Aplicación de la teoría de grafos como herramienta aplicada al análisis del comportamiento social ante un desastre en la red social Twitter.

Juan Guillermo Martínez Cano, Robinson Ortiz Sierra

Trabajo de Grado para Optar el título de Ingeniero Industrial

Director

Daniel Orlando Martínez Quezada

MSc. Ingeniería Industrial

Codirector

Henrry Lamos Díaz

PhD. Física-Matemática

Universidad Industrial de Santander

Facultad de Ingenierías Físico-Mecánicas

Escuela de Estudios Industriales y Empresariales

Bucaramanga

2018

Dedicatoria

A Dios, por ser mi guía en todos los momentos de mi vida, porque con su amor y su infinita misericordia me ha permitido consolidar las metas que me he propuesto a lo largo de mi vida.

A mi madre Veronica, por haber sido mi apoyo incondicional en todo momento, por ser la mejor maestra de mi vida, por enseñarme con su ejemplo y consejos a ser una persona de bien, por la motivación constante que me han permitido seguir adelante con mis propósitos y sobre todo por brindarme su inmenso amor.

A mi padre Alejandro, por ser mi mayor ejemplo de fortaleza y perseverancia, por enseñarme a no rendirme a pesar de las circunstancias, por sus innumerables enseñanzas e inigualable amor, lo que me ha permitido nunca desistir de mis sueños y luchar hasta el final para consolidar las metas que me he propuesto.

A mis hermanos Edimer, Albeiro, Alejandro y Albert, por ser mi ejemplo y brindarme su incondicional apoyo, cariño y alegrías a lo largo de mi vida.

A Juan Guillermo, por ser el mejor compañero del proyecto.

A nuestro director, por su constante apoyo, múltiples enseñanzas y disposición a lo largo de la realización del proyecto.

A los compañeros del grupo OPALO, por su apoyo y enseñanzas; que permitieron la realización de este proyecto.

Al grupo OPALO por incentivar y promover la investigación, y por brindar las herramientas necesarias para la realización del proyecto.

A mis amigos Fabián, Lina Lozano, Lina Sarmiento, Yesika, Leydi. Elizabeth, Wilmer, Heyder, Durley, Pedro, María, Carlos, Karol, Silvia y Mafe por acompañarme en ésta etapa y contribuir en mi formación

A la gloriosa Universidad Industrial de Santander la cual me ha brindado a través de sus docentes las herramientas necesarias para poder desenvolverme en la vida y ser un buen profesional.

Robinson Ortiz Sierra

Dedicatoria

A Dios.

Por haberme permitido llegar hasta este punto y haberme dado salud para lograr mis objetivos, además de su infinita bondad y amor.

A mi madre Nidia.

Por haberme apoyado en todo momento, por ser el pilar fundamental en todo lo que soy, en toda mi educación, tanto académica, como de la vida, por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien, pero más que nada, por su amor.

A mis amigos.

Que nos apoyamos mutuamente en nuestra formación profesional y que hasta ahora, seguimos siendo amigos

Juan Guillermo Martínez Cano

Tabla de contenido

Lista de Tablas

Lista de Figuras

Resumen

TÍTULO: APLICACIÓN DE LA TEORÍA DE GRAFOS COMO HERRAMIENTA APLICADA AL ANÁLISIS DEL COMPORTAMIENTO SOCIAL ANTE UN DESASTRE EN LA RED SOCIAL TWITTER*

AUTORES: ORTIZ SIERRA Robinson
MARTÍNEZ CANO Juan Guillermo**

PALABRAS CLAVES: *Redes Sociales, Twitter, Desastres, Análisis de Redes Sociales y Teoría de grafos.*

DESCRIPCIÓN:

Twitter se ha convertido en una herramienta importante para conocer en tiempo real lo que sucede en la sociedad en la que convivimos; de hecho, esta plataforma es cada vez más atractiva como medio de comunicación social lo que le permite tener un papel importante en situaciones de desastre. Éste interés en la plataforma puede usarse para ayudar a solucionar problemas secundarios al evento (ej: escases de suministros) o dar información que permita a terceros ayudar con los organismos de atención inmediata (qué, dónde y cómo) así como información para los afectados, como la posición de albergues y las rutas de evacuación.

En esta investigación se aplican métodos de minería de datos en Twitter con el propósito de analizar los mensajes relacionados con un evento de desastre sísmico, describiendo un proceso automatizado para el análisis de la información recolectada utilizando un enfoque de teoría de grafos. De esta forma se aborda un caso de estudio de desastres (erupción del volcán Sinabung en 2018) en Indonesia durante el 18 de febrero al 1 de marzo de 2018. Se identificaron usuarios relevantes de Twitter durante la ocurrencia del desastre a través de un análisis de red. Dicho análisis ofrece una vista general de las interacciones e impacto de los usuarios más influyentes durante el periodo de estudio, esto permitió observar que el flujo de información estaba controlado por diversos tipos de usuarios

---

* Trabajo de grado

** Facultad de Ingenierías físico-mecánicas. Escuela de estudios industriales y empresariales. Director: Daniel Orlando Martínez Quezada

Abstract

TITLE: APPLICATION OF GRAPH THEORY AS A TOOL FOR THE ANALYSIS OF SOCIAL BEHAVIOR IN THE FACE OF A DISASTER IN THE SOCIAL NETWORK TWITTER

AUTHORS: ORTIZ SIERRA Robinson
MARTÍNEZ CANO Juan Guillermo**

KEYWORDS: Social Networks, Twitter, Disasters, Social Network Analysis and Graph Theory.

DESCRIPTION:

Twitter has become an important tool to know in real time what happens in the society, which we live. In fact, this platform is increasingly attractive as a means of social communication, which allows it to play an important role in disaster situations. This interest in the platform can be used to help solve problems secondary to the event (eg, supply shortages) or give information to help immediate assistance organizations (answering questions such as: what do they need? where do they need it? how do they need it?) as well as information for those affected, as the position of shelters and evacuation routes.

In this research, data mining methods are applied on Twitter for the purpose of analyzing messages related to a seismic disaster/earthquake. We're describing an automated process for the analysis of the information collected using a graph theory approach. In this way, a case study of disasters (eruption of Sinabung volcano in 2018) that happened at Indonesia from February 18th to March 1st, 2018. Relevant users of Twitter were identified during the occurrence of the disaster through a network analysis. This analysis offers an overview of the interactions and impact of the most influential users during the study period, which allowed observing that the flow of information was controlled by different types of users.

---

* Bachelor Thesis

** Faculty of Engineering Phisics – Mechanical. School of Industrial and Business Studies. Director: Daniel Orlando Martínez Quezada

Introducción

A través de la historia en el mundo se ha presenciado un incremento marcado en sucesos de desastres naturales. Estos, tienen lugar en cualquier parte del mundo, pero son los países en vía de desarrollo los que han sufrido un impacto mayor debido a la capacidad para manejar estas situaciones por su baja calidad de las infraestructuras y servicios de emergencia, así como en las redes de comunicación.

En la década de los 90, las nuevas tecnologías han cambiado el modo de comunicarnos en diferentes aspectos. Las primeras redes sociales surgieron en esta década con el objetivo de facilitar la interacción de las personas entre sí, donde los usuarios comparten con otros usuarios todas las actividades que realizan en diferentes contenidos, estando en contacto constante y con actualizaciones en tiempo real, intercambiando información y dando lugares a debates y comentarios sobre el contenido en particular.

Entre estas redes sociales se encuentra Twitter, creada en marzo de 2006 se puede definir como una red social establecida para la participación y comunicación de los usuarios. Una principal característica es la limitación de sus mensajes a 280 caracteres, su facilidad de utilización y su rapidez en la comunicación de la información son dos de sus pilares básicos para su funcionamiento.

A principios de 2017, Twitter poseía 317 millones de usuarios mundiales únicos al mes, la mayoría de los usuarios está en un rango de edad entre 18 y 29 años, está disponible en más de 40 idiomas lo que facilita su difusión Toledano (2017). Twitter se postula, así como una de las

redes sociales más grandes en la actualidad debido a que representa con la gran cantidad de mensajes los cambios en la manera de comunicarnos, pero también los debates sociales o las noticias de actualidad.

De hecho, uno de los mayores atractivos de Twitter es su rapidez siendo un gran sitio donde se encuentran noticias de actualidad permitiendo obtener información en tiempo real de una manera fácil, donde los usuarios son capaces de recibir información y comentar los acontecimientos de última hora transmitidos en un periodo de tiempo muy pequeño. El uso de la red social Twitter se expande en otros ámbitos convirtiéndose en una potente herramienta complementaria a la hora de salvar vidas o alerta de desastres naturales o emergencias.

El presente trabajo trata de identificar y analizar el comportamiento humano utilizando teoría de grafos, para esto se pretende utilizar como herramienta dicha teoría que permite hacer un análisis descriptivo y análisis de red sobre una consolidación de información, examinando el comportamiento social de las personas durante un evento de desastre.

La teoría de grafos es una herramienta importante que analiza y ayuda a resolver una gran cantidad de datos considerando la forma de cómo los vértices se conectan por sus diferentes aristas; dichos elementos no solo se emplean para un solo tipo de análisis, la teoría de grafos es usada a menudo para diseñar conexiones de circuitos, en el estudio de algoritmos, que son las reglas a seguir para la resolución de problemas, también es utilizada para mostrar las redes de vuelo entre las ciudades e incluso para mostrar las estructuras químicas y moleculares de una sustancia. De esta forma la teoría de grafos permite analizar y procesar grandes volúmenes de datos que se encuentra en las redes sociales, con twitter como ejemplo un solo usuario puede tener miles de hashtags, urls, retweets... etc.; analizando todos los posibles usuarios que hacen

parte de una temática de interés, la gran cantidad de datos encontrados dificultará la comprensión de las tendencias y patrones.

En este documento se explicará el procedimiento para analizar e identificar el comportamiento analizando todos los posibles usuarios que hacen parte de una temática de interés, la gran cantidad de datos encontrados dificultará la comprensión de las tendencias y patrones, de igual manera los grafos simplifican y muestran de forma clara y sencilla las diferentes conexiones incluso de diferentes temas examinando el comportamiento social de las personas durante un desastre.

El presente documento está organizado de la siguiente manera: numeral 4 presenta una revisión de literatura relevante sobre de teoría de grafos y análisis de redes sociales y el uso en otras áreas; luego en el numeral 5 se presenta el Marco Teórico; en el numeral 6 se propone una metodología de investigación, junto con algunos estudios pertinentes. En los numerales 7, 8, 9, 10, 11 y 12 se aplica el marco de análisis de resultados y metadatos relacionados con el desastre, se presenta una gama de inteligencia descriptiva y técnicas de análisis de red en el contexto de desastres. Por último, en el numeral 13 y numeral 14 se muestran las conclusiones del presente proyecto y las recomendaciones para futuras investigaciones

Tabla 1.

Cumplimiento de los objetivos del proyecto

Objetivos	Cumplimiento
Realizar una revisión bibliográfica de la literatura sobre las aplicaciones de la teoría de grafos en análisis de redes sociales.	Numeral 4
Construir una base de datos sobre la red social ante un evento catastrófico.	Numeral 6
Realizar un análisis del comportamiento social mediante teoría de grafos que facilite la comprensión de la situación con una perspectiva más amplia.	Numeral 7 Numeral 8 Numeral 9 Numeral 10 Numeral 11 Numeral 12
Realizar un documento de carácter publicable donde se registren los resultados del trabajo de investigación.	Apéndice A Apéndice B

1. Planteamiento del problema

De acuerdo con la "International Federation of red Cross and red crescent societies" (PUBLICATIONS IFCR 2005002), se han identificado, probado y mejorado procesos para evaluar tanto el daño como las necesidades de respuesta ante situaciones repentinas que pueden alterar la estructura social; estar informado sobre dichos eventos es una forma vital de ayuda en sí misma, no solo para los organismos de atención primaria sino también para las personas afectadas quienes necesitan información de lo que está pasando así como agua, alimentos, medicamentos o refugio.

Estar debidamente informado de la situación puede ayudar a salvar vidas, en algunos casos puede ser la única forma de preparación ante los eventos que algunos pueden permitirse, aunque por lo general, los organismos de atención que se centran en la recopilación de los datos, lo hacen para sí mismos y no practican el intercambio de dicha información con las personas u otros entes que pueden apoyar la situación.

Ante un desastre natural diferentes organismos de atención primaria se tienen que desplegar para poder responder en las distintas etapas de este; la gestión internacional de desastres separa dichas etapas en cuatro componentes distintos: mitigación, preparación, respuesta y recuperación Coppola (2006). Al caracterizar el estudio se ve una gran importancia en la recuperación de datos en las etapas de respuesta y recuperación, dado que los organismos de atención a los desastres no pueden permitirse la espera en su toma de decisiones; muchas veces se actúa sin contar con la adecuada información, sin mencionar que en dichas etapas se encuentra también un límite a las 72 horas posteriores al desastre, debido a que es el tiempo necesario que tienen los cuerpos de socorro, búsqueda y rescate, policía y ejército para restablecer el orden y la normalidad. Sin embargo, en esos momentos la información para la toma de decisiones es casi nula, esto aumenta aún más la importancia sobre la recuperación de datos en dichas etapas, para la generación de información de flujo constante y que esté disponible lo más rápido posible David Sanderson (2016).

Hoy en día los medios de comunicación están dando más cobertura a diferencia del pasado y con las redes sociales a plena marcha esto no se puede negar. Sin embargo, como se mencionó anteriormente no todos los eventos provocan el mismo nivel de interés y respuesta, por lo que para poder abordar dichas situaciones en las redes sociales se deberá verificar su interés y palabras claves por las que se guían las conversaciones del mismo tema.

Los diferentes medios de comunicación en tiempo de crisis y desastres juegan un papel de vital importancia antes, durante y después los hechos. Los medios de comunicación, en particular se han convertido en un importante canal, desempeñando funciones complementarias a las que desempeñan los medios de comunicación tradicionales. Considerando esto importante, cabe mencionar que, en el servicio de las redes sociales, en el 2013 Twitter dio a conocer un nuevo servicio llamado Twitter Alertas diseñado para dar prioridad a la información de las organizaciones durante la crisis cuando los canales de comunicación no son accesibles.

Las redes sociales como Twitter pueden ayudar a que las comunidades afectadas por un desastre se transformen en comunidades intervinientes enviando solicitudes, información sobre la situación por la que se ven afectados, mensajes, evaluación de daños y peticiones de ayuda.

Existe una multitud de cuentas en la red social Twitter hablando sobre temas distintos cuando se presenta una emergencia, por lo cual, es difícil localizar cada opinión de forma aislada. En este punto el uso de Hashtag puede ayudar a unir lazos temáticos sobre conversaciones que parecen aisladas, pero siguen una misma línea temática. La disponibilidad de contenidos en la red social está condicionada por la participación de los diferentes usuarios, ellos son los que marcan las pautas en la calidad de los contenidos. En este punto podemos destacar la falta de calidad de algunos hashtags o publicaciones que no aportan nada a los usuarios.

Debido a la limitación de 280 caracteres en cada tuit, no es posible contar una historia concreta o ampliar un tema en un solo mensaje. Así pues, se deben utilizar varias trayectorias para ampliar un tema, lo cual puede hacer que se pierda el hilo de la conversación al usuario

que la sigue. Por el momento Twitter ha levantado su límite de caracteres para mensajes privados y para mensajes que incluyen retweets.

Aunque existen multitud de instituciones, entidades de rescate, entidades gubernamentales, no todas tienen cuenta en Twitte. Otro de los usos más importantes en el campo de los desastres reside en la utilización de la red Twitter como fuente de alerta temprana; su manejo sencillo, su llegada rápida e instantánea y su expansión a nivel mundial la convierte en una herramienta óptima para este uso.

De esta forma se pueden mencionar tres grandes puntos a tratar con la teoría de grafos aplicada a la plataforma de Twitter:

- Relevancia: No todos los eventos de desastre provocan el mismo interés, por lo que identificar y plantear un método que identifique los temas relevantes, así como los actores principales en un grafo nos permite un análisis sencillo y entendible de la situación para la toma adecuada de decisiones ante dicho evento.
- Velocidad: En muchos de estos casos el tiempo en que se responde es de gran importancia ya que son eventos en donde un minuto puede poner en riesgo vidas, medios y recursos por lo que se necesita información sobre lo que pasa segundo a segundo.
- Reconocimiento: De esta forma se puede ayudar en la logística, identificando los lugares más afectados a través de los diferentes datos suministrados, permitiendo a los agentes de atención inmediata la reacción oportuna para el despliegue de sus fuerzas o suministros de emergencia según se requiera.

2. Justificación del Proyecto

Con el análisis de datos en las redes se puede avanzar en la comprensión de fenómenos sociales, en las que se pueden presentar interacciones de individuos como organizaciones que forman estructuras que pueden ser observadas bajo una perspectiva diferente marcando las condiciones que permiten una mejor atención. Hoy en día, Twitter se ha convertido en una herramienta imprescindible para saber en tiempo real lo que sucede en la sociedad en la que convivimos. De hecho, esta plataforma es cada vez más interesante como medio de comunicación social que se utiliza para ayudar a mitigar los desastres como fuente de alerta temprana; su manejo sencillo, su llegada rápida e instantánea, permitiendo obtener información en tiempo real de una manera fácil, convirtiéndose así en una plataforma optima a nivel mundial para este uso. Esto se debe principalmente a las siguientes características de Twitter:

- Limitación de 280 caracteres como longitud de texto o tweet máximo, lo que convierte a esta plataforma en un generador de titulares.
- Capacidad de filtrar el contenido, es decir, el usuario elige a quién seguir, y éste decide si los contenidos que publica el otro usuario les resulta interesante o no.
- Portabilidad de la plataforma, ya que cualquier usuario puede publicar contenidos en cualquier lugar y momento mediante un teléfono inteligente.
- Acceso y publicación en tiempo real de la información, y es que cualquier suceso que ocurra en alguna parte del mundo puede ser transmitido en el momento por usuarios que tengan conciencia de dicha información. Por ejemplo, cuando ocurre un desastre, las personas hacen muchas publicaciones de Twitter (tweets) relacionadas con el desastre, lo

que permite detectar la ocurrencia de un terremoto rápidamente, simplemente observando los tweets.

Twitter en particular es visto como un medio de comunicación social crítico que se utiliza para la gestión de desastres. Las principales ventajas incluyen el crowdsourcing, la velocidad y la capacidad de acceder desde dispositivos móviles. Las desventajas potenciales incluyen el sesgo en la base de usuarios; información inexacta, falsa y desactualizada. Twitter ofrece una rápida aproximación a la recopilación de información de multitud de fuentes y una manera de llegar a muchas personas de la población Villegas y Álvarez (2016). Como tal, puede ser aprovechada para apoyar las actividades de alerta y respuesta para eventos extremos, tales como tsunamis.

La elección de Twitter como fuente de información es debido a su mayor facilidad para almacenar los datos, pero sobre todo la manera para acceder a ellos es más rápida. Por tanto, Twitter es una de las mayores fuentes de información actuales en Internet. Alimentada cada día con contenidos de todas partes del mundo y de muy distinta temática hace que se configure como un centro de información interesante para estudiar.

La teoría de grafos como medio de análisis no solo facilita la comprensión de la situación planteada, sino que también ayuda a construir grupos para estudiar su comportamiento y tendencias, por otro lado, permite también un componente social de investigación al contrastarlo con el comportamiento humano. Un ejemplo de esto se puede dar con los descubrimientos del psicólogo Kurt Lewin y su estudio del análisis de redes sociales, propuso en 1936 el "espacio vital" de un Individuo (Harary, 1969). Este punto de vista llevó a las personas a otra interpretación psicológica para el grafo, en la que las personas están

representadas por puntos y sus relaciones interpersonales por líneas, tales relaciones incluyen el amor, el odio, su comunicación, etc. (Reda Alhajj, 2014)

Por otro lado, las redes sociales, de tipo cibernético, son actualmente consideradas como nuevos modos de socialización y a partir de ellas puede obtenerse una fuente de interacción entre las personas que genere conocimiento en diferentes dominios de aplicación. (Ayala P., 2014). Actualmente podemos decir que el Análisis de Redes Sociales (ARS) es una metodología que busca mediante la aplicación de modelos extraídos de la teoría de grafos, predecir el comportamiento de una red social y aproximar las estrategias de los actores que la componen.

Los datos de las redes sociales se generan a gran velocidad y de manera continua y en forma de texto, audio, imágenes, números o hechos que son computables por un ordenador; siendo datos no estructurados. Un dato en particular es absolutamente inútil hasta que no se convierta en información útil; por lo tanto, es necesario analizar esta enorme cantidad de datos una vez sean estructurados para así poder hacer los respectivos análisis (Savita Kumari y Pratibha, 2017). De esta forma la minería de datos gana rápidamente la atención entre los investigadores debido a la posibilidad de tratar grandes fuentes de datos, que pueden ser incluso de tipo multimedia, lo que brinda una potencial rama de descubrimiento de información la cual puede impulsar diversos campos de investigación.

A medida que se genera un desastre los diferentes organismos de atención primaria se tienen que desplegar para poder responder en las distintas etapas de este; la gestión internacional de desastres separa dichas etapas en cuatro componentes distintos: mitigación, preparación, respuesta y recuperación Coppola (2006). Al caracterizar el estudio se ve una gran

importancia en la recuperación de datos en las etapas de respuesta y recuperación, dado que los organismos de atención a los desastres no pueden permitirse la espera en su toma de decisiones muchas veces se actúa sin contar con la adecuada información, en las posteriores horas al desastre el valor de la información para la toma de decisiones es casi nula, aumenta aún más la importancia sobre la recuperación de datos en dichas etapas (David Sanderson, 2016).

De esta forma el presente plantea utilizar la teoría de grafos y la consolidación de información de redes sociales, examinando el comportamiento social ante una situación de desastre y las respuestas sociales que se generan debido a este para encontrar tendencias o patrones de comportamiento que puedan ser de ayuda en la toma de decisiones de los organismos de atención.

3. Objetivos

3.1.Objetivo general

Aplicar modelos de grafos para estudiar el comportamiento de las personas ante un evento de desastre con base en datos proporcionados por redes sociales usando la teoría de grafos.

3.2. Objetivos específicos

- Realizar una revisión bibliográfica de la literatura sobre las aplicaciones de la teoría de grafos en análisis de redes sociales.
- Construir una base de datos sobre la red social ante un evento catastrófico.
- Realizar un análisis del comportamiento social mediante teoría de grafos que facilite la comprensión de la situación con una perspectiva más amplia.
- Realizar un documento de carácter publicable donde se registren los resultados del trabajo de investigación.

4. Revisión de la Literatura

El problema de los puentes de Königsberd, resuelto por Leonhard Euler, en 1736, fue el comienzo de la teoría de grafos, ya que es considerado el primer resultado de esta disciplina. Existieron otros pensadores que influyeron en el desarrollo, como Gustav Kirchhoff con las leyes de los circuitos para calcular el voltaje y la corriente en los circuitos eléctricos. Son estos ejemplos esenciales para el desarrollo de la teoría de grafos.

La teoría de grafos expone que la red se constituye por nodos conectados por aristas, donde los nodos son los individuos y las aristas, las relaciones que los unen, con las características de estos lazos donde puede ser usado para interpretar los comportamientos sociales de las personas aplicadas (Diestel, 2005).

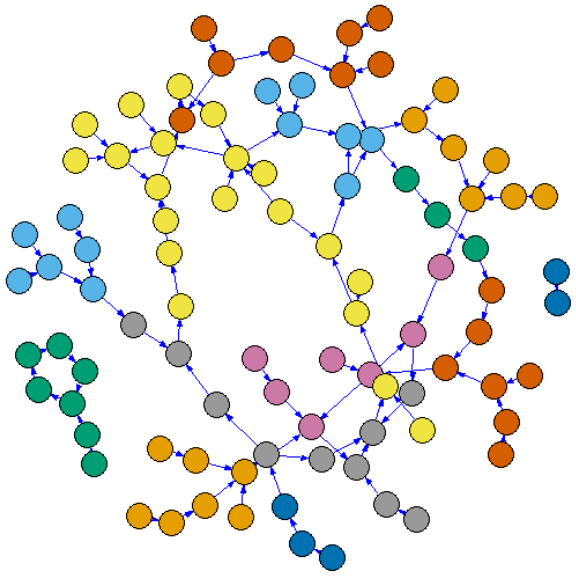


Figura 1. Diagrama de un grafo con vértices y aristas. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

Las posibilidades de la teoría de grafos es la descripción del comportamiento de una red social, a partir de sus aristas y nodos, esto tiene un gran valor para el análisis de las interacciones. Su aplicación en Twitter ha llevado (Smith, Rainie, Shneiderman, y Himelboim, 2014) a identificar algunos arquetipos de conversación con sus propias estructuras de comunidad:

- Multitudes cercanas: conversaciones entre usuarios con un alto grado de conexión y escasa presencia de participantes aislados, el usuario de cada red comparte información con frecuencia y mantiene conversaciones con otros usuarios.

- Multitudes polarizadas: son grupos grandes y densos en los que los usuarios comparten sus opiniones. Los participantes de un grupo generalmente no interactúan con personas de otros grupos donde cada miembro suele mencionar colecciones muy diferentes de URL de sitios web y usa etiquetas de palabras distintas en cada Tweet.

- Clúster de marca: redes de baja densidad con individuos pocos conectados entre sí. Cada usuario se agrupa por medio de temas interesantes que pueden atraer a grandes poblaciones fragmentadas de Twitter que twittean sobre el tema, pero no entre los usuarios.

- Clústeres comunitarios: redes estructuradas en pequeños subgrupos con sus propias audiencias, individuos influyentes y fuentes de información, donde las noticias globales atraen la cobertura de muchos medios de noticias, cada uno con sus seguidores.

- Redes de difusión: estructuras donde los usuarios siguen, difunden y comentan las noticias de última hora estableciendo apenas conversaciones con otros individuos. El centro de la red suele ser un medio de comunicación cuyo alcance es amplificado por la comunidad.

Según Aldrich en su trabajo de investigación denominado "*El poder de las personas: el papel del capital social en la recuperación del terremoto de Kobe en 1995"* examinaron literatura reciente para investigar el papel crítico del capital social y las redes en la recuperación de desastres. Demostraron que las agencias de desastres, los encargados de la toma de decisiones gubernamentales y las ONG necesitan fortalecer la infraestructura social a nivel de la comunidad para aumentar la resiliencia frente a los desastres. (Aldrich, 2011)

De acuerdo con Scott (2012) el análisis de redes sociales surgió por primera vez como un enfoque distintivo del análisis estadístico dentro de la antropología, la psicología social y la sociología. Luego influyó en los desarrollos teóricos y metodológicos en muchas otras áreas de las ciencias sociales. De esta forma, el análisis de las redes sociales se vinculó a temas más amplios promoviendo la investigación y explorando las implicaciones de otros modelos matemáticos, estadísticos y estocásticos.

La idea de una red de relaciones en un contexto sociológico se puede vislumbrar con el trabajo de Moreno JL (1934), el cual comenzó con estudios sobre la influencia en las elecciones de amistad en el desarrollo de la personalidad entre los escolares. Su innovación clave fue hacer uso de sociogramas, los cuales consisten en puntos, que representan individuos, y líneas, que representan las relaciones sociales entre ellos, para representar el patrón de amistad que pudo descubrir en grupos pequeños como en un salón de escuela o un grupo de vecinos; De esta

forma argumentó que estos puntos y líneas, llamados formalmente "vértices" y "aristas", podrían describirse como configuraciones sociales en las cuales se podrían registrar cosas tales como la dirección de las elecciones de amistad y su intensidad.

Citando a Ávila-Toscano, "el método de evaluación de las redes se denomina análisis de redes sociales (ARS) y en general es considerado como el estudio de la estructura social, y en un sentido más amplio se puede entender como un método cuantitativo por medio del cual se obtiene la estructura social a partir de las regularidades en el patrón de relaciones establecidas entre entidades sociales definidas como personas, grupos u organizaciones" (Toscano, 2012).

El análisis de Redes Sociales fue diseñado para descubrir las relaciones establecidas entre las entidades sociales, incluyendo cálculo de métricas que proporcionan una descripción local (*nivel de actor)* y global de la red *(nivel de red),* visualización gráfica de las redes y detección de comunidad para comprender la estructura de redes complejas y encontrando información útil del mismo. El análisis de las redes sociales evalúa las oportunidades de información para individuos o grupos de personas en términos de exposición y control de la información, teniendo conocimiento de las rutas existentes de intercambio de información (Márcia Oliveira, 2012).

Lu y Brelsford (2014) Construyeron un marco sistemático para el análisis de comunidades en línea en respuesta a desastres naturales, investigaron y compararon la estructura de las redes interactivas, la evolución de las comunidades en línea y el contenido de la comunicación antes y después de un evento adverso, mostrando cambios distintivos en los patrones de interacciones en las comunidades que han sido afectadas por un desastre en comparación con las comunidades que no fueron afectadas. Al aplicar el algoritmo de detección de la comunidad

del mapeo de información, encontraron que el comportamiento de unirse o abandonar una comunidad está lejos de ser aleatorio: los usuarios tienden a permanecer en un estado social actual y es menos probable que se unan a nuevas comunidades de su comunidad original.

