

UNIVERSIDAD ADOLFO IBAÑEZ

TESIS DE MAGÍSTER

**ANÁLISIS DE LAS CARACTERÍSTICAS Y
RAPIDEZ DE PROPAGACIÓN DE LA FUNA EN
TWITTER**

Autor:
Valentina Martínez Pastene

Profesor guía:
Sebastián Moreno Araya

*Tesis realizada acorde a los requerimientos para el grado de
Ingeniería Civil Industrial, Ingeniería Civil Informática y Magíster en Ciencias de
la Ingeniería, mención en Tecnologías de la Información
de la*

Facultad de Ingeniería y Ciencias

13 de enero de 2022

UAI

FACULTAD DE
INGENIERÍA Y
CIENCIAS

UNIVERSIDAD ADOLFO IBAÑEZ

Resumen

Facultad de Ingeniería y Ciencias

Ingeniería Civil Industrial, Ingeniería Civil Informática y Magíster en Ciencias de la Ingeniería, mención en Tecnologías de la Información

ANÁLISIS DE LAS CARACTERÍSTICAS Y RAPIDEZ DE PROPAGACIÓN DE LA FUNA EN TWITTER

por Valentina Martínez Pastene

Hoy en día las *funas* han conseguido tener un importante espacio en las redes sociales, dado que se han constituido como el medio para lograr el escarcamiento social del agresor, o bien, para advertir a la comunidad del posible peligro de relacionarse con él. Una de las redes sociales más utilizadas en este sentido es Twitter, dadas sus funcionalidades que empoderan al usuario para la búsqueda y difusión de información. A pesar de la notoriedad que han alcanzado las *funas* en el Chile de hoy, aún no se conoce el real alcance ni rapidez de su propagación en Twitter. En este sentido, a través del estudio de la difusión de publicaciones *tendencias* (*trend*) y *funas*, se busca responder a esta interrogante. Así, se plantea como hipótesis que la *funa* se difunde igual o más rápido que una *tendencia* promedio, dentro de su primera semana de propagación en Twitter. Para alcanzar tal objetivo, se utiliza como metodología la creación de una red por cada publicación analizada, relacionando a los usuarios que retweetearon con los que potencialmente se enteraron de la información. A medida que se iban añadiendo las relaciones (aristas) y las horas transcurridas, se analizaron las siguientes características: *degree distribution*, *cluster coefficient*, *hop plot* y *betweenness centrality*. En base al análisis, se agruparon las redes en: *trend* común, *funas* y publicaciones anómalas. Todas las difusiones inician con uno o más usuario principales que propaga la información a varias personas independientes. A medida que avanza el tiempo, el *trend* promedio mantiene el protagonismo de dicho usuario como difusor principal, y alejados de él, se forman grupos por los que se propaga la información. Mientras que las otras redes, refiriéndose a las *funas* y publicaciones anómalas, continúa su propagación primordialmente a través de un grupo principal y cohesionado, compuesto por personas que difunden la información. Además, se observa que dentro de las primeras 5 horas se alcanza el 50 % de las aristas, y ya a las 10 horas la forma de difusión se mantiene relativamente. Asimismo, el 70 % de las personas, retweetea las *funas* dentro de las 6 y 12 horas. Por tanto, se logra comprobar que las *funas* se difunden igual o incluso más rápido que un *trend* promedio en Twitter.

Palabras clave: funa, Twitter, trend, usuarios, aristas, horas, degree distribution, cluster coefficient, hop plot, betweenness centrality, red

Índice general

Resumen	I
1. Problema	1
2. Estado del arte	3
2.1. Modelos estructurales	4
2.2. Difusión de la información	8
3. Hipótesis	11
4. Metodología	12
4.1. Average degree distribution	14
4.2. Average hop plot	14
4.3. Average betweenes centrality	16
4.4. Average cluster coefficient	17
5. Análisis	19
5.1. Trend de la selección chilena de fútbol	20
5.1.1. Análisis de características: <i>Average Degree Distribution</i>	21
5.1.2. Análisis de características: <i>Average Hop Plot</i>	23
5.1.3. Análisis de características: <i>Average Cluster Coefficient</i>	24
5.1.4. Análisis de características: <i>Average Betweenes Centrality</i>	24
5.1.5. Recapitulación de análisis	25
5.2. Funa a Patricio Maturana	26
5.2.1. Análisis de características: <i>Average Degree Distribution</i>	26
5.2.2. Análisis de características: <i>Average Hop Plot</i>	28
5.2.3. Análisis de características: <i>Average Cluster Coefficient</i>	29
5.2.4. Análisis de características: <i>Average Betweenes Centrality</i>	30
5.2.5. Recapitulación de análisis	31
5.3. Trend de la entrevista de Mónica Rincón a ex ministro	31
5.3.1. Análisis de características: <i>Average Degree Distribution</i>	32
5.3.2. Análisis de características: <i>Average Hop Plot</i>	33
5.3.3. Análisis de características: <i>Average Cluster Coefficient</i>	34
5.3.4. Análisis de características: <i>Average Betweenes Centrality</i>	35
5.3.5. Recapitulación de análisis	35
5.4. Funa a Edwin Dimter	36
5.4.1. Análisis de características: <i>Average Degree Distribution</i>	36
5.4.2. Análisis de características: <i>Average Hop Plot</i>	38
5.4.3. Análisis de características: <i>Average Cluster Coefficient</i>	38
5.4.4. Análisis de características: <i>Average Betweenes Centrality</i>	39
5.4.5. Recapitulación de análisis	40
5.5. Trend de José Antonio Kast	40
5.5.1. Análisis de características: <i>Average Degree Distribution</i>	41

5.5.2.	Análisis de características: <i>Average Hop Plot</i>	42
5.5.3.	Análisis de características: <i>Average Cluster Coefficient</i>	43
5.5.4.	Análisis de características: <i>Average Betweenes Centrality</i>	43
5.5.5.	Recapitulación de análisis	44
5.6.	Funa a la primera línea	44
5.6.1.	Análisis de características: <i>Average Degree Distribution</i>	45
5.6.2.	Análisis de características: <i>Average Hop Plot</i>	46
5.6.3.	Análisis de características: <i>Average Cluster Coefficient</i>	47
5.6.4.	Análisis de características: <i>Average Betweenes Centrality</i>	47
5.6.5.	Recapitulación de análisis	48
5.7.	Discusión	48
5.7.1.	Características de redes trend esperadas de Twitter	51
5.7.2.	Características de redes tipo funa	51
5.7.3.	Características de redes anómalas	52
5.8.	Análisis de Usuarios	53
5.8.1.	¿Quién mantiene la difusión?	53
5.8.2.	¿Bots?	54
6.	Conclusiones Generales	56

Capítulo 1

Problema

La *funa* es un concepto muy popular en Chile. Si bien no es un término reconocido por la RAE, tiene real importancia en las comunidades de redes sociales. En Chile no es un término nuevo, su etimología es de raíz mapuche, que en su idioma (mapudungún) significa *podrido*, pero hoy en día se le da una connotación un tanto distinta. El estudio *La Funa: Aspectos Históricos, Jurídicos y Sociales* de la Universidad de Chile, define *funar* como “acción de denostación y repudio hacia alguien que se considera que ha incurrido en una acción injusta o ilegal” [1]. En otras palabras, la *funa* expone el comportamiento inadecuado de una persona en la sociedad, y usualmente es a través del relato de una historia. Esta vivencia puede ser un abuso físico y/o psicológico e incluso la exposición de acreedores morosos. Su difusión masiva es crucial para el escarcamiento público del *funado/a* o agresor/a. De este modo, se advierte a la comunidad del potencial peligro al relacionarse con él o ella, y/o también como reprimenda. A pesar de que el objetivo es claro, aún no se conoce el real alcance de las *funas* en las comunidades donde se divulga.

Si bien fue en el año 2018 cuando las *funas* ganaron relevancia, su uso data desde fines del siglo XX, con el fin de exponer las violaciones de derechos humanos incurridos en la dictadura de Chile. Luego de la detención de Augusto Pinochet, se utilizó la *funa* para destapar a los responsables de todas las desapariciones, ejecuciones y torturas que hubo bajo su mandato. A los transeuntes que circulaban cerca de las manifestaciones se les entregaban volantes informativos acerca del *funado* y el hecho por el que era acusado [1, 2, 3]. Hoy en día se ha extendido el alcance de la *funa*. No solamente abarca la revelación de violaciones a derechos humanos, sino también cualquier situación ilegal o injusta. Además con el tiempo ha cambiado su método de difusión, siendo la red social el principal medio de propagación, porque en esta es posible alcanzar un público masivo, gracias a la rapidez de difusión y la facilidad de publicación desde cualquier lugar.

Según el reporte *Digital 2020* elaborado por *We are Social*¹, las redes sociales se han convertido en un indispensable en el diario vivir de las personas alrededor del mundo. Actualmente, las redes sociales cuentan con 4.5 billones de usuarios en el mundo, y alrededor de 15 millones son de Chile. En la audiencia chilena, se concentra un 33 % entre los 23 y 34 años de edad. Dentro de las más famosas, se encuentran Facebook con 12 millones de chilenos, Instagram con 8.2, Twitter con 2.47, y Snapchat con 1.55². Las redes han causado un gran impacto a los medios de comunicación pasando a ser consideradas como “el canal de difusión de información más inmediato e

¹En colaboración con Hootsuite <https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media>

²<https://datareportal.com/reports/digital-2020-chile>

interactivo”[4]. En la presente investigación se utilizará Twitter por ser la red social que concentra comunidades de usuarios cuyas intenciones están más centradas en la difusión y búsqueda de información [5].

Es a través de las funcionalidades que Twitter ha adaptado y creado, se ha facilitado la difusión y búsqueda de información. En esta red social una persona puede suscribirse (o seguir) a otra para que prioritariamente le aparezcan sus publicaciones (o tweet). Para marcar un mensaje se crearon los hashtag, que son un conjunto de palabras anteceditos por un #. También son usados para unir discusiones públicas, categorizar tweet o construir comunidades en torno a un tópico de interés común [6]. En base a los hashtag Twitter realiza un ranking, permitiendo observar las tendencias del momento (trend), sin ser una limitante a quienes se sigue. Adicionalmente, se creó el retweet, que consta en la re-publicación del tweet de una persona. Al retweetear, el usuario se empodera difundiendo información a su elección, sin importar la relación con el autor de la publicación, sino más bien el interés del contenido [7]. Estas son algunas de las funcionalidades de Twitter que fomenta la difusión de información, convirtiéndose en una red social relevante para investigar, analizar y observar el real alcance de las *funas*.

Por las razones anteriormente descritas, Twitter se ha transformado en una importante alternativa para que los chilenos compartan y busquen información. Dentro de todo lo que se ha compartido, la *funa* ha tenido una gran relevancia y un importante lugar. Gracias al uso de hashtag, retweet y tweet, los diferentes grupos organizados y comunidades, han utilizado la *funa* como medio de acción, logrando importantes trends a nivel país. Sin embargo el problema está en que actualmente no se conoce con exactitud qué tan rápido se propagan las *funas* ni cuál es su alcance.

La presente investigación tiene como objetivo analizar la rapidez y las características de la propagación de las *funas* en la red social de Twitter. Se hace interesante el estudio de la *funa* al ser un tema prácticamente nuevo en las redes sociales y, como se mencionaba antes, en realidad no hay conocimiento si su divulgación es efectiva ni su real alcance.

Capítulo 2

Estado del arte

La llegada y el amplio desarrollo del internet, han cambiado muchos aspectos de la sociedad, dentro de los cuales se encuentra la forma de interacción y comunicación de las personas [8]. Anteriormente, los medios masivos de comunicación eran solo los oficiales, como el periódico, la televisión y la radio, entre otros. La única forma en que se difundía la información era unidireccionalmente. Después de la aparición del internet, en el año 2004 comenzaron a crearse los primeros sitios web donde se podía recibir retroalimentación del espectador [9]¹. Posteriormente, en el año 2005 las redes sociales empezaron a ganar importancia, permitiendo la interacción multidireccional, es decir, conversaciones de muchos a muchos, o uno a muchos [10]. Gracias al nuevo rol de las personas, hoy en día, las redes sociales son el medio para la comunicación e interacción entre distintos usuarios sin límite de tiempo ni lugar. Así se fueron estableciendo conversaciones de forma asincrónica y sin límites geográficos [11].

Las redes sociales son un espacio de comunicación entre las personas, representado a través de sus usuarios y las conexiones entre ellos. En las plataformas de las redes los participantes pueden: (1) proveer contenido, consumir el contenido proporcionado por otro usuario y/o contenido proporcionado por el sistema, (2) articular conexiones para llegar a otros usuarios; y (3) consumir, producir, y/o interactuar con usuarios que generan contenido [8]. Es así como el comportamiento de los usuarios permite obtener, difundir, y/o producir información. En efecto, cualquier persona puede ser un potencial participante, pues las redes sociales son relativamente fáciles de usar y no existe un costo monetario para iniciar sesión y ser parte de la comunidad en línea. Incluso los grupos que son minoría, que usualmente son pasados por alto, pueden utilizar esta herramienta para que su voz sea escuchada [10].

Dentro de las redes sociales más importantes se encuentra Twitter, un servicio de microblogging fundado en el año 2006. Es una red social que desde que fue creada, se centra en poder proporcionar el espacio de interacción entre las personas. En su comienzo, permitía compartir mensajes (tweet) de solo 140 caracteres entre personas, ya que como señala el director creativo de Twitter: *la creatividad nace de las limitaciones* [11]. Si bien esto incluía más caracteres que los SMS², aún seguían persistiendo limitaciones en la comunicación. Por lo cual, se empezaron a redefinir y crear más funcionalidades. Por ejemplo, se aumentó el tamaño de sus tweet a un máximo de 280 caracteres, y podían ser publicados de forma pública o privada. Lo público implica que cualquier persona que tenga acceso a la red social, puede leerlo. Mientras que lo privado solo lo visualiza un grupo restringido de personas. De

¹<https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>

²Short Message Service

forma similar, si una persona desea estar al tanto de los tweets de otros, se puede suscribir (seguir) al usuario para así priorizar la aparición de sus tweets. Adicionalmente, Twitter cuenta con otras características que permiten el flujo de información entre usuarios, tales como los hashtag y retweets.

Twitter tiene dos herramientas principales para que la información fluya a través de los usuarios: hashtags y retweets. El primero es una palabra antecedita por un # que se usa para marcar los tópicos de una publicación o indexar palabras claves, así otros usuarios pueden seguir la conversación centrado en un tema en particular [7, 11]. Otra manera de fluir la información es retweeteando, permitiendo atraer a nuevas personas a un particular *hilo*³ de conversación e invitándolos a participar de ella sin tener que llamarlos directamente. Usando estas funcionalidades, se podría llegar a nuevas audiencias y así validar la información según los intereses de otros usuarios [11, 12].

La difusión de información en una red social se puede analizar a través de la teoría de grafos. Un grafo consiste en un set de nodos y un set de aristas, que indica la relación entre ellos. En el caso de Twitter, los nodos corresponden a las cuentas de Twitter, que representan a los usuarios de la red social. Mientras que las relaciones entre los nodos, se da por la interacción entre usuarios mediante comentarios, retweet, y/o seguimiento, entre otras maneras de conectarse. En un comienzo, dichas aristas se representaban a través de grafos estáticos, considerando nodos y aristas que no cambian a lo largo del tiempo. Sin embargo, en la vida real muchos de estos grafos son dinámicos, en otras palabras, nuevos nodos son agregados a la red, nodos existentes pueden ser eliminados y las relaciones entre ellos también pueden ir variando [13]. Para modelar dichas redes y entender como se difunde la información, se ha ido desarrollando distintos modelos de grafos. Mientras los primeros modelos buscaban replicar la estructura del grafo con una gran cantidad de suposiciones, los nuevos modelos son cada vez más complejos, llegando a incorporar variables dinámicas como lo es el tiempo.

2.1. Modelos estructurales

Uno de los primeros modelos estructurales desarrollados corresponde al modelo aleatorio en [14, 15]. Este modelo desarrolla la creación de grafos aleatorios formado por nodos fijos, cuyas relaciones son completamente aleatorias, independientes entre otras conexiones, y con una probabilidad constante. En el caso de un grafo de amistad, esto implica que cualquier persona puede ser amigo de otra sin importar sus relaciones existentes. Dada sus suposiciones básicas, este modelo careció de una dimensión crítica para simular la realidad, entre ellas la variable tiempo [16].

Desde la creación de este modelo, la mayoría de los avances se dieron desde la teoría, siendo uno de los más importantes que el *degree distribution* (distribución de las aristas de los nodos) se distribuye como una log-lineal según los atributos de la estructura ([17]). Recién en 1997 se propone un nuevo modelamiento estructural para grafos basado en bloques [18]. Este asume que los vértices de los grafos están particionados entre dos grupos y que la probabilidad de un enlace entre dos nodos depende solo del grupo al que pertenece. La estimación de dichas probabilidades para predecir la estructura de los bloques se basa en la máxima verosimilitud. Posteriormente, en [19] se propone el exponential random graph model (ERGM) que modela un grafo

³serie de tweet conectados de una misma persona

basado en las diversas características estructurales del mismo, como por ejemplo el número de triángulos, o vecinos conectados a cada nodo. El mayor problema de los ERGM es su lentitud de cómputo, el cual solo es aplicable para grafos pequeños. Por último, en [20] se introduce el modelo de bloques de dos niveles de Erdős-Rényi (BTER). Este modelo tiene como ventaja que modela las estructuras de comunidades usando grupos con conexiones dentro de cada grupo y entre grupos. En una primera etapa los nodos se agrupan formando los bloques/grupos correspondiente y posteriormente se unen entre sí según cierta probabilidad intra e inter bloque.

Mientras que los modelos anteriores buscan replicar las características de un grafo estático, existen otros tipos de modelos que replican las cualidades de un grafo dinámico, es decir, que varía con el transcurso del tiempo. Uno de los modelos más conocido es el propuesto en [21], el modelo de apego preferencial. Este modelo comienza con un grafo con una baja cantidad de nodos, y a medida que avanza el tiempo nuevos nodos se agregan al grafo conectándose exclusivamente con los nodos existentes. El nuevo nodo selecciona a quien conectarse con una probabilidad proporcional al número de aristas de cada nodo existente en el grafo. Esto se conoce como efecto *rich-get-richer*, que se refiere a que a medida que un nodo tiene más conexiones⁴, tiene una mayor probabilidad de generar nuevas conexiones. Posteriormente, en [22] se propone un modelo de apego preferencial que es utilizado para redes, en donde el número de nodos permanecen fijos y la cantidad de relaciones crece en el tiempo; además de extender la creación de nuevos enlaces entre los nodos pre-existentes. A pesar de la creación de otros modelos durante este periodo de tiempo, no es hasta [23] que se mezcló el modelo preferencial y aleatorio. En [23] se estudiaron el crecimiento de conectividad en un grafo evolutivo aplicando la fusión del modelo aleatorio con el de apego preferencial. Descubrieron que la probabilidad de que dos nodos sean conectados es linealmente proporcional al grado (cantidad de aristas que inciden en el nodo) de ambos nodos.

Durante el mismo periodo se desarrolló otros tipos de modelos que contempla el apego preferencial, pero están basados en otras métricas particulares de los grafos, como el coeficiente de agrupación (*cluster coefficient*) o centralidad de cercanía (*betweenness centrality*). Este último indica el número de caminos más cortos entre dos nodos, en el que se debe pasar por un nodo particular. Mientras que el coeficiente de agrupación mide el grado en que los nodos de una red tienden a agruparse. Uno de los primeros modelos importantes de este tipo fue propuesto en [24], donde el apego preferencial es basado en la centralidad de cercanía para explicar redes de colaboración en investigación. Se descubrió que cuando un nuevo nodo llega a la red se conecta por apego preferencial, tal como se ha explicado anteriormente. Sin embargo, este nuevo nodo pasa a ser una fuente de conexión para los futuros nodos con respecto a los antiguos nodos, incrementando su centralidad de cercanía. Un ejemplo propuesto por el investigador es que los supervisores de proyectos de doctorado y postdoctorado actúan como intermediarios entre los nuevos participantes y la red ya existente y, por lo tanto, se vuelven focos de vinculación preferencial. Con respecto al coeficiente de agrupación, uno de los modelos más famosos que considera esta propiedad es el modelo *Mundo Pequeño* de [25], en el que se logra tener valores realistas de la interconexión entre nodos colindantes, debido a que cerraba las conexiones entre nodos cercanos, es decir, se creaban nuevas aristas si los nodos eran cercanos entre sí. Otro modelo importante en la literatura, corresponde al

⁴conocido también como concentradores

modelo desarrollado en [26] el cual está basado en el principio de máxima verosimilitud. Este es un modelo probabilístico donde la unión de nuevos nodos a la red se definen según distribuciones dadas, y los enlaces posteriores se crean según el modelo de cierre de triángulos. En otras palabras, al ingresar un nuevo nodo, este elige un nodo ya existente por apego preferencial, y luego crea relaciones con sus vecinos cerrando triángulos, es decir, si A está relacionado a B y B a C, existe una gran probabilidad que exista una relación entre A y C para cerrar el triángulo. De forma similar, en [27] se propone un modelo en el cual nuevos nodos crean sus aristas basados en una probabilidad que considera el coeficiente de agrupación. Este es utilizado para explicar distintos fenómenos como el envejecimiento, la explosión y la formación de estructuras comunitarias en redes. Por ejemplo, puede describir las relaciones de las personas al preferir formar parte de un grupo pequeño y cercano (dénsamente conectado), en vez de uno grande liderado por una persona (el concentrador).

Otros tipos de modelos se han enfocado en entender y/o replicar la distribución de las conexiones entre los nodos, es decir, el *degree distribution*. Uno de los primeros modelos existentes está en [28], el cual replica de forma muy precisa el *degree distribution*. Para ello, el modelo define la probabilidad de una arista entre nodos A y B, como un valor proporcional al número de aristas de los nodos A y B. En otras palabras, si dos nodos tienen una gran cantidad de aristas, existe una gran probabilidad de que exista una relación entre ellos. Posteriormente, en [29] se estudia como el *degree distribution* cambia según el tamaño y conexiones iniciales de un grafo usando el modelo de apego preferencial. En el estudio se determina como será el *degree distribution* final basado en los distintos escenarios como por ejemplo, una red altamente conectada con pocos nodos. Luego, en [30] se determinó que el *degree distribution* de una red en expansión está sesgada por el tiempo. Específicamente, determinó que los nodos iniciales, aunque influyan fuertemente en la creación de la red, su relevancia disminuye con el avance del tiempo.

Finalmente, existen otros tipos de modelos que replican las estructuras existentes de los grafos usando diversas técnicas. En [31] se desarrolla un modelo llamado *Iterated Local Transitivity*, basado en la transitividad y el conocimiento local en las interacciones sociales a medida que crece el grafo. Cada cierto tiempo se analiza las características existentes en el grafo y el proceso de creación es adaptado para replicar las características del grafo original. Luego, en [32] se propone el modelo GEO-P, en que los nodos de las redes son puntos en el espacio métrico y las relaciones son creadas según la distancia relativa entre los nodos y una función de ranking basada en la popularidad. Se menciona que las redes sociales online se desarrollan en un espacio social, cuyas variables representan los rasgos de los usuarios, como serían sus intereses o lugar de residencia. Por ejemplo, las personas que sean de un mismo lugar de trabajo estarían cerca en el espacio social, teniendo una mayor probabilidad de conectarse.

A pesar del gran desarrollo explicado con anterioridad, la ciencia carecía de modelos para describir casos de relaciones dinámicas, como serían las relaciones de amistad y divulgación de noticias en comunidades. En el año 2012, Kim y Anderson introducen el concepto de grafo *time-ordered*, asumiendo que el grafo dinámico (cambia en el tiempo) puede estar representado por una serie de grafos estáticos, siendo cada grafo estático una captura en un instante del tiempo [33]. Esto permite un primer paso en el análisis de redes temporales, extendiendo ciertas propiedades de las redes a casos dinámicos como el *betweenness centrality*. De forma similar, en [34] se examina el modelo de grafos con variación de tiempo (TVG) para estudiar las propiedades

de la evolución de una red, proponiendo diferentes técnicas dependiendo si el indicador es para propiedades atemporales o temporales. Luego, en [35] se propone una teoría que extiende la teoría de redes a tiempo dinámico adaptando las redes de neurociencia. Para ello, se usa una representación conceptual similar a la anterior, añadiendo una matriz de conectividad para representar las multiconexiones de una parte del grafo, en vez de múltiples capturas estáticas en el tiempo. Para saber más detalles de este tipo de modelos, referirse a [13].

