de la conmemoración del día internacional de la mujer el 8 de Marzo de 2021. Este evento tiene gran relevancia en la plataforma de Twitter, ya que es uno de los principales medios de convocatoria y documentación de las distintas manifestaciones públicas realizadas durante este día.

Para aplicar la metodología a este evento focal en particular, recopilamos a través de la API de Twitter un total de 2,524,873 tweets publicados entre 08/03/2021 y el 09/03/2021, que contenían diferentes hashtags relativos al día internacional de la mujer (#8M2021, #NiUnaMenos, #NiUnaMas, #VivasNosQueremos, entre otros).

→ 3.1. Detección y caracterización de comunidades

Con los datos obtenidos, construimos una red de retweets con un total de 1,004,778 cuentas participantes, en la que identificamos un total de 238 comunidades por medio del algoritmo de Girvan-Newman, de las cuales, nos enfocamos en las 10 comunidades con un mayor número de cuentas asociadas.

Una vez que hemos obtenido las comunidades de cuentas identificadas por el algoritmo de detección de comunidades en la red de retweets, procedemos a realizar un etiquetado de las mismas. Este proceso depende de las características específicas del evento de estudio, y puede realizarse de distintas maneras y criterios. Usualmente, una revisión manual de algunas de las cuentas representativas de la comunidad, el uso de expresiones idiomáticas en los textos, el uso de ciertos hashtags, o el uso de metadatos proporcionados por la API, tales como la geolocalización; además del conocimiento previo sobre la coyuntura o evento que se este analizando (entre otros), son los elementos necesarios para identificar y etiquetar a la comunidad detectada por el algoritmo, dentro del contexto del evento focal en cuestión.

En este caso en particular, una revisión de las cuentas involucradas y el uso de ciertas expresiones idiomáticas, nos permitió asociar a las 10 comunidades con distintos países de hispanoamérica, como: Argentina, Chile, España y México, además de una comunidad cuyos mensajes contenían expresiones de distintos países y que etiquetamos de manera general como hispanohablante.

En la figura 3.1, se muestra el número de cuentas en las 10 comunidades principales detectadas en el evento focal. La comunidad más grande (España 1) esta formada por cerca de 70,000 cuentas, mientras que la comunidad más pequeña (Chile 1) es de poco menos de 30,000 cuentas.

Como se describió en la sección 1.5, el algoritmo de Girvan-Newman, identifica y remueve sistemáticamente los enlaces con mayor betweenness, para determinar la estructura modular de la red. Dichos enlace pueden pensarse como puentes que conectan

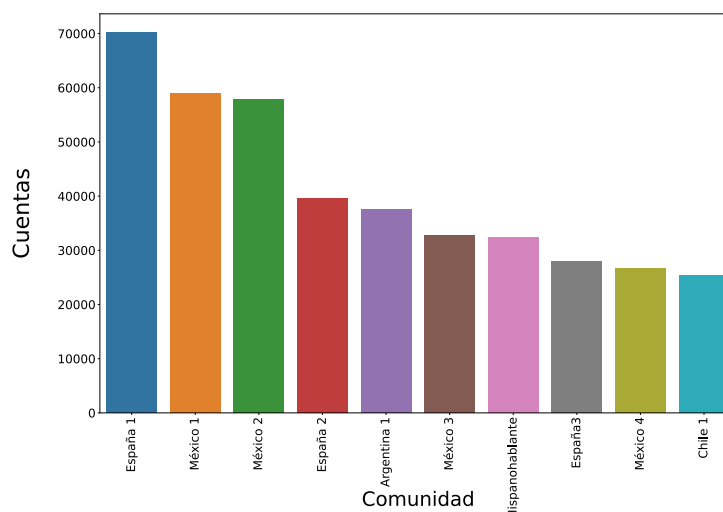


Figura 3.1: Número de cuentas en cada una de las 10 comunidades principales durante el evento focal del 8M 2021. Las etiquetas (España 1, México 2, etc) fueron asignadas tras una revisión manual de algunas de las cuentas participantes y del uso de ciertas expresiones idiomáticas.

a distintas (sub)redes. Es decir, cada comunidad se puede pensar así mismo como una red, que puede ser estudiada con todas la herramientas de la teoría de redes, como: distribución de grados, clustering coefficient, densidad de enlaces, etc.

