



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS

SEIS AÑOS DE COLABORACIÓN EN EL
COLOQUIO DE GRÁFICAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
ACTUARIA

PRESENTA:
NARDA CORDERO MICHEL

DIRECTORA DE TESIS:
BIBIANA OBREGÓN QUINTANA



Ciudad Universitaria, CDMX

2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de datos del jurado

1. Datos del alumno
Cordero
Michel
Narda
56890921
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
304684201
2. Datos del tutor
Dra.
Bibiana
Obregón
Quintana
3. Datos del sinodal 1
Dra.
Mucuy-Kak del Carmen
Guevara
Aguirre
4. Datos del sinodal 2
Dr.
Ilán Abraham
Goldfeder
Ortiz
5. Datos del sinodal 3
Dra.
Mika
Olsen
6. Datos del sinodal 4
M. en C.
Briseida Guadalupe
Trejo
Escamilla
7. Datos de la tesis
Seis años de colaboración en el coloquio de gráficas
110 p.
2019

Al gremio



Agradecimientos

A mi queridísima casa de estudios, la sin par UNAM, que me ha proveído de un biotopo, el primer sitio fuera de casa al que realmente pertenezco; y a la Facultad de Ciencias por formarme como *estudiante* y por confiarme un grupo para que experimente el ser docente.

Al Coloquio le agradezco su naturaleza: errante, amistosa y abierta; por iniciarme como ponente y por hacer de este estudio un divertimento.

A Bibiana le agradezco montones de cosas: haberme aceptado como alumna a sabiendas de no poder conservarme, su tolerancia con mis desapariciones, su paciencia y disposición para ayudarme en todo, sus continuas ofertas de café y dulces, su tiempo y, sobre todo, acompañarme a dar cierre a este tortuoso capítulo de mi existencia con un trabajo que me fue, extrañamente, interesante.

A Pablo por su ayuda con la tecnología.

A mis lectores por la singularidad de sus nombres, que hacen de mi hoja de datos algo excepcional.

A Mucuy por leerme.

A Mika e Ilán por compartirme su interés en mi trabajo, por sus reflexiones, por señalarme la carencia de matices, tan importantes en los estudios humanos, y por los agudos comentarios que me ayudaron a mejorar esta tesis.

A Mika por darme su ejemplar impreso en el que encontré un lío con los márgenes.

En especial, le agradezco a Ilán haberme empujado, en más de una ocasión, a titularme de actuaría y por su genuino gusto en verme conseguirlo.

A Briseida por su infatigable humor.

A Hortensia, por acogerme en el maravilloso mundo de sus colegas.

A mi madre por sugerirme el objeto de estudio de esta tesis, que vaya que fue bueno, y por sentarse conmigo a leer y corregir.

A mi empecinado encono que, aún con el espíritu abatido, me hizo salir victoriosa *desta* cruel batalla.

Índice general

| | |
|----------------------------------------------------|------------|
| Introducción | 1 |
| 1 Marco teórico | 5 |
| 1.1 Breve historia de la teoría de redes | 5 |
| 1.2 Definiciones, medidas y nomenclatura | 12 |
| 1.3 Modelos de redes complejas | 30 |
| 2 Metodología, análisis y resultados | 37 |
| 2.1 Metodología | 37 |
| 2.2 Análisis y resultados | 39 |
| Conclusiones | 71 |
| Anexos | 74 |
| A Nodos etiquetados | 75 |
| B Algoritmo BFS | 79 |
| C Algoritmos de Gephi | 81 |
| D Tablas de datos | 91 |
| Bibliografía | 107 |

Introducción

En el ámbito académico son comunes los eventos para compartir entre estudiosos de una misma área el trabajo, generalmente reciente, que se ha realizado de manera individual o en grupo. Con este fin se organizan congresos, coloquios, seminarios e, inclusive, escuelas de una semana. Estos eventos sirven, tanto para ponerse al corriente con las últimas aportaciones al conocimiento, como para contactar personalmente con los autores de estas contribuciones y, en ocasiones, crear nuevas relaciones colaborativas para la investigación.

En este trabajo se estudia la red de colaboración en las ponencias presentadas en las ediciones XXVIII, XXIX, XXX, XXXI, XXXII y XXXIII del Coloquio Víctor Neumann-Lara de Teoría de las Gráficas, Combinatoria y sus Aplicaciones (en adelante Coloquio VNL), que se llevaron a cabo en los años 2013, 2014, 2015, 2016, 2017 y 2018.

En el Coloquio VNL participan miembros de la comunidad de matemáticas discretas de diversas casas de estudio de nuestro país; asimismo brinda la oportunidad a nuevos estudiantes del área de conocer los temas de investigación actuales y quiénes los investigan.

Yo formo parte de esta comunidad. Dos veces participé como ponente en 2015 y 2018; y dos veces como asistente en 2016 y 2017. Sin embargo, ya sea por mi carácter o mi distracción, no había reparado en los grupos de trabajo y sus intereses. Me parece que a otro estudiante como yo, le será útil conocer la estructura de nuestra red, para saber qué proyectos y equipos de trabajo le son más afines y a quién aproximarse.

El objetivo de este trabajo es construir y caracterizar la red de colaboración en las ponencias de las últimas seis ediciones del Coloquio Víctor Neumann-Lara de Teoría de las Gráficas, Combinatoria y sus Aplicaciones, así como calcular las métricas de dicha red y hacer su análisis.

Para ello se calcularon las métricas que proporciona el *software* de Gephi¹ y se analizaron los resultados para saber cómo son los grupos de trabajo y determinar cuáles actores de la red son más importantes.

El Coloquio VNL

El Coloquio VNL es un evento dedicado a la divulgación de las matemáticas discretas, tanto en su lado teórico como en su lado aplicado, que se ha llevado a cabo anualmente desde 1985.

El Coloquio VNL fue un proyecto impulsado por un grupo de investigadores del área, entre los cuales destaca Víctor Neumann Lara, que fue el principal impulsor del estudio de la teoría de gráficas en México. Para el momento en que se fundó el Coloquio VNL, Víctor Neumann era investigador en el Instituto de Matemáticas (IMATE) de la Universidad Nacional Autónoma de México (UNAM).

Cuando comenzó el Coloquio VNL, se llamaba Coloquio de Teoría de las Gráficas, Combinatoria y sus Aplicaciones, fue hasta el año 2003, después de que Víctor Neumann falleciera, dando una plática en una de las emisiones del Coloquio VNL, que se le cambió el nombre a Coloquio Víctor Neumann-Lara de Teoría de las Gráficas y sus Aplicaciones.

Hay tres maneras de participar en el Coloquio VNL: las pláticas, ya sean las conferencias impartidas por invitados (de una hora) o las que dan los asistentes (de veinte minutos); los pósteres; y el planteamiento de problemas abiertos.

El Coloquio VNL es un evento itinerante, cada año cambia su localización, aunque con los años se han repetido sedes nunca se hace en el mismo lugar dos años consecutivos. Generalmente se realiza en las universidades de los estados o centros de investigación y ha recorrido dieciséis de las treinta y dos entidades federativas de la República Mexicana, además de una visita excepcional a la Universidad de La Habana, Cuba.

El Coloquio VNL ha tenido diversos patrocinadores para llevarse a cabo. En los últimos seis años ha obtenido el apoyo constante de la Sociedad Matemática Mexicana (SMM), de la UNAM y de la Universidad Autónoma Metropolitana (UAM). Esta última, de 2014 a 2017, lo brindó a través de las unidades Iztapalapa y Cuajimalpa (UAM-I y UAM-C, respectivamente). Además, recibió apoyos por parte del

¹Gephi es un *software* libre que sirve para ver y explorar gráficas y redes, mismo que se puede obtener en [43].

Consejo Nacional de Ciencia y Tecnología (CONACyT), de la Secretaría de Cultura del Estado de Michoacán y de la Universidad Michoacana de San Nicolás de Hidalgo en 2013; del IMATE, del Instituto Tecnológico Autónomo de México (ITAM) y de la Universidad Veracruzana en 2014; del IMATE y del ITAM en 2015; del IMATE y de la Universidad de Guanajuato en 2016; del IMATE y de la Universidad Autónoma de San Luis Potosí en 2017; y del IMATE, de la Universidad de la Habana, de la UAM-C y del Banco Santander en 2018.

Organización del texto

Este trabajo está organizado como sigue: En el primer capítulo, que consiste de tres apartados, se aborda la parte teórica sobre las redes, se comienza por una breve historia de la teoría de redes, seguida de un apartado de definiciones, medidas y cómo calcularlas y la nomenclatura del texto; a continuación se tiene una sección de modelos teóricos de redes en la que se explican tres de ellos para poder identificar cuál sirve para modelar la red del Coloquio VNL. En el segundo capítulo se habla sobre la parte práctica, es decir, la metodología para construir la red, seguida de los datos que se obtuvieron y el análisis de los mismos. Se finaliza con las conclusiones. Además, al final del texto se incluyen cuatro anexos: en el primero se muestran los nodos con etiquetas, en el segundo se explica el algoritmo de búsqueda en profundidad, en el tercero se encuentra la manera en que Gephi calcula la cifras y en el cuarto las tablas con los datos de la red.

Capítulo 1

Marco teórico

Este capítulo está dividido en tres secciones. En la primera se hace un breve recorrido por la historia de la teoría de redes, algunas situaciones que se quisieron estudiar y el uso de la teoría de redes para comprenderlas. En la segunda sección se presentan las definiciones generales, las medidas de una red y la nomenclatura pertinente, usadas a lo largo del trabajo. Finalmente, en la tercera sección, se explican tres de los modelos teóricos para redes complejas, con la finalidad de ver a cuál de ellos se asemeja la red del Coloquio VNL.

1.1 Breve historia de la teoría de redes

El origen del estudio de la teoría de redes (y de la teoría de gráficas) se sitúa, de acuerdo a la literatura del área [5], en 1736 con la pregunta que ahora se conoce como el problema de los Puentes de Königsberg. El problema es este: La ciudad de Königsberg está dividida por el río Pregel en cuatro áreas terrestres mismas que están conectadas entre sí por siete puentes, como se muestra en la figura 1.1 (i). Se quiere hacer un recorrido por toda la ciudad, que inicie en alguna de las áreas terrestres y que pase por cada uno de los siete puentes exactamente una vez. Hay dos variantes: la original es que el recorrido empiece y termine en el mismo lugar, y la alternativa es que los puntos de inicio y término no sean iguales. La pregunta es: ¿existe alguno de estos recorridos?

En 1735, Euler determinó que es imposible que exista un recorrido con estas características. Fue siglo y medio más tarde que se representó este problema con una

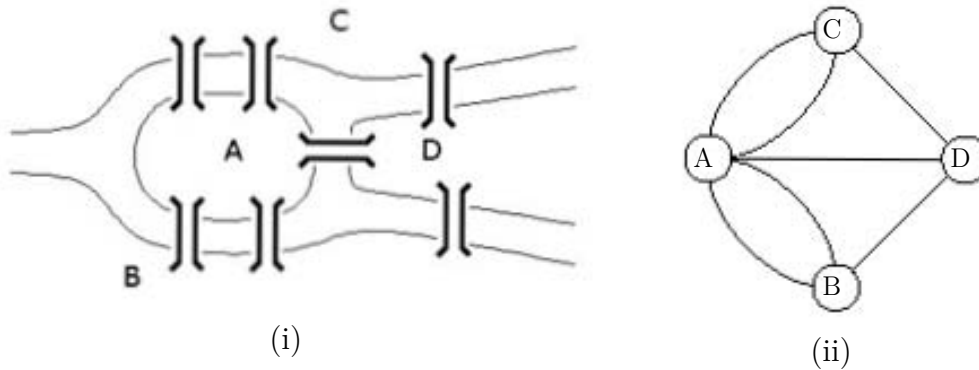


Figura 1.1: En (i) se muestra un mapa de la ciudad de Königsberg, con sus cuatro áreas de tierra firme, A, B, C y D, y sus siete puentes, obtenida en [42]. En (ii) se muestra la representación gráfica del problema de los Puentes de Königsberg, donde las áreas de tierra son representadas con vértices y ponemos una arista entre dos vértices por cada puente existente entre las zonas de tierra respectivas.

gráfica¹ [7], como en la figura 1.1 (ii). De hecho, Euler resolvió un problema más general, con esa solución se pudo hacer una caracterización de las gráficas en las que existe un recorrido que pasa por todos sus vértices, que pasa por cada una de las aristas exactamente una vez y que inicia y termina en el mismo vértice, y también se caracterizó el caso en el que el recorrido inicia y termina en vértices diferentes.

Después del episodio de los Puentes de Königsberg ha habido un buen número de problemas o situaciones que se han modelado con gráficas, como la red de telefonía, los circuitos eléctricos, los mapas de vialidades, los problemas de transporte y reparto, las relaciones sociales, las bases de datos (capítulo 2 y secciones 3.1, 3.6, 3.7, 4.3 y 19.2 de [27]), o bien las asignaciones de horarios [12]. Estos modelos con gráficas pueden tener atributos adicionales como colores en los vértices o en las aristas, pesos en las aristas o dirección en las aristas. Además, de acuerdo a la definición de gráfica que se utilice, se puede permitir que haya pares de vértices relacionados por más de una arista o vértices relacionados consigo mismos.

En particular, los modelos en los cuales las personas o grupos de personas son representadas por vértices, y las aristas representan relaciones sociales entre ellos como parentesco o amistad, se denominan *redes sociales*. El estudio de este tipo de redes se formalizó con las aportaciones de Jacob Moreno en los años 30, empezando por

¹Una *gráfica* o *red* es un estructura que consiste de dos conjuntos, el primero consiste de objetos llamados *vértices* y el segundo de parejas de vértices llamadas *aristas*.

1.1. Breve historia de la teoría de redes

una conferencia en 1933, seguida de la publicación de su libro *Who Shall Survive?*² en 1934, donde analiza grupos de personas y sus relaciones [23].

En los años 60, Stanley Milgram, quien era psicólogo experimental, realizó una investigación que ahora es conocida como el *experimento del mundo pequeño* [21, 34]. El objetivo de este experimento era saber cuántos intermediarios, en promedio, son necesarios para conectar a dos individuos de una red. Para ello hizo un experimento de mandar cartas por correo, 96 de ellas, cuyos destinatarios finales estaban prefijados (algunos amigos de Milgram que residían en Boston, Massachusetts) y él se las envió en un principio a 96 personas seleccionadas al azar de una guía telefónica de Omaha, Nebraska. En las cartas se informaba a los receptores quién era el destinatario final de dicha carta, y se les solicitaba que reenviaran la carta a una persona que pudiera conocer al destinatario final o que tuviera más posibilidades de hacerla llegar. Sólo 18 de las cartas llegaron a sus destinatarios finales y el promedio de pasos que tuvo que dar cada carta fue 5.9, de aquí surgió la idea de los seis grados de separación, que plantea que entre cualesquiera dos individuos en el mundo hay una manera de conectarlos con cinco intermediarios en promedio, es decir, seis pasos.

Derek J. de Solla Price, un físico e historiador contemporáneo a Milgram, estudió la red de citas de artículos académicos, escribió un artículo al respecto [29], pero su interés más grande era poder explicar cómo se forma una red y cómo se produce su distribución de grado que, en muchos ejemplos reales, sigue una ley de potencia. Más adelante, en los años 70, elaboró un modelo para crear redes de citas en artículos académicos con una distribución de grado³ que sigue una ley de potencia, mismo que se dio a conocer en [30]. Este modelo establece que un artículo nuevo cita publicaciones previas elegidas al azar, cuya probabilidad de ser elegidas es proporcional al número de citas con las que dichas publicaciones ya cuentan. Su planteamiento se basa en que los artículos que obtienen muchas citas desde su publicación hasta cierto momento, luego son altamente citados, mientras que, aquellos que no reciben mucha atención desde el principio, luego adquieren muy pocas citas.

También en los años 60, la socióloga Harriet Zuckerman, tras una serie de entrevistas que realizó a ganadores del Premio Nobel, observó que en el ámbito científico, ocurre que las personas que son ya reconocidas por su trabajo, como los ganadores del Premio Nobel, suelen conseguir más crédito por sus aportaciones que otras personas, aún cuando las contribuciones sean equiparables. Además, en los artículos

²El título en español es *¿Quién sobrevivirá?*

³En este caso, el *grado* del artículo es la suma de dos números: el número de citas que hace el artículo más el número de veces que dicho artículo es citado por otros.

cuya autoría es compartida entre varias personas, es frecuente que al ser referidos se olvide quiénes son los coautores que no son famosos. Zuckerman publicó su trabajo en el libro *Scientific Elite: Nobel Laureates in the United States*⁴. También se le reconoce por el efecto Matilda⁵, como coautora de [20] que está publicado a nombre de su marido, pues en ese artículo se analiza el fenómeno producido por los famosos, conocido como efecto Mateo, que veremos más adelante.

En 1973 y 1982, Mark Granovetter estudió otro fenómeno que ocurre en las redes sociales, que hoy se conoce como *el poder de los enlaces débiles*⁶, en el cual observó que para un individuo es menos probable que sus conocidos de distintos espacios o momentos de su vida, *enlaces débiles*, se conozcan entre sí. Mientras que sus amigos cercanos y familiares, *enlaces fuertes*, es más probable que estén conectados entre sí. Por lo que, si se fija la atención en la red que consiste de un individuo y sus enlaces débiles, es altamente probable que falten muchas aristas, es decir, la red tendrá una densidad pequeña, mientras que, si se fija la atención en la red que consiste de un individuo y sus enlaces fuertes, se espera que sea una red con casi todos los enlaces posibles, es decir, la red tendrá una densidad alta. En [15], Granovetter analizó las consecuencias que tienen los enlaces débiles en el flujo de información. Por ejemplo, él vio que muchas veces las ofertas de trabajo llegan vía un enlace débil. Se observa que si dos personas, i y j , son conocidos, cada una de ellas tiene un círculo de amigos cercanos que se espera sean ajenos entre sí, porque el enlace entre i y j es débil. Así, para que la información pase del círculo de amistades de i al círculo de amistades de j , el enlace entre i y j es la manera más inmediata, de otro modo la información podría no llegar o tardar mucho. Si se regresa a las ofertas de trabajo, se puede pensar que si en el trabajo de i hay una posición disponible, que él no necesita, se lo comunique a j , ya que no le representa un costo. En contraste, es difícil que la información que tiene un costo para el mensajero se llegue a transmitir a través de un enlace débil, para ese tipo de información se requiere de un enlace fuerte. Las personas sin enlaces débiles o con muy pocos quedarán restringidas a la información que sus enlaces cercanos puedan darles y sabrán poco de lo que ocurre en otras partes de la red. Además, en una vista panorámica de la red, si hubiera pocos enlaces débiles

⁴Título en español *Élite científica: Premios Nobel en los Estados Unidos*.

⁵Se le llama *efecto Matilda* al hecho de ignorar o desmerecer el trabajo de las mujeres en la ciencia, dándole el crédito a sus cónyuges o colegas masculinos [31]. Este fenómeno fue descrito por primera vez por Matilda Joslyn Gage en 1870 en su ensayo *Woman as an Inventor (La mujer como inventora* [11]), de ahí el nombre.

⁶Este nombre se debe al título de uno de sus artículos [14], *The strength of weak ties*.

1.1. Breve historia de la teoría de redes

en ella, entonces estaría dividida en grupos que no se relacionan entre sí, lo cual complicaría el flujo de nuevas ideas, de noticias, de artículos académicos o lo que sea que se esté analizando en la red.

En 2001, Mark E. J. Newman publicó una serie de artículos [24, 25, 26] sobre las redes de autoría y colaboración en trabajos científicos. Construyó dos redes: en la primera, de autoría, los nodos se dividen en dos conjuntos ajenos, uno conformado por los autores y otro por las publicaciones, luego un autor está ligado a una publicación si es autor de esa publicación, de este modo, todas las aristas tienen un extremo en los nodos-autores y otra arista en los nodos-publicaciones. A partir de esta red, se construye la segunda red, de colaboración, en la que todos los nodos son autores y dos de ellos son adyacentes si son coautores de una publicación. A esta red se le puede asignar un peso a las aristas que corresponda con el número de publicaciones comunes que tiene un par de autores, mismo que se obtiene fácilmente de la red de autoría.

En este trabajo se construye directamente la red de colaboración, no de publicaciones sino de trabajos presentados como plática en el Coloquio VNL, sin embargo no se da más peso a las aristas entre pares de personas que han trabajado más de una vez juntas.

En el estudio de redes reales, hay rasgos estructurales que no se pueden identificar a simple vista, pero que se pueden determinar con el análisis estadístico de sus medidas de centralidad. Se pueden identificar dos tipos de redes de gran escala de acuerdo a sus distribuciones estadísticas: las redes homogéneas y las redes heterogéneas.

En una red *homogénea* las distribuciones de los grados, de los pesos y de la centralidad intermedia tienen colas ligeras, es decir, que decrecen rápidamente, como las distribuciones de Gauss y Poisson. En una red *heterogénea* la conexidad y los patrones con peso tienen distribuciones sesgadas y con colas pesadas.

Consideremos la distribución de los grados $P(k)$: en una red homogénea, el grado de los vértices se puede aproximar con una constante igual al grado promedio, denotado por $\langle k \rangle$. Es decir, la mayoría de los vértices tienen el mismo grado o casi el mismo grado, y son pocos los que tienen un grado muy pequeño o muy grande, por lo que su distribución de grado se asemeja a una campana y se puede aproximar con una distribución Poisson. En una red heterogénea, se encontrarán muchos vértices cuyo grado es pequeño y pocos vértices cuyo grado es sumamente grande. Por ejemplo, en la red de los aeropuertos siempre aparecen *hubs*⁷, que son aeropuertos en ciudades

⁷Los *hubs* son vértices cuyo grado es extremadamente grande.

grandes, a las que muchas personas van y en las que muchas otras hacen escala para tomar otro vuelo en conexión, y por tanto, son aeropuertos a los que llegan o de los que salen vuelos desde muchos otros puntos; mientras que hay un gran número de aeropuertos en ciudades pequeñas, usualmente destinos finales de viajeros, a las que no se puede llegar desde muchos destinos. En este estilo de distribución, los grados de los demás vértices varían en este rango amplio entre el grado de los vértices con apenas unas cuantas conexiones y los que tienen muchas (cientos o miles) de ellas. Aquí, el grado tiene una distribución decreciente que se aproxima a una ley de potencia⁸, $P(k) \sim k^{-\gamma}$ donde $2 < \gamma \leq 3$, que se asemeja a una recta con pendiente negativa cuando la representamos en un plano de doble logaritmo, es decir, cuyos ejes representan los logaritmos: $\log(k)$ y $P(\log(k))$. En este tipo de distribuciones, si se toma un vértice al azar, la probabilidad de elegir uno cuyo grado esté por encima del grado promedio, $\langle k \rangle$, es mayor que cero y no despreciable.

Otra característica que se puede observar en las redes reales, en particular en las redes sociales, es que en muchas ocasiones las personas hacen vínculos con personas en las que reconocen características similares a las propias, como pueden ser la religión, el nivel educativo, la edad, el origen sociocultural, el poder adquisitivo, el fenotipo, entre otras. Los seres humanos tienden a juntarse con quienes sienten semejantes a sí mismos, a este comportamiento se le llama *homofilia*⁹. Cuando ocurre lo contrario, es decir, cuando se juntan con personas que no se parecen a ellos, se le llama *heterofilia*¹⁰, que es menos común que la homofilia. Un ejemplo de una red heterófila es la red de parejas sexuales donde la característica observada es el sexo de los individuos, pues en esta hay muchas más parejas heterosexuales que de otro tipo. Un ejemplo de una red homófila, es la red de citas de artículos académicos donde la característica observada es el campo de estudio, en esta red es frecuente que los artículos citen a otros de la misma área y menos común que citen artículos de otros campos.

En las redes sociales hay aún otra característica que se debe tomar en cuenta: la visibilidad de la red. Hay redes en las que se sabe exactamente quiénes son los actores y cuáles son los vínculos que existen entre ellos. Por ejemplo, si se considera la red en la que los nodos son actores de cine y dos actores están enlazados si han aparecido en la misma película, no hay actores sin nombre y podemos consultar si dos de ellos han trabajado juntos en los créditos de las películas. Sin embargo, hay redes sociales en las que se desconocen algunos actores o algunos enlaces. Por ejemplo, en

⁸En algunos textos se le llama distribución *Pareto* a una ley de potencia.

⁹En inglés *homophily* o *assortative mixing*, en algunos textos.

¹⁰En inglés *heterophily* o *disassortative mixing*.

1.1. Breve historia de la teoría de redes

una red de contrabandistas de sustancias ilegales, es posible que se conozcan algunos de los actores pero, seguramente, no a todos los integrantes de la red y, de los que se tienen identificados, puede que se ignoren enlaces que existen entre ellos. Para este último tipo de redes, cuyos miembros prefieren permanecer en secrecía, se han desarrollado algunas técnicas de recopilación de datos para poder estudiarlas. Una de ellas es la llamada *muestreo de bola de nieve*¹¹ que consiste en encontrar un miembro, digamos A , de la población que se desea estudiar, y entrevistarlo. Después, ganada la confianza de A , pedirle la identidad de otros miembros del grupo a estudiar. Ahora, contactar a estos nuevos miembros con la carta de presentación de ser conocido de A . Hacer lo mismo con los conocidos de A para obtener así contacto con otros miembros del grupo.

Otra técnica de recopilación de datos de poblaciones ocultas es el *rastreo de contactos*¹², que es un caso particular del muestreo de bola de nieve cuando lo que se estudia es la incidencia de infecciones, por ejemplo del Virus de la Inmunodeficiencia Humana (VIH). En este caso, cuando se detecta que una persona es portadora del virus, se desea hacer una reconstrucción de las personas que podrían haber sido expuestas a este, ya sea por relaciones sexuales o por el uso compartido de agujas, para poder rastrearlas, tratarlas y prevenirlas a fin de evitar más contagios. En principio, se hace el muestreo para cortar la epidemia, por así decirlo, pero teniendo la red construida se suelen hacer otros estudios, médicos, sociales o de algún otro tipo. Sin embargo, la red que surge del rastreo de contactos tiene mucha información sobre las personas portadoras del virus y poco o nada de aquellas que dieron negativo en las pruebas.

Cabe mencionar que estas dos técnicas producen redes parciales, por lo que no deben estudiarse rasgos estructurales en ellas porque los datos obtenidos pueden ser erróneos. Para lo que sirven las redes construidas así es para hacer hipótesis sobre las poblaciones ocultas.

La red del Coloquio VNL es una red visible, pues se sabe quiénes son los expositores y sus coautores en cada una de las pláticas de los años que se están estudiando. No obstante, es importante mencionar este tipo de muestreos de poblaciones ocultas, pues la teoría de redes ha sido una herramienta para estudiar las crisis epidemiológicas con la que se ha obtenido información interesante sobre las relaciones humanas.

En particular, en los años 1990, Edward O. Laumann y Yoosik Youm [19] hicieron

¹¹En inglés *snowball sampling*.

¹²En inglés *contact tracking*.

un estudio sobre las infecciones de transmisión sexual en los Estados Unidos, en el cual se percataron de que la tasa de personas afroamericanas portadoras de una serie de infecciones de transmisión sexual (ITS) era diez o veinte veces más grande que la de otros grupos étnicos o raciales. En principio no había una explicación clara de por qué ocurría esto. Después de estudiar la red de relaciones sexuales de distintos grupos se dieron cuenta que la manera de elegir parejas sexuales de los afroamericanos era, por una parte, mucho más variable que la de otros grupos: si se considera que los individuos que han tenido una sola pareja sexual el último año están en la periferia y aquellos que han tenido cuatro o más parejas sexuales el último año están en el núcleo, entonces ocurre que es cinco veces más probable que un afroamericano en la periferia se vincule con uno en el núcleo, a que un blanco en la periferia se relacione con un blanco en el núcleo. Por otra parte, las ITS permanecían dentro de la población afroamericana porque eran más selectivos para relacionarse con otras etnias.

Este tipo de estudios arrojan ideas interesantes: por ejemplo que, en apariencia, para tener relaciones sexuales con un blanco que ha tenido varias parejas sexuales hay que hacer currículum, mientras que, si el objetivo sexual es afroamericano, lo que parece necesario es ser también afroamericano. Si bien esta idea es un poco burda, el plantearse nuevas interrogantes a partir de los datos observados puede impulsar el estudio en una dirección que no se había contemplado.

En cualquier caso, las relaciones humanas son un objeto de estudio realmente interesante y la teoría de redes nos ayuda a obtener información que de otro modo sería difícil ver.

1.2 Definiciones, medidas y nomenclatura

En este apartado se introducen las nociones con las que se va a trabajar a lo largo del texto. Se utilizaron las definiciones y nomenclatura del libro de Mark E. J. Newman [27].