Recientemente, en [36] se propuso un modelo para generar redes estáticas o dinámicas. Para ello, se sigue tres ideas fundamentales: apego preferencial, asociatividad en los enlaces sociales de triángulo cerrado y la aleatoriedad en una red real. Esto permite crear grafos de cualquier tamaño y facilitan el entendimiento del crecimiento de conexiones entre nodos al focalizar el modelo en las conexiones y no en sus interacciones.

Si bien el desarrollo de modelos se pueden expandir en un sin fin de usos en la vida real, una de las áreas más importantes para esta tesis es su uso en las redes sociales online, por lo cual se procederá a describir trabajos relacionadas con esta área. En [37] se realizó uno de los primeros estudios en donde analiza el comportamiento de los usuarios en la redes sociales Flickr⁵ y Yahoo 360⁶. A cada nodo y relación se le asocia un tiempo en que fue originado y añadido al grafo. El análisis encontró que la red está compuesta por tres tipos de componentes:

1. Singletons: usuarios que se unen a la red pero nunca crean relaciones sociales.
2. Giant component: un grafo conectado que representa al mayor grupo de miembros, cuyas relaciones de amistades podrían ser de forma directa (arista entre dos nodos) o indirecta (existe un camino que puede llevar de un nodo a otro).
3. Middle region: grupos de usuarios que interactúan entre ellos formando pequeñas comunidades aisladas del Giant component.

En base a lo observado, el autor propone un modelo que toma en cuenta un apego preferencial sesgado dado que contrasta la relativa facilidad de encontrar posibles conexiones en línea dentro del componente gigante y la relativa dificultad de localizar posibles conexiones en las comunidades aisladas. Asimismo, en [38] se encontró observaciones similares al analizar la red social de Flickr. Observó un apego preferencial en la llegada de nodos y posterior creación de relaciones basado en el cierre de triángulos. De igual manera, en [39] se analizó la formación de relaciones en la evolución de la red en Twitter y notó que los usuarios populares con frecuencia cierran triángulos entre él y dos de sus seguidores.

Otros autores focalizaron sus investigaciones en la formación de triángulos cerrados en las redes dinámicas como son las redes sociales. Por un lado, en [40] se estudia la formación dinámica de triángulos cerrados en una red dinámica para luego proponer un modelo de predicción de estos. Para ello, incorporaron nuevos patrones y estructuras de la red descubiertas al investigar la red social de microblogueo Weibo. Uno de los patrones encontrados fue que si un *structural hole spanner*, individuo que se beneficia de cerrar las conexiones entre personas y grupos que están desconectados, es quien tiene las dos conexiones en un triángulo de usuarios, es muy difícil que se cierre la conexión. En cambio, si es quien tiene una de las conexiones, procederá a

⁵Red social donde se almacena, ordena, busca, vende y comparte fotografías o videos en línea

⁶Red social que integraba características de un blog y álbum de fotos de Flickr

cerrar el triángulo para tener más fuentes de recursos. En [41] se propone un modelo motivado por el cierre triádico. Postulan grafos *c-closed*, lo que quiere decir que para cada vértice u y v , al tener c vecinos comunes, entonces u y v están conectados. Este modelo facilitó encontrar *cliques* en una red, es decir, subconjuntos de la red en la que los actores están completamente vinculados entre sí. Recientemente, en [42] se logró ampliar el alcance de los grafos *c-closed* para el crecimiento de grafos dinámicos. Como en las redes sociales las relaciones entre dos nodos se hacen y deshacen de forma tan rápida, se propone un modelo que logra una generalización de *c-closed* evitando grandes costos computacionales.

2.2. Difusión de la información

Tal como se ha mencionado previamente, en las redes sociales los usuarios pueden interactuar de diferentes formas, y una de ellas es la difusión de información, la cual concierne a la presente investigación. A continuación se presentan algunos de los modelos más conocidos según [43]: el modelo de contagio, el modelo de influencia social y el modelo de aprendizaje social. El primero postula que la información puede potencialmente expandirse a través de la red social al igual que la propagación de una enfermedad infecciosa en una población. El modelo de influencia social propone que la información se propaga a través de los vecinos en una red. Se asume que un grupo de personas cree en la información cuando la cantidad de usuarios que difunden la información, supera cierto umbral. A diferencia de los modelos previos, el de aprendizaje social supone que los nodos son tomadores de decisiones racionales. Observan los resultados de sus vecinos que son adoptantes anteriores, y luego deciden si creen o no en la información [16].

Asimismo, existen modelos de difusión de información que se basan en el modelo SIR [44]. Dicho modelo es introducido inicialmente en la propagación de enfermedades. Se plantea que todos los nodos de una red son parte de uno de estos tres estados: susceptible (capaz de ser infectado), infectado, o recuperado (libre de dicha infección). En un lapso de tiempo, solo los nodos infectados pueden contagiar a los susceptibles con cierta probabilidad. Luego de dicho lapso, un nodo que estuvo en el estado infectado se mueve al estado de recuperación y deja de ser capaz de infectar o ser infectado. Existen variaciones del modelo SIR, como por ejemplo el modelo de cascada independiente, en donde en vez de que exista una única probabilidad de contagiar a los nodos susceptibles, existe una probabilidad por enlace [45].

Adicionalmente se han desarrollado más modelos de difusión de información. Uno de los modelos propuesto e está en [46], que propone un modelo de crecimiento logístico para la difusión de tópicos violentos por medios sociales. Este consiste en que el número de autores de una publicación violenta crece en función del número de personas que existan en la red, pero con un cierto límite. El número de autores se basa en tres características: los autores violentos actuales, la interacción entre sus seguidores y nuevos usuarios, y el decaimiento de la influencia de autores violentos en el tiempo. Luego, se encuentra el modelo de [47] que asume que los usuarios comparten o difunden información acorde a sus propias preferencias, y seleccionan diferentes tipos de relaciones sociales acordes a la información de preferencia. El modelo refleja el hecho de que los usuarios son activos y la información posee preferencia de propagación. Asimismo, en [48] se postula un modelo basado en el principio de cierre de triángulo y combina el factor de oportunidad, confianza social, y motivación de elección. El primer factor asume que la oportunidad de recibir determinada

información se puede medir por la similitud de intereses entre dos usuarios. El segundo factor determina un umbral de confianza social, combinando la influencia de la red de los usuarios y la contribución del contenido. Por último, el tercer factor hace referencia al reconocimiento de la motivación de los usuarios para difundir un mensaje dentro de un juego. Este factor determina que la motivación está dada por la elección de estrategias para beneficiarse lo mejor posible a sí mismo. Finalmente, se descubrió que al considerar simultáneamente la estructura de la red y el contenido de la información, la confianza social de los usuarios puede acelerar la difusión de información en las redes sociales de microblog como Twitter.

La teoría de la difusión de la información ha sido llevada a la práctica en diversos estudios. Por un lado, en [7] se concluyó que una vez que se retweeteaba un mensaje, el tweet era casi instantáneamente retweeteado en dos, tres y cuatro saltos de nodos alejados de la fuente, lo que significa una rápida difusión de la información luego del primer retweet. Esto implica que el tweet es retweeteado hasta por usuarios que están a tres personas de separación con el usuario que hizo la publicación. Complementando esta información, en [49] se analizó Twitter con el uso de cascadas (secuencia de activaciones generadas por un proceso de contagio, en el que los nodos hacen que los conectados se activen con cierta probabilidad) y detectó que el grafo tiene múltiples dendritas. Es decir, el grafo tiene forma de árbol, donde cada nodo sigue en promedio un usuario anterior. Asimismo, en [50] se estudió la evolución de Renren (red social en línea más grande de China) y analizó la propiedad temporal de la explosión en la creación de enlaces. Las explosiones fueron detectadas midiendo la aceleración del crecimiento de los grados del nodo. Una de las observaciones fue que cuando usuarios se unen por primera vez a la red, crean una relación basada en el mecanismo de apego preferencial; sin embargo, en estallidos posteriores, exploran varias regiones de la red para crear relaciones. Luego, en [51] se estudió la evolución de las redes sociales y su contenido de la red en Twitter. Se definió una red de contenido como una red de tweets de usuarios donde un usuario está vinculado a los tweets que ha publicado. Se descubrió que las características de una red social tienen una influencia fuerte en la red de contenido. Por ejemplo: el número de seguidores de un usuario está altamente influenciado a su nivel de actividad (número de tweets que publica). Asimismo, [52] monitoreó las elecciones de Irán en el año 2009 a través de Twitter. En este estudio se encontró que el grafo tiende a ser ancho, pero no muy profundo. Esto quiere decir que usualmente las personas que retweetean son seguidores del usuario que hizo el tweet original, entonces, la difusión se concentra principalmente en usuarios directos. Otra investigación relacionado a la difusión de la información es el análisis de tópicos explosivos.

En [53] se realizó una investigación en Twitter para encontrar patrones de tópicos explosivos, que corresponden a tópicos que ocurren en un lapsus de tiempo repentino y de corta duración. Para poder detectar estos tópicos, se utilizó el modelo desarrollado por [54, 55], los cuales detectan tópicos en un texto reducido, como lo son los tweet. El modelo utilizado logra descubrir los tópicos de forma más eficiente y asertiva que otros modelos existentes [56, 57, 58, 59]. Dentro de los patrones detectados se destaca:

1. La información tiende a propagarse desde usuarios influyentes a usuarios normales, por tanto, la razón principal de difusión es por dichos nodos principales.
2. La difusión de información de un tópico explosivo tiende a ser más amplia que profunda.

3. La difusión de información de múltiples tópicos explosivos tiende a propagarse entre los usuarios que son influyentes y llega a ser más profunda que la difusión de un solo tópico.

Por otro lado, el artículo [60] expone un estudio para entender cómo fluye la información en torno al *Universal Basic Income* (programa de ayuda social del gobierno de Estados Unidos). En su estudio categoriza las publicaciones y obtiene conclusiones de los tweets obtenidos de Twitter. Uno de los descubrimientos fue lo siguiente: las publicaciones informativas son las que más se comparten en Twitter. Mientras que los tweets que resuenan emocionalmente, es decir, publicaciones que los usuarios se sienten identificados con el sentimiento y creen que el resto debe conocerlo, son los que atraen nuevos usuarios a la discusión.

Por último, en [61] se estudió la viralización de las publicaciones verídicas y las que desinforman en Twitter, basado en los tweets publicados durante el huracán Harvey. Vale mencionar que lo viral hace referencia a que la información se difunde a una gran cantidad de personas de forma explosiva, y que no implica que estén relacionados con quien la publicó. Se descubrió que existe una alta viralización de publicaciones que desinforman, ya sea alguna novedad que transmite una emoción negativa o que tienen una baja complejidad léxica. Sin embargo, se concluyó que las noticias sociales reales son más propensas de ser virales que las noticias que desinforman.

Capítulo 3

Hipótesis

Hipótesis: La difusión de la *funa* a través de la red social de Twitter alcanza una rapidéz de propagación igual o mayor a un trend promedio durante su primera semana de difusión.

Los objetivos del presente trabajo son los siguientes:

- Adquirir datos de trends de Twitter a través de su api, durante un periodo de tiempo definido.
- Crear redes basadas en los datos extraídos de Twitter.
- Analizar el crecimiento de redes de trend a través del tiempo, usando las métricas *Average Degree Distribution*, *Average Hop Plot*, *Average Betweenes Centrality* y *Average Cluster Coefficient*.
- Comparar la difusión de la *funa* y de un trend promedio durante una semana, usando las métricas mencionadas anteriormente.

Capítulo 4

Metodología

Para analizar los datos de la difusión de la *funa* a través de la red social de Twitter, primero se debe realizar un proceso de captación y limpieza de datos, luego se procede a generar el grafo y, finalmente, modelar la propagación de la información. Con el fin de cumplir dichos objetivos, se recurrió a la API de Twitter para la recopilación de información y se utilizó el lenguaje de programación Python para el filtro de la data, formación de grafos y análisis.

En primer lugar, como se ha mencionado, se utilizó la API de Twitter para la extracción de la información. Esta API es abierta a desarrolladores inscritos en Twitter, en donde pueden hacer consultas de distinto tipo como: lectura de un tweet en particular, recopilación de las publicaciones de un usuario, *stream*¹ de tweets orientados a un tópico particular y obtención de amistades para usuarios particulares, entre otras peticiones. Particularmente, la información se extrae a través de las funciones *Filtered Stream v2* y *Friendship*. *Filtered Stream v2* permite estar escuchando constantemente a Twitter para obtener los tweets que cumplen una regla de contenido definida por el desarrollador. *Friendship* se utiliza para obtener a las personas que sigue cierto usuario. Gracias a esta fuente, se puede obtener todos los recursos necesarios para poder generar el grafo.

Con la API funcionando, se procede a la extracción de los tweets. Dentro de estas publicaciones capturadas se encuentra distinta información incluyendo: el identificador del usuario², el contenido de la publicación, el tipo de publicación (una palabra que indica si es un tweet, retweet, cita, comentario, entre otras) y fecha y hora del tweet. Además, existen otras variables que detallan más aún la publicación extraída, como sería el identificador de la interacción entre los usuarios, la ubicación geográfica, el lenguaje identificado y nombre de usuario, entre otra información. Sin embargo, para los alcances de este estudio, estas variables no son objeto de análisis, por lo cual se descartan. Una vez procesada esta información, se continúa con la creación del grafo.

Con la información ya extraída se genera el grafo a analizar. En primera instancia, se procede a la creación de los nodos del grafo, los que corresponden a todos los usuarios de los tweets extraídos. El proceso de generación de relaciones es más complejo, ya que dichas relaciones pueden ser de forma directa o indirecta con cualquiera de los nodos existentes en la red. Es decir, hay cierta probabilidad que el retweeteador no esté asociado a quien realizó el tweet original, dado que pudo haber visto la publicación a través de algún seguidor. Por lo tanto, la creación de la estructura de la

¹Ver y oír contenidos que se transmiten desde la api sin tener que descargar previamente los datos al dispositivo.

²Correspondiente a un número único del usuario

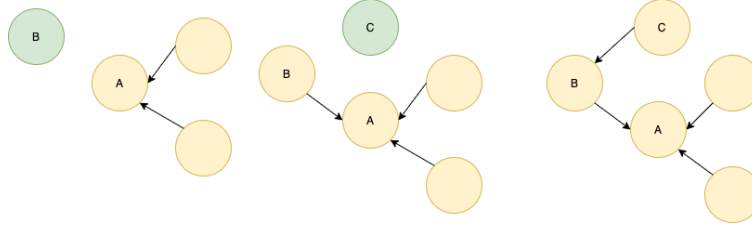


FIGURA 4.1: Ingreso de un nodo B seguidor de A, y C seguidor de B, a un grafo existente

red tiene como objetivo formar un grafo de retweets que relacione los nodos según los seguidores.

Prosiguiendo con la formación de la red, dicho desarrollo va en función del tiempo a medida que nuevas publicaciones van apareciendo. En un inicio se capturan todos los retweet que llegan a través del streaming que contengan un hashtag o palabra definida por el usuario, como por ejemplo *funa*. Luego, se comienza con la creación del grafo a través de un nodo central, correspondiente al usuario que publicó el tweet original (recordar que el resto de usuarios retweetea este tweet). Al analizar los retweet posteriores en forma temporal, se analiza si el usuario que retweeteo ya existe en la red en proceso de creación. En caso de ingresar un nuevo nodo (usuario), se analiza si sigue a algunos de los nodos ya existentes en la red con la función *Friendship* de la API. Una vez obtenido todos los nodos existentes que son sus amigos, se relaciona con cada uno de ellos. En el caso de que ningún amigo se encuentre en la red, entonces se enlaza con el nodo central.

Por ejemplo, la Figura 4.1 muestra la creación de enlaces en un grafo con dos nodos ya adyacentes, donde la publicación principal fue realizada por el nodo A. Dado que un nuevo nodo B realizó un retweet, este se agrega a la nueva red (imagen de la izquierda). Como B no sigue a ningún usuario de la red, este se relaciona directamente con A (imagen central). En cambio, cuando el nodo C ingresa a la red, se relaciona con B, dado que siendo seguidor del usuario B, se hace la suposición de que vio el tweet original a través de B.

Con el fin de comparar la difusión de la información existente en las *funas* versus otros tópicos, se realizará un contraste de los grafos generados para ambos casos (*funa* y otro tópico) durante un mismo periodo de tiempo. Esto será con el fin de poder contrastar el impacto de difusión que tiene una *funa* versus otros tópicos de rápida o lenta difusión. Específicamente, se guardarán los tweets que se obtengan durante una semana.

Luego, a partir del grafo creado con la información capturada, se procederá a analizar el flujo de los tweets según el tiempo y la red de usuarios. Específicamente, se analizará las características de *cluster coefficient*, *hop plot*, *betweenness centrality* y *degree distribution*.

Para poner bajo en contexto al lector, a continuación se entregará una definición, la interpretación y la expresión matemática de cada una de las características previamente señaladas para la siguiente definición de un grafo. Sea $G = (V, E)$ un grafo no dirigido donde $V = \{v_1, v_2, \dots, v_{N_v}\}$ es el conjunto de N_v vértices y $E \subseteq V \times V$ es un subconjunto de pares no ordenados de aristas, donde $(i, j) \in E$ implica una arista no dirigida entre los vértices v_i y v_j .

4.1. Average degree distribution

Para introducir lo que mide el *degree ditribution*, es necesario explicar qué es el *degree*. Sea d_i el grado de un vértice v_i , entonces d_i es la cantidad de conexiones que tiene el nodo v_i con otros nodos. El degree distribution dd_k es la fracción de nodos de un grafo con una cierta cantidad de grados, donde k varía entre uno y el máximo degree entre todos los nodos. Por lo tanto, si $dd_k = p$ implica que $p\%$ de los nodos tiene un degree igual a k .

Un típico ejemplo de un degree distribution se encuentra en la Figura 4.2. Como se puede observar es una distribución con una cola pesada. Esto significa que existe un gran número de nodos con bajas conexiones, mientras que pocos nodos tienen un gran número de conexiones, como se muestra en la figura 4.2b.

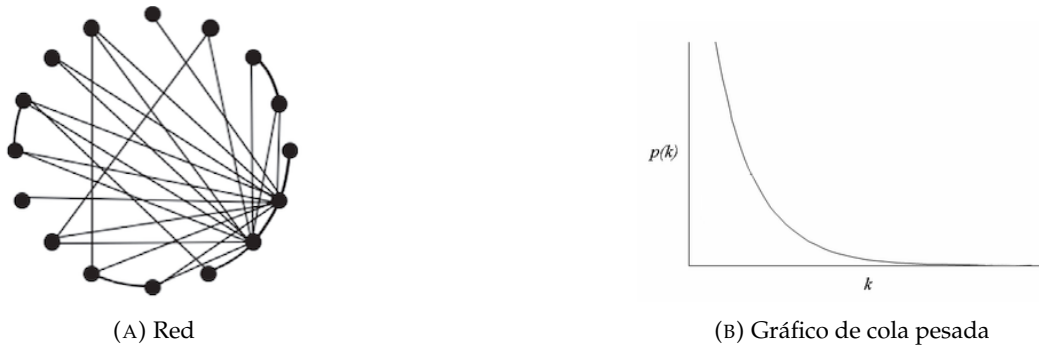


FIGURA 4.2: Ejemplo de *degree distribution* para un gráfico de cola pesada

En la presente tesis se desea evaluar como cambia la difusión a través del tiempo, por tanto, se considerará el average degree distribution sobre el tiempo para el análisis. Dado un grafo G con maximo degree M , el average degree distribution está dado por:

$$\overline{dd} = \sum_{i=1}^M i * dd_i \quad (4.1)$$

Esta expresión es el promedio de todos los degree del grafo. Por lo que se vuelve interesante hacer una comparación de su variación a través del tiempo, puesto que a medida que se irá capturando información el grafo irá creciendo. Vale mencionar que si en algún punto el grafo tiene un promedio pequeño, entonces es probable que la difusión se esté dando solo por el nodo principal o una cadena de usuarios conectados. En cambio si el promedio es un valor alto, se espera que los nodos tengan una alta cantidad de enlaces entre ellos, difundiendo la información entre sus seguidores.

4.2. Average hop plot

El *hop plot* es el número de nodos con una cierta distancia h entre ellos [62]. Por ejemplo, si el hop plot para $h = 2$ es igual a 10, entonces existen 10 pares de nodos con distancia 2 entre ellos ($hp_2 = 10$). Para entender mejor el ejemplo, la figura 4.3 muestra un grafo de 9 nodos y 8 aristas (las letras representan nodos y las líneas aristas), que muestra el caso de $hp_2 = 10$. En la imagen de la izquierda se puede ver que para el nodo A hay dos nodos a dos saltos (E y H). Luego se encuentra que para

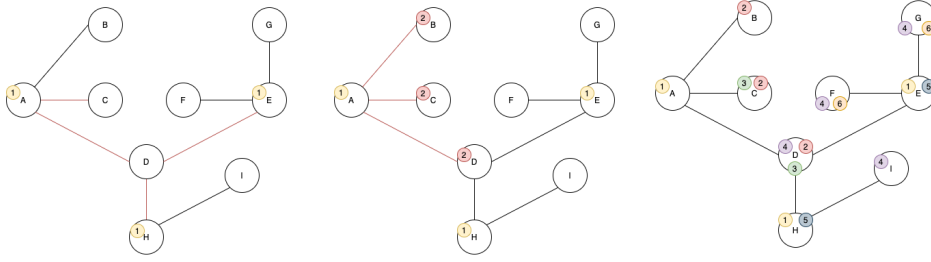


FIGURA 4.3: Grafo que posee diez pares de nodos que se encuentran a dos saltos de distancia. La numeración de los nodos muestra los pares de nodos. Por ejemplo, el nodo A tiene dos pares de nodos con distancia dos, reflejado a través del número 1.

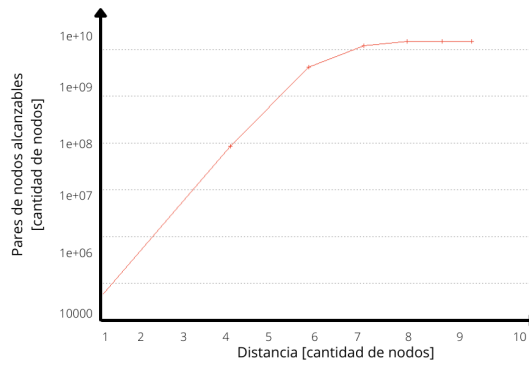


FIGURA 4.4: Ejemplo de un *Hop plot*: número de nodos alcanzables a cierta distancia entre ellos.

el vértice B, hay dos nodos a dos saltos (C y D). Luego si se sigue evaluando para el resto de nodos, se obtiene un *Hop plot* de 10, para $k = 2$.

Asimismo, en la figura 4.4 se puede observar el hop plot acumulado de un grafo con 10 mil millones nodos, donde el eje horizontal representa la distancia entre los nodos y el otro eje, los pares de nodos alcanzables. En el gráfico se muestra que la mayor cantidad de nodos se encuentra conectado a una distancia menor o igual a 7.

Para obtener un único valor que se pueda visualizar a través del tiempo, se considerará un promedio ponderado entre todos los nodos. Sea $H = \binom{N_v}{2}$ el número total de posibles caminos entre dos nodos, y sea D el diámetro del grafo (máxima distancia en un grafo), entonces el average hop plot de un grafo G está dado por:

$$HP = \frac{1}{H} \sum_{i=1}^D i * hp_i \quad (4.2)$$

Asumiendo una red con una gran cantidad de nodos, un valor bajo de HP implicaría que el grafo está altamente conectado o que todos los nodos están conectados al nodo central. En cambio, si se tiene un valor alto de HP , implica que los nodos no están muy conectados entre ellos, demostrando una lenta propagación de información.