Yan Jin (2014) Estudió canales de información de desastres (redes sociales vs medios tradicionales) y fuentes (agencias nacionales y medios vs agencias locales y medios) en términos de su capacidad para generar resultados públicos deseados (intenciones de buscar y compartir informaciones de emergencia). Descubrieron que los usuarios de las redes sociales tenían más probabilidades de buscar más información en Twitter cuando la información inicial sobre el desastre era en forma de tweet que de página web.

Twitter ha recibido una gran cantidad de atención de investigación en relación con otros medios utilizados en todas las fases de la gestión de desastres. Este medio está siendo aprovechado por las personas durante los desastres y organizaciones de socorro como agencias gubernamentales, policías, médicos y organizaciones de salud pública. Estos grupos generalmente tienen acceso a datos y herramientas analíticas que no están disponibles al público. Las personas en realidad ya pueden estar conectadas por medio de un teléfono inteligente, a la vez pueden funcionar en cualquier cantidad de plataformas de redes sociales. Las tecnologías para analizar las redes sociales no se limitan a una única plataforma, sino que aprovecha la mayor cantidad de datos posibles. (Landwehr P. M., 2014)

Cheng y Wicks (2014) en su trabajo plantearon una nueva metodología para la identificación de los desastres utilizando datos de la red social Twitter. Esto lo lograron utilizando el método de las estadísticas de análisis de espacio-tiempo (STSS), que se puede utilizar en diferentes tipos de desastres. Sin embargo, aún se requieren investigaciones

adicionales para mejorar esta técnica. En primer lugar, se debe analizar diferentes eventos de desastres para garantizar que el método de las estadísticas de análisis de espacio tiempo se pueda aplicar a diversos tipos de desastres. En segundo lugar, se debe investigar para explorar la posibilidad de aplicar este método a la vigilancia en tiempo real de los clústeres espaciales emergentes. Si el método es prospectivo para identificar eventos emergentes en Twitter, entonces el futuro de la detención y respuesta de desastres podrá volverse más eficiente, más dinámico y poderoso.

Los investigadores Bruno, Edson y Christine centraron su investigación en las formas en que los individuos afectados, por ejemplo, personas, periodistas, autoridades, etc. y organizaciones, por ejemplo, el gobierno, los medios de comunicación, ONG, etc. utilizan los medios sociales durante un desastre natural, ellos también examinaron un conjunto de factores que pueden explicar esos usos, tales como el tiempo de los tweets, la ubicación y las características de los usuarios. (Takahashi, Tandoc, y Carmichael, 2015)

Caragea , Silvescu y Tapia (2016) Presentaron un enfoque basado en las redes neuronales convolucionales para identificar mensajes informativos en las redes sociales durante los desastres. La contribución fue la mejora en la precisión de la identificación de los tweets informativos durante eventos de desastres, utilizando Redes Neuronales Convolucionales (CNN) que pueden predecir los Tweets informativos y filtrar los tweets que no son de naturaleza informativa. Este es un fuerte paso en el camino para proporcionar a los usuarios información verdaderamente procesable en tiempo real basada en datos de redes sociales. El uso de técnicas de adaptación de dominio junto con las redes neuronales convolucionales sería una dirección futura interesante por seguir.

Peter M. Landwehr (2016) Dan a conocer en su trabajo un sistema de Tsunamis y respuestas de redes sociales, TWRsms, un componente aplicativo web basado en Twitter para un sistema socio-técnico de apoyo a la toma de decisiones para eventos catastróficos. Los TWRsms recopilan y analizan tweets en tiempo real cuando está monitoreando la ciudad de Pandang, apoyando la alerta y la planificación. El uso de esta actividad proporciona información de los tweets generados en dicha área, lo cual es valioso cuando se planifica la evaluación y la respuesta. Estos datos en tiempo real, proporcionaron orientación sobre quiénes son los más influyentes en las publicación de temas relevantes con el evento, por lo cual es posible dirigir alertas a estos usuarios dentro de la comunidad dándole a conocer las alertas tempranas, finalmente cuando este sistema rastrea lo que los usuarios twittean, proporciona información de la hora, quien lo publica y en general en contenido del tweet, también proporciona información sobre los temas claves que se evidencian en los hashtags. Por lo tanto, esta información se puede usar para saber en qué región es poco probable que reciban información de Twitter. Los autores realizaron estudios por separado considerando varias estrategias de recolección de datos de Twitter para Indonesia en su conjunto y para Padang en particular. La primera estrategia que desarrollaron fue la utilización de un cuadro delimitador para recuperar primero los tweets y luego eliminar los tweets que no eran de interés utilizando palabras claves. La segunda estrategia fue la recuperación de palabras claves en conjunto con un cuadro delimitador, los autores plantearon su propio conjunto de desafíos utilizándolos para eliminar gran cantidad de tweets durante un periodo prolongado utilizando parámetros como palabras claves, cuadros delimitadores de ubicación y el ID del usuario ver figura 2.

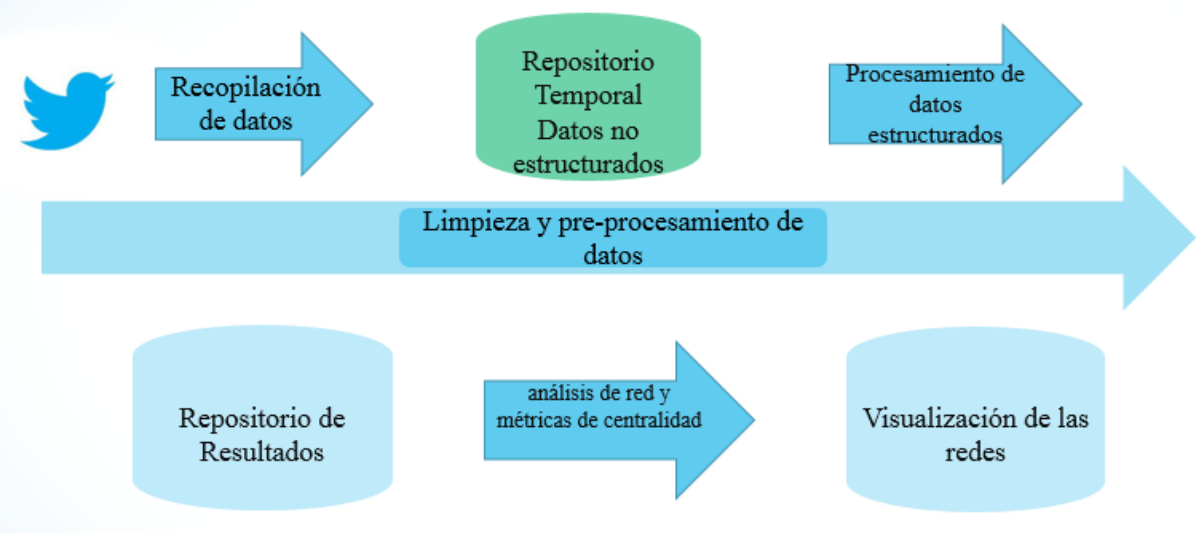


Figura 2. Flujo de recogida, análisis y presentación de los Tweets. Nota: Adaptado de Kathleen M. Carley, M. M. (13 de abril de 2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. safety science.

El uso de Twitter varía de acuerdo con el usuario: organizaciones periodísticas, fuentes gubernamentales, y ciudadanos. Durante el Tifón Haiyan en Filipinas las organizaciones periodísticas utilizaron Twitter para difundir informaciones de segunda mano, mientras que los ciudadanos hacían uso de la red social para rememorar a las víctimas. Entre tanto, el gobierno y las entidades no gubernamentales comunicaban esfuerzos de rescate, información sobre la situación publicando la cobertura informativa sobre el evento, y también ayudando a coordinar esfuerzos de rescate (Bruno Takahashi, 2016).

Jooho, Juhee y Makarand (2018) en su trabajo denominado "*Difusión de información de emergencia en las redes sociales en línea durante la tormenta Cindy en los EE.UU*" exploraron las conexiones y los patrones de la red social creada por las interacciones agregadas en Twitter durante las respuestas a desastres, analizaron las funciones críticas de cuatro tipos de usuarios en Twitter (agenda de noticias y meteorológicas) denominados como fuentes de información y difusores de información (el público y las organizaciones) en línea mediante el análisis de redes

sociales y el análisis de texto, usando múltiples métodos. Los resultados generados por los métodos proporcionan información sobre las agencias y organizaciones de emergencia con el fin de comprender características de las redes sociales en línea y su participación

También se han desarrollado múltiples software y herramientas para satisfacer la creciente necesidad de la tecnología de visualización y minería de datos de redes sociales como R y la biblioteca SNA, JUNG, NodeXL, Gelphi (Qiuju Luo, 2015).

La idea es utilizar la herramienta de la teoría de grafos para visualizar y modelar las relaciones sociales, de esta forma se podría ver las relaciones y el comportamiento de los individuos que comparten diferentes términos informales que surgen en la interacción de los medios virtuales.

En general, hay una gran cantidad de publicaciones que describen diferentes aspectos de los medios de comunicación social como una herramienta de ayuda para la gestión de desastres en diferentes escalas. Algunos investigadores aplicaron técnicas de análisis de datos y minería de datos de diferentes tsunamis y desastres naturales.

5. Marco Teórico

5.1. Big data

El término “Big data” se encuentra que está involucrado en muchas y diversas industrias, creando tendencias en otros campos, como la educación, el gobierno, la salud o la banca; su utilización generalizada en muchos ámbitos tiende a ocasionar confusiones, aunque por lo

general el término hace hincapié en un avance de la tecnología para el procesamiento de conjuntos de datos. De acuerdo con (James Manyika, 2011), Big Data hace referencia a conjuntos de datos cuyo tamaño está más allá de la capacidad de capturar, almacenar, administrar y analizar con las herramientas típicas de software de bases de datos.

Como tal una definición ampliamente extendida y generalizada del término aún no ha tenido lugar, aunque se encuentra un desarrollo del concepto con De Mauro, Greco, y Grimaldi, (2016) quienes lo describe como "el activo de información caracterizado por tener un volumen, velocidad y variedad tan elevados como para requerir una tecnología específica y métodos analíticos para su transformación en valor" dicha definición permite un entendimiento más coherente ya que se basa en los temas esenciales de las definiciones más populares actualmente utilizadas, mostrando características propias del Big data: volumen, velocidad y variedad. Sin embargo, otros atributos como la veracidad y el valor han sido añadidos por otros autores, por lo que la definición del término aún sigue formándose. Por otro lado, la tecnología y métodos necesarios para el análisis de la información están vinculados con el término debido a que son necesarios para el tratamiento de una gran cantidad de datos “Big data”.

5.2. Volumen

Cuando se habla del volumen que se maneja en el Big Data, se refiere a la cantidad de datos que se procesan en la transformación a información utilizable. Sin embargo, no se encontraron referentes confiables sobre una cantidad o rangos específicos que se utilicen para considerar un volumen propio del Big Data; a un es una característica apreciada ya que el entorno actual presenta un desarrollo continuo en áreas como la tecnología, que como resultado hace que se generen más y más datos debido a las muchas y distintas herramientas con mejores capacidades, tanto en velocidad como en capacidad de almacenamiento, de acuerdo con (IBM

, s.f.) corporación multinacional de tecnología informática y consultoría IBM “se convierten hasta 12 terabytes de Tweets que son creados cada día en un análisis de sentimiento de producto mejorado por la misma empresa” y “se convierten 350 mil millones de lecturas de medidores anuales para predecir mejor el consumo de energía”; por lo que la gran cantidad de datos que se generan segundo a segundo pueden llegar a manejarse de distinta forma dependiendo de los medios y la situación en la que es requerido el análisis.

5.3. Velocidad

La velocidad es la rapidez en la que los datos son procesados y analizados. Esta característica se vuelve cada día más importante debido a que muchas tomas de decisiones deben ser llevadas a cabo al mismo tiempo que se generan los datos, por lo que deben ser procesados en tiempo real. Un ejemplo de esto es dado por la corporación multinacional de tecnología informática y consultoría IBM quienes exponen que “algunas veces incluso 2 minutos pueden ser demasiado tiempo, debido a que existen procesos sensibles al tiempo, como la captura de fraudes cibernéticos en transacciones monetarias”, por lo que dependiendo del proceso, se buscará que la velocidad de respuesta sea lo más rápido posible para obtener la información precisa en el momento preciso dado que esta sería una ventaja competitiva frente a otras organizaciones.

5.4. Variedad

Cuando se habla de variedad se refiere a la gran diversidad de datos y sus distintos formatos, estos se pueden encontrar en los diferentes medios de información en donde son presentados por lo general como videos, audios, textos e imágenes. Sin embargo, cabe mencionar que los datos son categorizados como datos estructurados y no estructurados: los estructurados son aquellos archivos que son hallados con una mejor visualización de los datos, en filas y

columnas y, los no estructurados son aquellos datos en bruto que no están organizados. Teniendo esto en cuenta hay que considerar las diferentes formas en que los datos son encontrados, ya que se encuentran datos estructurados y no estructurados en los diferentes medios compartiendo una misma información. Un ejemplo de esto son los artículos y/o publicaciones en los que por lo general hay datos en tablas y que por lo tanto pueden ser procesables fácilmente a información, por otro lado, también se encuentran datos no estructurados principalmente en forma de texto, ya que el contenido del mensaje de dicho texto no se categoriza en un dato hasta que no es contextualizado por completo.

5.5. Veracidad

Las diferentes definiciones de los autores sobre Big data no solo abarcan el volumen, velocidad y variedad, Michael Schroeck (2012), la describen "como una combinación de volumen, variedad, velocidad y veracidad que crea una oportunidad para que las organizaciones obtengan una ventaja competitiva en el mercado digitalizado". Por lo que cuando hablamos de veracidad hablamos de la procedencia de los datos y en qué tan confiable son estos, debido a que se sacará información para la toma de decisiones que en caso de ser falsa o errada acarrearía pérdidas de todo tipo dependiendo la situación planteada.

5.6. Viabilidad, valor de los datos

La variedad se refiere a la capacidad que se tiene para generar un uso eficaz de los datos disponibles, de igual forma el valor de los datos guarda cierta relación con la viabilidad ya que dicho valor se obtiene de la transformación de los datos en información utilizable para la toma de decisiones; Sin embargo estas "V's." no son consideradas dado que como expone Grimes (2013) "La viabilidad y valor de los datos no son una propiedad de Big data, dado que son un aspecto de calidad que se determina a través del análisis de Big data.", dicho valor y viabilidad

sólo pueden ser apreciados luego de que se hallan hecho los respectivos análisis de los datos y se hallan convertido en información.

5.7. Minería de datos

Según Wang y Wang (2015) “La minería de datos es un proceso en el que se descubren o encuentran patrones o modelos de comportamiento a partir de una gran cantidad de datos”, de esta forma es utilizada para encontrar correlaciones en diferentes tipos de datos o bases de datos.

El *análisis de redes sociales* (ARS) es una herramienta que permite conocer las interacciones entre cualquier clase de personas. Es un conjunto de técnicas de análisis para el estudio formal de las relaciones entre actores y para analizar las estructuras sociales que surgen en la ocurrencia de esas relaciones o de la ocurrencia de determinados eventos. Es decir, el ARS permite el estudio de cómo la estructura de relaciones sociales alrededor de una persona, grupo u organizaciones afecta a su conducta y actitudes.

Debido a que el análisis de redes sociales requiere información de tipo cualitativa gracias a su propia naturaleza, se hace necesario seguir una serie de técnicas que nos permiten ordenar las interacciones de los individuos o grupos de personas de tal modo que dichas interacciones puedan ser representadas en una red o grafo. Así, las redes o grafos son herramientas principales para representar las interacciones entre actores o grupos de forma ilustrativa. El desarrollo del análisis de redes sociales se puede llevar a cabo gracias a una parte de las matemáticas denominada *Teoría de Grafos*

5.8. Teoría de grafos

La teoría de grafos es una disciplina que hoy en día presenta un desarrollo vertiginoso debido a su importancia en el diseño de diferentes programas. Estas ideas básicas las introdujo el matemático suizo Leonhard Euler en el siglo XVIII.

Euler en su gran mayoría de vida realizó aportes a la matemática, siendo uno de los más importantes en la teoría de grafos en 1736 con el trabajo de los puentes de Konigsberg que fue considerado el primer resultado de la teoría de grafos. Euler se dedicó por completo al estudio de los puentes, dando una solución simple e ingeniosa que servía también para un determinado número de puentes. En primer lugar, Euler reemplazó el mapa de la ciudad por un simple diagrama de puntos que construyó la gráfica de lo que posteriormente se conocía como un grafo, razón por lo que muchos autores consideran a Euler como el padre de la teoría de grafos Núñez *et al.*,(2004).

Existieron otros pensadores que influyeron en el desarrollo de la teoría de grafos, como Gustav Kirchhoff con las leyes de los circuitos para calcular el voltaje y la corriente en los circuitos eléctricos y Francis Guthrie que en 1852 demostró matemáticamente que con cuatro colores era suficiente para colorear cualquier mapa político, con la condición de que dos países adyacentes no pueden tener el mismo color.

El problema del cartero chino, conocido como el problema del circuito del cartero, el problema de los correos o problema de la inspección y selección de rutas, es el primer problema de rutas por arcos en el que se plantea la posibilidad de construir un ciclo euleriano con coste óptimo. Fue planteado originalmente por el matemático chino kwan Mei-Ko en 1960. Mei

planteaba el problema al que se enfrenta el cartero para repartir las cartas recorriendo la menor distancia posible.

El concepto de grafo y las nociones relacionadas son una parte central del análisis de las redes sociales, ya que la teoría de grafos proporciona un lenguaje formalizado apto para la descripción de las redes y sus características. Básicamente, un grafo es un conjunto de puntos interconectados por un conjunto de líneas. En teoría de grafos, estos elementos reciben la denominación de nodos y aristas respectivamente. Cuando un grafo representa una red social, los nodos representan a diferentes actores sociales, y las aristas representan la conexión entre los puntos del grafo (figura 3).

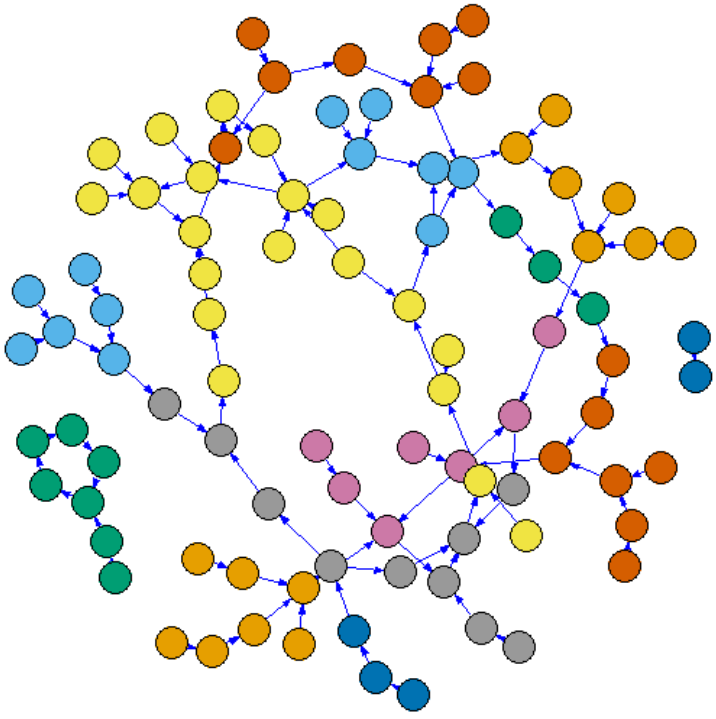


Figura 3. Diagrama de un grafo. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

5.8.1. Elementos (aristas, vértices). La teoría de grafos resulta una técnica apropiada para el estudio de todo tipo de redes, pues una gráfica dirigida representa un conjunto de nodos finitos denominados vértices, *E1, E2, …, En,* al lado de un conjunto de aristas dirigidas que unen un par ordenado de vértices distintos Kolman y Hill (2013). En las redes sociales esta correspondencia se logra haciendo que los usuarios simbolicen los vértices y las relaciones que se establecen entre ellos sustituyan los arcos dirigidos (aristas) ver figura 4.

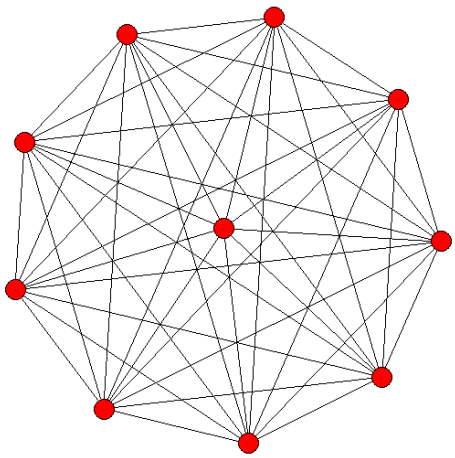


Figura 4. Elementos de un grafo (vértices y aristas). Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

5.8.2. Subgrafo. Un subgrafo de un grafo *I* es un grafo cuyo conjunto de vértices y aristas son subconjuntos de los de *I.* Se dice que un grafo *I* contiene a otro grafo *F* si algún subgrafo de *I* es *I2* o es isomorfo a *I.* En la figura 5 el subgrafo inducido de *I2* es un subgrafo *I* de *I*2 tal que contiene todas las aristas adyacentes al subconjunto de vértices de *I.*

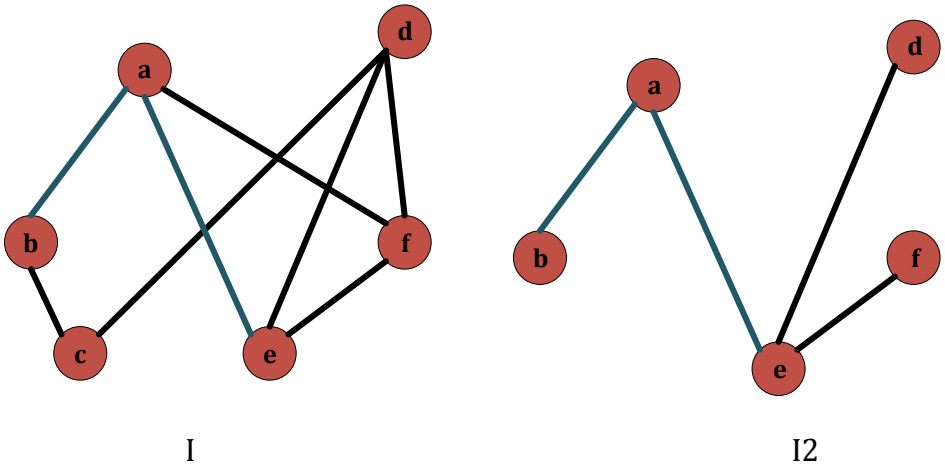
Definición:

Sea I= (V, A). I'= (V', A') se dice subgrafo de I si:

- $V' \subseteq V$
- $A' \subseteq A$

- *(V', A')* es un grafo

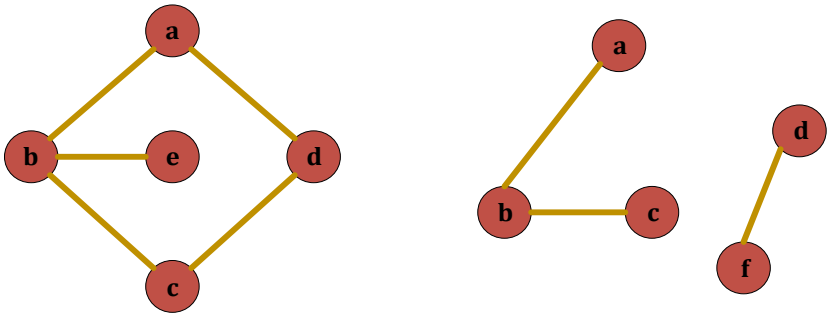
Si I' = (V', A') es subgrafo de I, para todo v ∈ G



I I2

Figura 5. Ejemplo de grafo I2 es un subgrafo de I. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

Por otro lado cuando hablamos de un grafo conexo nos referimos a un grafo que no está dividido, Caicedo Barrero , Wagner de garcía, y Méndez Parra (2010) aclara que para decir que un grafo es conexo para cada par de vértices "x" y "y" existe una trayectoria desde "x" hasta "y" la cual usa un máximo de "n-1" arcos"; en otras palabras un grafo NO conexo es aquel que posee dos o más piezas que son subgrupos del grafo original y por lo general son llamados componentes.



Grafo conexo Grafo no conexo

Figura 6. Grafo Conexo y No Conexo. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

Los grafos desde sus inicios se han empleado para resolver problemas de diversas áreas. Dicha teoría se puede utilizar en el área del análisis de redes sociales, ya que a través de los grafos se puede estudiar las complejas estructuras que se generan al analizar la gran cantidad de datos que se encuentran en las redes virtuales.

Estructuras de datos en la representación de grafos. Existen diferentes formas de representar un grafo (simple), además de la geometría y muchos métodos para almacenarlos en una computadora. La estructura de datos usada depende de las características de un grafo.

Entre las estructuras más sencillas y usadas se encuentran las listas y las matrices, aunque frecuentemente es una combinación de ambas, las listas son preferidas en grafos dispersos porque tienen un eficiente uso de la memoria. Por otro lado, las matrices proveen acceso rápido, pero consumen grandes cantidades de memoria.

Estructuras de lista.

- Lista de incidencia: las aristas son representadas con un vector de pares (ordenados si el grafo es dirigido), donde cada par representa una de las aristas.
- Lista de adyacencia: cada vértice tiene una lista de vértices los cuales son adyacentes a él Grafo

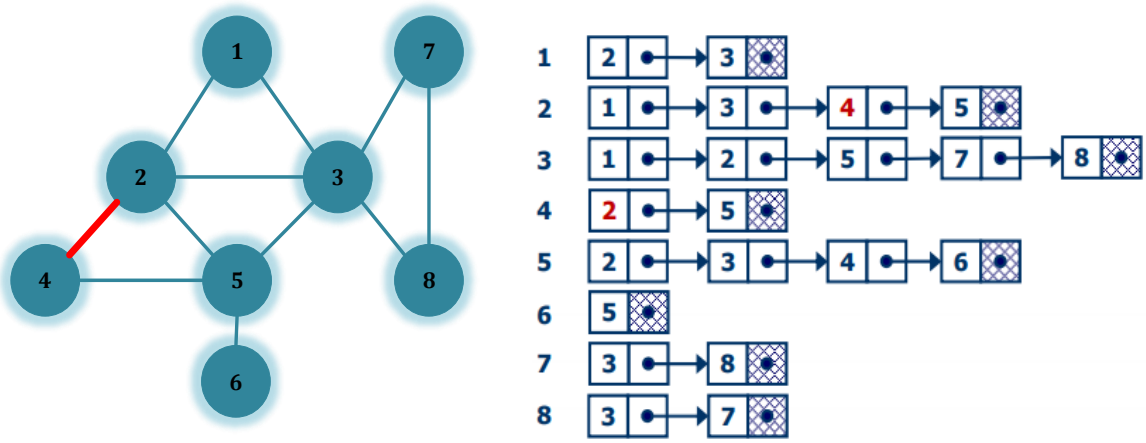
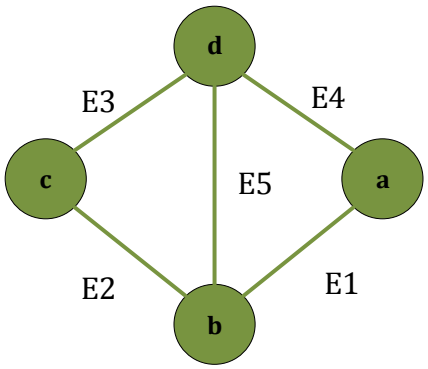


Figura 7. Representación mediante listas de adyacencia. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

Estructuras matriciales

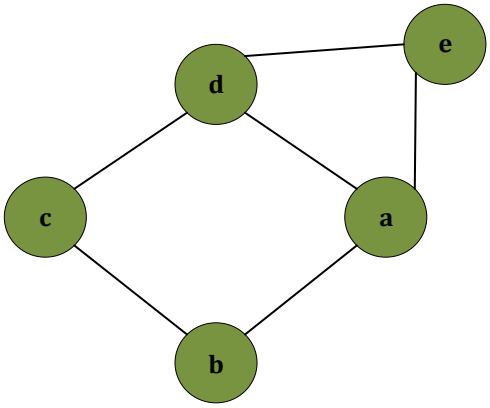
- Matriz de incidencia: el grafo está representado por una matriz A (aristas) por V (vértices), donde (arista, vértice) contienen la información de la arista (1-conectado, 0-no conectado).



	E1	E2	E3	E4	E5
a	1	0	0	1	0
b	1	1	0	0	1
c	0	1	1	0	0
d	0	0	1	1	1

Figura 8. Matriz de incidencia de un grafo. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

- Matriz de adyacencia: el grafo está representado por una matriz cuadrada M de tamaño n^2 donde n es el número de vértices.



	a	b	c	d	e	Grado
a	0	1	0	1	1	2
b	1	0	1	0	0	2
c	0	1	0	1	0	2
d	1	0	1	0	1	3
e	1	0	0	1	0	2

Figura 9. Matriz de adyacencia de un grafo. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

5.8.3. Centralidad. De acuerdo con Linton (1978), la centralidad puede calcularse de acuerdo con diferentes medidas, que dan lugar a diferentes conceptos de centralidad. La forma más simple e intuitiva de medir la centralidad es a través del grado ("*degree*") de los vértices del grafo. Un punto es central si está bien conectado con los demás puntos de su entorno, la centralidad es una medida de la importancia relativa de un nodo dentro del grafo. Esta importancia no es una propiedad intrínseca del nodo, si no que viene determinada por la posición de éste en el grafo.