De las muchas herramientas de las que podemos hacer uso de la teoría de redes, las medidas de centralidad, (véase la sección 1.4), son útiles para caracterizar a los nodos (cuentas) dentro de cada comunidad, e identificar a las que desempeñan un papel central en cada una de ellas.

Autoridad		
PageRank	Eigenectores	Betweenness
lopezdoriga	lopezdoriga	lopezdoriga
deniseramosm	latinus_us	deniseramosm
AleCrail	AleCrail	AleCrail
latinus_us	JLozanoA	latinus_us
vampire	ferbelaunzaran	Ilse_Asecas

Tabla 3.1: Cuentas con valores más altos de las medidas de centralidad asociadas con las características de autoridad de las cuentas tipo influencer, de la comunidad México 1.

En la Tabla 3.1 se muestran las 5 cuentas con los valores más altos de las medidas de centralidad asociadas con las características de autoridad de las cuentas tipo influencer. Una rápida inspección a estas cuentas, nos permite ver que esta comunidad está formada por aquellas cuentas que hicieron retweet a las publicaciones de cuentas de periodistas (@AleCrail, @lopezdoriga, @deniseramosm), así como también a algunas cuentas de medios de comunicación como @latin_us.

La popularidad, otra de las características de las cuentas tipo influencer de la que hablamos en el capítulo 2, puede ser cuantificada a través del alcance total (ecuación 2.6), alcance directo (ecuación 2.3) y el grado de entrada (ecuación 1.1b).

Popularidad		
Alcance Total	Alcance Directo	Grado de entrada
El_Universal_Mx	El_Universal_Mx	lopezdoriga
lopezdoriga	lopezdoriga	latinus_us
Excelsior	Excelsior	deniseramosm
NTelevisa_com	NTelevisa_com	AleCrail
CarlosLoret	CarlosLoret	azucenau

Tabla 3.2: Cuentas con valores más altos de las medidas de centralidad asociadas con las características de popularidad de las cuentas tipo influencer, de la comunidad México 1.

Al examinar la Tabla 3.2 en la que reportamos las cuentas con los valores más altos de la medidas de popularidad, se observa que la comunidad México 1 está compuesta por cuentas asociadas a medios de comunicación tradicionales como @El_Universal_Mx, @NTelevisa_com, @Excelsior.

Actividad		
Alcance Indirecto	Publicaciones	Grado de salida
lopezdoriga	Abundanza	r2d2_thx1138
fotoaccionuni	Excelsior	sandia19334752
latinus_us	lasillarota	Mviera56969722
El_Universal_Mx	EjeCentral	jr2d26212
azucenau	elsolde_mexico	Jorgelechef

Tabla 3.3: Cuentas con valores más altos de las medidas de centralidad asociadas con las características de actividad de las cuentas tipo influencer, de la comunidad México 1.

Finalmente, con el alcance indirecto (ecuación 2.4), el número de publicaciones de una cuenta y el grado de salida de un nodo (ecuación 1.1b), podemos caracterizar la Actividad que tienen las cuentas dentro de una comunidad. En la tabla 3.3 mostramos las cuentas más activas durante el 8M. Nuevamente aparecen cuentas relacionadas a medios de comunicación (@lasillarota, @elsolde_mexico, @EjeCentral).

De las tablas 3.1, 3.2 y 3.3 podemos concluir que la comunidad México 1, es una comunidad que reúne principalmente a cuentas que hicieron retweets a medios de comunicación que informaron sobre las manifestaciones del 8M en México. Vemos pues, que esta caracterización es útil para identificar la composición de las comunidades que participaron en el evento focal.

Las medidas de centralidad, que cuantifican las características de las cuentas tipo influencer (autoridad, actividad y popularidad), también pueden ser útiles para enfatizar de mejor manera a las cuentas relevantes de las comunidades en una visualización de las redes de interacción entre cuentas de las comunidades.

En la Figura 3.2, mostramos un par de representaciones de la red construida con los Retweets de la comunidad México 1. En la figura 3.2(a), el tamaño del nodo esta relacionado con el número de seguidores de una cuenta. En la figura 3.2(b), el tamaño de los nodos corresponde al alcance directo de cada una de las cuentas. Se puede observar como los nodos (cuentas) que resaltan, dependen de la característica que se desee enfatizar.