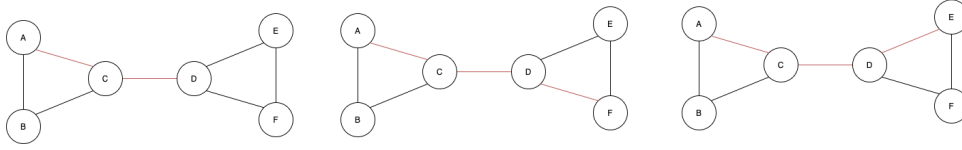


FIGURA 4.5: Grafo de seis nodos con $bc_C = 6$. Las rutas marcadas en rojo muestran los camino más corto entre los nodos A y D (izquierda), A y F (centro), y A y E (derecha).

pero una alta difusión. Es decir, la mayoría de los usuarios conocieron la noticia a través de uno de sus seguidores.

4.3. Average betweenness centrality

El *betweenness centrality* cuantifica la cantidad de veces que un nodo actúa como puente en todos los caminos más cortos [13]. Vale mencionar que el camino más corto es la ruta entre dos nodos con la menor cantidad de saltos, también llamado distancia geodésica, e implica que el nodo pertenece a ese camino más corto.

El *betweenness centrality* bc_i para un nodo v_i se calcula de la siguiente manera:

1. Para cada par de nodos (v_j, v_k) , con $v_j, v_k \neq v_i$, se obtiene todos los caminos más corto entre ellos.
2. Para cada par de nodos (v_j, v_k) , con $v_j, v_k \neq v_i$, determine la fracción de los caminos más cortos que pasen por el nodo v_i .
3. Sume las fracciones para cada par de nodos (v_j, v_k) .

Con apoyo de la figura 4.5, que cuenta con un grafo de 6 nodos y 7 aristas, se mostrará varias iteraciones del cálculo del *betweenness centrality* del nodo C. Para el vértice A se tiene los pares de nodos A-B, A-D, A-E y A-F; en donde para cada par se busca el camino más corto. Luego, se puede observar que solo en los pares de nodos A-D, A-E y A-F se encuentra el nodo C. Es decir, de los cuatro pares de nodos, solo tres incluye C en el camino más corto, por tanto, parte de la suma final es de $0 + 1 + 1 + 1$. Vale mencionar que como existe un único camino más corto para cada par de nodos mencionados, se suma la fracción completa en el caso que incluya a C. Posteriormente, al repetir este mismo proceso al partir del nodo B, se obtiene 3 nuevos rutas (B-D, B-E y B-F), donde el nodo C es puente, completando $bc_C = 6$.

El valor de *betweenness centrality* puede variar desde cero a la cantidad de pares de nodos que se están evaluando, donde en una red de difusión de información, el nodo que tenga un alto valor, es el usuario que más influencia tiene en el flujo de información. Esto implica que gran parte de la información pasa por ese nodo. Por el contrario, el usuario que tenga un bajo valor, implicará que la difusión de información no está centralizada en él, por tanto, existe una mayor dispersión de la información.

Para efectos prácticos de este trabajo, se utilizará el *average betweenness centrality* sobre el conjunto de nodos $\mathbf{V}$, lo cual se puede expresar de la siguiente manera:

$$BC = \frac{1}{|\mathbf{V}|} \sum_{i=1}^{|\mathbf{V}|} bc_i \quad (4.3)$$

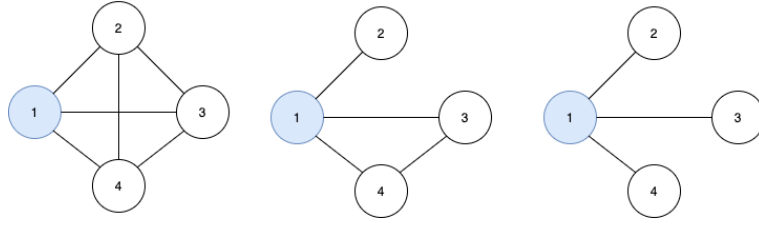


FIGURA 4.6: Grafos demostrativos de *cluster coefficient* del nodo uno con valores $\frac{3}{3}$, $\frac{1}{3}$ y $\frac{0}{3}$

Vale mencionar que para un bajo valor de *BC*, implicaría que se espera que hayan muchos usuarios con muy baja influencia, lo que implicaría que toda la información la concentra el tweet central, o todos los usuarios están fuertemente conectados. En cambio si se encuentra con un valor alto, significaría que hay muchos usuarios que influyen en gran medida a la difusión de la información por el grafo, por lo cual, los usuarios están difundiendo la información solo a un número limitado de sus contactos.

4.4. Average cluster coefficient

El *cluster coefficient* mide la probabilidad de que dos nodos que comparten un vecino estén conectados entre sí, es decir, que se forme un triángulo [13]. Sea c_i el *cluster coefficient* del nodo v_i , entonces se calcula a través de:

$$cc_i = \frac{\Delta_i}{\binom{d_i}{2}} \quad (4.4)$$

donde Δ_i es el número de triángulos cerrados para el nodo v_i y $\binom{d_i}{2}$ es el número de posibles triángulos que puede generar el nodo i .

Por ejemplo, en la figura 4.6 se observan tres grafos con 4 nodos, pero con 6, 4 y 3 aristas (de izquierda a derecha). En la imagen izquierda se puede observar que el nodo 1 en la primera imagen tiene un *cluster coefficient* de $\frac{3}{3}$, dado que logra cerrar todos los triángulos que podría formar con el resto de nodos. Es decir, de los $\binom{3}{2}$ posibles triángulos (1-2-3, 1-2-4, 1-3-4), todos los triángulos se han formado. Luego, para la imagen del centro, se observa que tiene un *cluster coefficient* de $\frac{1}{3}$ dado que de los tres posibles triángulos, solo cierra uno. Por último, en el grafo de la derecha, el nodo 1 tiene un *cluster coefficient* de 0, dado que no cierra ninguno de los posibles triángulos.

El cluster coefficient de cada nodo varía entre 0.0 y 1.0, donde 0.0 significa que no existe ningún triángulo, es decir, existe poca agrupación entre los nodos. Caso contrario, un valor de 1.0 implica que todos sus vecinos están conectados entre sí, mostrando un gran flujo de información.

Para el análisis del grafo, se utilizará el *average cluster coefficient*, que es el promedio de los *cluster coefficients* calculados. Dado un grafo con N_v vértices, su expresión es la siguiente:

$$CC = \frac{1}{N_v} \sum_{i=1}^{N_v} cc_i \quad (4.5)$$

Un alto valor de *average cluster coefficient* implicaría que el grafo esté agrupado e interconectado entre la mayoría de los nodos. En el caso de la difusión de información, se encontraría que las personas que retweetean la publicación, son usuarios que se siguen. En cambio, si se obtiene un bajo valor, implica que hay una baja interconexión entre los nodos.

Capítulo 5

Análisis

A continuación se presentarán los distintos análisis de las redes formadas por trends (figura 5.1, 5.13 y 5.25) y *funas* (figuras 5.7, 5.19 y 5.31). Mediante las siguientes características: *degree distribution*, *hop plot*, *cluster coefficient* y *betweenness centrality*, se estudiará la forma en que se difunde la información en Twitter. Se examinará la red a medida que transcurre las horas y la creación de aristas durante una semana.

En primera instancia, se analizó el tweet acerca de la selección chilena de fútbol. La noticia se viralizó por el sumario levantado al equipo, cuando estaban compitiendo por la copa América en Brasil. La red cuenta con alrededor de 500 usuarios y 1,000 aristas, su difusión se debe principalmente a un usuario influyente que propaga el tweet a varios usuarios independientes. Algunos de ellos, logran difundir el trend a otros grupos, que no están directamente relacionados al nodo principal.

Luego, se analizó la *funa* a Patricio Maturana, quien fue imputado de lesiones graves contra Fabiola Campillai. Esta *funa* nace bajo el contexto del levantamiento de la medida cautelar de prisión preventiva y su sustitución por el arresto domiciliario, mientras transcurría el proceso penal. La red formada cuenta con 870 usuarios y alrededor de 12,700 aristas. Inicialmente la difusión de la *funa* a las personas fue debida a un usuario principal. La mayoría estaban relacionados en forma de anillo o cadena, es decir, un usuario le informa a un amigo, y luego este a un amigo suyo y así sucesivamente. Cuando algunas personas propagaron la *funa* a grupos alejados del influyente, se observa cambios en el comportamiento. Se forma un grupo principal y cohesionado, compuesto por varias personas responsables de la difusión.

El siguiente trend analizado tiene como contexto la entrevista de Mónica Rincón a un ex ministro. El trend reproduce el comentario de la periodista cuando el entrevistado desliza un político imputado en el caso SQM¹ por emisión de boletas falsas. La red resultante cuenta con 1,797 personas y 42,438 aristas. Inicialmente la difusión se debe a usuarios principales. Si bien la mayoría de las personas solo se informa, algunas logran difundir en forma de cadena. Pronto la forma de propagación cambia: es un gran grupo central y cohesionado, conformado por usuarios que divulgan la publicación. Algunos de los que fueron influyentes, informan a conjuntos de usuarios que no logran propagar el trend. Se destaca que el comportamiento resultante es similar al de la *funa* anterior.

La siguiente *funa* analizada se dirige contra Edwin Dimter, quien es autor del secuestro y homicidio de Victor Jara² durante la dictadura militar. Actualmente la ejecución

¹Sociedad Química y Minera de Chile

²cantautor chileno

de su sentencia se encuentra suspendida, de ahí surge la censura pública. La red resultante consta de 594 personas y 7,779 aristas. Al inicio son los usuarios influyentes que difunden la información a las personas, y solo algunas de ellas logran propagar en forma de cadena. Posteriormente, se forma un grupo central e interconectado, en donde las personas influyen en la difusión de la *funa*. También se observa ciertos grupos a la periferia, que no propagan la publicación. La forma en que se difunde la información, es similar a la de la primera *funa* analizada y al trend relacionado con el ex ministro.

Después de analizó el trend de José Antonio Kast, político y candidato presidencial de Chile 2022. El tweet obtenido hace referencia a la despenalización del aborto hasta las 14 semanas, que Kast no apoya. La red resultante cuenta con 1,173 usuarios y 88,310 aristas. En un principio, la difusión se debe principalmente por usuarios influyentes que divulgan la información. Las personas pueden informarse a través de uno o varios de ellos. Prontamente, se forma un grupo central altamente cohesionado, cuya propagación se debe por la mayoría de los usuarios, amenguando el protagonismo de los influyentes. De igual manera siguen existiendo conjuntos de usuarios que solo se informan. El comportamiento llama la atención dada las abruptas variaciones de las características y la alta cantidad de relaciones que hay por usuario.

La última *funa* analizada se dirige contra a un grupo de manifestantes que intentaron herir a carabineros con fuegos artificiales en el contexto de una marcha en el *estallido social*. En comienzo la difusión se debe por varios usuarios influyentes. Las personas se enteraban de la *funa* a través de uno o varios de ellos, hasta que se interconectaron y empezaron a propagar también. Se formó un gran grupo central cohesionado, conformado por usuarios que difunden la *funa*. Los influyentes perdieron protagonismo, aunque igual seguían informando a conjuntos de usuarios independientes. Similar al tweet de Kast, se presencia cambios abruptos de las características, y una excesiva cantidad de relaciones por usuario.

5.1. Trend de la selección chilena de fútbol

El tweet tendencia surge en el contexto de la Copa América Brasil 2021 cuando la selección de fútbol de Chile se encontraba compitiendo. El escandaloso hecho fue el sumario levantado al equipo tras romper la burbuja sanitaria ingresando un peluquero brasileiro al hotel donde se hospedaban. Esto se filtró por las fotos subidas a las redes sociales de algunos futbolistas, como Gary Medel y Arturo Vidal³. En Twitter se hizo tendencia el hashtag #SelecciónChilena, y uno de los tweet con más retweet fue el analizado: "*La esposa de Bravo siempre tuvo razón. #SelecciónChilena*". Fue publicado el 20 de junio de 2021 a las 14:07, en el cual se logró capturar alrededor de 500 retweet y 1000 aristas aproximadamente. Una visualización de la red formada se observa en la figura 5.1. Vale destacar que en las primeras 20 horas, se realizaron el 98.6 % de las relaciones, es decir, casi el total de la difusión. Sin embargo, hubieron usuarios que retwitaron aisladamente luego de 28 horas.

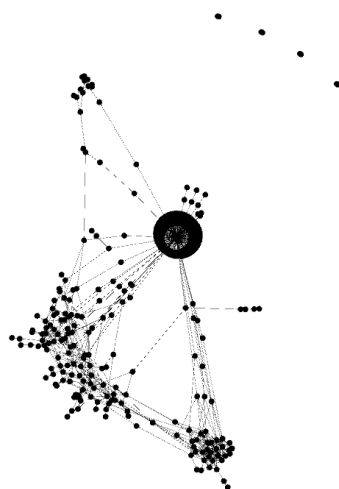
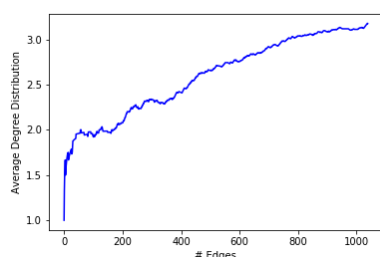
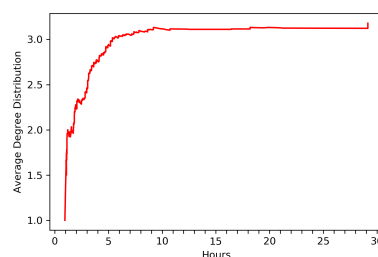


FIGURA 5.1: Red resultante del trend de la selección chilena de fútbol



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.2: Gráficos de *Average Degree Distribution* de la red tendencia de la selección chilena de fútbol

5.1.1. Análisis de características: *Average Degree Distribution*

En la figura 5.2 se visualiza dos gráficos correspondientes al cambio del *average degree distribution* según la creación de aristas (figura izquierda 5.2a), y el paso de las horas (figura derecha 5.2b). En la primera gráfica (figura 5.2a), se observa que el *average degree distribution* aumenta de 1 a 3 conexiones en promedio. Desde el inicio hasta las 100 aristas, existe un fuerte aumento de 1 a 2 conexiones, lo que podría indicar que el trend es re-retweeteado por otras personas (el retweet no es del original). Esto se podría interpretar que la difusión de información es en forma de cadena o línea de usuarios conectados. Dada la baja cantidad de relaciones promedio, se podría decir que las primeras personas que difundieron la información son de diferentes grupos. Una segunda forma de difusión que podría ser explicado por el bajo *average degree distribution*, es la de tipo estrella: todos retweetean la información de un usuario principal.

En el mismo gráfico (figura 5.2a), se mantiene un *average degree distribution* de 2 hasta las 200 aristas. Luego aumenta a 3 conexiones promedio en un ritmo paulatino hasta

³<https://www.emol.com/noticias/Deportes/2021/06/21/1024502/hubo-otro-peluquera-la-roja.html>

las 800 aristas. Esto podría indicar que las personas que ya retweetearon comenzaron a generar nuevas conexiones. O bien, que los usuarios que se hallaban aislados, empezaron a conectarse entre ellos tras el ingreso de nuevos nodos a la red.

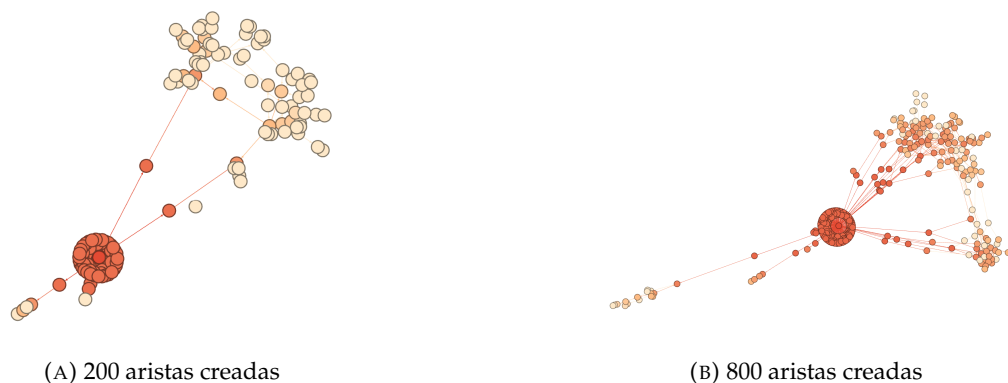


FIGURA 5.3: Red tendencia de la selección chilena de fútbol en distintas instancias temporales

Con el fin de corroborar las hipótesis planteadas, se utilizarán las figuras 5.3a y 5.3b. En ellas se observa la red con alrededor de 200 (izquierda) y 800 (derecha) aristas. Los nodos tienen un degradado de rojos. El color rojo intenso indica que están directamente conectados al usuario principal de la difusión, mientras que el más claro son los que están a más saltos de separación. Por tanto, es esperable que usuarios más claros hayan sido informados por otros usuarios, distinto al principal.

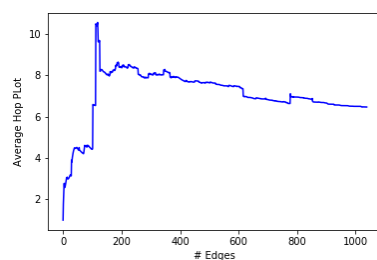
En la figura 5.3a (izquierda) se identifica un usuario como el principal difusor de información de la red, al tener varios usuarios conectados a este, formando la figura estrella previamente mencionada. Se observa que la mayoría están conectados directamente al nodo principal, y hasta el momento no han causado un impacto en la difusión al no lograr propagar el trend a otras personas. Sin embargo, se nota la existencia de ciertos usuarios que logran alcanzar nuevos grupos (nodos de color oscuro y rojo claro), formando una cadena de usuarios difusores de la información.

Luego, en la figura 5.3b (derecha) se observa que el nodo influyente ha difundido la información a más usuarios. Por tanto, dicho usuario sigue siendo predominante en la difusión del tweet, destacándose la figura estrella que forma. De igual forma, se muestra un aumento de usuarios que logran difundir la información a otros grupos, manteniendo la existencia de ciertas cadenas de usuarios difusores. Dichas personas lograron ampliar la red, alcanzando grupos diferentes a los que el nodo principal hubiera llegado directamente. Estos grupos se pueden identificar por estar agrupados a la periferia de la estrella, y por tener nodos de color rojo claro, lo cual indica que se enteraron del trend a través de sus amigos. En la figura existen cuatro grupos distintos, además del grupo asociado al nodo principal: el que tiene mayor cantidad de nodos a la izquierda, uno más pequeño en el lado inferior izquierdo, y los dos más diminutos que están a las derecha. Dentro y entre los dos grupos más grandes de la izquierda, se observa cierta conectividad similar a la que se observa dentro del grupo principal.

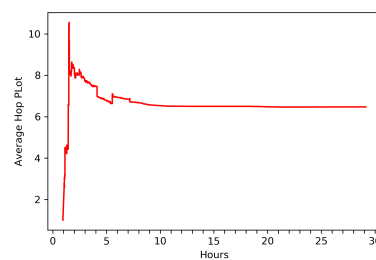
En el gráfico por hora (figura 5.2b), se observa un crecimiento de 1 conexión promedio a 3 en solo 5 horas, donde el resto de las 12.5 horas se mantiene estable. Esto muestra que la información se difundió con bastante rapidez. Por ejemplo, a las 2

horas ya existían 100 aristas en la red (average degree sobre 2 al ver figura 5.2a), lo que implica que una arista se forma cada 36 segundos en promedio. Adicionalmente, replicando el análisis a la quinta hora se formaron 600 nuevas aristas, lo que implica una nueva arista cada 18 segundos. Dado que la mayoría de las aristas de la red (800) se generan dentro de las primeras 5 horas, se podría decir que en aquel tiempo fue el momento de mayor difusión del tweet. Dentro de ese tiempo, las primeras dos horas fueron destinadas a que el nodo principal difundiera la información por diferentes grupos. Luego, en las tres horas restante la red empezó a interconectarse, formando la red principal. Finalmente, en las últimas horas se agregaron nuevas aristas que, al parecer, no modificaron en forma significativa la estructura de la red.

5.1.2. Análisis de características: *Average Hop Plot*



(A) Según la cantidad de aristas creadas



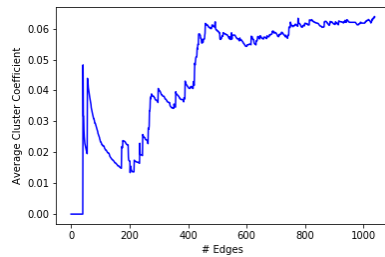
(B) Según las horas transcurridas

FIGURA 5.4: Gráficos de *Average Hop Plot* de la red tendencia de la selección chilena de fútbol

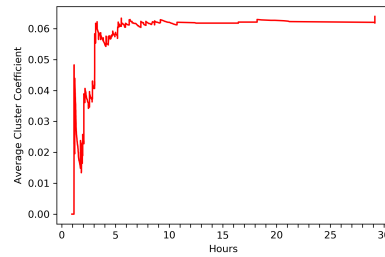
La figura 5.4 muestra el *average hop plot* según el aumento de la cantidad de relaciones (figura izquierda 5.4a), y según el crecimiento de la red en horas (figura derecha 5.4b). La figura 5.4a muestra que desde 0 a las 75 aristas creadas, la distancia promedio crece de 0 a 4 nodos de separación. Esto indica cercanía entre los usuarios, por lo cual es posible que la manera de propagación sea una combinación de cadenas de nodos y estrella. Un usuario principal debe otorgar la cercanía entre los nodos, y las relaciones tipo cadena, la separación. Al momento que la red tiene 175 aristas, se presenta un crecimiento del *average hop plot* hasta 10, lo que indica una mayor separación entre los usuarios. Se explica que el trend llega a varias personas de diferentes grupos, que no están conectados directamente al nodo central. En otras palabras, la información se difunde más allá del influyente. Entre las 200 y 1,000 aristas, se visualiza una decrecimiento entre 8 y 6 saltos de distancia promedio. Dicho comportamiento podría exponer la idea de que luego de propagar la información por diferentes grupos, los usuarios empiezan a conectarse entre ellos, disminuyendo la distancia promedio. Vale destacar que se cumple la propiedad del modelo de Mundo Pequeño de *Watts and Strogatz*, mencionado con anterioridad, obteniendo que cada usuario este separado por 6 personas en promedio.

La figura 5.4b representa el comportamiento del *average hop plot* por hora. Se observa que aproximadamente a la hora y media, la red formada alcanza su máxima distancia promedio de 10 nodos, alcanzando a diferentes grupos de personas que no estaban relacionadas directamente con el nodo central. Luego a las 5 horas ya se acerca al valor de 6 saltos promedio, al cual se estabiliza el resto de horas. Este comportamiento coincide con el análisis del *average degree distribution*, dado que luego de la hora y media, los usuarios se empiezan a relacionar entre ellos alcanzando una distancia menor.