Una *red* o *gráfica*, denotada por $G = (V, A)$, consiste de dos conjuntos, V y A , el primero consiste de objetos llamados *vértices* y el segundo de parejas de vértices llamadas *aristas*, que representan la relación entre dichos vértices. A los vértices también se les conoce como *nodos* o, en sociología, *actores* y a las aristas también se les llama *arcos* o, en sociología, *vínculos*. Cuando las relaciones que representan las aristas son asimétricas, se hace una distinción dando una dirección a la arista;

1.2. Definiciones, medidas y nomenclatura

cuando este es el caso se les llama *flechas* o *arcos dirigidos* y se dice que la red o gráfica es *dirigida*¹³. Siempre que se omita la palabra *dirigida* se está hablando de una red no dirigida.

Cabe señalar que, aunque se usen de manera indistinta las palabras red y gráfica, nodo y vértice o arco y arista, en general una red con sus nodos y arcos, se refiere a una estructura de la vida real, mientras que una gráfica con vértices y aristas, es un objeto abstracto que puede ser usado para modelar dicha estructura. Así, varias redes pueden tener la misma representación gráfica pero, según la naturaleza de la red, habrá características de la gráfica que la representa que sean más interesantes. En este trabajo se usan las palabras como sinónimos.

Se denota por n a la cardinalidad del conjunto de vértices y por m a la cardinalidad del conjunto de aristas. Cuando exista un arco entre dos nodos, u y v , se denotada por uv , se dice que u y v son nodos *adyacentes* o *vecinos* y que la arista uv *incide* tanto en u como en v . Se define la *vecindad* de un vértice u , denotada por $N(u)$, como el conjunto de vecinos del vértice u ; además, se dice que el conjunto $N[u] = N(u) \cup \{u\}$ es la *vecindad cerrada* de u . Al número de aristas que inciden en un vértice u se le llama el *grado* de u y se denotada por k_u . A los vértices cuyo grado es desproporcionadamente grande con respecto a los grados de los demás vértices se les denomina *hubs*. En el caso dirigido, si (u, v) es un arco dirigido, se dice que u es un *invecino* de v y que v es un *exvecino* de u . Además, se separa el grado de un vértice u en dos: k_u^{ex} denota el número de arcos dirigidos que salen desde el vértice u , *exgrado*, y k_u^{in} denota el número de arcos dirigidos que llegan al vértice u , *ingrado*. La suma $k_u^{ex} + k_u^{in}$ es el *grado* del vértice u , k_u ¹⁴. En el caso de las redes dirigidas, no es lo mismo que lleguen muchos arcos dirigidos a un nodo a que salgan de él, por lo que cabe hacer una distinción entre ambos. Para el caso dirigido se les llama *autoridades*¹⁵ a los vértices considerados importantes por las medidas de centralidad¹⁶, cuyo ingrado es grande, y se les llama *hubs* a los vértices que apuntan a las autoridades. Un ejemplo de esto es la red de publicaciones científicas. En esta, dos artículos son adyacentes si un artículo cita al otro y la adyacencia se representa por una flecha que inicia en el artículo que hace la cita y termina en el artículo que

¹³A las gráficas dirigidas también se les llama *digráficas*.

¹⁴En la teoría de gráficas, dado un vértice u , se suele denotar al exgrado de u por $d^+(u)$, al ingrado de u por $d^-(u)$ y al grado de u por $d(u)$.

¹⁵En inglés *authorities*.

¹⁶La *centralidad* de un vértice es un valor numérico que se le asigna a un vértice de acuerdo a su importancia dentro de la red.

es citado: un artículo con información relevante suele ser citado a menudo porque se reconoce la importancia de su contenido, lo cual le concede la calidad de autoridad, mientras que los artículos que hacen referencia a él son *hubs*, pues aunque no tengan la información directamente, ofrecen la ruta para obtenerla.

En este trabajo, debido al tamaño y tipo de red que se estudia, no se encontraron vértices con un grado extremadamente grande, de hecho, el grado más grande es 20 y el más pequeño es 1, por lo que la diferencia entre ellos es muy pequeña. El *software* de Gephi utiliza un algoritmo para determinar los *hubs* y las autoridades de la red y a cada nodo le asigna dos valores, un peso de *hub* y un peso de autoridad, mismos que coinciden cuando la red es no dirigida (que es, aquí, el caso). A los vértices con el mayor peso de *hub*, de acuerdo a los resultados que proporciona dicho *software*, es a los que se les llamará *hubs*, que no necesariamente coinciden con los vértices de grado mayor por las características de la red.

Se trabajó con una red en la que no se permite que existan dos o más aristas entre la misma pareja de vértices (*aristas múltiples*), ni aristas que conecten a un vértice consigo mismo (*lazos* o *rizos*). Es decir, la red será lo que se conoce en algunos libros como una *gráfica simple*. Cabe señalar que, si bien no se le da dirección a las aristas, se puede hacer un estudio diferente en el que la dirección sale a relucir. Generalmente, en las redes de colaboración en publicaciones académicas ocurre un fenómeno: los estudiantes y sus asesores publican o presentan como pláticas trabajos en conjunto. Pero no es lo mismo ser asesor que asesorado, por lo que se le puede dar una dirección a la arista. Además, así se podría distinguir cuáles relaciones son resultado de encuentros académicos entre investigadores, de igual a igual poniendo arcos dirigidos simétricos, y cuáles provienen de la formación académica para encontrar lo que se llama “escuelas” de pensamiento o, de manera coloquial, los parentescos académicos.

Una *subgráfica* de una gráfica es en sí misma una gráfica y está contenida en la gráfica original, es decir, su conjunto de vértices es un subconjunto de los vértices de la original y cualquier arista de la subgráfica es una arista en la original. Dado un subconjunto de vértices U de una gráfica G , llamaremos *subgráfica inducida* por U en G , denotada por $G\langle U \rangle$, a la subgráfica de G cuyo conjunto de vértices es U y cuyo conjunto de aristas consiste de todas las aristas de G que tienen ambos extremos en U . Además, si una subgráfica contiene a todos los vértices de la gráfica, diremos que es una subgráfica *generadora*.

1.2. Definiciones, medidas y nomenclatura

Una *ruta*¹⁷ es una sucesión de vértices en la que cualesquiera dos consecutivos son adyacentes. Es decir, es un recorrido que nos permite llegar de un vértice a otro a través de las aristas de la red, pasando por ciertos vértices intermediarios. La *longitud* de una ruta, es simplemente la cantidad de aristas que atraviesa (la cantidad de pasos que da la ruta), si pasamos por una arista más de una vez, la contamos tantas veces como es atravesada por la ruta.

En principio, una ruta puede visitar algún vértice más de una vez, o recorrer una arista en múltiples ocasiones, a las rutas que no repiten vértices (y en consecuencia, tampoco aristas) se les llama *autoevasivas*¹⁸. Otro tipo de rutas son los *ciclos*, que inician y terminan en el mismo vértice y en los cuales todos los vértices intermedios son diferentes entre sí y al vértice inicial/final. Se tiene interés en los ciclos de longitud tres y se les distingue llamándolos *triángulos*. A una gráfica que no contiene ciclos se le llamará *acíclica*.

Una *ruta geodésica* es una ruta entre dos vértices, tal que no existe ninguna otra ruta entre ellos cuya longitud sea menor. A la longitud de una ruta geodésica se le suele conocer como la *distancia geodésica* o, simplemente, *distancia* entre el par de vértices que conecta. Una ruta geodésica es siempre autoevasiva, pues si la ruta se intersecta a sí misma, ésta contiene un bucle¹⁹ y podemos construir, entre la misma pareja de vértices, una ruta de longitud menor cortando dicho bucle. Por convención, se dirá que dos vértices están a distancia geodésica infinita si no existen rutas geodésicas entre ellos. Se denotará la distancia entre los vértices i y j por d_{ij} o, si no se presta a confusión, por d simplemente. Se dirá que una gráfica es *conexa* si entre cualesquiera dos de sus vértices existe una ruta. De otro modo se dirá que es *disconexa*. En una gráfica desconexa, llamaremos *componente* a una subgráfica conexa para la que no exista ninguna otra subgráfica conexa que la contenga propiamente. A las gráficas conexas que son acíclicas se les conoce como *árboles*, mientras que a las gráficas desconexas que no contienen ciclos se les llama *bosques*, pues cada una de sus componentes es un árbol.

En una gráfica conexa, la *excentricidad* de un vértice es igual a la longitud de la ruta geodésica más larga que inicia (o termina) en él, es decir, es igual al máximo de las distancias de dicho vértice a cualquiera de los otros vértices; y el *diámetro* de la gráfica es el máximo de las excentricidades, en otras palabras, el diámetro es igual a

¹⁷Una ruta en teoría de redes es lo que llamamos camino en teoría de gráficas.

¹⁸En inglés *self-avoiding paths* que, en otras traducciones, aparecen como *autoevitantes*. En la teoría de gráficas estas son las llamadas *trayectorias*.

¹⁹Por bucle se alude a una ruta que inicia y termina en el mismo vértice.

la longitud de la ruta más larga entre todas las rutas más cortas entre todo par de nodos. Mientras que en una gráfica desconexa, la excentricidad de cada vértice será infinito y el diámetro de la gráfica será igual a la longitud de una ruta geodésica lo más larga posible que exista entre las parejas de vértices de la gráfica para las cuales exista una ruta que los conecte, es decir, el diámetro de la gráfica es el máximo de los diámetros de sus componentes conexas.

Dada una gráfica conexa, a veces interesa saber qué tan fácilmente pierde esta propiedad, es decir, si quitando uno o varios vértices (o bien, una o varias aristas) se obtiene una gráfica que sea desconexa. Se llama un *conjunto de corte* (o en ocasiones, un *conjunto de corte de vértices*) a un subconjunto de vértices que al ser removidos de la gráfica, se obtiene una nueva en la que dos vértices específicos están desconectados y por tanto entre ellos no existe ninguna ruta que los comunique. Dados dos vértices fijos en una gráfica conexa, se llama *conjunto de corte mínimo* a un conjunto de corte que separe dichos vértices y que tenga cardinalidad lo más pequeña posible. Equivalentemente, un *conjunto de corte de aristas* es un subconjunto de aristas que al ser removidas de la gráfica, se obtiene una en la que una pareja de vértices específica está desconectada. Se llama *vértice de corte* a un vértice que al ser removido de una gráfica, produce una gráfica con un número mayor de componentes conexas y se llama *punte* a una arista que al ser removida de una gráfica, produce una nueva gráfica con exactamente una componente conexa más.

Si una gráfica es simple, entonces la cantidad máxima de aristas que puede tener es $\binom{n}{2} = \frac{1}{2}n(n-1)$, que es el total de subconjuntos de cardinalidad 2 que se pueden formar con n elementos. Al cociente del número de aristas que tiene la gráfica, entre el número de aristas posibles se le llama la *densidad* de la gráfica y se denota por ρ , así $\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}$. A una gráfica simple que posee $\binom{n}{2}$ aristas se le llama gráfica *completa*.

Un valor de densidad alto en una red social se traduce en una visibilidad alta entre los integrantes de la misma, es decir, los actores pueden ver qué están haciendo los demás y monitorearse unos a otros. Si una red tiene una densidad grande refleja cohesión, apoyo social, confianza y una transmisión eficaz de ideas, rumores o enfermedades [16]. En las redes sociales pequeñas es más fácil tener densidades altas, mientras que en las redes grandes suele ser pequeña²⁰.

Se utilizarán dos maneras de representar una red: 1) la representación gráfica

²⁰Las redes reales (que no son sociales), suelen tener una densidad baja aún si tienen pocos nodos, pues cada enlace tiene un costo y tener muchos enlaces sería muy costoso.

1.2. Definiciones, medidas y nomenclatura

intuitiva en la que se dibuja un punto por cada vértice y se unen dos puntos con una línea siempre que haya una arista entre ellos; 2) la *matriz de adyacencias*, que es una matriz A de $n \times n$ con entradas en el conjunto $\{0, 1\}$. Para construirla se etiqueta a los vértices con los enteros del 1 al n , de manera arbitraria pero fija, de modo que todos los vértices tengan etiquetas diferentes. Si se le llama A_{ij} a la entrada (i, j) de la matriz A , entonces las entradas de A están dadas por:

$$A_{ij} = \begin{cases} 1 & \text{si los vértices } i \text{ y } j \text{ son adyacentes} \\ 0 & \text{en otro caso.} \end{cases}$$

Se observa que la matriz de adyacencias es simétrica en el caso no dirigido. La definición de la matriz de adyacencias para el caso dirigido, difiere ligeramente, las entradas de A están dadas de la siguiente forma:

$$A_{ij} = \begin{cases} 1 & \text{si hay un arco dirigido que inicia en } j \text{ y termina en } i \\ 0 & \text{en otro caso.} \end{cases}$$

Se puede saber el grado de un vértice en una gráfica no dirigida sumando las entradas del renglón (o columna) que le corresponde. Así, el grado del vértice i , k_i , es $k_i = \sum_{j=1}^n A_{ij} = \sum_{j=1}^n A_{ji}$. Mientras que en el caso dirigido la suma de las entradas del i -ésimo renglón es el ingrado del vértice i , $k_i^{in} = \sum_{j=1}^n A_{ij}$, y la suma sobre las entradas de la i -ésima columna es el exgrado del vértice i , $k_i^{ex} = \sum_{j=1}^n A_{ji}$. Para el caso dirigido, en general, el ingrado y el exgrado son números diferentes.

Las potencias de la matriz de adyacencias también pueden ayudar a determinar el número de rutas de longitud r que existen entre los vértices i y j , denotado por $N_{ij}^{(r)}$. La manera de hacerlo para el caso no dirigido es la siguiente (el caso dirigido no es muy diferente):

Si hay una arista entre los vértices i y j , entonces hay exactamente una ruta de longitud 1 del vértice i al vértice j . De hecho, la entrada ij de la matriz de adyacencias es 1 si y sólo si hay una ruta de longitud 1 del vértice i al vértice j y es cero en otro caso. Por lo tanto $N_{ij}^{(1)} = A_{ij}$.

Ahora, si hay una arista entre los vértices i y k y otra arista entre los vértices k y j , entonces existe una ruta de longitud 2 del vértice i al vértice j y que pasa por el vértice k . Así, el producto $A_{ik}A_{kj}$ es 1 si existe una ruta como esa y es cero en otro caso. Si se consideran todos los posibles vértices intermedios, se obtienen todas las posibles rutas de longitud 2 desde el vértice i hasta el vértice j . Entonces

$N_{ij}^{(2)} = \sum_{k=1}^n A_{ik}A_{kj} = [A^2]_{ij}$, donde $[\cdot]_{ij}$ es la entrada ij de la matriz en cuestión.

De manera similar, el producto $A_{ik}A_{kl}A_{lj}$ es 1 si hay una ruta de longitud 3 del vértice i al vértice j que pasa por los vértices k y l , en dicho orden, y es cero en otro caso. Así, sumando sobre todos los vértices intermedios posibles, se obtiene que $N_{ij}^{(3)} = \sum_{k=1}^n \sum_{l=1}^n A_{ik}A_{kl}A_{lj} = \sum_{k=1}^n A_{ik}[A^2]_{kj} = [A^3]_{ij}$.

Si se continúa de esta manera, se obtiene que $N_{ij}^{(r)} = [A^r]_{ij}$, para cualquier entero positivo r . Se observa que, en una gráfica conexa, la distancia entre los vértices i y j es igual al mínimo entero positivo r para el cual la entrada ij de la matriz A^r es distinta de cero.

Otro uso que se le da a la matriz de adyacencias, es para determinar si una gráfica es conexa y si no lo es, saber cuántas componentes tiene y qué vértices pertenecen a cada componente. Para ello, haciendo uso de herramientas de álgebra lineal, se cambia la apariencia de la matriz de adyacencias de modo que las entradas no nulas queden acomodadas por bloques en la diagonal y cualquier otra entrada sea cero, como se muestra en el ejemplo de la figura 1.2. Sin embargo, pocas veces se utiliza

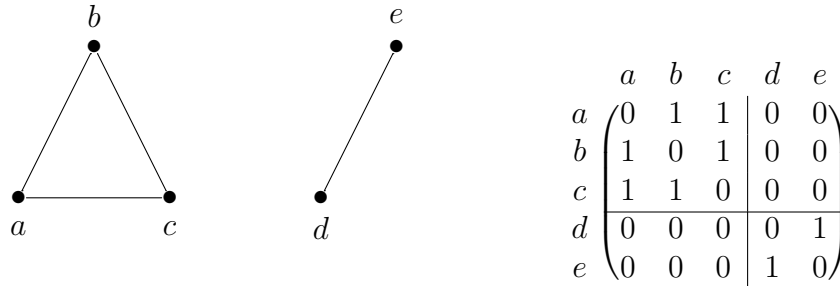


Figura 1.2: Una gráfica y su matriz de adyacencias por bloques

la matriz de adyacencias para determinar si una gráfica es conexa, en general se usa el algoritmo de búsqueda exhaustiva (BFS por sus siglas en inglés²¹), que provee un bosque generador tal que cada componente de la gráfica es generada por un árbol del bosque. Para más detalles sobre el algoritmo BFS consultar el anexo B.

Cuando se tiene una red real y no muy grande, a partir de la representación gráfica podemos encontrar algunas características de la red, por ejemplo, si tiene vértices de corte o si es acíclica, sin embargo hay propiedades que no se perciben a simple vista. Es por esta razón que se estudian algunas propiedades numéricas de la red, para saber más sobre su estructura.

²¹*Breadth-First Search.*

1.2. Definiciones, medidas y nomenclatura

Las medidas que se presentan a continuación fueron introducidas inicialmente para el análisis de redes sociales, aunque ahora se utilizan también en otras áreas. Muchas de ellas serán medidas de *centralidad*, es decir, darán parámetros para decidir cuáles vértices son los más importantes (o centrales) dentro de una red. Habrá variaciones en los resultados dependiendo de la definición de “ser importante” que se tome. Este tipo de medidas son interesantes pues ayudan a saber qué tan influyente es un individuo en la red. En lo que sigue, para una red de n nodos, x_i denotará la centralidad del vértice i , para cada $i \in \{1, 2, \dots, n\}$. La manera en que se define x_i variará en cada apartado, según el tipo de centralidad que sea introducida.

Se comenzará por la más sencilla, la *centralidad de grado*, que se determina por el grado del vértice en la red. Los vértices más centrales serán aquellos que tengan un grado mayor. En la red de colaboración, los vértices más centrales, serán aquellos que hayan trabajado con el mayor número de personas. Sin embargo, puede pensarse que no todas las conexiones tienen la misma relevancia. Por ejemplo, en las publicaciones algunas veces conviene colaborar con una persona que sea más visible en el área que publicar solo, sobretodo para aquellos que no se han dado a conocer, como estudiantes recién egresados, pues el nombre famoso puede atraer lectores. En [20], Merton nombró este fenómeno como *efecto Mateo* o *los ricos se hacen más ricos*²², pues en el Evangelio según San Mateo, en el versículo 25:29, se afirmaba que “porque a todo el que tenga, se le dará y le sobrará; pero al que no tenga, aún lo que tiene se le quitará”²³. Este fenómeno es muy común en las redes de citas de artículos académicos, si un acreedor al Premio Nobel o a la Medalla Fields hace una publicación nueva, enseguida será consultada y citada. Mientras que, las primeras publicaciones de un estudiante promedio, recibirán poca o nula atención, ya no digamos citas en artículos nuevos.

Con la idea de que no todos los vínculos son igual de importantes, se han estudiado otras medidas. Una de ellas es la *centralidad del vector propio* o *eigenvector*, que sí hace una distinción entre las relaciones. Juntarse con personas bien comunicadas, ya sea por su grado o por su posición dentro de la red, da mayor peso a la centralidad que juntarse con personas mal comunicadas.

Para calcular la centralidad del vector propio se empieza por dar a cada vértice i una centralidad inicial $x_i = 1$ para cada i en $[n]$, donde $[n] = \{1, 2, \dots, n\}$. Es decir, al principio todos los vértices son iguales. A partir de la centralidad inicial se calcula

²²En inglés *rich gets richer*.

²³La traducción del versículo se obtuvo en [1].

una nueva x'_i en la que se comienzan a diferenciar, según sus relaciones con los otros vértices. La nueva centralidad se define como:

$$x'_i = \sum_{j=1}^n A_{ij}x_j, \quad (1.1)$$

donde A_{ij} es la entrada ij de la matriz de adyacencias. Si x es el vector cuya i -ésima entrada es x_i , podemos escribir la ecuación (1.1) en notación matricial como $x' = Ax$, donde x' y x son vectores columna. Si se repite este proceso de manera recursiva se obtiene en el paso t la expresión:

$$x^{(t)} = A^t x. \quad (1.2)$$

La i -ésima entrada de $x^{(t)}$ es la centralidad correspondiente al vértice i .

Sean $\kappa_1, \kappa_2, \dots, \kappa_n$ los valores propios de la matriz de adyacencias (ordenados de manera decreciente) y denótese por v_i al vector propio correspondiente a κ_i . Si se escribe a $x = x^{(0)}$ como combinación lineal de los eigenvectores:

$$x^{(0)} = \sum_{i=1}^n c_i v_i, \quad (1.3)$$

donde c_i es una constante para cada $i \in [n]$. Ahora, se puede reescribir la ecuación (1.2) en términos de los vectores propios como:

$$x^{(t)} = A^t \sum_{i=1}^n c_i v_i = \sum_{i=1}^n c_i \kappa_i^t v_i = \kappa_1^t \sum_{i=1}^n c_i \left(\frac{\kappa_i}{\kappa_1} \right)^t v_i. \quad (1.4)$$

Puesto que κ_1 es el valor propio más grande, se tiene que $\kappa_i/\kappa_1 < 1$ para cada $i \neq 1$. Así, cuando se hace tender t a infinito, todos los términos de la suma salvo el primero tienden a cero. Entonces, el límite de $x^{(t)}$ cuando t tiende a infinito es $c_1 \kappa_1^t v_1$. De este modo, el límite del vector cuyas entradas son las centralidades de los vértices, es proporcional al vector propio correspondiente al valor propio principal. Se observa que cambiando las centralidades iniciales (y por tanto el vector de centralidades inicial $x^{(0)}$), el resultado sólo cambia en la constante c_i , que depende del vector de inicio. Entonces el vector de centralidades que se toma es aquel que satisfaga:

$$Ax = \kappa_1 x \quad (1.5)$$

1.2. Definiciones, medidas y nomenclatura

y así la centralidad correspondiente al vértice i es:

$$x_i = \kappa_1^{-1} \sum_{j=1}^n A_{ij} x_j. \quad (1.6)$$

En este tipo de centralidad un nodo puede ser importante, ya sea porque tiene muchas relaciones, o bien, porque sus contactos están bien conectados.

La *centralidad pagerank* es otra medida que pone en consideración la calidad de las relaciones, normalmente está definida para redes dirigidas, pero se usó una adaptación para el caso no dirigido pues la red que se estudia en esta tesis es no dirigida. Esta medida fue introducida por Google en 1999 para determinar, en una búsqueda en Internet, qué páginas son las más importantes de acuerdo al tema y mostrar los resultados de manera jerarquizada [32].

La definición de la centralidad *pagerank* está dada por:

$$x_i = \alpha \sum_{j=1}^n A_{ij} \frac{x_j}{k_j^{ex}} + \beta, \quad (1.7)$$

La manera de adaptar la definición al caso no dirigido es considerar que una arista entre dos vértices es equivalente a tener el par de arcos dirigidos entre esos vértices en las dos direcciones posibles (*arcos dirigidos simétricos*), y así podemos sustituir k_i^{ex} en la ecuación por el grado (no dirigido) del vértice i , k_i .

El problema con esta definición es que si algún vértice tiene exgrado (respectivamente, grado) cero, entonces tenemos una división indefinida (este no es el caso aquí, porque únicamente se analiza la componente conexa más grande de la red y todos los vértices en ella tienen grado positivo). Sin embargo, en la literatura se hace una corrección de esta situación definiendo una matriz diagonal D de $n \times n$ cuyas entradas no nulas estén dadas por $D_{ii} = \max\{k_i^{ex}, 1\}$ (resp. $D_{ii} = \max\{k_i, 1\}$), un vector de $n \times 1$ con todas sus entradas iguales a 1 y es denotado por $\mathbf{1}$. Entonces se define la centralidad en su forma matricial como:

$$x = \alpha AD^{-1}x + \beta\mathbf{1}, \quad (1.8)$$

donde la i -ésima entrada de x es la centralidad correspondiente al vértice i , x_i , y α y β son dos constantes. Observe que si el exgrado (resp. grado) del vértice j es cero, entonces $A_{ij} = 0$ ²⁴, por lo que $A_{ij}/\max\{k_j^{ex}, 1\} = 0$ (resp. $A_{ij}/\max\{k_j, 1\} = 0$) y

²⁴En el caso dirigido $A_{ij} = 1$ si hay un arco dirigido que inicia en j y termina en i .

no aporta nada a las centralidades de los otros vértices, que es lo que se espera de un vértice que no apunta a otros.

Al despejar x de la ecuación (1.8), se obtiene la expresión $x = \beta(I - \alpha AD^{-1})^{-1}\mathbf{1}$. Escrita de esta forma podemos ver que β no contribuye en esencia a la centralidad pues es sólo una constante que multiplica a cada término, por esta razón suele considerarse que $\beta = 1$, y se obtiene la ecuación:

$$x = (I - \alpha AD^{-1})^{-1} \mathbf{1} = D(D - \alpha A)^{-1} \mathbf{1} \quad (1.9)$$

Antes de hacer el cálculo se debe elegir un parámetro α , el cual debe ser menor que $1/\kappa_1$ donde κ_1 es, como antes, el eigenvalor más grande de la matriz AD^{-1} , que en el caso no dirigido es 1 y su eigenvector correspondiente es (k_2, k_3, \dots) . Entonces para la red usaremos $\alpha < 1$.

La centralidad *pagerank* se puede generalizar suponiendo que el término constante en la ecuación (1.7) es diferente para cada i , es decir x_i estará dada por:

$$x_i = \alpha \sum_{j=1}^n A_{ij} \frac{x_j}{k_j^{ex}} + \beta_i, \quad (1.10)$$

o en la forma matricial:

$$x = D(D - \alpha A)^{-1}\beta, \quad (1.11)$$

donde $\beta = (\beta_1, \beta_2, \dots, \beta_n)$.

En adelante llamaremos *nodos dominantes* a los vértices que tengan la mayor centralidad *pagerank*.

Ahora introduciremos una medida de centralidad que decide qué tan bien está relacionado un nodo por la distancia media de este a cualquier otro nodo en la red, de ahí el nombre *grado de cercanía*²⁵. Una primera aproximación diría que para encontrar la distancia media de un nodo i , ℓ_i , a cualquiera de los otros, basta calcular la media aritmética de las distancias, d_{ij} , del vértice i a cada vértice j de la red, es decir:

$$\ell_i = \frac{1}{n} \sum_{j=1}^n d_{ij}. \quad (1.12)$$

Sin embargo, el término d_{ii} no contribuye en nada a la suma, pues la distancia de un vértice a sí mismo es nula. Por lo que, en otros textos, a veces se define a la distancia

²⁵En inglés *closeness centrality*.

1.2. Definiciones, medidas y nomenclatura

media como:

$$\ell_i = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n d_{ij}, \quad (1.13)$$

lo cual no afectará la noción de qué tan central es un nodo, ya que todas las centralidades en la ecuación (1.12) quedan multiplicadas por el factor constante $n/(n-1)$ y para determinar qué tan importante es un actor lo que interesa es su centralidad respecto a las de los otros, no por el valor en sí.

Observe que ℓ_i es menor cuando un nodo está bien vinculado y es mayor cuando no lo es. Lo cual tiene sentido porque se puede pensar que un actor que dista menos de los demás es un punto clave en el paso de información, tanto en una posición de emisor, como en una de receptor: como emisor, sus ideas llegarán más rápidamente a los demás, entonces será un actor influyente; como receptor, es más fácil que le lleguen ideas de los otros simplemente por estar cerca, o bien, por ser un nodo de tránsito en rutas más largas, así será un actor “conocedor”.

En este sentido, la distancia media funciona diferente de otras medidas de centralidad, porque en las centralidades que se han definido hasta ahora, se considera que son nodos centrales aquellos cuya centralidad es mayor, aquí funciona exactamente al revés, los más centrales serán aquellos con distancia media menor. Para definir el grado de cercanía, C_i , se tomarán los inversos de las distancias medias, para obtener así una medida de centralidad como las otras:

$$C_i = \frac{1}{\ell_i} = \frac{n}{\sum_{j=1}^n d_{ij}}. \quad (1.14)$$

El problema que tiene esta medida es que, para redes conexas, la diferencia entre la distancia máxima (que es a lo más $n-1$) y la distancia mínima (que es al menos 1) es muy pequeña y así las distancias medias, y por lo tanto los grados de cercanía, no difieren en mucho. Entonces, cuesta trabajo diferenciar cuáles vértices son centrales y cuáles no, sobretodo para redes grandes, pues la diferencia entre dos centralidades estará en las colas decimales. Como consecuencia de esta cercanía en los valores de las centralidades, un pequeño cambio en la red (como añadir algún arco) puede afectar drásticamente el orden de las centralidades.

