



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

REDES COMPLEJAS: DINÁMICA DEL COMPORTAMIENTO
EXPLOSIVO EN TENDENCIAS DE TWITTER.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO APLICADO

P R E S E N T A :

ERICK MUÑIZ MORALES

TUTORA

DRA. BIBIANA OBREGÓN QUINTANA



CIUDAD UNIVERSITARIA, CDMX, 2022

Índice general

Introducción

La comunicación es un fenómeno complejo. Esta nace de la necesidad del ser humano por el intercambio efectivo de ideas. Con el avance de la tecnología, se ha ido reduciendo el costo de la mensajería de un contacto lejano, mejorando la calidad de transmisión del mensaje y reduciendo el tiempo de respuesta entre el comunicante emisor y el comunicante receptor [?]. Las *redes de medios sociales* (RMS) son el resultado más reciente de comunicación en medios no tangibles. Sin embargo, el interés político actual de las mismas puede manipular el fin de la comunicación hacia una beneficencia meramente privada o hacia un mérito no loable [?]. Hoy en día, las RMS han modificado a la sociedad: Ejemplo de ello, el portal LinkedIn ha mostrado una nueva forma de generar y crear líderes para el avance de los negocios¹. De la misma manera, la sociedad ha podido modificar leyes internas de las RMS : Twitter notifica a sus usuarios sobre las medidas y restricciones realizadas a través de su blog de su compañía².

El ser humano, por su propia naturaleza, toma sus decisiones de forma interna donde lista cada una de sus actividades por orden de prioridad. Esto último, siendo consecuencia de un intento eficiente de asignación de recursos [?]. Por lo tanto, no es algo inesperado encontrar diversos patrones y eventos con comportamientos similares a la naturaleza de la comunicación humana. Uno de estos es aquel donde entre periodos largos de tiempo sin interacción, existe un pequeño lapso de tiempo con una alta cantidad de interacciones. A este tipo particular de comportamiento se denomina *comportamiento explosivo* [?]. Un ejemplo de ello lo podemos ver en la Figura ??, donde la imagen de lado izquierdo muestra una alta interacción en un lapso inferior a una hora.

Una modelación sencilla de este fenómeno puede verse al considerar dos estados por cada individuo del sistema: Activo e inactivo. Para ello, la población puede fluctuar en una única dirección, es decir, pasar del estado inactivo al activo. Un simple sistema de ecuaciones diferenciales lineales como el siguiente puede representar el fenómeno:

¹Osman, M. (2021, July 20). *Mind-Blowing LinkedIn Statistics and Facts* (2022). Extraído el 10 de enero de 2022 de Kinsta. Sitio web <https://kinsta.com/blog/linkedin-statistics/>

²Blog oficial de Twitter.

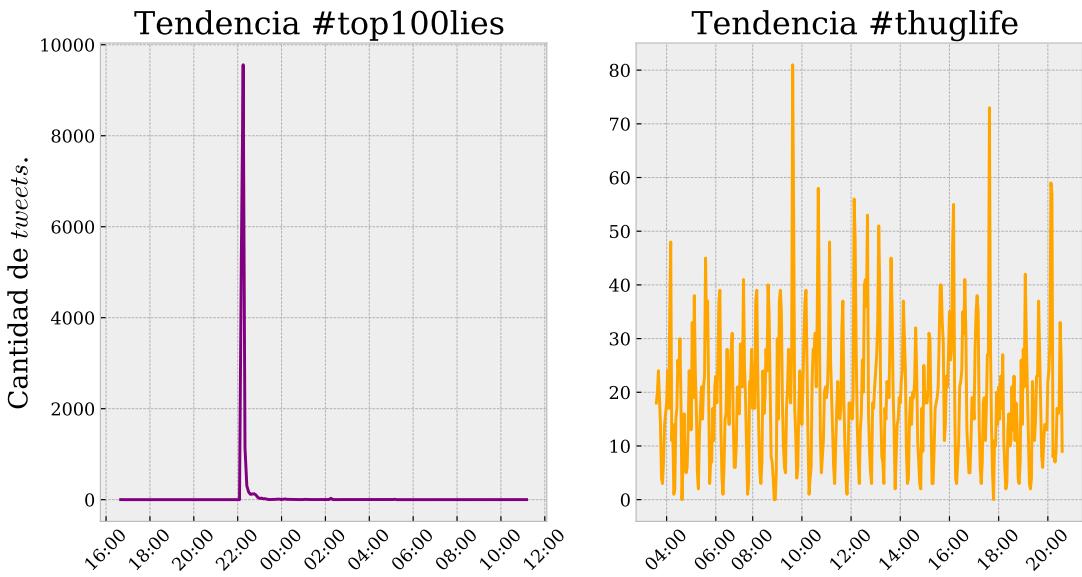


Figura 1: Series de tiempo de dos tendencias a través del tiempo. En el eje de las ordenadas, se contabiliza la cantidad de interacciones que ocurrieron en el periodo de tiempo marcado por el eje de abscisas. La imagen del lado izquierdo es de comportamiento explosivo debido a la alta interacción en un periodo de tiempo muy corto, mientras que la imagen del lado derecho no lo es debido a su parcial y constante interacción a través del tiempo. Fuente: Elaboración propia.

$$\dot{x} = -\alpha xy,$$

$$\dot{y} = \alpha xy,$$

donde x puede representar nuestra población activa, y la población no activa y α una tasa de cambio de un estado a otro. La interacción entre las poblaciones está dada por el producto de x con y . El significado del producto implica una interacción homogénea entre los individuos y es constante durante el tiempo de análisis. Esta interacción, si bien es simplista en su interpretación, se vuelve imprecisa para detallar interacciones con comportamientos explícitos entre los individuos del sistema. Como consecuencia de esta modelación, se restringe el estudio de la dinámica del sistema considerando el posible impacto de la interacción social de los mismos usuarios.

Por lo que, el **objetivo principal** es: Caracterizar el comportamiento explosivo utilizando únicamente la topología del sistema. Para alcanzar el objetivo, se seguirá una metodología basada en redes complejas y analizando una base de tendencias de Twitter que refle-

jan el comportamiento explosivo. Parte de la bibliografía actual, dedicada a este particular problema, se enfoca en la dinámica de propagación de tipo cascada siguiendo una cadena de interacciones (véase la sección de antecedentes para un mayor detalle).

Si bien tenemos que la bibliografía muestra que existe un interés para pronosticar cuando una tendencia pueda exhibir un comportamiento explosivo [??], no se enfocan en la interacción y organización de los usuarios partícipes. Un interés principal en esta problemática parte desde el interés social [?], hasta el económico y el político [?].

Este trabajo está organizado como sigue:

Capítulo 1. Antecedentes. En este capítulo se mostrará un breve panorama de las investigaciones previas en torno al comportamiento explosivo en la dinámica de poblaciones.

Capítulo 2. Marco teórico. Se mostrarán los fundamentos teóricos utilizados en este trabajo. Se comienza con teoría de redes y termina con una síntesis de la entropía de Shannon.

Capítulo 3. Metodología. Se mostrará el procedimiento de la generación de las redes usadas en el trabajo. Asimismo, se presentan los supuestos principales que llevaron a esta modelación.

Capítulo 4. Análisis y resultados. En este capítulo, se explican los resultados de cada uno de los modelos de redes presentados en el capítulo de metodología.

1 Antecedentes

En este capítulo se mostrará un breve panorama de las investigaciones previas en torno al comportamiento explosivo en la dinámica de poblaciones.

Las secciones de este capítulo son la dinámica de poblaciones, ecuaciones diferenciales y redes en sistemas dinámicos, redes de media social sobre comportamiento de comunidades y, por último, el comportamiento explosivo en redes de media social.

1.1. Dinámica de Poblaciones

Las especies son producto de la evolución que, en diferentes escalas, conforman las comunidades de organismos vivientes del planeta Tierra [?]. Al conjunto de varias especies de individuos se le denomina *población*. Las poblaciones pueden mostrar un rango de patrones dinámicos en su estructura; por ejemplo, cambiar el número de individuos [??].

Ahora bien, estos patrones dinámicos de población pueden ser vistos desde dos perspectivas: cuantitativa y cualitativa. Por un lado, una descripción cuantitativa involucra datos descriptivos para medir tendencias. Por otro lado, una descripción cualitativa permite encontrar relaciones de causalidad entre procesos físicos y la población.

El aspecto principal del estudio de la dinámica de poblaciones es identificar factores dependientes de la cantidad de individuos y la tasa de crecimiento de este número. En términos prácticos, el estudio de la dinámica de poblaciones se ha desarrollado desde un punto de vista demográfico [?]. Es decir, definir una serie de estados donde los individuos pueden fluctuar de un estado a otro.

Sin embargo, las relaciones entre los estados donde fluctúan los usuarios no consideran de forma explícita las interrelaciones entre el ambiente y el individuo. En efecto, los factores exógenos pueden ser una causa de la dinámica de población [?].

Finalmente, la capacidad de entender patrones de comportamiento, a través del comportamiento de especies individuales y cómo estos se relacionan en su ambiente, permitiría predecir los cambios la población.

1.1.1. Ecuaciones Diferenciales y Redes en Sistemas Dinámicos

La bibliografía muestra que los primeros modelos sobre la dinámica de poblaciones utilizan sistemas de ecuaciones diferenciales [???]. Estos primeros modelos se basan en el mismo principio de la fluctuación de personas a través de estados.

El uso de ecuaciones diferenciales permite aplicar esta metodología de fluctuación de estados en diferentes campos de investigación. Por ejemplo, en el modelo de epidemias SIR, los estados son susceptible, infectado y recuperado [?], donde una persona puede transitar de un estado susceptible a infectado o de infectado a recuperado. En otro ejemplo, basado en migración, los estados pueden representar localidades, estados o naciones [???], donde las personas pueden transitar al cambiar de ubicación.

Varias de estas modelaciones generan patrones sobre el comportamiento de la población al modificar los parámetros del sistema [?]. Sin embargo, los patrones no siempre resultan cuando el modelo se aplica a un sistema de otra población con características similares. Esto último en consecuencia de la omisión de varios supuestos; en particular, no todas las comunidades, ni los individuos pertenecientes a ellas, tienen el mismo comportamiento [?].

No obstante, la implementación de la teoría de redes complejas permite solucionar estos problemas. Esto es debido a que las conexiones de cada población son explícitas en la red definida, mientras que la red que se genera proporciona información sobre la dinámica epidemiológica [?].