5.1.3. Análisis de características: *Average Cluster Coefficient*



(A) Según la cantidad de aristas creadas



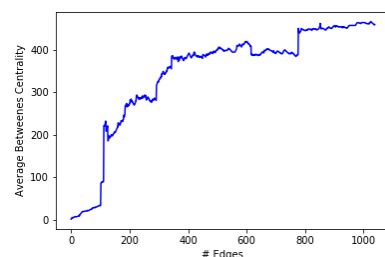
(B) Según las horas transcurridas

FIGURA 5.5: Gráficos de *Average Cluster Coefficient* de la red tendencia de la selección chilena de fútbol

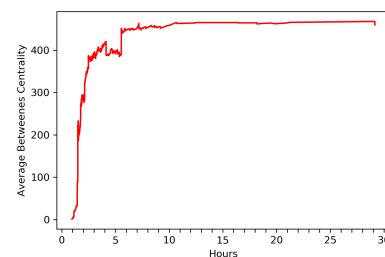
La figura 5.5 muestra el *average cluster coefficient* ($\bar{cc}$) versus la cantidad de aristas en la red, y según las horas transcurridas desde el primer retweet capturado. La figura 5.5a muestra que hasta las 50 aristas formadas, hay una cohesión nula entre los nodos. De lo cual se puede concluir que la red en un principio tiene una forma de estrella, cadena o de anillo, tal como se mencionó en el análisis del *average degree distribution*. Luego, desde las 50 a las 200 aristas, $\bar{cc}$ varía entre un 5 % y un 1.5 %. Esto implica que la red comenzó a generar aristas entre nodos existentes, logrando unificar en cierta medida la red. Sin embargo, al bajar su valor, la red se expande y los nuevos no se relacionan con los nodos existentes, o bien, solo se informan. Después de llegar a la depresión de 1.5 %, se empieza a visualizar un aumento hasta un 4 % de agrupación, lo que nuevamente implica la generación de aristas entre nodos cercanos. Finalmente, en las 450 aristas, la agrupación fluctúa alrededor de un 6 %, donde se estabiliza al alcanzar las 800 aristas. Esto podría indicar que las nuevas relaciones ayudan tanto en la expansión como en el cierre de triángulos en la red.

Por otro lado, la figura 5.5b muestra un crecimiento continuo de un 0 % a un 6 % del *average cluster coefficient* en las primeras 5 horas. Esto implica que las interconexiones dentro la red se forman rápidamente en dicho periodo. En el resto de horas, las nuevas personas que difunden la red mantienen el mismo porcentaje de agrupación, por lo que no se siguió difundiendo entre más gente conocida.

5.1.4. Análisis de características: *Average Betweenness Centrality*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.6: Gráficos de *Average Betweenness Centrality* de la red tendencia de la selección chilena de fútbol

En la figura 5.6 se muestra dos gráficas del *average betweenness centrality* (BC), la primera muestra su variación según la cantidad de aristas agregadas a la red, y la otra

según las horas transcurridas. En la figura 5.6a, se observa que hasta las primeras 100 aristas agregadas, el promedio de actuar como puente va de 0 a 50 veces. Esto indica la existencia de algunos nodos relevantes en la difusión de información, dado que para llegar de una persona a otra, es necesario pasar por estos nodos. Entonces, la mayoría de los usuarios no son necesarios para difundir el trend. En coincidencia con los análisis previos, también se difunde formando la estrella. Luego, alrededor de las 125 aristas formadas en la red, se alcanza rápidamente un BC de 200 veces. Esto muestra como varios de los nodos actúan como puentes, en comparación a la cantidad de aristas que existen en la red. Por tanto, a medida que se agregan nuevos nodos, existen más usuarios influyentes en la difusión. Según los descubrimientos previos, también hace sentido que la información comienza a difundirse a partir de un nodo central y posteriormente, se propaga a través de más personas. Finalmente, el valor se estabiliza luego de las 400 aristas, en este caso se asume un nuevo comportamiento de la red. Las nuevas relaciones empiezan a unificar la red, por lo tanto, hay nuevos caminos que unifican los diferentes grupos. Esto podría indicar que se mantuvieron los mismos influyentes, y que los nuevos usuarios unificaron los grupos.

Al analizar la figura 5.6b, se observa un crecimiento dentro de las dos primeras horas y media hasta un BC de 400, lo que coincide con la creación de 400 aristas en la red. Luego, a las 5 horas con 800 aristas, se estabiliza en las 450 veces. Además, se puede observar que a las dos horas se alcanza un índice de 200, lo cual muestra el ingreso de nuevos usuarios relevantes para la difusión de información. En las tres horas siguientes, ingresan usuarios que solamente se informan a través de estos nodos relevantes. Finalmente, se estabiliza en un BC cercano a 450, lo cual indicaría el momento en que los usuarios empiezan a unificar la red, agregando nuevos caminos entre ellos. Este comportamiento también se ve reflejado en los análisis anteriores.

5.1.5. Recapitulación de análisis

Recapitulando los análisis previos, se caracteriza la difusión de información con lo siguiente: la existencia de usuarios cruciales que propagan la información a diferentes personas y grupos, y la interconexión entre las agrupaciones formadas. Se inicia con un *average degree distribution* de 2 conexiones, un *average hop plot* de 4 saltos, un *average cluster coefficient* de 0% y un *average betweenes centrality* de 50 veces, lo que indica un comportamiento de estrella inicial. Es decir, existe un nodo central que difunde la información al resto de usuarios. Luego, entre las 100 y 200 aristas formadas en la red, todas las características sufren un aumento en sus índices: un *degree distribution* posee un valor de 2 conexiones, un *hop plot* de 8 saltos, un *cluster coefficient* de 1.5% y un *betweenes centrality* de 200 veces. En dicho momento, algunos nodos empiezan a ganar relevancia en la difusión de la información, logrando ampliar la red al alcanzar grupos que no están directamente relacionados con el nodo principal. Luego, las características siguen aumentando hasta las 800 aristas: el *average degree distribution* es de 3 conexiones, el *average hop plot* es de 7 saltos, el *average cluster coefficient* es de 6% y el *average betweenes centrality* es de 400 veces. A este punto la red comienza a conectar nodos ya existentes debido al ingreso de nuevos usuarios, y de igual forma se mantienen los usuarios que logran difundir la información a grupos que no están directamente relacionados al nodo principal.

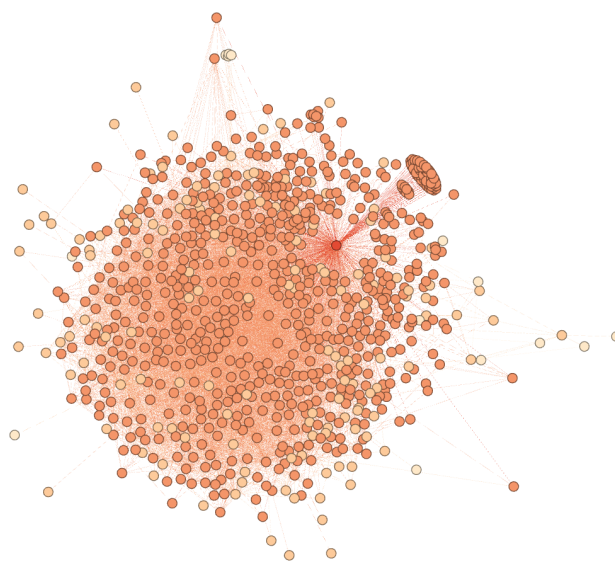


FIGURA 5.7: Red resultante del funa a Patricio Maturana

5.2. Funa a Patricio Maturana

Será analizada la *funa* contra el ex capitán de Carabineros de Chile, Patricio Maturana, quien está imputado por apremios ilegítimos con lesiones graves contra Fabiola Campillai. La situación data en el año 2019 (estallido social de Chile⁴), cuando la mujer pierde la visión, el gusto y olfato por el impacto de una lacrimógena lanzada por Patricio Maturana. Se posee como pruebas la cámara que llevaba el ex comandante, y como testigos están los compañeros del imputado y civiles, siendo uno de ellos la hermana de Fabiola⁵. Al momento que fue levantado el caso, el ex oficial fue llevado a prisión preventiva, sin embargo, en el mes de julio del año 2021 la corte revoca su prisión preventiva para que cumpla con arresto domiciliario⁶. El tweet encontrado presenta una imagen de Patricio Maturana con el siguiente escrito: “Se busca Patricio Maturana Mutilador. Si no hay justicia, hay funa”. Esta publicación fue realizada el 2 de agosto del 2021 a las 13:29.⁷ La red resultante cuenta con 870 usuarios y 12,682 aristas, la cual se muestra en la figura 5.7. Similar al trend anterior, solo se muestra el 97 % de las aristas formadas, dado que el resto de usuarios retweetearon al cuarto o quinto día.

5.2.1. Análisis de características: *Average Degree Distribution*

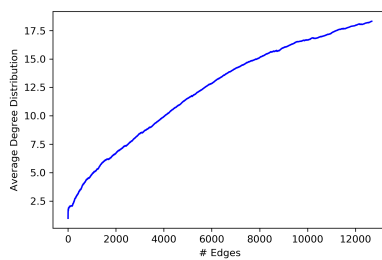
En la figura 5.8 se observa dos gráficas de *average degree distribution* de la *funa*, mostrando su variación según la cantidad de aristas agregadas a la red, y las horas transcurridas. En la imagen 5.8a, se observa que a las 100 aristas creadas la red ya tiene

⁴masivas manifestaciones y disturbios originados en Santiago y propagados a todas las regiones de Chile

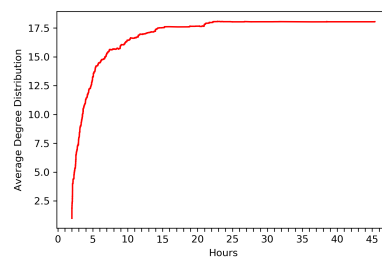
⁵<https://www.latercera.com/la-tercera-pm/noticia/patricio-maturana-los-rastros-que-dejo-el-capitan-de-carabineros-imputado-por-dejar-ciega-a-fabiola-campillai/CHSRPIZ7DJBL3HXMVLMINIJBHE/>

⁶https://www.cnnchile.com/pais/revocan-prision-preventiva-carabinero-fabiola-campillai_20210730/

⁷<https://twitter.com/EPInforma/status/1422248209238433795>



(A) Según la cantidad de aristas creadas

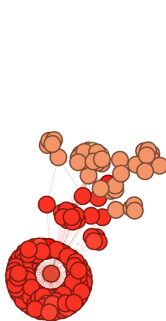


(B) Según las horas transcurridas

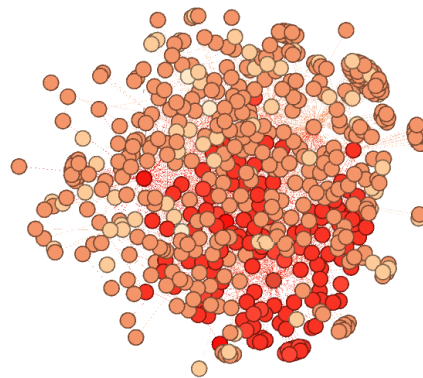
FIGURA 5.8: Gráficos de *Average Degree Distribution* de la red *funa* a Patricio Maturana

alrededor de 2.5 conexiones promedio por usuario. Posteriormente, a medida que aumenta la cantidad de aristas, se va incrementando las conexiones promedio. Se obtiene un *average degree distribution* de 5, 9, 15, y 17.5 para las 1,000, 4,000, 8,000, y 12,000 aristas respectivamente. Una posible hipótesis es que mientras incrementa la cantidad de aristas y usuarios en la red, pierden protagonismo los nodos principales, dado que los distintos usuarios que conforman la red propagan la *funa*. En otras palabras, los usuarios principales jugaron un papel importante en un comienzo, difundiendo la información a diferentes personas de distintos grupos, hasta un punto que los nodos no solamente se informaban, sino que lograron informar dentro de su mismo grupo. Dada la cantidad final de conexiones promedio por usuario alcanzadas (17.5), claramente existe una gran cantidad de relaciones entre las personas que participaron en el proceso. Mostrando una clara diferencia con la red previamente analizada (subsección 5.1).

La figura 5.8b muestra que las primeras 100 conexiones de la red se generaron antes de la primera hora. Además, se observa que a las 5 horas ya se tenía 8,000 aristas en la red (más de la mitad), dado que coincide con las 15 conexiones promedio por nodo. En otras palabras, se creó 26.66 aristas por minuto dentro de las primeras 5 horas, aproximadamente. Luego, se observa una estabilización del *average degree distribution*, similar a lo observado en el análisis del grafo anterior (subsección 5.1). En resumen, para esta red, las primeras 5 horas es donde se produce la mayor difusión de la información entre usuarios que están interconectados entre ellos. Posteriormente, la red crece por conexiones más esporádicas.



(A) 200 aristas creadas



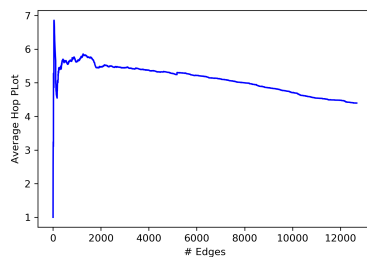
(B) 2,000 aristas creadas

FIGURA 5.9: Red *funa* contra Patricio Maturana en distintas instancias temporales

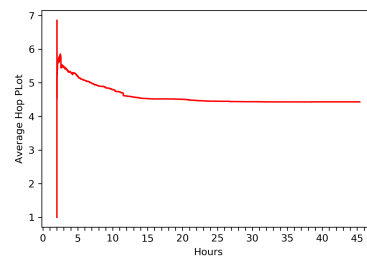
Las imágenes 5.9a y 5.9b muestran la red en los primeros minutos cuando posee 200 y 2,000 aristas, respectivamente. En la figura 5.9a, se observa el usuario original de la difusión, que informa a los nodos de color rojo que lo rodean. Además, se observa cierta agrupación entre los usuarios de la periferia del nodo principal. Por tanto, es posible que las personas se hayan enterado de la funa por el usuario influyente u otros. Además, se observan nuevos grupos de nodos desconectados del usuario central, expandiendo la noticia a nuevos grupos de personas.

En la figura 5.9b se observa un cambio de comportamiento, dado que se redujo la cantidad de grupos y aumentó la interconexión entre los usuarios. Se muestra un grupo principal, por el cual se difunde la funa a través de distintas personas. En la periferia de dicho grupo, se observan aglomeraciones de usuarios conectados entre sí, lo que implicaría la difusión de la información a nuevos grupos de usuarios. Por tanto, existen nodos que informan a gran parte de los usuarios, además de grupos más pequeños y cerrados.

5.2.2. Análisis de características: *Average Hop Plot*



(A) Según la cantidad de aristas creadas



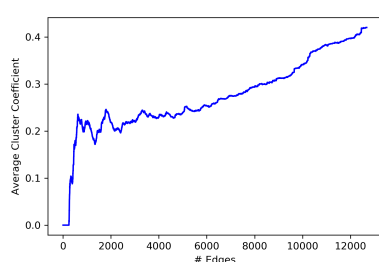
(B) Según las horas transcurridas

FIGURA 5.10: Gráficos de *Average Hop Plot* de la red funa contra Patricio Maturana

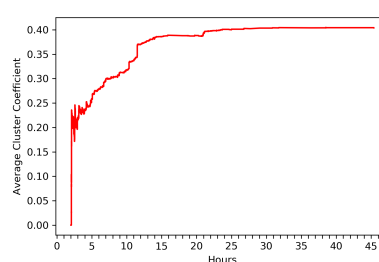
En las figuras 5.10a y 5.10b se muestran dos gráficas del *average hop plot* según la cantidad de aristas creadas en la red (izquierda), y el tiempo transcurrido (derecha). En la primera figura mencionada, se observa que la mayor distancia entre los usuarios se alcanza al comienzo de la propagación de la funa. Alrededor de las 100 y 200 aristas, se indica una distancia de 7 saltos promedio entre los usuarios. Puede coincidir con la idea de que los nodos estén relacionados con forma de anillo, cadena, o árbol. Se puede postular la idea de que las personas están alejadas unas de otras y con una baja interconexión entre ellas. Entonces, en un comienzo la funa, alcanza a diferentes personas de distintos grupos, que van conociendo el tweet por alguien que siguen y no por quien hizo la publicación. También, se plantea la idea de la existencia de un nodo o nodos inicializadores de la difusión que en un inicio informan a las personas, quienes luego notifican a su grupo en forma de cadena o anillo. Sin embargo, luego de alcanzar esos 7 saltos de separación promedio, se reduce fuertemente hasta llegar a 4.5. En este pequeño rango de crecimiento de la red, se observa que los nuevos usuarios agregados, empezaron a unificar a las personas de la red. Sin embargo, rápidamente este valor vuelve a subir alrededor de 5.5 saltos de distancia (2,000 aristas), implicando una nueva ampliación del grupo de usuarios. Finalmente, la distancia esperada entre los nodos empieza gradualmente a descender hasta alcanzar 4 saltos promedio en las 12,000 aristas. Esto implica que se siguen agregando aristas a la red que conectan grupos de usuarios desconectados entre ellos, disminuyendo la distancia entre los nodos.

En la figura 5.10b (derecha) se observa que se alcanza el máximo *hop plot* alrededor de la primera hora. Es el momento de más amplitud de la red, por tanto, la *funa* se divulga por distintas personas de distintos grupos. A pesar de que coincide con el punto de menor interconexión, la red se comporta según la propiedad del *Mundo Pequeño*. Luego, la distancia va disminuyendo dentro de las siguientes 9 horas hasta alcanzar 4.5 personas promedio, lo cual coincide con la creación de 10,000 aristas. Esto indica que la red se empieza a caracterizar por personas que se enteran de la *funa* a través de más de un amigo, conectando grupos de personas dentro de la red. Finalmente, en las horas restantes, la red lentamente se estabiliza en una distancia cercana a 4 personas promedio, sin presentar un cambio significativo en la forma en que se difunde la *funa*.

5.2.3. Análisis de características: *Average Cluster Coefficient*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

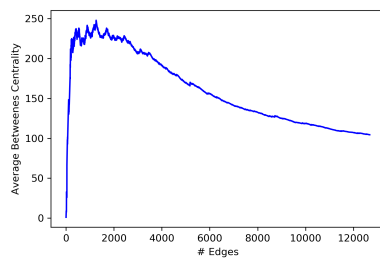
FIGURA 5.11: Gráficos de *Average Cluster Coefficient* de la red *funa* contra Patricio Maturana

En las figuras 5.11a y 5.11b se observan los gráficos que muestran el *average cluster coefficient* versus la cantidad de aristas creadas en la red (izquierda), y según las horas transcurridas (derecha). En la figura de la izquierda se muestra que la red inicia con una cohesión nula, lo cual implica que las primeras personas que se han enterado de la *funa* no se conocen entre ellos. Esto apuntaría a que la red posee cierta forma de estrella, anillo, línea, o bien, combinaciones de ella. Por tanto, podría existir usuarios principales que difunden la información, donde luego las personas continúan informándole a amigos. Luego, entre las 200 y 900 aristas, se observa un rápido aumento a un 25 % de *clustering coefficient* promedio, lo que indica una fuerte interconexión de los nuevos usuarios que retweetearon la *funa*. Por lo visto, existe una tendencia de que los usuarios se informen a través de retweets de dos o más amigos, en vez de encontrar la presencia de nodos difusores cuyos seguidores no están conectados entre sí (como en el caso del trend de la selección chilena de fútbol de la sección 5.1). Entre las 1,000 y 4,000 aristas el $\bar{cc}$ fluctúa entre un 25 % y 19 %. Por tanto, existe una mezcla entre la aparición de usuarios importantes que difunde la *funa* a otros grupos, y la difusión entre grupos de amigos. Finalmente, posterior a las 4,000 y 12,000 aristas, se observa un crecimiento desde un 24 % a un 40 % de $\bar{cc}$. Dicho incremento en el $\bar{cc}$, se podría explicar por la llegada de nuevos usuarios que conectan fuertemente los grupos existentes. De igual manera, pero en menor medida, deben existir usuarios que no pertenecen a los grupos recién señalados.

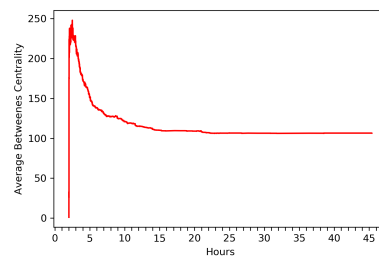
La figura 5.11b muestra que dentro de la primera hora se alcanza un $\bar{cc}$ de un 23 % aproximadamente, el cual aumenta a un 30 % en las 4 horas siguientes, e incrementa paulatinamente en un 41 % en las últimas horas restantes. En las primeras 4 horas se presencia la transición del comportamiento de la red. Parte con una agrupación

cercana a un 0 %, y aumentar rápidamente a un 23 %. Esto podría indicar que la *funa* se difunde entre personas de distintos grupos, los cuales, rápidamente, comienzan a formar grupos cerrados entre ellos. Por tanto, gana mayor protagonismo la difusión por varias personas de la red. En las siguiente 4 horas, debido al aumento importante del $\overline{cc}$, se podría estar en presencia de usuarios que empezaron a interconectar los distintos grupos. Sin embargo, en las siguientes 5 horas se observa un estancamiento del $\overline{cc}$, que podría deberse al ingreso de nuevos usuarios conectados a pocos nodos, mientras se mantiene la interconexión debido a otros. Finalmente, en las 15 horas restantes, el $\overline{cc}$ aumenta paulatinamente hasta un 41 % aproximadamente. En dichas horas no se presencia un nuevo cambio del comportamiento ya descrito. Sigue aumentando el divulgación de la *funa* a través de las personas de un mismo grupo (aumentando la interconexión entre nodos y por ende el $\overline{cc}$), y una pequeña participación de nodos que difunden la información a otras personas que solo logran informarse.

5.2.4. Análisis de características: *Average Betweenes Centrality*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.12: Gráficos de *Average Betweenes Centrality* de la red funa contra Patricio Maturana

Las figuras 5.12a y 5.12b muestran el *average betweenes centrality* (BC) versus la cantidad de aristas creadas en la red (izquierda), y según las horas transcurridas (derecha). En la imagen de la izquierda se observa que dentro de las 100 primeras aristas, el BC aumenta rápidamente de 0 a 225. Esto podría implicar que la red tiene formas de línea o cadena, dado que se tendría que pasar por los mismos usuarios para llegar de una persona a otra. Una segunda opción podría ser que ciertos usuarios difunden la información a diferentes grupos de personas, donde dichas personas continúan difundiéndola *funa*. Hasta las 2,000 aristas creadas, la cantidad de veces que un nodo actúa como puente fluctúa alrededor de las 225 veces. La fluctuación se puede deber a que la red debe empezar a tener más interconexiones entre las personas, disminuyendo el valor del BC, y la existencia de nuevos nodos en los extremos que aumentan el BC. Finalmente se observa un decrecimiento hasta las 12,000 aristas, alcanzando un valor de 100. Este decrecimiento indica que la interconexión gana importancia en la red, por tanto, los nuevos nodos conectan distintos grupos de usuarios, generando nuevas opciones para encontrar el camino más corto, y no necesariamente por el o los nodos principales de la difusión.

En figura 5.12b, se observa que a las 2 horas se alcanza el *average betweenes centrality* de 225, coincidiendo con la llegada de los usuarios principales que difunden la *funa*. Luego, el descenso más pronunciado se presencia hasta las 5 horas, donde el BC alcanza un valor de 125. El cambio de comportamiento en esas 4 horas se podría deber a un aumento en la interconexión en la red y un incremento en la importancia

de más usuarios en la red para difundir la funa. En las siguiente 5 horas hay un descenso más paulatino hasta alcanzar un valor de BC igual a 100. En este punto, se debe tener un número reducido de grupos o un grupo, dado que los usuarios deben estar fuertemente conectados. En las horas restantes, los usuarios que ingresaron a la red no tienen un comportamiento distinto al comentado previamente, dado que el valor del *average betweenes centrality* se mantiene relativamente constante.

5.2.5. Recapitulación de análisis

Recapitulando los análisis previos de *average degree distribution*, *average hop plot*, *average cluster coefficient* y *average betweenes centrality* se puede observar el comportamiento dinámico de la red durante sus 47 horas. En un comienzo cuando la red alcanza a tener alrededor de 100 o 200 aristas, se caracteriza con un *average degree distribution* de 2.5, un *average hop plot* de 7, un *average cluster coefficient* de 0 % y un *average betweenes centrality* de 225. Estos indicadores muestran un ingreso de varios usuarios principales que se conectan a la red, pero casi sin interconexión entre ellos. A la vez, estos usuarios logran compartir la funa a otras personas de diferentes grupos manteniendo la desconexión de la red ($\overline{cc} \approx 0$ y $\overline{dd} \approx 2,5$). Este es el momento de mayor desconexión de la red, casi sin existir agrupaciones cerradas, donde se observa que la funa inicialmente se difunde por personas de distintos grupos.