Sin embargo, la centralidad no tiene que ver sólo con la identificación de los puntos más centrales en el grafo de una red, sino también con la de los puntos periféricos, que igual los centrales pueden caracterizarse como puntos localmente periféricos y globalmente periféricos.

Existen múltiples métricas que permiten determinar la centralidad de un nodo. Estas métricas se clasifican en dos grupos: métricas radiales y mediales. En las métricas radiales el nodo en cuestión es origen o destino de caminos generados con un cierto criterio, y la métrica de centralidad determina la cantidad de caminos existentes (métrica de volumen) o la longitud

de estos (métrica de longitud). Sin embargo, en las métricas mediales se emplea para determinar la centralidad de los nodos la cantidad de caminos que atraviesan dicho nodo(Vega Bayo).

Algunas de las métricas de centralidad más importantes son la centralidad de grado, de cercanía, de proximidad y de vector propio.

5.8.3.1. ***Métricas de centralidad*** En teoría de grafos existen diversas métricas de centralidad que buscan explicar o representar la importancia relativa que tiene un nodo dentro de un grafo, esta importancia está en función de diversos factores como: la posición dentro del grafo, los nodos que tiene a su alrededor, el número de veces que se ubica entre caminos geodésicos, entre otros. Desde la formulación realizada por Bavelas, se ha propuesto varias medidas de centralidad de un nodo. Existen cuatro medidas que son ampliamente usadas en el análisis de redes; la centralidad de grado y centralidad de vector propio que son medidas de volumen, la cercanía que es una medida radial de longitud y la intermediación que es una medida medial.

Centralidad de grado. Este parámetro considera la centralidad de cada nodo asociado a su grado. Es decir, es una medida radial de volumen que se calcula como el número de aristas incidentes en el vértice. Si bien el análisis de centralidad se centra principalmente en la conectividad de los nodos individuales. Es decir, un mayor grado indica mayor centralidad en el grafo y cuantifica aspectos como la influencia directa del nodo y su acceso a información de primera mano. Va de 0 a $n-1$ en un gráfico simple y, por lo tanto, puede ser normalizado por $n-1$

$$\boldsymbol{C'D(x) = \frac{deg(x)}{n-1}}$$

Lo que intenta explicar esta medida es el potencial de comunicación que tiene un nodo dentro de un grafo (ver figura 10) nodo J, por ejemplo, qué tan importante es una persona dentro de un grupo de amistad al momento de hacer circular una información, o en el caso de un grafo de *retuits* el grado de entrada de un nodo representa el número de usuarios que le han *retuiteado* y el grado de salida representa el número de usuarios a los que ha *retuiteado*. En redes con enlaces dirigidos, el grado se divide en dos: Grado de Entrada *(In-degree):* cuantifica aquellos enlaces con dirección hacia un nodo específico, pero procedentes de otros de la red. Grado de salida *(out-degree):* por el contrario, el nodo constituye el origen de enlaces hacia el resto de la red. En consecuencia, cuando más enlaces se establecen con un nodo más importante es el mismo porque la circulación en la red lo convierte en un centro de recolección y distribución de información, el nodo puede cumplir dos funciones: la receptora y la emisora. Ambas características pueden usarse como medidas de centralidad.

Centralidad de intermediación (betweeness centrality). La centralidad de intermediación es una medida de centralidad medial que viene dada por la fracción de caminos más cortos entre nodos del grafo que atraviesan el nodo en cuestión. Por lo tanto, esta medida sirve para indicar la centralidad de un nodo basada en el flujo entre los demás nodos del grafo enfocándose en el control de la comunicación, y se interpreta como la posibilidad que tiene un nodo para intermediar las comunicaciones entre pares de nodos. Un nodo con un valor alto indica que este nodo forma parte de muchos caminos cortos del grafo, con lo cual será un nodo clave en la estructura del grafo ver figura 10 nodo A. Su eliminación puede repercutir de forma muy directa en la conectividad del grafo.

Bavelas (1948) Dice: "*Una persona estratégicamente localizada en un camino de comunicación entre dos puntos, es una persona central, una persona en tal posición puede influenciar el grupo manteniendo o distorsionando la transmisión de la comunicación*". Este potencial de control de información que posee un punto que está dentro del geodésico se intenta modelar con una medida de centralidad que se calcula de diversas formas que fueron desarrolladas por autores como Anthonisse en 1971 y Feeman en 1977.

Formalmente, la centralidad de intermediación $\boldsymbol{C_{BET}(i)}$ de un nodo $\boldsymbol{i}$ en una red se define como:

$$\boldsymbol{C_{BET}(i) = \sum_{j,k} \frac{bjik}{bjk}}$$

Donde bjk es el número de caminos más cortos desde el nodo j hasta el nodo k y $bijk$ el número de caminos más cortos desde j hasta k que pasa a través del nodo i. (Shimbel, 1953)

Los nodos con una alta medida de centralidad de intermediación, al igual que las aristas, juegan un papel crítico en la red, cuando hay grandes flujos de información que es transportada por las aristas y nodos pertenecientes de la red. Los nodos que tienen una posición de intermediación son llamados controladores o reguladores de información.

Centralidad de cercanía (Closeness Centrality). La centralidad de cercanía es una métrica radial de longitud calculada como la inversa del valor medio de las distancias mínimas desde un nodo al resto de nodos del grafo. Se presenta normalizada en el rango (0,1). Por tanto, cuanto mayor sea la distancia media menor será la centralidad del nodo y viceversa, mide que tan cercano es cada individuo a los demás en la red.

Según Sabidussi (1966) esta medida tiene un foco distinto, puesto que evalúa la importancia relativa de un nodo en función de los otros nodos del grafo y además se puede considerar como una medida descentralizada o centralidad inversa dado que mientras más apartado del resto está un nodo, mayor es el valor de su centralidad.

Formalmente, la medida de cercanía $\boldsymbol{C_{CLO}(i)}$ de un nodo i se define como:

$$\boldsymbol{C_{CLO}(i) = e_i^T S 1 = \sum_{j=1}^{n} (S) ij}$$

Donde **S** es la matriz de distancias de la red, es decir, aquella matriz cuyos elementos $(\boldsymbol{i, j})$ corresponden a la distancia más corta desde el nodo $\boldsymbol{i}$ hasta el nodo $\boldsymbol{j}$, mientras menor sea el valor, se puede decir que el valor está más cercano al centro de la red. Esta medida se puede interpretar como la rapidez que tomará la propagación desde un nodo a los demás, midiendo la accesibilidad de un nodo en la red ver figura 10 nodo A. Esta medida es utilizada en teoría de grafos, donde es aplicada especialmente al análisis de red.

Centralidad de vector propio (Eingenventor Centrality). Otra medida de centralidad radial de volumen es la centralidad de vector propio que mide la influencia de un nodo en la red que corresponde al principal vector propio de la matriz de adyacencia del grafo analizado tabla 2.

Esta métrica mide la influencia de un nodo en el grafo que posee un valor alto, es decir, están conectados a muchos nodos que a su vez están bien conectados entre sí, ver nodo E en la figura 10, los cuales se pueden definir como los nodos más centrales para difundir información,

estos nodos corresponden a centros de grandes grupos cohesivos en la red. Mientras que, en el caso de la centralidad de grado, cada nodo pesa lo mismo dentro de la red, en este caso la conexión de los nodos pesa de forma diferente

Tabla 2.

Matriz de Adyacencia del grafo no dirigido

	B	A	C	D	H	E	F	G	I	J	K	L	M	N	O	P	Q
B	0	1	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0
A	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
C	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0
D	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
H	0	1	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0
E	0	1	0	0	1	0	1	1	0	0	0	0	0	1	1	0	0
F	0	0	0	0	2	1	0	1	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	0
I	1	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0
J	1	0	1	0	0	0	0	0	1	0	1	1	1	0	0	0	1
K	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1
M	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
N	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0
O	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	2	0
P	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0
Q	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0

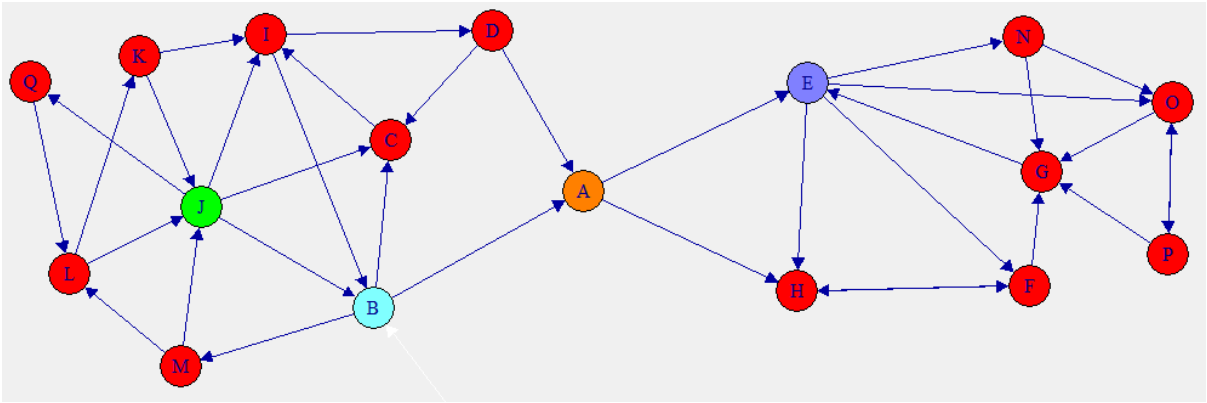


Figura 10. Ejemplo de centralidad de intermediación (B), centralidad por cercanía (A), centralidad de grado (J) y centralidad de vector propio (E). Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

Los resultados de la tabla 3 muestran las centralidades de todos los nodos del grafo de la figura 10.se puede ver que el nodo de mayor centralidad de grado es el nodo J con un grado de 7 siendo el nodo que presenta mayores conexiones adyacentes. La métrica de centralidad de intermediación, el nodo que está en todos los posibles caminos geodésicos entre todos los pares posibles es el nodo B con un valor de centralidad de 64.833. El grado de cercanía es la capacidad que tiene un nodo de llegar a todos los nodos de la red, podemos observar que cada nodo tiene un valor distinto, cabe mencionar que valores altos indican una mejor capacidad de los nodos para conectarse con los demás nodos del grafo, así podemos observar que el nodo A es quien posee al más alto grado de cercanía de 0.03125. Por el contrario, encontramos que el nodo P tiene un valor de cercanía de 0.01667, indicando que este nodo no se encuentra bien posicionado en el grafo. Por último, otra medida de centralidad que se destacó en el grafo es la de vector propio *(Eingen-vector)* indicando que un nodo es central cuando tiene muchas conexiones centrales en el grafo, el nodo E es el que está bien conectado con los demás nodos dentro de la red.

Reciprocidad: Para ilustrar con mejor detalle el concepto se tiene el grafo de la figura 10, al cual se le hallaron sus distintas medidas como la **reciprocidad** la cual describe la relación de nodos que tienen naturaleza mutua, siendo esta relación descrita por las aristas, sería la probabilidad de que si existe una arista en un sentido también exista una en el sentido opuesto. Es decir, que cuando B declara estar en relación con A, se puede esperar que A también esté en relación con B, en el ejemplo dado es claro que si bien B se relaciona con A esta última no se relaciona con B y lo mismo para la mayoría de los nodos exceptuando a P y O por lo que una ***reciprocidad*** *baja de* ***0.1111111*** *era de esperar.*

Tabla 3.

Resultados de centralidad para cada nodo

	Métricas de centralidad del grafo dirigido			
Nodos	Grado	Cercanía	Intermediación	Eigen-vector
A	4	**0,03125**	63,000	0,697
B	5	0,03030	**64,833**	0,847
C	4	0,02439	9,000	0,712
D	3	0,02632	11,167	0,503
E	6	0,02778	61,500	**1,000**
F	4	0,02083	10,333	0,745
G	5	0,02174	19,000	0,840
H	4	0,02500	7,500	0,723
I	5	0,02500	33,333	0,810
J	**7**	0,02632	48,667	0,980
K	3	0,02000	2,000	0,529
L	4	0,02000	21,000	0,542
M	3	0,02326	16,000	0,537
N	3	0,02083	2,833	0,600
O	5	0,02128	17,833	0,805
P	3	0,01667	0,000	0,556
Q	2	0,01923	3,000	0,345

*Transitividad**:*** Una *transitividad* describe también un tipo de relación pero con un nodo de más, esta se puede describir como la observación de que si a “M” se relaciona con “L” y “L” se relaciona con “J” entonces hay una alta probabilidad de que a “M” también se relacione “J”, siendo cierto para estos tres nodos en el grafo presentado, esta relación no está presente en todos los nodos ni de la misma forma, es decir, que los nodos presentados cumplen la propiedad por el orden dado ya que “M” se relaciona con “J” pero este último no se relaciona con “L”, *entonces la probabilidad de que esto se cumpla en el grafo dado es de **0.4956522** siendo baja*

.

Densidad: La medida más básica de densidad cuantifica cuántos bordes de todos los bordes posibles se realizan en un gráfico G:

$$n(G) = \frac{m}{n(n-1)/2}$$

La densidad gráfica clásica se define como la fracción total de los bordes realizados de todos los bordes posibles. Para una partición C dada, la densidad intergrupal se define como la fracción de bordes realizados entre nodos del mismo grupo de todos los bordes posibles entre nodos del mismo grupo.

La **densidad** de red cuantifica la probabilidad de relación que puede haber entre los nodos de una misma red, es decir, que en una red donde todos los nodos se relacionan con cada nodo presente en la red, en la figura 10 la **densidad** de red es de 0.1323529 o del 13,23%.

La **conectividad** muestra la relación entre los diferentes nodos del grafo ya que para que se comuniquen los diferentes nodos debe existir una conexión o camino que lo haga posible, de esta forma cuando hablamos de conectividad del grafo nos referimos a que existe un "recorrido" a través de los diferentes nodos que los conectan a los unos con los otros, de la misma forma se pueden hallar dos tipos de conectividad.

De esta forma se encuentran varias métricas de conectividad como, por ejemplo:

- *Conectividad fuerte en grafos dirigidos*: ya que en este tipo de grafos las aristas tienen un sentido definido esta conectividad consiste en estudiar si hay un camino dirigido conectando cada par de nodos del dígrafo (grafo dirigido). Inclusive se puede estudiar si existe un camino dirigido entre cada par de vértices, en ambas direcciones.

- *Conectividad General:* Consiste en encontrar el número mínimo de aristas que al ser removidas separarán el grafo en dos partes disjuntas (conectividad de aristas). También

se puede encontrar el número mínimo de nodos que al ser removidos separarán el grafo en dos partes disjuntas (conectividad de nodos).

- *Conectividad fuerte en grafos dirigidos:* dado que el grafo presentado es dirigido, es esta conectividad la que se analiza.En este tipo de grafos las aristas tienen un sentido definido, esta conectividad consiste en estudiar si hay un camino dirigido conectando cada par de nodos. Con una ***longitud del camino*** igual a 10 se puede decir que el camino más largo que separa un par de nodos es de 10 relaciones y con un promedio de relaciones entre cada par de nodos o una ***longitud media del camino*** *igual a* ***3.***

Longitud media del camino: se define la longitud de camino medio como el promedio de aristas de los caminos más cortos entre todos los posibles nodos de la red. Es una característica que mide la eficiencia de las relaciones de información en una red. Una red muy densa tenderá a tener una longitud media menor, puesto que existen muchos más caminos para llegar de un nodo a otro (Zaki y Jr., 2014).

Dado un grafo $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ sin etiquetar, es decir, sin peso alguno de sus aristas, se define el camino más corto entre dos nodos $vi, vj \in V$ como $d(vi, vj)$. Se asume que cuando $d(vi, vj) = 0$, entonces $vi = vj$. Cuando vj no es alcanzable por $vi, d(vi.vj) = \infty$. La longitud del camino medio se define por la siguiente ecuación.

$$lG = \frac{1}{n(n-1)} \sum_{i,j} d(vi, vj)$$

Donde *n* es el número total de nodos existentes en el grafo *G.*

Diámetro: antes de definir el diámetro de una red primero definamos la excentricidad en una red. La *excentricidad* ϵ de un nodo *v* es la mayor distancia geodésica o camino más corto entre los nodos de la red. En otras palabras, cuál es el camino corto más grande entre los nodos de la red.

Según Zaki y JR, (2014) el *diámetro* de una red equivale a la máxima excentricidad de cualquier nodo de la red. Esto es, el camino corto más grande entre cualquier par de nodos que conforman la red y luego encontrar el camino más grande existente. Representa el tamaño de la red y permite saber lo grande que es el mismo, las redes más dispersas suelen tener un mayor diámetro que las más densas al existir menos caminos entre cada par de nodos. Dado un grafo $G = (V, E)$, el diámetro del grafo *G* se define como:

$$DG = \max\{d(vi, vj)\} \forall\, vi, vj \ \in V$$

El *grado* es claramente un atributo local de cada nodo. Uno de los atributos globales más simple es el *Grado medio* de la red, como la medida del grado de los nodos. El grado de un nodo es el número de enlaces conectados a ese nodo

$$\mu d = \frac{\sum_i di}{n}$$

Las definiciones anteriores se pueden generalizar fácilmente para gráficos dirigidos (ponderados). Por ejemplo, podemos obtener el grado de entrada y el de salida tomando la suma de las aristas entrantes y salientes, el grado medio se define como:

$$id(vi) = \sum_{j} A(j,i)$$

$$od(vi) = \sum_{j} A(i,j)$$

5.8.4. Conglomerados. Un conglomerado o “clúster” se puede definir como la división de los nodos de una red en subgrupos dentro de los cuales, las conexiones o aristas entre los nodos que la forman son densas o numerosas, pero las conexiones intercomunitarias son exactas.

Los conglomerados o clúster pueden definirse como subconjuntos de una muestra S, satisfaciendo ciertas condiciones. En este enfoque, el término de "grupo" está bien definido, ya que los conglomerados se definen como subconjuntos de S, que presentan determinada homogeneidad. Tales clústeres predefinidos generalmente se descubren utilizando medidas de similitud, disparidad, disimilitud o distancia, respectivamente; en raras ocasiones, los datos brutos se usan directamente para buscar clases. En la figura 11 se puede apreciar un ejemplo de un grafo con tres comunidades bien diferenciadas entre sí. Cada comunidad está formada por un conjunto de vértices con muchas uniones entre ellos, y en cada comunidad está conectada con las demás con pocos enlaces, estos enlaces pueden denominarse puentes entre comunidades.

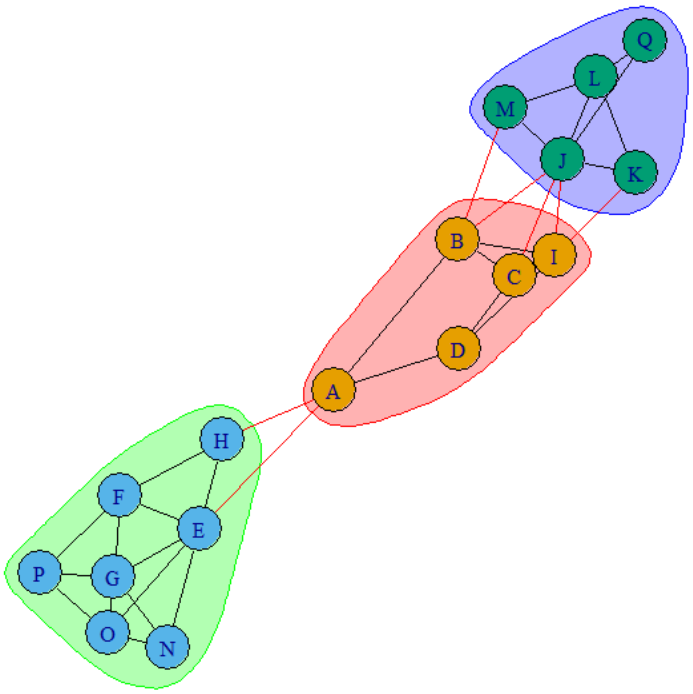


Figura 11. Ejemplo de conglomerados en un grafo de densidad. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

Godehardt presenta definiciones de clúster que se basan en la teoría de grafos, estas definiciones permiten el enunciado y la prueba de hipótesis sobre una estructura de clase de una muestra. De esta forma la distancia entre dos objetos se puede plantear como el peso de conexión de los dos vértices *(nodos)* en el grafo. Los resultados se utilizan para derivar estadísticas de prueba y para probar la hipótesis de aleatoriedad de tales grupos, suponiendo que la aleatoriedad de los clústeres significa la aleatoriedad de las distancias entre los objetos que se agruparán. (Godehardt, 1988)

Formalmente, dado un conjunto de datos, el objetivo de la agrupación es dividir el conjunto de datos en comunidades de modo que los elementos asignados a una determinada comunidad sean similares o estén conectados en algún sentido predefinido. Sin embargo, no todos los gráficos tienen una estructura con clústeres naturales. No obstante, un algoritmo de

agrupamiento genera un clúster para cualquier gráfico de entrada. Si la estructura del gráfico es completamente uniforme, con los bordes distribuidos uniformemente sobre el conjunto de vértices, la agrupación calculada por cualquier algoritmo será bastante arbitraria.

Algunos métodos de agrupamiento de comunidades que se utilizan para esta técnica son: agrupamiento por densidad, agrupamiento jerárquico y agrupamiento global diviso. (Elisa Schaeffer, 2007)

Agrupación por densidad: El primer algoritmo que emplea este enfoque para dividir el conjunto de datos es DBSCAN (*Density Based Spatial Clustering of Aplications with Noise*), en este aparecen los conceptos: punto central, borde y ruido los que serán empleados para determinar los diferentes clústeres.

El algoritmo comienza seleccionando un nodo i arbitrario, si i es un nodo central, se comienza a construir un grupo y se ubican en su grupo todos los nodos denso-alcanzables desde i. Si i no es un nodo central se visita otro objeto del conjunto de datos. El proceso continúa hasta que todos los nodos han sido procesados. Los nodos que quedan fuera de los grupos formados se llaman nodos de ruido, los nodos que no son ni ruido ni centrales se llaman puntos bordes. De esta forma DBSCAN construye grupos en los que sus nodos son o nodos centrales o nodos de borde, un grupo puede tener más de un nodo central, (ver figura 11).

Los algoritmos de este tipo, dado un conjunto de datos D, definen un criterio de densidad para un grupo y tratan de encontrar grupos o divisiones D1, D2, ..., Dk de este conjunto de datos de forma tal que las densidades de estas divisiones sean las más cercanas posibles. En el mejor de los casos, estos algoritmos son capaces de detectar automáticamente el número de

grupos k en los que se deben dividir los datos, así como son capaces de descubrir grupos de diferentes formas y tamaños.

Agrupación jerárquica: Los métodos de agrupación en clúster que producen agrupaciones multinivel se denominan algoritmos de agrupación jerárquica, en oposición a agrupaciones planas que comprenden una única partición o cubierta. Un clúster jerárquico generalmente se construye generando una secuencia de particiones, donde cada subcluster pertenece a un supercluster en su entidad. El clúster raíz contiene como máximo todos los datos, y cada uno de los clústeres contiene al menos un elemento de datos; los conglomerados semánticamente relevantes generalmente aparecen en niveles intermedios.

Los métodos jerárquicos usan una matriz de distancia como entrada para el algoritmo de agrupamiento. La elección de una métrica adecuada influirá en la forma de los clústeres, y que algunos elementos pueden estar cerca uno del otro según la distancia. Para la salida de agrupamiento se visualiza por medio del dendrograma (ver figura 12).

Esta agrupación depende de los datos de entrada, para poder calcular una jerarquía de clústeres, cada nivel de agrupamiento define un subconjunto diferente y, por lo general, los conglomerados definidos por los niveles superiores contienen los conglomerados de los niveles inferiores como subgrafos.

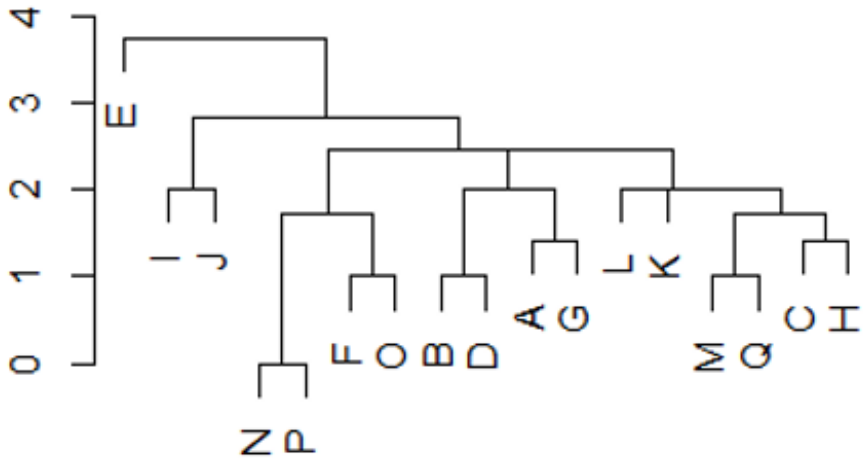


Figura 12. Ejemplo de agrupamiento por dendrograma. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2.

Agrupamiento global divisivo: Los algoritmos de agrupamiento divisional son una clase de métodos jerárquicos que funcionan desde arriba, partición recursiva del gráfico en clústeres. La división en cada iteración suele ser en dos conjuntos, pero no hay ninguna razón por la que un algoritmo de agrupación no pueda dividir un conjunto de vértices en más de dos conjuntos para la siguiente iteración. Los agrupamientos de este tipo asumen inicialmente que todos los nodos pertenecen a un mismo grupo. A partir de este nodo y aplicando técnicas particionales se realiza un proceso en que iterativamente se selecciona un grupo para ser dividido en dos.

Este proceso termina cuando se satisface algún criterio de convergencia. Generalmente los criterios utilizados pueden ser: *(i)* se alcanza una cantidad de grupos determinados o *(ii)* los grupos formados satisfacen algún criterio de cohesión. El algoritmo divisivo más utilizado es el *Bisecting K-means,* que utiliza el algoritmo particional K-means para iterativamente ir dividiendo un grupo seleccionado.

6. Recolección de datos y metodología

En este estudio se aplica el análisis de las redes sociales en particular Twitter para convertir los datos de la red social en conocimiento. Proporciona información para comprender el papel fundamental de la red en la propagación de información de emergencias, es necesario disponer de aplicaciones para la recolección sistemática y procesamiento de datos. Las fases de este estudio son las siguientes.

- Recolección de datos de la red social Twitter de Indonesia durante el periodo de erupción del volcán Sinabung 2018, de 18 de febrero al 2 de marzo.
- Limpieza de la base de datos, en el cual se eliminaron los bost (tweets repetidos)
- Procesamiento de los datos para identificar las conexiones y los patrones creados por las interacciones de los usuarios de Twitter durante la erupción.
- Generación de grafos generales creados por las interacciones agregadas de los usuarios de Twitter durante y después de la erupción del volcán.
- Cálculo y análisis de métricas de centralidad a nivel de red y global entre las relaciones y nodos de cada grafo.
- Visualización y análisis de conglomerados por método de agrupamiento por densidad.

La elección de Twitter como una herramienta de mayor fuente de información es debido a su mayor facilidad para almacenar los datos, pero sobre todo la manera para acceder a ellos es más rápida, donde el contenido es público y accesible. Además, ofrece una API sencilla y amigable de utilizar para realizar búsquedas muy específicas incluyendo algunos parámetros predefinidos en la consulta. Una API es una interfaz de programación de aplicaciones

(Application Programming Interface) que facilita la interacción humano-software. En este caso es utilizada para extraer la información de la red social Twitter a partir de la cuenta del usuario y contraseña de la red social. Una vez registrado hay que dar de alta la aplicación, obteniendo una serie de claves *(Consumer Key, Consumer Secret, etc)* mediante las cuales se realiza la conexión con Twitter desde el software R-studio.

La elección del evento se realizó por medio de la página El Centro para la Investigación sobre la Epidemiología de los Desastres CRED, permitiendo la facilidad de encontrar eventos que hayan ocurrido últimamente. Por otra parte, se tuvo en cuenta la mayor cobertura de las redes sociales y en particular Twitter en los países que presentan mayor sismicidad y vulcanismo.

Indonesia se estima que es el país con mayor actividad sísmica y volcánica en los últimos años donde hay alrededor de 127 volcanes activos de los cuales el monte *Sinabung*, en el norte de *Sumatra*

6.1. Extracción y recolección de datos

Para investigar la dinámica de la red social Twitter y la formación y evolución de comunidades en línea en respuesta durante un desastre, se descargó un conjunto de datos de Twitter durante y después de la erupción del volcán *Sinabung* de 2018 en Indonesia empleando la API que permite incursionar en el núcleo de Twitter, para recopilar los Tweets sobre temas relacionados con el desastre. El procedimiento de la aplicación es el siguiente: la descarga de tweets se ha orientado a búsqueda por palabras claves (los hashtags). Dada una búsqueda, Twitter filtra los tweets que contengan alguna de las palabras claves, devolviendo a través de la API una lista de

tweets. Los hashtags utilizados para la extracción de datos del estudio durante la erupción del volcán Sinabung en Indonesia fueron *#Sinabung #Sumatra*.