El otro problema con esta medida es cuando se trabaja con una red desconexa, porque así todos los vértices tienen distancia media infinita y C_i igual a cero. Esta situación se resuelve de dos maneras: la primera es considerar por separado cada componente y calcular la centralidad de acuerdo a la componente, pero esta es una

mala solución porque los nodos en las componentes pequeñas tienen distancias medias menores que los nodos en las componentes más grandes y esto es contrario a lo que se quiere, pues lo intuitivo es pensar que los vértices en componentes pequeñas están mal conectados con respecto a la red; la segunda posibilidad es redefinir el grado de cercanía como el promedio de los inversos de las distancias:

$$C'_i = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{d_{ij}}. \quad (1.15)$$

A esta medida se le conoce como *grado de cercanía armónico*. Aquí, si los vértices i y j están en distintas componentes, es decir, $d_{ij} = \infty$, el término $1/d_{ij} = 0$ y por tanto no aporta nada a la suma. Además, cuanto más cerca está j de i , mayor es $1/d_{ij}$, e inversamente, cuanto más alejado está j de i , menor es $1/d_{ij}$. De este modo, los vértices que estén más alejados de los demás tendrán C'_i muy pequeñas y los mejor comunicados tendrán C'_i grandes. Sin embargo, esta definición se usa rara vez.

Hasta ahora se ha definido la distancia media de un vértice fijo, ahora se introducirá la distancia media en la red, ℓ , que será el promedio de las distancias medias:

$$\ell = \frac{1}{n} \sum_{i=1}^n \ell_i = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}. \quad (1.16)$$

Para redes disconexas, esta distancia media tiene, a su vez, los mismos problemas que el grado de cercanía y en consecuencia puede definirse de dos maneras adicionales, una por cada solución que dimos para el grado de cercanía: la primera, sea $\{\mathcal{C}_m\}$ el conjunto de todas las componentes de la red, con $m = 1, 2, \dots$ y n_m el número de vértices en \mathcal{C}_m para cada m . Se define ℓ como sigue:

$$\ell = \frac{\sum_m \sum_{i,j \in \mathcal{C}_m} d_{ij}}{\sum_m n_m^2}, \quad (1.17)$$

la segunda, se toma el promedio de las C'_i :

$$\frac{1}{\ell'} = \frac{1}{n} \sum_{i=1}^n C'_i = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{d_{ij}}, \quad (1.18)$$

o equivalentemente:

$$\ell' = \frac{n}{\sum_{i=1}^n C'_i}. \quad (1.19)$$

1.2. Definiciones, medidas y nomenclatura

Sin embargo, rara vez se usan ℓ' y C'_i .

A continuación, la centralidad intermedia, que mide qué tanto aparece un vértice en las vías de comunicación entre otros vértices bajo las siguientes hipótesis: Se tiene una red por la que fluye algo a través de las aristas de la gráfica, pueden ser mensajes en una red social, mercancías en una red de comercio o alguna otra cosa de acuerdo a la red que se tenga (se trabajará con mensajes pues la red es social); se asumirá que cada par de vértices conectados por una ruta, puede hacer un intercambio de mensajes con igual probabilidad en una unidad de tiempo; y los mensajes viajan de un vértice a otro únicamente por rutas geodésicas, cuando hay más de una geodésica entre dos vértices, se elige alguna de ellas al azar. Si se hacen fluir mensajes entre todas las parejas de vértices durante algún tiempo razonable, dado que los mensajes viajan por las geodésicas a la misma velocidad, el número de mensajes que pasa a través de un vértice es proporcional al número de geodésicas que lo contienen. Se llama *centralidad intermedia*²⁶ al número de rutas geodésicas que pasan por un vértice. Para esta medida, una vez más, tener un valor numérico mayor estará asociado a una mejor posición en la red, ya que si un vértice está en muchas rutas (geodésicas), un mayor número de mensajes pasan a través de él y en consecuencia el vértice adquiere poder respecto de los otros, puede ser un poder de la información porque se le permite saber el contenido de los mensajes o quizá un poder económico porque reciba una cuota por hacerlos fluir. Si se quitan los vértices que tienen una centralidad intermedia mayor, lo que se obtiene es un entorpecimiento en el flujo. Cabe decir, que en la realidad, las tres hipótesis no siempre se satisfacen, sin embargo la medida proporciona una buena aproximación de la influencia que puede tener un vértice en el flujo de la información entre otros vértices.

Para la definición numérica suponga que se tiene una red no dirigida en la que entre cada pareja de vértices existe a lo más una ruta geodésica y considere el conjunto de todas las rutas geodésicas de la red. La centralidad intermedia del vértice i es el número de rutas en dicho conjunto que pasan por i .

Dados tres vértices i , s y t , se define el número n_{st}^i como sigue: n_{st}^i es igual a 1 si el vértice i está en la geodésica entre s y t , y es igual a 0 si i no está en la geodésica entre s y t , o bien, si s y t están en distintas componentes y por lo tanto no hay geodésicas entre ellos. La centralidad intermedia x_i está dada por:

$$x_i = \sum_{st} n_{st}^i. \quad (1.20)$$

²⁶En inglés *betweenness centrality*.

En esta definición, para redes no dirigidas, se cuenta cada geodésica por un vértice dos veces, una en cada dirección, por lo que en algunos libros se divide entre 2 dicho valor, lo cual no cambia el orden de magnitudes entre las centralidades de los vértices. Sin embargo, la expresión en (1.20) es más general pues puede usarse para redes dirigidas donde n_{st}^i no necesariamente es igual a n_{ts}^i . Otra variación que se puede encontrar en la literatura, es que no se cuentan las rutas de un vértice en sí mismo, y así $x_i = \sum_{s \neq t} n_{st}^i$, lo cual hace disminuir en 1 todas las centralidades calculadas con la ecuación (1.20), por lo que el orden de magnitudes permanece invariante.

La última condición que puede variar es si consideramos que s y t pertenecen a una geodésica entre ellos. En los libros de redes sociales se suelen incluir dichos vértices, porque se considera que los vértices s y t tienen control sobre la información que sale o llega a ellos, pero en sociología se suelen excluir, lo cual hace disminuir el número obtenido, para cada vértice en una componente conexa dada, en un valor constante igual a dos veces el número de vértices en la componente. Observe que dentro de cada componente el orden de las magnitudes es invariante. Cuando el orden de magnitudes cambia vértices en distintas componentes, no se le da importancia, ya que no es de interés comparar vértices que tienen ámbitos de ajenos.

Hasta ahora se ha considerado que entre dos vértices existe una única ruta geodésica, sin embargo esto en general no se cumple para redes arbitrarias. La manera natural de extender la definición de centralidad intermedia a una red con múltiples rutas geodésicas, es darle a cada una de las rutas geodésicas entre dos vértices dados un peso igual a 1 entre el número de rutas geodésicas entre dicho par de vértices, es decir, elegimos una ruta entre dos vértices con probabilidad uniforme. Así, se definirá la centralidad intermedia de un vértice como la suma de los pesos de todas las rutas geodésicas que pasan a través de él, donde dos o más de ellas pueden tener más de un vértice en común.

Para expresarlo matemáticamente, denote por n_{st}^i al número de rutas geodésicas que hay del vértice s al t y que pasan por el vértice i ; y por g_{st} al número total de rutas geodésicas que van de s a t . Así, la centralidad intermedia del vértice i está dada por:

$$x_i = \sum_{st} \frac{n_{st}^i}{g_{st}}, \quad (1.21)$$

donde, por convención, $n_{st}^i/g_{st} = 0$ si tanto n_{st}^i como g_{st} son cero.

De esta manera, volviendo a la idea de la transmisión de mensajes entre parejas

1.2. Definiciones, medidas y nomenclatura

de vértices, x_i es proporcional a la tasa promedio en la que los mensajes pasan a través de i .

La centralidad intermedia difiere de otras medidas, en que no evalúa qué tan bien conectado está un vértice, sino qué tan frecuentemente es un intermediario en las comunicaciones entre otros vértices. Es por esto que hay vértices que tienen valores grandes de centralidad aunque tengan grado pequeño o excentricidad grande. Pues, puede ocurrir que un vértice sea el único vínculo entre dos grupos y así, todas las rutas entre un grupo y otro deban pasar a través de él. Así, su centralidad intermedia será grande, aunque otras medidas puedan desfavorecerle. Sin embargo, el hecho de tener una posición clave en el flujo de información puede darle mucha influencia en la red. A este tipo de vértices se les suele llamar *brokers* en el ámbito sociológico [27].

Los valores de la centralidad intermedia pueden ir desde $2n - 1$ hasta $n^2 - n + 1$: el primer valor se obtiene de la idea de que cada vértice está al menos en todas las rutas que inician o terminan en él, esto ocurre por ejemplo cuando un vértice es una hoja, es decir, está conectado al resto de la red con una única arista, por lo que se cuentan $n - 1$ rutas geodésicas que inician en la hoja y terminan en cada uno de los otros vértices, $n - 1$ rutas geodésicas que inician en uno de los vértices restantes y terminan en la hoja y la ruta de la hoja en sí misma; el segundo valor ocurre para el centro de una *estrella*, que es una gráfica con un vértice (el centro) unido a todos los demás por aristas simples y cualesquiera dos de los otros no son adyacentes, aquí todas las rutas geodésicas entre vértices distintos pasan por el centro, hay $n(n - 1) = n^2 - n$ rutas de estas, y se suma un 1 por la ruta que hay del centro en sí mismo.

Cabe señalar que aún cuando se incorporan nuevos nodos y/o vínculos a una red, los cambios en los valores de la centralidad intermedia rara vez cambian de orden.

Usualmente dentro de las redes se pueden encontrar grupos pequeños de vértices que están más relacionados entre sí, como sucederá en la red de colaboración con la que se trabajará, donde se reflejan los equipos de trabajo donde habrá más vértices conectados. Estas pequeñas comunidades también se estudian en la teoría de redes, por lo que introduciremos nuevas definiciones.

Se llama *clique* o *clan* a un subconjunto de vértices tal que cada pareja de vértices ahí es adyacente y es máximo con esta propiedad, es decir, si se agrega cualquier otro vértice al conjunto habrá una pareja de vértices no adyacentes. Observe que puede haber dos cliques cuya intersección sea no vacía. Un clique en la red del Coloquio VNL representará un grupo de personas que han presentado un trabajo en conjunto,

o bien, que todas ellas han presentado alguna vez un trabajo con cualquiera de los otros y no habrá otro individuo de la red que haya trabajado con todos ellos, de cualquier manera será un grupo que probablemente tenga intereses comunes dentro de la teoría de gráficas.

Cuando la intersección de cliques sea no vacía, veremos que hay personas que no han trabajado juntas (o por lo menos presentado un trabajo en conjunto en el Coloquio VNL) aunque tengan en sus círculos de trabajo muchas personas en común.

Dada una red, se dice que es *transitiva* si satisface que, dados tres vértices, u , v y w , tales que u es adyacente a v y v es adyacente a w , entonces también es cierto que u es adyacente a w . Para que una red sea transitiva se requiere que cada una de sus componentes conexas sea una gráfica completa, cosa que rara vez sucede, por lo que consideraremos una transitividad parcial que es natural en redes sociales, pues es más probable que una persona se vincule con otra si tienen un “amigo” en común, a que se vincule con una persona tomada al azar del conjunto de nodos de la red.

Se dice que una ruta uvw de longitud dos es *cerrada* si u es adyacente a w , es decir si los tres vértices forman un triángulo en la red. En redes sociales, se dice que u , v y w forman una *tríada cerrada*.

Se define ahora el *coeficiente de agrupamiento* o *aglomeración*²⁷, C , como sigue:

$$C = \frac{(\text{número de rutas cerradas de longitud 2})}{(\text{número de rutas de longitud 2})}. \quad (1.22)$$

Si $C = 1$ entonces la red es transitiva y si $C = 0$ entonces la red es libre de triángulos, es decir, puede ser acíclica, o bien, todos los ciclos tienen longitud al menos 4 y así todos los cliques tienen exactamente dos vértices.

Si se observa que en el conteo de trayectorias de longitud 2, las trayectorias uvw y wvu , se cuentan separadamente, así las trayectorias cerradas de longitud 2 se cuentan separadamente por cada vértice de inicio y cada dirección, es decir, seis veces: uvw , wvu , wuv , uvw , uvw y wvu . Por lo que se puede reescribir el coeficiente de agrupamiento como:

$$C = \frac{(\text{número de triángulos}) \times 6}{(\text{número de trayectorias de longitud 2})}, \quad (1.23)$$

El factor de seis proviene de la multiplicidad del conteo de cada triángulo, es contado una vez por cada una de las seis rutas de longitud dos que se pueden formar con sus

²⁷En inglés *clustering coefficient*.

1.2. Definiciones, medidas y nomenclatura

tres vértices.

También se puede definir el *coeficiente de agrupamiento local* para un vértice i , C_i , como:

$$C_i = \frac{\text{número de pares de vecinos de } i \text{ que están conectados entre sí}}{\text{número de pares de vecinos de } i}. \quad (1.24)$$

Cuando dos nodos tienen un vecino común, pero entre ellos no están conectados, decimos que la red contiene un *agujero estructural*. Los agujeros estructurales se pueden percibir como obstáculos o como ventajas, según la perspectiva. Por ejemplo, en el caso de flujos de información un agujero estructural indica que la información tiene menos vías para circular pues hay menos vínculos, lo cual puede ser pensado como una desventaja. Sin embargo, el nodo que es vecino común de dos nodos no adyacentes puede verlo como una ventaja, ya que para que la información fluya entre ese par, es necesario que esta pase a través de él (pensando en una ruta lo más directa posible), por lo que se convierte en un intermediario y adquiere poder sobre el flujo de la información entre los otros dos.

Dentro de una red se puede identificar un tipo de subestructuras llamadas *comunidades*, estas son subgráficas en las que los nodos están muy conectados entre sí y tienen pocos lazos con otros nodos fuera de la subgráfica. Una manera de contar estos grupos es buscando el número de cliques bipartitos²⁸ en la gráfica, lo cual nos lleva a la modularidad en las redes. La *modularidad* es una medida de la proporción con la que dos nodos similares se conectan. Esta proporción será denotada por Q y para calcularla se procede como sigue:

Sea c_i la clase o el tipo de un vértice i , que es un número entero en $\{1, 2, \dots, n_c\}$, donde n_c es el número total de clases. De este modo el número total de aristas que existen entre vértices de un mismo tipo es:

$$\sum_{\text{aristas}(i,j)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j), \quad (1.25)$$

donde la primera suma es sobre todas las parejas no ordenadas de vértices adyacentes (distintos), $\delta(m, n)$ es la delta de Kronecker y el factor $\frac{1}{2}$ se debe a que en la suma cada pareja de vértices i, j se cuenta dos veces.

²⁸Un *clique bipartito* es lo que en teoría de gráficas se conoce como *gráfica bipartita completa*, que es una gráfica, $K_{n,m}$, cuyo conjunto de vértices está dividido en dos subconjuntos, uno con n vértices y otro con m vértices, tiene todas las aristas posibles de un subconjunto al otro y ninguna arista con ambos extremos en el mismo subconjunto.

Ahora se calculará el número esperado de aristas entre los vértices, si las aristas se colocan al azar. Considere una arista dada que incide en un vértice i , cuyo grado es k_i . Observe que en la gráfica total hay $2m$ extremos de aristas, ya que hay m aristas, y las posibilidades de que el otro extremo de esta arista sea uno de los k_j extremos conectados al vértice j es entonces $k_j/2m$, si las aristas se ponen al azar²⁹. Si se cuenta sobre las k_i aristas que inciden en i , el número esperado de aristas entre todos los pares de vértices de un mismo tipo es:

$$\frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j), \quad (1.26)$$

donde el factor $\frac{1}{2}$ se debe, otra vez, al doble conteo de cada par de vértices.

Ahora tome la diferencia de las ecuaciones (1.25) y (1.26) para obtener una expresión para la diferencia entre el número exacto y el número esperado de aristas en la red que unen vértices del mismo tipo:

$$\frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j) - \frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j). \quad (1.27)$$

Ahora, por convención, se calcula no el número de aristas entre vértices de un mismo tipo, sino la proporción de ellas, que se obtiene de dividir la expresión en la ecuación (1.27) por el número de aristas, m , obteniendo así:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j). \quad (1.28)$$

La modularidad, Q , es estrictamente menor que 1, toma valores positivos si existen más aristas entre vértices semejantes de las que se esperan y toma valores negativos si hay menos.

1.3 Modelos de redes complejas

En este apartado se presentan tres modelos de redes que se han desarrollado de manera teórica, los cuales fueron consultados en [3], con el fin de identificar a cuál de ellos se asemeja la red real con la que se trabajó.

²⁹Las conexiones se hacen al azar pero preservando el grado de los vértices.

1.3. Modelos de redes complejas

Modelos de redes aleatorias (modelo ER)

Una *gráfica aleatoria* es un modelo de una red en el cual algunos parámetros quedan fijos de entrada, mientras que otros son aleatorios.

Los modelos de redes aleatorias fueron introducidos por Paul Erdős y Alfréd Rényi a mediados del siglo pasado en [8, 9, 10] y, posteriormente, en 1995 Michael Molloy y Bruce Reed hicieron un algoritmo de generación de redes en este sentido en [22].

En los modelos de redes aleatorias que van aumentando el número de enlaces entre un número de nodos fijo y en los que no se conoce de qué manera se generan los nuevos enlaces, se plantean desde el principio de que un par de nodos se puede conectar de manera aleatoria con probabilidad p . En el modelo de Erdős-Rényi se construye una gráfica $G_{n,m}$ a partir de un conjunto de n vértices, creando m aristas cuyos extremos son seleccionados al azar en el conjunto de vértices. Una variante de este modelo, consiste en construir una gráfica $G_{n,p}$ a partir de un conjunto de n vértices de manera que cada una de las $\binom{n}{2} = \frac{n(n-1)}{2}$ aristas posibles, está presente con probabilidad p y está ausente con probabilidad $1 - p$, este modelo se debe a Edgar Nelson Gilbert [13]. En ambos modelos se pide que no haya aristas múltiples ni lazos.

Se pueden relacionar ambos modelos de la siguiente manera: Cuando se construye una gráfica $G_{n,p}$, la probabilidad de obtener una gráfica $G_{n,m}$ con n vértices y m aristas está dada por:

$$P(G_{n,m}) = p^m (1 - p)^{\binom{n}{2} - m}.$$

El grado de los vértices de las gráficas $G_{n,p}$ siguen dos comportamientos distintos: en el primero, cuando el número n es pequeño, el grado se distribuye como una variable aleatoria binomial; en el segundo, cuando n es grande, el grado de los vértices se distribuye como una variable aleatoria Poisson, por lo cual cuando n crece mucho, el grado de los vértices va tendiendo a una constante, es decir, casi todos los vértices tendrán el mismo grado si el número de vértices es grande. En este último tipo de gráficas es sencillo calcular el coeficiente de agrupamiento, $\langle C \rangle$, pues la probabilidad de que una arista aparezca es independiente de la presencia de las otras aristas. Para un vértice arbitrario, la probabilidad de que dos de sus vecinos estén conectados entre sí es p . Así, el coeficiente de agrupamiento promedio es:

$$\langle C \rangle = p = \frac{\langle k \rangle}{n},$$

donde $\langle k \rangle$ es el grado promedio de los vértices. Cuando $\langle k \rangle$ queda fijo y n tiende a infinito, el coeficiente de agrupamiento tiende a cero.

Además, para una gráfica conexa de grado promedio $\langle k \rangle$ (que no se comporta como un árbol), se tiene que para cualquier vértice i , el número promedio de vecinos que están a un paso es $\langle k \rangle$. Por lo cual, el número promedio de vértices que están a d pasos de i , sin importar si se forman ciclos, se puede aproximar por $\langle k \rangle^d$. Si se define r_G como el número tal que $\langle k \rangle^{r_G} \simeq n$. Observe que $\langle k \rangle^d$ crece exponencialmente respecto a d , por lo que una gran mayoría de vértices estarán a r_G pasos del vértice i . Así, se puede aproximar el promedio de las longitudes de las rutas con el mínimo número de pasos, $\langle l \rangle$, por r_G como:

$$\langle l \rangle \simeq r_G \simeq \frac{\log n}{\langle k \rangle}.$$

Modelos de redes de mundo pequeño (modelo WS)

En los modelos de mundo pequeño, a diferencia de los de redes aleatorias, que dos vértices tengan un vecino común será significativo para determinar si están o no conectados. Recuerde que, en el modelo $G_{n,p}$, el coeficiente de agrupamiento tiende a cero, es decir, la proporción de triángulos a los que pertenece un vértice es insignificante. En términos sociales, esto significa que si dos personas están conectadas a una tercera, la posible relación entre ellas es completamente independiente de la que tienen con su vecino, lo cual es falso en la vida real pues muchas relaciones nuevas se hacen a través de amigos, familiares o compañeros de trabajo. En los modelos de mundo pequeño se procurará que esta manera de vincularse sea significativa.

En 1998, Duncan Watts y Steven Strogatz [35] propusieron un *modelo de mundo pequeño*³⁰ al que se denominará *modelo WS*. Para este modelo se inicia con el *modelo circular* que es una estructura circular de n vértices, en la que cada vértice tiene grado $c = 2m$, sus vecinos son los m vértices anteriores y los m vértices posteriores en el círculo, como se muestra en la figura 1.3 (i). Después, se recorren las aristas a lo largo de la estructura circular y, con probabilidad p , se cambia uno de los extremos de la arista por un vértice elegido al azar de manera que no se creen lazos ni aristas múltiples, o bien, con probabilidad $1 - p$, la arista permanece como está. A las aristas obtenidas de este proceso se le conoce como *atajos*³¹, pues las nuevas aristas conectan vértices en distintos extremos del modelo circular, como en la figura 1.3 (ii).

³⁰Título original: *small-world model*.

³¹En inglés *shortcuts*.

1.3. Modelos de redes complejas

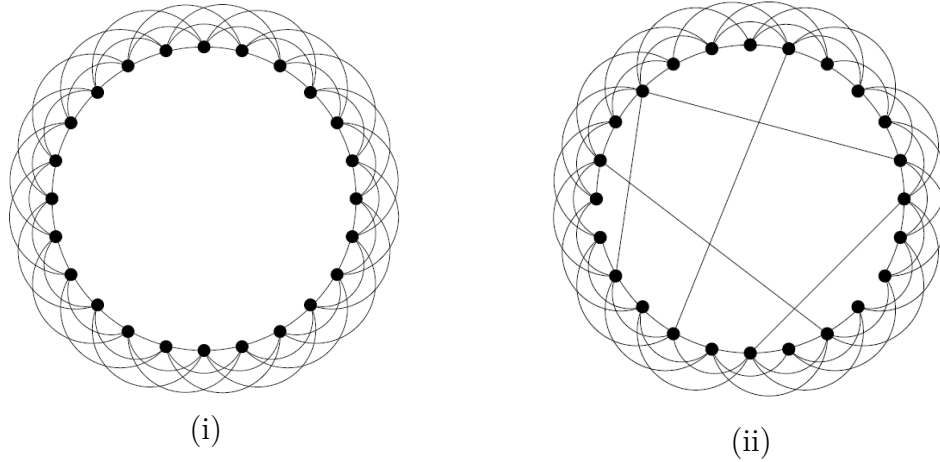


Figura 1.3: En (i) se muestra el modelo circular con $n = 24$ y $2m = 6$. En (ii) se muestra la gráfica obtenida de cambiar aristas al azar en el modelo circular con $n = 24$, $2m = 6$ y $p = 0.07$. Imágenes obtenidas de [27].

Cuando $p = 0$ ninguna arista cambia, por lo que la gráfica que se obtiene es el modelo circular y cuando $p = 1$ todas las aristas cambian, por lo que la gráfica obtenida es una gráfica aleatoria. Cuando p es un valor en el intervalo $(0, 1)$ la gráfica obtenida es una gráfica intermedia entre esos mundos. Por lo cual el modelo de WS es una interpolación entre el modelo circular y una gráfica aleatoria. Cabe señalar que el número de aristas en la gráfica obtenida por este proceso es igual al número de aristas de la original, por lo cual el grado promedio de un vértice en la gráfica nueva es $\langle k \rangle = 2m$.

La distribución del grado para este modelo está dada por la ecuación:

$$P(k) = \sum_{n=0}^{\min(k-m, m)} \binom{m}{n} (1-p)^n p^{m-n} \frac{(pm)^{k-m-n}}{(k-m-n)!} e^{-pm}, \text{ para } k \geq m.$$

En el límite cuando p tiende a 1, se obtiene la siguiente expresión:

$$P(k) = \frac{(m)^{k-m}}{(k-m)!} e^{-m},$$

misma que corresponde con una distribución Poisson para la variable $k' = k - m$, cuyo valor promedio es $\langle k' \rangle = m$.

Cuando $p = 0$, este modelo tiene coeficiente de agrupamiento no nulo si $2m > 2$

pero no tiene *efecto de mundo pequeño*³². Mientras que, cuando $p = 1$, hace lo contrario. Pero, para cierto intervalo de valores de p , se tienen tanto un coeficiente de agrupamiento grande como un comportamiento de mundo pequeño, este último quiere decir que el promedio de las longitudes de las rutas con el mínimo número de pasos, $\langle l \rangle$, es pequeño y se aproxima a $\log n$.

La hipótesis que se plantea es que la red del Coloquio VNL se comportará como un modelo de mundo pequeño. Esta suposición tiene sentido puesto que, en los trabajos colaborativos, muchas veces ocurre que se trabaja con colegas de nuestros colegas. Se toma por ejemplo mi caso: en mi tesis de matemáticas tuve dos tutores, Ilán A. Goldfeder y Hortensia Galeana. Mi vínculo con Ilán se produjo por nuestras relaciones con Hortensia, yo tomé muchos cursos con ambos pero le pedí a Hortensia que fuera mi asesora de tesis e Ilán era su estudiante de doctorado, por lo que me propusieron trabajar juntos, formando así un nuevo triángulo en la red (sin embargo, esta manera de vincularse no es única, siempre se pueden crear lazos con personas que no están relacionadas con un contacto directo, como ahora que trabajo con Bibiana en esta tesis, que hasta donde sé no estaba a dos pasos de mí, por lo menos en la red del Coloquio VNL). Entonces el comportamiento de los vínculos se asemeja a esta estructura circular con muchos triángulos y vínculos que cruzan de un lado a otro la estructura que representan esas relaciones imprevisibles.

Modelos de redes de libre escala (modelo BA)

Los modelos de redes de libre escala, modelan los mecanismos por los cuales se crean las redes. Lo que se hace es suponer qué hace que una red surja para ver qué tipo de estructura se produce con dicha hipótesis. De esta manera se puede suponer que si la estructura obtenida se asemeja a la de una red real, entonces dicha red real podría haber surgido a partir de un mecanismo parecido. El modelo mejor conocido de este tipo es el de *enlace preferencial*³³ que se debe a Albert-László Barabási y Réka Albert [2].

En este modelo la red crece, se agregan vértices uno a uno y cada vértice se conecta con un conjunto de vértices ya existentes, elegidos con ciertas condiciones. Las conexiones serán no dirigidas y el número de lazos con los que se conecta cada

³²El promedio de las longitudes de las rutas geodésicas es pequeño, es decir, la mayoría de las parejas de vértices están a distancia geodésica pequeña.

³³En inglés *preferential attachment*. En los años 60, Price lo llamaba *cumulative advantage*, sin embargo el término de Price cayó en desuso después de las aportaciones de Barabási y Albert.

1.3. Modelos de redes complejas

vértice nuevo es c . La manera de elegir sus vecinos es al azar, donde cada vértice existente es elegido con una probabilidad proporcional al grado que tiene hasta ese momento, lo cual propicia la aparición de *hubs*. Cabe señalar que se agregan vértices y se agregan aristas, pero nunca se quitan, por lo cual todos los vértices tienen grado al menos c y en cada paso el grado mínimo es c , pues al menos el último vértice agregado tiene dicho grado. Las redes obtenidas a partir de este procedimiento tienen una distribución de grado que sigue una ley de potencia de la forma $P(k) \sim k^{-\gamma}$ donde $2 < \gamma < 3$.

En este modelo la longitud de la ruta con el mínimo número de pasos promedio, $\langle l \rangle$, se aproxima a $\frac{\log n}{\log \log n}$, donde n es el número final de nodos. Esta longitud es pequeña y es consecuencia de la existencia de *hubs*, que reducen las distancias.