Cuando la red alcanza las 2,000 aristas aproximadamente, se muestra un *average degree distribution* de 6.25, un *average hop plot* de 5.5, un *average cluster coefficient* de 20 % y un *average betweenes centrality* de 230. Este es el momento cuando la red empieza a aumentar su interconexión, aumentando la cercanía de los usuarios (hop plot disminuye) y comienza la agrupación de usuarios ($\overline{cc}$ aumenta considerablemente). Por lo tanto, existen varios usuarios que influyen en la difusión de la funa, y no es el nodo principal el difusor más importante.

En las 8,000 aristas se observan las siguientes características de la red: *average degree distribution* de 15, un *average hop plot* de 5, un *average cluster coefficient* de 28 % y un *average betweenes centrality* de 125. Por lo tanto, la red sigue aumentando su interconexión a medida que nuevos usuarios retweetean la funa. Los nuevos nodos siguen uniéndose y fortaleciendo grupos (disminución del BC y aumento del *clustering coefficient*). De igual manera, siguen existiendo algunas personas que no pertenecen a los grupos principales, que tampoco están relacionados entre ellos.

Finalmente, luego de 47 horas, la red se cierra con alrededor de 12,000 aristas, obteniendo un *average degree distribution* de 17.5, un *average hop plot* de 4.5, un *average cluster coefficient* de 40 % y un *average betweenes centrality* de 100. En este punto la red se encuentra bastante interconectada, dado que las personas están separadas a cuatro saltos de distancia, y cuenta con altos valores de *average degree distribution* y *cluster coefficient*. Por tanto, a este punto ya se encuentra un número reducido de grupos y es difícil de distinguir a los usuarios que fueron principales al principio de la difusión.

5.3. Trend de la entrevista de Mónica Rincón a ex ministro

El tweet tendencia que será analizado es una extracción de la entrevista de Mónica Rincón al ex ministro de Defensa del Interior de Chile, Francisco Vidal. El entrevistado muestra apoyo a Rodrigo Peñailillo, quien fue uno de los políticos involucrados en el caso SQM (abril del 2015, cuando se descubren boletas y facturas falsas por

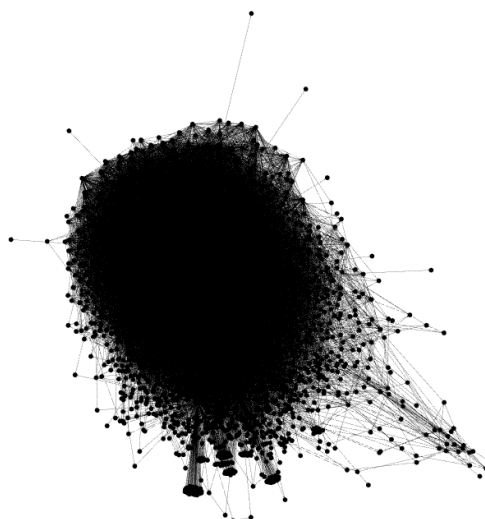
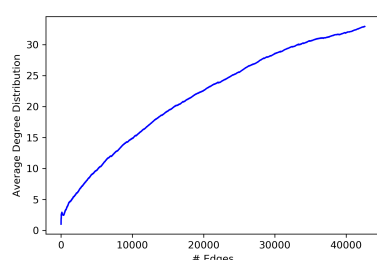


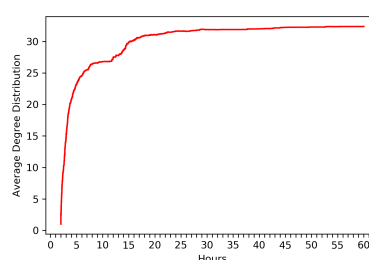
FIGURA 5.13: Red resultante del trend de la entrevista al ex ministro

alrededor de 4,000) millones⁸. Luego de 6 años, en agosto del 2021, la fiscalía decidió no perseverar el caso contra algunos involucrados, dentro de ellos Rodrigo Peñaillo, dado que el Servicio de Impuestos Internos no prosiguió con la denuncia⁹. Por tanto, la periodista fue destacada por su respuesta al ex ministro: “*son pasadas las 12, la hora en que somos tontos*”, cuando este menciona apoyo a Peñaillo dado que no fue sentenciado ni querellado. El tweet completo es el siguiente: “*La base de la impunidad (1-2). Francisco Vidal dice apoyar a Peñaillo en su carrera senatorial xq no fue formalizado ni querellado y @tv_monica le dice: “Sr Vidal, son las 7, son pasadas las 12 la hora en que somos tontos”*”; fue publicado el 3 de agosto del 2021 a las 20:07. La red resultante posee 1,797 nodos y 42,438 relaciones (figura 5.13). Las gráficas muestran el 97 % de las aristas, dado que el resto de de retweet se encuentran entre el segundo y casi quinto día de difusión, sin existir un cambio significativo de las tendencias.

5.3.1. Análisis de características: *Average Degree Distribution*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.14: Gráficos de *Average Degree Distribution* de la red del trend de la entrevista al ex ministro

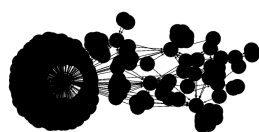
En la figura 5.14 se observa los gráficos de *average degree distribution* según el tiempo y la cantidad de aristas. En la figura 5.14a se muestra que al inicio a las 1,000 aristas,

⁸<https://www.eldesconcierto.cl/nacional/2021/08/04/video-monica-rincon-a-francisco-vidal-tras-su-defensa-a-penaillo-son-las-7-pasadas-las-12-es-la-hora-en-que-somos-tontos.html>

⁹<https://www.cnnchile.com/lodijeronencnn/gajardo-fiscalia-caso-sqm-clase-politica-20210803/>

se aprecia un incremento cercano de 0 a 2.5 conexiones promedio. Podría indicar la existencia de un usuario principal en la difusión del tweet, o una cadena de usuarios relevantes. Posterior a alcanzar dicho valor, se presencia cambios de comportamientos en la red. La red empieza a incrementar desde los 2.5 a 32 conexiones promedio, alcanzando las 15, 20 y 30 conexiones promedio en las 10,000, 20,000 y 30,000 aristas creadas. En este comportamiento se puede apreciar que las personas que se informaron, logran propagar la información dentro de sus grupos (aumentando fuertemente el degree distribution).

Con respecto a la imagen de la derecha 5.14b, se observa que dentro de las 5 horas ya se alcanza un *average degree distribution* de 25, lo que implica que se encuentra más de la mitad de las aristas creadas en la red. Dicho periodo coincide con el mayor crecimiento de la red, creciendo a una tasa de 73,33 aristas por minuto. Vale mencionar que dentro de la primera hora se presencian los nodos principales de la difusión. Luego, se observa que las siguientes 5 horas se mantienen alrededor de las 26 aristas, por lo que el aumento no es más que alrededor de 1,000 aristas. En las siguientes 2 horas se acerca al comportamiento final de la red, alcanzando las 30 conexiones promedio. Entonces, el resto de 48 horas aumenta paulatinamente hasta las 32 aristas promedio.



(A) Grafo con 300 aristas creadas



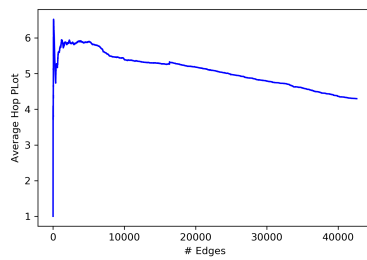
(B) Grafo con 2,000 aristas creadas

FIGURA 5.15: Red del trend de la entrevista al ex ministro en distintas instancias temporales

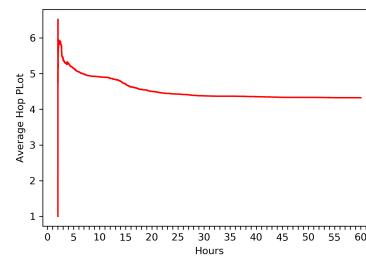
En la figura 5.15 se observa los cambios de comportamiento de la red a medida que va creciendo. Al inicio se muestra la red con 300 aristas (figura 5.15a), que cuenta con un nodo principal que logra difundir la publicación a varios usuarios, los cuales la mayoría solo se informa. Solo algunos de estos usuarios logran difundir a otras personas formando una cadena de propagación. Luego, en la figura 5.15b, se observa un gran grupo central, con otras pequeñas agrupaciones en la periferia (informadas por usuarios claves).

5.3.2. Análisis de características: *Average Hop Plot*

En la figura 5.16 se muestra dos gráficos de *average hop plot* según la cantidad de aristas y las horas transcurridas. En la imagen 5.16a se observa que dentro de las primeras 1,000 aristas se alcanza el máximo de distancia promedio (6.5 average hop plot). Esto muestra una red dispersa con personas alejadas de los nodos principales. Rápidamente, el valor disminuye a 4.6, lo cual marca un cambio de comportamiento, los usuarios empiezan a relacionarse entre ellos (menos de 2,000 aristas en la red). En las siguientes 6,000 aristas vuelve a subir la distancia promedio a 6.0, ampliando



(A) Según la cantidad de aristas creadas



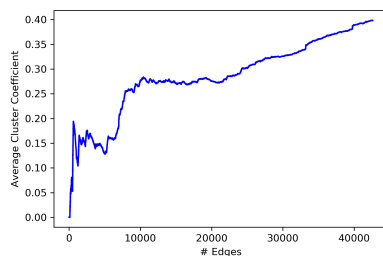
(B) Según las horas transcurridas

FIGURA 5.16: Gráficos de *Average Hop Plot* de la red del trend de la entrevista al ex ministro

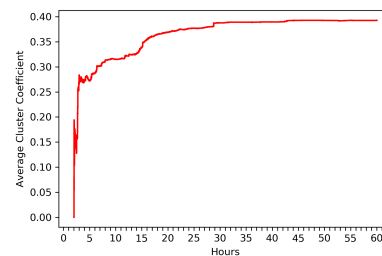
la dispersión de la red. Finalmente, la agregación de nuevas aristas comienza a unir distintos grupos de personas hasta alcanzar un *average hop plot* cercano a 4.

En la figura 5.16b se observa la distancia promedio según el paso de las horas. Dentro de la primera hora, se encuentra la red más amplia. En las siguientes 4 horas, baja la distancia hasta un 4.9 aproximadamente, donde se mantiene hasta las 10 horas. Se presencia menor separación entre las personas y gana importancia la difusión entre la mayoría de los usuarios. En las siguientes 10 horas baja hasta 4.2, aproximadamente. Por tanto, ya desde las 5 horas se observó el mismo comportamiento, donde los usuarios están cercanos entre ellos y logran propagar la información.

5.3.3. Análisis de características: *Average Cluster Coefficient*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

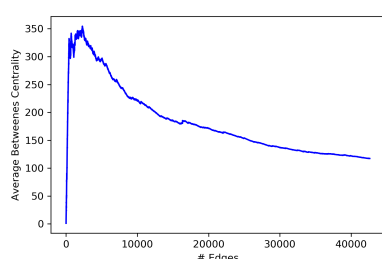
FIGURA 5.17: Gráficos de *Average Cluster Coefficient* de la red del trend de la entrevista al ex ministro

En la figura 5.17 se observa dos gráficos de *average cluster coefficient*, uno con respecto a la cantidad de aristas creadas y otro según el tiempo transcurrido. En la imagen 5.17a se muestra que hasta las primeras 1,000 aristas creadas, hay un crecimiento de un 0 % a un 20 % de agrupación entre los nodos. Se aprecia que en un comienzo se experimentó una difusión debida a nodos principales, como se ha mencionado previamente. Sin embargo, mientras más personas conocían la publicación, los nodos comenzaron a generar agrupaciones entre ellos. Hasta las 6,000 aristas se presencia fluctuaciones en el índice: varía entre un 10 % y un 17.5 %, aproximadamente. Es posible que se deba a la creación o mayor importancia de distintos grupos en la red, que experimentan también la interconexión entre ellos. Luego a las 7,000 aristas se observa una alza de un 25 % de agrupación, que se mantiene por 13,000 aristas. Esto explica una predominancia en la interconexión de los grupos, por sobre varias agrupaciones distintivas. Finalmente, en las 20,000 aristas restantes, se presencia un

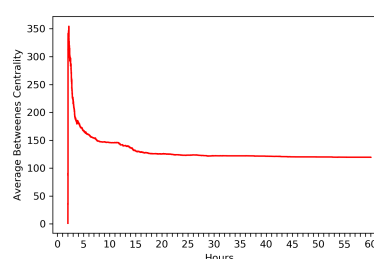
crecimiento paulatino hasta alcanzar un 40 % de agrupación, aproximadamente. Esto muestra que los nuevos usuarios conocían a varios de los nodos que ya habían transmitido la información, generando triángulos cerrados.

En la figura 5.17b se muestra el *average cluster coefficient* según el tiempo transcurrido. A la hora se alcanza el 26 % de agrupación, por tanto, en dicho periodo se concentra la formación de los distintos grupos. Las siguientes 9 horas, se alcanza un 30 % de agrupación, entonces los nuevos nodos comienzan a conectarse a grupos de personas ya conocidos. Finalmente, a las 20 horas transcurridas, se acerca al 40 % de agrupación, momento en que la red está más interconectada y la difusión se debe por varios usuarios que transmiten la misma información.

5.3.4. Análisis de características: *Average Betweenes Centrality*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.18: Gráficos de *Average Betweenes Centrality* de la red del trend de la entrevista al ex ministro

En la figura 5.18 se observa los gráficos de *average betweenes centrality* (BC) según la cantidad de aristas creadas y las horas transcurridas. En primera instancia, la figura 5.18a muestra que en las primeras 1,000 aristas hay un incremento de 0 a alrededor de 345 veces en que un usuario actúa como puente en los caminos entre usuarios. Esto tendería a mostrar una red con pocas conexiones entre grupos, donde los usuarios principales forman cadena entre ellos. Luego se empiezan a formar los grupos que interactúan entre ellos, disminuyendo el *average betweenes centrality* hasta un valor cercano a 125. En la figura 5.18a podemos observar que a las 2 horas ya se alcanza el máximo valor de BC. A partir de este momento, los nuevos nodos difunden la información a través de sus grupos, o interconectando cada vez más grupos y reduciendo su valor.

5.3.5. Recapitulación de análisis

Recapitulando los análisis previos, se observan distintos comportamientos que se desenvuelven a lo largo del tiempo. En un inicio se encuentra la red con 1,000 aristas con un *average degree distribution* de 2.5 conexiones, un *average hop plot* de 6.5, *average cluster coefficient* de un 20 % y un *average betweenes centrality* de 345 veces. Se presencia una red con ciertos usuarios principales difusores de la publicación. Además, algunos de los usuarios, informados por estas personas principales, logran propagarlo por distintos grupos en forma de cadena. Hace sentido que es el momento más amplio de la red. Luego, a las 6,000 aristas, la red se amplió con la formación de distintos grupos a la periferia del grupo central. Estas agrupaciones, a diferencia de lo que se espera de una persona de la red, solo se informa sin propagar la información. En las 10,000 aristas disminuye la cantidad de grupos en la red al aumentar la

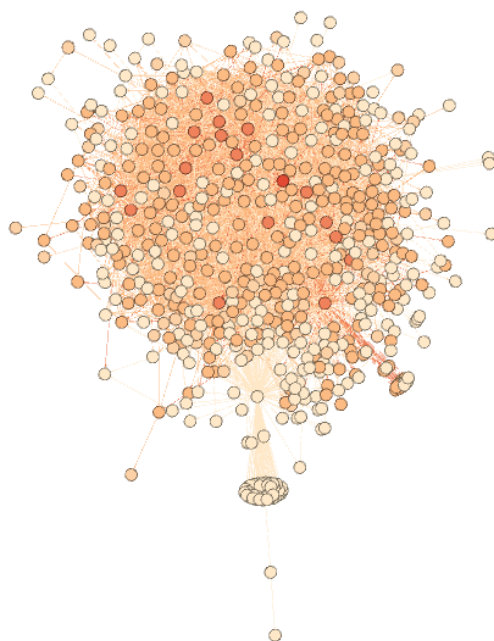


FIGURA 5.19: Grafo resultante de la funa a Edwin Dimter

interconexión entre ellos. Finalmente, en la red conformada por 42,000 aristas, predomina un grupo central que está interconectado entre ellos, por lo que los usuarios logran propagar la información entre sus seguidores y las personas tienen varios caminos por lo que enterarse. Se conectaron los grupos que se encuentran a la periferia con otros grupos principales.

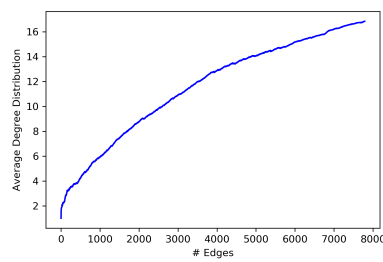
5.4. Funa a Edwin Dimter

La publicación analizada es una funa a Edwin Dimter Bianchi, ex militar, quien fue condenado por violación a los derechos humanos. En el año 2018, el ministro de la Corte de Apelaciones condenó a Edwin por ser el autor del homicidio y delito de secuestro simple de Víctor Jara (cantautor chileno) y Littré Quiroga (miembro del Partido Comunista y director general de Servicio de Prisiones), en el año 1973. La pena es de 15 años y un día 3 años de presidio¹⁰. Hoy en día, el cumplimiento de su sentencia está suspendido, dado que Edwin ha recurrido a distintas instancias permitidas por el Estado. El tweet fue publicado el 3 de noviembre del 2021, y dice lo siguiente: "Si no hay justicia, hay funa" junto a una imagen que indica: "Edwin Dimter Bianchi, asesino de Víctor Jara", su rut y la dirección donde vive actualmente. La red que se muestra en la figura 5.19 fue formada en 22 horas aproximadamente, alcanzando 594 usuarios y 7,779 relaciones.

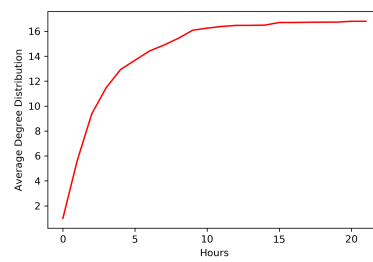
5.4.1. Análisis de características: *Average Degree Distribution*

En la figura 5.20 se observan los gráficos del *average degree distribution* según la variación de cantidad de aristas y horas transcurridas. En la figura 5.20a se observa que en un comienzo, alrededor de las 100 aristas, existe un promedio de una relación,

¹⁰<https://www.t13.cl/noticia/nacional/victor-jara-muerte-47-investigacion-judicial-16-09-20>



(A) Según la cantidad de aristas creadas

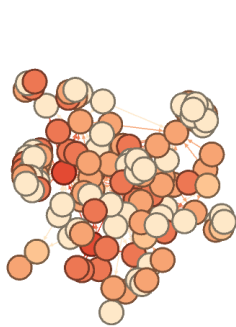


(B) Según las horas transcurridas

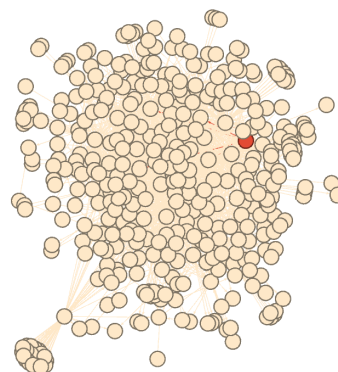
FIGURA 5.20: Gráficos de *Average Degree Distribution* de la red funa a Edwin Dimter

lo que implica que uno o varios usuarios influyentes difunden la información. Luego hasta las 500 aristas, aumenta a un promedio de 4 relaciones. La red empieza a interconectarse y en este punto ya hay en promedio 4 personas a las que informas o con las que te informaste. Este número continua incrementando observando 9, 13 y 16 conexiones promedio con 2,000, 4,000 y 8,000 aristas, respectivamente. La red empieza a cohesionarse cada vez más, lo que significa que dentro de las personas que difundieron la información, se espera que cada usuario este relacionado con dichas cantidades. Por tanto, la participación de todos los usuarios debe ser primordial para la propagación. Asimismo, deben existir ciertos grupos aislados que no logran difundir la información.

En la figura 5.20b se muestra el *average degree distribution* según el avance de las horas. Como ha sido la tendencia en las redes pasadas, dentro de las primeras 4 horas se observa un incremento progresivo, alcanzando, en este caso, las 12 aristas promedio. Incluso en la primera hora, la red empieza a interconectarse, ganando protagonismo los diferentes usuarios para la difusión. Dentro de las 5 horas ya se encuentra más de la mitad de las relaciones creadas (4,500 aristas) y al llegar a dicho punto, la red sigue aumentando la cantidad de relaciones, pero de forma más paulatina. Finalmente, a las 10 horas ya se mantiene un comportamiento más estable, por tanto, en el resto de las 12 horas, se espera que los usuarios nuevos formen parte de manera colaborativa: informando o informándose de distintas fuentes.



(A) Grafo con 200 aristas creadas

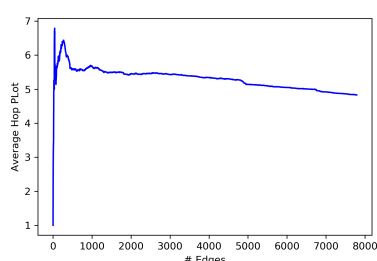


(B) Grafo con 2,000 aristas creadas

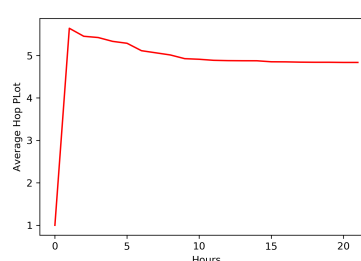
FIGURA 5.21: Red funa a Edwin Dimter en distintas instancias temporales

En la figura 5.21 se observa la red en dos puntos distintos de formación. La primera es a las 200 aristas creadas, y la segunda con 2,000. En la imagen 5.21a se muestra que hay varios usuarios influyentes que logran difundir la información a ciertos grupos. Dichas personas también están conectados entre ellos de cierta manera (directa o indirectamente). En la figura 5.21b se observa un gran grupo principal donde los usuarios están interconectados entre ellos. Por tanto, el protagonismo de la difusión lo tienen los distintos usuarios de la red. De igual forma, se pueden observar grupos aislados a la periferia, que no logran difundir la funa más allá de ellos. Vale destacar que estos son informados por personas particulares que podrían coincidir con los que fueron influyentes.

5.4.2. Análisis de características: *Average Hop Plot*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

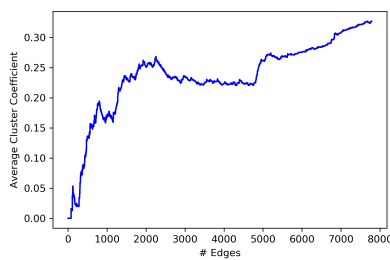
FIGURA 5.22: Gráficos de *Average Hop Plot* de la red funa a Edwin Dimter

En la figura 5.22 se observan los gráficos de *average hop plot* con variación según las aristas agregadas y las horas transcurridas. En un comienzo se puede observar el mayor distanciamiento de la red, donde se encuentran los usuarios influyentes que informan a varias grupos que no logran difundir la información. Entonces, de cierta forma debe existir una cadena de usuarios para llegar a los distintas personas aisladas. Hasta las 500 aristas se muestra dicho comportamiento, alcanzando un distanciamiento de 6.5. Luego, se observa una baja en el distanciamiento a 5.5, aproximadamente. La distancia no presenta una gran variación luego, dado que a las 4,000 y 8,000 aristas se observa un average hop plot de 5.25 y 4.9, respectivamente.

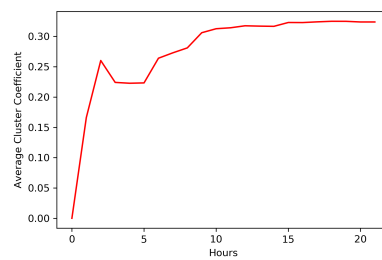
En la figura 5.22b se muestra el gráfico según la variación en el tiempo. Se observa que a las primeras 2 horas y medias ya se alcanza el máximo de distancia entre las personas. Por tanto, en dicho periodo se encuentra la red con los usuarios influyentes como principales medios de difusión. A las 5 y 10 horas ya se reduce a una distancia de 5.2 y 4.9, respectivamente. Por tanto, luego de las 10 horas es cuando se mantiene su comportamiento final.