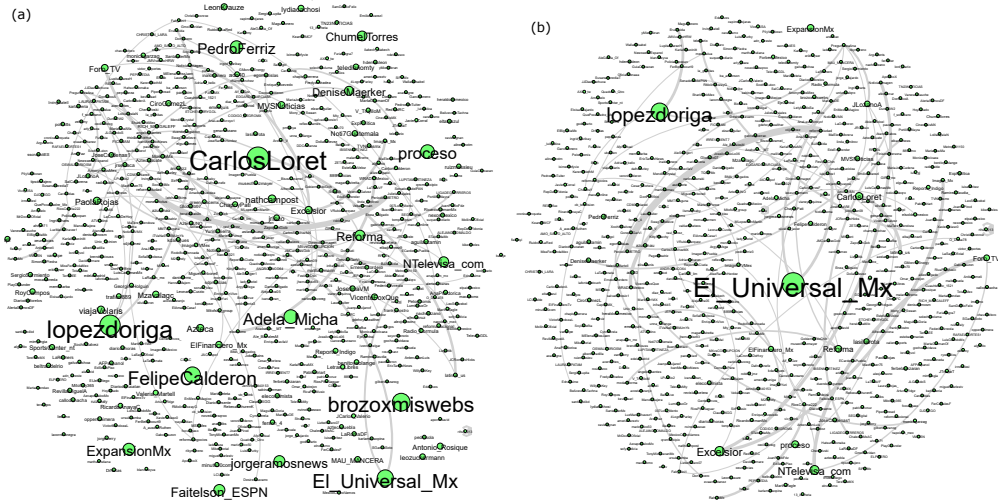


Figura 3.2: Red de Retweets para la comunidad México 1. En la figura (a) hemos codificado el tamaño del nodo con el número de seguidores de cada cuenta. En la figura (b), la codificación se ha hecho para dar cuenta del alcance directo de la cuenta.

➔3.2. Audiencia y Alcance de las cuentas

Ya que hemos caracterizado a las comunidades, caracterizamos ahora a las cuentas que participaron en el evento focal con el objetivo de obtener información relevante de cada una de ellas. Para ello, los conceptos de audiencia 2.2 y alcance total 2.6, son claves para entender el papel que las cuentas juegan en las comunidades.

Por un lado, en la figura (3.3) (a) observamos el alcance directo e indirecto para cada una de las comunidades. En este caso podemos observar que el alcance indirecto (azul), contribuye muy poco al alcance total, y que en cada comunidad, el alcance total esta dado principalmente por el alcance directo.

Por otro lado, en la figura (3.3) (b), mostramos el tamaño de las audiencias de cada una de las comunidades. Esta información nos permite tener una noción del número de cuentas a las que un mensaje llega de manera directa. Por ejemplo, la comunidad España 1, a pesar de ser la comunidad más grande, en términos de cuentas participantes, tiene una audiencia menor que la comunidad México 1 (que es más pequeña). Esto es debido, probablemente, a que los mensajes de la comunidad México 1 son retuiteados con más frecuencia, ya que provienen principalmente de medios de comunicación. Como veremos más adelante, la comunidad España 1, parece pertenecer a cuentas de personas comunes más que a medios de comunicación.