Los modelos basados en redes engloban varios enfoques sobre la topología de la red. Por ejemplo, se presentan modelos teóricos que abarcan coeficientes de agrupamiento y modularidad necesarios para obtener una transmisión de información con menor costo de relaciones [??].

1.1.2. Redes de Media Social sobre Comportamiento Explosivo

La relación entre los modelos citados anteriormente y la dinámica de las redes de media social (RMS), es considerar dos estados: usuario activo y usuario inactivo. Es decir, un usuario inactivo se convierte en activo cuando realiza una interacción.

Esta idea ha llevado a algunos a implementar diversos análisis generales. Por un lado, se desarrollaron análisis estadísticos en función de la cantidad de interacciones específicas de un cierto tópico [?]. Por otro lado, usando sistemas de ecuaciones diferenciales, se ha implementado un modelo para ajustar la serie de tiempo sobre el comportamiento de las interacciones activas e inactivas [?].

Los modelos de redes aplicados en RMS resultan ser más efectivos para estudiar y ana-

lizar registros precisos de flujos de información [?]. De hecho, los registros y los modelos de redes llevaron a investigaciones sobre el comportamiento de comunidades [??], enfocados en la estructura de la red generada por los usuarios activos [?].

1.2. Comportamiento Explosivo en Redes de Media Social.

Un comportamiento en RMS interesante es aquel que, sin una causa aparente, adquiere un alto volumen de actividad en poco tiempo y, de forma gradual, su intensidad se hace nula. A este comportamiento se le denomina comportamiento explosivo.

El comportamiento explosivo es un fenómeno existente en diversos campos de la ciencia, cabe mencionar, es contrario a modelaciones de procesos de Poisson [?]. En estudios de ciencias sociales, se han encontrado dos posibles causas de este tipo de comportamiento que son el limitado tiempo de realizar tareas cotidianas [?] y la priorización de las mismas [?].

Ahora bien, las redes de media social no son la excepción ante este comportamiento y, mucho menos, al estudio del mismo. Se ha mostrado que el comportamiento explosivo puede provenir de un ataque coordinado entre varios iniciadores de distintas comunidades [???]. Sin embargo, estas últimas modelaciones se basan en una construcción local sobre los primeros usuarios relacionados con usuarios activos.

Aunado a esto último, también son necesarias interacciones de menor esfuerzo para que el comportamiento sea explosivo [??].

2 Marco Teórico

En este capítulo, se mostrarán los fundamentos teóricos utilizados en este trabajo. Se comienza con teoría de redes y termina con una síntesis de la entropía de Shannon.

2.1. Teoría de Redes.

El problema de los puentes de Königsberg resulta ser una de las primeras motivaciones en la teoría de gráficas. La ciudad de Königsberg fue fundada por los caballeros teutónicos en el año 1255 y, debido a su asentamiento sobre el río Pregolia, se convirtió en un importante centro de intercambio. Por ello, la ciudad fue dividida en cuatro regiones conectadas por siete puentes (ver figura ??).

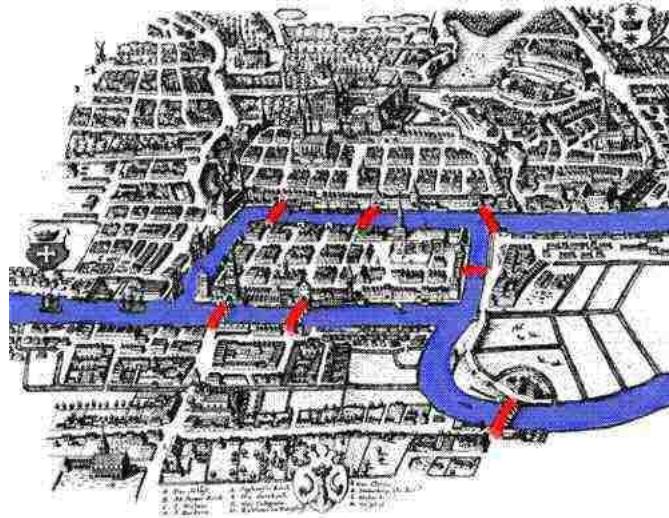


Figura 2.1: Una vista de Königsberg mostrando en rojo los siete puentes sobre el río Pregolia.

El problema de los siete puentes de Königsberg que conectan las cuatro regiones intenta resolver la siguiente pregunta: ¿es posible dar un paseo por todas las regiones, comenzando desde cualquiera de estas regiones, pasando por todos los puentes una sola vez y regresando al mismo punto de partida?

Fue hasta el año 1736 cuando se hace público el artículo de Leonhard Euler dando una solución particular y general a este problema. Lo más esencial de este texto de 21 párrafos es la simplicidad de la solución. Euler propone enfocarnos en las conexiones generadas por los puentes; explícitamente, identificar a cada segmento de tierra con una letra mayúscula y cada puente con una letra minúscula (ver figura ??). De la representación de Euler, se obtuvieron resultados generales a través de varios experimentos.

Es curioso mencionar que debido a la sencillez del planteamiento del problema de los puentes, Euler no creía necesaria la intervención de una visión matemática [?], sin embargo, la abstracción que realizó dio inicio una rama más de la Matemática: Teoría de Gráficas.

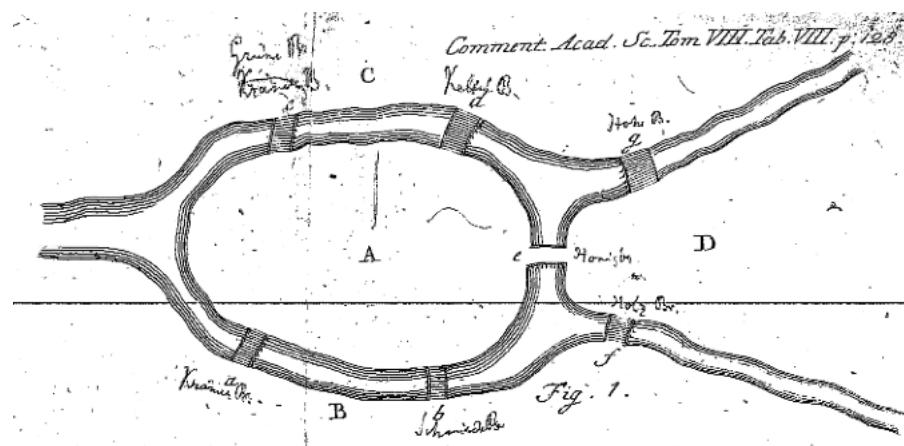


Figura 2.2: Esbozo original de Euler sobre el problema de los puentes de Königsberg.

Una abstracción sencilla para un pensamiento actual es extrapolar los puentes hacia las relaciones humanas, algo innovador para el pensamiento del siglo XIX. Jacobo Levy Moreno, influenciado por diversas culturas y diferencias sociales, consideró la afinidad y preferencia de la gente por relacionarse ante distintas adversidades [?]. Esta observación dio origen a la sociometría que fue impulsada por los sociogramas: representaciones gráficas de relaciones sociales.

Al seguir la línea del análisis del comportamiento social, en 1967, Stanley Milgram introdujo una pregunta relacionada con situaciones cotidianas, asociada al fenómeno de *mundo pequeño*: el efecto que cualesquiera dos personas en el mundo están a pocas personas de conocerse. La pregunta que planteaba era sencilla: ¿cuántas conexiones entre dos personas son necesarias para que estas se conocieran? El problema y solución considerarían la estructura matemática de la sociedad. La parte experimental de este proceso concluyó que la cantidad de personas intermedias es apenas un poco más de cinco personas [?].

No fue sino hasta finales del siglo XX que comenzó una formalización de las observaciones antes vistas. En 1998, Duncan Watts y Steven Strogatz, a través de un modelo que

itera probabilísticamente las aristas de una red regular, encontraron una dinámica interesante en función de la estructura topológica de la red [?]. En específico, se mostró un umbral donde la red tiene un coeficiente de agrupamiento alto pero con una corta distancia media entre los nodos. Si bien esta conclusión de la propiedad de mundo pequeño nunca fue definitiva, motivó una formalización e investigación del fenómeno; aunque, muchas investigaciones son en función de lo mostrado por Watts y Strogatz [?].

Poco después del trabajo de Watts y Strogatz, la investigación de Barabási dentro del mundo de los sistemas complejos fomentó aun más la investigación en redes. En 1999, poco después de la publicación de Watts y Strogatz, Albert-László Barabási y Réka Albert analizaron la red de citas de páginas de Internet. El aspecto más relevante del trabajo de citas de portales de Internet era inducir una nueva dinámica en redes desde un paradigma distinto al aleatorio: Las *redes de libre escala* [?]. Este tipo de red presenta una propiedad interesante en su distribución de grado, pues se distribuye como una ley de potencia; de manera formal, si $f(x)$ es la distribución de grado, entonces f es ley de potencia si $f(x) \sim x^{-\alpha}$, con $\alpha > 0$. Esta caracterización de la distribución de grado induce a que la red en cuestión contenga *hubs* (nodos con un grado muy alto). Esto último consecuencia de dos propiedades relevantes: el continuo nacimiento de nuevos enlaces y la preferencia de relacionarse a un nodo en particular [?].

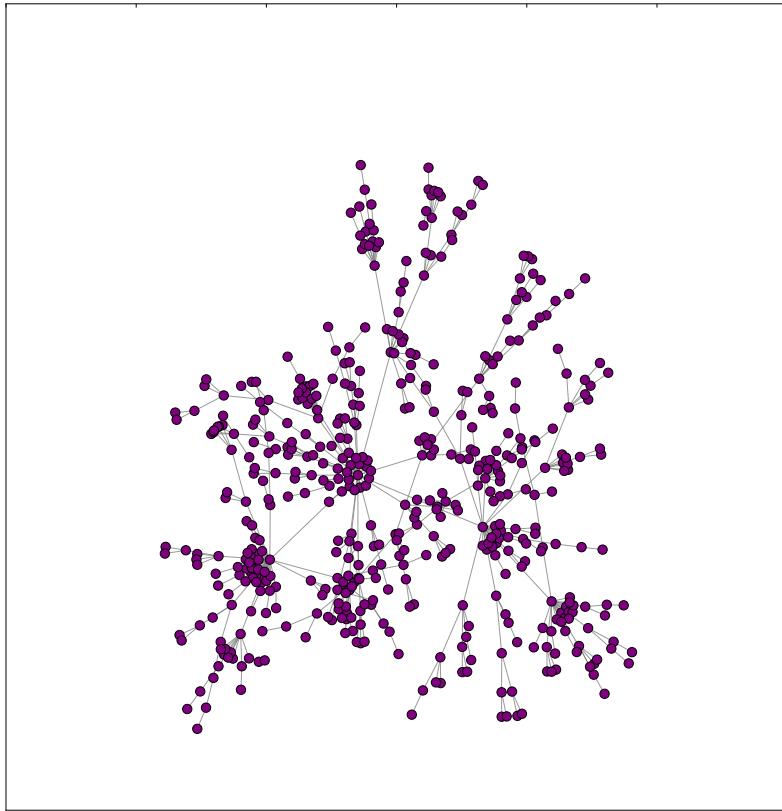


Figura 2.3: Ejemplo de una red de libre escala. Fuente: Elaboración propia.