5.4.3. Análisis de características: *Average Cluster Coefficient*

En la figura 5.23 se visualiza el *average cluster coefficient* según las aristas que se han agregado a la red y las horas transcurridas. En la imagen 5.23a se observa que inicia con un 0 % de agrupación, manteniéndose en dicho valor hasta alrededor de las 200 aristas. Esto muestra una agrupación nula, por tanto, se esperaría que usuarios solo se informen, o bien, que se formen cadenas de difusión. Luego, se observa un alza a un 5 %, formándose los primeros grupos dentro de la red. A las 500 relaciones se observa una baja que aumenta rápidamente hasta alcanzar un 20 % de agrupación a las



(A) Según la cantidad de aristas creadas



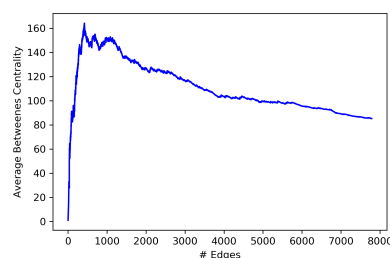
(B) Según las horas transcurridas

FIGURA 5.23: Gráficos de *Average Cluster Coefficient* de la red funa a Edwin Dimter

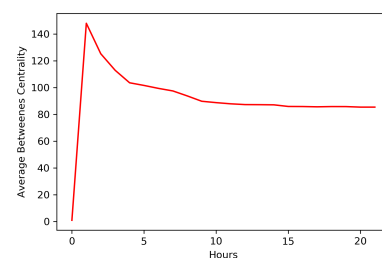
1000 aristas, aproximadamente. La red ya se observa interconectada, dado que se espera que un 20 % de los amigos de cualquier persona, sean amigos. Luego de una baja, se observa un alza hasta un 25 % en las 2,500 aristas. A las 3,000 aristas vuelve a disminuir a un 20 %. Por tanto, se debe presenciar la creación de nuevos grupos que se mantienen aislados o se agrupan al gran grupo principal. A las 5,000 aristas vuelve a aumentar a un 25 %, que luego sigue aumentando paulatinamente hasta alcanzar un 32 % aproximadamente. Se presencia la agrupación más alta del transcurso del tweet, por tanto, se espera un grupo en el cual hay varias formas de enterarse de la funa.

En la figura 5.23b se muestra la variación del *average cluster coefficient* según el tiempo. Se aprecia que dentro de las 2 horas y media se alcanza rápidamente un alza a un 25 %, donde luego se disminuye y mantiene en un 20 % hasta las 5 horas. En este punto se debe observar un gran grupo principal junto a varios más pequeños. Hasta las 10 horas se observa un aumento en la agrupación a un 30 %, donde dicho comportamiento no varía en gran porcentaje en las horas restantes. En este punto ya se encuentra en la agrupación máxima de la red.

5.4.4. Análisis de características: *Average Betweenes Centrality*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.24: Gráficos de *Average Betweenes Centrality* de la red funa a Edwin Dimter

En la figura 5.24 se observa las imágenes del *average betweenes centrality* versus la cantidad de aristas y el tiempo. En la figura 5.24a se observa un inicio en 0 que sube rápidamente alcanzando un valor de 160 veces promedio a las 500 aristas. En dicho punto se debería presenciar distintos grupos que logran informarse, pero que aún no tienen una alta interconexión entre ellos, entonces algunos usuarios deben ser usado como uniones entre los grupos. Por tanto, el flujo de la funa a este punto debe ser

aún lenta. Luego, hasta las 1,500 aristas el *average betweenes centrality* fluctúa entre las 140 y 150 veces. Por tanto, el comportamiento de la red debe estar pasando por una transición hasta lograr la interconexión. Finalmente, a medida que nuevos usuarios se conectan a la red y se genera una interconexión entre los grupos, el valor baja hasta 80.

En la figura 5.24b se encuentra el gráfico del *average betweenes centrality* variando en el tiempo. Dentro de las primeras dos horas, se alcanza un alza de 150 veces, por tanto, durante dicho periodo los usuarios influyentes siguen actuando como puente entre los usuarios. Ya a las 5 horas se observa un *average betweenes centrality* de 100, lo que indicaría que la red se encuentra más cohesionada, formando un gran grupo principal. En las 5 horas siguientes, el índice disminuye levemente a 80 veces, lo cual se mantiene relativamente constante las horas restantes. Entonces, se podría encontrar la red aún más cohesionada en un gran grupo junto a otros pequeños colindantes que no logran difundir más allá la funa.

5.4.5. Recapitulación de análisis

Dentro de la primera hora se alcanza a tener 200 aristas con las siguientes características: *average degree distribution* de 3 conexiones, un *average hop plot* de 7, *average cluster coefficient* de un 5 % y un *average betweenes centrality* de 100 veces. En este punto, se encuentran varios usuarios influyentes que logran difundir la información a distintas personas que no generan gran impacto en la difusión de la funa. Por esta razón se encuentra el momento de mayor distensión de la red. Aunque sean varias las personas que se enteran de la funa, la divulgación aún no se hace entre grupos cohesionados.

Al transcurrir cuatro horas más, se alcanzan 5,000 aristas y se aprecian las siguientes características: *average degree distribution* de 14 conexiones, un *average hop plot* de 5, *average cluster coefficient* de un 20 % y un *average betweenes centrality* de 100 veces. La red está más cohesionada, por tanto, los usuarios empezaron a difundir la funa entre ellos, perdiendo protagonismo los influyentes del comienzo. También se puede observar menor distanciamiento entre las personas y mayor cantidad de relaciones, lo que indica que hay varias formas de que se enteren.

Finalmente a las 21 horas transcurridas la red se encuentra con su total de 8,000 aristas, con las características: *average degree distribution* de 17 conexiones, un *average hop plot* de 4.9, *average cluster coefficient* de un 31 % y un *average betweenes centrality* de 80 veces. Se observa que la red continuó incrementando la cohesión, entonces hay más formas que los usuarios se enteren de la funa. Se observa un gran grupo principal, junto a otros que se encuentran a la periferia que no logra informar a nuevos usuarios.

5.5. Trend de José Antonio Kast

La publicación analizada es de José Antonio Kast, abogado y político chileno, que en el año 2021 postuló como presidente. En septiembre del año 2021 se aprobó la despenalización del aborto hasta las 14 semanas¹¹, lo cual Kast no apoya. Entonces, twiteo lo siguiente: "Han pasado 24 horas y Sebastián Piñera aún no dice nada por la

¹¹<https://cnnespanol.cnn.com/2021/09/28/camara-de-diputados-de-chile-debate-despenalizacion-del-aborto-sin-causales-hasta-las-14-semanas-de-gestacion/>

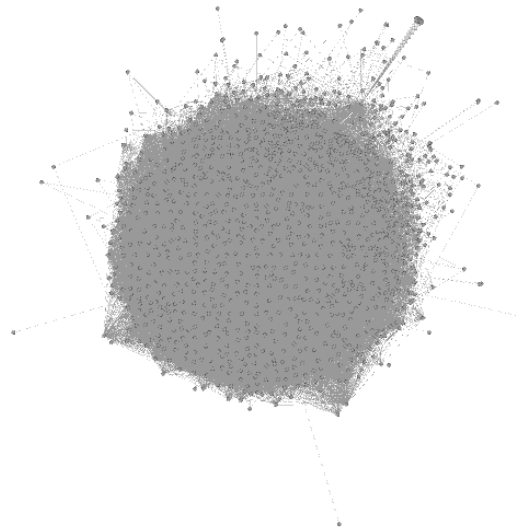
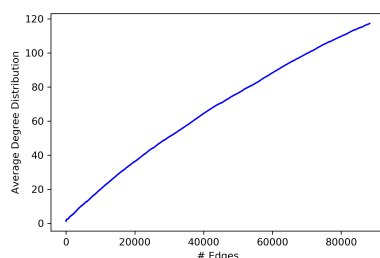


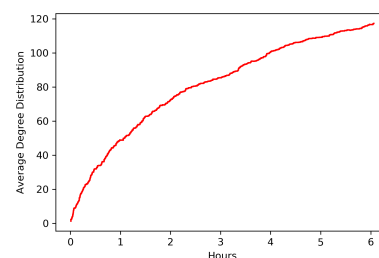
FIGURA 5.25: Grafo resultante del trend de José Antonio Kast

aprobación de la Ley de Aborto en Chile.". Aludiendo que el presidente de turno, debería decir algo dado que son de una misma coalición política. Esta publicación se realizó el día 29 de septiembre del año 2021 y fue divulgada por 1,363 usuarios. Dada la complejidad de análisis, se detuvo el programa a los 1,173 usuarios con 88,310 aristas. En la imagen 5.25 se puede observar la red resultante.

5.5.1. Análisis de características: *Average Degree Distribution*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.26: Gráficos de *Average Degree Distribution* del trend de José Antonio Kast

En la figura 5.26 se observa los gráficos de *average degree distribution* según la cantidad de aristas creadas (figura 5.26b) y el tiempo (figura 5.26a). Dentro de la primera hora se observa que la red inicia con una baja cantidad de relaciones promedio. Esto implica que puede estar involucrado uno o varios usuarios influyentes. Sin embargo, ya dentro de la primera media hora y con 10,000 aristas, se alcanza un promedio de 20 relaciones. Llama la atención que en comparación a las otras redes analizadas, se alcanza rápidamente un valor alto de *degree distribution*. Esto podría implicar por un lado que los usuarios influyentes, impactaron en gran medida la difusión. Entonces, los nuevos usuarios se enteran por los principales, o bien, son nuevos influyentes. También cabe la posibilidad que la red se cohesiona rápidamente, por tanto, la mayoría de las personas logran propagar la información. La red continúa aumentando las relaciones: a la hora y con 20,000 aristas, se encuentra con una esperanza de 40

usuarios conectados. A las dos horas con 45,000 aristas, con una esperanza de 60 usuarios, y así continúa creciendo hasta alcanzar una esperanza de 120 usuarios a las 5 horas y con alrededor de 85,000 aristas. En comparación al resto de redes analizadas, se destaca el alto valor que alcanza al final, dado que se espera que un usuario este conectado con otros 120 amigos para difundir o se informe de la publicación. Un valor no alcanzado en las redes anteriores.

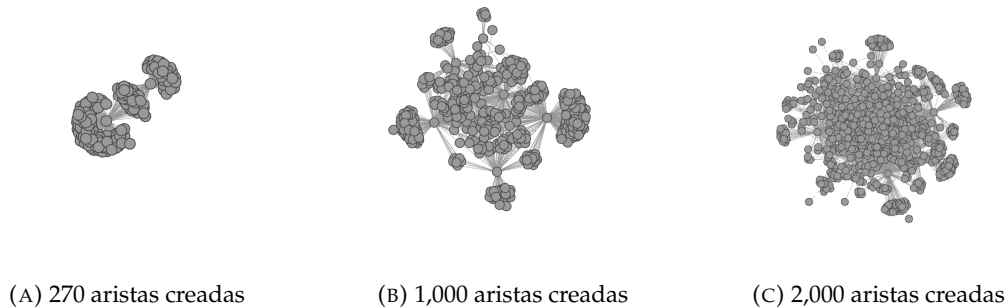
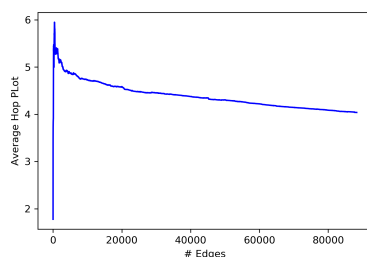


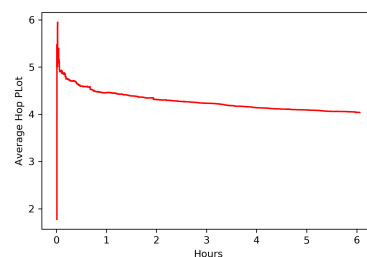
FIGURA 5.27: Red del trend de José Antonio Kast en distintas instancias temporales

En la figura 5.27 se observa el comportamiento inicial de la red. Se comienza con dos usuarios principales en la difusión (figura 5.27a). Las personas se enteran de la publicación solo a través de uno o ambos usuarios principales. Luego aumenta la cantidad de usuarios que difunden la información (figura 5.27b). Algunas de las personas nuevas se transforman en influyentes, logrando difundir información, y el resto solo se enteran a través de una o más de esas personas principales. En la imagen 5.27c ya se no se puede identificar visualmente a usuarios que son influyentes, pero aún existen varios usuarios que logran difundir la publicación a conjuntos de personas que se ven a la periferia. Por otro lado, deben existir usuarios que ya están participando de la difusión en el grupo central cohesionado.

5.5.2. Análisis de características: *Average Hop Plot*



(A) Según la cantidad de aristas creadas



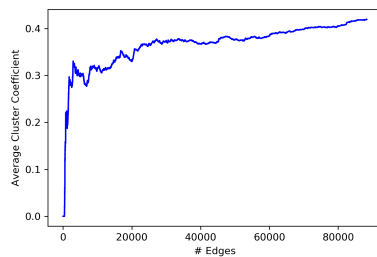
(B) Según las horas transcurridas

FIGURA 5.28: Gráficos de *Average Hop Plot* del trend de José Antonio Kast

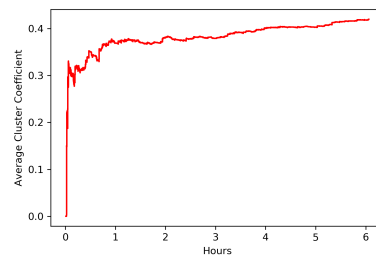
En la figura 5.28 se observa el *average hop plot* según la cantidad de aristas creadas (figura 5.28a) y las horas transcurridas (figura 5.28b). Se puede observar que dentro de la primera hora y antes de las 1,000 aristas, se alcanza el máximo de la distancia de la red. Esto se puede deber a que la información se comparte principalmente por los usuarios influyentes, por tanto, el resto que solo se informan, se encontraban más

alejados entre ellos. Una vez alcanzado el máximo, se muestra una rápida disminución hasta las 10,000 aristas aproximadamente, lo que sería dentro de la primera hora. En las siguientes 5 horas, disminuye de forma más paulatina hasta alcanzar una distancia de 4 saltos promedio. Por tanto, se mantiene estable el comportamiento. La red debe estar más cohesionada y deben existir usuarios que logran la difusión de información.

5.5.3. Análisis de características: *Average Cluster Coefficient*



(A) Según la cantidad de aristas creadas

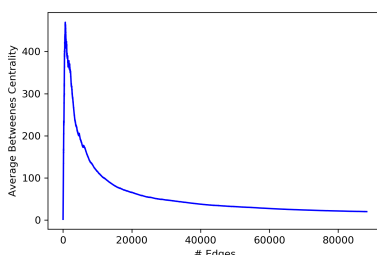


(B) Según las horas transcurridas

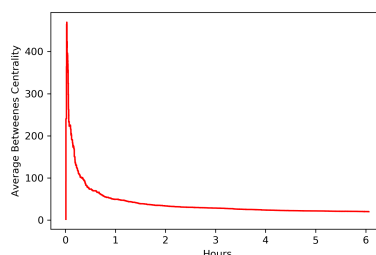
FIGURA 5.29: Gráficos de *Average Cluster Coefficient* del trend de José Antonio Kast

En la figura 5.29 se observa el *average cluster coefficient* según la cantidad de aristas creadas (figura 5.29a) y la cantidad de horas transcurridas (figura 5.29b). En un comienzo, se observa que la agrupación sube de un 0 % a un 20 %, lo se puede deber a los usuarios influyentes que están relacionados entre ellos, y al mismo tiempo que varias personas que se informan a través de ellos. A medida que más personas compartían la información, más se potenciaba este comportamiento. Antes de la primera hora y las 10,000 aristas, se alcanza un 30 % de agrupación. Luego, se observa que la agrupación aumenta de forma más gradual en las siguientes 5 horas hasta alcanzar un 41 %. Esta agrupación podría ser explicada con la cohesión de la red y la mayor participación de las distintas personas en la difusión.

5.5.4. Análisis de características: *Average Betweenes Centrality*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.30: Gráficos de *Average Betweenes Centrality* de la publicación de José Antonio Kast

En la figura 5.30 se observa el *average betweenes centrality* según la cantidad de aristas creadas (figura 5.30a) y la cantidad de horas transcurridas (figura 5.30b). Se inicia con un BC de 500 veces, lo que se debe a la existencia de los usuarios influyentes que informan a conjuntos de usuarios que cuentan con una baja interconexión. Sin

embargo, luego de alcanzar el máximo valor, el BC empieza a disminuir en forma considerable e irregular. Específicamente, dentro de la primera hora se observa el aumento del *average betweenes centrality* y su disminución pronunciada hasta un BC de 100 veces a las 12,000 aristas. Las siguientes 5 horas el BC sigue decreciendo paulatinamente: a las 20,000 aristas se observa un BC de 90 veces promedio, a las 30,000 aristas un BC de 50 veces, y a las 60,000 aristas, un BC de 20 veces. Esto indica un comportamiento bastante irregular, ya que la mayoría de los nodos se encuentran conectados con casi todos el resto de los nodos. Es decir, ya no existen nodos influyentes iniciales que siguen propagando la información, en cambio todos los nodos ahora sirven para propagar la información.

5.5.5. Recapitulación de análisis

Dentro de la primera hora y antes de alcanzar las 1,000 aristas, la red se encuentra con siguientes características: el *average degree distribution* es de 3 conexiones, el *average hop plot* es de 6 saltos, el *average cluster coefficient* es de 30 % y el *average betweenes centrality* es de 500 veces. Esto podría indicar la presencia de una difusión de tipo funa, donde hay un grupo fuertemente conectado. Sin embargo, llama la atención de la rapidez en que se logra estas características. Luego,

alcanzando la primera hora de difusión, la forma en que propaga la publicación cambia. Son los diversos usuarios la que la divulgan, por tanto, se observa un cambio en las características: el *average degree distribution* es de 40 conexiones, el *average hop plot* es de 4.5 saltos, el *average cluster coefficient* es de 38 % y el *average betweenes centrality* es de 60 veces. Aquí ya se observa un comportamiento bastante extraño en comparación al resto de las redes analizadas. En el resto de de horas estas características cambian de forma más paulatina, mostrando que se potencia el comportamiento señalado. A las 6 horas transcurridas se alcanza las siguientes características: el *average degree distribution* es de 120 conexiones, el *average hop plot* es de 4 saltos, el *average cluster coefficient* es de 40 % y el *average betweenes centrality* es de 20 veces. Existen varias comportamiento extraños en este tipo de red, incluyendo la cantidad de conexiones promedio que tiene una personas en la red, el extremadamente bajo valor de *average betweenes centrality*, y el bajo valor de *hop plot*. En palabras sencillas, pareciera ser que en esta red la mayoría de los nodos se conocen entre si. Algo que no se había observado hasta ahora.

5.6. Funa a la primera línea

La última funa analizada es contra los manifestantes que dañaron a una civil y carabineros en una marcha. A inicios de octubre del año 2021 en Santiago de Chile, unas personas usaron fuegos artificiales para atacar a carabineros, lesionando a varios de ellos y causando la muerte de una manifestante ¹². Por tanto, el tweet hace alusión a la primera línea, identificando a las personas que protestan al inicio del grupo. La publicación es la siguiente: "El lanzamiento de fuegos artificiales hacia los carabineros era constante por parte de los violentistas. No era de extrañar que tarde o temprano ocurriría una tragedia. #PrimeralineaAsesina". La red cuenta con un total de 1635 usuarios que difundieron la funa, pero dado al extenso tiempo de análisis de las características de la funa, se analizaron 1394 personas. Vale mencionar que se lograron formar alrededor de 55,000 relaciones.

¹²<https://www.emol.com/noticias/Nacional/2021/10/10/1035016/manifestaciones-plaza-baqueda-domingo.html>

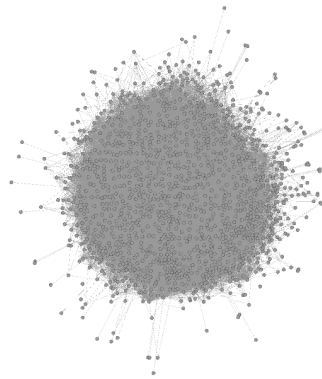
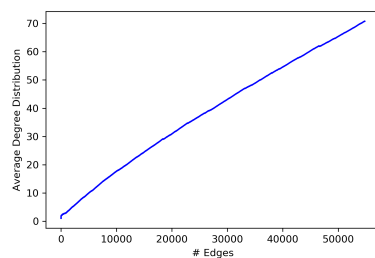
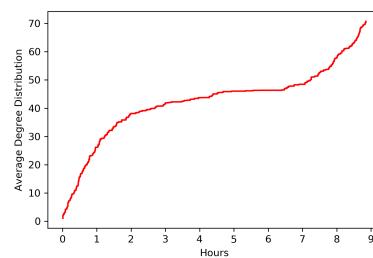


FIGURA 5.31: Red resultante del a funa a primera línea

5.6.1. Análisis de características: *Average Degree Distribution*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.32: Gráficos de *Average Degree Distribution* de la funa a la primera línea

En la figura 5.32 se observa la variación del *average degree distribution* según el tiempo (figura 5.32b) y la cantidad de aristas (figura 5.32a). Dentro de los primeros minutos de la difusión, con alrededor de 200 aristas, se observan menos de 3 relaciones por usuario promedio. Esto podría indicar la existencia de usuarios influyentes que difunden la funa, o bien, la difusión en forma de cadena. Dentro de la primera hora y hasta las 3,000 aristas aproximadamente, la red sigue aumentando la cantidad de usuarios influyentes, siendo unos pocos los principales que propagan la funa. Desde dicho punto, más personas empezaron a ganar protagonismo, por tanto, la funa empezó a ser difundida por varias personas. A la hora se alcanza un índice de 20 conexiones promedio, y en este punto es difícil concluir si se debe por muchos usuarios influyentes que estén relacionados entre ellos o se debe a la cohesión entre las personas de la red, por lo que todos propagan la funa. A las dos horas con 25,000 aristas, se alcanza un *average degree distribution* de 40 conexiones. Luego se observa que en las siguientes 5 horas, aproximadamente, crece paulatinamente hasta alcanzar un *degree distribution* de 50. Desde las 7 horas, continúa creciendo hasta alcanzar 70 conexiones promedio. Llama la atención que se espera que una persona en la difusión, sea amiga con otras 70 personas.

En las imágenes de la figura 5.33, se observan cuatro instantes de la red. Al inicio se cuenta con dos usuarios influyentes que logran informar al resto de personas (figura

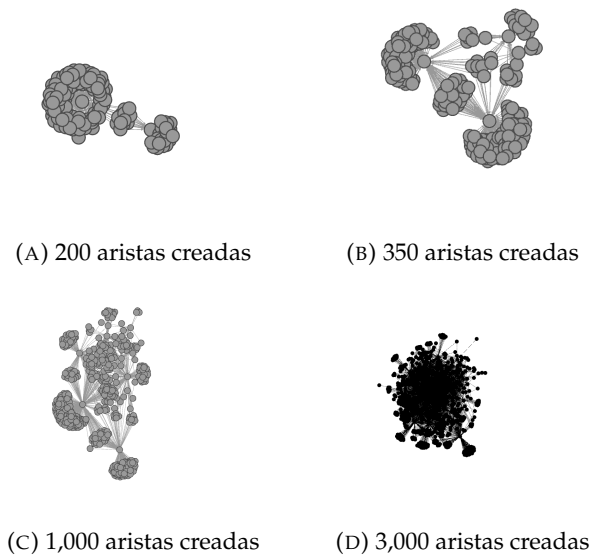
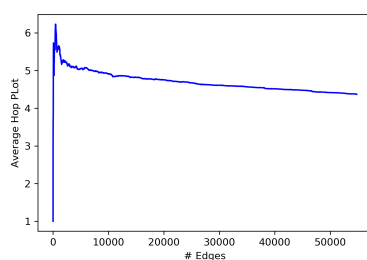


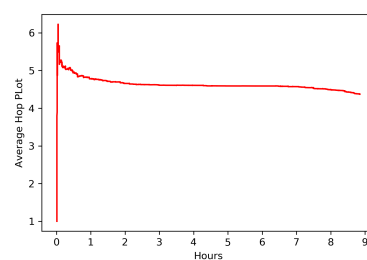
FIGURA 5.33: Red tendencia en distintas instancias temporales

5.33a). Se espera que el usuario se entere de la funa a través de uno o dos influyentes. Al crecer la red, aumenta la cantidad de usuarios principales, que comparten la información a las otras personas (figura 5.33b). A las 1,000 aristas se muestran más influyentes, pero de igual forma hay usuarios que empezaron a difundir la funa dentro de un grupo (figura 5.33c). A las 3,000 aristas ya no se visualiza tan claro a quien se interconecta el usuario, dada la alta cohesión de la red (figura 5.33d). Se identifica un gran grupo central cohesionado, que tiene a la periferia grupos que no logran propagar la funa.