Capítulo 2

Metodología, análisis y resultados

En este capítulo se explica cómo se construyó la red del Coloquio VNL, y se exponen sus características cualitativas y cuantitativas para hacer una interpretación de las mismas, lo cual contribuye a conocer cómo se relaciona esta población.

2.1 Metodología

En cada edición del Coloquio VNL se dan entre 35 y 50 pláticas, las cuales pueden ser de trabajos individuales o en colaboración con otras personas. Para tener una representación más significativa de quiénes participan en la actualidad, se decidió tomar el periodo de los últimos seis años, pues en un año puede haber faltas de personas que se presentan regularmente.

Para construir la red, se comenzó formando una gráfica en la que se pone un nodo por cada persona que ha participado, ya sea como expositor o bien como coautor en una plática, en alguna de las seis ediciones del Coloquio VNL con las que se trabajó, donde dos nodos (personas) son adyacentes si han colaborado en un proyecto que se haya presentado como plática (en alguna de las seis ediciones del Coloquio VNL). Se tomaron todos los datos de ponentes y coautores de las páginas web de los seis coloquios. Se usaron: para 2013, los archivos de las presentaciones que se encuentran en [36]; para 2014, los resúmenes en [37]; para 2015, las memorias del Coloquio en [38]; para 2016, los resúmenes en [39]; para 2017, los resúmenes contenidos en el programa en [40], y; para 2018, los resúmenes en [41].

Como resultado se obtuvo una gráfica no dirigida, sin pesos en las aristas, en la que dos personas están relacionadas si trabajaron juntas al menos una vez y sin

distinción si lo hicieron más veces. Cabe señalar que si los participantes colaboraron en proyectos que se publicaron o presentaron en otro sitio, las relaciones no se vieron reflejadas en la gráfica, pues únicamente se consideró lo que ocurrió en los coloquios de 2013 a 2018.

La gráfica que se obtuvo en principio, consistía de 268 nodos en 36 componentes conexas. Para esta tesis se trabajó únicamente con la componente más grande que contenía 194 nodos, a la que llamaremos R .

Se introdujeron la tabla de nodos y la tabla de adyacencias al *software* de Gephi y ahí se calculó para la red: el grado medio, el diámetro y la modularidad. Además, para cada nodo se obtuvo: la excentricidad, el coeficiente de agrupamiento, el número de triángulos a los que pertenecía, el grado de cercanía, el grado de cercanía armónico y las centralidades *pagerank*, intermedia y de eigenvector.

Es de interés saber si R es una gráfica robusta, es decir, si la mayoría de los vértices siguen formando una comunidad conexa cuando los vértices centrales desaparecen. Para verificarlo se usaron, independientemente, dos parámetros: la centralidad *pagerank* y la centralidad intermedia. Se probó quitando los diez nodos más centrales de acuerdo a estas medidas, sin embargo no se quitaron los diez nodos de una vez. Se eliminó el nodo o el conjunto de nodos (si un conjunto tenía exactamente las mismas medidas) con el valor más alto en una de estas medidas, luego se volvió a calcular el valor de dicha medida para todos los nodos en la nueva gráfica y se quitó el nodo o el conjunto de nodos con la medida más alta en la componente gigante de la nueva gráfica. Se siguió así hasta haber eliminado diez nodos que es, aproximadamente, el 5% de los nodos de R . El motivo por el cual se eliminaron los vértices de este modo y no todos a la vez es que, si uno de los actores más centrales hubiera faltado (por un sabático, jubilación u otra razón), entonces las medidas de los demás nodos habrían cambiado y el orden de los nodos más centrales no necesariamente se habría preservado tras modificar la red.

En el caso de la centralidad intermedia es clara esta diferencia como veremos más adelante.

En la sección 2.2 se exponen los datos que se obtuvieron con Gephi de la red de colaboración de las ponencias del Coloquio VNL. Para que algunas de las interpretaciones queden más claras se explica con un ejemplo qué significan. El ejemplo consiste en lo siguiente: En la próxima emisión del Coloquio VNL, en lugar de enviar una convocatoria abierta, se le enviará una invitación individual a una persona, seleccionada al azar, de la lista de ponentes de 2013 a 2018, se le pedirá que com-

2.2. Análisis y resultados

parta la invitación, de manera anónima, únicamente con sus colaboradores de las seis emisiones de 2013 a 2018. A cada persona que reciba la invitación se le pedirá que proceda de la misma forma¹ hasta agotar el cupo (que se aclarará enseguida).

El Coloquio VNL tiene un margen de asistentes con los que debe contar. Al menos debe haber 35 pláticas para llevarse a cabo (que es el mínimo número que se ha dado en los últimos años) y dado que los auditorios tienen un cupo limitado, podrán asistir a lo más 110 personas (o un número menor, lo más cercano posible a 110, de manera que todos los nodos que reciben invitación en el mismo tiempo hacen la misma acción: enviar o no enviar la invitación²). Se supondrá el caso ideal en el que todos los nodos que reciban la invitación harán el reenvío correspondiente y asistirán al congreso. Cabe señalar que, para modelar la realidad, habría que considerar la probabilidad de que alguien con invitación no la envíe, pues es fácil que alguna persona no pase la invitación: por distracción, por algún imprevisto o por un error en la comunicación, puede que la invitación quede atascada en algún nodo.

2.2 Análisis y resultados

La red del Coloquio VNL, tiene 268 nodos repartidos en 36 componentes (figura 2.1), de las cuáles sólo una de ellas tiene más de 20 nodos. Por lo cual, el Coloquio VNL se podrá realizar únicamente si el vértice seleccionado pertenece a la componente gigante. Se supondrá que este es el caso y más adelante se verá qué podría ocurrir si algunas personas están ausentes por algún imprevisto.

La componente gigante de la red del Coloquio VNL, R , tiene 194 nodos y 432 aristas y así su densidad es 0.0236, que es el 2.36% de los enlaces que podrían existir entre los integrantes de R . En la teoría de redes una densidad baja, como esta, indica que la red estudiada no conforma una comunidad cohesionada, refleja desconfianza o falta de apoyo entre los distintos grupos de trabajo y, como consecuencia, una mala fluidez de ideas e información. Así, la transmisión de la invitación marchará con dificultad dentro de la red. Sin embargo, cabe señalar que esta es una inter-

¹Esta manera de enviar las invitaciones arroja una lista de invitados igual a la que se obtiene usando el algoritmo de búsqueda exhaustiva BFS (ver anexo B), pues el proceso es semejante salvo porque una persona invitada después, puede enviar una invitación a una persona que ya está enterada, por la cuestión del anonimato. Sin embargo, los nodos enterados serán los mismos que los visitados por el algoritmo BFS y la primera persona de quien alguien reciba la invitación corresponderá con el vértice desde el cual el algoritmo llega a explorarlo.

²En el árbol producido por el algoritmo BFS, significa que todos los nodos en un mismo nivel (a la misma distancia de la raíz) estarán en la misma condición: invitados o no invitados.

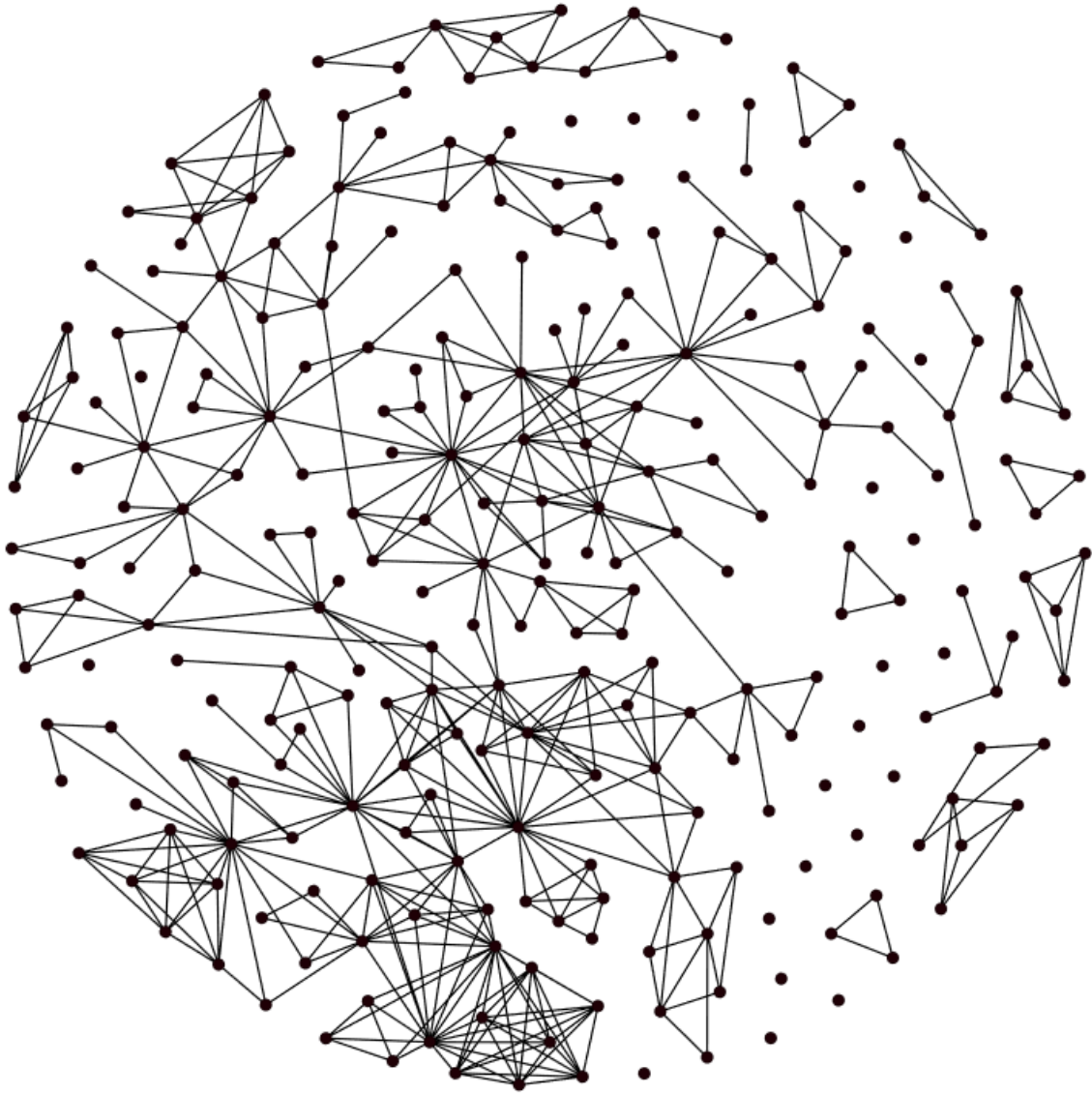


Figura 2.1: La red del Coloquio VNL de 2013 a 2018

2.2. Análisis y resultados

pretación parcial pues la red R considera únicamente las relaciones de colaboración en las ponencias que se presentaron en el Coloquio VNL entre 2013 y 2018. En la comunidad del Coloquio VNL existen otros vínculos de comunicación que no se están considerando, como las amistades, los intercambios que se pueden dar en seminarios, las conversaciones entre los miembros del comité organizador y los comités locales, las relaciones de colaboración en trabajos que no se presentaron en el Coloquio VNL u otros.

El grado medio en R es de 4.454, es decir, en promedio cada uno de estos participantes ha presentado un trabajo con 4 o 5 personas, ya sea en un proyecto común o en proyectos separados. En el ejemplo esto significa que, en promedio, cada invitado reenviará cuatro o cinco veces la invitación (porque no sabe de quién la recibió).

La modularidad de R es 0.754 que, al ser mayor que cero, indica que hay más aristas de las esperadas entre parejas de nodos similares, por lo que nodos similares se invitarán entre sí. Se pueden identificar diez comunidades cuya representación gráfica está en la figura 2.2. Además, se puede ver cómo se relacionan dichas comunidades en la gráfica de la figura 2.3. A continuación se clasifican los nodos por comunidad³:

Comunidad 0 (beige, densidad local 0.209):

17, 18, 37, 38, 39, 71, 105, 114, 115, 210, 211, 212, 217 y 230.

Comunidad 1 (verde limón, densidad local 0.202):

28, 29, 40, 41, 42, 43, 44, 64, 92, 93, 94, 95, 96, 164, 165, 166, 167, 172, 192, 215, 237, 238, 239, 240, 253, 254 y 255.

Comunidad 2 (naranja, densidad local 0.162):

0, 1, 2, 3, 4, 5, 6, 51, 55, 56, 69, 70, 148, 173, 185, 196, 200, 201, 205, 257 y 267.

Comunidad 3 (azul agua, densidad local 0.167):

72, 74, 75, 82, 119, 120, 128, 129, 135, 161, 162, 163, 168, 193, 197 y 256.

Comunidad 4 (azul claro, densidad local 0.114):

27, 30, 62, 86, 87, 106, 107, 108, 111, 116, 117, 123, 145, 169, 170, 175, 181, 187, 208, 209, 218, 232, 250, 251, 252, 258 y 259.

Comunidad 5 (negro, densidad local 0.17):

7, 8, 9, 10, 19, 31, 32, 33, 34, 35, 36, 90, 109, 133, 134, 137, 138, 147, 171, 182, 183, 184, 190, 191 y 241.

Comunidad 6 (rojo, densidad local 0.7):

67, 68, 139, 140 y 141.

Comunidad 7 (lila, densidad local 0.123):

³En el anexo A se encuentran las gráficas de las diez comunidades con los nodos etiquetados.

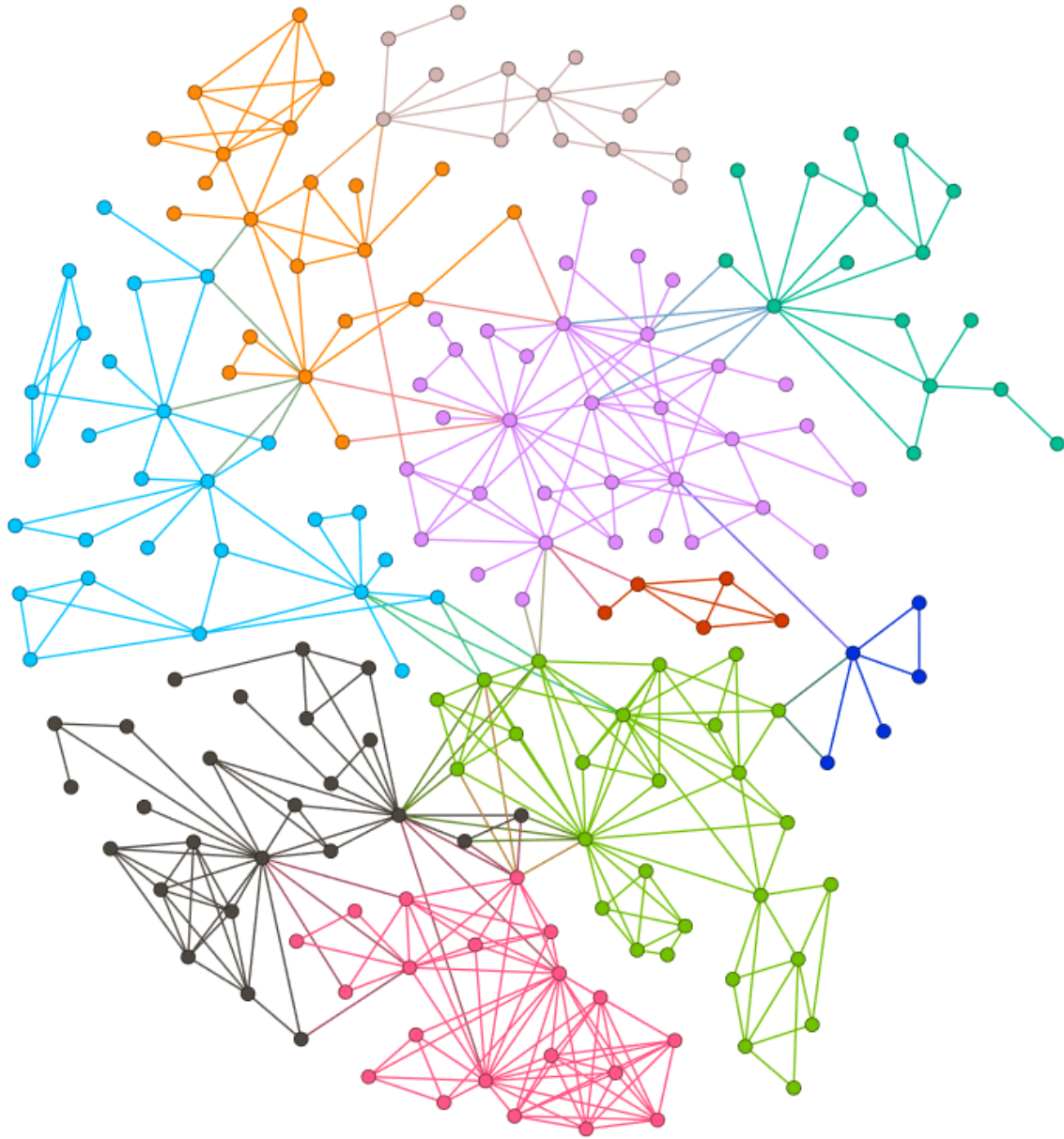


Figura 2.2: Las comunidades de R

2.2. Análisis y resultados

48, 49, 50, 52, 53, 54, 61, 65, 66, 76, 77, 89, 97, 98, 112, 113, 122, 126, 127, 131, 132, 155, 176, 177, 178, 179, 180, 194, 195, 213, 233, 234, 235 y 249.

Comunidad 8 (azul rey, densidad local 0.5):

88, 144, 206, 207 y 216.

Comunidad 9 (rosa, densidad local 0.368):

20, 21, 22, 23, 24, 25, 26, 45, 46, 47, 63, 243, 244, 260, 261, 262, 263, 264, 265 y 266.

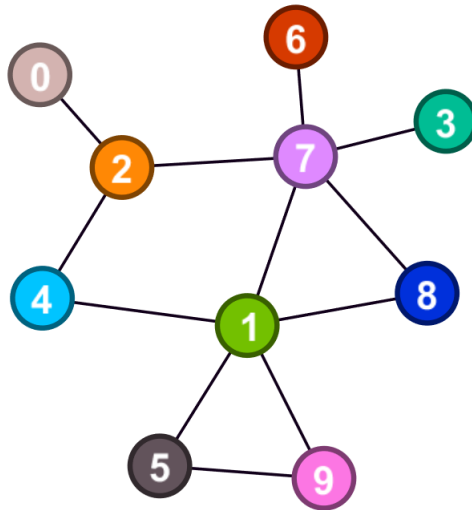


Figura 2.3: Gráfica de las comunidades y sus relaciones. Cada vértice representa a una comunidad y dos comunidades son adyacentes si existe una arista en R que tenga un extremo en cada una de ellas.

Las diez comunidades tienen estructuras muy diversas y tamaños diferentes. Por ejemplo, las comunidades 6 y 8 cuentan apenas con cinco nodos y sus formas son muy diferentes: la comunidad 6 es una gráfica completa de orden 4 a la que se le pega una hoja en uno de sus vértices, mientras que la comunidad 8 consiste de un triángulo al que se le pegan dos hojas en uno de los vértices por lo que, aún cuando ambas comunidades tienen el mismo número de vértices, la comunidad 6 está mejor conectada que la comunidad 8.

En las comunidades más amplias, también hay una diversidad de estructuras. La comunidad 3 tiene un nodo, 75, que es el que la aglutina como comunidad, aquí todos los clanes son de dos o tres vértices, es decir, nunca trabajan más de tres personas juntas y siete de los nueve nodos adyacentes a 75 son o fueron sus alumnos. Las comunidades 0, 3 y 4, son “alargadas” ya que están divididas en pequeños clanes de

dos, tres, cuatro y hasta uno de cinco vértices, que se conectan unos con otros en una estructura de árbol.

La comunidad 5 tiene dos partes, una de ellas consta de dos clanes de seis vértices traslapados en cinco de ellos, que consisten de todos los alumnos del nodo 32 que es el que los comunica con el resto de la comunidad y él, 32, se conecta con el resto de la red a través de uno de sus colegas, 9. La otra mitad de la comunidad está más dispersa en pequeños grupos (poco conectados) adyacentes a 9 o a 32.

La comunidad 7 tiene nodos de grado alto adyacentes entre sí, que producen una comunicación más eficiente entre sus integrantes. Mientras que la comunidad 1 tiene sólo un vértice de grado muy alto, que es el punto de intersección de varios cliques de al menos cuatro nodos, mismos que son la estructura más gruesa de esa comunidad.

La comunidad 9 es la más aglutinada, aquí hay tres cliques medianos de cinco, seis y nueve vértices que tienen en común a los nodos 20 y 21 y tres nodos que cuelgan de uno de los vértices. Esta es la comunidad (de más de cinco vértices) que tiene el menor diámetro, los nodos más alejados aquí están a distancia tres.

En la gráfica de las comunidades (figura 2.3), la comunidad 7 está a distancia a lo más 2 de las otras, las comunidades 1, 2, 3, 4, 6 y 8 están a distancia a lo más tres de las demás y, finalmente, las comunidades 0, 5 y 9 están a distancia a lo más 4 de las demás. Por lo que la comunidad 7 es la más central en esta pequeña red que, casualmente, coincide con la comunidad que contiene tres integrantes del comité organizador en las últimas cinco emisiones del Coloquio VNL⁴.

En general, una vez que un integrante de una comunidad es invitado, en breve los demás nodos de su comunidad estarán enterados pues, en cada comunidad, los vértices están cerca (a distancia a lo más seis). Puede que algún nodo de una comunidad no reciba una invitación aunque otros miembros de la misma sí la reciban, pero esta situación se dará sólo si el cupo se satura antes de que dicho nodo la reciba. Lo que sí es absolutamente seguro es que la comunidad completa del primer nodo seleccionado será invitada al Coloquio VNL.

La red R cuenta con 403 triángulos y un coeficiente de agrupamiento promedio de 0.74, que es muy alto. En la red, este número se puede interpretar como sigue: dado un participante en las pláticas y dos de sus colaboradores, estos son colaboradores

⁴El comité organizador ha variado de integrantes durante estos años, sin embargo, los nodos 54, 65 y 89 han sido constantes en los últimos cinco coloquios. En 2014 estaban además los nodos 3 y 98, en 2015 los nodos 9 y 98, en 2016 y 2017 los nodos 9 y 162, en 2018 los nodos 9, 27 y 71. Mientras que en el coloquio de 2013 el comité consistió de los nodos 37, 38, 210, 217 y alguien que no está en la red.

2.2. Análisis y resultados

entre sí con probabilidad 0.74.

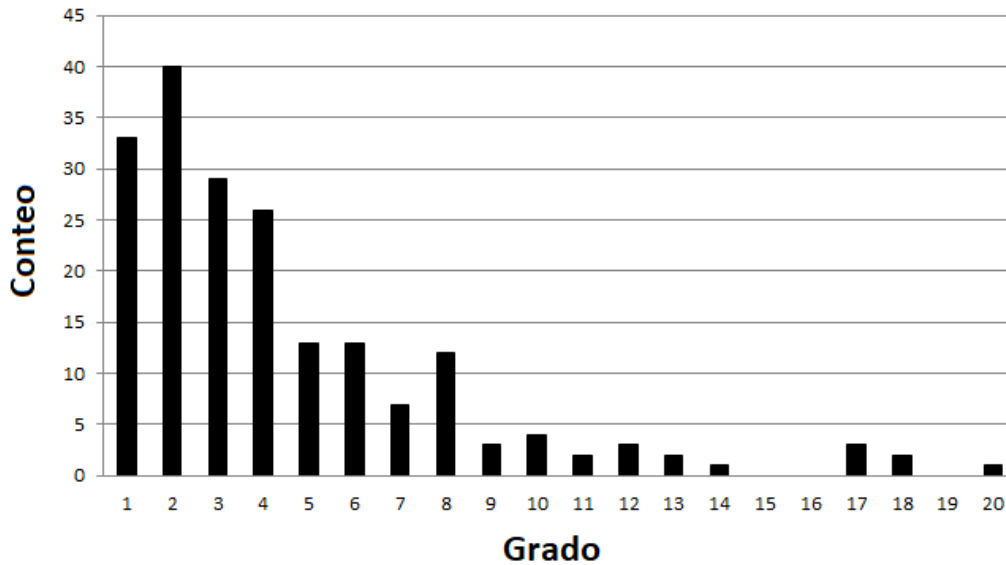


Figura 2.4: Distribución de grado

R tiene diámetro 11 y radio 6, lo cual significa que para comunicar a un par de personas lo más alejadas posible hay diez intermediarios; mientras que una persona lo mejor comunicada posible en el sentido de distancia (que es cercana a todos los demás nodos) requiere de a lo más cinco intermediarios para comunicarse con cualquier otro nodo de R . Además, el promedio de las longitudes de las rutas geodésicas es 5.117515, así que, en promedio, hay cuatro o cinco intermediarios entre dos nodos de R . Se observa que $\log(194) \sim 5.2678$, donde 194 es el número de nodos de la red, de este modo se tiene que el promedio de las longitudes de las rutas geodésicas no difiere mucho de $\log(194)$, por lo que se observa un comportamiento de mundo pequeño (sección 1.3). Así, con un coeficiente de agrupamiento alto y un comportamiento de mundo pequeño, la red del Coloquio VNL tiene las características del modelo de mundo pequeño, que es lo que se esperaba. La distribución de grado (figura 2.4) no sigue la propuesta del modelo WS, puesto que en dicho modelo se hace tender el número de nodos a infinito y aquí, con 194 nodos, no se puede hacer ese análisis.

En la red R el 24.74% de los integrantes son mujeres y el 75.26% son hombres. En la figura 2.5 podemos observar que existen muchos grupos de trabajo en los que no hay mujeres y que casi todas las mujeres que hay están repartidas en las comunidades 2, 3 y 7, cuya mayoría reside en la Ciudad de México (entre las sedes de la UAM

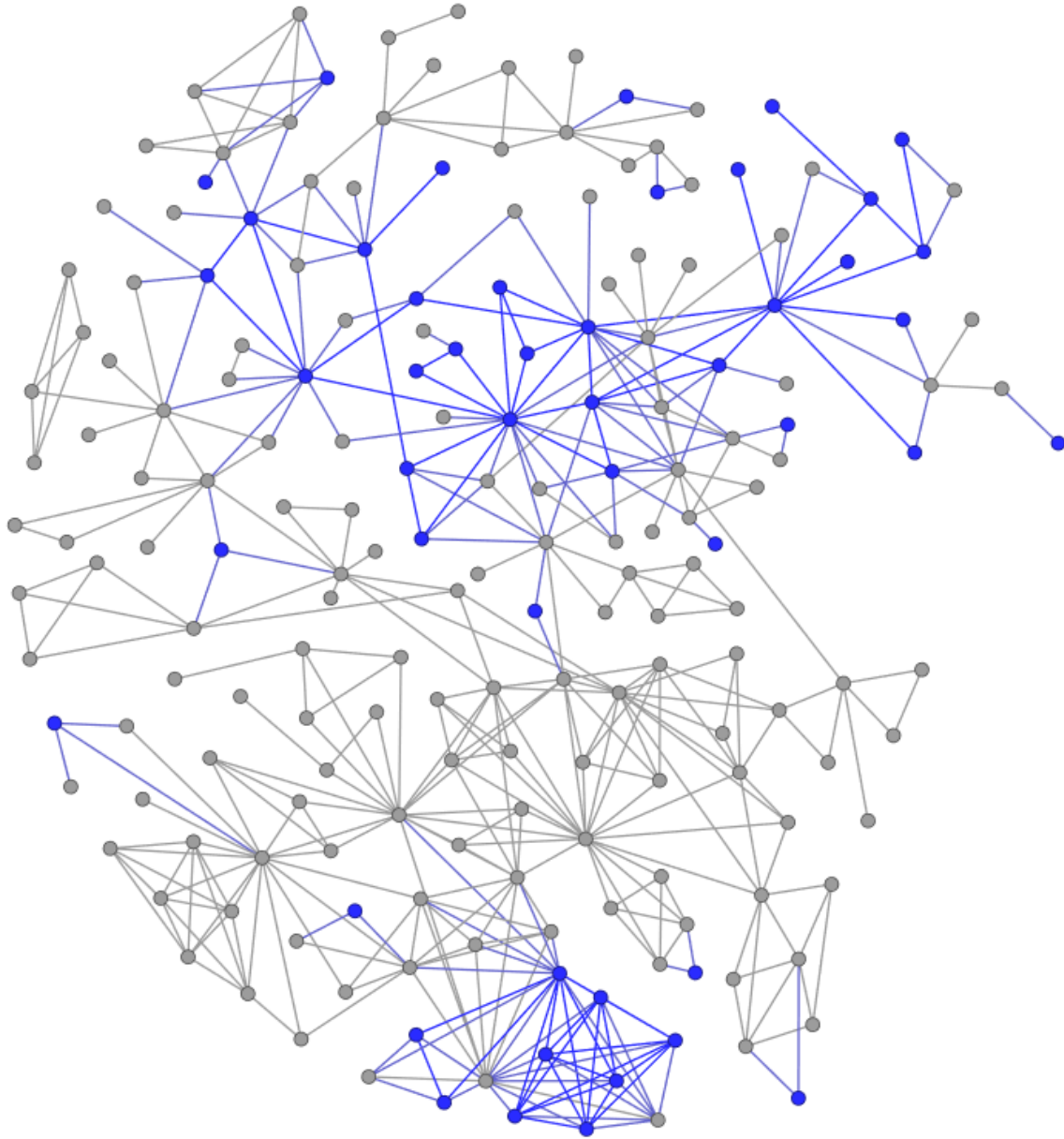


Figura 2.5: Distribución por sexo. En la gráfica las mujeres tienen color azul y los hombres tienen color gris.