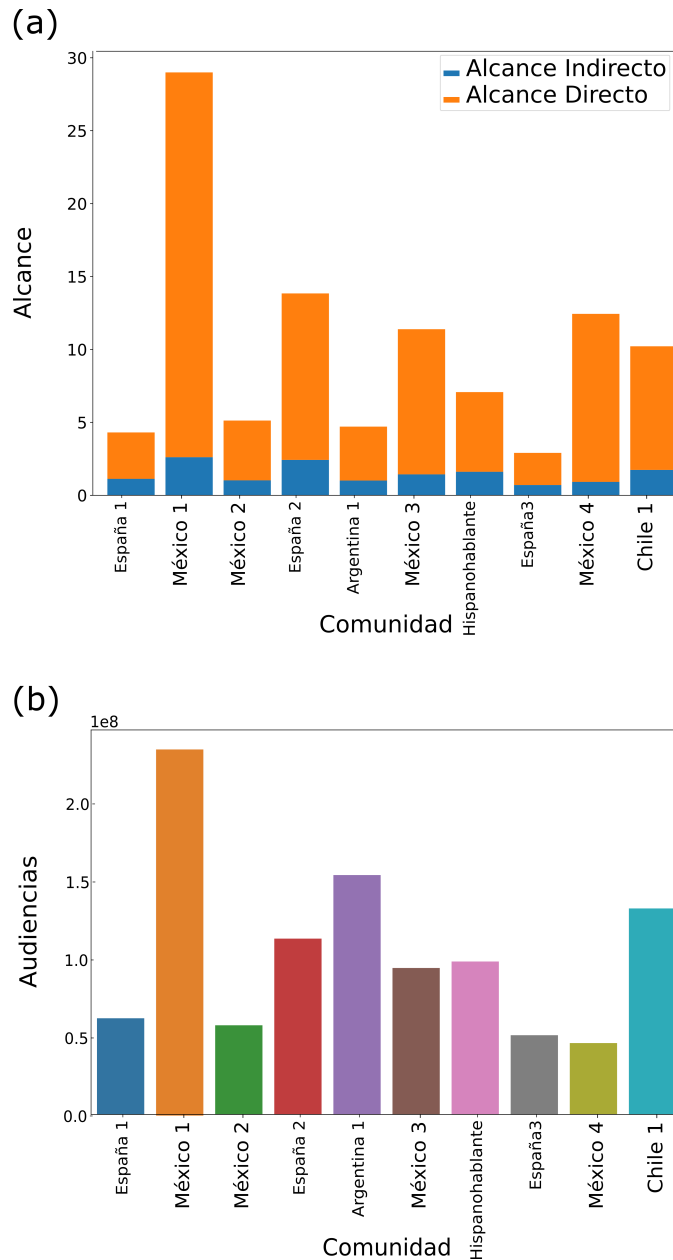


Figura 3.3: Comunidades como función de su alcance y audiencia. (a) Gráfico de barras del alcance directo (naranja) e indirecto (azul) para cada comunidad. (b) Figura donde se muestra las audiencias para cada una de las comunidades.

Asimismo, en la figura 3.4, mostramos la relación entre alcance directo, alcance indirecto y alcance total para cada una de las cuentas de una comunidad. Las líneas rojas continuas representan a las medias para los alcances directo e indirecto. Es importante notar que éstas líneas dividen la figura el tres cuadrantes, que pueden usarse para determinar la importancia de las cuentas en términos de sus alcances. Con base en lo anterior, clasificamos las cuentas en una comunidad en tres tipos:

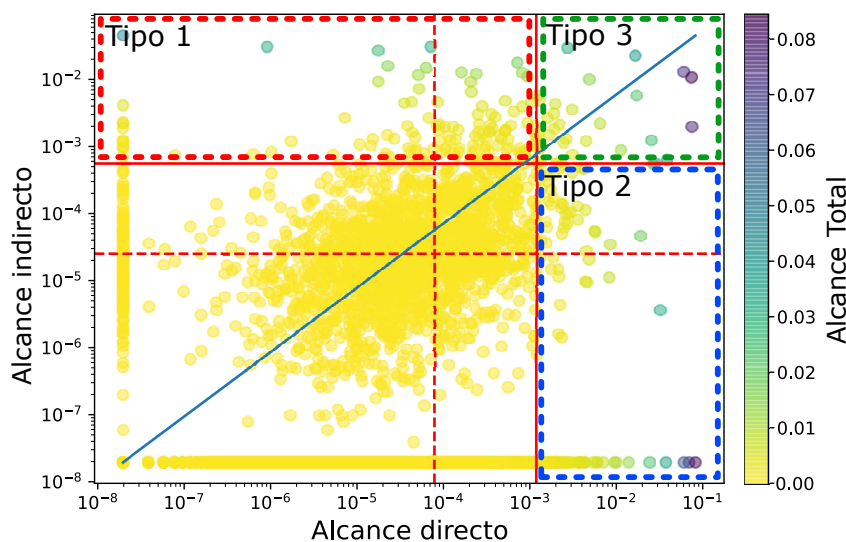


Figura 3.4: Relación entre alcance directo y alcance indirecto, para cada una de las cuentas de la comunidad España 1 (en escala log-log). Los puntos representan a cada una de las cuentas de la comunidad; las líneas continuas representan las medias de los alcances, y las líneas punteadas son las desviaciones estándar de cada uno de los alcances. El color de cada punto codifica el valor del alcance total, siendo amarillo, un valor bajo y púrpura un valor alto del alcance total. Esta información nos permite distinguir entre 3 tipos de cuentas.