5.6.2. Análisis de características: *Average Hop Plot*



(A) Según la cantidad de aristas creadas



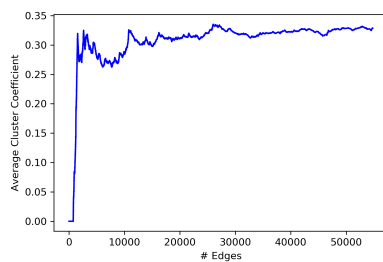
(B) Según las horas transcurridas

FIGURA 5.34: Gráficos de *Average Hop Plot* de la funa a la primera línea

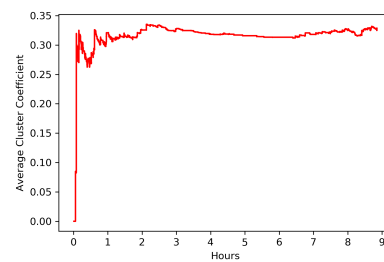
En la figura 5.34 se observa el *average hop plot* según la cantidad de aristas (figura 5.34a) y el tiempo en horas (figura 5.34b). Dentro de la primera hora y las primeras aristas creadas, se alcanza una distancia promedio de 6 saltos. Esto puede ser muy común, dado la teoría de los 6 grados de separación de Milgran (en promedio estamos a 6 saltos de cada persona). Sin embargo, antes de la primera hora, con alrededor de 4,000 aristas, la distancia disminuyó pronunciadamente, hasta una distancia promedio de 5 saltos. En las siguientes 8 horas, la distancia sigue disminuyendo de

forma más paulatina, hasta una distancia de 4.5 saltos. Se observa una red cohesionada, por lo que distintos usuarios son los difusores de la funa. De igual modo, los usuarios que están a la periferia solo se enteran de la funa través de algún usuario que fue influyente. Nuevamente, llama la atención la forma de la red, atípica en comparación a la mayoría de los casos anteriores.

5.6.3. Análisis de características: *Average Cluster Coefficient*



(A) Según la cantidad de aristas creadas

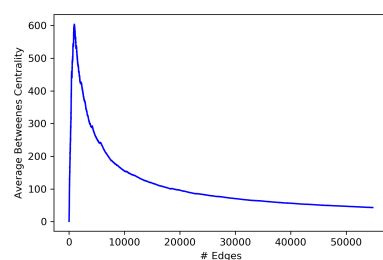


(B) Según las horas transcurridas

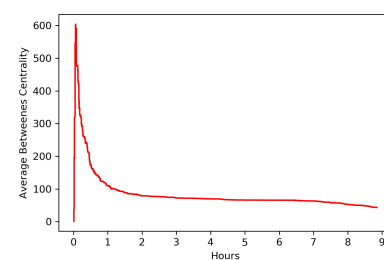
FIGURA 5.35: Gráficos de *Average Cluster Coefficient* de la funa a la primera línea

En la figura 5.35 se observa el *average cluster coefficient* según la cantidad de aristas (figura 5.35a) y según el tiempo (figura 5.35a). Se inicia con una agrupación muy baja, lo cual se explica por que la mayoría de usuarios solo se relacionan a través de los influyentes. Indicando un comportamiento típico de red social. Sin embargo, antes de la primera hora y de las 5,000 aristas, la agrupación aumenta repentinamente a un 30 %, esto podría explicarse por la inclusión de usuarios que empiezan a interconectarse entre ellos, y no exclusivamente con influyentes, lo que corresponde a una anomalía. Hasta las 20,000 aristas, la agrupación fluctúa entre un 25 % y un 30 %, y luego hasta las 45,000 aristas, la fluctuación es entre un 30 % y 35 %. Esto muestra que el comportamiento de agrupación se mantiene, y que la difusión se puede deber a todos los usuarios que forman parte del grupo cohesionado central.

5.6.4. Análisis de características: *Average Betweenes Centrality*



(A) Según la cantidad de aristas creadas



(B) Según las horas transcurridas

FIGURA 5.36: Gráficos de *Average Betweenes Centrality* de la funa a la primera línea

En la figura 5.36 se observa el *average betweenes centrality* según la cantidad de aristas (figura 5.36a) y las horas transcurridas (figura 5.36b). Dentro de la primera hora, se observa un índice de 600 veces, lo que explica la existencia de varios usuarios influyentes que difunden al resto de personas. A la hora y media, empieza a disminuir

significativamente, hasta alcanzar un BC de 100 veces. Esto podría indicar el aumento de la cantidad de usuarios que propagan la funa, disminuyendo la importancia de los influyentes. En las siguientes horas, el índice continúa disminuyendo hasta alcanzar un BC de 60 veces. Nuevamente, algo atípico en este tipo de redes.

5.6.5. Recapitulación de análisis

En los primeros minutos, la red se encuentra con 300 aristas y las siguientes características: el *average degree distribution* es de 3 conexiones, el *average hop plot* es de 6.1 saltos, el *average cluster coefficient* es cercano a 0.0 % y el *average betweenes centrality* es de 600 veces. La difusión se debe principalmente por usuarios influyentes, es decir, la principal forma que se entere el resto de personas, es a través de uno o más influyentes. Todas estas características corresponde a una red casi promedio.

A la hora, la red explota y ya se encuentra con 18,000 aristas y las siguientes características: el *average degree distribution* es de 30 conexiones, el *average hop plot* es de 4.9 saltos, el *average cluster coefficient* es de 30 % y el *average betweenes centrality* es de 100 veces. Si bien en un inicio la difusión se debió a influyentes, llegó un momento en que las personas también difundían la funa. Por tanto, disminuyó el protagonismo de los usuarios principales, al ir aumentando la agrupación e interconexión entre ellos, y la cantidad de personas a las cuales compartían la funa. Estas características, ya muestran un comportamiento atípico con respecto a lo esperado en redes sociales.

Desde las 3 a las 7 horas los índices no variaron en forma significativa. Al alcanzar 36,000 relaciones en la red, se tuvo las siguientes características: el *average degree distribution* es de 50 conexiones, el *average hop plot* es de 4.5 saltos, el *average cluster coefficient* es de 31 % y el *average betweenes centrality* es de 90 veces. Se observa que el tweet se puede difundir por cualquiera de los usuarios de la red que conforman un grupo altamente cohesionado. De igual forma, se observan conjuntos de usuarios que no logran difundir la funa, los cuales se encuentran a la periferia.

A las 9 horas con 55,000 aristas se detiene el análisis, debido a la complejidad de análisis de las usuarios en la red. Las características alcanzadas son las siguientes: el *average degree distribution* es de 70 conexiones, el *average hop plot* es de 4.2 saltos, el *average cluster coefficient* es de 31 % y el *average betweenes centrality* es de 60 veces. Finalmente, se observa el mismo comportamiento de propagación.

5.7. Discusión

Se analizaron seis redes de difusión formadas por los retweet de publicaciones obtenidas de la red social Twitter. Se tomaron en consideración la cantidad de aristas que se van formando a medida que transcurrían las horas. A continuación se presentarán resúmenes de cada uno de sus análisis.

El primer trend evaluado dice lo siguiente: "*La esposa de Bravo siempre tuvo razón #SelecciónChilena*". Esta publicación se viralizó cuando el equipo chileno de fútbol rompió la burbuja sanitaria, estando compitiendo en Brasil. La difusión inicia con un usuario influyente, quien informa a varios usuarios. Sin embargo, la mayoría de las personas solo se informan y son independientes, lo que implica la ausencia de relación entre ellas. A excepción de unas pocas personas que logran propagar la información a grupos que no están directamente relacionado con el usuario principal. Dicho comportamiento se mantiene por el resto de la difusión, entonces, la

mayoría de los casos corresponde a personas aisladas que logran difundir la información entre otros usuarios desconectados entre sí. El resto de personas difunden la información a través de grupos aislados del influyente. Según la literatura, el comportamiento recién descrito es el esperado dentro de las redes sociales. Se cuenta con una distancia de 6 saltos, lo cual se postula en la propiedad del modelo del *Mundo Pequeño de Watts y Strogatz*. También, la red cuenta con un *degree distribution* entre 1 a 3 personas, lo que está dentro de lo esperado en Twitter, dado que el 92.46 % de los usuarios sigue entre 1 a 100 personas [63]. Además, hay usuarios que difunden la información a través de los grupos, que no siguen directamente a quien publicó [64].

El siguiente tweet analizado corresponde a la *funa* descrita a continuación: “*Se busca Patricio Maturana Mutilador. Si no hay justicia, hay funa*”. El *funado* es el imputado por apremios ilegítimos con lesiones graves de vista, gusto y olfato contra Fabiola Campillai. La *funa* surge dado que se revoca su prisión preventiva a arresto domiciliario. En un comienzo, la manera principal de difusión de la *funa* es por un usuario influyente, quien informa a distintas personas independientes. Algunas pocas logran propagar la *funa* a usuarios alejados del influyente. Sin embargo, esta tendencia cambia y se observa un gran grupo con alta interconexión entre ellos. Esto hace que el usuario original pierda protagonismo y la información comience a difundirse por distintas fuentes. Dicho comportamiento se mantiene, aumentando el protagonismo de las diferentes personas en la difusión y la interconexión de la red. Llama la atención que la forma de difusión difiere del trend anterior, dado las distintas características resultantes. En este caso, se destaca la formación de un grupo cohesionado (alto *average degree distribution*, *cluster coefficient* y *hop plot* bajo seis), donde cualquiera de los usuarios podría realizar la difusión.

El siguiente trend analizado fue: “*La base de la impunidad (1-2). Francisco Vidal dice apoyar a Peñalillo en su carrera senatorial xq no fue formalizado ni querellado y @tv_monica le dice: “Sr Vidal, son las 7, son pasadas las 12 la hora en que somos tontos”*”. El tweet muestra lo sucedido en la entrevista que hizo Mónica Rincón (periodista) a Francisco Vidal¹³, quien muestra apoyo a un político imputado en el caso SQM por emisión de boletas falsas. Finalmente no siguió siendo investigando al no presentarse querrela. Inicialmente se observan varios usuarios principales que difunden la información a personas que, primordialmente, solo se informan. Algunos de los usuarios logran difundir en forma de cadena, es decir, le informan a un amigo, y luego este a un amigo suyo y así sucesivamente. Dentro de la primera hora, la forma de propagación cambia: se forma un gran grupo principal, que quienes lo conforman divulgan la publicación. Por tanto, los pocos usuarios influyentes van perdieron protagonismo, pero aún siguen informando a conjuntos de usuarios independientes. Dicho comportamiento se mantiene hasta el final de la difusión. Se destaca que su forma de propagación es parecida a la de la *funa* anterior. Además, sus características resultantes son similares: bajos BC y *hop plot*, y altos *average degree distribution* y *cluster coefficient*. Además, si bien es una cita de la entrevista, el motivo de la publicación es parecido al de una *funa*, dado que de cierto modo se expone al ex ministro.

La siguiente *funa* analizada es contra Edwin Dimter, autor del homicidio y secuestro de Victor Jara¹⁴ en la dictadura militar. Actualmente el procedimiento de ejecución de la sentencia se encuentra paralizado, por lo cual se realiza una *funa* con el siguiente contenido: “*Si no hay justicia, hay funa*” junto a una imagen que indica: “*Edwin Dimter*

¹³ex ministro de Defensa del Interior de Chile

¹⁴cantautor chileno

Bianchi, asesino de Víctor Jara". Al inicio se encuentran varios usuarios influyentes que difunden la información a distintas personas, de lo cuales algunos no logran difundir la información más allá. Mientras que otros propagan la *funa* en forma de cadena. Por tanto, hay usuarios que tienen varias fuentes para informarse. Posteriormente, aumenta la interconexión y agrupación entre los usuarios, formando un gran grupo central, en donde la mayoría de las personas son los que influyen en la propagación. También se observan ciertos grupos a la periferia, que no difunden la publicación. Dicho comportamiento se mantiene hasta el final de la difusión, solo sigue aumentando la interconexión y cercanía entre los usuarios. La forma de difusión es similar a la de la primera *funa* analizada y al trend relacionado con el ex ministro. Se forma un grupo central cohesionado, donde las personas que lo conforman difunden la *funa*.

La siguiente publicación analizada es la de José Antonio Kast, postulante a presidente de Chile en el año 2021. El trend obtenido hace referencia a la despenalización del aborto hasta las 14 semanas, que Kast no apoya. Hace mención al presidente de turno, quien es de su coalición, por no haber comentado respecto a la aprobación. La publicación dice lo siguiente: *"No me sorprende, pero me decepciona aún más. Han pasado 24 horas y Sebastián Piñera aún no dice nada por la aprobación de la Ley de Aborto en Chile."* En un principio, la difusión se debe principalmente por usuarios influyentes que divulgan la información a conjuntos de personas independientes. Estas personas pueden enterarse del trend a través de uno o varios influyentes. Luego, se observa un cambio en el comportamiento, dado que aumenta la interconexión de los usuarios, estando más cerca entre ellos. La publicación es propagada por la mayoría de los usuarios, amenguando el protagonismo de los influyentes. De igual forma se observan conjuntos de usuarios a la periferia, que solo se informan. Llama la atención el comportamiento recién descrito, dada las abruptas variaciones de las características (especialmente del BC) y el alto valor del *average degree distribution*. En el trend se espera que una persona difunda o se informa a través de otras 120, y en Twitter solo se espera que un 3.52 % siga entre 101 y 200 personas [63]. Además, el *betweenes centrality* disminuye bruscamente, mostrando que muchos de los usuarios se encuentran conectados entre sí de un momento a otro.

La última *funa* analizada es la siguiente: *"El lanzamiento de fuegos artificiales hacia los carabineros era constante por parte de los violentistas. No era de extrañar que tarde o temprano ocurriría una tragedia. #PrimeralíneaAsesina"*. A diferencia de las anteriores, esta *funa* es contra los manifestantes que intentaron herir a carabineros con fuegos artificiales. Resultado del ataque, salieron varios policías heridos y una mujer fallecida. En un comienzo la difusión se debe a varios usuarios influyentes. Las personas se enteraban de la *funa* a través de uno o varios influyentes. Hasta que llegó un punto que las personas se interconectaron entre ellas y empezaron a propagar también. Los influyentes perdieron el protagonismo, aunque igual seguían informando a conjuntos de usuarios que solo lograron informar. Se formó un gran grupo central cohesionado, conformado por usuarios que difunden la *funa*. Similar al trend de Kast, se presencia cambios abruptos de las características de la red, y llama la atención el excesivo *degree distribution* y bajo *betweenes centrality*.

Tras haber realizado los análisis respectivo de cada red y sus características (*average cluster coefficient*, *average degree distribution*, *average hop plot* y *average betweenes centrality*), se fueron observando características similares entre los siguientes conjuntos de redes: (1) trend esperados, (2) las *funas* y trend consideras *funas* y (3) los tweet anómalos.

5.7.1. Características de redes trend esperadas de Twitter

Según la literatura se considera el trend de la selección chilena de fútbol como una esperada dentro de Twitter. Algunas características que se destacan son las siguientes:

- La difusión inicial del trend se debe por un usuario influyente, por tanto, juegan un papel crucial para la propagación de la información.
- Dentro de la primera hora, la red cuenta con bajos *degree distribution*, *clustering coefficient*, *betweenes centrality* y *hop plot*.
- Dentro de las primeras dos horas de la difusión, la información alcanza a grupos alejados del usuario central, entonces se encuentra en la máxima distancia.
- Dentro de las primeras 5 horas, se alcanza el 50 % o más, de las aristas en la red.
- Se estabilizan las características en un bajo *average degree distribution* (conexiones) y *average cluster coefficient* (agrupación), y un alto *average betweenes centrality* y un *hop plot* (distancia) esperado de seis personas.
- Se observa conjuntos de usuarios que solo se informan y no logran difundir la información a otras. Esto genera varios grupos que difunden la información dentro del mismo.

Complementando con lo ya señalado, se realizó un análisis de la cantidad de usuarios que retweetearon el trend según el transcurso del tiempo. El 50 % de las personas retweetearon dentro de las primeras 18 horas de iniciado el primer tweet. Y a las 20 horas transcurridas, se alcanza el 95 %.

5.7.2. Características de redes tipo funa

A continuación, se presentarán puntos en común de las siguientes redes: las *funas* a Patricio Maturana y Edwin Dimter, y la publicación de la entrevista al ex ministro. Una de las dificultades de encontrar *funas* en Twitter, es que no es tan común encontrar alguna que sea difundida por más de 600 personas y que, además, contenga la palabra "*funa*". Además, la api de Twitter limitaba la captura de información con un rango de 7 días, por tanto, era necesario estar atenta constantemente a la red social. Asimismo, llamó la atención que la *funa* a Maturana fue posteriormente eliminada de Twitter, lo cual pudo ser debido al vocabulario usado ("*mutilador*"), o bien, alguien reportó la publicación.

Luego de realizar los análisis de sus características, se encuentran los siguientes puntos en común de las *funas* y la publicación de la entrevista de Mónica Rincón al ex ministro:

- La difusión inicial del tweet se debe por usuarios influyentes, por tanto, juegan un papel crucial para la propagación de cualquier publicación.
- Dentro de las primeras horas de difusión, las redes tienen un comportamiento similar: bajo *degree distribution* y *clustering coefficient*, *hop plot* cercano a 6.0 y alto *betweenes centrality*. Estas características son muy similares a las mostradas en el trend de la selección chilena.

- Dentro de las primeras dos horas de difusión, la red se encuentra en la máxima distancia. Lo que coincide con que en un inicio las personas solo se conectan por el influyente.
- Luego de algunas horas se forma un gran grupo central y cohesionado, cuyo núcleo de difusión son las personas.
- Se observa que hasta alrededor de las 10 horas, aumenta la cantidad de conexiones promedio y la agrupación (entre un 30 % y 40 %), bajo BC (cercano a 100) y distanciamiento (5 - 4 saltos). Vale destacar que las características de la publicación de la entrevista, siguen con la misma tendencia hasta las 20 horas.
- Existen conjuntos de usuarios en la periferia que no logran propagar la funa. Ellos se informan principalmente por los usuarios que fueron influyentes.

Complementando con lo ya señalado, se realizó un análisis de la cantidad de usuarios que retweetearon la *funa* según el transcurso del tiempo. Destaca el hecho que dentro de las primeras seis horas el 70 % del total de personas ya la habían retweeteado. Por otro lado la publicación de la entrevista alcanza el 70 % cerca de las 12 horas.

Vale mencionar que en comparación a las funas, las características del tweet del ex ministro varía de forma muy similar. Solo con la salvedad que las variaciones surgen entre un tiempo más largo: de 10 a 20 horas, mientras que las *funas* entre 5 a 10 horas. En ambos casos se inicia la difusión con usuarios influyentes que luego pierden protagonismo, formándose un gran grupo central cohesionado

5.7.3. Características de redes anómalas

A continuación, se presentarán las características en común de las siguientes redes: trend de Jose Antonio Kast y la *funa* a la primera línea asesina. La recolección de información en este caso no fue complicado, a excepción de su análisis. Llevó más de un mes para poder avanzar lo máximo posible de su análisis.

Luego de realizar los análisis de sus características, se encuentran los siguientes puntos en común:

- La difusión inicial del tweet se debe por usuarios influyentes, por tanto, juegan un papel crucial para la propagación de información.
- Dentro de la primera hora, la red cuenta con las siguientes características: baja cantidad de conexiones promedio, la máxima distancia alcanzada, bajo porcentaje de agrupación y alto BC.
- Las variaciones de las características se concentran entre la primera y tercera hora, con la salvedad del *average degree distribution*. En tres horas, la distancia varía de 6 a 4 saltos, y en una hora, el BC disminuye a un valor extremadamente pequeño y la agrupación aumenta considerablemente. El BC cambia de 550 a 60 y el porcentaje de agrupación de un 0 % a un 30 %. Mientras que la cantidad de relaciones está en un aumento pronunciado: en la publicación de kast alcanza las 120 conexiones, y en la *funa* 70.
- Luego de la primera hora de difusión, la red forma un grupo central y cohesionado por el cual se difunde la información a través de los distintos usuarios que lo conforman.

- Se observan conjuntos de personas que solo se informan y no logran difundir la información a otras.

Complementando con lo ya señalado, se realizó un análisis de la cantidad de usuarios que retweetearon la publicación según el transcurso del tiempo. En el trend de José Antonio Kast, el 50 % de las personas, difunde dentro de las primeras 4 horas. En las siguientes 6 horas, alcanza el 90 %. A diferencia de la funa, el 50 % de las personas retweetea dentro de las primeras 11 horas y el 90 % se alcanza a las 20 horas.

5.8. Análisis de Usuarios

Complementando los análisis de las redes, se desarrolla un estudio de los usuarios que las conformaron. Se busca responder la pregunta: ¿son los mismos usuarios que difunden las funas? Para esto se realizó una matriz que indica las publicaciones que difundió cada usuario. En total son seis redes analizadas con un acumulado de 5,993 usuarios distintos.

Además, llamó la atención la cantidad esperada de amistades y la agrupación de las redes de José Antonio Kast y de la entrevista a Francisco Vidal. Entonces, se realizó un análisis para poder identificar si la difusión tiene involucrado bots. Los bots en Twitter son cuentas controladas por una aplicación para difundir o publicar cosas con cierto propósito [65]. Una de las características para identificarlos, es que retweetea muchas publicaciones de un mismo autor. Por tanto, se extrajeron otras 3 publicaciones de cada autor y se comparó la cantidad de usuarios que se mantuvieron en la difusión de una publicación a otra (subsección 5.8.2).

5.8.1. ¿Quién mantiene la difusión?

A continuación se analizará la frecuencia en la que participa cada uno de los usuarios de las seis redes analizadas. Los análisis se basan en dos conjuntos distintos de personas: en el primer gráfico (figura 5.37a) se considera a todos los usuarios que participaron en la difusión de las seis redes analizadas, y en el resto de gráficas se considera solo a los usuarios que participaron en la difusión de las *funas*.

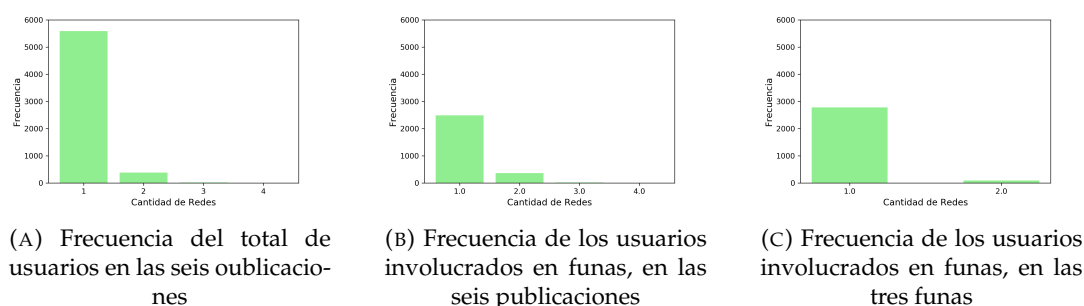


FIGURA 5.37: Frecuencia en la participación de los usuarios

En la izquierda de la figura 5.37a, se observa que la mayoría de los usuarios participaron solo en una de las redes analizadas. Siendo un grupo muy minoritario los que participaron en dos o más de las redes. Este comportamiento se repite cuando se analiza solo a los usuarios que difundieron alguna funa (figura 5.37b central). En resumen, la cantidad de personas que difunden dos publicaciones (ya sea funa o noticia) es bastante reducida.