2.1.1. Definiciones de Teoría de Redes

Una *gráfica* es una estructura matemática compuesta de dos partes: los objetos y las relaciones que se tienen entre ellos. Es decir, una red es una colección de puntos unidos de dos en dos por líneas [?]. Así, una *red* es una gráfica con alguna interpretación. Los objetos y las relaciones son nombrados de diferentes maneras dependiendo del contexto y la interpretación de la red del sistema (véase la tabla ?? para algunos ejemplos y su interpretación, para más información ver [?]).

En general, los objetos pueden ser distinguidos como *nodos* y las relaciones como *aristas*.

Tabla 2.1: Comparativo de redes respecto a los nombres de los nodos y aristas. (Fuente: [?])

Red	Nodos	Enlaces	Tipo de red
Internet	Routers	Conexión de internet	No dirigida
World Wide Web	Páginas web	Enlaces	Dirigida
Red eléctrica	Centrales eléctricas	Cables	No dirigida
Red de actores	Actores	Coactuación	No dirigida

De manera formal, una *red* es el par ordenado de dos conjuntos

$$G = (V, E), \quad (2.1)$$

donde V y E son el conjunto de nodos y aristas, respectivamente.

Una propiedad que podría definir el tipo de red es si la arista tiene o no dirección. Una *arista dirigida* es aquella donde el orden de los nodos es importante; de forma contraria, una *arista no dirigida* es aquella donde el orden de los nodos es irrelevante.

Lo anterior conlleva a que una *red dirigida* tiene todas sus aristas dirigidas. De forma contraria, una *red no dirigida* es aquella donde todas sus aristas son no dirigidas.

Es decir, una red dirigida cumple con

$$(a, b) \neq (b, a), \quad \text{con } a, b \in V,$$

Y, por otra parte, una red no dirigida cumple con

$$(a, b) = (b, a), \quad \text{con } a, b \in V,$$

El *grado de un nodo* es el número de aristas que tiene el nodo con los demás nodos [?]. Este valor es denotado por k_v donde v es un nodo.

Para una red dirigida, el grado de un nodo se extiende a dos casos: grado interior y grado exterior. El *grado interior* de un nodo es el número de aristas donde el nodo en cuestión es el nodo final. El *grado exterior* de un nodo es el número de aristas donde el nodo en cuestión es el nodo inicial. De forma explícita, consideremos una arista,

$$\begin{aligned} (s, t) &\in E, \quad \text{donde } s, t \in V, \\ s &\text{ es el nodo inicial,} \\ t &\text{ es el nodo final.} \end{aligned}$$

El grado exterior de un nodo v es denotado por k_v^{out} y el grado interior de un nodo es denotado por k_v^{in} .

Así, de manera formal, el grado exterior e interior de un nodo se definen como:

$$k_v^{out} = |\{(x, v) \in E \mid x \in V\}|,$$

$$k_v^{in} = |\{(v, x) \in E \mid x \in V\}|,$$

donde $|A|$ representa la cardinalidad del conjunto A .

Cabe mencionar que para el caso de una red no dirigida, donde el orden las aristas no es relevante, el grado de un nodo es igual a cualquiera de los casos anteriores, es decir,

$$k_v = k_v^{out} = k_v^{in},$$

Por otra parte, la matriz de adyacencia de una red es una matriz donde las entradas de la misma representan los enlaces entre los nodos; las columnas y renglones son la representación de los nodos. De manera formal, para cualesquiera dos nodos $i, j \in V$, tenemos que:

$$(A)_{i,j} = \begin{cases} 1 & \text{si } (i, j) \in E, \\ 0 & \text{En caso contrario.} \end{cases}$$

Esta matriz es útil para otras métricas de interés que requieren conocer las relaciones entre los nodos.

Cabe mencionar que para una red no dirigida se cumple que

$$A = A^T,$$

mientras que para un red no dirigida, en general, se cumple que

$$A \neq A^T.$$

Por otro lado, una de las propiedades principales para el cálculo de métricas de centralidad de nodos es la distancia entre ellos.

Para definir la distancia entre nodos, es necesario definir los caminos entre los nodos.

Un *camino* es una sucesión ordenada y alternada de nodos y aristas; esta sucesión puede repetir tanto nodos como aristas. Cuando el camino no repite nodos, se dice que es una *ruta*. Cuando el camino no repite enlaces, se dice que es una *trayectoria* [?].

La *longitud* de un camino es la cantidad de aristas que tiene el camino. Así, la *distancia entre nodos* es la longitud de su camino. En particular, un camino relevante es el *camino más corto*: El camino de longitud mínima; e incluso, puede existir más de un camino.

El valor de esta longitud puede depender del tipo de enlace que se defina (ver [?] para mayor información).

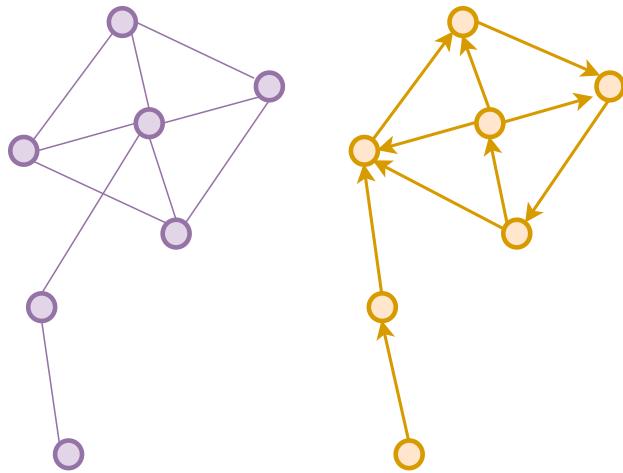


Figura 2.4: Ejemplos de una red dirigida y no dirigida. De lado izquierdo tenemos una red no dirigida. De lado derecho, una red dirigida.

2.1.2. Red Multicapa y Red Temporal

Una característica compartida entre una red dirigida y una no dirigida es la *propiedad estática* de los nodos y enlaces: La definición de los nodos y enlaces no es mutable [?]. En consecuencia, este tipo de modelación tiene una limitación ante cambios del sistema.

Indexar los nodos y aristas con una secuencia definida permite recuperar la mutabilidad del sistema. Cada uno de estos índices se define para representar los múltiples niveles o cambios del sistema. Una *capa* (L_t) es una tupla de conjuntos de nodos y aristas

$$L_t = G_t(V_t, E_t),$$

donde $E_t \subseteq E$ y $V_t \subseteq V$,

donde t es un índice para la capa L_t . Cabe mencionar que una capa puede ser vista como una sola red. Así, una *red multicapa* es la sucesión de capas $\{L_t\}_1^n$, donde n es el número de capas.

Un caso particular de las redes de multicapa es indexar el parámetro t a un intervalo de tiempo. Este tipo de red es una red temporal.

Una *red temporal* es una red multicapa donde el índice de las capas está asociado a un intervalo de tiempo. Este tipo de red permite una propuesta para describir la mutabilidad de una red a través del tiempo. Es decir, esta red captura los cambios entre los nodos y enlaces en períodos de tiempo definidos, a través de redes más simples. En la Figura ?? se presenta un ejemplo de una red en distintos períodos de tiempo.

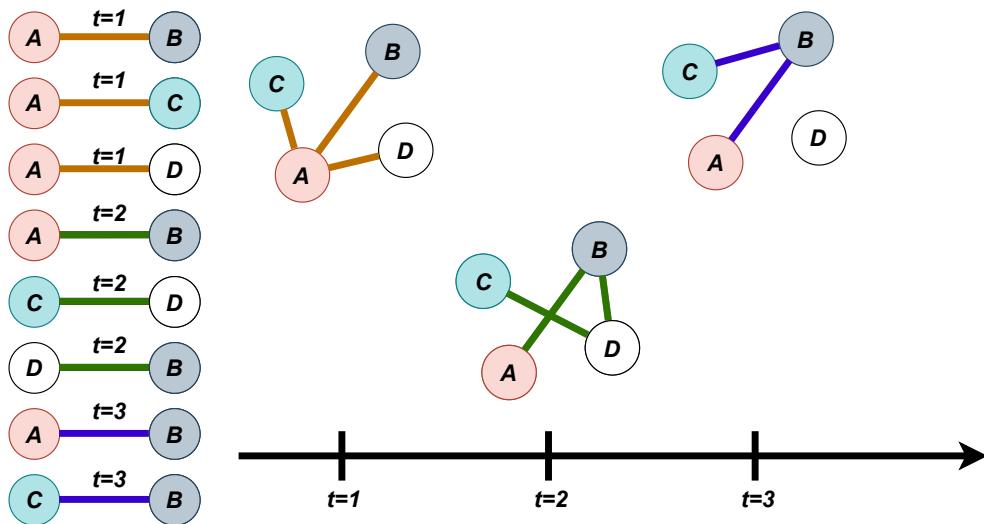


Figura 2.5: Representación de una red temporal. Se observa como el enlace (A, B) se mantuvo en cada periodo de tiempo.

La extensión y definición de la red multicapa es extensa; puesto que permite añadir enlaces entre capas, entre nodos, etcétera. Este trabajo se basó en la red multicapa para definir la Red de la Dinámica Total (véase la sección 3.1).

Para mayor información sobre la red multicapa, se puede consultar [?]

2.1.3. Métricas

En esta sección, se presentan diversas métricas aplicadas en esta tesis.

Coeficiente de agrupamiento local (*clustering*)

Uno de los primeros patrones que hay en las redes es la aparición de *ciclos*. Un ciclo es una *trayectoria cerrada*; es decir, el nodo de inicio del camino es el mismo que el nodo final.