- **Tipo 1:** son aquellas cuentas que están por encima de la desviación estándar del alcance indirecto y debajo de la desviación estándar del alcance directo. Estas son cuentas cuyo alcance resulta de las interacciones de otras cuentas con sus publicaciones; es decir, depende de la difusión generada por el resto de las cuentas en la comunidad.
- **Tipo 2:** son aquellas cuentas que están por debajo de la desviación estándar del alcance indirecto y por encima de la desviación estándar del alcance directo. Son cuentas cuyo alcance está dominado por sus propios seguidores.
- **Tipo 3:** son las cuentas que se encuentran por encima tanto de la desviación estándar del alcance directo como del alcance indirecto. Estas cuentas tienen el mayor alcance en la comunidad, debido a un alto número de seguidores, pero también por la difusión generada por las cuentas de la comunidad.

Las cuentas tipo 3, debido a su alcance e influencia, suelen ser importantes para investigar la propagación de la información en las redes socio-digitales, y la construcción de narrativas alrededor de un evento focal.

Por otro lado, al comparar el alcance total y el número de seguidores de las cuentas que participaron en el evento focal (véase la figura 3.5), observamos que el alcance total no crece necesariamente con el número de seguidores de la cuenta. Esto sugiere que la noción de que una cuenta es **influencer** solamente debido a su número de seguidores no es necesariamente cierta. Hacer esta distinción entre la noción popular de influencer y la descripción que hemos desarrollado en este documento, es fundamental para la identificación de los actores claves y sus narrativas en un evento focal.

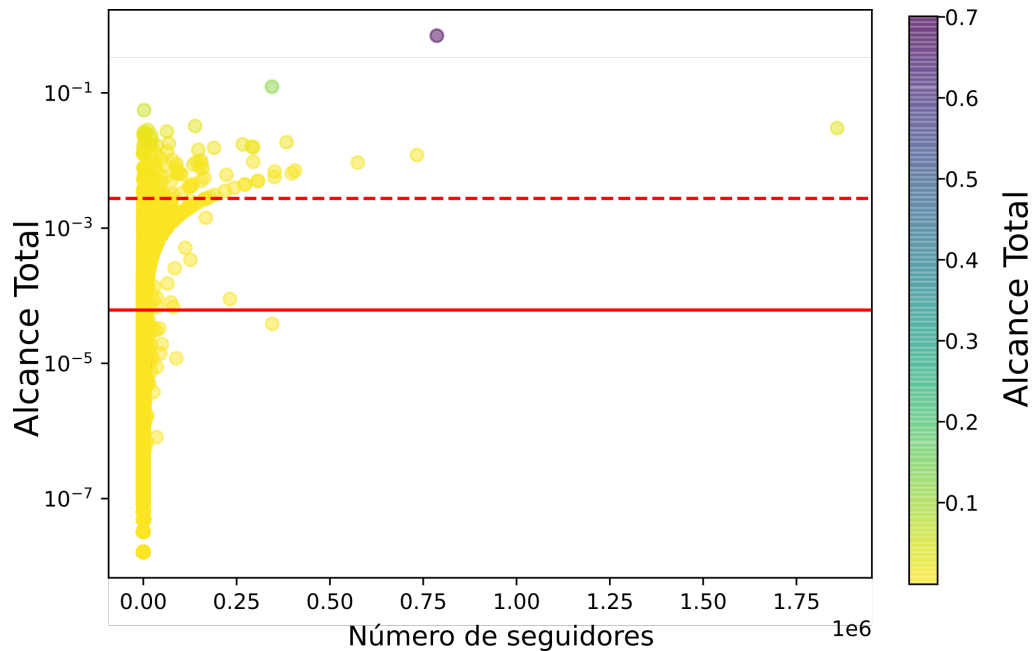


Figura 3.5: Alcance total como función del número de seguidores para cada una de las cuentas de la comunidad España 1. La línea horizontal continua de color rojo representa la media del alcance total, mientras que la línea punteada representa la desviación estándar del alcance.