Los datos de Twitter tienen unas características especiales que se describen a continuación, una vez obtenidos. Existen varias librerías de *R* que permiten la facilidad y el manejo de extraer los datos de texto como el libro *twitteR*. Los datos de texto, como ocurre con los tweets, tienen una estructura muy simple. El modo más útil de imponer una estructura a este tipo de datos es convertirlos en lo que denominaremos matriz de término-documento.

La recolección de datos se realizó entre el 18/02/2018 al 02/03/2018. El conjunto de datos final incluye un total de 42.087 tweets provenientes de los 31.450 usuarios relacionados con el hashtag *#Sinabung.* Una vez obtenida la información se procedió a la depuración de ésta para su posterior análisis, aplicando para ello estadísticas y técnicas de análisis de datos como: análisis descriptivo, análisis de red.

Como resultado tenemos una base de datos con 16 columnas y 42.087 filas. Cada columna es una variable y cada fila contiene la información de todas las variables de cada tweet. En la figura 13 se describe las variables de la base de datos creada.

Campo	Descripción
Created_at	Fecha y hora de cuando se creó el Tweet
Id	Identificación del tweet
Text	Texto del tweet publicado
User(screenName)	Nombre del usuario que ha publicado el tweet
Favorite_count	Número de favoritos que contiene el tweet
retweet_count	Número de veces que el tweet ha sido retweetteado
Favorited	Indica si algún usuario lo ha marcado
ReplyToSN	Si el tweet es respuesta a otro, contiene el nombre del usuario del tweet original.
Truncated	Es un valor lógico que señala si el texto es truncado, es decir, si supera los 140 caracteres aparecen puntos suspensivos.
replyToSID	Contiene la identificación del tweet original si éste en respuesta a otro
replyToUID	Si el tweet es respuesta a otro contiene el identificador del usuario del tweet original
statusSource	Dirección HTML en la que se ha publicado el tweet
isRetweet	Variable de tipo lógico, indica si el tweet es un retweet o no
Retweeted	Indica si el tweet ha sido retweeteado por el usuario que se autentica
Longitude, Latitude	Indica la ubicación geográfica del tweet

Figura 13. Descripción de las variables generadas en la base de datos.

7. Análisis descriptivo de los Tweets

7.1. Ubicación de los tweets que comprende la muestra

La muestra de tweets, como se ha mencionado previamente, se compone de 42.087 publicaciones. Debido a las limitaciones encontradas con la geolocalización, tan solo 40 tenían la ubicación exacta activada. Por tanto, se ha ubicado únicamente el 0.095% de la muestra inicial. En el mapa de Indonesia se han ubicado 34 *tweets* (figura 14) que contenían geolocalización exacta, permitiendo así, una visualización en los distintos lugares cercanos al desastre.

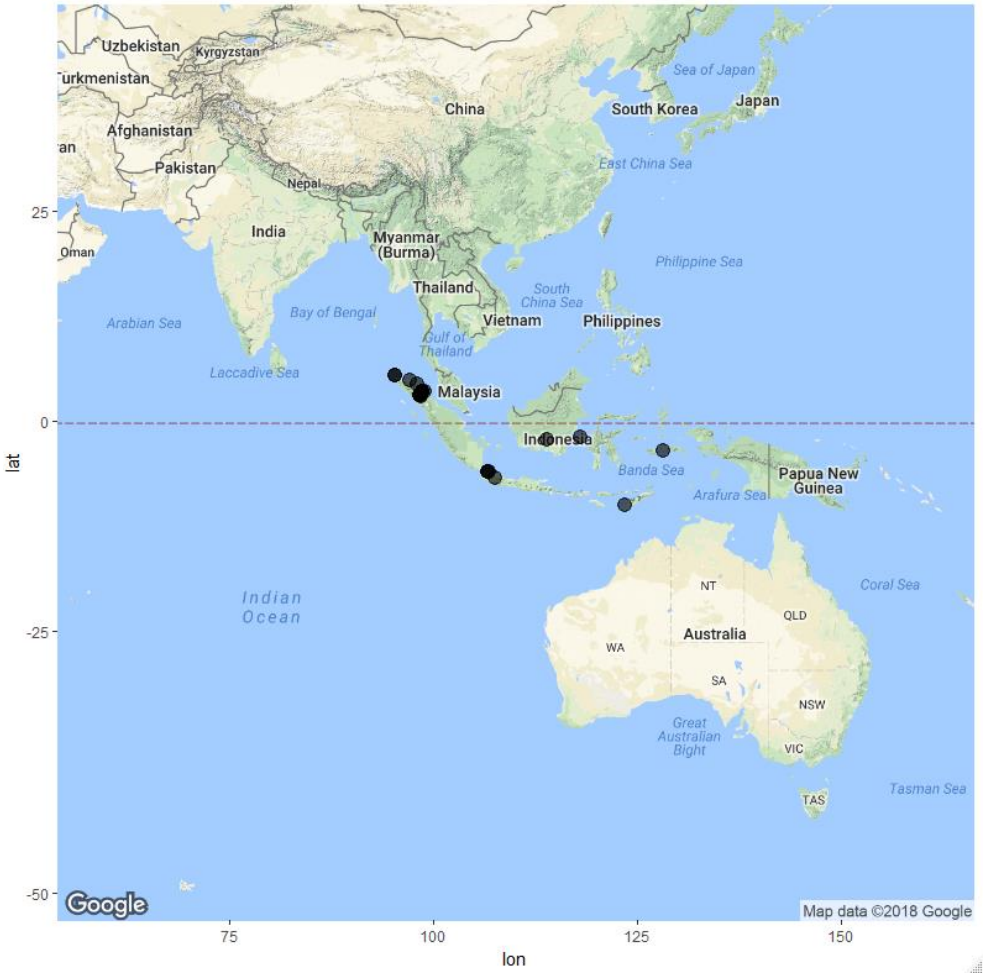


Figura 14. Distribución de los tweets geo localizados en Indonesia. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En la figura 15 se observa 6 tweets ubicados en diferentes partes del mundo, el contenido de los tweets hace mención a noticias publicadas por los diferentes organismos de desastre a nivel mundial sobre el desastre.

Figura 15. Distribución de los tweets geo localizados a nivel mundial. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

7.2.Métricas de los tweets

Entre 42.087 tweets, los tweets originales y retuits los cuales representan el 19.69%(8.288), y el 80.31%(33.799), respectivamente se encontró más de 939 hashtags diferentes en los tweets 102 urls y 31.450 usuarios. Entre los hashtags más populares se encuentran: *#sinabung, #sumatra, #indonesia, #volcano* y *#eruption.*. En la tabla 4 se muestra una descripción de los datos en cada uno de los escenarios a analizar.

Tabla 4.

Información del conjunto de datos

Datos	Tweets	Retuits	Total, Tweets	Hashtags	Usuarios	URLS
Erupción del volcán Sinabung (Indonesia)	8.288	33.799	42.087	939	31.450	102

7.3. Análisis de los usuarios

La suma de los tuits realizados cuenta los tuits originales y los retuits conjuntamente dando un total de 42.087 observaciones, en las cuales se originaron 28.273 conversaciones o replicas, también hubo 28.717 citas en la conversación (figura 16). Por otro lado, los datos recabados fueron generados utilizando la *API twitteR* por lo que se cuenta con restricciones de capacidad en el momento de producir una muestra para su estudio.

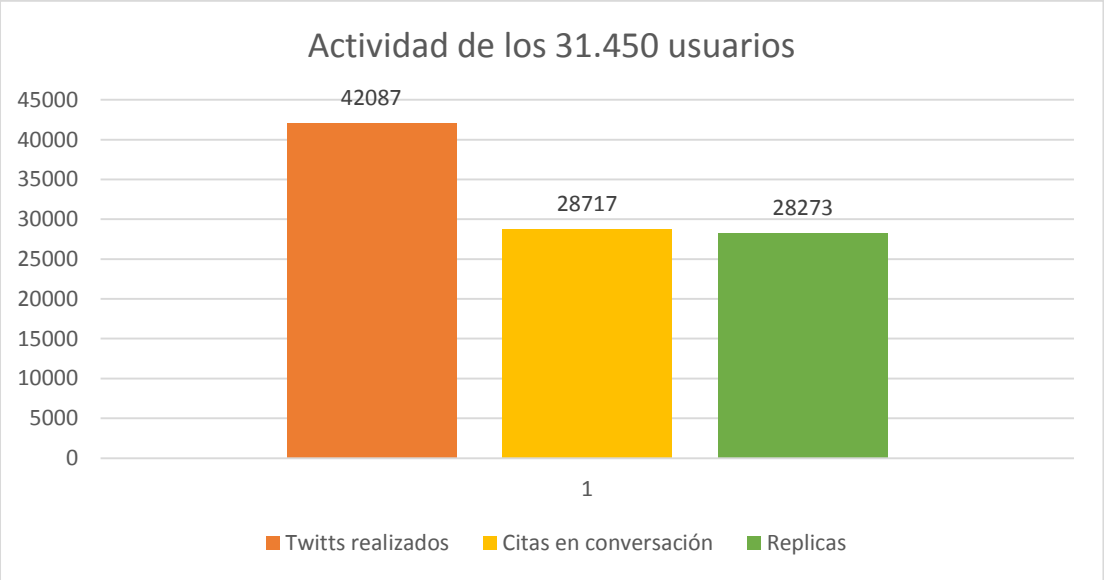


Figura 16. Resumen de actividad de los usuarios. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

7.4.Análisis de los usuarios más activos

En la figura 17 presenta una distribución exponencial y contiene información de los usuarios más activos en *#Sinabung*. Se encontraron 31.450 usuarios únicos en el conjunto de datos. Los usuarios activos se calculan en función del número de tweets. La visibilidad de los usuarios se puede calcular por el número de respuestas recibidas.

El usuario más activo *@infobencana* con un total de 145 tweets realizados, de los cuales 125 son tweets originales y 20 retuit, es un usuario registrado como un centro de medios y habla en sus tweets sobre desastres, derechos humanos y medio ambiente; además compartió contenidos como: imágenes, videos, enlaces, etc. En torno al evento transmitió alertas desde que empezaron a notar nubes de calor eruptivo hasta el momento de la erupción

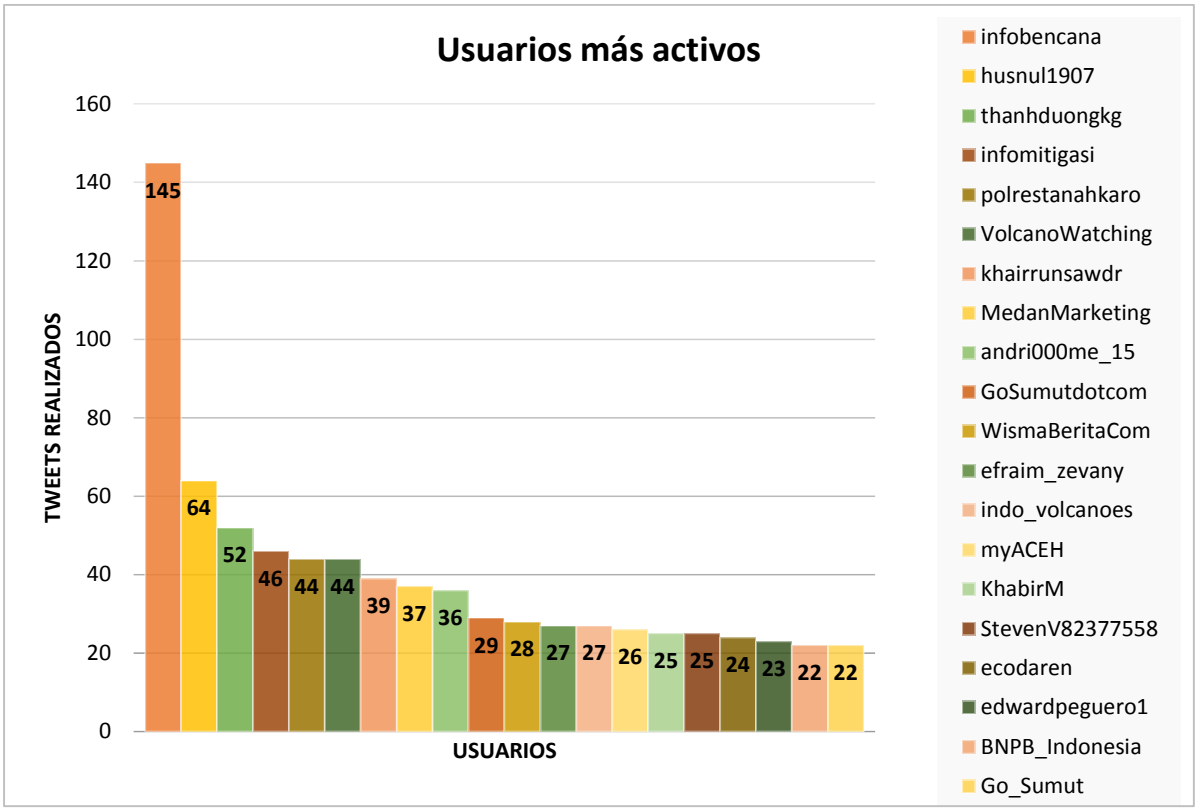


Figura 17. Usuarios más activos en las publicaciones. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Con el gráfico *"usuarios más visibles"* de la figura 18 podemos encontrar aquellos usuarios con mayor actividad en los tweets, encontrándose 20 usuarios más visibles que han hecho retuits sobre temas específicos relacionados con el desastre durante la línea de tiempo seleccionada, se puede ver que el usuario que presenta mayor actividad en los tweets es *@Sutopo_pn* su nombre Sutopo Purwo Nugroho, portavoz y jefe del centro de datos de BNPB.

En el siguiente análisis se tendrá en cuenta el uso de los retuits, hashtags, réplicas y urls. De esta forma, los grafos construidos uno por cada tipo de relación cuyos enlaces serán dirigidos en algunos grafos. En las relaciones de los retuits y replicas, el peso de estos enlaces será establecido por el número de veces que un usuario ha replicado o realizado retuit a otro usuario. En cambio, en el grafo generado a través de la relación de hashtag, los enlaces serán

establecidos por la relación que existe en las publicaciones y los hashtags más relevantes serán aquellos más frecuentes en las publicaciones.

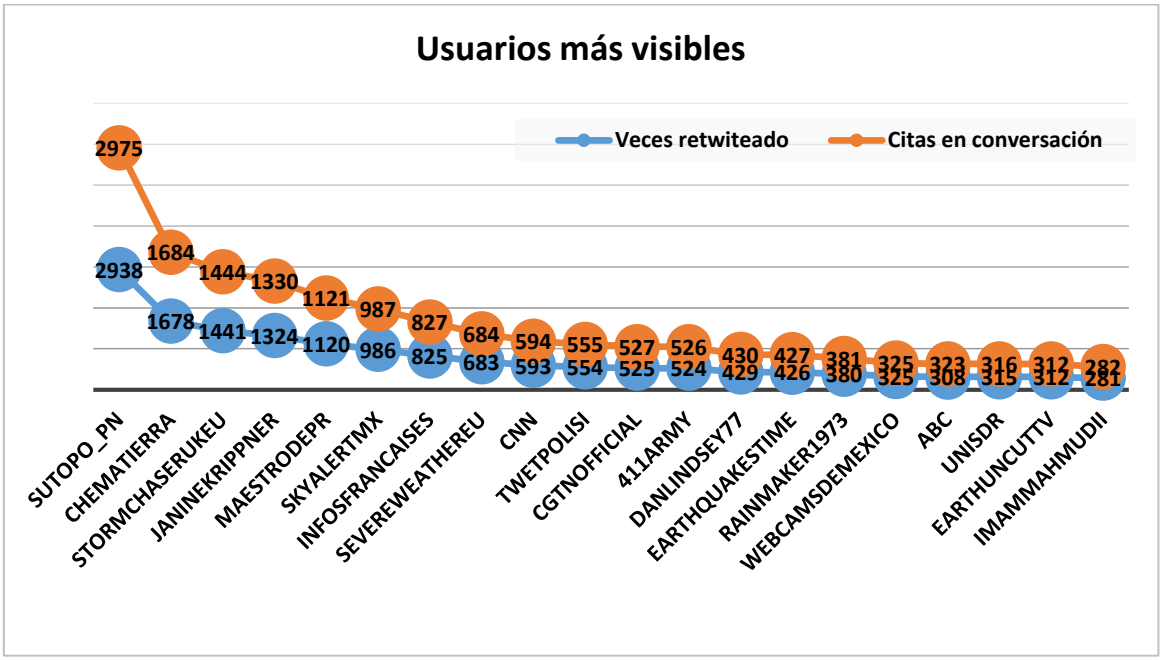


Figura 18. Usuarios más visibles en cuanto a citas y veces de retuits. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

7.5.Análisis de resultados Retweets

En la figura 19 indica la frecuencia de los retweets durante el periodo de observación que fue de 11 días. Asimismo, se observa una periodicidad alta en febrero 20, donde los usuarios aumentaron su actividad en línea durante la erupción, retuiteando así los tweets generados a partir del día 18 de febrero donde se dieron las primeras alertas del volcán. E3n este día solo hubo un tweet generado, a las 22 horas del día 18 de febrero donde se empezó a notar fuego eruptivo. En febrero 19 (figura 20) el día con mayor periodicidad con un total de 770 tweets generados por los usuarios de medios y comunicación.

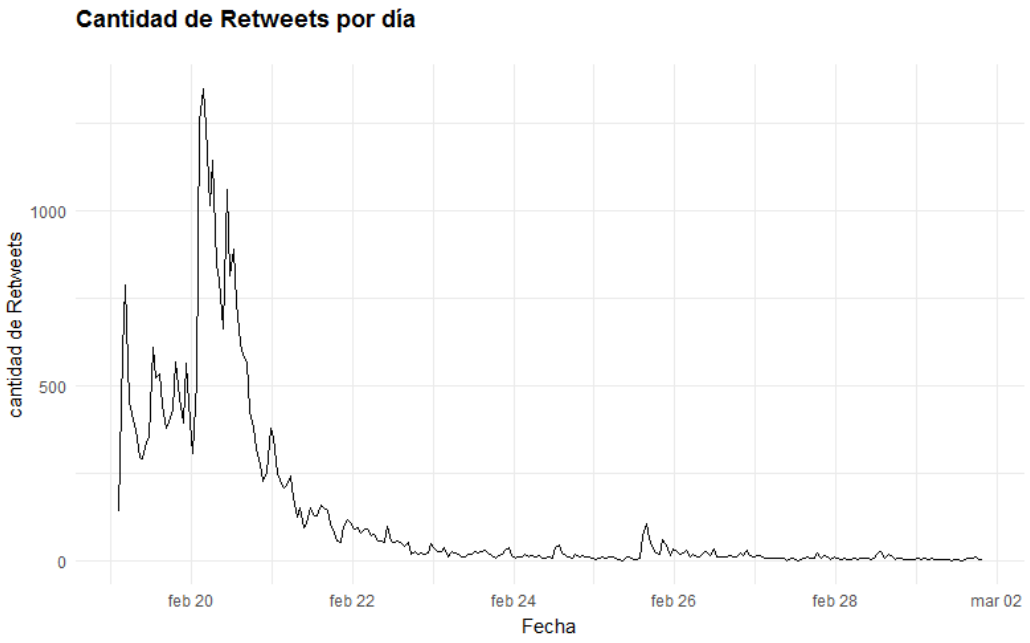


Figura 19. Cantidad de Retweets por fecha. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

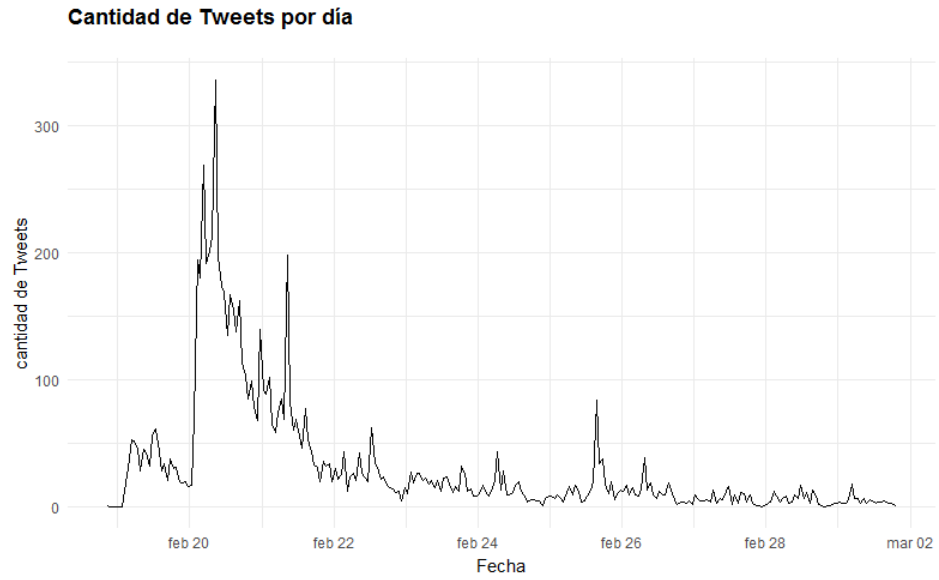


Figura 20. Cantidad de Tweets por fecha. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

7.6.Análisis de usuarios que han hecho Retweets

Teniendo en cuenta los usuarios que dieron retuits se encuentra que el usuario *@Sutopo_pn* con una cantidad de 2.938 retuits, este usuario es un portador y jefe del centro de datos de la Agencia Nacional de Gestión de Desastres *(BNPB)* en Indonesia, siguiéndole el usuario *@chematierra* con 1.678 retuits, geólogo de México la mayoría de los tweets hecho por él trataban de riesgos geológicos sobre el volcán *Sinabung.*

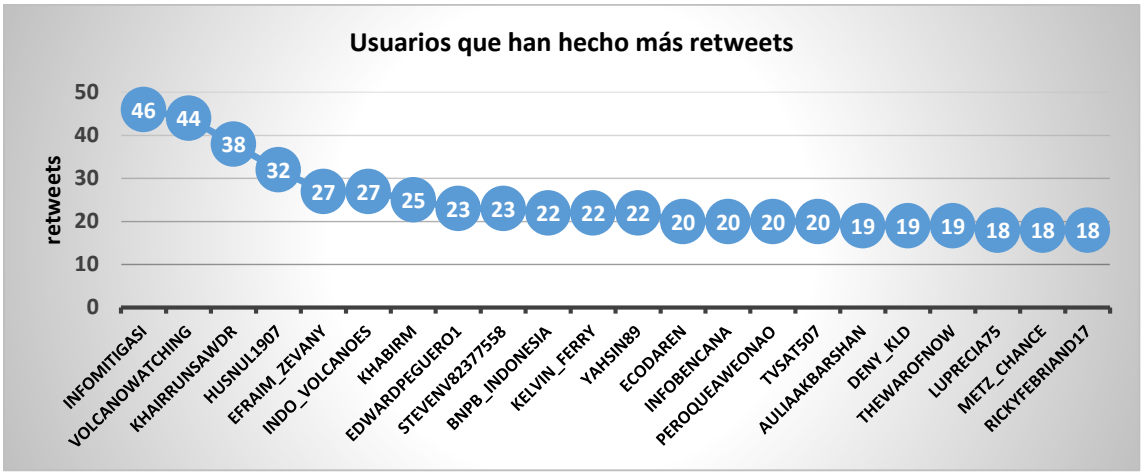


Figura 21. Principales usuarios destacados en retuits.Extraído de software para análisis estadístico R_Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En un análisis más detallado de tendencia central (figura 21), se muestra los 22 principales usuarios más destacados en sus publicaciones. El usuario más influyente en los retuits *@infomitigasi* con 46 de los cuales el 21.74% de los retuits corresponden a los tweets de *@infobencana,* y el 8.7% de los retuits son hechos al usuario *@sutopo_pn.* La media de los datos que nos indican la cantidad de retuits hechos por cada usuario es de 1,267, donde la mayoría de usuarios en el conjunto de datos está en un rango de 1 a 10 retuits.

En la figura 22 se muestra la frecuencia de los datos expresado en un histograma de frecuencias, se filtraron los usuarios que tuvieron más de 10 retuits para una mejor comprensión de los datos, donde la mayor acumulación se encuentra sesgada a la derecha. 48 usuarios, es decir el 80% realizaron de 10 a 20 retuits; solo 8 usuarios (el 13.33%) realizaron de 20 a 30 retuits y; el 6.66% de los usuarios que retuitearon con mayor frecuencia, es decir, estos usuarios son los más centrales en el grafo de retuits.

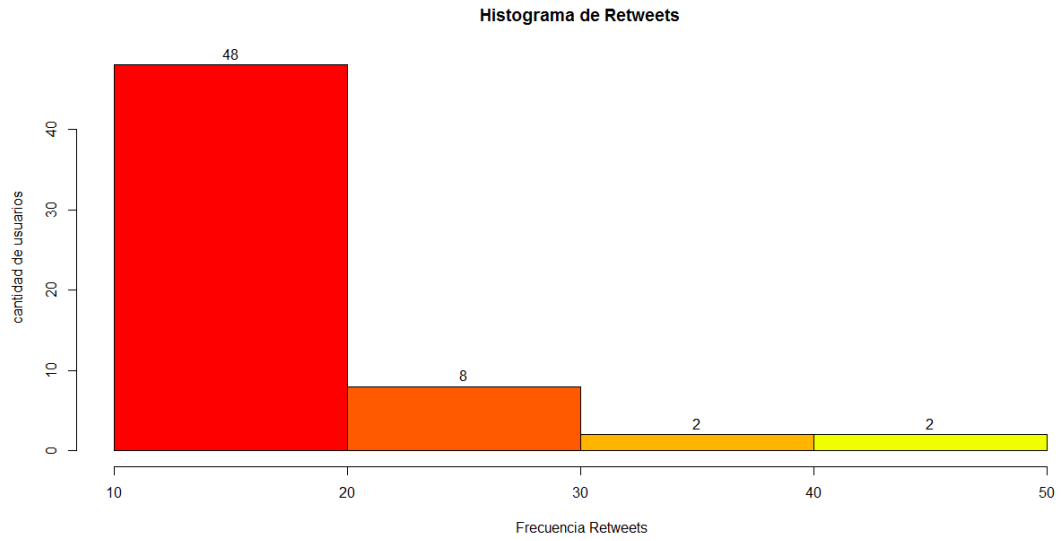


Figura 22. Frecuencia de Retweets. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

8. Análisis de red de los Retweets

8.1.Análisis de red método de visualización

Una vez creada la base de datos lo siguiente que se realizó fue buscar la forma de poder visualizarla en forma de grafo. La elaboración de esta red se efectuó, en una primera instancia sin restricciones respecto a la centralidad de los nodos. Como resultado se obtuvo un diagrama general figura 23. Se observa algunos grupos muy conglomerados alrededor de un nodo en

particular, asimismo se visualiza una posible existencia de clústeres asociados a determinados nodos, como se ve reflejado en los vecindarios de los nodos de color naranja

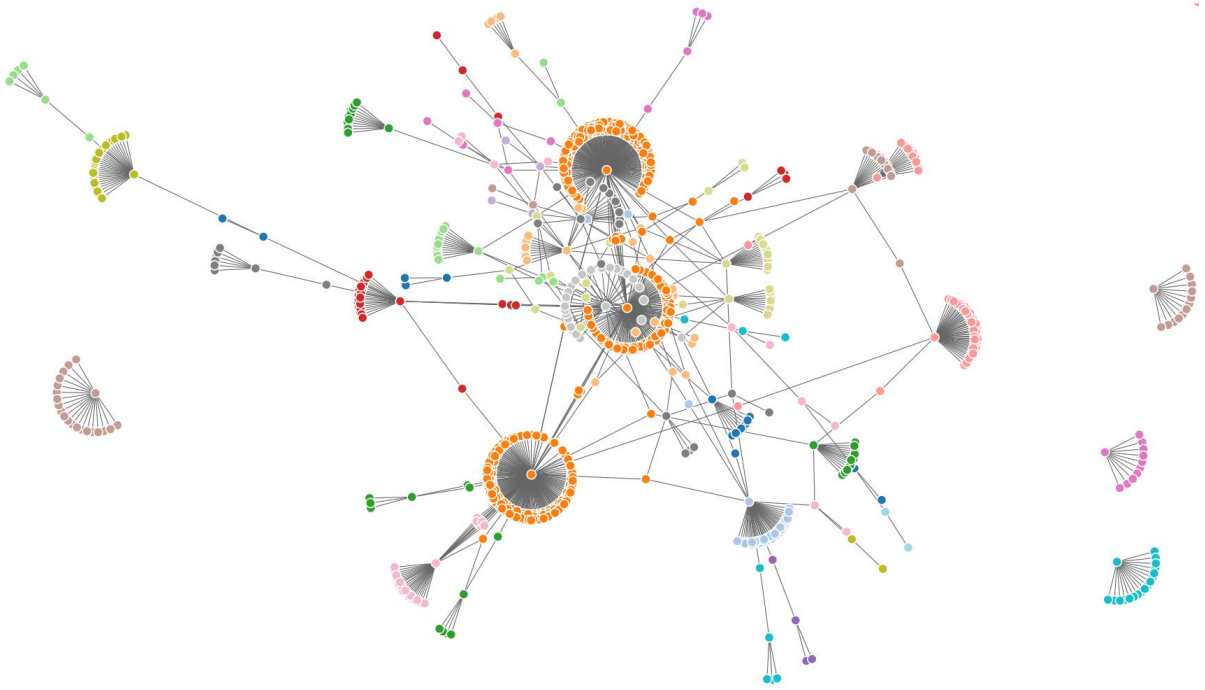


Figura 23. Estructura general de red de retuits.Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