El interés en la métrica nace por la medición de agrupación de nodos por sus enlaces [?]. El ciclo de longitud tres es de particular importancia en el estudio de redes sociales, debido a que permite cuantificar la densidad local; es decir, la conexión de un nodo con los vecinos más cercanos [?].

El coeficiente de agrupamiento es el conteo de aquellos triángulos existentes en la red sobre los triángulos que puedan existir.

De manera formal, se fija a un nodo v . Se define t_v como el número de triángulos donde el nodo v es parte de éste y d_v es el grado del nodo. Entonces, el coeficiente de agrupamiento (C_v) es:

$$C_v = \frac{t_v}{\frac{d_v(d_v-1)}{2}} = \frac{2t_v}{d_v(d_v-1)}. \quad (2.2)$$

El denominador de la expresión después del primer símbolo igual es el número de triángulos posibles entre sus vecinos. De la propia definición de la ecuación ??, esta es no negativa y el valor máximo es uno.

Valor k – núcleo (k -core)

La obtención de este valor se basa en la agrupación de nodos por su grado.

En palabras generales, este valor se obtiene al considerar el valor de k más grande en los k – core donde un nodo pertenece. Un k – core de una gráfica es una subgráfica donde todos los nodos tienen un grado mayor o igual que k .

De manera formal, el k -core D_i está definido como

$$D_i = \{v \in V | k_v \geq i\}, \quad (2.3)$$

Entonces, para un nodo $u \in V$, tenemos que su valor k -core es

$$k - core(u) = \max \{k \mid v \in D_k\} \quad (2.4)$$

Como una consecuencia de la definición, se puede ver a $V = \bigcup_{i=1}^k D_i$. Esto último permite una visualización de la red en forma de capas con en la Figura ??.

Métricas de Centralidad

El uso de las métricas de centralidad, para el enfoque social, tiene como objetivo identificar los nodos más importantes del sistema.

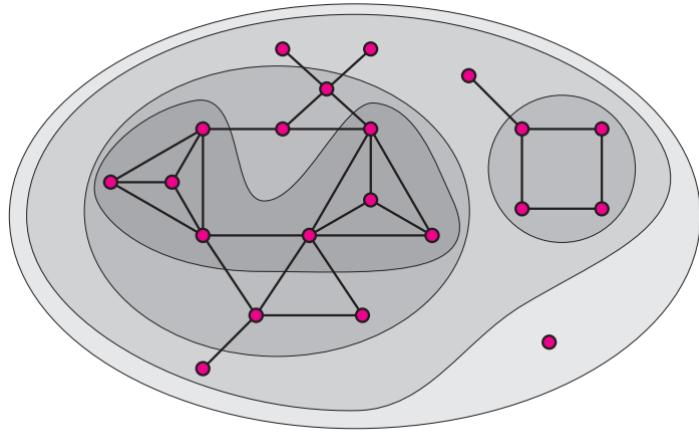


Figura 2.6: Ejemplo visual de la aplicación de la métrica k – núcleo. (Fuente: [?])

Centralidad intermedia (*betweenness*)

Es de suma importancia considerar los múltiples caminos que puede existir entre dos nodos. En particular, aquellos caminos donde los nodos partícipes juegan un papel relevante en la red: los caminos más cortos.

De manera formal, esta métrica se define como:

$$B(v) = \sum_{v \neq j, i \in V} \frac{\sigma(i, j|v)}{\sigma(i, j)}, \quad (2.5)$$

donde $v \in V$, $\sigma(i, j)$ representa el número de caminos más cortos entre los nodos i y j ; y $\sigma(i, j|v)$ representa el número de caminos más cortos entre los nodos entre i y j que pasan por el nodo v .

La interpretación de esta métrica varía del objeto de estudio en cuestión. Sin embargo, en general, su interpretación denota la importancia de algún nodo en la comunicación con los demás nodos en la red [?].

Grado de cercanía (*closeness*)

Es el coeficiente de cercanía de un nodo hacia los demás nodos. Esta métrica pondera la suma de las distancias de un nodo fijo contra los demás nodos de la red y de este valor se obtiene su recíproco. Así, un valor alto en la suma de las distancias simboliza un grado bajo y, entonces, una importancia menor.

De manera formal, si $d(u, v)$ representa la longitud del camino más corto entre u y v : y n es la cardinalidad del conjunto V . Entonces, el grado de cercanía de un nodo u es

$$close(u) = \left(\frac{\sum_{v \neq u} d(u, v)}{n - 1} \right)^{-1}. \quad (2.6)$$

Centralidad de vector propio (*eigenvector*)

La noción de esta centralidad es medir la importancia de un nodo por la importancia de sus nodos vecinos.

El cálculo de la misma es resolver una operación matricial:

$$Ax = \lambda x, \quad (2.7)$$

donde A es la *matriz de adyacencia* de la red y $\lambda > 0$ es el valor propio más grande de la matriz de adyacencia.

Índice de Katz

Una problemática con la expresión en la ecuación (??) es que sólo involucra a los vecinos en la primera *vecindad* de cada nodo. La n -ésima vecindad de un nodo se define como el conjunto de nodos donde hay un camino de longitud n .

El índice de Katz extiende la noción de la influencia de nodos que pertenecen más allá de la primera vecindad. En efecto, el cálculo de la misma pondera su importancia a partir de todos los posibles caminos de longitud k entre los nodos.

$$Katz(i) = \left[\sum_{k=0}^{\infty} (\alpha A)^k e \right]_i, \quad (2.8)$$

donde e es el vector que en todas las entradas tiene el valor 1. La implementación de considerar todos los caminos de longitud k está llevada a cabo por el producto matricial A^k . Nótese que, como en la métrica de vector propio, la ponderación es lineal debido a la definición de producto matricial.

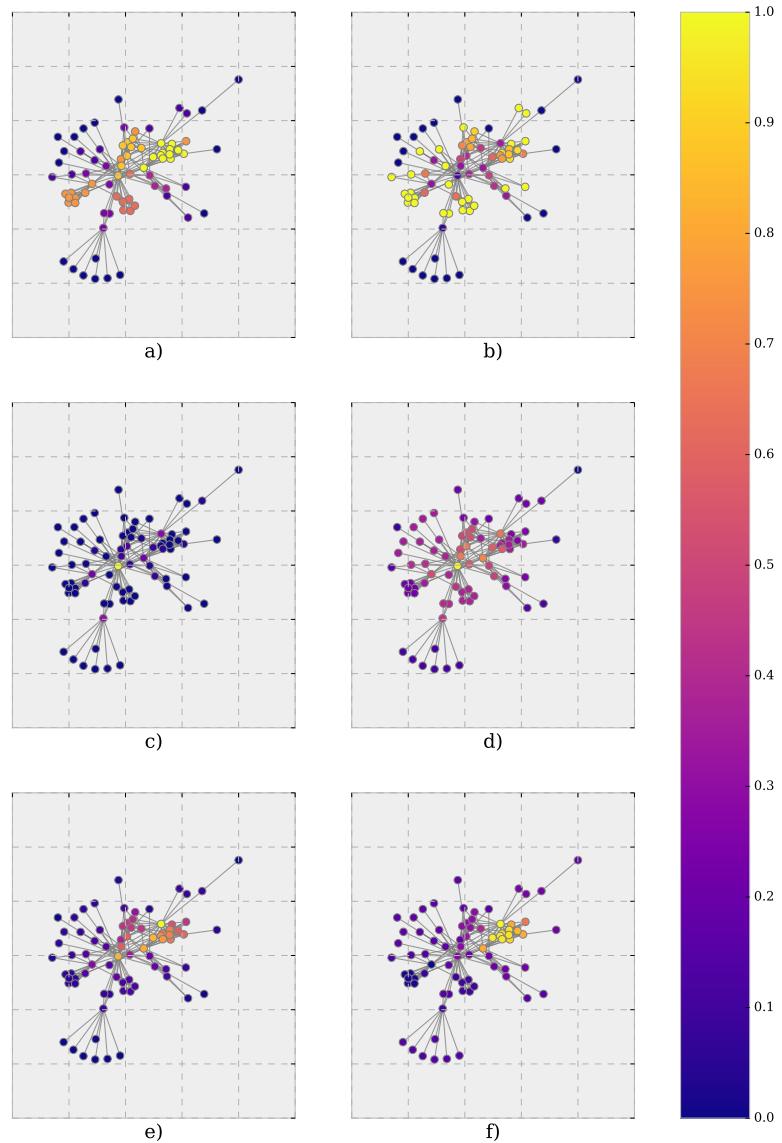


Figura 2.7: Comparativo de métricas para la red del Karate Club de Zachary [?]. Las métricas que se muestran son a) k-núcleo, b) clustering, c) betweenness, d) closeness, e) eigenvector y f) centralidad de Katz. Fuente: Elaboración propia.

2.2. Sistemas Complejos.

Manlio de Domenico y Hiroki Sayama, en Complexity Explained [?], mencionan que los sistemas complejos pueden espontáneamente auto-organizarse y presentar estructuras globales y comportamientos no-triviales a mayores escalas, sin intervención externa, autoridad central o líderes que determinen el comportamiento colectivo [?].

Cabe mencionar que no es posible hablar de una definición formal de los sistemas complejos, sin embargo, sí es posible identificar ciertas características, que a continuación se explican [?].

1. Sistémica y sinérgica. La propiedad sistémica indica que es necesario explorar las relaciones o interconexiones de un sistema a fin de buscar patrones en las conexiones, sin embargo, dichas relaciones pueden ser lineales. En cambio, si las interacciones son no lineales entre las partes de un sistema, entonces se habla de la propiedad sinérgica.

2. Multi escala. Esta propiedad indica que para estudiar o entender un sistema es necesario considerar las diferentes escalas, incluyendo el tiempo, que forman parte del sistema. Además, las características que se presentan en una escala no deben asumirse en las demás, es decir, el estudio de una escala no permite entender el comportamiento de las demás escalas, lo cual indica que no debe considerarse *auto similaridad*.

3. Resiliencia y adaptabilidad

La resiliencia de un sistema indica la capacidad que dicho sistema tiene para responder a variaciones internas y externas. Para el estudio de la resiliencia y la adaptación es necesario precisar dos conceptos: *variación*, que hace referencia a un cambio discreto a nivel individual; y, *fluctuación*, que hace referencia a un cambio en el fenómeno, como por ejemplo la densidad poblacional. Ahora bien, cuando las variaciones son entre los agentes de una misma especie de un sistema se conocen como *microdiversidad*; mientras que variaciones de especies entre los agentes son conocidas como *macrodiversidad*.