Modelado de temas y búsqueda inversa

Con la identificación de comunidades que participan de un evento focal, podemos obtener información muy útil para entender las discusiones sociales en las redes socio-digitales. Sin embargo, para comprender mejor éstas discusiones es clave, saber qué están diciendo los usuarios que participan en un evento focal en Twitter. En la sección 2.4, vimos que el Latent Dirichlet Allocation puede usarse para determinar los temas subyacentes en un conjunto de documentos.

Al recopilar todos los (re)tweets que aparecen en el evento focal, es posible construir un conjunto de documentos, o corpus, que pueda ser estudiado con LDA. Por ejemplo, al analizar el corpus de textos con el que fue construida la red que permitió identificar a la comunidad España 1, identificamos cerca de 90 temas diferentes. Sin embargo, los cuatro temas con mayor probabilidad fueron los siguientes:

De la figura 3.6 se puede observar, que los temas 1 y 2 se refieren en términos desobligantes (y típicos de el idiolecto de España) a las manifestaciones del 8M. Esto último fue clave, por ejemplo, para asignar las etiquetas para cada comunidad (España 1, México 1, etc). La red temática construida con la búsqueda inversa (centro de la figura), permite identificar a aquellas cuentas que participan en las discusiones más probables de esta comunidad; además de permitir observar la interrelación de las discusiones de la comunidad y ayudando a dar el contexto necesario analizar las discusiones.

Esta red temática, junto con las medidas de centralidad, alcances, edad, etc, es un

elemento fundamental para el análisis de las narrativas y actores durante eventos focales.

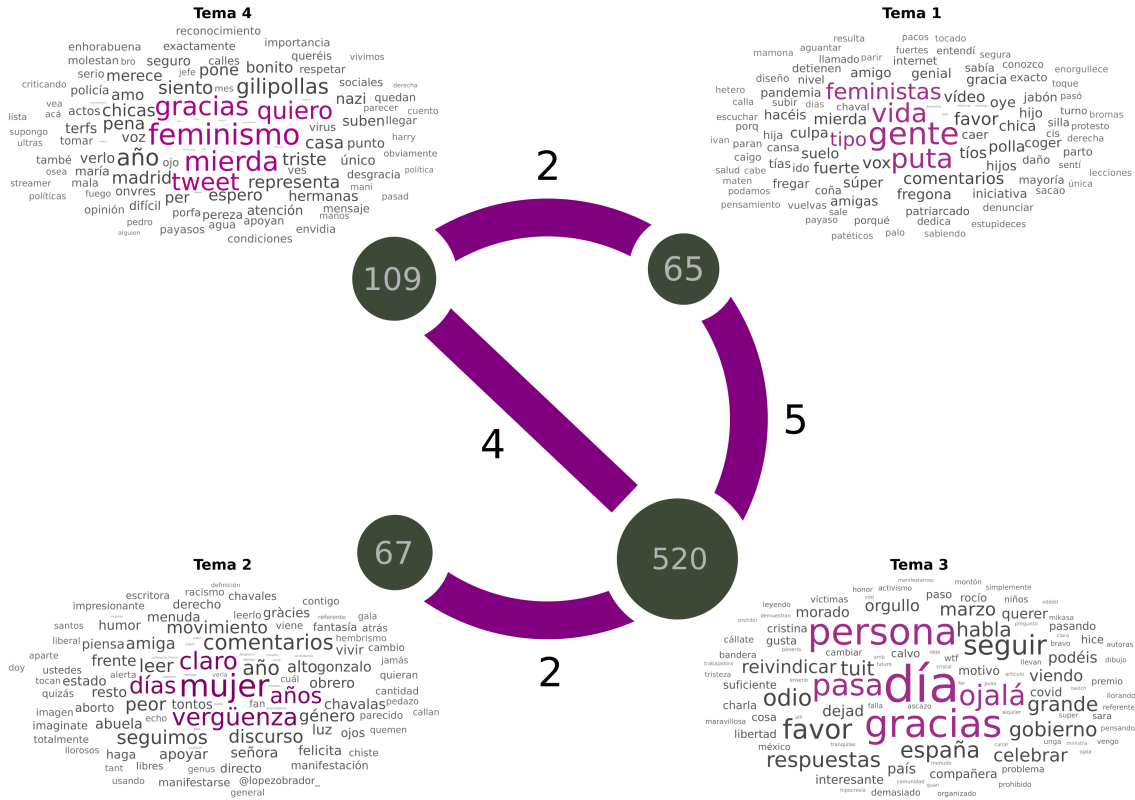


Figura 3.6: Temas determinados por LDA para la comunidad España 1. En el centro se observa la representación de la red temática construida con la búsqueda inversa. El tema 3 concentró al mayor número de usuarios (520). Dos de estos usuarios participan también en la discusión del tema 2, 5 del tema 1 y 4 del tema 4.