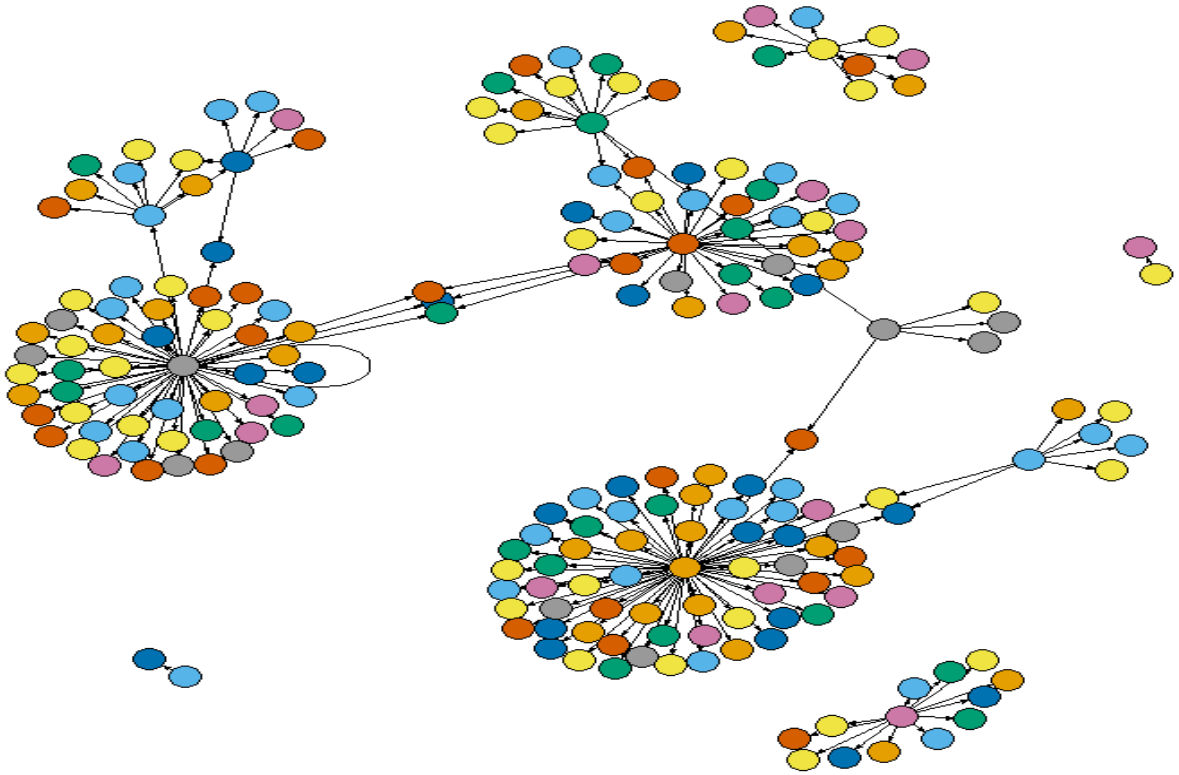


Figura 24. Grafo de los principales usuarios destacados en los retweets. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En la figura 24, se muestra un subgrafo a partir de la figura 23 extrayendo los nodos más centrales en el diagrama general de la red retweet incluyendo 264 enlaces y 282 vértices lográndose una representación más clara. El gráfico de red ofrece una vista general de las interacciones de los usuarios más influyentes durante el periodo del desastre, los nodos son aquellos usuarios que dieron retuits a los tweets y los enlaces son las relaciones que tuvieron en común. La transitividad de la red, que es la probabilidad de que dos nodos sean a su vez vecinos, es baja de 0.075%. La longitud de la trayectoria de la mayoría de los nodos de la red es de 10, el diámetro es 13, que es el camino más largo entre dos nodos del grafo. El diámetro y las longitudes de camino en el grafo generado son relativamente pequeños respecto al número de enlaces y nodos de los mismos. La reciprocidad del grafo es nula, esto ocurre a causa de las

configuraciones de la red social Twitter. Las relaciones suelen estar orientadas a una sola dirección, es decir, un usuario puede hacer a otro un retuit pero no necesariamente al revés. La densidad de la red es de 0.003319753, que es la proporción de todos los nodos posibles en la red, por tanto, es una medida de cohesión en la red, este valor es bajo lo que convierte en un grafo de tipo disperso.

8.2.Análisis de medidas de centralidad

Para medir las estadísticas a nivel de vértice no se consideró los nodos con valores de centralidad de grado en cero. En primera instancia es interesante ver cuáles son los usuarios que presentaron mayores grados de centralidad. Además, en la figura 25 se observan las estadísticas a nivel de nodos, tales como la centralidad de grado que es una medida clave para el estudio de los usuarios con mayor número de relaciones con otros usuarios en la red.

Centralidad de grado (local) Aplicando la métrica de centralidad de grado como se observa en la figura 25, se evidencia los nodos *(usuarios)* más importantes en la red. El grado de un nodo se define como el número de conexiones directas que tiene con otros nodos. Como se ilustra, el usuario *@sutopo_pn,* los tweets generados por este usuario trataron temas relacionados con el desastre y *@chematierra,* los tweets destacados por este usuario trataron temas de riesgos geológicos y desastres naturales que generaron información de interés para muchos usuarios, pero el volumen de conexiones que tiene sugiere que en realidad ocupan roles de distribuir la información, tales como las comunicaciones internas.

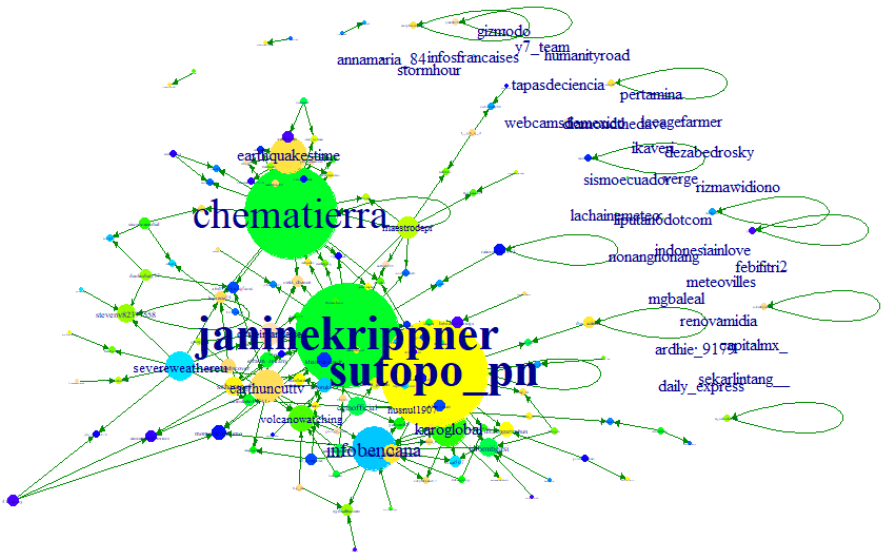


Figura 25. Principales nodos (usuarios) con mayor centralidad de grado. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En la figura 26 se muestra la frecuencia de los datos expresado en un histograma de frecuencias, la mayor acumulación se encuentra sesgada a la izquierda; 27.803 usuarios, es decir el 99.9% de los usuarios presentan menor centralidad de grado *(degree),* solo 8 usuarios, es decir el 0.028% de usuarios.Los usuarios *@sutopo_pn, @chematierra, @janinekrippner* y *@maestrodepr,* son los más relevantes en la red destacándose por su centralidad mayor, es decir, estos usuarios trataron temas de interés sobre el desastre.

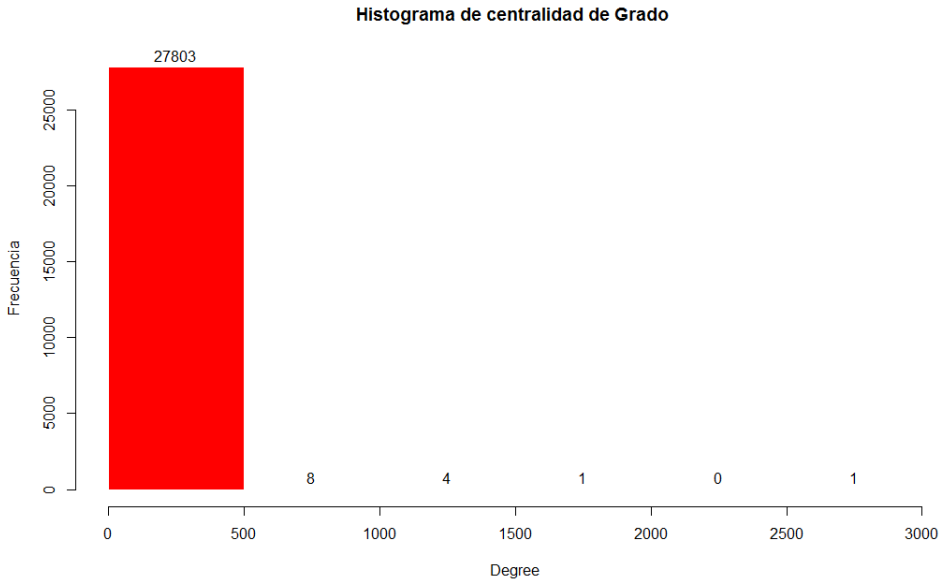


Figura 26. Histograma de centralidad de grado. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

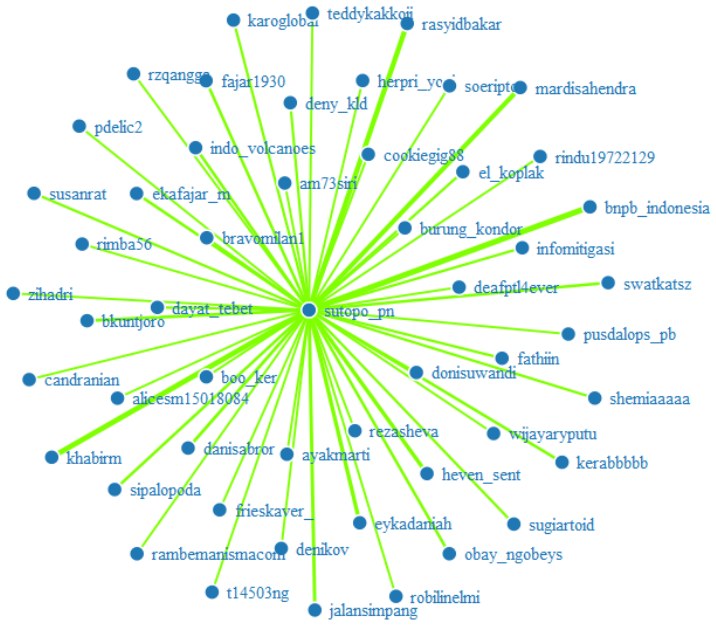


Figura 27. Sección de grafo general de mayor centralidad de grado. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Al ver solo los retuits, se nota la formación de un grupo alrededor del nodo *@sutopo_pn (*figura 27); para esta red se tuvo en cuenta los usuarios más influyentes que han retuiteado los tweets de este usuario. Los usuarios que tuvieron interés sobre sus publicaciones fueron *@bnpb_indonesia* con un total de 20 retuits, *@khabirm* con 17 retuits, *@rasyidbakar* con 15 y *@mardisahendra* con 13 retuits, estos usuarios formaron un grupo alrededor del nodo y no interactuaron entre ellos. La comunicación entre redes también fue limitada; muy pocos retuitearon contenidos.

8.3.Medidas de centralidad global

Centralidad de cercanía En la figura 28 se ilustra una de las medidas de centralidad global más ampliamente utilizadas como la centralidad de cercanía. Esta medida puntúa cada nodo según su cercanía con todos los otros nodos dentro del grafo. Según la cercanía, los resultados de centralidad, a diferencia de la centralidad de grado, los usuarios *@chematierra* y *@sutopo_pn* tienen los valores de centralidad de cercanía más alto a través de todo el grafo, esto significa que estos usuarios tienen roles igualmente importantes en el flujo de la red.

En la figura 29 se muestra la frecuencia de los datos expresado en un histograma de frecuencias; la mayor acumulación se encuentra sesgada a la derecha; 1034 usuarios, es decir el 77.69% de los usuarios presentan mayor grado de cercanía solo 299 usuarios, es decir el 22.43% de usuarios presentan menor valor. Los usuarios *@sutopo_pn, @chematierra,y @janinekrippner,* son los más relevantes en la red destacándose por su grado de cercanía.

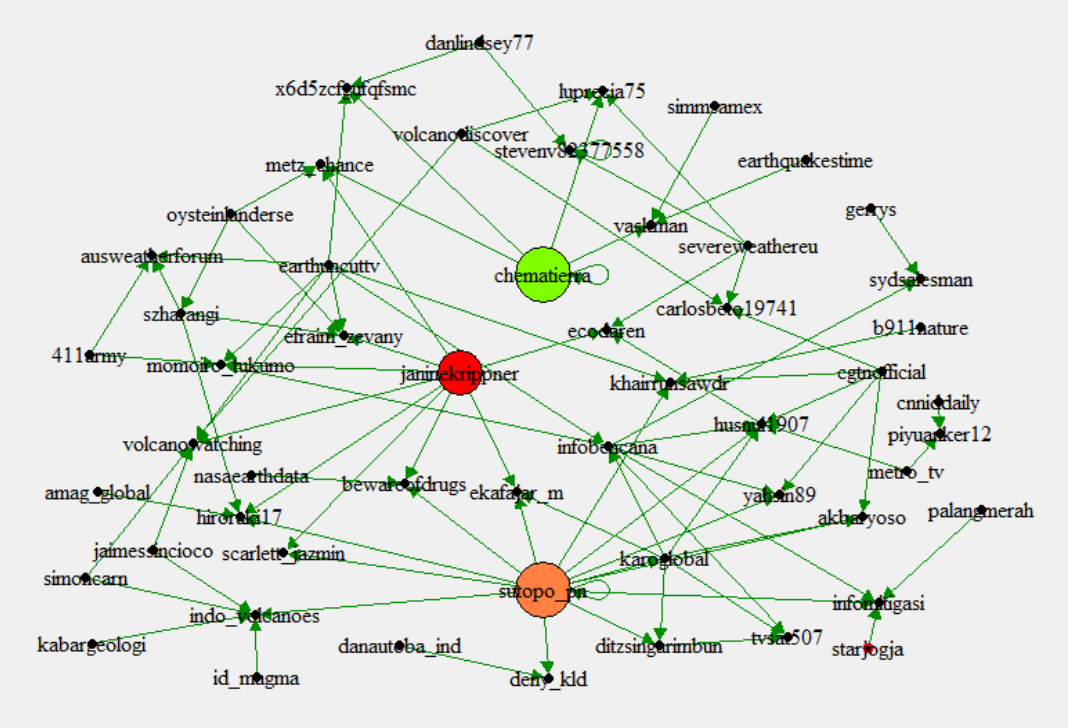


Figura 28. Principales nodos (usuarios) con mayor centralidad de cercanía. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

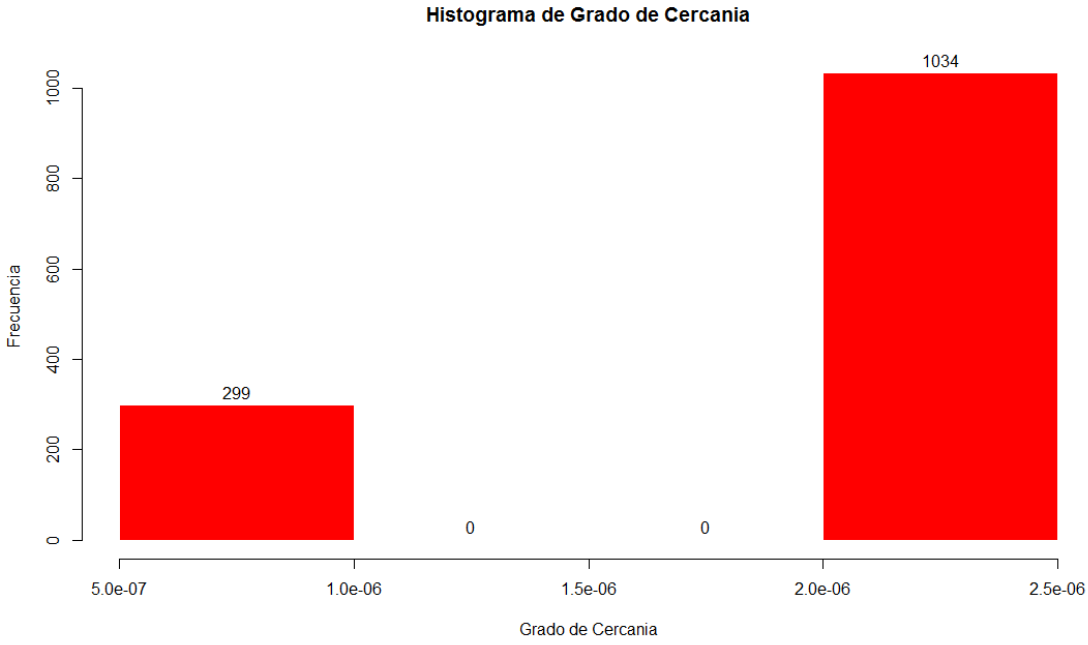


Figura 29. Histograma de centralidad de cercanía. Extraído de software para análisis estadístico R-.Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Centralidad de intermediación Otro nivel de nodo que se destacó es la centralidad de intermediación que se presenta en la figura 30. Los usuarios que presentan mayor centralidad de intermediación son vitales para conectar otras comunidades para la difusión de la información. Se identifica varios usuarios que desempeñan un papel importante en la red, que presentan caminos más cortos entre ellos. También se denominan guardianes, ya que tienden a controlar el flujo de información entre las comunidades. Un gráfico de red describe claramente el papel de los vértices con mayor centralidad de intersección de los vértices de la red.

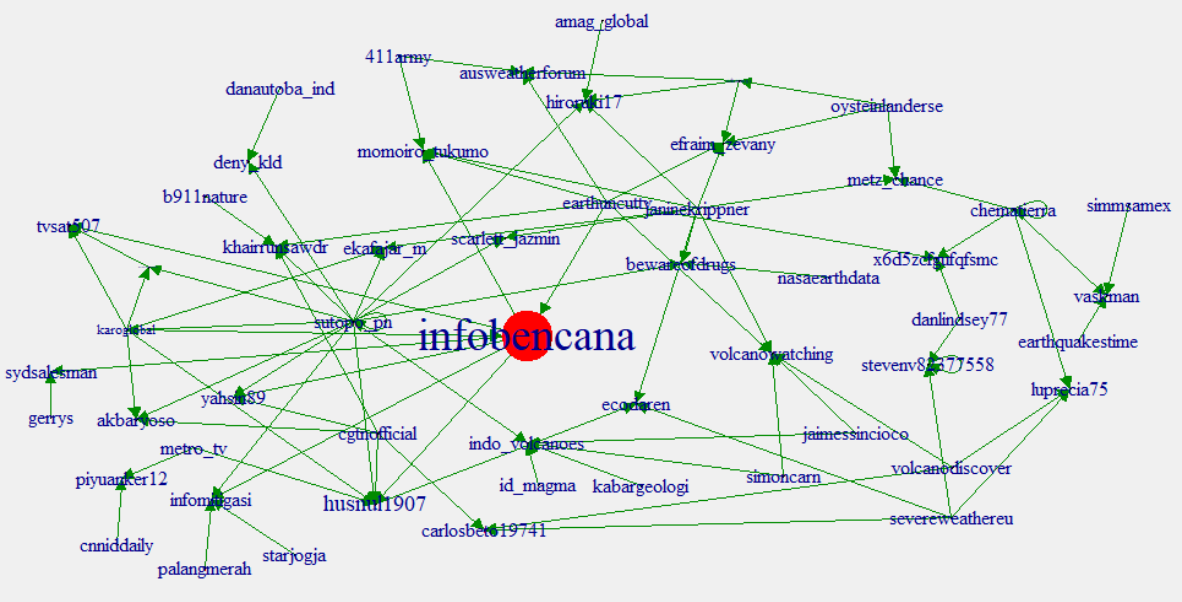


Figura 30. Principales nodos (usuarios) con mayor centralidad de Intermediación. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En la figura 31 se muestra la frecuencia de los datos expresado en un histograma de frecuencias, la mayor acumulación se encuentra sesgada a la izquierda; 1330 usuarios, es decir el 99.77% de los usuarios presentan menor centralidad de intermediación y solo 3 usuarios, es decir el 0.022% de usuarios, presentan mayor intermediación. Los usuarios *@infobencana,*

@Karoglobal, y *@ szharangi* estos son los más relevantes en la red destacándose por su centralidad mayor.

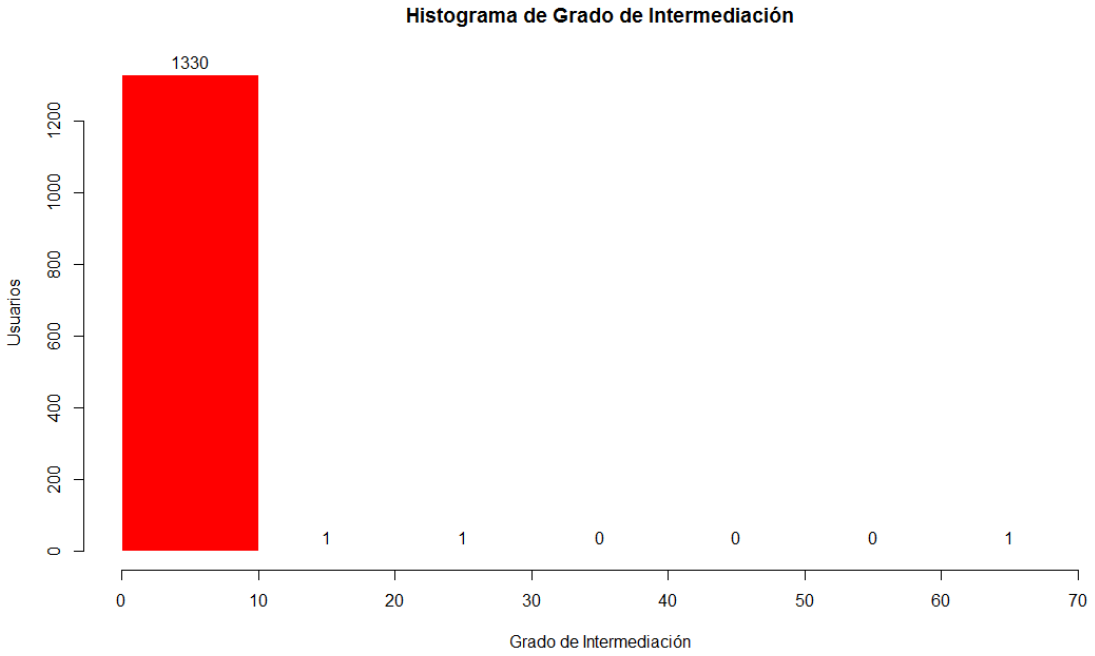


Figura 31. Histograma de grado de intermediación. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Centralidad de vector propio (eigenvector centrality): en la figura 32 se observa los nodos más relevantes en la red, los usuarios con altos puntajes de centralidad como *@sutopo_pn, @karoglobal y @janinekrippner* son usuarios que tienen mayores conexiones, que a su vez están bien conectados con otros usuarios, estos usuarios fueron los más influyentes en la red en cuanto a la difusión de información entre las comunidades. Los usuarios más centrales en este sentido corresponden a centros de grandes grupos cohesivos en la red.

En la figura 33 se muestra la frecuencia de los datos expresado en un histograma de frecuencias, la mayor acumulación se encuentra sesgada a la izquierda; 1332 usuarios, es decir el 99.92% de los usuarios presentan menor centralidad de vector propio mientras que el usuario

@sutopo_pn presenta un valor de centralidad de vector de 1, siendo el más relevante en el grafo.

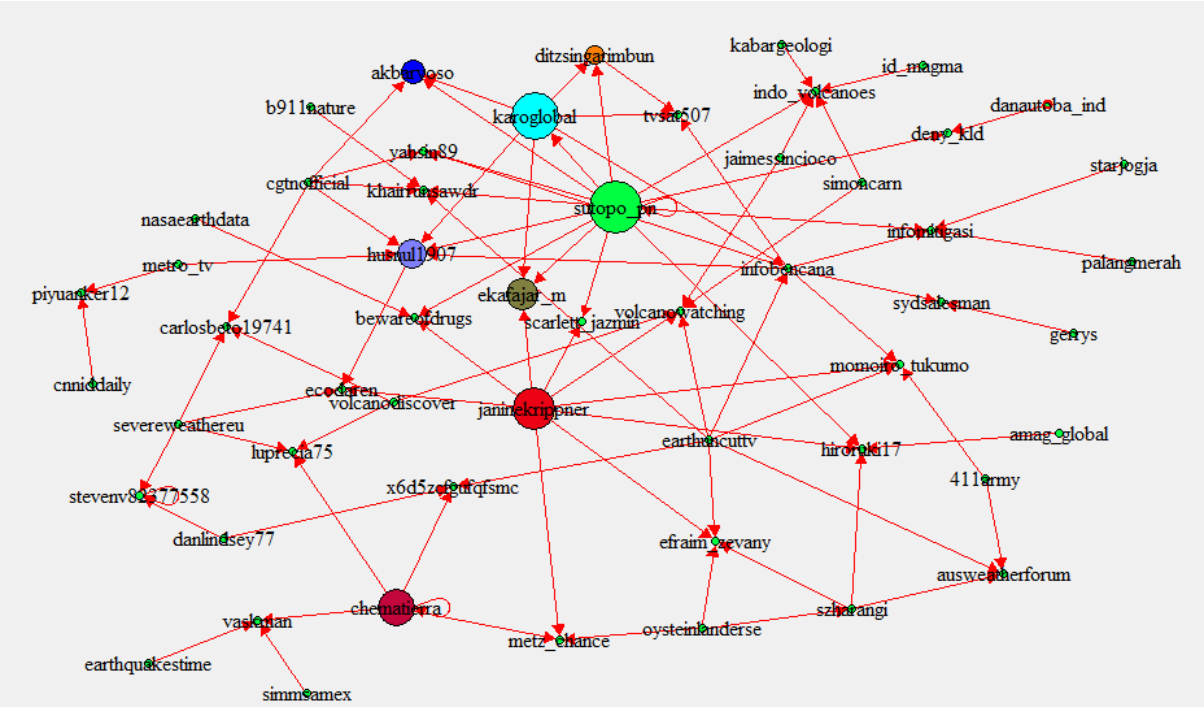


Figura 32. Principales nodos (usuarios) con mayor centralidad de vector propio (eigenvector centrality). Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

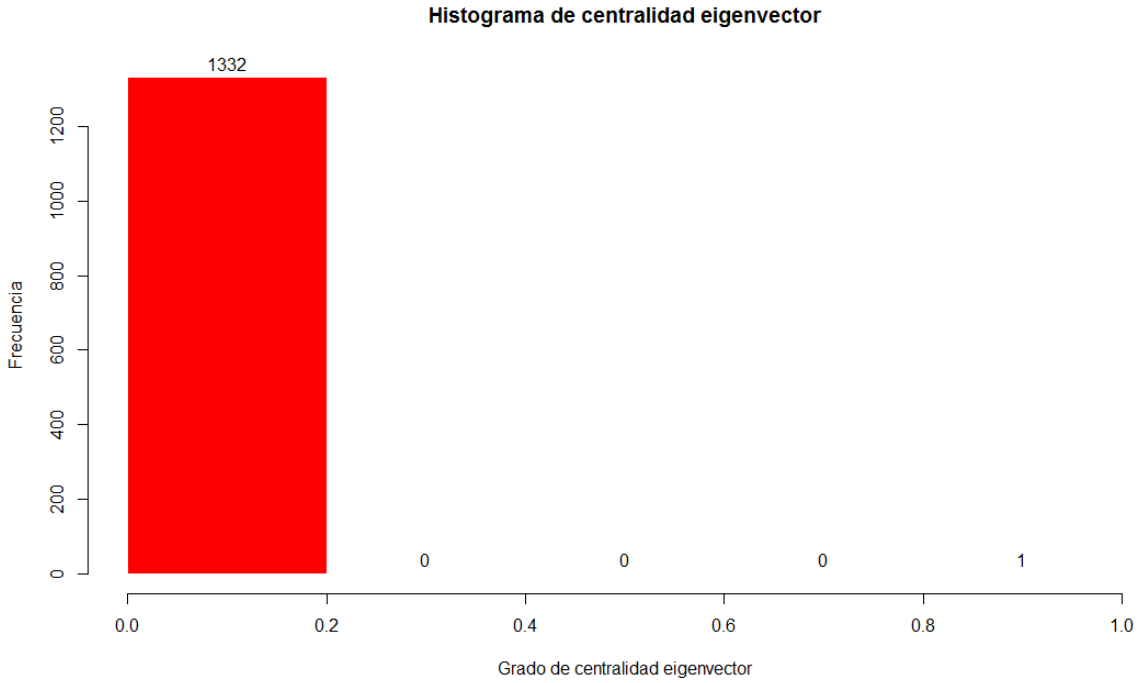


Figura 33. Histograma de centralidad de vector propio (eigenvector centrality). Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

8.4.Análisis de conglomerados en retweets

Se identificó cierta tendencia a formar clústeres en torno a usuarios en común. En la figura 34, donde se puede ver claramente las comunidades centrales donde están la mayoría de usuarios más relevantes. Este grafo está dividido principalmente en 8 comunidades, de las cuales se observa aquellos nodos *(usuarios)* principales en cada comunidad. La comunidad más numerosa está formada por el usuario *@chematierra (*clúster 5). Esta comunidad es un ejemplo de densa actividad de interacciones que corresponden al grado de entrada, es decir no establecen ninguna interacción entre ellos. Este clúster está relacionado con el clúster 6 que está formado por el usuario *@erthquakestime.* Los usuarios *@resilirel y @peroqueaweonao* tuitearon tuits generados por los usuarios de los clústeres 5 y 6, centrándose específicamente en publicaciones relacionadas con los geólogos que comprenden esas comunidades.