2.2. Análisis y resultados

y la UNAM), y en la comunidad 9, residente en Estados Unidos. Mientras que las comunidades que tienen una mayor proporción de nodos en los estados tienen una o dos mujeres. Además, el sector que estudia temas más cercanos a la programación (como comunicación con robots, el problema de la galería, entre otros) no cuenta con ninguna mujer. Por el contrario, los sectores que estudian temas de coloración, dominación e independencia y número de cruce tienen un número balanceado de hombres y mujeres.

Dentro de la red algunas personas tuvieron el rol de ponentes en las ediciones del Coloquio VNL, hay otras que únicamente participaron en la elaboración de los contenidos que se presentaron (figura 2.6) y, algunas veces, los colaboradores no asistieron al Coloquio VNL, hecho que podría ser tomado en cuenta en un ejemplo más complicado del envío de invitaciones: ya sea evitando invitar a quienes no asisten o bien, si se envían las invitaciones a quienes no asisten para no romper el flujo de información, hacer un registro de quién asiste y quién no, para poder aprovechar el cupo lo mejor posible. Sin embargo, en este trabajo no se consideran dichas posibilidades.

En el ejemplo, se tiene que si el primer individuo que recibe la invitación tiene excentricidad mínima, de 6, entonces en seis pasos o tiempos de reenvío de la invitación todos los integrantes de R quedarían invitados, mientras que, si el primer individuo que la recibe tiene excentricidad máxima, de 11, entonces tomaría once tiempos el proceso de informar a todos los nodos de R . Como se tiene la restricción de cupo, habrá quienes no la reciban, pero en cualquiera de los casos quienes no la reciban son los vértices a mayor distancia del primer invitado.

Se calcularon los nodos con el mayor peso de *hub*, el grado para cada vértice y el número de triángulos a los que pertenecían, las medidas de grado de cercanía, grado de cercanía armónico, centralidad intermedia, *pagerank* y del eigenvector. Los valores obtenidos para cada vértice se encuentran en el anexo D. En la tabla 2.1, se organizaron en orden decreciente, para cada tipo de medida, cuáles nodos son los más centrales.

Se puede observar que los nodos 20 y 21 aparecen en las dos primeras posiciones de peso de *hub*, que son los dos vértices que pertenecen a un mayor número de triángulos y cuya centralidad de eigenvector es mayor, además de ocupar las posiciones 4 y 5 en el valor de grado. Estos dos vértices son, además, adyacentes entre sí y comparten comunidad, la número 9. Ambos nodos tienen grado 17 y pertenecen a 51 triángulos. De hecho, las vecindades cerradas de 20 y 21, coinciden, pues $N(20) = \{9, 21, 22,$

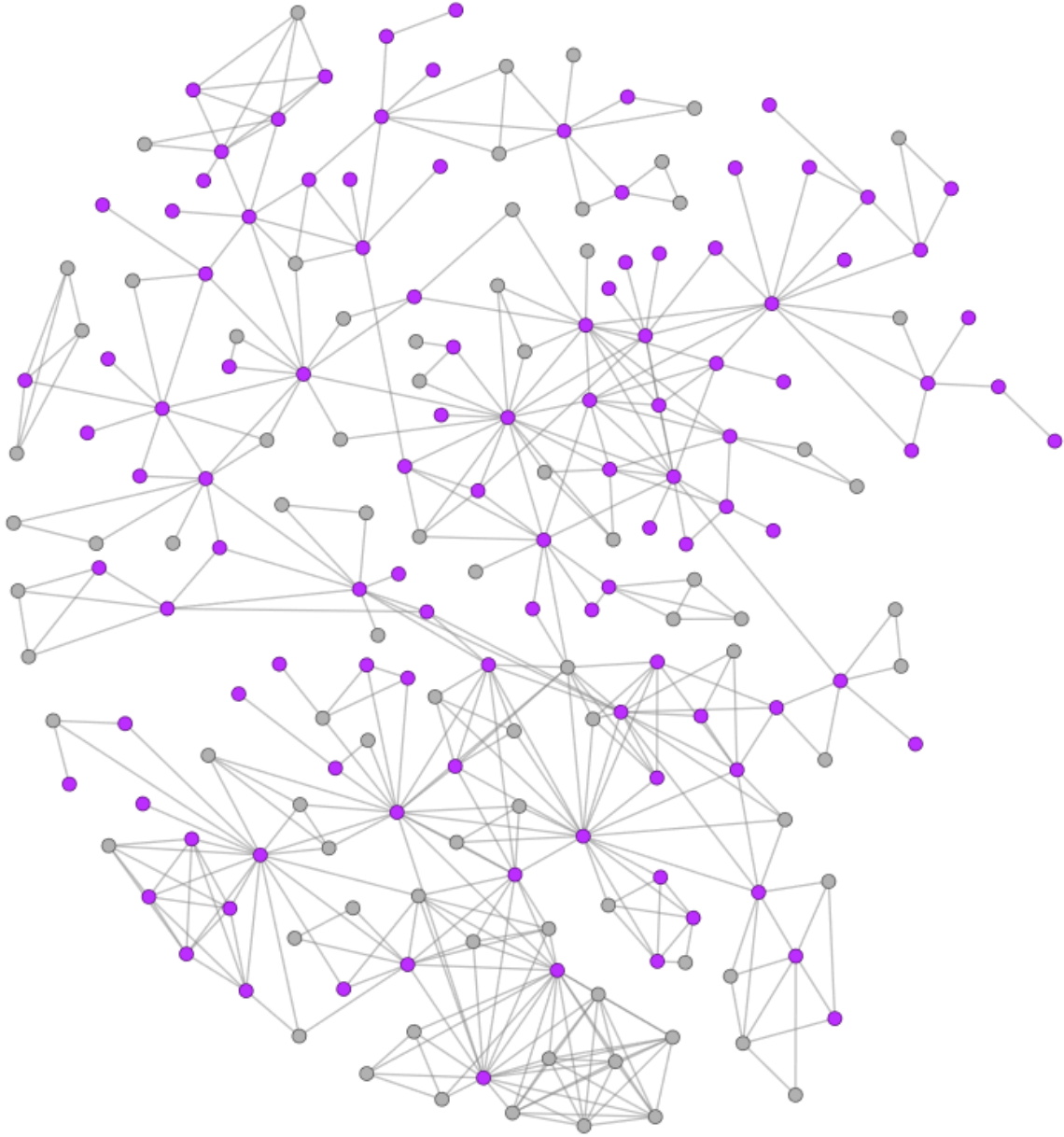


Figura 2.6: Ponentes y colaboradores. En la gráfica los ponentes tienen color morado y los demás tienen color gris.

2.2. Análisis y resultados

23, 24, 25, 26, 45, 46, 47, 260, 261, 262, 263, 264, 265, 266} y $N(21) = \{9, 20, 22, 23, 24, 25, 26, 45, 46, 47, 260, 261, 262, 263, 264, 265, 266\}$, por lo cual $N[20] = N[21]$ y así, dado un triángulo de colaboración en el que uno de ellos está, existen dos posibilidades: i) también el otro está en el triángulo colaborativo, o bien, ii) las otras dos personas del triángulo, junto con 20 y 21, forman un clan de cuatro personas. La pareja de nodos 20 y 21 funciona como una unidad en la red y dicho par forma un conjunto de corte de vértices.

Tabla 2.1: Los 10 nodos más centrales

| | Posición | | | | | | | | | |
|----------------------------|----------|----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Peso de <i>hub</i> | 20 | 21 | 24 | 260 | 261 | 262 | 263 | 264 | 265 | 266 |
| Grado | 42 | 9 | 52 | 20 | 21 | 32 | 29 | 66 | 75 | 24 |
| Excentricidad | 52 | 49 | 177 | 176 | 178 | 66 | 53 | 3 | 43 | 112 |
| Grado de cercanía | 49 | 43 | 52 | 28 | 9 | 54 | 53 | 42 | 29 | 176 |
| Grado de cercanía armónico | 52 | 9 | 42 | 49 | 43 | 28 | 29 | 53 | 3 | 54 |
| Centralidad intermedia | 49 | 43 | 9 | 52 | 3 | 30 | 111 | 196 | 75 | 176 |
| <i>Pagerank</i> | 52 | 75 | 42 | 9 | 32 | 3 | 66 | 53 | 38 | 112 |
| Número de triángulos | 20 | 21 | 42 | 260 | 261 | 262 | 263 | 264 | 265 | 266 |
| Centralidad de eigenvector | 20 | 21 | 9 | 42 | 24 | 25 | 260 | 261 | 262 | 263 |

Si bien los nodos 20 y 21 no son los vértices con mayor grado, sí son los que tienen un mayor peso de *hub* según el cálculo que proporcionó el *software* de Gephi. Esto quiere decir que ambos son adyacentes a vértices que tienen, a su vez, un buen peso de *hub*. Esto se puede corroborar con los datos de la tabla 2.1, en la cual se señala que las siguientes ocho posiciones en peso de *hub* las ocupan los vértices 24, 260, 261, 262, 263, 264, 265 y 266, todos ellos vecinos tanto de 20 como de 21.

Como ya se dijo, los nodos 20 y 21 tienen las dos primeras posiciones en la centralidad de eigenvector. Esta centralidad da un orden de importancia local, por lo que un nodo puede ser central ya sea por tener un número alto de relaciones, o bien porque sus contactos tienen una centralidad de eigenvector grande, lo cual hace que en R los nodos con mayor centralidad de eigenvector sean aquellos con grados altos y que también sean adyacentes a otros nodos con grados altos. En la red del

Coloquio VNL, del total de nodos sólo existen tres personas (9, 42 y 52) con un grado mayor que el de 20 y 21, por lo cual no es de extrañar que tengan una buena posición en esta centralidad y, como ambos colaboraron con el nodo 9 (que tiene el segundo grado más alto y el tercer lugar en la centralidad de eigenvector), su posición queda afianzada, pues se proveen numéricamente en esta centralidad unos a otros.

El nodo número 52 tiene la menor excentricidad en la red R (que es de 6) y así, este actor es el que requiere menos intermediarios para comunicarse con cualquier otro, a lo más necesita cinco de ellos. El nodo 52 tiene el mayor valor en el grado de cercanía armónico y en la centralidad *pagerank*, además de ocupar la tercera posición en el grado y el grado de cercanía, y la cuarta en la centralidad intermedia. Por lo que en términos de varias medidas es un vértice central.

El grado de cercanía, en principio, se calcula como un promedio de las distancias de un nodo a cualquiera de los otros, y el más central es aquel cuya distancia media a los demás sea menor; por lo que, aún cuando 52 es el nodo cuya distancia máxima a cualquier otro es la más pequeña, el nodo 49 tiene un mejor grado de cercanía, lo cual quiere decir que del total de rutas geodésicas de los nodos 52 y 49 a cualquier otro, las que conectan a 49 son, en conjunto, más cortas que las que conectan a 52, en conjunto.

El grado de cercanía armónico es un promedio de los inversos de las distancias de un vértice a cualquiera de los otros. Así, si dos vértices están muy alejados, el recíproco de la distancia es un número muy pequeño. Si un vértice requiere pocos intermediarios para comunicarse con cualquier otro (estará a una distancia pequeña de cualquier otro), entonces los recíprocos aportarán a dicho promedio y será central con esta medida, de ahí que los nodos 49 y 52 tengan dos de las tres primeras posiciones en el grado de cercanía armónico, pues dichos nodos tienen las dos menores excentricidades y posiciones altas en el grado de cercanía.

En la tabla 2.1 de los diez nodos más centrales, se tiene que los nodos 49 y 43, en ese orden, tienen las más altas centralidades intermedia y de grado de cercanía y, en el mismo orden, las posiciones 4 y 5 del grado de cercanía armónico. La centralidad intermedia de un nodo contabiliza el número de rutas geodésicas a las que dicho nodo pertenece, es decir, qué tan frecuente es un intermediario en la comunicación. De estos valores en las tres centralidades podemos decir que, tanto el nodo 49 como el nodo 43, no sólo están a una distancia media pequeña de los demás, sino que muy frecuentemente están en el camino para que otras parejas de nodos se comuniquen. Así, los nodos 49 y 43 tienen una posición privilegiada en la red respecto al flujo de

2.2. Análisis y resultados

información y pueden influir en el curso de esta.

Por ejemplo, si se observa el nodo más grande en la figura 2.7, este corresponde al nodo 49 que está en la comunidad 7 (lila). Dicho nodo es un broker entre la comunidad 6 (roja) y el resto de los nodos de la red R , puesto que es la única vía de comunicación. En el ejemplo, el nodo 49 puede decidir dejar fuera a todos los integrantes de la comunidad 6, le basta con no pasar la invitación en esa dirección. O, si la invitación le llegara desde la comunidad 6, podría no pasarla al resto de la red y, de este modo, dar por cancelado el Coloquio VNL pues no habría suficientes asistentes.

Otro broker en la gráfica es el nodo 210 en la comunidad 0 (beige), que es el único vínculo entre su comunidad y el resto de la gráfica. Él también podría decidir no pasar la información en cualquiera de las dos direcciones, sin embargo, por su posición respecto a las otras comunidades y al resto de la gráfica, es menos probable que la invitación llegue a manos del nodo 210. Por el contrario, el nodo 49 está relacionado con tres comunidades, 1, 6 y 7, y tiene altos valores en varias centralidades, por lo que es altamente probable que la invitación sí llegue a sus manos y pueda tomar este tipo de decisiones.

Es bastante probable que los nodos 43, 49 y 52 reciban la invitación porque, con frecuencia, están en la vía de comunicación entre otros nodos y requieren de pocos intermediarios para comunicarse con quien sea. Así, ya sea por estar a pocos pasos del primer invitado, o bien por ser intermediarios entre el primer invitado y otros nodos, estos tres tendrán fácilmente acceso a la invitación.

La centralidad *pagerank* da un orden de importancia global, un vértice consigue relevancia de sus enlaces de acuerdo a la importancia de sus vecinos ponderada por el grado que tiene cada uno de ellos. De los diez nodos más relevantes de acuerdo a la centralidad *pagerank* en la red R , nueve de ellos tienen dos características en común, son vértices adyacentes a un buen número de vértices dentro de su comunidad y son adyacentes a al menos un vértice de otra comunidad, como se puede ver en la figura 2.8. A diferencia de la centralidad de eigenvector, que da un orden de relevancia local, la centralidad *pagerank* da un orden de relevancia global gracias a la ponderación por el grado que se hace en el cálculo. Dentro de las comunidades muy conectadas, la centralidad de eigenvector crece mucho (como en la comunidad 9) pero no se debe a una buena comunicación con el resto de la red, sino a la cantidad de enlaces dentro de la misma comunidad.

Una manera de entender cómo da relevancia la centralidad *pagerank* es con el

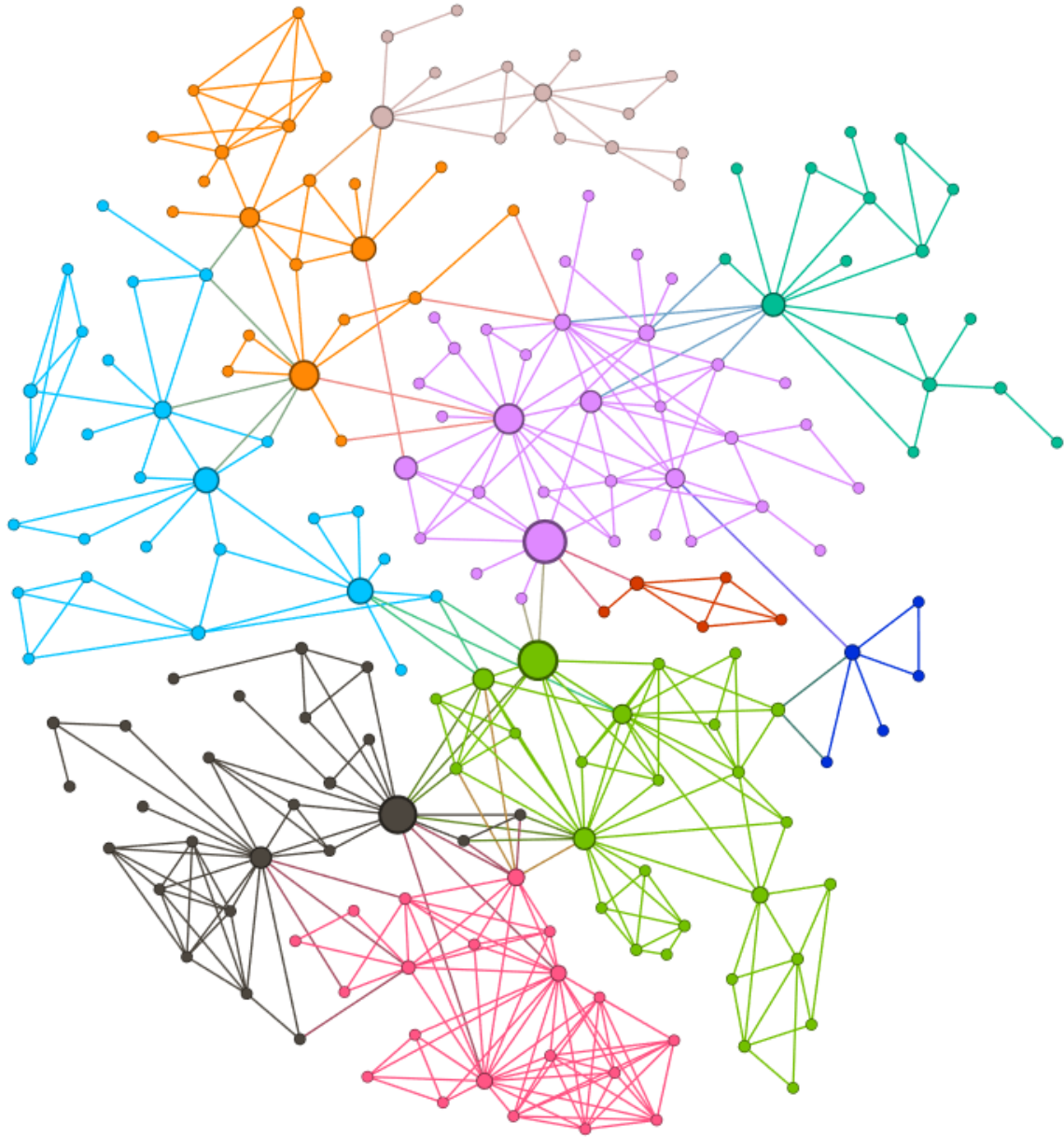


Figura 2.7: Centralidad intermedia. El tamaño de cada nodo es proporcional a su centralidad intermedia, por lo que los más grandes son también los más centrales con esta medida.

2.2. Análisis y resultados

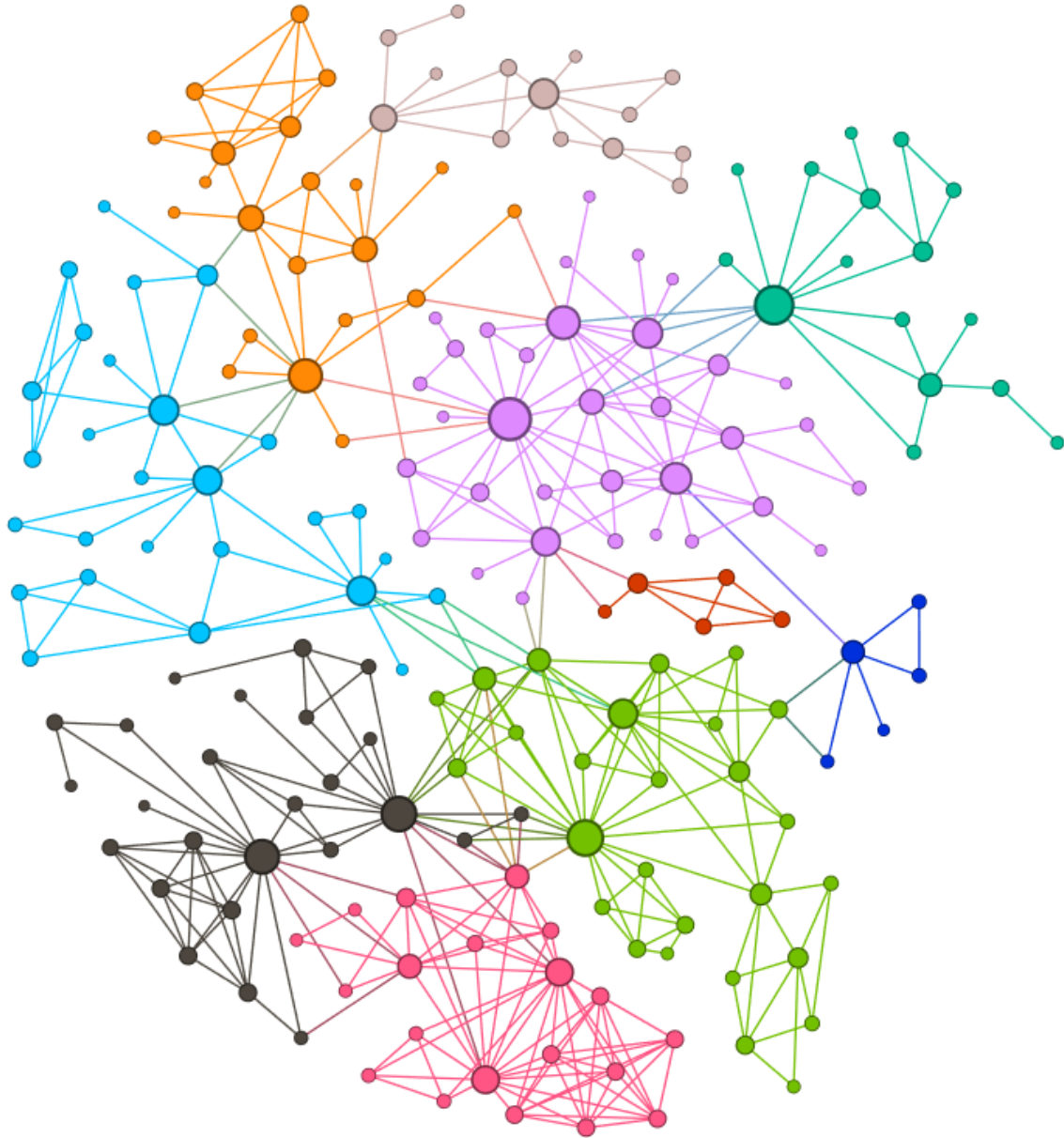


Figura 2.8: Centralidad *pagerank*. El tamaño de cada nodo es proporcional a su centralidad *pagerank*, por lo que los más grandes son también los más centrales con esta medida.

siguiente ejemplo: En una búsqueda en Internet, el orden en que se muestran los resultados sigue un orden de relevancia dado por una centralidad *pagerank*. Se muestran aquellos resultados que más personas han consultado en un rastreo similar. En una búsqueda de los colaboradores del coloquio entre 2013 y 2018, estos aparecerán en el orden dado por el valor de su centralidad *pagerank*.

En la red del Coloquio VNL los nodos con mayor centralidad *pagerank* son investigadores que ya tienen una buena posición dentro de la comunidad, a los que acuden otras personas para trabajar con ellos. Esto es muy claro con los nodos 75 y 32: son investigadores muy conocidos en el área y tienen, en la centralidad *pagerank*, la segunda y quinta posición, respectivamente, dentro de la red del Coloquio VNL. Sin embargo, fuera del grado, no aparecen en las primeras diez posiciones con ninguna otra centralidad. Esto se puede interpretar de la siguiente manera: los nodos 75 y 32 tienen una posición tan reconocida dentro de la red, que no necesitan buscar nuevas relaciones con los demás investigadores y estudiantes, ni esforzarse en hacerse visibles dentro de la comunidad, pues serán los demás participantes quienes intenten aproximarse a ellos. Por el contrario, los otros actores con centralidad *pagerank* alta, como 52, 42, 9, 3 y 53 son investigadores que, aún cuando son conocidos dentro de la comunidad, tienen más movimiento dentro de la red que 75 y 32, lo que podría atribuirse a que son más jóvenes. El nodo 66 corresponde a una investigadora extranjera que fue tutora de algunas personas del Coloquio VNL cuando hicieron estancias de investigación en su país; además ella hizo una estadía en México en la que estableció otras relaciones. Entre quienes colaboraron con ella se encuentran vértices con centralidades altas como 52, 53, 54 y 75, por lo que el nodo 66 obtuvo una posición privilegiada en la red aunque no esté físicamente cerca.

En la figura 2.9, se dio una coloración de los nodos para indicar el número de años, entre 2013 y 2018, en los que han participado como colaboradores en alguna de las pláticas. En esta misma figura se puede observar que los nodos más importantes de la red del Coloquio VNL son también los que han participado como colaboradores en al menos cuatro de las seis ediciones del Coloquio VNL. En la red no se distinguen unos años de otros, por lo que algún nodo podría obtener todos sus enlaces en una sola emisión. Sin embargo, en el análisis se observa que la constancia de algunos participantes influye en su importancia dentro de la red, donde esta no es en asistencia ni siendo ponente, sino colaborando en los trabajos que se presentan en el coloquio año con año. Así, nodos que han dado sólo una plática en seis años, como 32 y 75, figuran en las pláticas cada año por el trabajo que hacen con sus colegas y alumnos.

2.2. Análisis y resultados

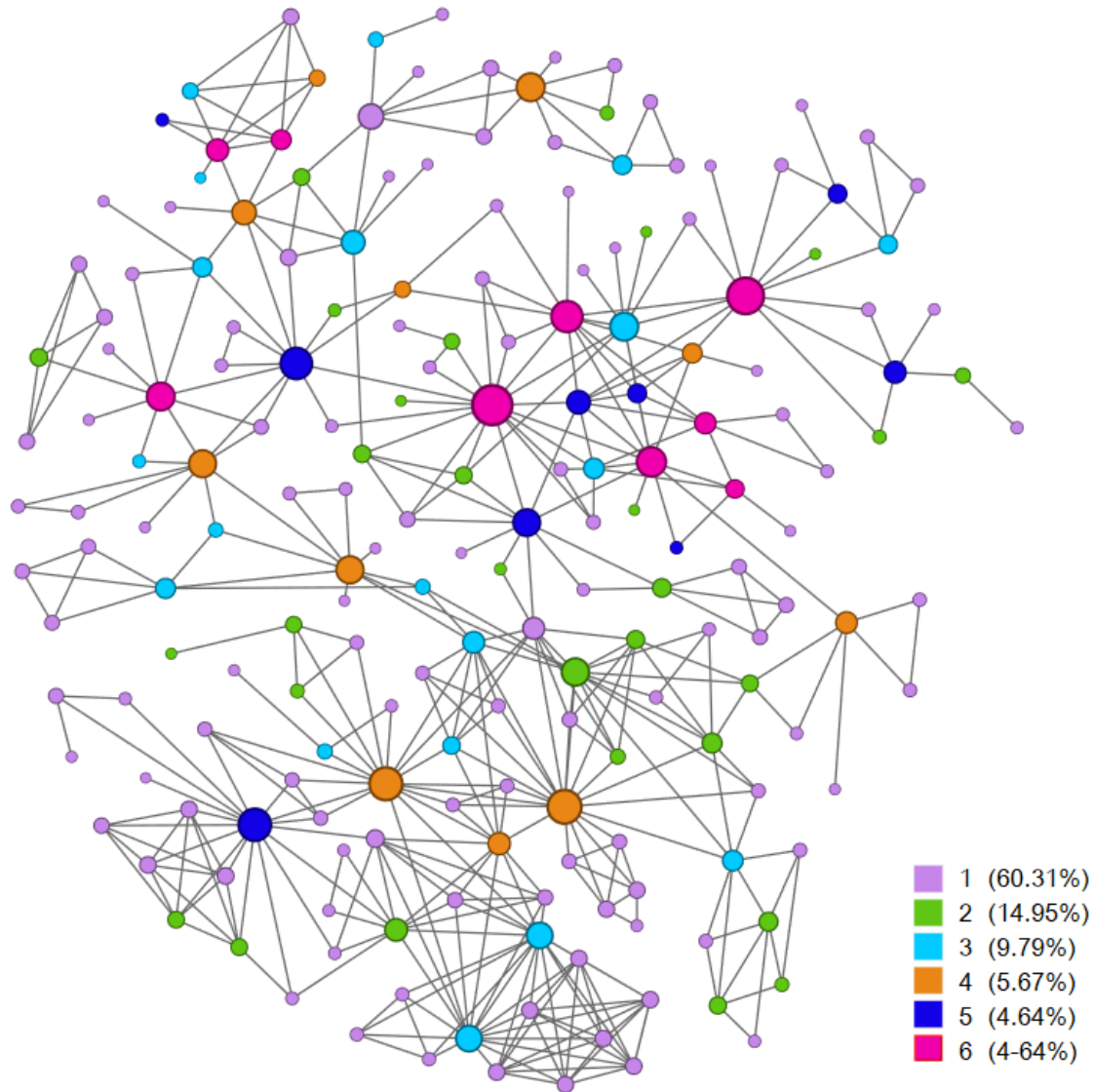


Figura 2.9: Presencia anual. El color de cada nodo indica en cuántas de las ediciones del Coloquio VNL, entre 2013 y 2018, dicho nodo ha aparecido como colaborador en alguna de las pláticas presentadas. El tamaño de cada nodo es proporcional a su centralidad *pagerank*, por lo que los más grandes son también los más centrales con esta medida.