En la figura 5.37c se observa las mismas conclusiones: la mayoría mantiene la difusión de una sola funa, y la minoría mantiene dos. Por tanto, son pocas las personas que divulgan más de una funa. Surge la duda si la difusión de dos funas corresponden a las en contra de ex oficiales. Entonces, se identifican a las personas que retweetearon la funa contra los civiles, y se observa que solo difunden una sola funa, y su cantidad corresponde a más de la mitad de los usuarios que propaga una sola publicación.

5.8.2. ¿Bots?

Cuando se analizó la red de la entrevista de Mónica Rincón a Francisco Vidal, sorprendió las características alcanzadas. Al final del análisis, se encuentra que el promedio de relaciones que tiene un usuario es de 30 personas, y además, se cuenta con una agrupación de un 40 %. Por tanto, se procedió a obtener otras dos publicaciones del mismo autor (Mr Wolf), para evaluar si es que se mantenía las mismas personas en su difusión.

Dado el tiempo adicional que podría tomar, no se hicieron análisis de las características de la red. Sin embargo, se obtuvieron cuántos fueron los usuarios que participaron de la difusión. Las publicaciones dicen lo siguiente:

1. *"La base de la impunidad (1-2). Francisco Vidal dice apoyar a Peñalillo en su carrera senatorial xq no fue formalizado ni querellado y @tv_monica le dice: "Sr Vidal, son las 7, son pasadas las 12 la hora en que somos tontos"*
2. *"Así o más patético. Diego Schalper no para de dar jugo y hoy en la previa para discutir el proyecto q despenaliza el aborto pide la palabra para exigir un 1 min de silencio por la víctimas del aborto. Pase y vea este vergonzoso show. #AbortoLegal "*
3. *"Aprobado en gral el proyecto que despenaliza el aborto hasta las 14 semana y @camila_vallejo tuvo una d esas notables intervenciones para enmarcar #AbortoLegal "No puedo tolerar q se hable q tenemos un pacto con la muerte, q estamos aprobando el asesinato y q somos asesinos"*

La primera publicación, fue una de las analizadas con anterioridad, hacía mención a la frase de la entrevistadora, Mónica Rincón, cuando Francisco Vidal dijo apoyar a una persona que había sido investigada por boletas falsas. Las otras dos publicaciones se sitúan en las discusiones que se hicieron en torno a la despenalización del aborto en 14 semanas. Dado el hashtag #AbortoLegal, se concluye que apoya la causa. Las publicaciones cuentan con 1797, 1298 y 1697 usuarios, respectivamente. Del total de usuarios, 77 publicaron las tres publicaciones (correspondientes al 4.2 %, 5.9 %, y 4.5 % del total de usuarios de cada red). Sin embargo, este número aumenta en forma considerable cuando se analiza si han retweeteado al menos dos publicaciones. Un 15.8 % difunde el primer tweet y una de las otras dos publicaciones, mientras que estos valores cambian a 24.6 % y 19.8 % cuando se considera la segunda publicación y la tercera publicación como obligatoria y una de las otras como posibles.

En comparación al análisis previo de la frecuencia de todas las redes, no se observan grandes anomalías en los porcentajes de difusión de las tres. Sin embargo, se observan altos valores en el porcentaje de frecuencia de dos publicaciones. Por tanto, a primera vista, no se sospecha de una clara presencia de bots, pero sí sería interesante un análisis más profundo para mayor certeza.

Asimismo, cuando se analizó la publicación de José Antonio Kast, también hubo sospecha de bots. Entre la primera y tercera hora, se muestra variaciones repentinas de la mayoría de las características analizadas. Además, en el periodo final de difusión, se espera que una persona esté conectada con otras 120, y un 41 % de agrupación, características muy extrañas. Por tanto, se procede a analizar otras publicaciones realizadas por el autor.

Se replicó el experimento mencionado anteriormente. Es decir, se capturaron otros tres tweets, sin analizar las características de la red, pero sí la información de los usuarios. Las publicaciones dicen lo siguiente:

1. *"No me sorprende, pero me decepciona aún más. Han pasado 24 horas y Sebastián Piñera aún no dice nada por la aprobación de la Ley de Aborto en Chile."*
2. *"La Diputada Camila Vallejos justifica el asesinato de niños pequeños diciendo que estar en contra del aborto es totalitarismo religioso. Esto no tiene nada que ver con religión, son derechos humanos y el derecho a la vida es el más importante de todos."*
3. *"Dos modelos de país: Boric, que promueve el aborto libre, que elimina la libertad de enseñanza, destruye las pensiones y abre las fronteras del país. Kast, en contra del aborto, a favor de la libertad de enseñanza, fortalece las pensiones y protege las frontera"*

En los dos primeros tweet, Kast hace alusión a la aprobación de la despenalización del aborto en 14 semanas, y se observa su desaprobación al usar las palabras "asesinato de niño" y "me decepciona". El último tweet, se refiere a su contrincante en las postulaciones a presidente 2022 (Gabriel Boric). Cada tweet cuenta con 1195, 1817 y 1299 personas en la red, respectivamente.

Del total de usuarios, 189 retweetearon las tres publicaciones (correspondientes al 15.8 %, 10.4 % y 14.5 % del total de usuarios de cada red). Sin embargo, este número aumenta en forma considerable cuando se analiza si han retweeteado al menos dos publicaciones. Un 29.8 % difunde la primera y una de las otras dos publicaciones, un 20.4 % difunde la segunda y una de los otros tweet, y un 32.4 % la tercera y otra más.

En comparación al análisis de la publicación de Mónica Rincón, el porcentaje de difusión de las tres publicaciones, crece entre dos a cuatro veces. Esto podría indicar la existencia de bots, lo cual sería interesante de analizar en más profundidad. Además, el porcentaje de participación de dos publicaciones también aumenta, lo cual apoya la idea.

Capítulo 6

Conclusiones Generales

Se comienza la investigación manifestando que las *funas* han ganado importancia en las redes sociales, dado que su difusión es crucial para advertir a las comunidades del posible peligro al relacionarse con el/la agresor/a. Una de las redes sociales más utilizadas con el propósito de difundir rápidamente información es Twitter, dada la funcionalidad que ofrece la aplicación y el interés de las personas que lo conforman. También se sostuvo que a pesar de que sea un tema contingente en nuestra actual sociedad, en realidad, no se conoce su real alcance ni sus características de difusión. He ahí la razón de la investigación, que a través del análisis de trend y *funas* se busca comprobar que la funa se propaga de igual o mayor rapidez que un trend promedio en Twitter.

Se analizó las características de difusión de seis publicaciones de Twitter: tres corresponden a *funas* y las otras a publicaciones tendencias. Las *funas* analizadas se realizaron en contra de un ex militar, un ex carabinero y civiles en una marcha. Estas publicaciones fueron difundidas entre 600 y 1,400 usuarios, y contaron entre 8,000 y 55,000 aristas. Se observó que la mayoría de las personas retweetean la funa dentro de las 19 horas. Por otro lado, los trend constaron de una publicación de un postulante a presidente durante el año 2021 (José Antonio Kast), una entrevista en apoyo a un ex ministro de interior y de la selección chilena de fútbol. Estas publicaciones fueron difundidas entre 500 y 1,800 usuarios y formaron entre 1,000 a 88,000 relaciones. Se observó que la mayoría de los usuarios retweetearon los trend dentro de 21 horas.

Para realizar este análisis se utilizó la API de Twitter. A partir de ella se capturaron a las personas que propagaron cada una de las publicaciones. Para generar las conexiones de la red, que representan la difusión como entrega o recepción de información, se utilizó a los seguidores de cada usuario. Cada vez que una persona retweeteaba uno de los tweets seleccionados, se conectaba al nuevo usuarios con todos los usuarios que sigue que habían publicado este tweet (generando una o más relaciones). Una vez con la red generada, se procedió a analizar sus características a través del tiempo y formación de aristas.

Las características analizadas para esta tesis fueron el *average degree distribution*, el *average cluster coefficient*, el *average hop plot* y el *average betweenes centrality*. El objetivo, como se sostuvo, fue observar el comportamiento de difusión de las funas en Twitter a través del tiempo, y comprobar si se propaga de igual manera que un tweet cualquiera, o incluso más rápido.

Según los análisis realizados, se descubrió que todas las difusiones son iniciadas por usuarios influyentes, generando en las primeras horas de difusión una red con

baja cantidad de conexiones, baja cohesión, un alto distanciamiento y un alto BC. Vale destacar que en un principio se observa que las redes cuentan con 6 grados de separación, lo que coincide con la propiedad del modelo del *Mundo Pequeño de Watts and Strogatz*. Luego se observó que dentro de las primeras 5 horas se alcanza, en promedio, el 50 % de las aristas finales de la red, y a las 10 horas se observa que las características de la red se estabilizan o sufren variaciones menos abruptas. Además, en general, se advierte un aumento en la cantidad de conexiones y en el porcentaje de agrupación a medida que aumenta el tiempo. Por último, en todas las redes se observó una fuerte presencia de usuarios que solo se informan y no logran informar al resto.

Al comparar los distintos comportamientos, se advirtió que luego de la primera hora, la difusión del trend común de Twitter se mantiene por un usuario principal. Sin embargo, en las *funas* y tweet anómalos, la difusión se da por un gran grupo altamente interconectado, donde no se puede determinar a través de quién un nuevo usuario se informa.

Al analizar los tiempos de difusión, el trend común tiene una propagación lenta, donde recién a las 18 horas el 50 % de los usuarios retweetearon. Esto varía en comparación a las *funas*, particularmente contra el ex militar y ex carabinero, y a la entrevista de Mónica Rincón, donde a las 4 horas ya se alcanzaba el 50 % de retweets. Luego, en el caso de las *funas* referenciadas, a las 6 horas ya se alcanzó el 70 % de retweets, mientras que el otro tweet (entrevista) alcanzó dicho porcentaje a las 12 horas. En comparación a las redes consideradas anómalas (*funa* de la primera línea y Kast), la primera alcanzó el 50 % de los usuarios a las 11 horas, mientras que la red de Kast alcanza a las 4 horas el 50 %, y a las 10 horas ya tenía un 90 % de los usuarios. En ambos tweet se observa una variación abrupta de sus características.

Finalmente la investigación se dirigió a analizar la cuestión de si acaso las distintas *funas* fueron difundidas por los mismos usuarios. Se observó que solo un 3 % del total de usuarios difundieron las tres *funas*, y solo un 7 % difunde las *funas* contra ex oficiales. Estos valores son bastante bajos comparados con las personas que retweetearon diferentes tweet de un mismo autor. Al analizar otros tres tweets de los usuarios de Mr Wolf (autor del tweet de la entrevista al ex ministro) y Jose Antonio Kast, se obtuvo que un 24 % difunde al menos dos tweets de Mr Wolf, y un 32 % los de Kast. Ambos son porcentajes más altos de los esperados, pero llamó la atención que el trend de Kast sufre cambios bruscos de sus características, por lo que se presume el uso de bots para su difusión. Se identifica un interesante camino de investigación para definir la difusión de información mediante bots.

Recapitulando, la *funa* puede difundirse con mayor o igual rapidez que un trend común. Las difusiones inician con usuarios influyentes, y, a diferencia de un trend común, se observa que se propaga gracias a varios usuarios de la red. Se advierte que a medida que avanza el tiempo, se forma un gran grupo central cohesionado, el cual esta integrado por varios usuarios. Además, son pocos los que solo se informan y no logran difundir la información.

Se espera que la presente investigación haya logrado responder con fundamentos las interrogantes: ¿cómo se propaga una *funa* en Twitter? ¿cuál es su real alcance? ¿es más rápido que un trend promedio? Gracias a la metodología utilizada, la respuesta anticipada es que la *funa* se difunde en grupo con igual o incluso con mayor rapidez que un trend promedio. Se cree que esta investigación abre puertas para profundizar

en algún modelo de propagación de *funas* en Tiwtter, o derivar a lo que presagiamos de bots en tendencias políticas chilenas.

Bibliografía

- [1] C. Schmeisser, "La funa: Aspectos históricos, jurídicos y sociales," tech. rep., Universidad de Chile, 2019.
- [2] S. Trafilaf, ""si no hay justicia, hay funa": Se reactiva la "comisión funa"," *diarioUchile*, 2014.
- [3] J. Kovalskys, "Trauma social, modernidad e identidades sustraídas: Nuevas formas de acción social," *Psykhé*, vol. 15, pp. 13 – 24, 2006.
- [4] I. Bayo, J. Fuertes, and O. Menéndez, "La comunidad científica ante las redes sociales," tech. rep., Universidad Complutense Madrid, 2019.
- [5] A. Java, X. Song, T. Finin, and B. Tseng, *Advances in Web Mining and Web Usage Analysis*, ch. Why We Twitter: An Analysis of a Microblogging Community, pp. 118–138. Springer Berlin, 2007.
- [6] P. Ferragina, F. Piccinno, and R. Santoro, "On analyzing hashtag in twitter," tech. rep., University of Pisa, 2015.
- [7] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," in *19th International Conference on World Wide Web*, p. 591–600, 2010.
- [8] N. Ellison and D. Boyd, *The Oxford Handbook of Internet Studies*, ch. Sociality Through Social Network Sites, pp. 151–172. Oxford University Press, 2013.
- [9] F. Freire, "Las redes sociales trastocan los modelos de los medios de comunicación tradicionales," *Revista Latina de Comunicación Social*, vol. 11, pp. 277–268, 2008.
- [10] C. Steinmetz, H. Rahmat, N. Marshall, K. Bishop, S. Thompson, M. Park, L. Corkey, and C. Tietz, "Liking, tweeting and posting: An analysis of community engagement through social media platforms," *Urban Policy and Research*, pp. 1–21, 2020.
- [11] D. Boyd, S. Golder, and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter," in *43rd Hawaii International Conference on System Sciences*, pp. 1 – 10, 2010.
- [12] K. Weller, E. Dröge, and C. Puschmann, "Citation analysis in twitter: Approaches for defining and measuring information flows within tweets during scientific conferences," in *Workshop on 'Making Sense of Microposts': Big things come in small packages (ESWC)* (M. Rowe, M. Stankovic, A. Dadzie, and M. Hardey, eds.), vol. 718, pp. 1–12, 2011.
- [13] M. Cordeiro, R. Sarmiento, P. Brazdil, and J. Gama, *Social Media and Journalism - Trends, Connections, Implications*, ch. Evolving Networks and Social Network Analysis Methods and Techniques, pp. 101–134. IntechOpen, 2018.

- [14] P. Erdős and A. Rényi, "On random graphs i.," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.
- [15] P. Erdős and A. Rényi, "On the strength of connectedness of a random graph," *Acta Mathematica Hungarica*, vol. 12, pp. 261–267, 1961.
- [16] M. Panda, *Social Networking Mining, Visualization, and Security*, vol. 65, ch. Structure and Evolution of Online Social Networks, pp. 23–41. Springer International Publishing, 2014.
- [17] O. Frank and D. Strauss, "Markov graphs," *Journal of the American Statistical Association*, vol. 81, pp. 832–842, 09 1986.
- [18] T. Snijders and K. Nowicki, "Estimation and prediction for stochastic block-models for graphs with latent block structure," *Journal of Classification*, vol. 14, pp. 75–100, 1997.
- [19] T. Snijders, "Markov chain monte carlo estimation of exponential random graph models," *Journal of Social Structure*, vol. 3, pp. 1–40, 2002.
- [20] C. Seshadhri, T. Kolda, and A. Pinar, "Community structure and scale-free collections of erdős-rényi graphs," *Physical Review E*, vol. 85, p. 056109, 2012.
- [21] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–12, 1999.
- [22] P. Krapivsky, G. Rodgers, and S. Redner, "Degree distributions of growing networks," *Physical Review Letters*, vol. 86, pp. 5401–5404, 2001.
- [23] E. Ben-Naim and P. Krapivsky, "Popularity-driven networking," *Europhysics Letters (EPL)*, 2011.
- [24] A. Abbasi, L. Hossain, and L. Leydesdorff, "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks," *Journal of Informetrics*, vol. 6, pp. 403 – 412, 2011.
- [25] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, p. 440–442, 1998.
- [26] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," in *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 462–470, 2008.
- [27] J. Bagrow and D. Brockmann, "Natural emergence of clusters and bursts in network evolution," *Physical Review X*, vol. 3, p. 021016, 2013.
- [28] W. Aiello, F. Chung, and L. Lu, "A random graph model for power law graphs," *Experimental Mathematics*, vol. 10, pp. 53–66, 2001.
- [29] P. Guimarães, M. de Aguiar, J. Bascompte, P. Jordano, and S. dos Reis, "Random initial condition in small barabasi-albert networks and deviations from the scale-free behavior," *Physical review E: Statistical, nonlinear, and soft matter physics*, vol. 71, p. 037101, 2005.
- [30] M. Medo, G. Cimini, and S. Gualdi, "Temporal effects in the growth of networks," *Physical Review letters*, vol. 107, p. 238701, 2011.
- [31] A. Bonato, N. Hadi, P. Horn, P. Pralat, and C. Wang, "Models of online social networks," *Internet Mathematics*, vol. 6, pp. 285–313, 2009.

- [32] A. Bonato, J. Janssen, and P. Pralat, "A geometric model for on-line social networks," in *Algorithms and Models for the Web-Graph*, vol. 6516, pp. 110–111, 2010.
- [33] H. Kim and R. Anderson, "Temporal node centrality in complex networks," *Physical review E: Statistical, nonlinear, and soft matter physics*, vol. 85, p. 026107, 2012.
- [34] A. Casteigts, P. Flocchini, N. Santoro, and W. Quattrociocchi, "Time-varying graphs and dynamic networks," *International Journal of Parallel, Emergent and Distributed Systems*, Taylor & Francis, vol. 6811, pp. 387–408, 2012.
- [35] W. Thompson, P. Brantefors, and P. Fransson, "From static to temporal network theory: Applications to functional brain connectivity," *Network Neuroscience*, vol. 1, pp. 1–56, 2017.
- [36] M. Yousuf and S. Kim, "Generating graphs by creating associative and random links between existing nodes," *Journal of Statistical Physics*, vol. 179, pp. 1–32, 2020.
- [37] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 611–617, 2006.
- [38] A. Mislove, H. Koppula, K. Gummadi, P. Druschel, and B. Bhattacharjee, "Growth of the flickr social network," in *1st Workshop on Online Social Networks (WOSN)*, p. 25–30, 2008.
- [39] D. Romero and J. Kleinberg, "The directed closure process in hybrid social-information networks, with an analysis of link formation on twitter," in *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.
- [40] H. Huang, J. Tang, S. Wu, L. Liu, and X. Fu, "Mining triadic closure patterns in social networks," in *23rd International Conference on World Wide Web*, pp. 499–504, 2014.
- [41] J. Fox, T. Roughgarden, C. Seshadhri, F. Wei, and N. Wein, "Finding cliques in social networks: A new distribution-free model," *SIAM Journal on Computing*, vol. 49, p. 55, 2018.
- [42] S. Ahmadian and S. Haddadan, "The wedge picking model: A dynamic graph model based on triadic closure." preprint by S. Haddadan supported by a Google Focused Research Award, SIR Grant RBSI14Q743, and NSF Award CCF-174074, 2020.
- [43] H. Young, "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning," *American Economic Review*, vol. 99, pp. 1899–1924, 2009.
- [44] R. Anderson and R. May, "Population biology of infectious diseases: Part i," *Nature*, vol. 280, pp. 361–7, 1979.
- [45] W. Royce, "Managing the development of large software systems," tech. rep., Institute of Electrical and Electronics Engineers, 1970.
- [46] J. Woo, J. Son, and H. Chen, "An sir model for violent topic diffusion in social media," in *IEEE International Conference on Intelligence and Security Informatics*, pp. 15–19, 2011.

- [47] C. Zhao, Z. Zhang, H. Li, and S. Zhao, *Genetic and Evolutionary Computing. Advances in Intelligent Systems and Computing*, vol. 834, ch. An Adaptive Model Parameters Prediction Mechanism for LCU-Level Rate Control, pp. 275–282. Springer Singapore, 2014.
- [48] J. Wan, X. Chen, Y. Du, and M. Jia, “Information propagation model based on hybrid social factors of opportunity, trust and motivation,” *Neurocomputing*, vol. 333, pp. 169 – 184, 2019.
- [49] K. Lerman, R. Ghosh, and T. Surachawala, “Social contagion: An empirical study of information spread on digg and twitter follower graphs,” *Journal of Clinical Orthopaedics and Related Research*, pp. 1–17, 2012.
- [50] S. Gaito, M. Zignani, G. Rossi, A. Sala, X. Wang, H. Zheng, and B. Zhao, “On the bursty evolution of online social networks,” in *First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, pp. 1–8, 2012.
- [51] P. Singer, C. Wagner, and M. Strohmaier, “Understanding co-evolution of social and content networks on twitter,” in *2nd Workshop on Making Sense of Microposts held in conjunction with the 21st World Wide Web Conference*, pp. 57–60, 2012.
- [52] M. ten Thij, T. Ouboter, D. Worm, N. Litvak, H. Berg, and S. Bhulai, “Modelling of trends in twitter using retweet graph dynamics,” in *Algorithms and Models for the Web Graph, 11th International Workshop*, pp. 132–147, 2015.
- [53] G. Dong, W. Yang, F. Zhu, and W. Wang, “Discovering burst patterns of burst topic in twitter,” *Computers & Electrical Engineering*, vol. 58, pp. 551–559, 2017.
- [54] X. Yan, J. Guo, Y. Lan, and X. Cheng, “A biterm topic model for short texts,” in *22nd International Conference on World Wide Web*, pp. 1445–1456, 2013.
- [55] X. Yan, J. Guo, Y. Lan, J. Xu, and X. Cheng, “A probabilistic model for bursty topic discovery in microblogs,” in *29° AAAI Conference on Artificial Intelligence*, p. 353–359, 2015.
- [56] M. Mathioudakis and N. Koudas, “Twittermonitor: Trend detection over the twitter stream,” in *ACM SIGMOD International Conference on Management of Data*, pp. 1155–1158, 2010.
- [57] C. Li, A. Sun, and A. Datta, “Twevent: Segment-based event detection from tweets,” in *21st ACM international conference on Information and knowledge management*, pp. 155–164, 2012.
- [58] T. Hofmann, “Probabilistic latent semantic analysis,” 2013.
- [59] D. Blei, A. Ng, and M. Jordan, “Latent dirichlet allocation,” *Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [60] J. Hemsley, M. Garcia-Murillo, and I. MacInnes, “Tweets that resonate: Information flows and the growth of twitter’s universal basic income discussion space,” *Policy & Internet*, vol. 10, pp. 324–346, 2018.
- [61] K. King and B. Wang, “Diffusion of real versus misinformation during a crisis event: A big data driven approach,” *International Journal of Information Management*, p. 102390, 2021.

-
- [62] D. Chakrabarti, C. Faloutsos, and M. McGlohon, *Managing and Mining Graph Data*, vol. 40, ch. Graph Mining: Laws and Generators, pp. 69–123. Springer, Boston, MA, 2010.
 - [63] Sysomos, “Inside twitter: An in-depth look inside the twitter world,” tech. rep., Sysomos, 2009.
 - [64] H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?,” in *Proceedings of the 19th International Conference on World Wide Web*, p. 591–600, 2010.
 - [65] J. Rodriguez, J. Mata, R. Monroy, O. Loyola, and A. Lopez, “A one-class classification approach for bot detection on twitter,” *Computers and Security*, vol. 91, p. 101715, 2020.