4. Incertidumbre. El futuro de un sistema es incierto, es decir, la secuencia o el orden de los eventos y el contexto en que se presentan son necesarios para determinar qué podría ocurrir.

5. Autorregulación, auto-organización y emergencia.

Estas propiedades están relacionadas con el estado del sistema y cómo cambios, tanto a pequeña como a gran escala, afectan el estado del sistema final.

La propiedad de autorregulación permite al sistema fluctuar en un umbral de estabilidad; es decir, oscilar algunos valores cuantitativos sin cambiar estados cualitativos. La propiedad de auto-organización permite al sistema modificar su estructura con base en interacciones locales.

La propiedad de emergencia evoca un cambio radical y cualitativamente diferente partiendo del estado inicial. Este permite al sistema cambiar en su estructura permitiendo una nueva ramificación de estados y, por tanto, forzando la aparición de nuevos patrones de comportamiento. Las tres propiedades se basan en el mismo proceso, variaciones microscópicas que permiten cambios macroscópicos no lineales.

2.2.1. Entropía

La entropía es una medida formal de incertidumbre. El valor obtenido permite determinar tanto el nivel de incertidumbre como la información mostrada. Es decir, los sistemas que presentan un valor de entropía baja, indican que tienen baja incertidumbre y proporcionan poca información, por el contrario, cuando el sistema tiene un valor alto, indica mayor incertidumbre y mayor información revelada [?].

Entropía de Shannon

Claude Shannon publicó en 1948 un artículo titulado *A Mathematical Theory of Communication*, con el cual transformó el entendimiento de la información, al hacerla una cantidad bien definida y medible [?].

La teoría de la información de Shannon proporciona una definición matemática de información, además, indica cuánta información se obtiene entre las diferentes partes o componentes de un sistema.

Entonces, la entropía se define como un promedio de la incertidumbre del verdadero estado de un fenómeno. Esto es, para una variable aleatoria X , con función de distribución de probabilidad $f(x)$, la entropía de Shannon se define como:

$$H(X) = - \sum_x f(x) \log_2(p(x)). \quad (2.9)$$

El valor resultante, para el caso de \log_2 , se llama *bits* de información.

La interpretación de esta métrica es sencilla, por ejemplo: ¿Cuántas preguntas con respuestas binarias se necesitan para saber el estado actual del sistema? Pues, entre mayor número de preguntas se necesiten mayor es la incertidumbre. Por el contrario, cuando X tiene un sesgo para ciertos valores, la entropía es baja; ya que, el valor de X es más probable que esté en la parte sesgada.

Cabe mencionar que, se puede cambiar la base del logaritmo en esta métrica, obteniendo posibles resultados distintos (véase [?] para algunas propiedades). En este trabajo, por convención, únicamente se usará la entropía de Shannon con logaritmo en base 2.

3 Metodología

En esta sección se muestra el procedimiento de la generación de las redes usadas en el trabajo. Asimismo, se presentan los supuestos principales que llevaron a la modelación utilizada.

Se presentan dos tipos de redes: por un lado, se modela la dinámica total de los usuarios comunicantes en todo el periodo de tiempo observado; por otro lado, se modela la dinámica de los usuarios antes de la hora pico o el momento de explosividad.

Tabla 3.1: Las 10 tendencias más importantes respecto a la cantidad de *tweets* recabados.

Tendencia	Cantidad de tweets
#oomf	363,518
#np	158,608
#nf	128,920
#ff	123,875
#teamfollowback	83,064
#bahrain	59,277
#rt	51,735
#thoughtsduringschool	50,927
#yolo	50,256
#dearoomf	40,894
#retweet	39,097
#lrt	35,357
#nw	35,323
#taurus	35,309
#lt	32,226

Cabe mencionar que la modelación se hace con el fin de analizar un comportamiento específico en la dinámica de comunicación de los usuarios en función de la red social. Por lo tanto, el tema inducido por cada nombre de tendencia no se consideró para este trabajo y tampoco se consideró la dependencia entre ellas. Por ejemplo, se tienen tendencias muy relacionadas que registraron una buena actividad de los usuarios; como las tendencias #ff (acrónimo en inglés de *Follow Friday*) y #teamfollowback que refieren a una invitación a

seguirse entre los usuarios que participen en la dinámica (ver tabla ??).

3.1. Red de la Dinámica Total

Es una red temporal multicapa (*snapshot*), con dos capas, indexadas por tiempo para cada tendencia. Cada capa estudia un tipo de actividad realizada por los usuarios; en específico, una por *tweets* y otra por *retweets*.

A esta red de multicapa se nombró como $NC_t^\alpha(h)$ donde h representa el nombre de la tendencia, t es un periodo de tiempo y $\alpha \in \{T, R\}$ es un índice para la capa de *tweets* y *retweets* respectivamente.

Para la explicación de la construcción de cada una de las capas, se fijó un periodo de tiempo t . Las capas se construyeron de la siguiente forma:

- $NC_t^T(h)$ es una red no dirigida que representó la dinámica de los *tweets* en la red social de los primeros adoptadores o usuarios.

En específico, el usuario A realizó un *tweet* en el periodo de tiempo t . Se tiene que A y B se seguían mutuamente. Entonces, se definió el enlace (A, B) . Cabe recalcar que bastaba con que uno de los dos haya hecho una interacción (*tweet*) para que se generara este mismo enlace en la red.

- $NC_t^R(h)$ es una red dirigida que representó la dinámica de los *retweets* en la red social de los primeros usuarios.

Es decir, el usuario A realizó un *tweet* y un usuario B realizó un *retweet* al *tweet* de A en el periodo de tiempo t . Entonces, el enlace (A, B) pertenecía en la red.

En la figura ?? se observa una pequeña representación de este modelo. Se hizo referencia a $NC(h) = \cup_t NC_t(h)$ como la dinámica total de la tendencia h .

Cabe recalcar que esta red multicapa estaba indexada por tiempo. Es decir, se discretizó la serie de tiempo en intervalos fijos de tiempo. Para este trabajo, la discretización se realizó por intervalos de tiempo de una hora. Esta elección, principalmente, se fundamenta para dar un seguimiento a la cadena de comunicación generada por los usuarios. Dicha cadena no se esperaba con un comportamiento fijo entre usuarios [?]; por un lado, si el periodo de tiempo era muy corto, se perdía el seguimiento por el alto detalle de análisis; por otro lado, si este periodo era muy largo, se volvía a perder por la gran escala del mismo al no saber de donde nació el *tweet* inicial. Al tratar diversas longitudes de tiempo, se optó el intervalo de tiempo de una hora como el más significativo.

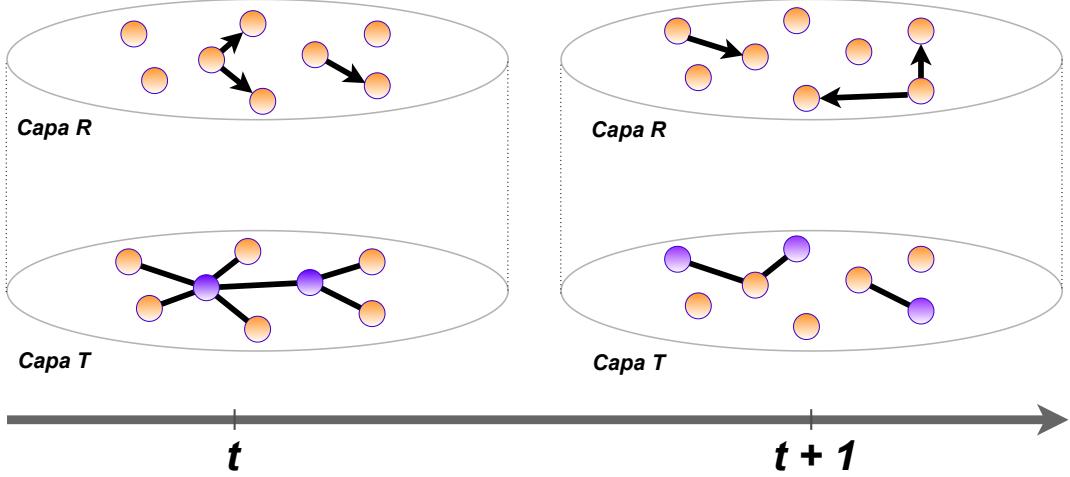


Figura 3.1: Ejemplo sencillo de la red $NC_t^\alpha(h)$ para alguna tendencia de nombre h . Los nodos en color morado representan a los primeros adoptadores de la tendencia. Nótese que no hay enlaces entre capas.

3.2. Red Pre-Explosividad.

Es una red no dirigida inducida por los usuarios que adoptaron la tendencia momentos antes del comportamiento explosivo. A esta red se definió como $\mathcal{G}(h)$ donde h , siguiendo la notación antes vista, es una tendencia.

A continuación, se presentará el procedimiento explícito de la construcción de esta red. De manera formal y para seguir la notación, sea $G(V, E)$ la red social inicial de todos los usuarios registrados. Cabe recordar que $(a, b) \in E$ si y sólo si a y b se seguían mutuamente.

El proceso que se menciona en las siguientes líneas se repetirá por cada una de las tendencias. Como en la red anterior, se discretizó la serie de tiempo de la cantidad de *tweets* en intervalos de longitud de 25 minutos. En términos sencillos, se generó la red a partir de la red inducida por los verdaderos primeros comunicantes. De manera formal, sea $\{\tau_i\}_{i=0}^n$ el conjunto de dichos intervalos. Sea τ_j el intervalo de tiempo donde ocurrió la mayor cantidad de *tweets* de la tendencia h . El conjunto de intervalos de interés es $T = \{\tau_i\}_{i=\max(j-404, 0)}^{\min(j+404, n)}$. Esto último fue con el fin de obtener un análisis máximo de 14 días. Al cambiar los índices de este último conjunto como $T = \{t_i\}_{i=0}^k$. Sea U_{t_i} el conjunto de usuarios que *tweetearon*, al menos una vez, en el periodo de tiempo t_i . Por lo que $U_T = \bigcup_{i=0}^k U_{t_i}$ es el conjunto de todos los usuarios que participaron en el intervalo de interés T .