La comunidad formada entorno al clúster 2, es la segunda comunidad más numerosa en cuanto a usuarios que dieron retuits a las publicaciones realizadas por el geólogo *@sutopo_pn*, encontrándose relacionado con el clúster 3 que está conformado por el usuario *@janinekrippner*, usuario que se encargó de difundir información sobre vulcanología; los usuarios que generaron tuits tanto para el clúster 3 como para el clúster 2 fueron usuarios que pertenecían a cuentas relacionadas con medios de comunicación. *@ekafajar_m, @shemia y @khabirm,* fueron los más influyentes en cuanto a las publicaciones de los tuits transmitidos por los demás usuarios, ellos se encargaron de retransmitir la información relacionada con vulcanología y sismología generada por los usuarios de los clústeres 2 y 3.

La comunidad formada en el clúster 1 son usuarios que publicaron sus tuits relacionados con el desastre; el usuario más relevante en este clúster es *@infobencana* que es un usuario registrado como un centro de medios y habla en sus tweets sobre desastre; este clúster se caracteriza por la poca afluencia de usuarios que compartieron información sobre los tuits generados por el usuario, este clúster se encuentra relacionado con el clúster 2 por medio de los usuarios *@karoglobal y @infomitigasi*, que fueron cuentas de medios de comunicación encargados de difundir la información a sus seguidores suministrada por los usuarios *@infobencana y @sutopo:pn.* Los nodos conformados en el clúster 4 son usuarios que trataron temas relacionados con las publicaciones generadas por los usuarios que conforman los clústeres 3 y 5, es un clúster de baja densidad. Los clústeres 7 y 8 no tienen relación con los demás clústeres debido a su poca influencia en la red social.

Las comunidades de cada agrupación mantienen vínculos con otras comunidades por medio de usuarios que se encuentran relacionados con dos comunidades entre sí, la comunidad

que presenta mayor centralidad de grado, es aquella donde los usuarios hicieron más retuits a las publicaciones hechas por sismólogos o vulcanológicos, siendo así la comunidad en torno al nodo *@chematierra* uno de los usuarios más destacados en sus publicaciones.

Las interacciones entre los clústeres no siguen un patrón específico, no están relacionadas todas si no que están conectados directamente por pares, todo esto es por la poca densidad de relaciones entre los usuarios que representan intereses comunes muy específicos de información relacionada con el desastre, por ejemplo, alertas sobre la erupción, información actual del volcán y evacuación de instituciones educativas

Las comunidades de cada agrupación mantienen vínculos con otras comunidades por medio de usuarios que se encuentran relacionados con dos comunidades entre sí, la comunidad que presenta mayor centralidad de grado, es aquella donde los usuarios hicieron más retuits a las publicaciones hechas por sismólogos y vulcanológicos, siendo así la comunidad en torno al nodo *@chematierra* uno de los usuarios más destacados en sus publicaciones.

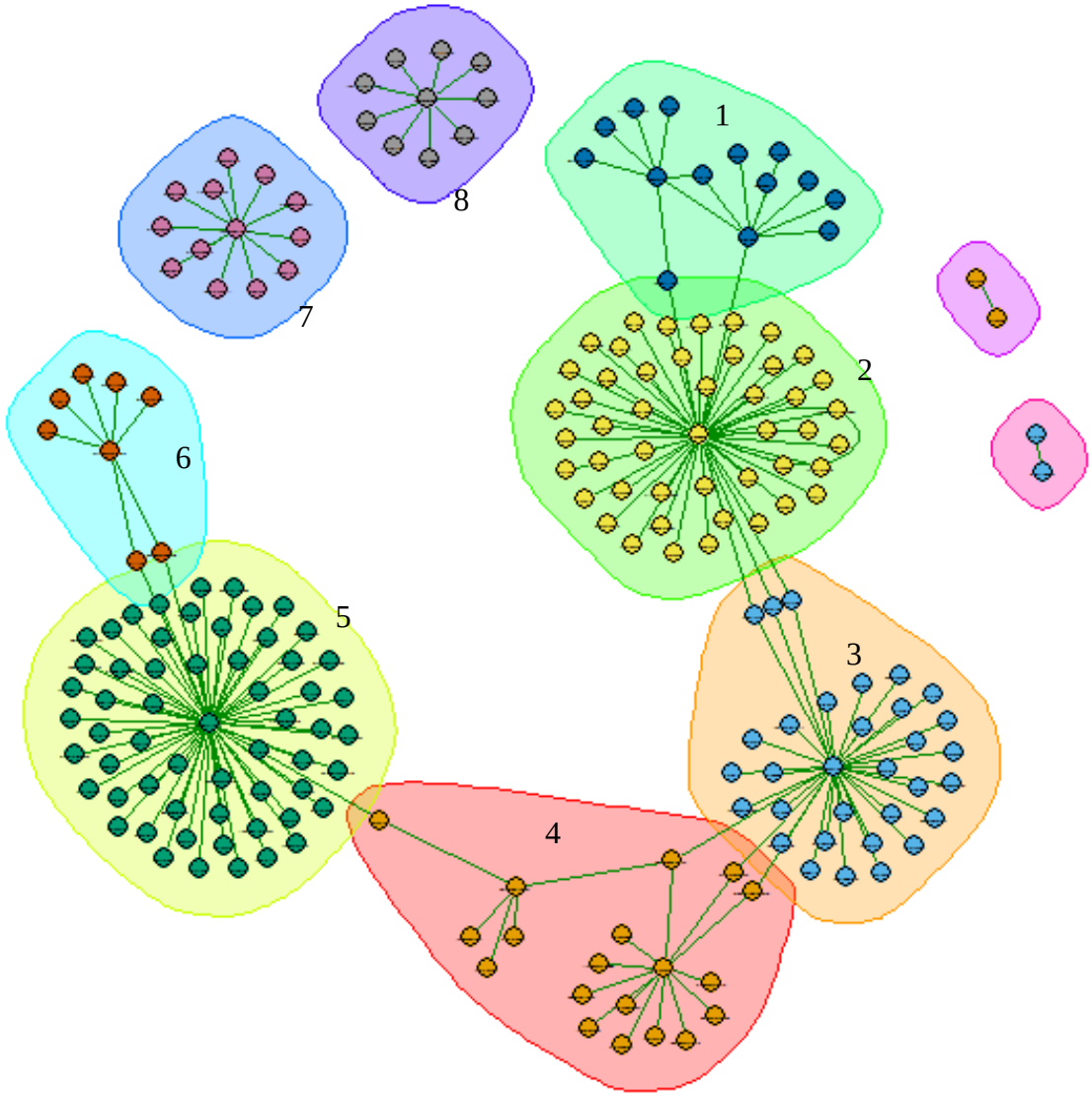


Figura 34. Conglomeración de comunidades en la red de retweets. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

9. Análisis de red de los Hashtags

9.1.Análisis de red método de visualización

En la figura 35, se muestra el diagrama general de red incluyendo 1428 enlaces y 939 nodos.

En este gráfico de red ofrece una vista general de las interacciones de los hashtags más

influyentes en los tweets; cada hashtag está representado por un nodo y un color que se asemeja a la comunidad de red que pertenece y los enlaces son las relaciones que tienen entre ellos en los tweets. Los hashtags más comunes en los tweets: *#sinabung*, *#volcano* y *#sumatra*, entre otros. La longitud más larga entre los nodos de la red es 8. El diámetro y las longitudes de camino en el grafo generado son relativamente pequeños respecto al número de nodos y enlaces de los mismos. Las relaciones entre los nodos están orientadas a una sola dirección, es por esto, la reciprocidad de la red es nula. La densidad del grafo es 0.00162128, este valor es bajo lo que convierte en una red de tipo disperso. En la red se forman clústeres de hashtags fuertemente conectados entre sí, esto es debido a las mayores relaciones que existen entre ellos.

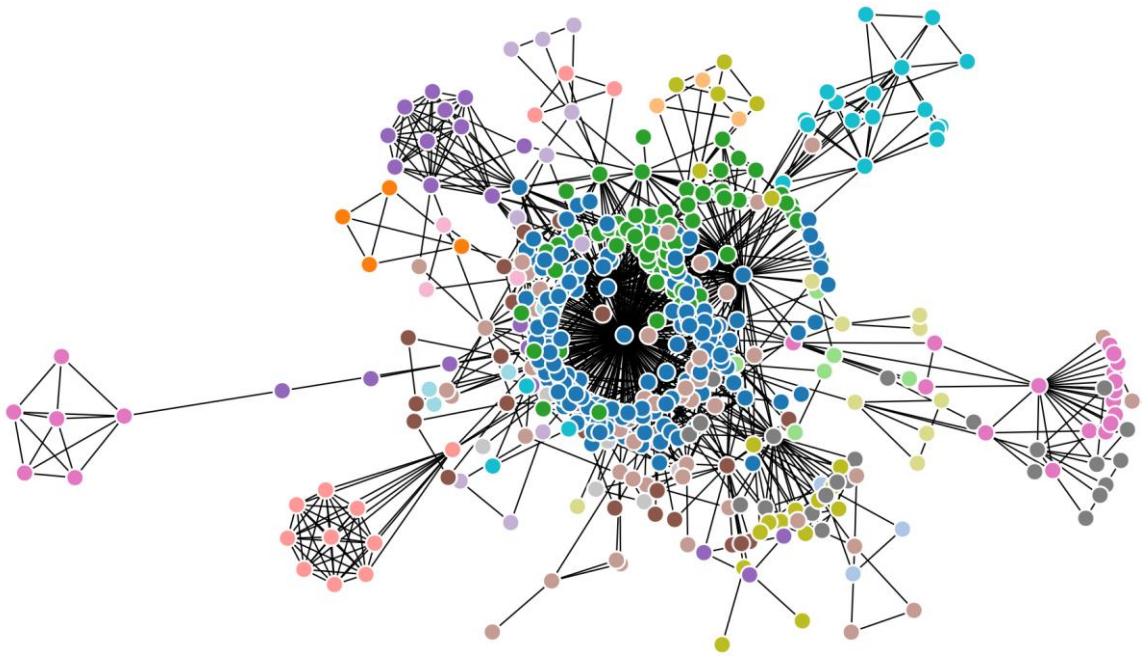


Figura 35. Estructura general de red de hashtags. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

9.2.Análisis de medidas de centralidad

En la figura 36 se visualiza el grafo completo donde se encuentran nodos con mayores relaciones, cada hashtag está representado por un nodo con un tamaño proporcional al número

de veces que se twitteó. El grupo formado en torno al hashtag *#sinabung*, constituye el centro del grafo principal. Es el nodo con más interacciones en el grafo, con un valor de centralidad de grado de 236. Siendo así, el principal hashtag que se utilizó para la búsqueda de los tweets. Otros hashtags más destacados por su centralidad de grado son: *#volcano* (92), *#indonesia* (81), *#eruption* (50) y *#sumatra* (48). Estos hashtags presentan relaciones entre ellos, ilustrando la frecuencia con la que se mencionaron conjuntamente en los tweets. Esta red destaca tanto los temas más populares, como la forma en que se relacionan entre sí.

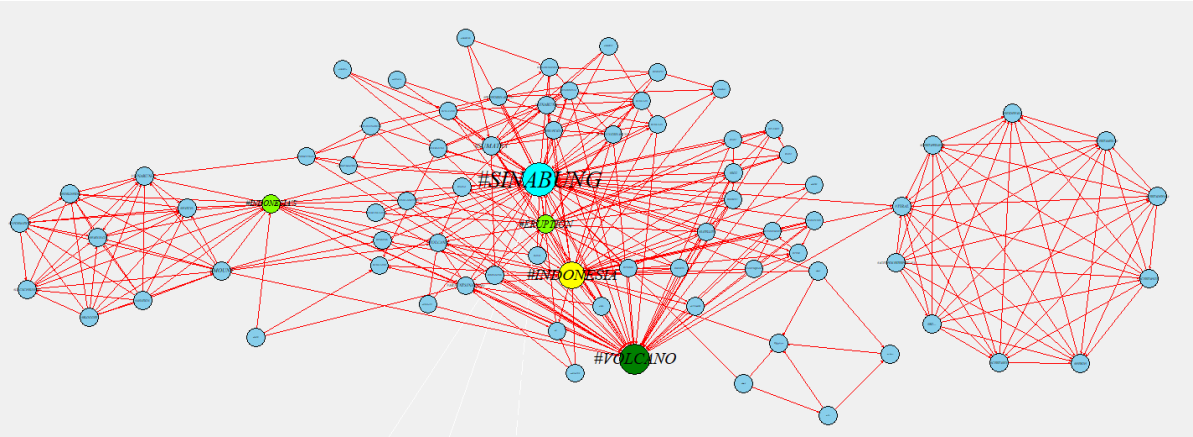


Figura 36. Estructura de grafo con mayor centralidad de grado y relaciones entre los nodos. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En la figura 37 se observa el grado de entrada de los principales hashtags con más relaciones destacados en los tweets. Se observa que el hashtag *#sinabung* con (177 relaciones) es el más influyente en la red; este hashtag fue uno de los influyentes que se tuvieron en cuenta para la descarga de los tweets.

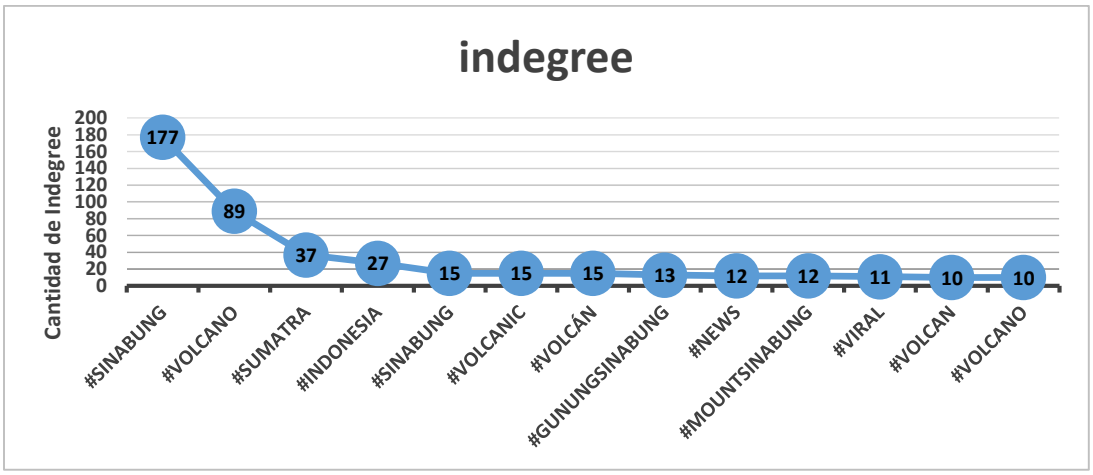


Figura 37.Frecuencia de los principales hashtags. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En la red de hashtag se puede ver que hay muchos grupos dispersos que se encuentran relacionados directamente. En la figura 38 se observa un conglomerado de 10 hashtags con mayor conectividad entre ellos, y a su vez, tiene la misma centralidad de grado 9, excepto *#viral* por su centralidad de grado de 12. Este último hashtag se encuentra relacionado con los nodos que tienen mayor centralidad de grado en la red. Su relación entre ellos es por la información que trataron los usuarios sobre el desastre en los tweets, estos contenidos giraron sobre la misma temática relacionados en torno al estado del volcán y su erupción transmitiendo noticias y contenidos como imágenes y videos.

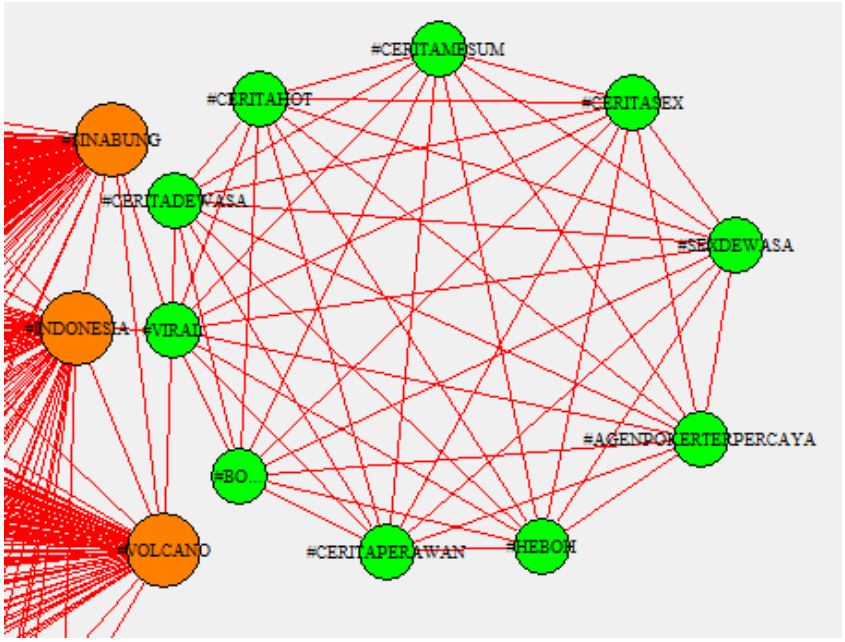


*Figura 38.*Sección del grafo con mayor conectividad entre los nodos. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

9.3.Medidas de centralidad global

Centralidad de cercanía En la figura 39 muestra una versión de la figura 35, detallándose más concretamente la medida de cercanía de los nodos en a la red. Esta medida no se puede aplicar a grafos con nodos desconectados, es por esto que se muestra un grafo donde todos los nodos se encuentran relacionados; Los colores utilizados en la red representan la centralidad de cercanía: amarillo una centralidad baja, verde una centralidad superior, azul indica la mayor centralidad.

En la figura se observa las flechas sin puntas, con el fin de mostrar la estructura de la red con más claridad. Se puede observar en el grafo que los nodos con mayores centralidades de cercanía se encuentran en el centro del grafo.

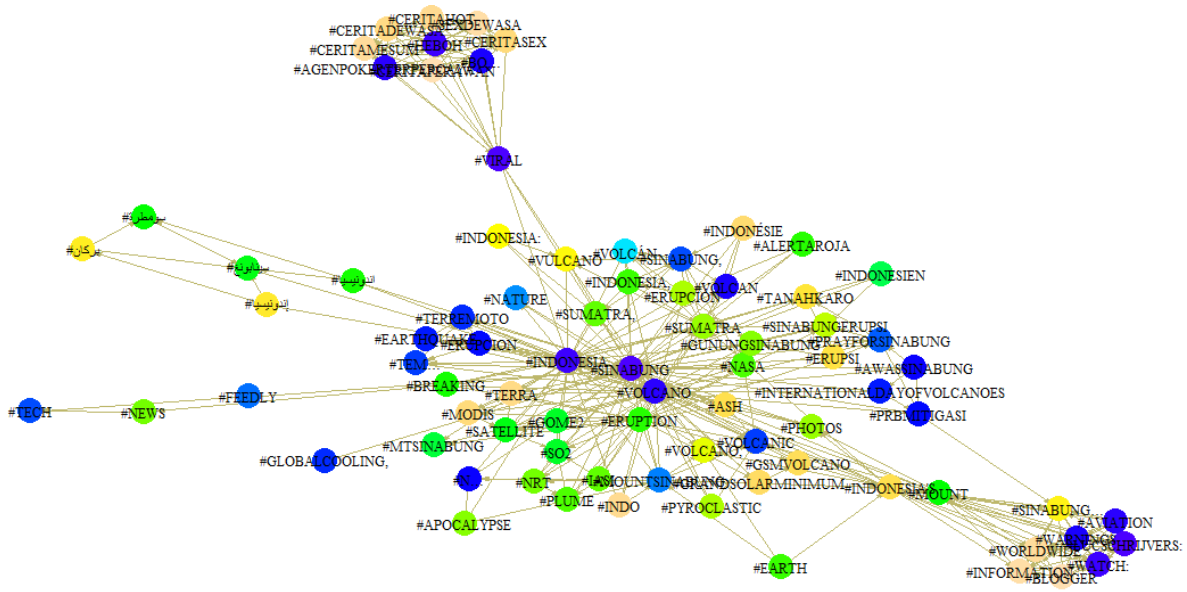


Figura 39. Estructura de grafo con centralidad de cercanía. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En el grafo se puede ver detalladamente los nodos lejanos en la red figura 40; estos nodos que presentan menos de cuatro conexiones en la red, están representados de color azul y a la vez con mayor tamaño de letra, permitiendo así la facilidad de localizarlos en la red. Se evidencia que los nodos más lejanos son los hashtags menos frecuentes en las publicaciones de los usuarios (fueron publicaciones menos importantes para los usuarios). Se puede observar en la red que los nodos con mayor centralidad de lejanía se encuentran localizados en la periferia de la red.

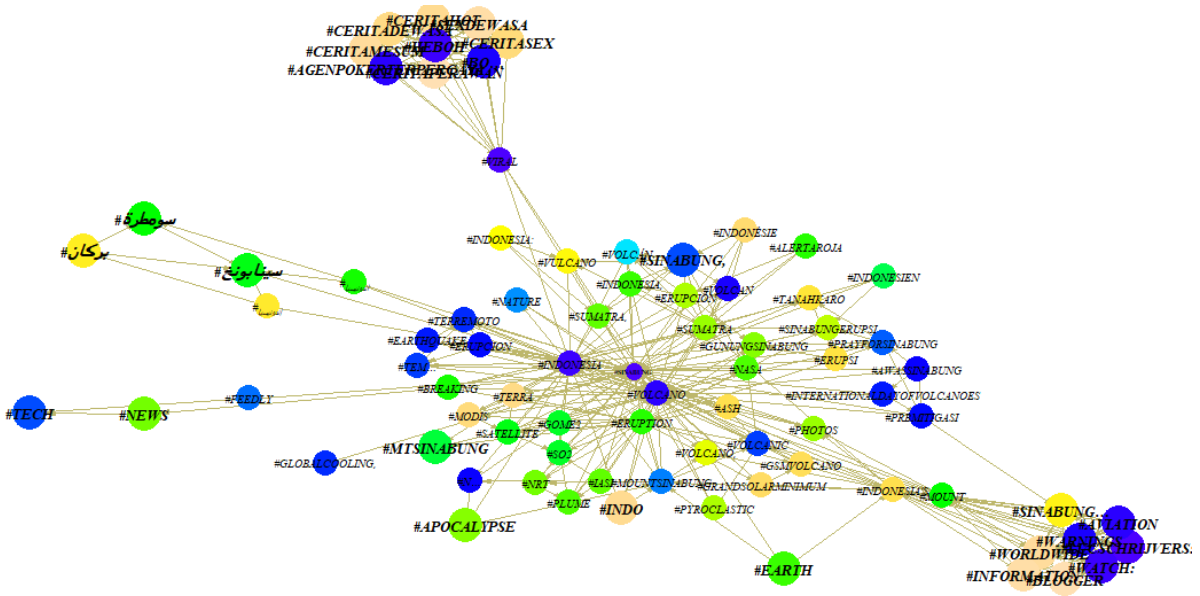


Figura 40. Estructura de grafo con centralidad de lejanía. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Centralidad de intermediación: En la figura 41, se muestra los nodos con mayor centralidad de intermediación (*betweenness centrality)* donde se observa el hashtag *#sinabung*, este nodo de color azul que representa un valor máximo de centralidad de intermediación, tiene más conexiones entre los nodos más cercanos de la red, es un nodo de alta interdependencia que actúa como puente a lo largo del camino más corto entre los nodos de la red que necesariamente están conectados a él.

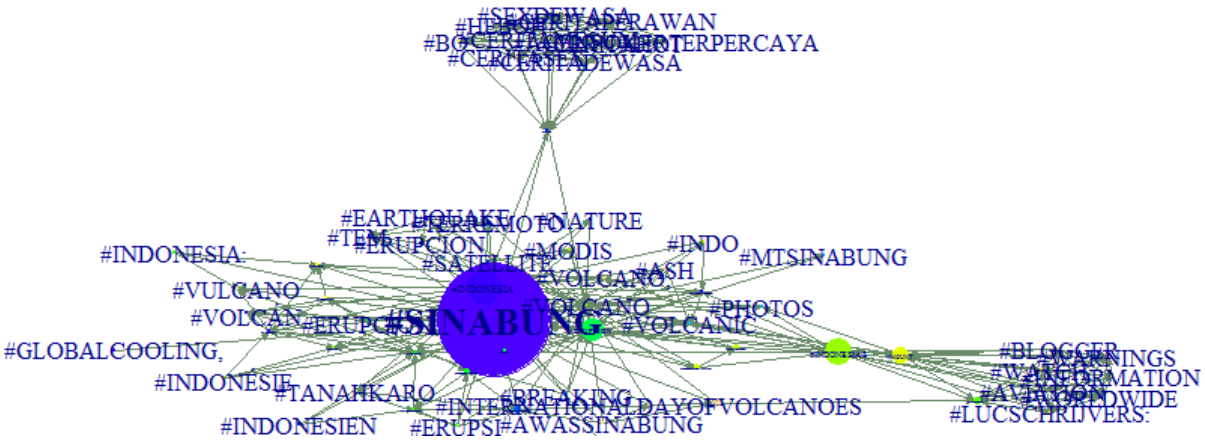


Figura 41. Estructura de grafo con centralidad de intermediación. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

9.4. Análisis de conglomerados en la red de hashtags

Mediante el análisis de red se identificó una estructura de comunidad. En la figura 42 se puede ver 21 comunidades centrales de los cuales se observan aquellos nodos (*hashtags*) más utilizados por los usuarios en sus publicaciones de los tweets en cada clúster, cada nodo representa un hashtag y su color indica el cluster al que pertenece. Los clústeres más numerosos están formados por los hashtags *#sinabung*, *#volcano*, *#indonesia*, *#eruption* y *#sumatra*; todos ellos constituyen el centro de la estructura por su centralidad de grado, son nodos de alta interdependencia que actúan como guardianes a lo largo de los caminos más cortos entre los nodos que están conectados a ellos. En la red se forman clustering de hashtags fuertemente conectados entre sí, esto es debido a las mayores relaciones que existen entre ellos.

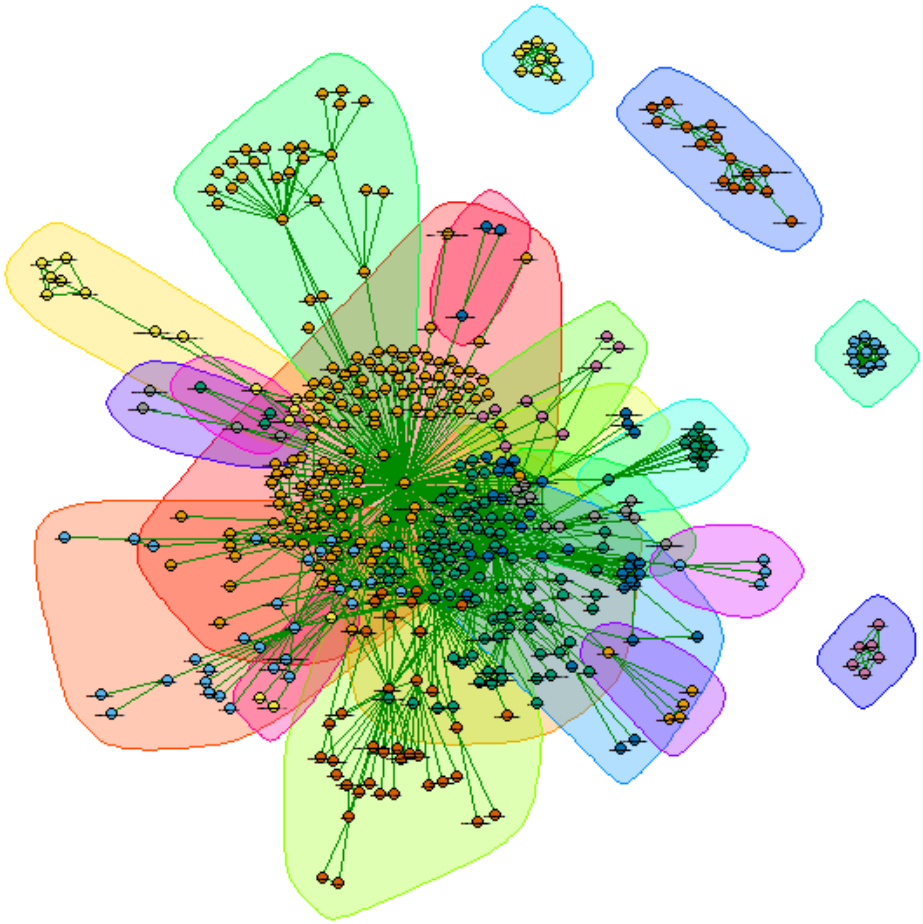


Figura 42. Conglomeración de comunidades en la red de hashtags. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

10. Análisis de red de las replicas

10.1. Análisis de red método de visualización

En la figura 43, se muestra un diagrama de red de réplicas generadas en los tweets que representan las conversaciones por los usuarios, incluyendo 125 enlaces y 200 nodos. Este grafo tiene una transitividad nula, que es la probabilidad de que dos nodos cercanos sean vecinos. La reciprocidad de la red es de 0.02040816, hay algunas relaciones orientadas en

ambos sentidos, es decir dos usuarios se replican conjuntamente; esto se observa en la red donde los usuarios *@hendrik_ps* y *@jaiswaa* se replican entre ellos las publicaciones.