►4. Conclusiones

La naturaleza de la dinámica que se desarrolla en las redes socio-digitales, presenta tanto retos como oportunidades para su estudio. Como hemos discutido en este documento, a diferencia de los medios de comunicación tradicionales, en las redes socio-digitales, una gran parte del contenido se genera y difunde mediante la acción colectiva de las cuentas que participan en las plataformas, por lo que las cuentas no tienen un papel pasivo sino activo tanto en la generación como en la amplificación de las lecturas y narrativas alrededor de eventos focales.

Estas características de las plataformas, así como la posibilidad de obtener, almacenar y procesar la actividad de las cuentas por medio de herramientas computacionales, nos han permitido analizar los patrones de interacción entre cuentas haciendo uso de herramientas de la ciencia de redes (véase el capítulo 1), lo cual, como hemos descrito y desarrollado en este documento, es útil para identificar a las distintas comunidades, y las cuentas relevantes presentes en ellas mediante su alcance directo e indirecto, que se organizan en las redes socio-digitales, alrededor de eventos focales. En general, estas comunidades pueden corresponder a distintas lecturas y narrativas, y representar afinidades ideológicas, cuya identificación y análisis es relevante en el para el estudio de la conversación e identificar polarización y disputa de narrativas en la esfera digital.

Por otro lado, aunque no es tratado en este documento, las herramientas metodológicas que hemos desarrollado, son la base de partida para la identificación y descripción de mecanismos de manipulación del tráfico, mediante patrones de interacción de cuentas que interactúan preferentemente entre ellas, realizando retweets mutuos. Esta es una de las líneas de investigación a ser desarrolladas por el Tlatelolco Lab.

Así mismo, es importante mencionar que, los resultados que presentamos aquí, pueden complementarse con análisis de las figuras más compartidas. Todo lo expuesto en el presente documento, en conjunto con una interpretación multidisciplinaria que considere aspectos sociológicos, permitirá realizar una descripción amplia y tener un mejor entendimiento de las culturas políticas y la construcción de sentidos comunes en las redes socio-digitales.

Bibliografía

- [1] Rafael E. Banchs. *Text Mining with MATLAB*. Springer, 2013.
- [2] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016.
- [3] Michael I. Jordan David M. Blei Andrew Y. Ng. «Latent Dirichlet Allocation». En: *Journal of Machine Learning Research* 3 (2003), págs. 993-1022.
- [4] M. Girvan y M. E. J. Newman. «Community structure in social and biological networks». En: *Proceedings of the National Academy of Sciences* 99.12 (2002), págs. 7821-7826. ISSN: 0027-8424. DOI: 10.1073/pnas.122653799. eprint: <https://www.pnas.org/content/99/12/7821.full.pdf>. URL: <https://www.pnas.org/content/99/12/7821>.
- [5] Ackerman John y col. *Fundamentos teóricos de Tlatelolco Lab. Nueve dimensiones para el estudio de las luchas por la democracia*. Inf. téc. UNAM, Mexico: Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad, 2021.
- [6] Mark EJ Newman. «The mathematics of networks». En: *The new palgrave encyclopedia of economics* 2.2008 (2008), págs. 1-12.
- [7] Ben Nimmo. *Measuring Traffic Manipulation on Twitter*. Inf. téc. Computational Propaganda Research Project, 2019.
- [8] Lawrence Page y col. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report 1999-66. Previous number = SIDL-WP-1999-0120. Stanford InfoLab, 1999. URL: <http://ilpubs.stanford.edu:8090/422/>.
- [9] Lorena Recalde y col. «Who You Should Not Follow: Extracting Word Embeddings from Tweets to Identify Groups of Interest and Hijackers in Demonstrations». En: *IEEE Transactions on Emerging Topics in Computing* 7.2 (2019), págs. 206-217. DOI: 10.1109/TETC.2017.2669404.
- [10] Stanley Wasserman y Katherine Faust. *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994. DOI: 10.1017/CB09780511815478.