Sea $\Gamma(X)$ la función que devolvía el conjunto de vecinos del conjunto X . Se definió la siguiente recursión $\Gamma^p(X) = \Gamma(\Gamma^{p-1}(X))$ con $\Gamma^0(X) = X$ para un p entero positivo. Se definió $\mathcal{E} = \bigcup_{p=0}^l \Gamma^p(U_T)$ donde l sería el entero máximo tal que $|\mathcal{E}| \leq 20,000$. La elección

de limitar a 20,000 nodos se hizo para que la comunidad participe de nodos representara entre el 10 y 20 por ciento de la red. Para finalizar, $\mathcal{G}(h)$ era igual a la subred inducida por los nodos \mathcal{E} .

Esta red se definió para analizar la interacción de los usuarios que verdaderamente realizaron, al menos, un *tweet* en una aproximación a la red social de la que pertenecían; con ello, ver este comportamiento desde las posibles comunidades iniciales. En resumen, se extiendió la red social generada por los usuarios que *tweetearon* hasta un límite de usuarios por cuestiones computacionales.

La principal justificación y diferencia con la red en la sección anterior es la escala. Por un lado, la red social que abarcaba a los partícipes en el intervalo de una hora donde hubo la mayor interacción (como se definió $NC_{t^*}^T(h)$ donde t^* es el intervalo de tiempo donde hubo mayor interacción) tenía la sutileza en eliminar a los verdaderos pioneros de la tendencia así como en la cantidad de veces que realizan dicha acción. Esto claramente se puede apreciar en el gráfico ???. Por otro lado, era necesario escalar la red para abstraer la idea de *usuarios comprometidos* [?] en distintas comunidades y ver en qué *capa* tenían una interacción clave. Dicha modelación tiene un argumento similar al comportamiento de los sismos para caracterizar comportamientos endógenos y exógenos ?.

3.3. Datos e Implementación.

La base de datos utilizada en este trabajo fue recabada por el Centro de Investigación de Redes y Sistemas Complejos (CNetS) [??].

Esta base cuenta con, aproximadamente, 500 MB de información. La información es un compendio de varios seguimientos del portal *Twitter* desde el 24 de marzo de 2012 al 26 de abril de 2012. Cabe recalcar que, por las políticas de privacidad de *Twitter*, todos los usuarios que aparecen tienen un identificador único y diferente a su cuenta original; por lo que no es posible hacer búsqueda de dicha cuenta a través de otras paqueterías u obtener una información sobre los usuarios u obtener explícitamente lo que se haya publicado. Hay 14,273,311 usuarios recabados con 1,345,913 de tendencias y 12,054,205 *tweets* en total.

Para esta modelación, se realizó una limpieza de toda la base. El primer filtro fue considerar sólo aquellas que tuvieran, al menos, 4000 *tweets* registrados. Así, se analizaron únicamente 143 tendencias.

Los modelos que se mencionaron se implementaron con el lenguaje de programación Python; haciendo uso de diversas paqueterías, entre ellas NetworkX para el trato y manipulación de las redes y de NetworkIt para el cálculo de diversas métricas. La implementación se encuentra alojada en el repositorio Github.

4 Análisis y Resultados

En este capítulo, se explican los resultados de cada uno de los modelos de redes presentados en el capítulo de metodología.

El análisis se enfoca en dos perspectivas: i) se analizaron las redes $NC(h)$ y se calcularon sus métricas (sección de métricas del capítulo 2); y ii), en la red de pre-explosividad $\mathcal{G}(h)$, se realizó un análisis en las comunidades obtenidas y cómo favorecieron el comportamiento explosivo.

En esta sección solamente se presentan los principales resultados obtenidos en la investigación. Para un análisis más detallado, véase el Anexo A.

4.1. Análisis preliminar de las tendencias.

La cantidad de usuarios partícipes es diferente en las tendencias con comportamiento explosivo de las tendencias sin comportamiento explosivo. En las tablas del comparativo ??, al tener una mayor cantidad de usuarios es esperable obtener una mayor cantidad de tweets; sin embargo, el hecho de tener una gran audiencia no es suficiente para caracterizar las tendencias por su propiedad de explosividad. De los datos, la figura ?? no muestra un diferenciador para las tendencias con comportamiento explosivo.

Tabla 4.1: Comparativo entre las tendencias por el número de usuarios totales que interactuaron en ella. En la tabla de la izquierda corresponden a tendencias con comportamiento explosivo (**CE**); del lado izquierdo aquellas sin comportamiento explosivo (**sCE**).

Tendencias CE	No. de usuarios	Tendencia sCE	No. de usuarios
<i>wheniwaslittle</i>	20,150	<i>fml</i>	9,467
<i>yougetmajorpointsif</i>	15,219	<i>wtf</i>	8,467
<i>top100lies</i>	13,129	<i>fail</i>	7,283
<i>nationalbestfriendday</i>	12,569	<i>20mf</i>	6,957
<i>mythoughtsduringsschool</i>	10,662	<i>ohwell</i>	6,915

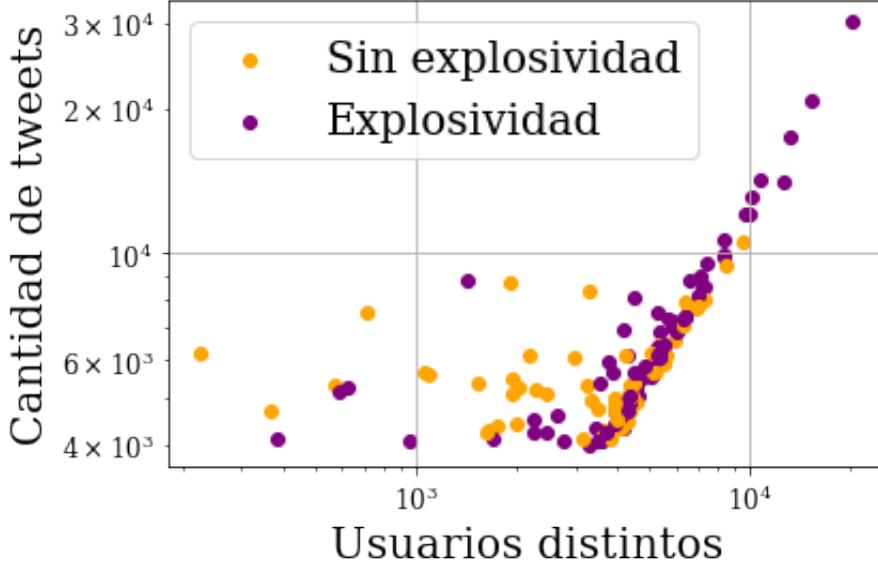


Figura 4.1: Comparativo de tendencias por la cantidad de usuarios registrados respecto a la cantidad de tweets recabados.

4.2. Análisis de Comunidades.

En esta sección, se analizarán dos momentos importantes de la interacción de los usuarios por cada tendencia a través del tiempo: Minutos previos a la mayor interacción en la red social $\mathcal{G}(h)$ y el momento de la mayor interacción en la red $NC_{t^*}(h)$.

Cabe recalcar que la principal diferencia entre las redes $\mathcal{G}(h)$ y $NC_{t^*}^T(h)$ es la escala y el tiempo de análisis. Es decir, al analizar la red $NC_{t^*}^T(h)$ interesa la dinámica del comportamiento explosivo en el momento del pico de interacción (t^*). En cambio, de la red $\mathcal{G}(h)$ interesa conocer más sobre la estructura de la red social que llevó al comportamiento explosivo. Considerar esta diferencia permite analizar los resultados desde una perspectiva cualitativa y desde un enfoque predictivo.

4.2.1. Comunidades con comportamiento similar.

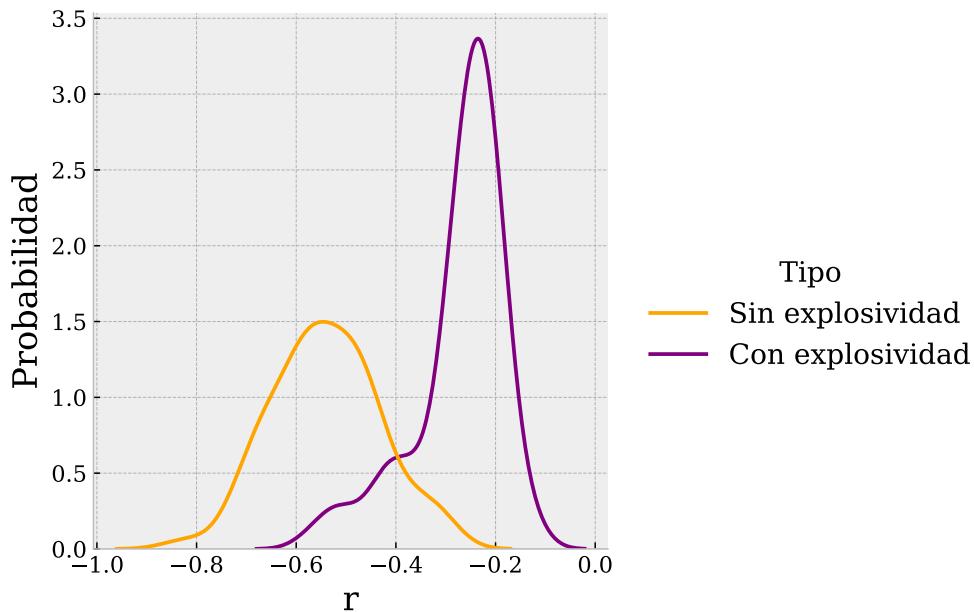
El primer paso fue determinar el grado de asortatividad (r) de la red $NC_{t^*}^T(h)$. Dicha métrica denota distribuciones de grado heterogéneas cuando el valor es negativo; es decir, hay nodos de grado alto enlazados con nodos de grado bajo indicando que la red es dissortativa. Además, se puede denotar la existencia de enlaces puente cuando r es negativo.

En los resultados, un menor grado de dissortatividad de las redes $NC_{t^*}^T(h)$ etiquetadas por la propiedad de explosividad, resulta ser significativa (95 % de confianza por Kolmogorov-Smirnov). Es decir, la red social de los comunicantes de tendencias con comportamiento

Tabla 4.2: Comparativo de las estadísticas de la cantidad de comunidades considerando la propiedad explosividad y sin explosividad.

	Explosividad	Sin explosividad
Promedio de la cantidad de comunidades	167.951807	99.416667
Menor cantidad de comunidades	8.000000	6.000000
Máxima cantidad de comunidades	363.000000	418.000000

explosivo resultan ser menos assortativas. Esta diferencia es visible en la figura ??.