En esta red se representan los usuarios más destacados en las publicaciones más influyentes durante la línea de tiempo en la cual se dio el desastre, los nodos son aquellos usuarios que publicaron tweets y las conexiones son las réplicas que realizaron a las publicaciones. Se ubicaron mayoritariamente en los nodos *@cayslue*, *@_leroy_pauline* y *@aron0076* a quienes se le respondieron y desarrollaron un diálogo, es decir, que el usuario principal no respondió a las réplicas generadas en sus tweets, en cuanto a la conectividad del grafo la mayor parte de los nodos tienen una baja conectividad y muy pocos nodos una alta conectividad, siendo una característica habitual en las redes libres. El diámetro de la red igual a dos saltos necesarios para ir de un nodo a otro, producido por una réplica del usuario *@mattyoung* a *@szharangi* la cual a su vez replicó *@internetremove*; por otro lado, se nota que varios usuarios se auto-replicaron con motivo de actualizar o aclarar sus propios tweets, sin embargo, estos no generaron una réplica con otros usuarios. La densidad de la red es de 0.003125 que es una medida de cohesión en la red, esta medida explica que el grafo es de tipo disperso.

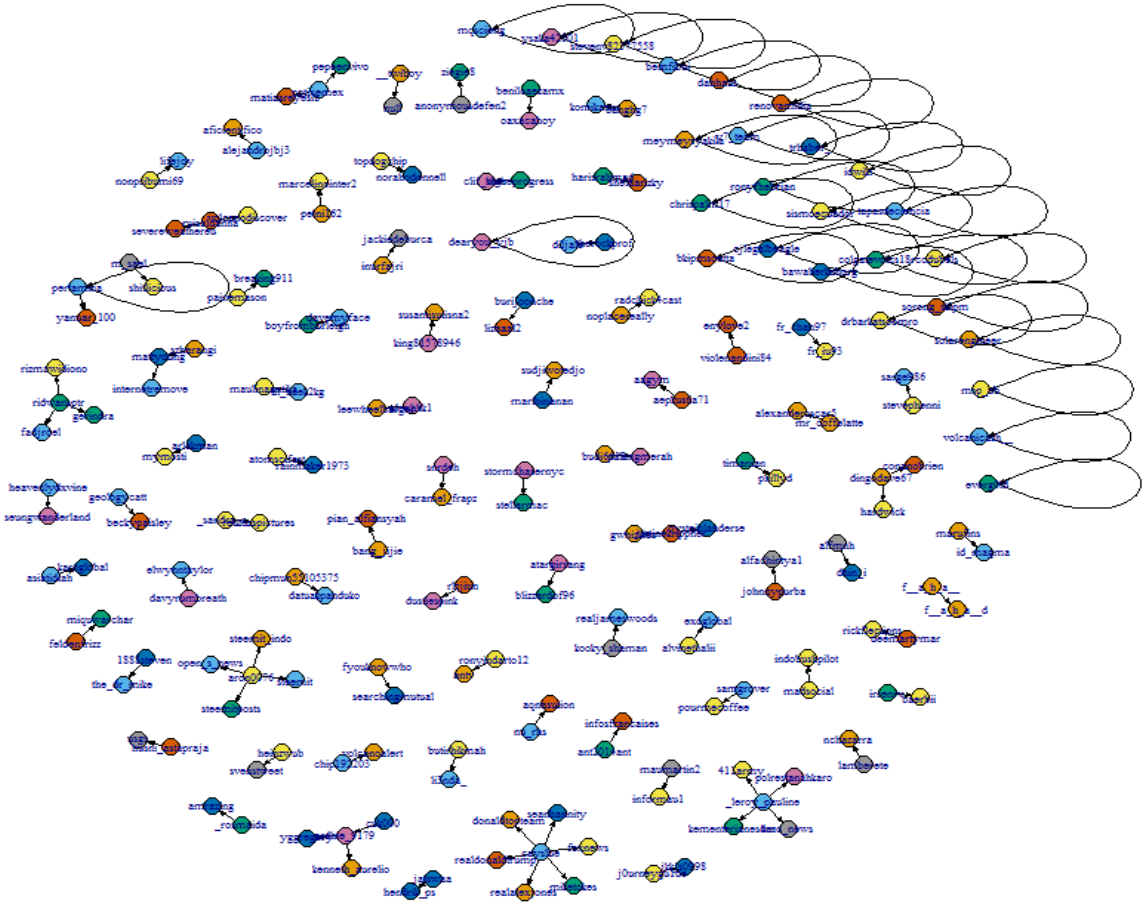


Figura 43. Grafo general de réplicas. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Como se aprecia en la figura 44 del histograma, se encontró que el 97% de los nodos, es decir, 194 vértices tienen una centralidad de grado igual a dos, 3 con centralidad tres, 2 con centralidad cuatro y uno con centralidad igual a seis, por lo que debido a esto se observa los clústeres que se formaron, figura 45, en donde se observa con más detalle la relación entre los usuarios que dieron réplicas y los que hicieron las publicaciones.

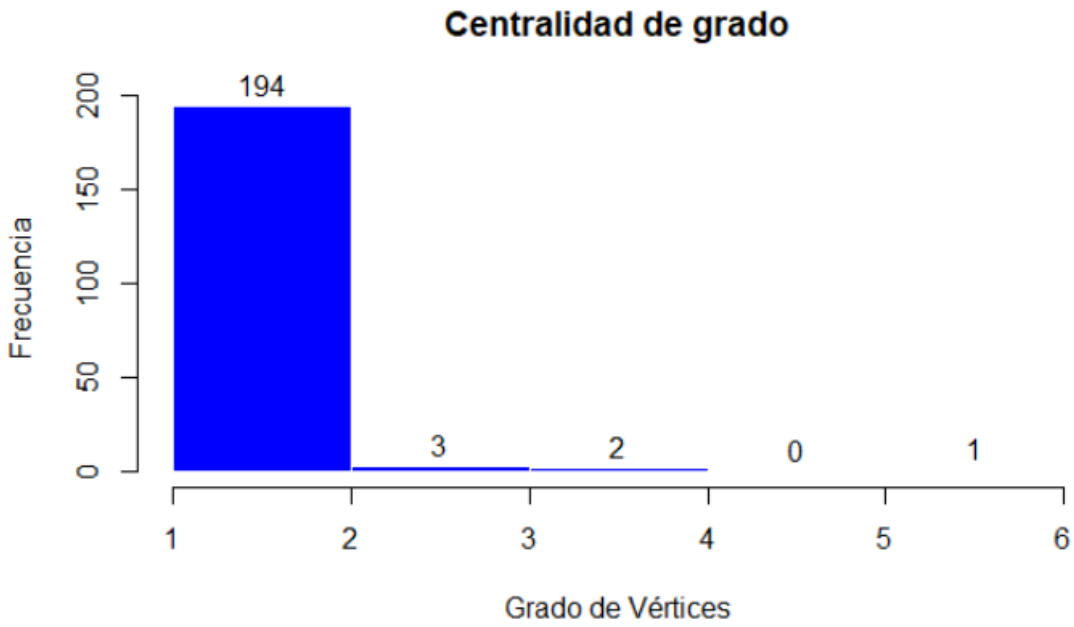


Figura 44. Distribución de centralidad de grado. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

10.2. Análisis de conglomerados en la red de replicas

En el conglomerado de la red de réplicas figura 45 podemos evidenciar los diferentes clústeres que se formaron; esta formación es debida a temas que compartieron en común entre usuarios, se ilustra una formación alrededor del usuario *@_leroy_pauline* que es un periodista francés para niños, los temas que compartió el usuario se relaciona con la búsqueda de niños durante las erupciones y cómo viven, compartió el mismo mensaje a usuarios como *@polrestanahkaro, @411Army, @karo_news* y a *@KementerianESDM.*

Se puede evidenciar la comunidad alrededor del usuario *@cayslue*, las publicaciones que compartió están relacionados con la información primaria sobre el volcán fue uno de los primeros tweets que trató sobre la erupción, en la publicación mencionó a usuarios como: *@donaldtopteam, @foxnews, @miketokes, @realalexjones, @realdonaltrump.*

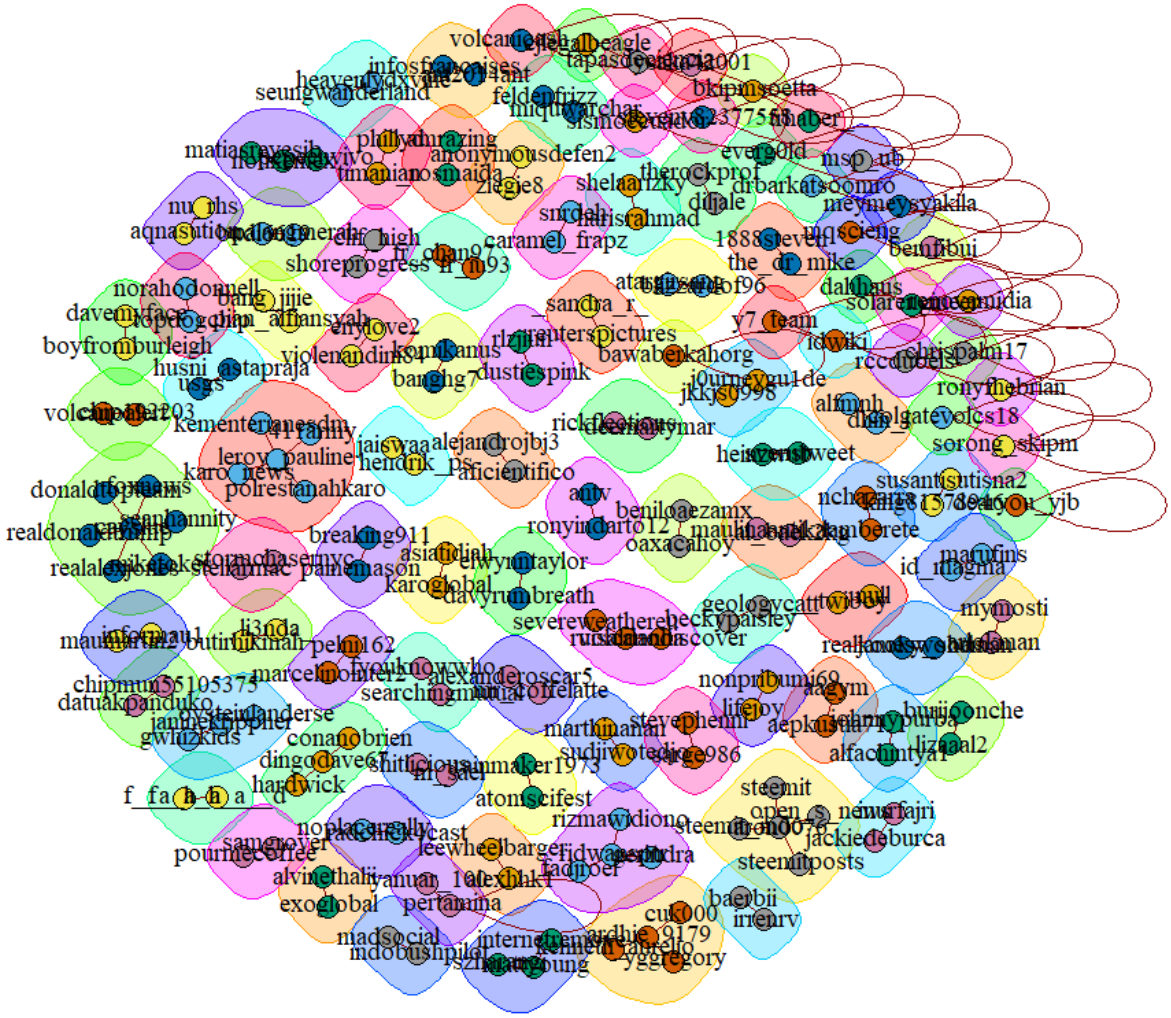


Figura 45. Conglomerado de la red de réplicas. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Otra comunidad relevante en el conglomerado alrededor del usuario *@aron0076*, este usuario compartió temas como "*la erupción ocurrió nuevamente por segunda vez en sinabung. Siendo la erupción más podadora que haya ocurrido en el volcán",* fotos, videos y noticias de interés sobre la erupción, hizo mención a usuarios como *@open_s_news, @steemit, @steemit_indo* y *@steemitpost.*

La comunidad formada por el usuario *@Ardhie_9179,* compartió noticias y videos sobre la erupción, en los tweets hizo mención a usuarios como periodistas y vulcanológicos como *@kenneth_autrelio,* y *@yggregory.* La comunidad formada entorno al usuario *@ridwansptr,* compartió tweets sobre el desacuerdo que tiene las entidades de rescate al no actuar en forma inmediata en el lugar de los hechos, en el tweet menciona a usuarios como *@fadjroel* investigador en temas relacionados con la vulcanología, *@gerindra* su nombre oficial Partai Gerindra página oficial de indonesia y *@RizmaWidiono*

11. Análisis de red de Urls

11.1. Análisis de red método de visualización

En la figura 46 se muestra el diagrama general de red de urls que se utilizaron en los tweets; en este se representan con nodos los usuarios tanto como las urls por lo que se incluyen un total de 920 vértices, por otro lado, sus interacciones se representan con aristas las cuales incluyen un total de 1105 enlaces, debido a que los usuarios se pueden vincular a varias urls, pero estas últimas no se pueden vincular entre sí; ocurre que la reciprocidad del grafo es nula y que la longitud media del camino es de 1, de usuario a url. El diámetro del grafo representa la mayor distancia entre los nodos que se puede encontrar y es de 10, por lo que es de esperar que los usuarios no solo se referencian en una misma página web sino en varias. La densidad del grafo es de 0.000754147 que es significativamente baja ya que no todos los nodos se relacionan entre sí dando como resultado un grafo disperso y por último el grado medio es de 2 por lo que las relaciones de conexiones fueron mayores a 2.

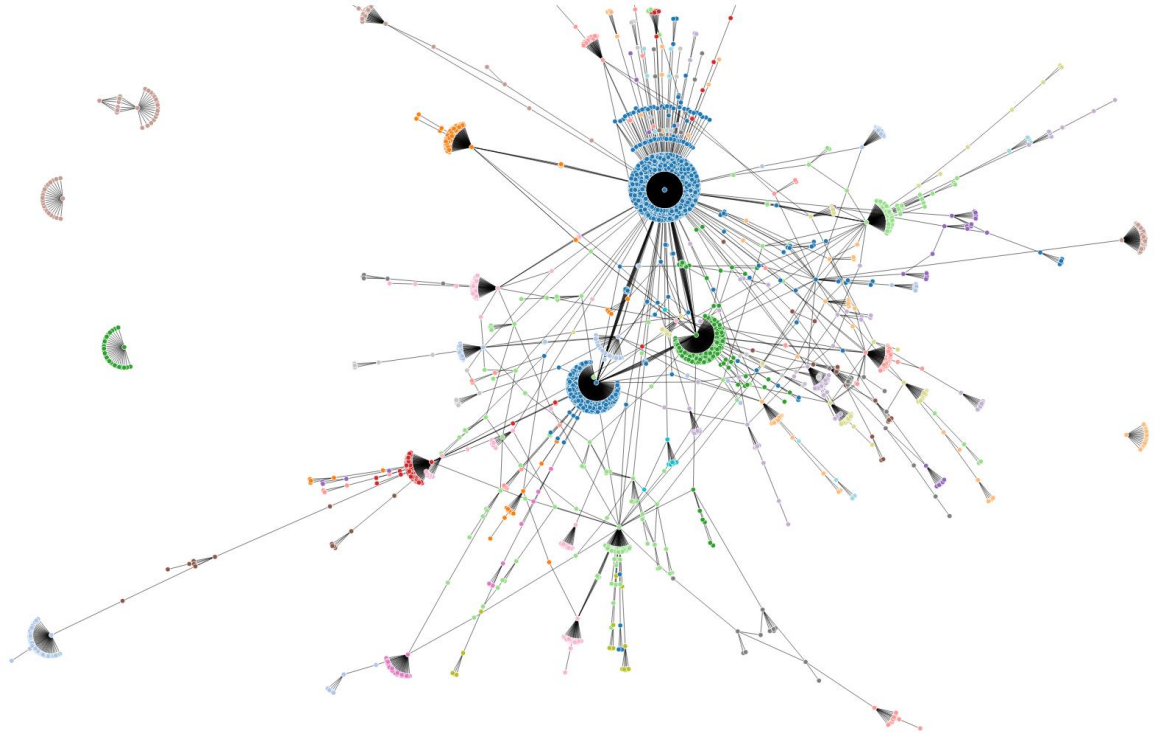


Figura 46. Estructura general de red de urls. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Tabla 5.
Principales enlaces con mayor centralidad de grado.

Fuente	Centralidad de grado (*degree*)
Twitter.com	1598
www.youtube.com	477
Bit.ly	118
Fb.me *(Facebook)*	389
www.instagram.com	100
Feeds.feedburner.com	89
Gizmodo.com	64
Mashable.com	56
News.dekit.com	55
www.viv.co.id	49
www.theverge.com	38

En la tabla 5, se enumeran 11 enlaces principales en los mensajes de Twitter durante la erupción del volcán Sinabung. Se puede observar que la fuente de mayor participación por los

usuarios es Twitter.com, ya que 1598 usuarios compartieron sus tuits referenciados por la misma página.

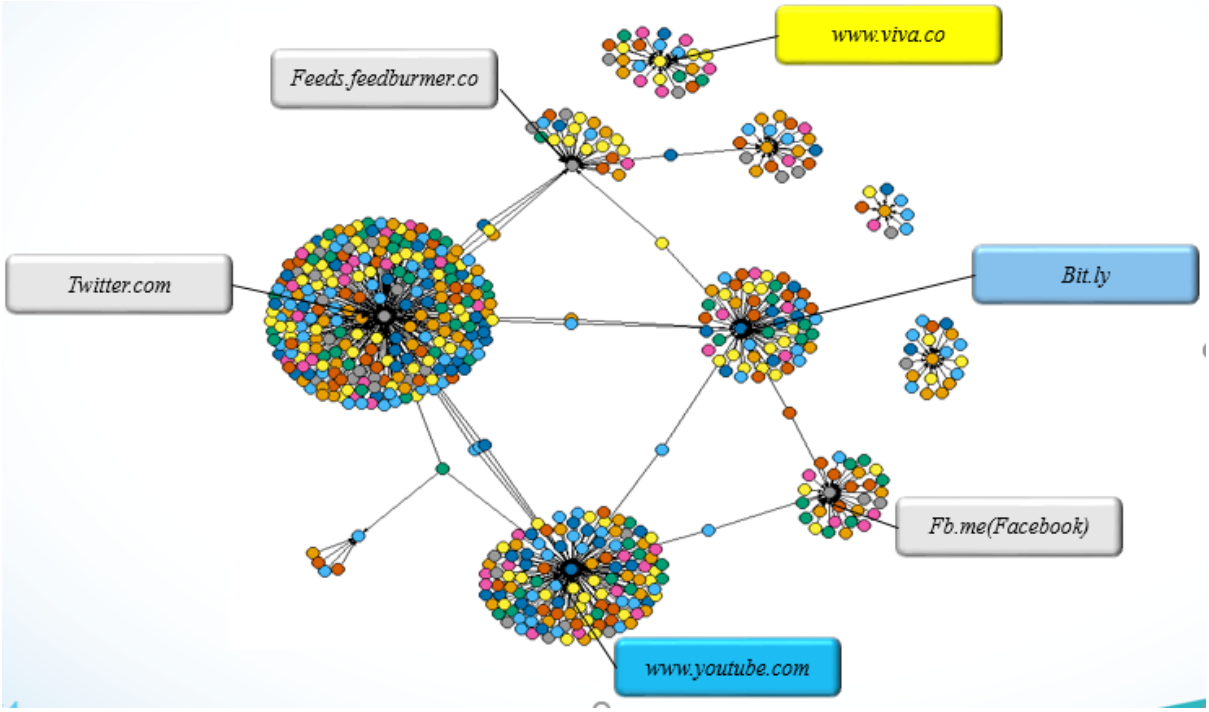


Figura 47. Estructura general de red de urls más importantes. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En la figura 47 se visualiza con mayor detalle la estructura general de red de las urls en donde se puede apreciar las páginas web más compartidas por los usuarios; *twitter* aparece como uno de los favoritos ya que en muchos tuits se referenciaba contenido alojado por la misma página; por otro lado, era de esperar urls que redirigiera a páginas de información como noticieros, como lo son los enlaces de "*gonews.co*" o *"Aceh.tribunnews.com",* por otro lado mucho del contenido publicado contenía videos que mostraban de una forma más cerca la situación del evento, es por esto que *"youtube*" es representativo al ser un popular sitio web dedicado a compartir vídeos; sin embargo, se aprecian otras urls que comparten videos como *"VIVA.co.id"* (anteriormente VIVAnews) el cual es un portal de noticias en línea , por último

es importante mencionar la utilización "*Feedburner*" y páginas similares las cuales son gestores que ofrecen entre otras cosas la posibilidad de crear una lista de suscriptores para que reciban por *email* todo tipo de actualizaciones de las redes sociales.

11.2. Análisis de medidas de centralidad

Centralidad de grado: En la figura 48 se encuentra representado la centralidad de grado de las *urls* extraídas de los tweets, en la cual se puede apreciar las relaciones entre los usuarios y las páginas web conjuntamente con la dirección de las aristas que ahora son visibles, por lo que cabe mencionar que la interacción de las urls entre sí está intermediada por los usuarios que compartieron sus contenidos en varias urls. Por otro lado, la distribución de grado de los nodos se encuentra sesgada a la derecha. Representan mayores nodos que presentan menor centralidad de grado figura 49, 1102 nodos (urls y usuarios) tuvieron una centralidad de grado de 50, y tres representativas tuvieron centralidad de grado de 100, 150 y 250 las cuales fueron "*twitter.com", "youtube.com" y "feefs.fedburner.com"* respectivamente.

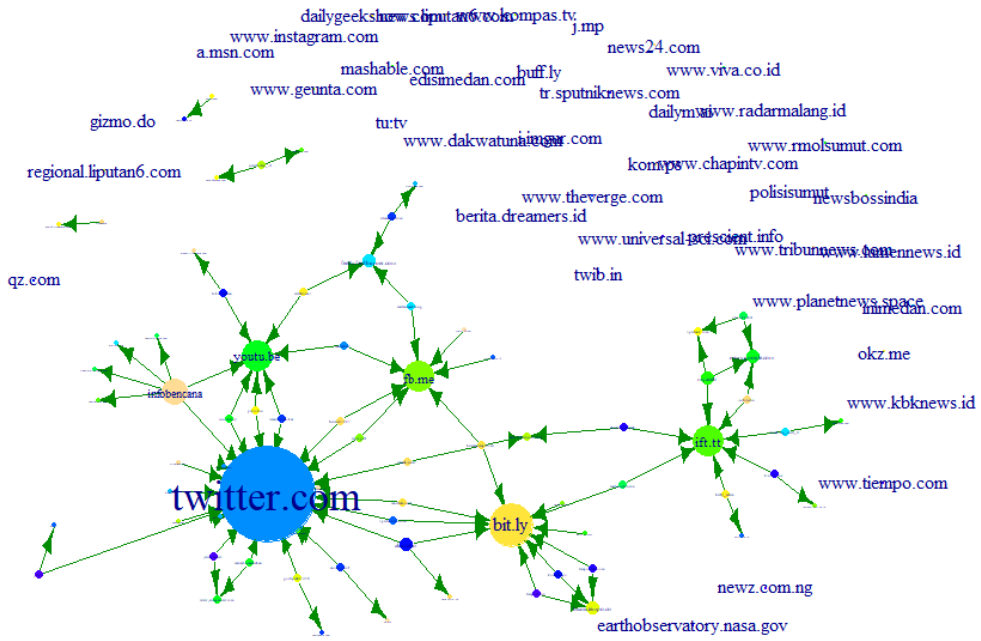


Figura 48. Principales nodos (urls) con mayor centralidad de grado. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

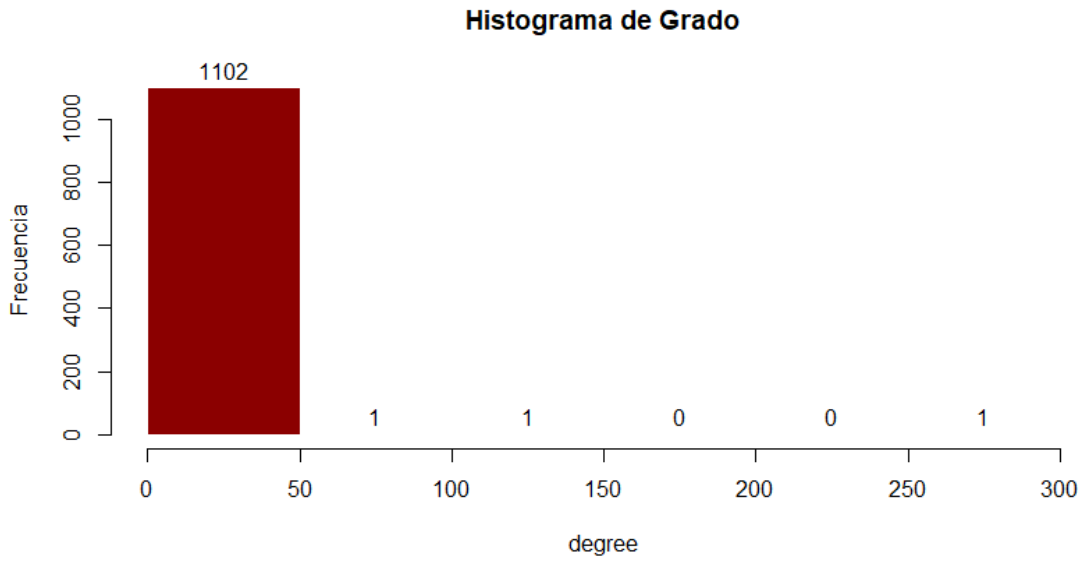


Figura 49. Histograma de centralidad de grado. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

Al ver sólo una sección de la red figura 50, se nota la formación de un grupo alrededor del nodo *twitter.com,* donde se puede ver el número de usuarios que tuitearon en esta página. Los usuarios *@chematierra, @infobencana* y *@sutopo_pn* registrados como centros de medio y portadores de voz en medios de comunicación, presentan contenidos en sus tuits relacionados con noticias de última hora sobre el sismo, tuitearon opiniones de vulcanólogos.