El coeficiente de aglomeración mide con qué frecuencia dos colaboradores de un individuo son colaboradores entre sí. Los nodos cuyo coeficiente de agrupamiento es igual a uno, que es el valor más alto que se puede tener, son los nodos: 1, 4–8, 22, 26, 34, 37, 39, 41, 44–48, 56, 63, 67, 69, 70, 92, 93, 95, 97, 106–108, 114–117, 127–129, 137, 139–141, 147, 163, 168, 170, 172, 175, 178–184, 190–195, 200, 201, 206–209, 211, 212, 216, 230, 233, 234, 238, 240, 243, 244, 250–256, 260–266.

Que un nodo u tenga coeficiente de agrupamiento igual a uno, significa que la gráfica inducida por $N[u]$ es una gráfica completa. En el caso del Coloquio VNL, se observa que cualquier pareja de personas, en el conjunto de u y sus colaboradores, ha presentado algún trabajo colaborativo en conjunto.

A continuación verificaremos si la red del Coloquio VNL es robusta, como se planteó en la sección 2.1.

De acuerdo a la centralidad *pagerank*, el orden de eliminación de los nodos centrales es como sigue:

Paso 1: Se quita el nodo 52. La nueva red tiene 3 componentes y 189 vértices en la componente más grande, que es el 97.93% de los vértices.

Paso 2: Se quita el nodo 75. La nueva red tiene 7 componentes y 174 vértices en la componente más grande, que es el 90.62% de los nodos.

Paso 3: Se quita el nodo 42. La nueva red tiene 8 componentes y 168 vértices en la componente más grande, que es el 87.96% de los nodos.

Paso 4: Se quita el nodo 9. La nueva red tiene 10 componentes y 160 vértices en la componente más grande, que es el 84.21% de los nodos.

Paso 5: Se quita el nodo 32. La nueva red tiene 13 componentes y 152 vértices en la componente más grande, que es el 80.42% de los nodos.

Paso 6: Se quita el nodo 3. La nueva red tiene 15 componentes y 148 vértices en la componente más grande, que es el 78.72% de los nodos.

Paso 7: Se quita el nodo 66. La nueva red tiene 18 componentes y 141 vértices en la componente más grande, que es el 85.4% de los nodos.

Paso 8: Se quita el nodo 53. La nueva red tiene 19 componentes y 139 vértices en la componente más grande, que es el 74.73% de los nodos.

2.2. Análisis y resultados

Paso 9: Se quita el nodo 29. La nueva red tiene 19 componentes y 138 vértices en la componente más grande, que es el 74.59% de los nodos.

Paso 10: Se quita el nodo 30. La nueva red tiene 22 componentes y 133 vértices en la componente más grande, que es el 72.28% de los nodos (figura 2.10).

De aquí se puede decir que la mayor parte de los vértices permanece unida aún cuando los nodos dominantes⁵ no están presentes. Se llamará R_p a la componente gigante obtenida tras la eliminación de estos diez vértices, esta red cuenta con 133 vértices y 257 aristas, su grado medio es 3.865 que es poco menor al 4.454 de la original; la densidad tiene un valor de 0.029, misma que es apenas mayor al 0.0236 de la original.

R_p tiene diámetro 14, radio 8 y el promedio de las longitudes de las rutas geodésicas es 6.461, es decir, la longitud de la ruta promedio tiene un intermediario más que en R . La modularidad de R_p es de 0.794, es decir, hay más aristas de las esperadas entre parejas de vértices semejantes y tiene 10 comunidades. R_p cuenta con 223 triángulos y su coeficiente de agrupamiento es 0.721, que es ligeramente menor al coeficiente de agrupamiento de R , que vale 0.74, pero en ambos casos podemos decir que es frecuente que dos colaboradores de un tercero sean adyacentes entre sí.

Cabe señalar que los primeros ocho vértices eliminados coinciden (y en el mismo orden) con los ocho vértices con mayor centralidad *pagerank* en la tabla 2.1, cosa que diferirá con la centralidad intermedia. Esto refleja que la centralidad *pagerank* es menos susceptible a cambios en la red, mientras que la centralidad intermedia se ve fácilmente afectada tras la eliminación de nodos.

De acuerdo a la centralidad intermedia, el orden de eliminación es como sigue:

Paso 1: Se quita el nodo 49. La nueva red tiene 3 componentes y 187 vértices en la componente más grande que es el 96.89% de los vértices.

Paso 2: Se quita el nodo 3. La nueva red tiene 4 componentes y 184 vértices en la componente más grande que es el 95.83% de los nodos.

Paso 3: Se quita el nodo 88. La nueva red tiene 6 componentes y 180 vértices en la componente más grande que es el 94.24% de los nodos.

⁵En la sección 1.2, se definieron los nodos dominantes como los vértices que tienen la mayor centralidad *pagerank*.

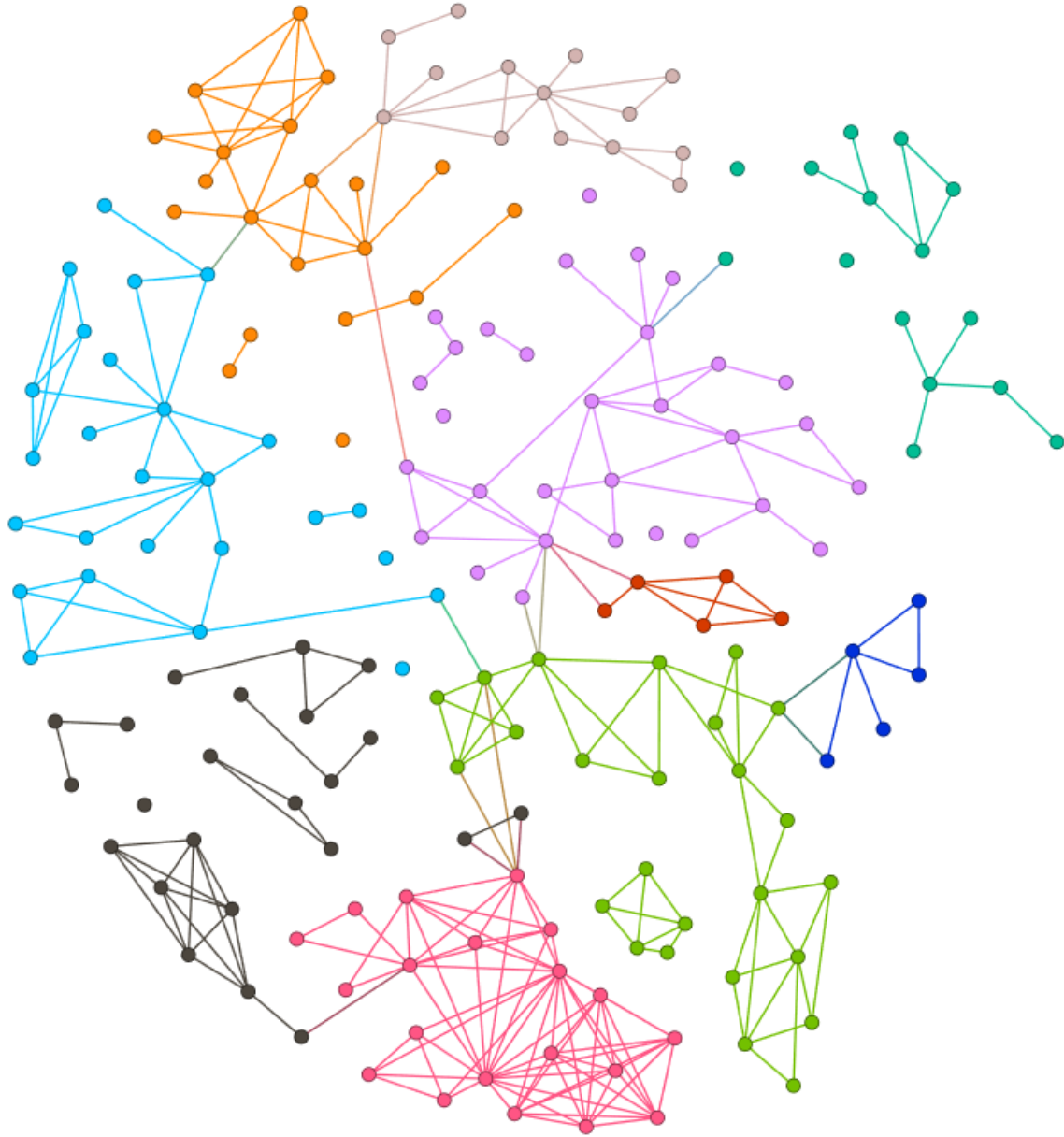


Figura 2.10: Paso 10 de la eliminación por centralidad *pagerank*. La red obtenida tras eliminar los nodos 3, 9, 29, 30, 32, 42, 52, 53, 66 y 75.

2.2. Análisis y resultados

Paso 4: Se quita el nodo 62. La nueva red tiene 10 componentes y las dos componentes más grandes tienen 82 y 91 vértices, respectivamente, la componente más grande tiene el 47.89% de los nodos (figura 2.11).

Paso 5: Se quita el nodo 9. La nueva red tiene 12 componentes y las dos componentes más grandes tienen 82 y 83 vértices, respectivamente, la componente más grande tiene el 43.92% de los nodos.

Paso 6: Se quita el nodo 196. La nueva red tiene 15 componentes y 83 vértices en la componente más grande que es el 43.92% de los nodos.

Paso 7: Se quita el nodo 24. La nueva red tiene 16 componentes y 51 vértices en la componente más grande que es el 27.27% de los nodos (figura 2.12).

Cabe señalar que si se hubieran eliminado los diez nodos con mayor centralidad intermedia en la gráfica original: 49, 43, 9, 52, 3, 30, 111, 196, 75 y 176; la red resultante tendría una componente gigante con 103 vértices, que es el 55.98% de los nodos. Esta diferencia fundamental se debe a que las rutas geodésicas se modifican tras la eliminación de cada nodo. Se decidió hacer la eliminación de la otra manera pues, para pensar qué es lo que ocurre con la red del Coloquio VNL cuando un actor deja de participar, tenemos que considerar la red que no tiene a dicho actor y sus respectivas medidas.

A diferencia de lo que ocurre con los nodos dominantes, quitar nodos con la mayor centralidad intermedia desconecta la gráfica rápidamente. Si se eliminan tan sólo siete vértices se pierde la comunicación en la red.

Por lo que se puede decir que, aunque la red es aparentemente robusta por lo observado tras la eliminación de vértices dominantes, la red es vulnerable a ataques específicos ya que, al eliminar nodos seleccionados, se pierde por completo la comunicación.

Cuando se eliminaron los vértices de acuerdo a su centralidad intermedia, el primer momento en el cual se ve seriamente afectada la red es en el paso 4 (figura 2.11) después de haber quitado cuatro nodos (49, 3, 88 y 62), en el cual quedan ocho componentes conexas pequeñas y dos componentes grandes, una con 82 vértices y otra con 91 vértices, que dividen, a grandes rasgos, a las comunidades en dos grupos: las comunidades 0, 2, 3 y 7 (y tres vértices de la comunidad 4) por un lado y las comunidades 1, 4, 5 y 9 (y dos vértices, uno de la comunidad 7 y uno de la 8) por otro. Esta es una primera señal de lo frágil que es la comunicación entre las comunidades

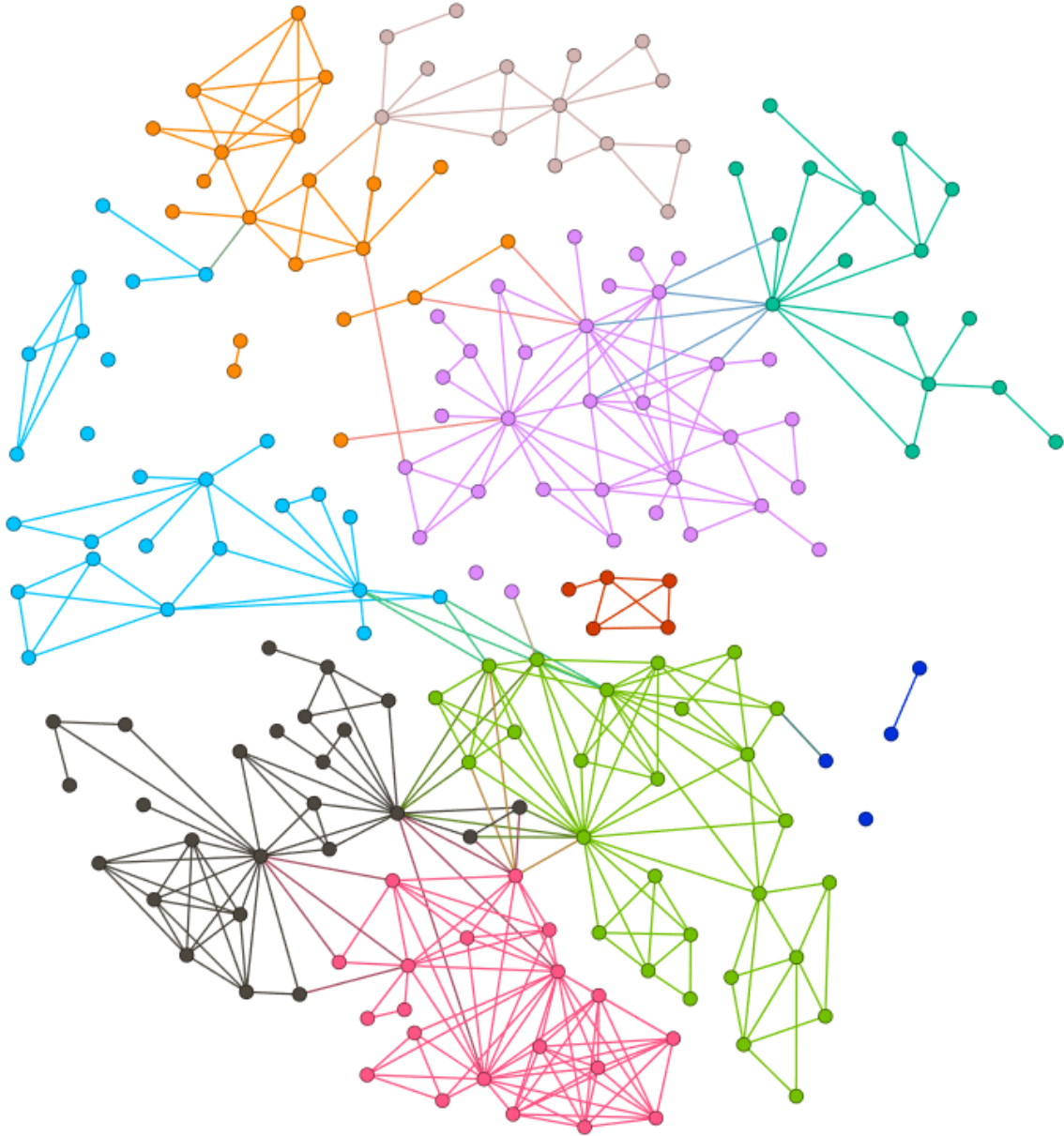


Figura 2.11: Paso 4 de la eliminación por centralidad intermedia. La red obtenida tras eliminar los nodos 3, 49, 62 y 88.

2.2. Análisis y resultados

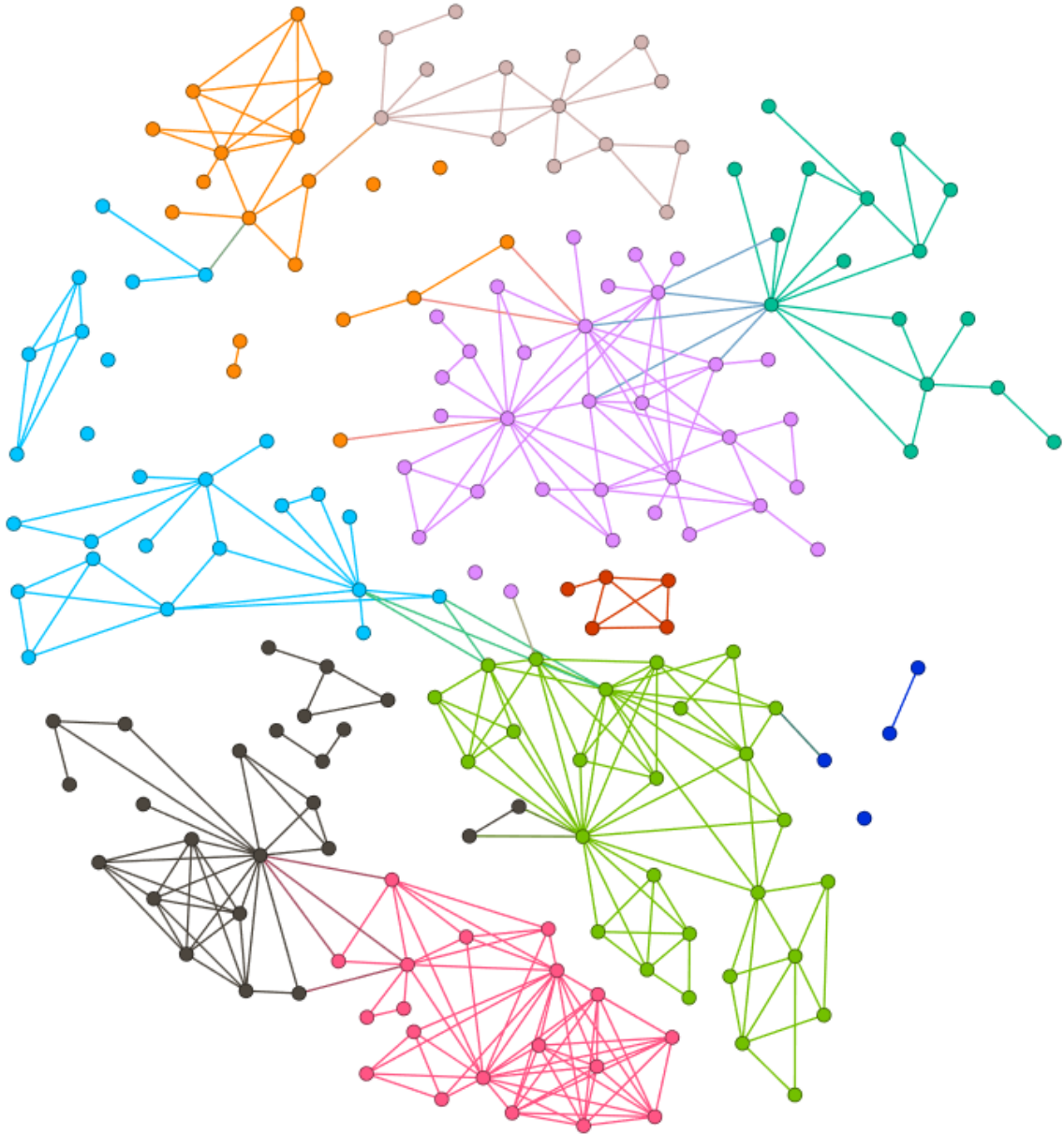


Figura 2.12: Paso 7 de la eliminación por centralidad intermedia. La red obtenida tras eliminar los nodos 3, 9, 24, 49, 62, 88 y 196.

de la red. Si se considera el ejemplo de las invitaciones, se tiene que la transferencia de la invitación entre estos dos grandes grupos pende de cuatro individuos. Por lo que si ellos no envían la invitación, la asistencia queda limitada a personas en una sola de estas componentes grandes.

Después se quitaron tres nodos más (9, 196 y 24), se obtuvieron 16 componentes: doce de ellas con a lo más 5 personas y cuatro componentes medianas (de 28, 34, 48 y 51 nodos). Aquí las comunidades se ven aún más divididas y la comunicación está cortada, pues aún en la componente de 51 personas no se encuentra ni el 30% de la población. En el ejemplo de las invitaciones, muy pocas personas estarían enteradas y si el Coloquio VNL pudiera realizarse sería con apenas 48 o 51 personas, es decir, con la cantidad de ponentes de las últimas ediciones. Las siete personas que quitamos afectan fuertemente el flujo de información en la red.

La mala comunicación en la red, que se intuye en el proceso de eliminación de los vértices con mayor centralidad intermedia, se ve claramente reflejada en una simulación del ejemplo que se planteó en esta sección.

A continuación se harán tres simulaciones. Para elegir el vértice de inicio se usará el generador de números aleatorios de WolframAlpha [45].

En la simulación 1, el primer nodo que recibió la invitación fue el nodo 170, mismo que pertenece a la comunidad 4. A continuación se presentarán los vértices que reciben la invitación en cada paso (los nodos que reciben la invitación por primera vez en el paso i , coinciden con los vértices en el i -ésimo nivel de un árbol BFS enraizado en el nodo de inicio: 170), hasta agotar el cupo.

Paso 0: 170,

Paso 1: **111**, 62, **3**,

Paso 2: **30**, 86, 259, 117, 116, 175, 258, 187, 218, 181, 123, 5, 4, 148, 205, 56, 55, **52**, 201,

Paso 3: 209, 232, 27, **29**, **28**, 145, 87, 208, 252, 250, 251, 169, 173, 2, 0, 185, **196**, 200, **66**, 127, 126, 179, 180, **112**, **54**, **53**, 98, 234, 233, **49**, **177**, **178**, **176**, 61,

Paso 4: **43**, 40, 93, 215, 92, 94, 95, 96, 44, **42**, 41, 255, **24**, 64, **9**, 254, 108, 107, 106, 51, 1, 69, 70, 6, 210, 267, 257, 155, **75**, 89, 76, 95, 235, 113, 132, 131, 256, 77, 88, 122, 97, 68, 67, 48, 50.

2.2. Análisis y resultados

Al terminar el paso 4 hay 102 nodos enterados del Coloquio VNL, por lo que nadie más puede enviar la invitación.

Se subrayaron en letras negritas los nodos más centrales de acuerdo a la tabla 2.1. Los nodos centrales que no reciben invitación en esta simulación son: 20, 21, 32, 38, 260–266. Los nodos 32 y 38, que tienen centralidad *pagerank* alta, y los nodos 20 y 21, que tienen centralidad de eigenvector y peso de *hub* altos se enterarían, si hubiera un envío más de invitaciones. Sin embargo, los nodos 260–266 no la recibirían hasta que 20 y 21 estuvieran enterados. De aquí se observa que tener un peso alto de *hub* y estar en una comunidad muy conectada no necesariamente implica recibir una invitación, como ocurre para los nodos 260–266, pero parece que, en general, los nodos centrales deberían recibirla.

En la figura 2.13 podemos ver en rojo los nodos que estarían presentes en el Coloquio VNL suponiendo que todos pueden recibir y enviar sin problemas la invitación.

Se puede observar, en las figuras 2.2 y 2.13, que en esta simulación la comunidad inicial, 4, está invitada por completo, así como la comunidad 2 (que es adyacente a la anterior), las comunidades 1 y 7 están invitadas en casi su totalidad. En las comunidades 3 y 6 hay dos invitados y en las comunidades 0, 5, 8 y 9 hay un invitado. Es decir, las comunidades en la periferia de la gráfica de las comunidades (figura 2.3), tienen pocos invitados en esta simulación (8 invitados entre seis comunidades), mientras que las comunidades en el ciclo de longitud 4, $\eta = (4, 2, 7, 1, 4)$, de la gráfica de las comunidades son las que estarán invitadas (94 invitados entre cuatro comunidades). Es decir, las invitaciones quedan repartidas en cuatro de las diez comunidades.

En la segunda simulación se eligió un vértice al azar que no está en la comunidad 4, pues habría poca diferencia con la anterior. Se iniciará con el nodo 90, en la comunidad 5. A continuación se presentarán los vértices que reciben la invitación en cada paso.

Paso 0: 90,

Paso 1: 31, **32**, 33, 35, 36, 241

Paso 2: 34, **9**, 23, **25**, 31, 63, 109, 134, 137, 182, 183, 184,

Paso 3: 7, 8, 10, 19, **20**, **21**, **24**, **28**, **42**, **43**, 64, 147, 190, 191, 22, 26, 243, 244, 133,

Paso 4: 171, 138, 45, 46, 47, **260**, **261**, **262**, **263**, **264**, **265**, **266**, 27, **29**, **30**, 254, 255, 40, 41, 44, 94, 95, 96, 237, 238, 239, 240, 48, **49**,

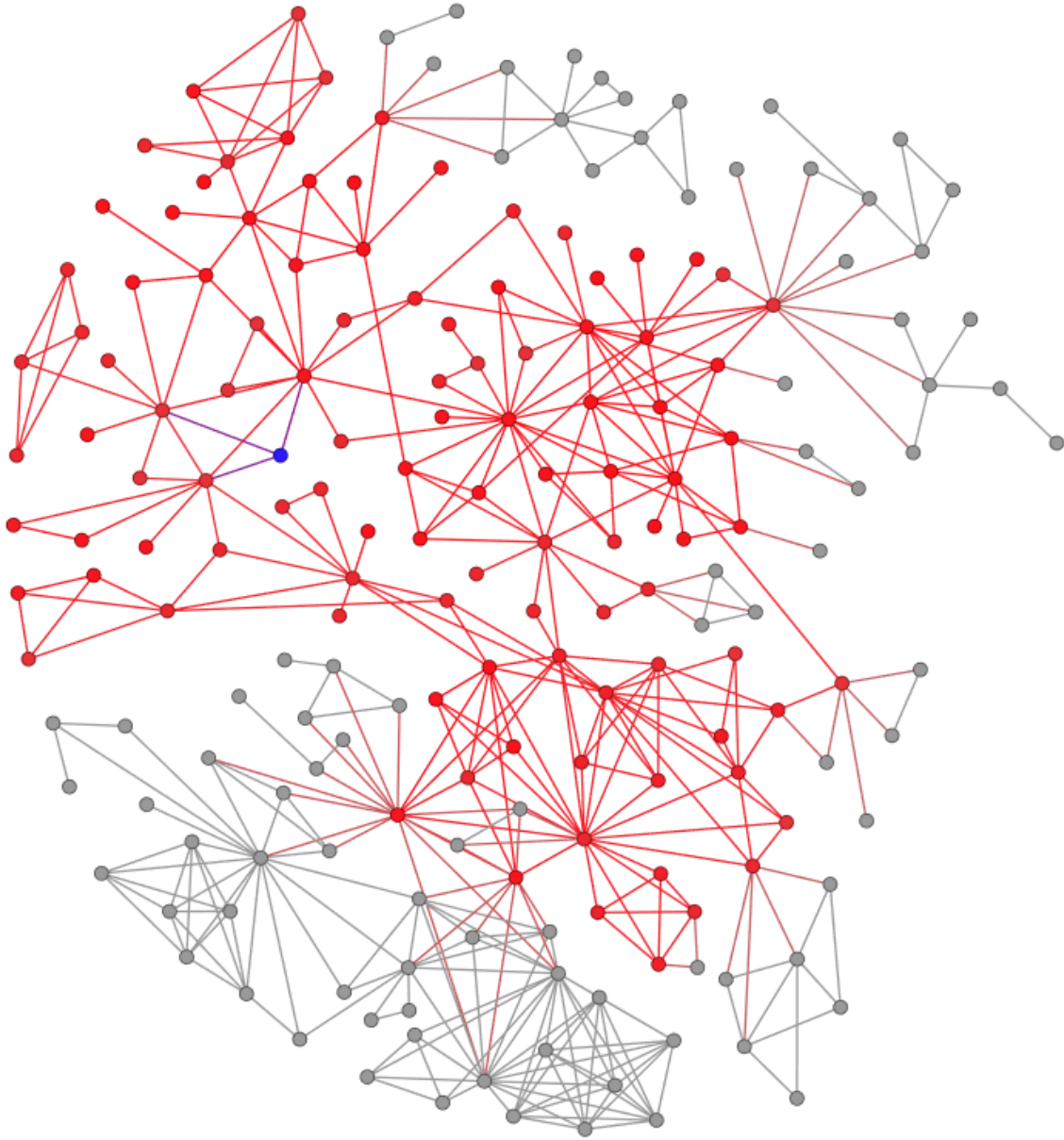


Figura 2.13: Simulación 1. El nodo 170, que fue el primero en recibir la invitación, está marcado en azul. Los nodos invitados son todos aquellos en rojo y los grises son los que no recibieron la invitación.