Por otro lado, una mayor cantidad de comunidades partícipes marca la diferencia de tendencias con comportamiento explosivo de las que lo son (95 % de confianza por Kolmogorov-Smirnov). En la tabla ??, se aprecia como la media del número de comunidades es mayor en tendencias con comportamiento explosivo. Este resultado se puede visualizar de la figura ??.

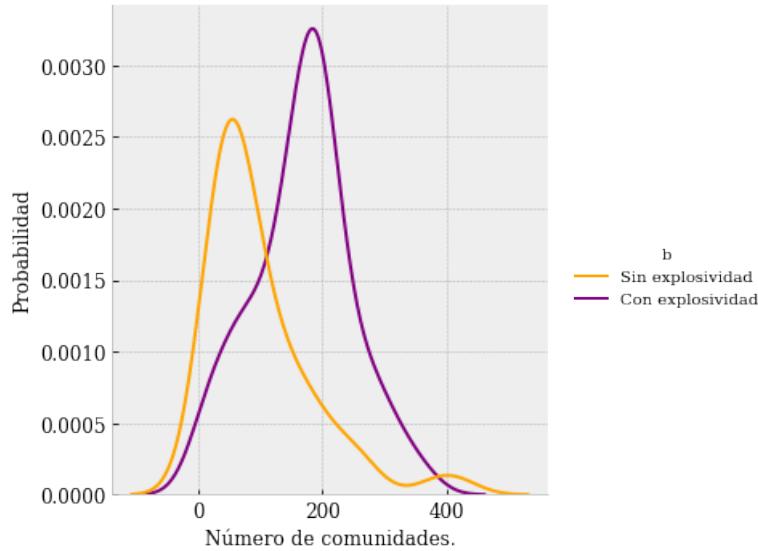


Figura 4.3: Histograma del número de comunidades de las redes $\mathcal{G}(h)$ por su tipo de comportamiento. Se observa la existencia de un sesgo significativo que caracteriza a las tendencias.

4.2.2. Incremento de comunidades respecto al tiempo

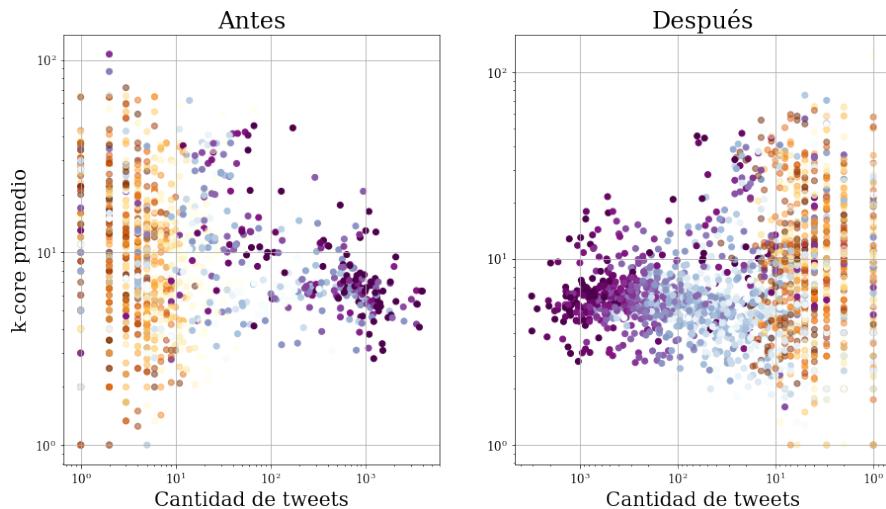


Figura 4.4: Comparativo del k – core de los usuarios partícipes antes y después del momento de mayor interacción (**peak**). De color rojo son aquellas tendencias con comportamiento explosivo.

Un paso relevante fue el cálculo del k -core en las redes $\mathcal{G}(h)$. Este cálculo permite identificar qué tan lejanos están los nodos que publicaron un tweet, respecto al centro de su comunidad. Es decir, se espera encontrar una capa en la comunidad donde se realice la mayoría de las interacciones.

Los resultados se obtuvieron con el siguiente procedimiento. De la red $\mathcal{G}(h)$, se generó

el valor k-core para cada nodo. Después, se identificaron aquellos nodos que realizaron algún *tweet*. Así, para generar una sucesión de números, se agruparon los nodos en intervalos de tiempo de 15 minutos antes y después del momento de mayor explosividad. Finalmente, se obtiene la media de los valores recolectados por cada intervalo de tiempo. En la figura ??, se aprecian dos gráficas de los resultados obtenidos en este proceso.

La interpretación de los resultados indica una capa donde se caracteriza la interacción relevante. Es decir, hay un valor del valor k-core que distingue las tendencias de comportamiento explosivo de las que no son.

Antes de la mayor interacción, el k-core, no tiene un valor establecido para las tendencias. Sin embargo, en el último intervalo de tiempo antes de la mayor interacción, el valor k-core de los nodos es cercano a 10.

Después de la mayor interacción, el valor k-core se dispersa con una tendencia a la disminución. En otras palabras, la interacción ya no es gobernada por nodos donde su k-core es cercano a 10.

4.3. Análisis Global de la Dinámica de las Tendencias.

En esta sección, se explicará con detalle el cálculo que se realizó para la red $NC_t^T(h)$ por cada tendencia h . Cabe recordar que la red $NC(h)$ es una red temporal multicapa, en este caso (ver la sección de metodología), $NC^T(h)$ es una red no dirigida y $NC_t^R(h)$ es una red dirigida donde t es un periodo de tiempo fijo. Cada una en estas redes representa a un usuario que participó con algún tweet o retweet en el periodo de tiempo t .

Con esto en mente, es factible aplicar diversas métricas conocidas de redes. Por ejemplo, para la red $NC_t^T(h)$ se puede calcular el coeficiente de assortatividad y para NC_t^R se puede obtener la media del grado exterior.

Con esto en mente, se realizaron los siguientes cálculos, especificando en qué red fue aplicada.

- Para cada NC_t^T donde t es una ventana de tiempo.
 1. Grado promedio.
 2. Media del camino más corto.
 3. Cálculo del coeficiente de centralidad clustering
 4. Cálculo del coeficiente de centralidad betweenness
 5. Cálculo de la entropía de Shannon de la distribución de grado.
 6. Grado promedio

- Para cada NC_t^R
 1. Grado exterior más grande.
 2. Diámetro más grande.

De cada tendencia se obtuvo una serie de tiempo por cada una de las métricas. Para inferir un valor característico en cada una de estas métricas, se procedió a calcular un estadístico en cada una de estas serie de tiempo. Por un lado, para las métricas relacionadas a la red no dirigida (NC_t^T) se obtuvo la media de cada una de las series de tiempo. Por otro lado, para las redes dirigidas, se procedió a calcular el máximo valor de esta serie de tiempo.

A partir del cálculo de estos estadísticos, se empiezan a notar diferencias entre las tendencias. Se puede notar de la tabla ??, que la diferencia entre el comportamiento de tendencias resulta ser estadísticamente significativo a través de la prueba de Kruskal-Wallis

Tabla 4.3: Valores p obtenidos al aplicar la prueba no paramétrica de Kruskal-Wallis.

Estadística	$p - value$
Media de la cantidad de Tweets por hora	1.585502x10 ⁻¹⁶
Media de la cantidad de retweets	2.797278x10 ⁻¹⁸
Media del grado promedio	5.785002x10 ⁻⁰²
Media del promedio de longitud del camino más corto	2.423004x10 ⁻²⁰
Media del clustering	1.419858x10 ⁻⁰⁷
Media del betweenness	2.560294x10 ⁻⁰¹
Media de la entropía	2.288698x10 ⁻¹⁸
Diámetro más grande	4.843335x10 ⁻¹⁴
Vecindad más grande	1.283606x10 ⁻¹⁵

4.3.1. Valores Significativos en la Dinámica Total y en el Pico de Interacción.

La entropía de Shannon (??) es una medida de incertidumbre para una variable aleatoria. Por su propia construcción, es positiva y no se enfoca en los valores de la variable aleatoria; se da énfasis en la probabilidad de los eventos más no el evento en sí. Se especifica en la incertidumbre del estado actual.

En las figuras ?? y ??, se observa la relación entre el comportamiento explosivo y la entropía. Si bien algunos datos son atípicos, las tendencias con comportamiento no explosivo tuvieron un valor mínimo en la entropía (0.25 aprox.). Con base en la entropía de Shannon,

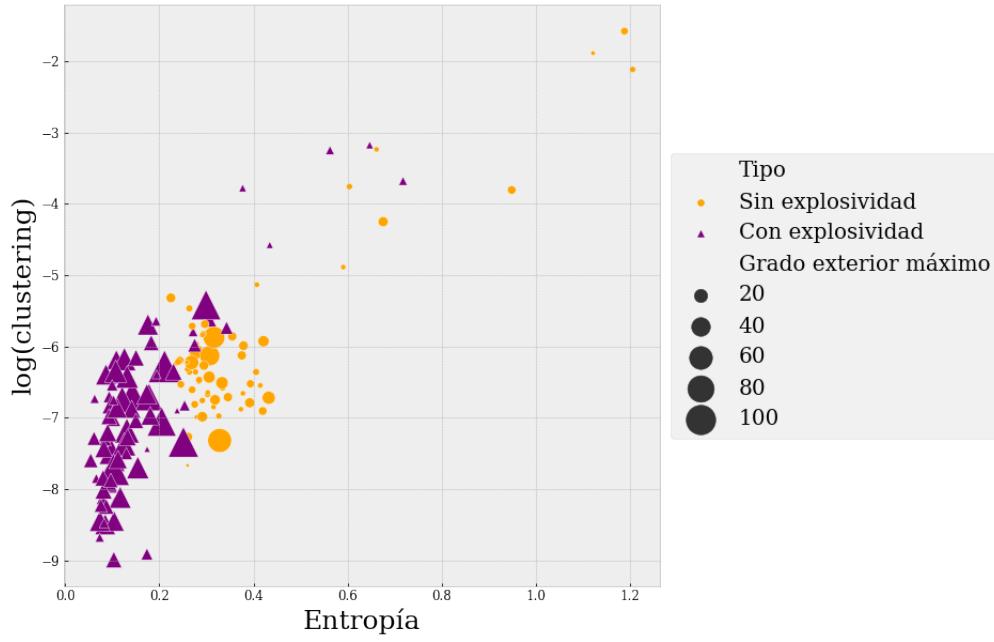


Figura 4.5: Comparativo entre las tendencias respecto a las métricas calculadas. El tamaño de cada uno de los marcadores es proporcional al máximo del grado exterior máximo de cada una de las tendencias.

la red de las tendencias con comportamiento explosivo son homogéneas en la distribución de grado.