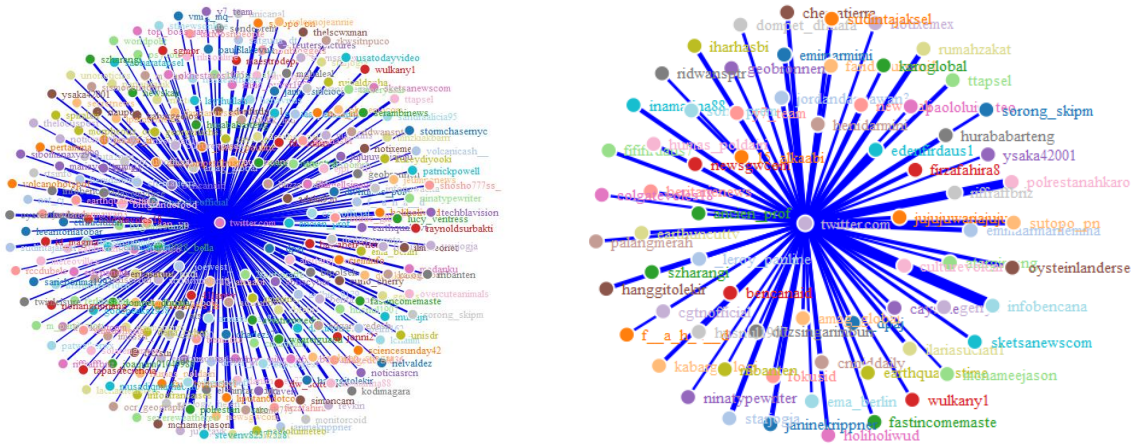


Figura 50. Sección de grafo general de mayor centralidad de grado. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

11.3. Medidas de centralidad global

Centralidad de cercanía En la figura 51 muestra un subgrafo de la red general figura 46, detallándose más concretamente la medida de cercanía de los nodos en la red, esta medida no se puede aplicar a grafos con nodos desconectados, es por esto que se muestra un grafo donde todos los nodos se encuentran relacionados son los que presentan mayor centralidad de cercanía, los colores utilizados en la red representan más detalladamente la centralidad, amarillo una centralidad baja, verde una centralidad superior, azul indica la mayor centralidad. En la figura se observa las conexiones con una dirección, con el fin de mostrar la estructura de la red con más claridad, especificando la relación usuarios-urls. Por otro lado, la distribución de grado de los nodos figura 52, se encontró el 68.87% de los nodos, es decir, 3849 nodos (urls y usuarios) representan una centralidad mayor en la red, el 31.13% de los nodos, 1740 nodos tienen centralidad menor en la red. Entre los nodos que representan una centralidad mayor encontramos a *twitter.com, Facebook, youtube.com* y el usuario *@infobencana.*

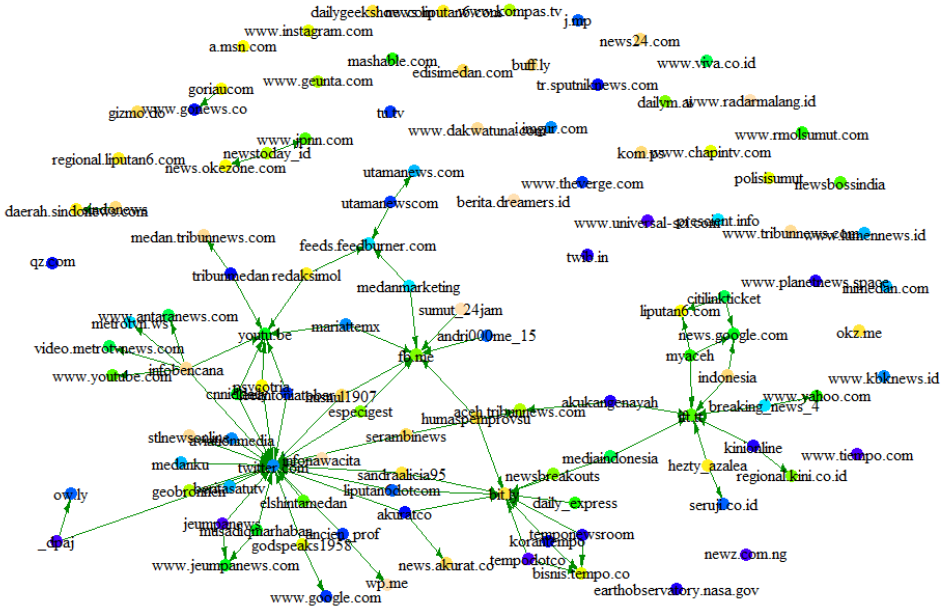


Figura 51. Principales nodos con mayor centralidad de cercanía. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

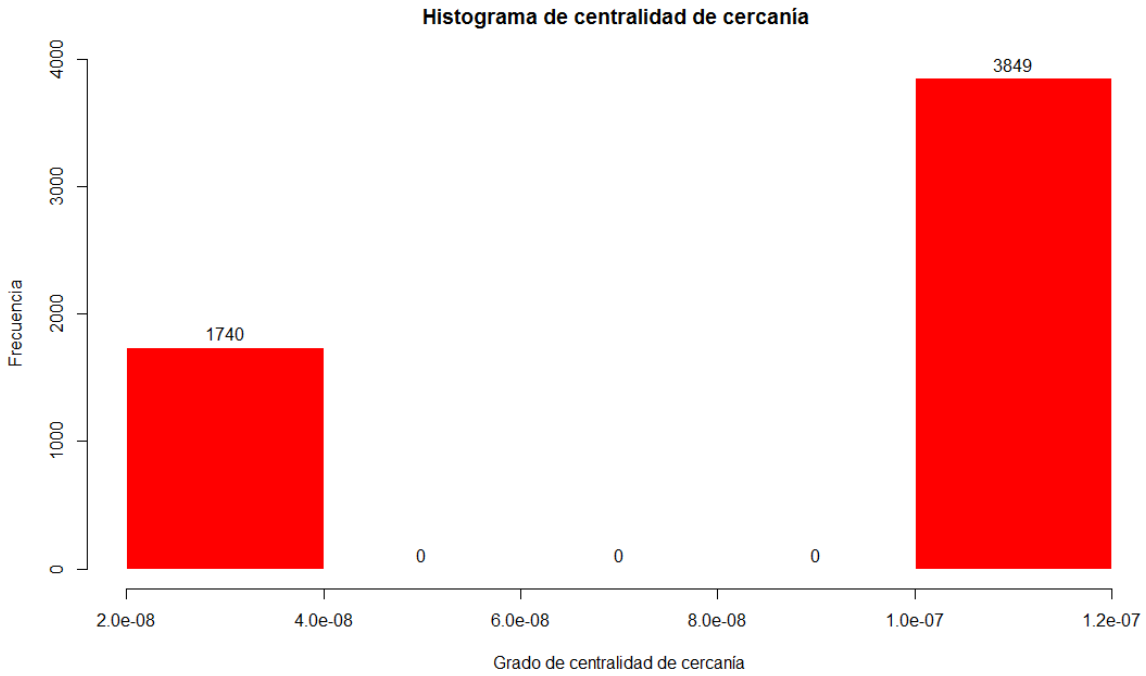


Figura 52. Histograma de centralidad de cercanía. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

En el grafo se puede ver detalladamente que hay muchos nodos lejanos en la red figura 53, estos nodos que presentan menos de dos conexiones en la red, están representados de color azul y a la vez con mayor tamaño de letra, permitiendo así la facilidad de localizarlos en el grafo. Se evidencia que los nodos más lejanos son las urls menos usadas por los usuarios.

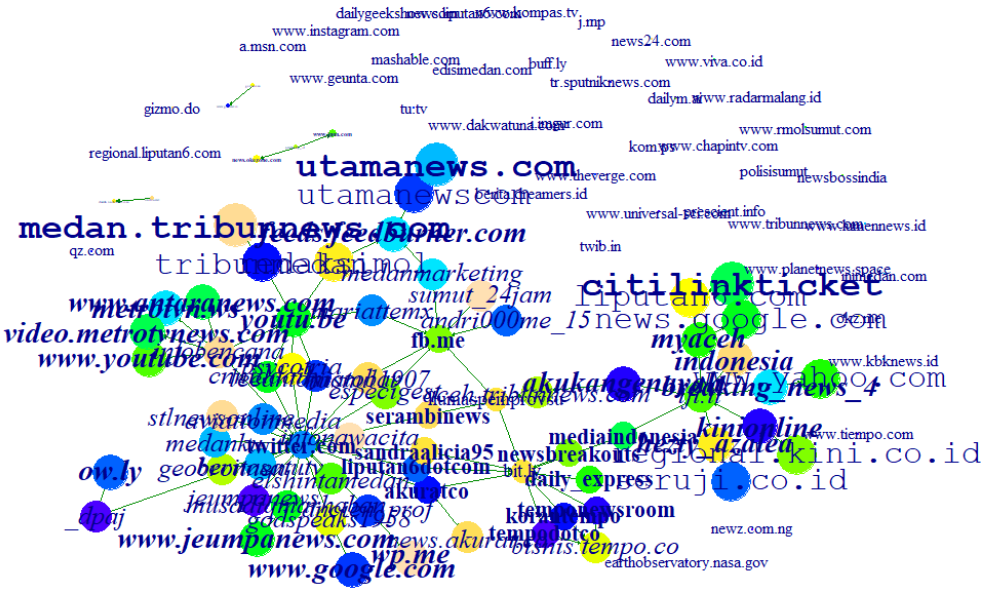


Figura 53. Principales nodos con mayor centralidad de lejanía. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

11.4. Análisis de conglomerados en urls

Se identificó cierta tendencia a formar clústeres en torno a los nodos en común. En la figura 54, donde se puede ver claramente las comunidades centrales donde están la mayoría de las urls con los usuarios más relevantes, este conglomerado está dividido principalmente en 10 comunidades, de las cuales se observa aquellos nodos (*urls y usuarios*) principales en cada clúster. El clúster más numeroso (1) formado por la fuente *twitter.com,* siendo el nodo con mayor centralidad de grado; éste se relaciona con los clústeres formados por las fuentes de *Facebook.com,* 5 *youtube.com* 3 *y bit.ly* 4, estas fuentes fueron las más utilizadas por los principales usuarios que publicaron sus tuits; estas comunidades publicaron información relevante sobre la erupción volcanica encontrándose relacionadas entre sí por su participación

en la red social Twitter. Los usuarios que hicieron sus publicaciones en las fuentes son cuentas relacionadas con sismología, vulcanología y medios de comunicación. Las demás comunidades formadas por los clústeres 6 y 7 constan de usuarios que publicaron sus tuits en las fuentes *ift.tt y news.google.com*, estos grupos se caracterizan por poca influencia de usuarios que compartieron sus publicaciones en las urls.

Las comunidades de cada agrupación mantienen vínculos con otras comunidades de la red, pero presentan mayor conectividad entre ellas, como es el grupo uno formado por la fuente *twitter.com*, este nodo presenta mayor grado de centralidad por la mayor cobertura de usuarios como *@infobencana* que en su mayoría de tuits estuvieron relacionadas con noticias sobre el evento.

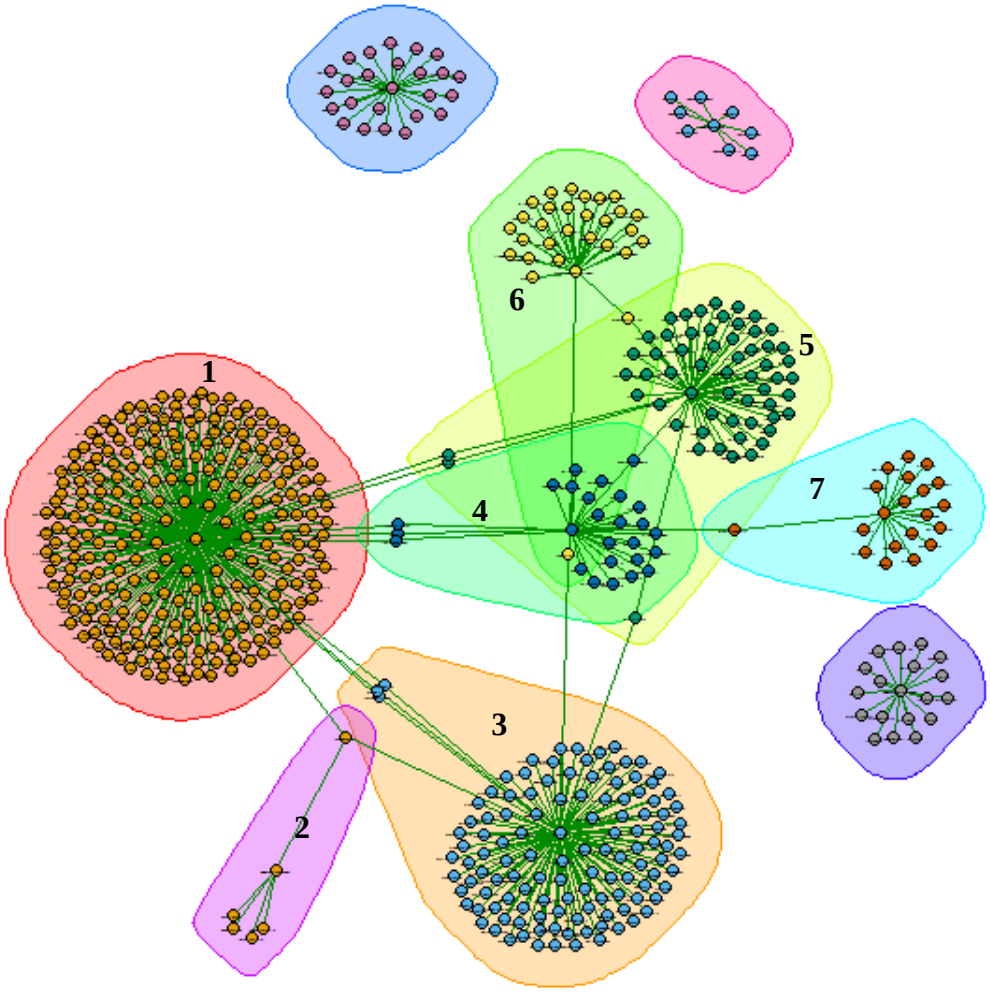


Figura 54. Conglomeración de comunidades en la red de urls. Extraído de software para análisis estadístico R-Studio 2018 versión 3.4.2. Disponible en https://github.com/INDUSTRIALOPALO/TEORIA-DE-GRAFOS-EN-REDES-SOCIALES

12. Análisis generales de las métricas de los grafos

Una vez obtenidas todas las métricas de las diferentes redes en el conjunto de datos, se analizan de forma global comparándolas entre sí. Para facilitar la comprobación de las métricas, a continuación, en la tabla 6 se puede ver a modo resumen las métricas vistas para una mejor comparación.

Como complemento del análisis topológico de las diferentes redes anteriores, la tabla 6 muestra algunas características básicas de las redes. Los datos utilizados en las redes anteriores indican que el diámetro generado en los grafos es relativamente pequeño respecto al número de nodos y enlaces de los mismos. Este fenómeno se conoce comúnmente como red de mundo pequeño. Este tipo de red es aquel en el que no todos los nodos son vecinos entre sí, pero en el cual se puede llegar a un nodo desde cualquier nodo a través de un corto número de saltos. Este fenómeno se observa en los diferentes grafos.

Tabla 6.

Comparación de las métricas extraídas de los distintos grafos analizados

Relación	**Re tuits**	**Hashtags**	**Replicas**	**urls**
Nodos	282	939	200	920
Enlaces	264	1428	125	1105
Diámetro	13	8	2	10
Densidad	0.003319753	0.001621287	0.003125	0.000754147
Reciprocidad	0	0	0.02040816	0
Transitividad	0.075%	10.68%	0	0
Longitud de camino	10	10	10	10
Longitud media de camino	1	2	1	1
Grado medio	2	3	1	2

En cuanto a la densidad de los grafos, tal como se vio en el análisis de cada red, los grafos obtenidos tienen valores muy bajos, lo que los convierte en grafos de tipo disperso. El grafo de los retweets es una red completa de máxima densidad (todos los nodos están conectados entre sí). En cambio, las redes de réplicas, hashtags y urls tienen una densidad relativamente baja. Por otra parte, los valores para la reciprocidad se encuentran muy pequeños en todos los tipos

de grafos generados. Esto ocurre a causa de la configuración de la red. Las relaciones de los grafos de retuits, hashtags y urls suelen estar orientadas en una sola dirección, es decir, un usuario puede hacer a otro retuits, mención y publicaciones. En la red de réplicas hay algunas relaciones orientadas en ambos sentidos, en decir, los usuarios se replicaron a la vez.

Para finalizar la comparativa del conjunto de datos, la red de hashtags se diferencia mucho en el promedio de grado siendo significativamente alto, en comparación con el resto de los grafos con un valor de 3 enlaces por nodo y transitividad de 10.68%. Del resto se diferencia por el bajo coeficiente de densidad obtenido, con apenas 0.001621287, siendo así la red dispersa. El diámetro es bajo comparándolo con las redes de retuits y urls, con tan sólo 8 saltos requeridos.

Los gráficos de red que se muestran en las secciones 8, 9, 10 y 11 permitieron explorar el conjunto de datos a gran escala. La mayoría de los clústeres tenían topologías de árboles. Se identificó que cada comunidad se conectaba directamente con otras comunidades por medio de actores que intervienen entre ellas. Comprender estas interacciones entre las comunidades ayudará a las agencias de emergencia a detectar centros críticos y utilizarlos para distribuir rápidamente cualquier mensaje urgente.

13. Conclusiones

En esta investigación, se examinó la difusión de información de emergencia a través de noticias, agencias meteorológicas, organizaciones y el público, usando múltiples métodos de

minería de datos. Estos resultados proporcionan información sobre las agencias y organizaciones de emergencia con el fin de comprender las características de la red social Twitter en línea y su participación.

El método empleado resultó acertado para estudiar este fenómeno; su carácter iterativo posibilitó el desarrollo de esta investigación. Esta forma de trabajo promovió una identificación más completa de requerimientos, así como la producción de códigos enfocados a resolver las necesidades del análisis. En consecuencia, se considera que la investigación puede replicarse o servir como guía en otras investigaciones similares.

Sobre las aplicaciones de la teoría de grafos en el análisis de redes sociales, se encuentran grandes contribuciones que no solo se enfocan en la red social twitter, páginas como Facebook e Instagram también son analizadas por teoría de grafos para definir comportamientos y gustos con un sin fin de métodos variantes que pueden involucrar, por ejemplo, software de código abierto y programación para obtener los datos e información necesaria para poder emplear la teoría de grafos.

Construir una base de datos sobre la red social ante un evento catastrófico requiere estar constantemente informado, saber que palabras claves se deben buscar para que sea relevante la muestra. Además, obtener los datos en el menor tiempo posible se vuelve indispensable debido a la naturaleza del evento; por otro lado, la base de datos debe estar actualizándose o definir un periodo de tiempo para obtener los datos dependiendo de la velocidad en la que se requieran los análisis.

Twitter durante los desastres se utiliza para transmitir información por lo general de segunda mano; esto tiene relación con los tweets procedentes de otros usuarios ya sean de cuentas de noticieros o información de sitios web que reportan el comportamiento de las entidades de rescate y el seguimiento de alertas.

Los usuarios utilizan la red principalmente para transmitir datos descriptivos sobre otras categorías que comúnmente lo identifican, como una plataforma para expresar opinión, para llamar a la acción social y como una red para la manifestación de emociones que surgen de la interacción, en etapas reactivas de las personas, en situaciones de emergencia y desastres.

14. Recomendaciones

Como se ha comentado en la introducción, existe un gran interés en el procesamiento de datos suministrados por las redes sociales, y en particular en Twitter, por lo que se podría seguir trabajando en la búsqueda de otros posibles usos prácticos de las redes sociales.

Se necesita un estudio futuro para comprender las características y la efectividad de los diferentes medios sociales, incluyendo Facebook, Google+, YouTube e Instagram en respuesta a desastres, como su viabilidad y confiabilidad como difusores de información en caso de emergencia. La mayoría de las preguntas podrían responderse mediante un enfoque de estudio de casos múltiples que compararía el uso y la efectividad de las redes sociales en una amplia gama de desastres. También se necesita otro estudio futuro para nodos aislados donde la información podría no ser alcanzada a través de una red en línea existente

Referencias Bibliográficas

Aldrich, D. P. (Marzo de 2011). The power of people: social capital's role in recovery from the 1995 Kobe earthquake. *Natural Hazards, 56,* 595-611. doi:https://doi.org/10.1007/s11069-010-9577-7

Ayala P., T. (2014). Redes sociales, poder y participación ciudadana. *Austral de Ciencias Sociales*(26), 23-45. Obtenido de http://www.redalyc.org/articulo.oa?id=45931862002

Bavelas, A. (1948). A mathematical model for group structures. *Human organization, 7*(3), 16-25. Obtenido de https://doi.org/10.17730/humo.7.3.f4033344851gl053

Blashfield RK, A. M. (Julio de 1978). The Literature On Cluster Analysis. *Multivariate Behavioral Research, 13,* 271-295. doi:10.1207/s15327906mbr1303_2

Bruno Takahashi, E. C.-P. (Julio de 2016). Barreras en la comunicación durante momentos de crisis: Lecciones de tres estudios sobre el tifón Hiyan en la Filipinas. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones", 10*(2), 5-6. doi:http://dx.doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.4730

C.Freeman, L. (1979). Centrality in social networks conceptual clarification. *Social Networks,* 215-239. Obtenido de www.sciencedirect.com/science/article/pii/0378873378900217?via%3Dihub#!

Caicedo Barrero , A., Wagner de garcía, G., & Méndez Parra, R. M. (2010). *Introducción a la Teoría de Grafos.* Armenia , Quindio: EDICIONES ELIZCOM.

Caragea , C., Silvescu, A., & Tapia, A. H. (Mayo de 2016). Identifying informative messages in disaster events using convolutional neural networks. *International Conference on Information Systems for Crisis Response and Management* , 137-140.

Cheng, T., & Wicks, T. (3 de Junio de 2014). Event Detection using Twitter: A Spatio-Temporal Approach. *Journal Citation Reports, 9*(6). Obtenido de https://doi.org/10.1371/journal.pone.0097807

Coppola, D. P. (2006). *Introduction to International Disaster Management.* United States of America. doi:https://doi.org/10.1016/C2014-0-00128-1

David Sanderson, A. S. (2016). *Word Disasters Report.* Swaziland: International Federation of Red Cross and Red Crescent Societies. Obtenido de http://www.ifrc.org: http://www.ifrc.org/Global/Publications/disasters/WDR/69001-WDR2005-english-LR.pdf

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 122-135. Obtenido de https://search.proquest.com/docview/1774841295?accountid=29068

De Nooy, W. M. (2006). Exploratory Social Network Analysis with pajek. *Cambridge University Press.*

Diestel, R. (2005). *Graph theory (Graduate texts in mathematics)* (Vol. 173). New York: Springer.

Duncan J, W., & Steven H, S. (04 de Junio de 1998). Collective dynamics of 'small-world' networks. *Nature*, 440-442. doi:10.1038 / 30918

Elisa Schaeffer, S. (2007). Graph clustering. *Computer Science Review*, 27-64. Obtenido de https://doi.org/10.1016/j.cosrev.2007.05.001

Godehardt, E. (1988). Graph-Theoretic Methods of Cluster Analysis. En *Graphs as Structural Models* (págs. 75-96). Wiesbaden. Obtenido de http://link.springer.com/10.1007/978-3-322-96310-9

Grimes, S. (8 de Agosto de 2013). Big Data: Avoid 'Wanna V' Confusion. *Informationweek - Online*. Obtenido de https://bibliotecavirtual.uis.edu.co:2171/docview/1428641245?accountid=29068

Hanneman, R., & Riddle, M. (2005). *Introduction to Social Network Methods.* California . Obtenido de http://www.faculty.ucr.edu/~hanneman/nettext/

Harary, F. (1969). *Graph Theory.* Canada: Addison wesley.

Houston J. Brian, J. H. (2015). Social media and disasters: a functional framework for social media use in disaster planning, response, and research. *Disaters , 39*(1), 1-20. Obtenido de https://doi.org/10.1111/disa.12092

IBM . (s.f.). Obtenido de International Business Machines : https://www-01.ibm.com/software/in/data/bigdata/

James Manyika, M. C. (Mayo de 2011). Big Data: The next Frontier for Innovation, Competition and productivity. *McKinsey Global Institute*. Obtenido de https://iapp.org/media/pdf/knowledge_center/MGI_big_data_full_report-1.pdf

Jooho, K., Juhee , B., & Makarand, H. (Junio de 2018). Emergency information diffusion on online social media during storm Cindy in U.S. *International Journal of Information Management, 40,* 153-165. Obtenido de https://doi.org/10.1016/j.ijinfomgt.2018.02.003

Kathleen M. Carley, M. M. (13 de Abril de 2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. *safety science,* 48-60. Obtenido de https://doi.org/10.1016/j.ssci.2016.04.002

Kelm B, M., McCallum, A., & Pal, C. (2006). *papers Combining Generative and Discriminative Methods for Pixel Classification with Multi-Conditional Learning.* Obtenido de College of Information and Computer Sciences: http://people.cs.umass.edu/~mccallum/papers/mclearn-icpr06.pdf

Kolman, B., & Hill , D. (2013). *Álgebra Lineal. Fundamentos y aplicaciones.* Bogotá: Pearson.

Kryvasheyeu, Y. C. (11 de Marzo de 2016). Rapid assessment of disaster damage using social media activity. *Science Advances.*

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Hentenryck, P. V., Fowler, J., & Cebrian, M. (11 de Marzo de 2016). Rapid assessment of disaster damage using social media activity. *Science Advances*.

Landwehr, P. M. (2014). Social Media in Disaster Relief. En W. W. Chu, *Data Mining and Knowledge Discovery for Big Data* (págs. 225-257).

Landwehr, P. M., Wei, W., Kowalchuck, M., & Carley, K. M. (Diciembre de 2016). Using tweets to support disaster planning, warning and response. *Safety Science*, 33-47. Obtenido de https://doi.org/10.1016/j.ssci.2016.04.012

Linton C, F. (1978). Centrality in social networks Conceptual clarification. *Social Networks, 1*. Obtenido de https://doi.org/10.1016/0378-8733(78)90021-7

Lu, X., & Brelsford, C. (2014). Network structure and community evolution on Twitter: Human behavior change in response to the 2011 Japanese earthquake and tsunami. *Scientific Reports, 4*, 1-12. Obtenido de http://dx.doi.org/10.1038/srep06773

Márcia Oliveira, J. G. (17 de Febrero de 2012). An overview of social network analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 90-105. doi:https://doi.org/10.1002/widm.1048

Matiyasevich, Y. V. (2004). One Probabilistic Equivalent of the Four Color Conjecture. *Theory of Probability & Its Applications*, 368–372. Obtenido de http://bibliotecavirtual.uis.edu.co:2311/doi/10.1137/S0040585X97980476

Michael Schroeck, R. S.-M. (2012). Analytics: The real-world use of big data. *Shanghai Jiaotong University (Science)*, 1-20. Obtenido de https://doi.org/10.1007/s12204-016-1714-3

Mohar, B. (2008). Part V: Theorems and Problems - 12. The Four-Color Theorem. *Princeton: Princeton University Press. Retrieved from*. Obtenido de https://bibliotecavirtual.uis.edu.co:2171/docview/189253682?accountid=29068

Núñez Valdés, J., Alfonso Pérez, M., Bueno Guillén, S., & Diánez del Valle, M. d. (2004). Siete puentes, un camino: Königsberg. *Suma*, 70-75. Obtenido de http://hdl.handle.net/11441/45044

Osmar, Z. (1998). Mining Multimedia Data. *Meeting Of Minds*.

Peter M. Landwehr, W. W. (Diciembre de 2016). Using tweets to support disaster planning, warning and response. *Safety Science*, 33-47. Obtenido de https://doi.org/10.1016/j.ssci.2016.04.012

Qiuju Luo, D. Z. (Febrero de 2015). Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tourism Management, 46*, 274-280. Obtenido de https://doi.org/10.1016/j.tourman.2014.07.007

Reda Alhajj, J. R. (05 de octubre de 2014). *Encyclopedia of Social Network Analysis and Mining* (Vol. 2). Springer Reference. doi: https://doi.org/10.1007/978-1-4614-6170-8_79

Repiso Caballero, R., Torres Salinas, D., & Delgado Lópes, E. (2011). :una aproximación a través del Análisis Bibliométrico y de Redes Sociales de las tesis doctorales defendidas en España entre 1976-2008. 420-421.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*. doi:https://doi.org/10.1007/BF02289527

Savita Kumari, S., & Pratibha, Y. (2017). Threats Detection using Big Data Analytics. *International Journal of Advanced Research in Computer Science, 8*(1), 126-128. doi:https://doi.org/10.26483/ijarcs.v8i1.2864

Scott, J. (2012). Social Network Analysis and Mining. *Computational Complexity. Springer*. doi:https://bibliotecavirtual.uis.edu.co:2236/ 10.1007 / 978-1-4614-1800-9_178

Shimbel, A. (1953). Structural parameters of communication networks. *The bulletin of mathematical biophysics, 15*, 501-507. doi:https://doi.org/10.1007/BF02476438

Smith, M., Rainie, L., Shneiderman, B., & Himelboim, I. (2014). Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters. *Pew Research Center, 20*, 1-56. Obtenido de http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/

Takahashi, B., Tandoc, E., & Carmichael, C. (Septiembre de 2015). Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Computers in Human Behavior, 50*, 392-398. Obtenido de https://doi.org/10.1016/j.chb.2015.04.020

Thales Botelho de Sousa, C. E. (2014). *An overview of the advanced planning and scheduling systems*. Obtenido de https://dialnet.unirioja.es/servlet/articulo?codigo=5680305

Toledano, B. (Julio de 2017). *El Mundo*. Obtenido de El Mundo: http://www.elmundo.es/tecnologia/2017/07/27/5979dc3146163fc6568b4674.html

Toscano, J. H. (2012). *Redes sociales y análisis de redes (aplicaciones en el contexto comunitario y virtual).* Corporación Universitaria Reformada.

Vega Bayo, M. (s.f.). Aplicación de teoría de grafos a redes con elementos autónomos. 16-22.

Villegas, J. S., & Álvarez, J. C. (2016). Los dilemas deontológicos del uso de las redes sociales como fuentes de información. Análisis de la opinión de los periodistas de tres países. *Latina de Comunicación Social*, 66-84. doi: 10.4185/RLCS-2016-1084

Wang, L., & Wang, G. (2015). Data Mining Applications in Big Data. *Computer Engineering and Applications Journal*, 143-152. Obtenido de https://search.proquest.com/docview/1942218903?accountid=29068

Yan Jin, J. D. (Diciembre de 2014). How disaster information form, source, type, and prior disaster exposure affect public outcomes: Jumping on the social media bandwagon? *Journal of Applied Communication Research, 43*(1), 44-65. doi:10.1080 / 00909882.2014.982685

Zahra Ashktorab, C. B. (2014). Tweedr: Mining twitter to inform disaster response. *ISCRAM*.

Zaki, M. J., & Jr., W. M. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms.* New York: Cambridge. Recuperado el 03 de Mayo de 2018, de www.cambridge.org/9780521766333

Zweig, K. A. (27 de Octubre de 2016). Graph Theory, Social Network Analysis, and Network Science. *Network Analysis Literacy*, 23-55. Obtenido de https://link.springer.com/chapter/10.1007%2F978-3-7091-0741-6_2