2.2. Análisis y resultados

Paso 5: 92, 93, 215, 86, 87, **111**, 145, 208, 209, 232, 165, 166, 167, 192, 50, **52**, **53**, **54**, 67, 68, **176**, **177**, **178**.

Al finalizar el paso 5 hay 90 invitados y más de 20 vértices en el siguiente nivel de un árbol BFS enraizado en el nodo 90, por lo que el proceso de invitación se detiene.

Los nodos en letras negritas son, otra vez, los más centrales de acuerdo a la tabla 2.1. Los nodos centrales que no reciben invitación en esta simulación son: 3, 38, 66, 75 y 112. Los cinco tienen centralidad *pagerank* alta y todos ellos, salvo 38, se enterarían en el siguiente paso. Aunque no es raro que 38 no se entere pues este nodo sólo tiene centralidad alta por ser el único intermediario entre una porción de su comunidad (la comunidad 0) y el resto de la red.

Las comunidades 5 y 9, tienen cliques grandes, nueve enlaces entre ellas y dependen de la comunidad 1 para comunicarse con el resto de la red (figura 2.3). En esta simulación, estas tres comunidades pronto saturan los lugares del Coloquio VNL, por lo que sólo alcanzan invitación doce personas fuera de ellas: dos de ellas en la comunidad 2, dos en la comunidad 6 y nueve en la comunidad 7 que es la comunidad mejor conectada. Mientras que, en las comunidades 0, 2, 3 y 8, ninguna persona recibe la invitación. En esta simulación, una vez más, la invitación queda repartida en cuatro comunidades y seis quedan fuera, para efectos prácticos.

En la figura 2.14 podemos ver en rojo los nodos que estarían presentes en el Coloquio VNL suponiendo que todos pueden recibir y enviar sin problemas la invitación.

En la tercera simulación se eligió un vértice al azar que no estuviera en las comunidades 4, 5 y 9 (se eliminó la comunidad 9 por su relación tan cercana con la comunidad 5 con la cual ya tenemos un ejemplo ilustrativo) para obtener un ejemplo que sea diferente a los anteriores, de manera significativa. Se comenzó con el nodo 89, en la comunidad 7. A continuación se presentan los vértices que reciben la invitación en cada paso.

Paso 0: 89,

Paso 1: **53**, **54**, 65, **66**, **75**, 213,

Paso 2: **49**, **52**, 77, 88, 97, 98, **112**, 122, 76, 55, 155, 179, 180, 200, 72, 74, 82, 120, 163, 168, 193, 197, 256,

Paso 3: **43**, 48, 50, 67, 68, **176**, **177**, **178**, **3**, 61, 126, 127, 201, 233, 234, 249, 144, 206, 207, 215, 216, 113, 131, 132, 194, 195, 56, 135, 128, 129, 119, 162,

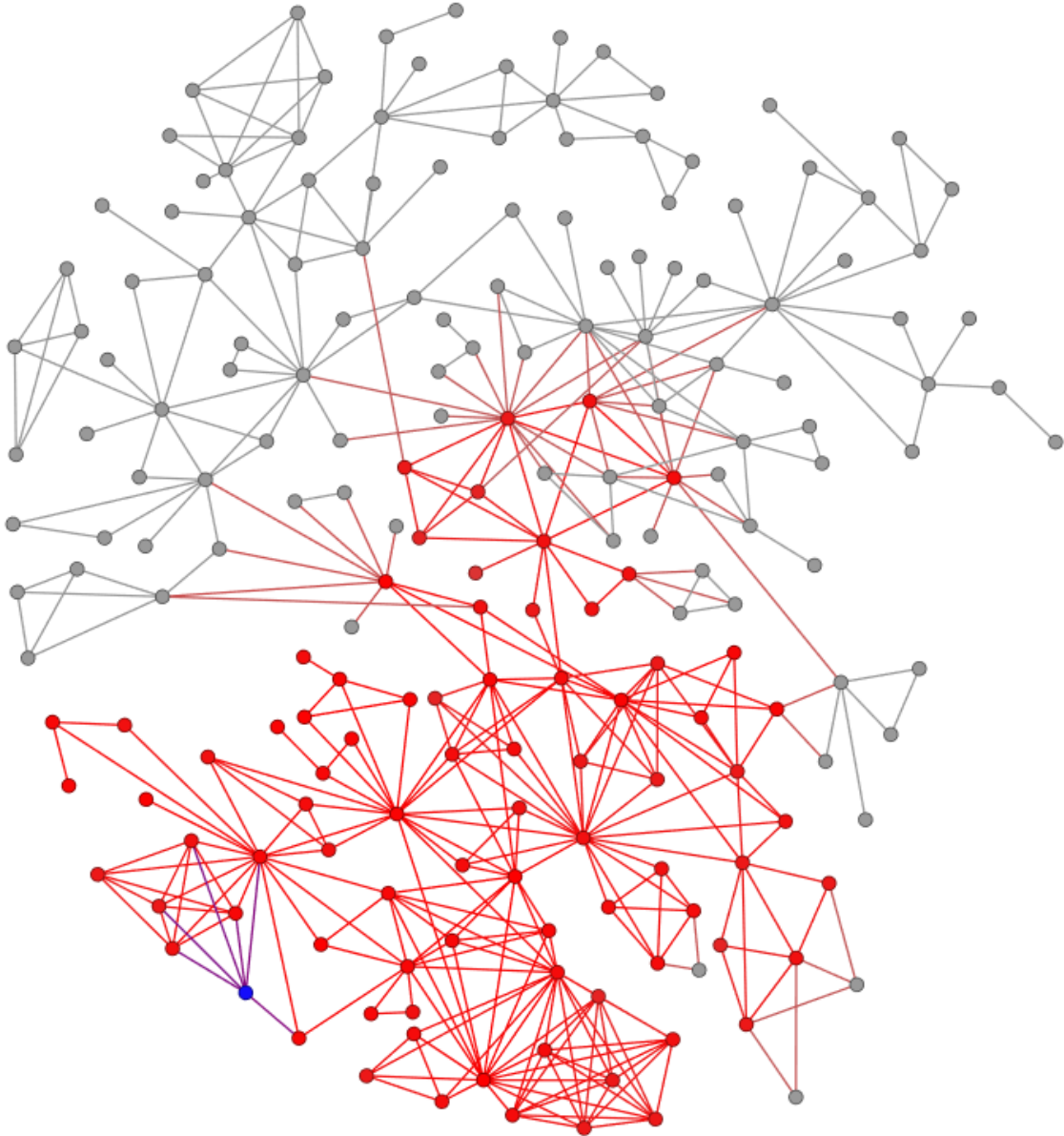


Figura 2.14: Simulación 2. El nodo 90, que fue el primero en recibir la invitación, está marcado en azul. Los nodos invitados son todos aquellos en rojo y los grises son los que no recibieron la invitación.

2.2. Análisis y resultados

Paso 4: **9, 28, 29**, 40, 41, **42**, 44, 64, 139, 140, 141, **196**, 4, 5, 62, **111**, 123, 148, 170, 205, 235, 94, 161.

Al finalizar el paso 4 hay 85 invitados y más de 25 vértices en el siguiente nivel de un árbol BFS enraizado en el nodo 89, por lo que el proceso de invitación se detiene.

Una vez más los nodos en letras negritas son los más centrales de acuerdo a la tabla 2.1. Los nodos centrales que no reciben invitación en esta simulación son: 20, 21, 24, 32, 38, 260–266. En esta simulación iniciamos en la comunidad 7, cuyos nodos están bien conectados entre sí, aunque no tanto como en la comunidad 9; además, es la mejor comunicada respecto a las otras. Por su estructura interna y su posición en R , el efecto en las invitaciones es que se esparcen en la totalidad de cuatro comunidades (3, 6, 7 y 8), llegan a diez nodos en cada una de las comunidades 1 y 2, cuatro en la 4 y un nodo en la 5, y se deja fuera a las comunidades 0 y 9. Es decir, las comunidades adyacentes a la 7 quedarán representadas en esta emisión hipotética del Coloquio VNL, mientras que las cuatro que están a distancia 2 en la gráfica de las comunidades, reciben apenas cinco invitaciones para los 86 vértices que hay juntando las cuatro.

En la figura 2.15 podemos ver en rojo los nodos que estarían presentes en el Coloquio VNL suponiendo que todos pueden recibir y enviar sin problemas la invitación.

En las tres simulaciones se puede observar que la red R no está tan bien comunicada porque, en las tres ocasiones, hubo comunidades con uno o ningún invitado y las comunidades adyacentes a la primera en obtener la invitación son las que saturan el cupo. Otra característica en común es que la mayoría de los nodos centrales respecto a la excentricidad, los grados de cercanía usual y armónico, y las centralidades intermedia y *pagerank* sí reciben la invitación, pues estas medidas reflejan la posición de los nodos en la red. No así con el peso de *hub* y la centralidad de eigenvector, que se fijan en el nodo y sus vecinos de manera local, por lo que en dos de las simulaciones los nodos con mayores valores en estas medidas no resultan invitados. De hecho podemos prever (por la posición de las comunidades en la figura 2.3) que si el primer nodo invitado estuviera en alguna de las comunidades 0, 2 o 3, los nodos centrales de acuerdo al peso de *hub* y la centralidad de eigenvector tampoco estarían invitados.

En los tres ejemplos también se puede observar que las comunidades 5 y 9 están muy relacionadas entre ellas, pero están alejadas del sector superior en la figura 2.2. Es curioso que los integrantes de la comunidad 5, que están en su mayoría en la Ciudad de México, estén igual de incomunicados de los otros miembros en la ciudad que los integrantes de la comunidad 9 que están, en su mayoría, en Estados Unidos.

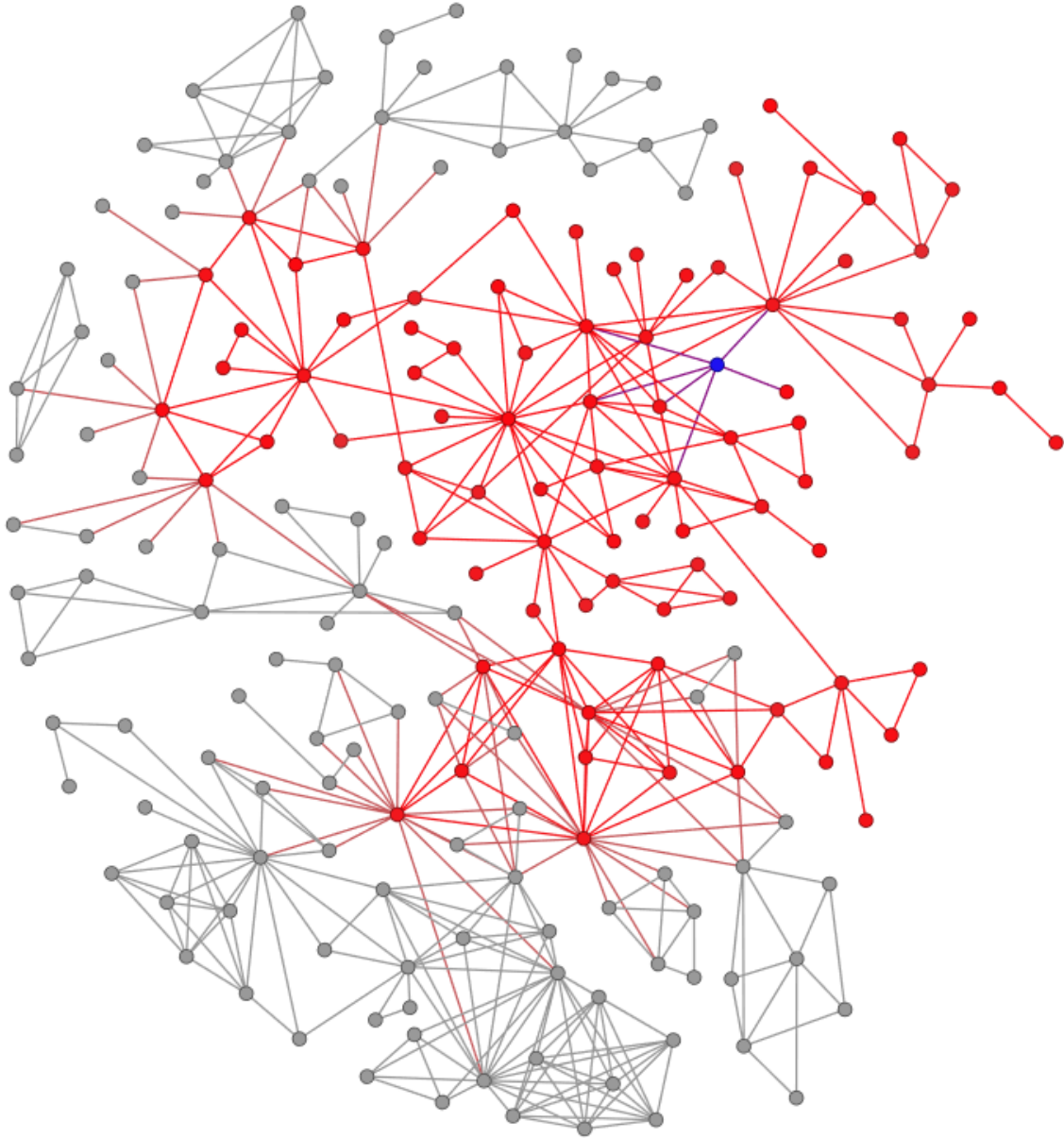


Figura 2.15: Simulación 3. El nodo 89, que fue el primero en recibir la invitación, está marcado en azul. Los nodos invitados son todos aquellos en rojo y los grises son los que no recibieron la invitación.

2.2. Análisis y resultados

Es decir, la cercanía geográfica no garantiza la comunicación. La comunidad 0 es otra que está muy aislada y, salvo que el proceso de invitaciones inicie en ella o en la comunidad 2, sus miembros no estarán invitados al Coloquio VNL, salvo quizá por el nodo 210, como en la primera simulación. Pero, en general, quedarán fuera.

El esparcimiento de la invitación suele llegar a una nueva comunidad por un nodo de centralidad intermedia alta (figura 2.7) y luego se difunde entre los demás, que es algo que ya se podía predecir con el ejercicio de eliminar nodos con alta centralidad intermedia, pues estos son los que habilitan la comunicación en R .

Al combinar los tres ejemplos vemos que, aunque la comunicación en la totalidad de la red puede cortarse con relativa facilidad, en grupos de dos o hasta tres comunidades hay más posibilidades de intercambio de información. Por lo que, si se envía un número pequeño de invitaciones, una por cada grupo de comunidades bien interconectadas, todos los integrantes de la red tienen la posibilidad de recibirla.

Conclusiones

En este trabajo se analizaron las relaciones entre los participantes del Coloquio VNL a partir de la colaboración en las ponencias presentadas entre 2013 y 2018. Debido a ello, las reflexiones expuestas a continuación no son generalizables a todos los vínculos que se establecen entre los miembros del Coloquio VNL.

Dentro de las características que se pudieron observar en la red del Coloquio VNL, una que vale la pena destacar es su baja densidad. Como ya se dijo en la sección 1.2, cuanto mayor es la densidad en una red social, mayor es la visibilidad que tienen unos de otros en la red, además de reflejar cohesión, confianza y apoyo. En la red de colaboración de las ponencias del Coloquio VNL observamos lo contrario: las relaciones están bastante restringidas, los grupos están diferenciados de manera marcada y hay poca visibilidad entre los mismos. De acuerdo a estos datos, la comunicación es pobre y la información no fluye bien en la totalidad de dicha red, lo que la haría vulnerable, como ya se vio en la sección 2.2, cuando analizamos la robustez. Se puede suponer que tanto la división en grupos como la falta de confianza o seguridad en la red, pueden ser consecuencia de la competitividad del medio académico derivada de los sistemas de estímulos por número de publicaciones. Otra posibilidad es que algunas personas suelen trabajar juntas debido a su buen entendimiento y que rara vez incluyan a alguien distinto.

En contraste, de manera local, las comunidades tienen un mejor número de enlaces (en siete, al menos el 11% de ellos y en tres, al menos el 35%) por lo que, en la organización hipotética del Coloquio VNL basada en invitaciones de mano en mano, nos basta enviar pocas en comunidades estratégicas a individuos centrales para que este pueda realizarse. Una invitación sería al nodo 9 de la comunidad 5, que es el nodo más central dentro de las comunidades 5 y 9, que están comunicadas entre sí con nueve enlaces e internamente tienen muchos enlaces. Otra sería al nodo 3 de la comunidad 2, que es el nodo más central en su comunidad, con una centralidad intermedia alta en la red, y es adyacente a cuatro nodos en la comunidad 4, incluido

el nodo 111 que es el más central en dicha comunidad, así el nodo 3 es un buen punto de propagación. Una última sería al nodo 54 en la comunidad 7, miembro del comité organizador, a distancia 1 de la comunidad 3 y a distancia 2 de las comunidades 1, 6 y 8. Si bien no es el nodo más central de su comunidad, tiene la sexta posición en el grado de cercanía y su ubicación es más conveniente que la de los nodos 49 y 52, por la elección de las dos invitaciones anteriores.

Los vértices 3, 27 y 210 fueron organizadores una vez y el 9 en cuatro ocasiones; los cuatro son vértices que tienen vínculos con dos o tres comunidades diferentes. Además los nodos 54 y 89 de la comunidad 7, miembros del comité organizador por cinco y seis años respectivamente, tienen enlaces con el nodo 75 en la comunidad 3. Se puede inferir que los miembros del comité organizador fomentan la vinculación entre comunidades; además se puede suponer que promueven la asistencia de sus alumnos y colegas lo cual facilita la creación de enlaces nuevos. Una propuesta para sacar partido de estas hipótesis es rotar una o dos personas por año: que descansen un año o dos pero vuelvan después para que otros descansen y así sucesivamente. De esta forma ayudarían a que la comunidad del Coloquio VNL estuviera más integrada, pues el incremento de vínculos (y por tanto en la densidad), se traduce en más apoyo entre los integrantes de la red.

La información que tengo como participante de la red, me permite darme cuenta de que los investigadores jóvenes, que ya cuentan con trayectorias profesionales, colocados en alguna institución y que trabajan en conjunto, han logrado crear lazos entre grupos diferentes. Por el contrario, los estudiantes trabajan sólo con sus asesores, individualmente o en grupo, y los investigadores con carreras consolidadas asisten cada vez menos a este evento (sus alumnos y colegas sí, pero ellos cada vez menos, como es el caso de los nodos 2, 32, 75 y 185). Entonces, puede pensarse que los estudiantes y los investigadores consolidados contribuyen menos a la vinculación de la red que los investigadores jóvenes.

Otra característica que se puede destacar es que la red no es vulnerable en la cuestión de la importancia de los nodos, a diferencia de lo que ocurre con la densidad y la centralidad intermedia, pues en el análisis de robustez en el que se eliminaron nodos importantes (con centralidad *pagerank* alta), la gráfica permaneció en su mayoría conectada. Esto tiene que ver con el hecho de que los nodos importantes son más estables ante los cambios en la red y, de esta manera, los enlaces entre comunidades hechos entre personas con menos importancia permanecen aún cuando quitamos los vértices importantes. Es decir, aunque las cabezas de los distintos grupos dejaran

2.2. Análisis y resultados

de trabajar, los vínculos que se hubieran formado entre el resto de los integrantes de los grupos con personas de otras comunidades no dejarían de existir. De hecho, es probable que se crearan nuevos lazos si no estuvieran los “ascendientes académicos” de los demás. Dada la apertura del Coloquio VNL a aceptar pláticas de estudiantes, sería previsible que la falta de participación de los investigadores consolidados no impidiera su realización, pues siempre hay estudiantes que quieren participar y se les da el espacio. Además, es un proceso natural que los nuevos miembros de cualquier comunidad tomen terreno; en todas las profesiones llega un momento en el que los mentores ya no están y sus sucesores continúan la labor.

En la sección 1.3, se expusieron tres modelos teóricos para modelar redes reales. Se mencionó que, de esos tres modelos, se esperaba que el modelo de mundo pequeño, diseñado para redes sociales, resultaría el adecuado para modelar la red del Coloquio VNL. Tras los resultados obtenidos en la sección 2.2 podemos concluir que, en efecto, la red del Coloquio VNL es una red de mundo pequeño.

Dentro de las cosas que quedan pendientes para estudiar en la red del Coloquio VNL hay cuatro inmediatas:

La primera es darle peso a las aristas, contabilizando el número de veces que dos personas trabajaron juntas, pues no todos los vínculos son de la misma calidad. Hay con quienes uno se entiende más que con otros y se hace frecuente trabajar juntos, por lo que cabe hacer una distinción en este sentido.

La segunda es considerar aristas dirigidas, para diferenciar si uno de los nodos es alumno del otro, o bien, si la relación es de igualdad en el sentido académico. De esta manera se podría tener una mejor visión de las “escuelas” de pensamiento y de quiénes las llevan.

La tercera es incluir las colaboraciones fuera del Coloquio VNL, pues varias relaciones de trabajo que existen en artículos publicados en revistas indexadas, no salieron en la red del Coloquio VNL por no haber presentado ahí el contenido de dichos textos. Para este trabajo se decidió acotar a lo que se presentó en el Coloquio VNL porque la información está organizada y al acceso dentro de la página web del mismo.

La cuarta es hacer un registro de las incorporaciones anuales a la red de actores o vínculos nuevos para estudiar, por un lado, la creación de lazos nuevos entre nodos existentes y, por otro, cómo se vinculan los nodos nuevos con los nodos preexistentes.

Anexo A

Nodos etiquetados

A continuación se presentan las diez comunidades con sus nodos etiquetados.

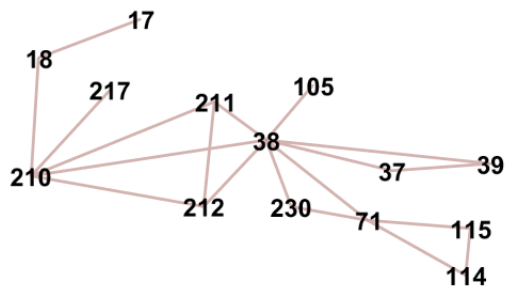


Figura A.0: Comunidad 0

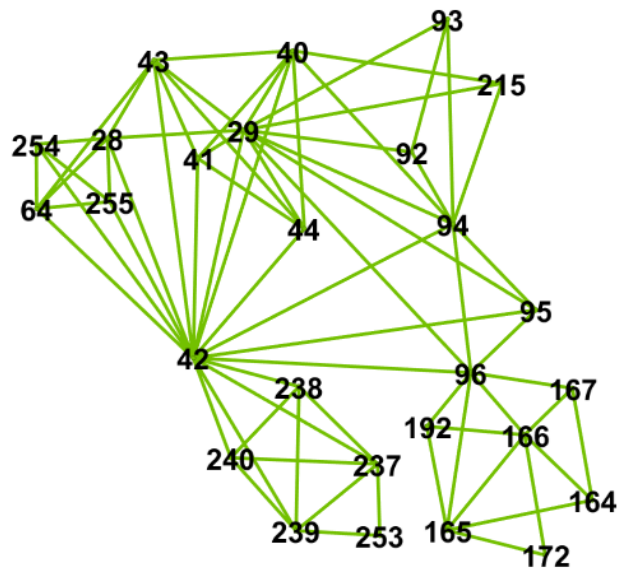


Figura A.1: Comunidad 1

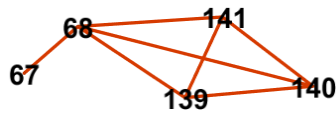


Figura A.6: Comunidad 6

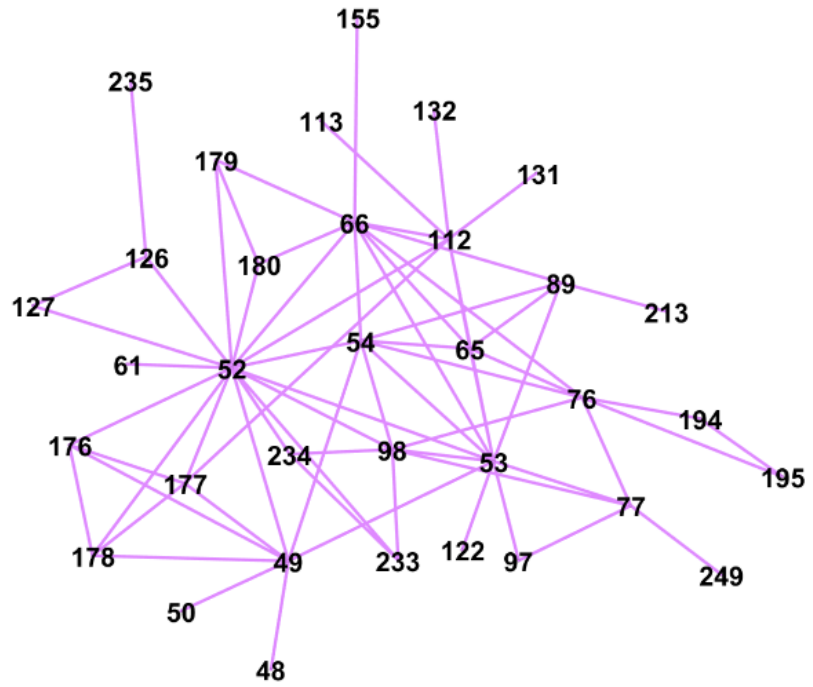


Figura A.7: Comunidad 7



Figura A.8: Comunidad 8

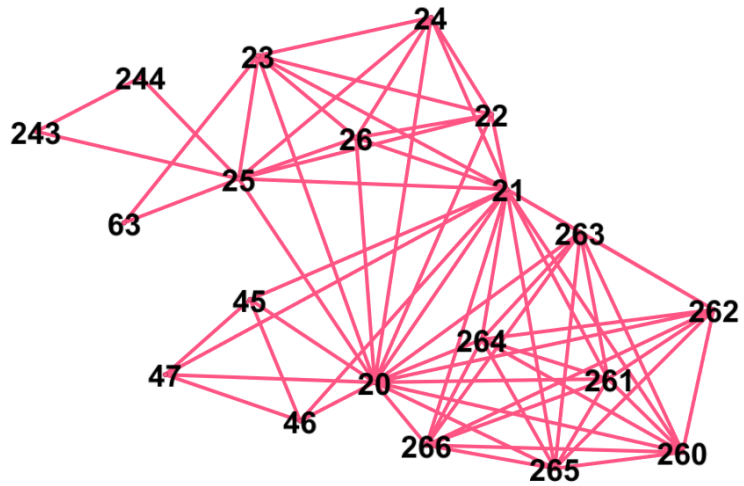


Figura A.9: Comunidad 9

Anexo B

Algoritmo BFS

El algoritmo de búsqueda exhaustiva BFS (consultado en [5]) es un algoritmo que sirve para explorar los vértices de una gráfica conexa $G = (V, A)$, construyendo un árbol generador T , en un orden dado por las listas de adyacencias de los vértices explorados. Para hacer dicha exploración de manera eficiente, el algoritmo guarda los vértices que visita en una *cola*¹, que es una lista Q que se actualiza a cada paso, ya sea añadiendo un elemento al final de ella, o bien, removiendo el primer elemento de la lista. En cualquier momento, la cola Q contiene a todos los vértices desde los cuales se puede hacer crecer el árbol construido hasta el momento.

Para comenzar, en el tiempo $t = 0$ la cola Q es vacía. En el momento que un vértice se añade al árbol, este se agrega al final de Q . En cada paso, la lista de vértices adyacentes al vértice que está al inicio de Q se revisa en busca de un nuevo vértice que se pueda añadir al árbol. Si todos los vértices en dicha lista están en el árbol, entonces se elimina el vértice de Q . El algoritmo termina cuando Q vuelve a ser vacía. Este algoritmo provee de un árbol (dado por la función *predecesor*, p , que asigna a cada vértice, el vértice desde el cual fue visitado por el algoritmo) y también da dos funciones: $\ell: V \rightarrow \mathbb{N}$ que guarda el nivel de cada vértice en el árbol, que es la distancia de la raíz r del árbol a dicho vértice en la gráfica; $t: V \rightarrow \mathbb{N}$ que guarda el tiempo en el que cada vértice se agrega al árbol. Para saber cuáles vértices ya fueron agregados al árbol se suelen colorear. Denotaremos por $G(x)$ a la gráfica G con un vértice distinguido x que será la *raíz* del árbol obtenido por el algoritmo BFS. Llamaremos *x-árbol* a un árbol enraizado en x .