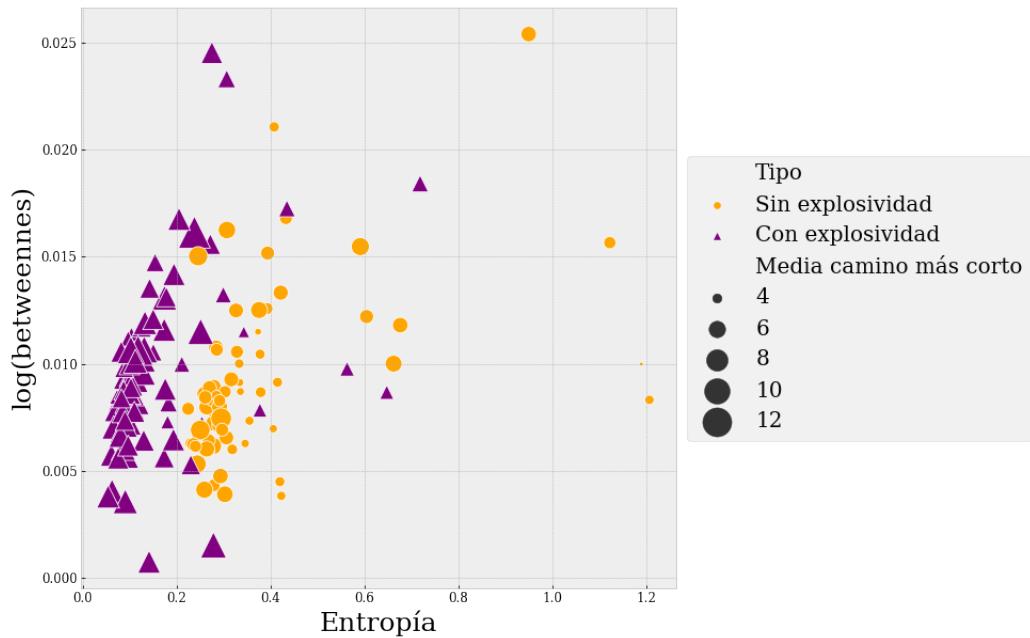


Figura 4.6: Comparativo de tendencias entre betweenness (media) y entropía. El tamaño de cada uno de los marcadores del gráfico está en proporción de la media del camino más corto.

También se obtuvieron relaciones interesantes entre las métricas betweenness y entropía. En la figura ??, se observa una correlación lineal positiva entre entropía y clustering. Una correlación esperable ya que al incrementar el clustering, se incrementa el número de triángulos y, por tanto, se incrementa el grado de los nodos. Además, en la misma figura, se puede apreciar como el beneficio de actividades de costo mínimo, como los retweets, propician que la tendencia sea de comportamiento explosivo. Esto último puede ser consecuencia del tipo de tendencia.

En resumen, se necesita cierta homogeneidad entre comunidades para generar el comportamiento explosivo. Específicamente, que esa homogeneidad no rebase una entropía de 0.25. Por otro lado, también debe existir la necesidad de acciones de menor costo y con tiempos específicos. La dinámica de *Twitter* es una comunicación sencilla y simple, y la acción de *retweet* es la que recapitula la mayor cantidad de interacciones.

5 Discusión

Es importante mencionar que, los datos del estudio son del año 2012 por lo que no es posible generalizar a dinámicas actuales. La tecnología y necesidades hoy en día han cambiado considerablemente, así como la sociedad ha modificado las políticas permitiendo una alteración al comportamiento [?]. Con base en lo anterior, para hacer un análisis actualizado, es importante considerar información como geolocalización o algoritmos basados en preferencias de los usuarios, dado que son nuevas dinámicas en las relaciones.

Por otra parte, el factor tiempo, entre cada interacción y la repetición o preferencias, se muestra en la figura ??, donde se observa un umbral de interacciones repetidas, es decir, que un mismo usuario repita interacciones. De manera específica, en la figura ??, se observan los usuarios en los primeros mil tweets ordenados antes del periodo t^* para cada tendencia; siendo $n = 1$ el tweet más antiguo y $n = 1000$ el más reciente.

De manera formal, si $\{U_i\}_{i=1}^{1000}$ es la sucesión de usuarios de los tweets, entonces la figura ?? es $\frac{|\cup_{i=1}^k U_i|}{k}$ para cada $1 \leq k \leq 1000$, a esta función se le conoce como *la razón de la cantidad de usuarios distintos hasta el n -ésimo tweet*.

Finalmente, se concluye que valores cercanos a uno son tweets hechos por usuarios distintos; mientras que valores cercanos a cero son tweets hechos por un mismo usuario. El grosor de cada línea es proporcional al tiempo que hay entre cada tweet. Por lo que se reconoce la importancia del tiempo entre las interacciones y el número de interacciones únicas, influye en la explosividad de la tendencia.

Razón de usuarios distintos.

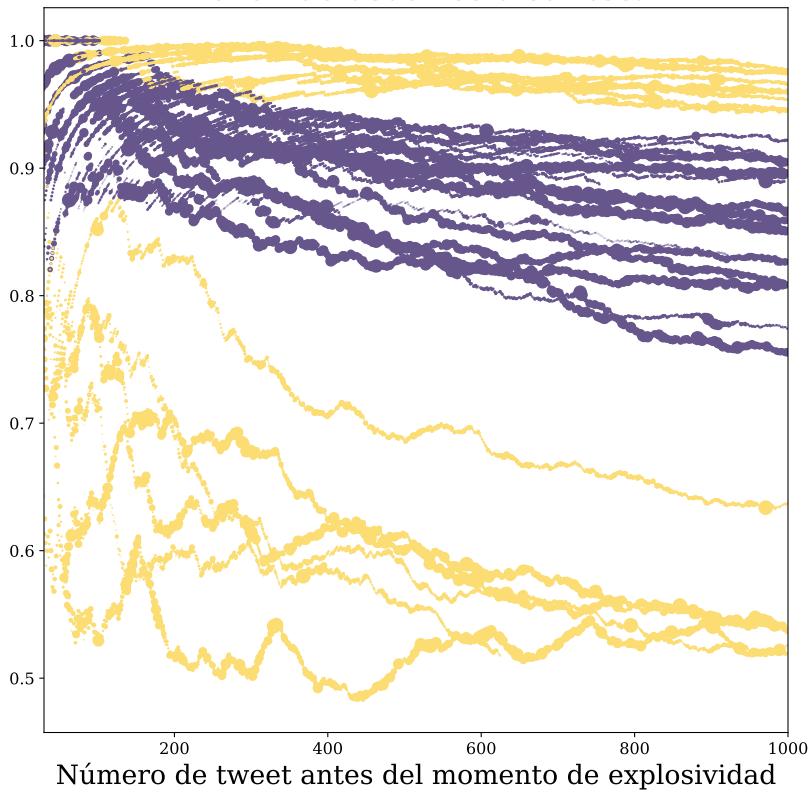


Figura 5.1: Comparativo de tendencias por los primeros 1000 tweets. Razón de usuarios nuevos en cada tweet de los 1000 tweets antes del periodo t^* . Las líneas de color morado corresponden a tendencias con comportamiento explosivo.

Conclusiones

El objetivo de esta tesis era encontrar una diferencia en las redes sociales de cada tendencia por su propiedad de explosividad. En otras palabras, se buscaba un patrón en la red a fin de identificar la explosividad de la tendencia en función de la red social de los participantes.

De los resultados, se tienen varios diferenciadores entre las tendencias con comportamiento explosivo de las que no lo son. Estas diferencias se pueden resumir con el siguiente cuadro.

Tabla 5.1: Comparativo de los resultados por propiedad de explosividad

Comportamiento explosivo	Comportamiento no explosivo
La entropía es menor	La entropía es mayor
El número de comunidades es mayor	El número de comunidades es menor
El <i>K-core</i> menos variable	<i>K-core</i> variable

Una de las conclusiones importantes es que pudieron identificarse las características más relevantes como: el tiempo entre cada interacción, el número de interacciones y la cantidad de comunidades. Dichas características permiten diferenciar tendencias con comportamiento explosivo o sin comportamiento explosivo.

Por último, se concluye que el comportamiento explosivo es influenciado por la cantidad de pequeñas comunidades con características similares; y el tiempo en el cual están interactuando; es decir, está influenciada por la topología e interacción coordinada de la red generada de los comunicantes. Las características más relevantes son el tiempo entre cada interacción, número de interacciones y la cantidad de comunidades. Los resultados obtenidos en esta tesis son consistentes con la literatura reportada ??.

Trabajo futuro

En esta sección se hace una propuesta considerando las limitantes encontradas en el trabajo que se realizó.

La primera limitante fue la sobre simplificación al considerar los enlaces no dirigidos. Sin embargo, los *influencers* tienen una gran cantidad de seguidores que ellos no necesariamente siguen. Por lo tanto, una propuesta de investigación es analizar estos casos con relaciones dirigidas.

La segunda limitante fue el análisis de la actividad de la red social en períodos de tiempo establecidos, sin importar la cantidad de interacciones en dicho periodo. Cabe mencionar, la importancia del tiempo entre interacciones y usuarios únicos que participan (ver sección de Discusión). Entonces, se propone una modelación de red multicapa, donde los nodos son los usuarios, cada capa representa una comunidad y los enlaces intercapa podrían representar la unión de las comunidades.

La tercera limitante fue que el trabajo se enfocó únicamente en el servicio de media social Twitter. Así que, una propuesta, al considerar las nuevas dinámicas en redes sociales, sería integrar otras redes de media sociales.

Finalmente, la característica principal de este trabajo fue el enfoque en la estructura de las comunidades, es decir, no eran relevantes las características particulares de nodos importantes. Lo anterior conlleva a considerar si ese comportamiento era influenciado únicamente con la actividad de un nodo importante o no. Por tanto, la última propuesta sería analizar la actividad de los nodos importantes para propiciar un comportamiento explosivo en función de métricas de centralidad y el tiempo entre sus actividades en una red multicapa dimensional.