¹En inglés *queue*.

Algoritmo 1: Búsqueda exhaustiva

Entrada: una gráfica conexa $G(r)$ establece $i = 0$ y $Q = \emptyset$;incrementa i en 1;colorea r de rojo;establece $\ell(r) = 0$ y $t(r) = i$;añade r al final de Q ;**mientras** Q es no vacío **hacer** considera el inicio x de Q ; **si** x tiene un vecino sin colorear **y entonces** incrementa i en 1; colorea y de rojo; establece $p(y) = x$, $\ell(y) = \ell(x) + 1$ y $t(y) = i$; añade y al final de Q ; **si no** elimina a x de Q **fin****fin****devolver** (p, ℓ, t) **Salida:** un r -árbol T en G con una función predecesor p , una función nivel ℓ tal que $\ell(v) = d_G(r, v)$ para todo $v \in V$, y una función tiempo t

Anexo C

Algoritmos de Gephi

La manera en la que se definieron las medidas en la sección 1.2 es puramente teórica, sin embargo, para obtener los valores numéricos de estas medidas para una red como la que aquí se estudia es necesario hacer uso de una computadora y no siempre es posible para la computadora hacer el cálculo como se plantea en la teoría, ya sea por el tiempo que le tomaría calcularlo o por la cantidad de memoria de la que dispone para hacer las operaciones, por lo cual la computadora calcula aproximaciones por medio de operaciones basadas en algoritmos que sí son computables. En este apartado se encuentran los algoritmos que usa el *software* de Gephi para calcular la componente más grande, las centralidades intermedia y *pagerank*, los *hubs* y autoridades, los módulos, el coeficiente de agrupamiento.

Se comenzará por explicar el algoritmo 2, obtenido en [33], que encuentra las componentes conexas y que proporciona la componente grande que es con la que se trabaja en este texto. Para entenderlo se necesitan algunas definiciones.

Dada una gráfica $G = (V, A)$, para cada vértice $v \in V$ una lista con todos los vértices w tales que $vw \in A$. A una lista así se le llama una *lista de adyacencias* de v . A un conjunto de listas de adyacencias, una por cada vértice de G , se le llama una *estructura de adyacencias* G .

Algoritmo 2: Búsqueda en profundidad (DFS)

```

inicio
  Sea  $i$  un entero;
  Procedimiento  $\text{DFS}(v, u)$ ; Comentario el vértice  $u$  es el padre del vértice
     $v$  en el árbol generador que se está construyendo, i. e.  $(u, v)$  es una flecha
    del árbol;
  inicio
     $L(v) \leftarrow i$ ;
     $i \leftarrow i + 1$ ;
    para  $w$  en la lista de vértices adyacentes a  $v$  hacer
      si  $w$  aún no ha sido etiquetado entonces
        construye la flecha  $v \rightarrow w$  en  $P$ ;
         $\text{DFS}(w, v)$ ;
      fin
      si  $L(w) < L(v)$  y  $w \neq u$  entonces
        construye la flecha  $v \dashrightarrow w$  en  $P$ ;
      fin
    fin
  fin
   $i \leftarrow 0$ ;
   $\text{DFS}(s, 0)$ ;
fin

```

Al inicio el estado de todos los vértices de G' es “por explorarse”. Se empieza en algún vértice s de G' y se elige una arista para avanzar. Al pasar por la arista se llega a un vértice nuevo. Se continúa de esta manera; en cada paso se elige una arista sin explorar que inicia en un vértice explorado y se avanza por ella. La arista lleva a un vértice, que puede ser nuevo o uno ya explorado; si el vértice al que llega es nuevo se construirá una *flecha continua* del vértice inicial al vértice final, y si el vértice al que llega ya había sido explorado se construirá una *flecha discontinua* del vértice inicial al vértice final. Cuando se acaban las aristas por explorar que inciden en vértices ya explorados, se elige un vértice nuevo sin explorar, si existe, y se empieza una nueva búsqueda. Eventualmente todas las aristas de G' serán transitadas exactamente una vez. La regla que se usa a la hora de elegir una arista para avanzar, será que esta salga del vértice que fue explorado más recientemente y que aún tenga aristas que salen de él sin ser exploradas.

El algoritmo 2 construirá una gráfica dirigida $P = (V, F)$ a partir de G , que consiste de dos conjuntos de arcos dirigidos ajenos, denotados por $v \rightarrow w$ y $v \dashrightarrow w$, respectivamente, el primero consistirá de todas las flechas continuas y el segundo de todas la flechas discontinuas. La gráfica dirigida P satisface las siguientes propiedades:

- (i) La subgráfica T que contiene a los arcos dirigidos $v \rightarrow w$ es un bosque generador, es decir, es una digráfica cuyas componentes conexas son árboles y su conjunto de vértices es V .
- (ii) Cada arco dirigido que no está en el bosque generador T de P conecta a un vértice con uno de sus antecesores en T , es decir, existe una ruta dirigida autoevasiva que va del antecesor al vértice en cuestión.

El valor de la centralidad intermedia obtenida por el algoritmo 3 debe dividirse entre dos pues nuestra gráfica es no dirigida, por lo cual cada ruta geodésica es contada dos veces.

El algoritmo 3 no sólo obtiene la centralidad intermedia sino que obtiene, para cada vértice, la distancia geodésica a cada uno de los otros vértices y una lista de los demás vértices cuyo orden (no creciente) está dado por la distancia geodésica a cada uno de ellos. Así, con los cálculos hechos por este algoritmo se obtienen también el diámetro y el radio de la red, la longitud promedio de las rutas geodésicas, la excentricidad de cada vértice, el grado de cercanía y el grado de cercanía armónico.

El procedimiento para encontrar la modularidad, obtenido en [4], hace una partición de los vértices en comunidades y a cada comunidad le asigna un número entre -1 y 1 que es la *modularidad de la partición*, misma que mide la densidad de enlaces dentro de las comunidades respecto de los enlaces que van de una comunidad a otra. Se define, en redes con peso, como:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

donde A_{ij} es el peso asociado a la arista entre i y j , $k_i = \sum_j A_{ij}$, c_i es la comunidad a la que es asignado el vértice i , la función $\delta(u, v)$ es 1 si $u = v$ y 0 en otro caso, y $m = \frac{1}{2} \sum_{i,j} A_{ij}$. En la gráfica del Coloquio VNL todas las aristas tendrán peso 1.

Algoritmo 3: Centralidad intermedia en gráficas sin peso obtenido en [6]

```

 $C_B[v] \leftarrow 0, v \in V;$ 
para  $s \in V$  hacer
     $S \leftarrow$  pila vacía;
     $P[w] \leftarrow$  lista vacía,  $w \in V;$ 
     $\sigma[t] \leftarrow 0, t \in V; \sigma[s] \leftarrow 1;$ 
     $d[t] \leftarrow -1, t \in V; d[s] \leftarrow 0;$ 
     $Q \leftarrow$  cola vacía;
    agregar a la cola  $s \rightarrow Q;$ 
    mientras  $Q$  es no vacío hacer
        quitar de la cola  $v \leftarrow Q;$ 
        insertar a la pila  $v \rightarrow S;$ 
        para cada vecino  $w$  de  $v$  hacer
            // ¿cuándo encontró a  $w$  por primera vez?
            si  $d[w] < 0$  entonces
                agregar a la cola  $w \rightarrow Q;$ 
                 $d[w] \leftarrow d[v] + 1;$ 
            fin
            // ¿cuál es la ruta con menos pasos a  $w$  que pasa por  $v$ ?
            si  $d[w] = d[v] + 1$  entonces
                 $\sigma[w] \leftarrow \sigma[w] + \sigma[v];$ 
                agregar al final de la lista  $v \rightarrow P[w];$ 
            fin
        fin
    fin
     $\delta[v] \rightarrow 0, v \in V;$ 
    //  $S$  devuelve los vértices en orden de distancia no creciente de  $s$ 
    mientras  $S$  es no vacío hacer
        quitar de la pila  $w \rightarrow S;$ 
        para  $v \in P[w]$  hacer  $\delta[v] \leftarrow \delta[v] + \frac{\sigma[v]}{\sigma[w]} (1 + \delta[w]);$ 
        si  $w \neq s$  entonces  $C_B[w] \leftarrow C_B[w] + \delta[w];$ 
    fin
fin

```

Para calcular la modularidad habrá dos fases. Iniciaremos con una red $G = (V, A)$ en la que cada arista tiene asociado un peso:

Fase 1: Se asigna cada nodo a una comunidad diferente, es decir, iniciamos con tantas comunidades como nodos. Para cada nodo i consideramos a cada vecino j de i y evaluamos la ganancia en la modularidad que obtendría dicho nodo al quitar a i de su comunidad y ponerlo en la comunidad de j . El nodo i se mueve a la comunidad en la que la ganancia es máxima, siempre que la ganancia sea positiva. Si no existe una ganancia positiva posible, i permanece en la comunidad en la que estaba. Este proceso se hace repetidas veces, siguiendo una secuencia por todos los nodos hasta que no se puede mejorar. Aquí termina la primera fase, cuando el movimiento de un solo individuo no mejora la modularidad. Cabe señalar que un nodo puede ser considerado más de una vez, de hecho, esto ocurre con frecuencia.

Fase 2: Construir una nueva red en la cual los nodos son las comunidades que se encontraron en la fase 1. Las aristas entre dos comunidades tendrán un peso asignado que está dado por la suma de los pesos de las aristas de la red original que tienen un extremo en cada una de las dos comunidades correspondientes. Además, las aristas entre nodos de la misma comunidad producirán lazos.

La nueva red se somete al procedimiento de la fase 1.

Llamaremos *paso* a una iteración completa de las dos fases. En cada paso el número de comunidades decrece y, de este modo, en el primer paso se utiliza el mayor tiempo de cómputo. Se continúa iterando pasos hasta que ya no se producen cambios en las comunidades y la modularidad alcanza un máximo.

Para calcular los *hubs* y las autoridades, Gephi utiliza el método iterativo propuesto en [17], que fue diseñado para páginas de Internet, donde los enlaces son dirigidos y van de una página a otra si en la primera hay un vínculo que lleve a la segunda. Este método actualiza numéricamente el peso de cada página para determinar cuáles de ellas son autoridades y cuáles son *hubs*.

Antes de escribir el algoritmo se requieren algunas definiciones y dos operaciones: Sea S_σ una colección de páginas de Internet con cierta relación prefijada¹, $G[S_\sigma]$ la

¹En el artículo trabajan con la *world wide web*, que tiene una cantidad de páginas con la que no se puede trabajar con un algoritmo computacional entonces, para determinar los *hubs* y las autoridades, acotan el problema a una subgráfica con cierta cantidad de páginas relacionadas, S_σ , sin embargo no se necesita hacer eso en este texto porque la red es pequeña y el *software* puede usar el algoritmo directamente sin recortar la red. No se explicará aquí el algoritmo para obtener S_σ , pero puede consultarse en [17].

gráfica dirigida inducida por este conjunto de páginas. En $G[S_\sigma]$ hay dos tipos de enlaces: los *transversales* que vinculan páginas con nombres de dominio diferentes y los *intrínsecos* que vinculan páginas con el mismo nombre de dominio (donde el *nombre de dominio* es el primer nivel de la URL asociada a la página). Los enlaces intrínsecos sirven para navegar dentro de un mismo dominio, por lo cual no aportan nada en el sentido de ser autoridad o *hub*. Por ello, se borran todos los enlaces intrínsecos y se obtiene una nueva gráfica dirigida, G_σ , que es con la que se trabaja. Así, el conjunto de vértices de G_σ es S_σ y se llamará E a su conjunto de arcos dirigidos. A cada página $p \in S_\sigma$ se le asignan dos números no negativos: el *peso de autoridad*, $x^{(p)}$, y el *peso de hub*, $y^{(p)}$. Durante el proceso estos dos números serán normalizados de manera que la suma de los cuadrados sobre todas las páginas sea igual a 1, es decir, $\sum_{p \in S_\sigma} (x^{(p)})^2 = 1$ y $\sum_{p \in S_\sigma} (y^{(p)})^2 = 1$. Al final del proceso, las páginas cuyo valor $x^{(p)}$, *x-peso*, sea más grande serán “mejores” autoridades y aquellas cuyo valor $y^{(p)}$, *y-peso*, sea más grande serán “mejores” *hubs*. Donde la palabra “mejores” se refiere a que los *hubs* y las autoridades tienen una relación en la que se refuerzan mutuamente: un buen *hub* es una página que apunta a muchas autoridades buenas y una buena autoridad es una página a la que apuntan muchos buenos *hubs*. Así, para que un *hub* sea mejor, no basta con tener un exgrado muy grande, también tiene que cumplir que varios de sus exvecinos sean autoridades y para que una autoridad sea mejor, no bastará con tener un ingrado muy grande, también tendrá que tener varios invecinos que sean *hubs*. Sean $N^+(p) = \{q \mid (p, q) \in E\}$ y $N^-(p) = \{q \mid (q, p) \in E\}$ los conjuntos de exvecinos e invecinos, respectivamente, de la página p en G_σ .

Ahora se definen las operaciones \mathcal{F} y \mathcal{O} que actualizan los *x*-pesos y los *y*-pesos, respectivamente. La operación \mathcal{F} actúa de la siguiente manera:

$$x^{(p)} \leftarrow \sum_{q \in N^-(p)} y^{(q)}.$$

La operación \mathcal{O} actúa de la siguiente manera:

$$y^{(p)} \leftarrow \sum_{q \in N^+(p)} x^{(q)}.$$

Ahora sí se puede definir el algoritmo (en pseudocódigo):

Algoritmo 4: $\text{Iterar}(G, k)$ para determinar x -pesos y y -pesos

Entrada: G una colección de n páginas enlazadas y k un número natural

Sea z el vector $(1, 1, \dots, 1) \in \mathbb{R}^n$;

$x_0 = z$;

$y_0 = z$;

$i = 1$;

repetir

 Aplicar la operación \mathcal{S} a (x_{i-1}, y_{i-1}) , para obtener los nuevos x -pesos x'_i ;

 Aplicar la operación \mathcal{O} a (x'_{i-1}, y_{i-1}) , para obtener los nuevos y -pesos y'_i ;

 Normalizar x'_i y asignar el valor normalizado a x_i ;

 Normalizar y'_i y asignar el valor normalizado a y_i ;

$i = i + 1$;

hasta que $i = k$;

Salida: (x_k, y_k)

El algoritmo $\text{Iterar}(G, k)$ proporciona dos vectores x_k y y_k cuyas entradas son los x -pesos y los y -pesos, respectivamente. Se pueden filtrar las mejores c autoridades y los mejores c hubs de la siguiente manera:

Algoritmo 5: $\text{Filtrar}(G, k, c)$ para determinar las mejores c autoridades y los mejores c hubs

Entrada: G una colección de n páginas enlazadas y k y c dos números naturales

$(x_k, y_k) = \text{Iterar}(G, k)$;

Reportar las páginas con las c coordenadas más grandes de x_k como autoridades;

Reportar las páginas con las c coordenadas más grandes de y_k como hubs.

Para calcular la centralidad *pagerank* Gephi se basa en el artículo [28], en el cual se da un orden de relevancia a las páginas de la red de Internet de acuerdo al interés que tienen para los usuarios y la atención que efectivamente les prestan.

Sea u una página de la red de Internet. Sea F_u el conjunto de páginas a las que u apunta (su conjunto de exvecinos) y sea B_u el conjunto de páginas que apuntan a u (su conjunto de invecinos). Sea $N_u = |F_u|$, el exgrado de u , y $c < 1$ un factor de normalización para que el rango total de todas las páginas sea constante.

Primero se define un orden de relevancia simple, R , de manera recursiva como

sigue: $R(u) = c \sum_{v \in B_u} \frac{R(v)}{N_v}$. Ahora, sea A una matriz cuadrada donde los renglones y las columnas corresponden a las páginas web. Sea $A_{u,v} = 1/N_u$ si hay una flecha de u a v y $A_{u,v} = 0$ en otro caso. Si se trata a R como un vector sobre las páginas, entonces tenemos que $R = cAR$. Así, R es un eigenvector de A y su eigenvalor correspondiente es c . Del conjunto de eigenvectores de A , el que interesa es el eigenvector dominante, que se puede calcular multiplicando repetidas veces la matriz A por un vector de inicio no degenerado².

Sea $E(u)$ un vector sobre las páginas web que corresponden a una fuente de rango. Entonces, el *pagerank* de un conjunto de páginas es una asignación, R' , a las páginas que satisface:

$$R'(u) = c \sum_{v \in B_u} \frac{R'(v)}{N_v} + cE(u)$$

tal que c está maximizada, $\|R'\|_1 = 1$ (donde $\|R'\|_1$ denota la norma L_1 de R') y $E(u)$ es un vector de las páginas que corresponde a una fuente de rango. Note que si E tiene entradas positivas, c debe reducirse. En notación de matrices se tiene que $R' = c(AR' + E)$. Como $\|R'\|_1 = 1$, se puede reescribir la ecuación anterior como $R' = c(A + E \times \mathbf{1})R'$ donde $\mathbf{1}$ es el vector que tiene todas sus entradas iguales a 1. Así R' es un eigenvector de $A + E \times \mathbf{1}$.

El vector E corresponde a la distribución de las páginas web que un usuario tomado al azar visita periódicamente. Los autores del artículo proponen usar un vector E uniforme sobre todas las páginas con valor α ($E(u) = \alpha$ para toda página u) o bien, que E consista de una sola página (o sea que el usuario sólo visite una página u y jamás vaya a otra por lo que $E(v) = 0$ si $v \neq u$).

Algoritmo 6: *Pagerank*

Entrada: S una colección de casi todos los vectores sobre páginas web, por ejemplo E

$R_0 \leftarrow S$;

mientras $\delta > \epsilon$ **hacer**

| |
|--------------------------------------------|
| $R_{i+1} \leftarrow AR_i$; |
| $d \leftarrow \ R_i\ _1 - \ R_{i+1}\ _1$; |
| $R_{i+1} \leftarrow R_{i+1} + dE$; |
| $\delta \leftarrow \ R_{i+1} - R_i\ _1$; |

fin

El factor d aumenta la tasa de convergencia y mantiene $\|R\|_1$.

²Para más detalles se puede consultar [28].

Ahora se verán los algoritmos de conteo de triángulos en una gráfica simple $G = (V, A)$ con n vértices y m aristas obtenidos en [18], que son los que usa Gephi para determinar el coeficiente de agrupamiento. Se llama *conteo por vértices* al conteo de triángulos a los que un vértice pertenece, para cada vértice de la gráfica. Recuerde que k_v es el grado del vértice v y $N(v)$ es el conjunto de vecinos de v .

En el algoritmo se usa una estructura de adyacencias como la del algoritmo 2, que es un objeto computacional que consiste de n listas (presentadas como arreglos), en las que se almacena la información de las adyacencias de G de una manera que economiza memoria respecto al que utilizaría la matriz de adyacencias.

Algoritmo 7: *ayz-conteo por vértices*. Cuenta, para cada $v \in V$, el número de triángulos de G que contienen a v .

Entrada: la representación de arreglo de las adyacencias de G , la matriz de adyacencias A y un entero K

para cada $v \in V$ **hacer** $T[v] = 0$;

para cada $v \in V$ **tal que** $k_v \leq K$ **hacer**

para cada par $\{u, w\}$ de vecinos de v **hacer**

si $A[u, w]$, *i. e.*, si u y w son vecinos **entonces**

 incrementa $T[v]$ en 1;

si $k_u > K$ y $k_w > K$ **entonces** incrementa $T[u]$ y $T[w]$ en 1;

si no, **si** $k_u > K$ y $w > v$ **entonces** incrementa $T[u]$;

si no, **si** $k_w > K$ y $u > v$ **entonces** incrementa $T[w]$;

 sea G' la subgráfica de G inducida por $\{v \mid k_v > K\}$;

 construye la matriz de adyacencias A' de G' ;

 calcula $(A')^3$ usando el producto de matrices rápido;

para cada $v \in V$ **tal que** $k_v > K$ **hacer**

 | incrementa $T[v]$ en $\frac{1}{2}(A')^3_{vv}$

fin

fin

fin

fin

Salida: T tal que $T[v]$ es el número de triángulos en G que contienen a v

El algoritmo 7 proporciona el número de triángulos a los que pertenece cada vértice, ahora se dará una modificación de dicho algoritmo que cuenta todos los triángulos de la gráfica. Para el cual se requieren dos procedimientos, los algoritmos 8 y 9, que se encuentran a continuación.

Algoritmo 8: Listado por vértices. Enumera todos los triángulos de G que contienen a un vértice dado.

Entrada: la representación de arreglo de las adyacencias de G , la matriz de adyacencias A y un vértice v

para cada par $\{u, w\}$ de vecinos de v **hacer**
 | **si** $A_{uw}=1$ **entonces devolver** el triángulo $\{u, v, w\}$;
fin

Salida: todos los triángulos de G que pasan por v

Algoritmo 9: Listado por aristas. Enumera todos los triángulos de G que contienen a una arista dada.

Entrada: la representación de arreglo de las adyacencias de G , la matriz de adyacencias A y una arista uv de G

para cada $w \in N(u) \cap N(v)$ **hacer**
 | **devolver** el triángulo $\{u, v, w\}$
fin

Salida: todos los triángulos de G que pasan por uv

Algoritmo 10: *ayz*-listado. Enumera todos los triángulos de G .

Entrada: la representación de arreglo de las adyacencias de G , la matriz de adyacencias A y un entero K

Salida: todos los triángulos en G

para cada $v \in V$ **tal que** $k_v \leq K$ **hacer**

para cada par $\{u, w\}$ de vecinos de v **hacer**
 | **devolver** todos los triángulos que contienen a v obtenidos del listado por
 | vértices (algoritmo 8), sin duplicados
fin

 sea G' la subgráfica de G inducida por $\{v \mid k_v > K\}$;

 construye la representación de arreglo de las adyacencias G' ;

 enlista todos los triángulos de G' usando el listado por aristas (algoritmo 9)

fin

El cálculo de la centralidad de eigenvector, a diferencia de las demás, no está basado en un artículo sino en la aplicación publicada por Gephi que puede consultarse en [44].

Anexo D

Tablas de datos

Tabla D.1: Primera tabla de medidas

| Inicio de la tabla | | | | | | |
|--------------------|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
| 0 | 0 | 0.000033 | 6 | 9 | 0.18381 | 0.226538 |
| 1 | 0 | 0.000007 | 2 | 10 | 0.155645 | 0.180796 |
| 2 | 0 | 0.000033 | 7 | 9 | 0.183985 | 0.229129 |
| 3 | 1 | 0.001197 | 12 | 7 | 0.263302 | 0.332519 |
| 4 | 1 | 0.000146 | 2 | 8 | 0.208874 | 0.240692 |
| 5 | 1 | 0.000146 | 2 | 8 | 0.208874 | 0.240692 |
| 6 | 0 | 0.000009 | 4 | 10 | 0.155897 | 0.185977 |
| 7 | 6 | 0.02882 | 3 | 9 | 0.212789 | 0.256999 |
| 8 | 6 | 0.02882 | 3 | 9 | 0.212789 | 0.256999 |
| 9 | 6 | 0.20777 | 18 | 8 | 0.268802 | 0.363718 |
| 10 | 6 | 0.029129 | 4 | 9 | 0.213024 | 0.25959 |
| 17 | 0 | 0.000001 | 1 | 10 | 0.139754 | 0.155196 |
| 18 | 0 | 0.000005 | 2 | 9 | 0.162321 | 0.187923 |
| 19 | 6 | 0.025581 | 3 | 9 | 0.212789 | 0.256999 |
| 20 | 7 | 0.369681 | 17 | 9 | 0.220824 | 0.308812 |
| 21 | 7 | 0.369681 | 17 | 9 | 0.220824 | 0.308812 |

| Continuación de la tabla D.1 | | | | | | |
|------------------------------|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
| 22 | 7 | 0.162013 | 6 | 10 | 0.198764 | 0.260303 |
| 23 | 7 | 0.177246 | 8 | 10 | 0.201883 | 0.276279 |
| 24 | 7 | 0.230123 | 12 | 9 | 0.238566 | 0.32266 |
| 25 | 7 | 0.185002 | 11 | 10 | 0.202518 | 0.284051 |
| 26 | 7 | 0.162013 | 6 | 10 | 0.198764 | 0.260303 |
| 27 | 4 | 0.022585 | 4 | 9 | 0.233939 | 0.28269 |
| 28 | 2 | 0.105007 | 10 | 8 | 0.274929 | 0.341969 |
| 29 | 2 | 0.073096 | 14 | 8 | 0.26511 | 0.335665 |
| 30 | 4 | 0.024381 | 10 | 8 | 0.256649 | 0.318807 |
| 31 | 6 | 0.024568 | 6 | 10 | 0.182075 | 0.232266 |
| 32 | 6 | 0.105 | 17 | 9 | 0.221077 | 0.309244 |
| 33 | 6 | 0.024568 | 6 | 10 | 0.182075 | 0.232266 |
| 34 | 6 | 0.022047 | 5 | 10 | 0.181904 | 0.229676 |
| 35 | 6 | 0.024568 | 6 | 10 | 0.182075 | 0.232266 |
| 36 | 6 | 0.024568 | 6 | 10 | 0.182075 | 0.232266 |
| 37 | 0 | 0.000001 | 2 | 10 | 0.141288 | 0.166077 |
| 38 | 0 | 0.000007 | 8 | 9 | 0.164255 | 0.209511 |
| 39 | 0 | 0.000001 | 2 | 10 | 0.141288 | 0.166077 |
| 40 | 2 | 0.04902 | 7 | 8 | 0.247119 | 0.298174 |
| 41 | 2 | 0.043177 | 5 | 8 | 0.244924 | 0.28997 |
| 42 | 2 | 0.147584 | 20 | 8 | 0.266207 | 0.358019 |
| 43 | 2 | 0.085206 | 10 | 7 | 0.293313 | 0.352171 |
| 44 | 2 | 0.043177 | 5 | 8 | 0.244924 | 0.28997 |
| 45 | 7 | 0.102406 | 4 | 10 | 0.181562 | 0.226826 |
| 46 | 7 | 0.102406 | 4 | 10 | 0.181562 | 0.226826 |
| 47 | 7 | 0.102406 | 4 | 10 | 0.181562 | 0.226826 |
| 48 | 3 | 0.010636 | 2 | 7 | 0.258713 | 0.286886 |
| 49 | 3 | 0.012859 | 11 | 6 | 0.302983 | 0.35544 |
| 50 | 3 | 0.001395 | 1 | 7 | 0.232811 | 0.254071 |

Continuación de la tabla D.1

| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
|----|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| 51 | 0 | 0.000004 | 1 | 10 | 0.15552 | 0.178205 |
| 52 | 3 | 0.004149 | 18 | 6 | 0.28979 | 0.366062 |
| 53 | 3 | 0.003499 | 12 | 7 | 0.267313 | 0.333259 |
| 54 | 3 | 0.003114 | 9 | 7 | 0.268428 | 0.328337 |
| 55 | 3 | 0.000407 | 4 | 8 | 0.225204 | 0.273458 |
| 56 | 3 | 0.000174 | 2 | 8 | 0.214444 | 0.248415 |
| 61 | 3 | 0.00045 | 1 | 7 | 0.224942 | 0.254577 |
| 62 | 1 | 0.000682 | 9 | 7 | 0.235366 | 0.288194 |
| 63 | 7 | 0.050678 | 3 | 10 | 0.184512 | 0.234857 |
| 64 | 2 | 0.093259 | 7 | 8 | 0.256308 | 0.314927 |
| 65 | 3 | 0.001368 | 6 | 8 | 0.220068 | 0.270984 |
| 66 | 3 | 0.002112 | 13 | 7 | 0.244613 | 0.320466 |
| 67 | 3 | 0.001572 | 2 | 7 | 0.233939 | 0.259252 |
| 68 | 3 | 0.001639 | 5 | 7 | 0.234793 | 0.267024 |
| 69 | 0 | 0.000009 | 4 | 10 | 0.155897 | 0.185977 |
| 70 | 0 | 0.000009 | 4 | 10 | 0.155897 | 0.185977 |
| 71 | 0 | 0.000001 | 4 | 10 | 0.141703 | 0.172985 |
| 72 | 3 | 0.00014 | 4 | 9 | 0.183985 | 0.224603 |
| 74 | 3 | 0.000111 | 1 | 9 | 0.182938 | 0.21424 |
| 75 | 3 | 0.001024 | 13 | 8 | 0.223638 | 0.295164 |
| 76 | 3 | 0.00098 | 7 | 8 | 0.217587 | 0.266839 |
| 77 | 3 | 0.000706 | 5 | 8 | 0.213969 | 0.255268 |
| 82 | 3 | 0.00013 | 4 | 9 | 0.183985 | 0.224603 |
| 86 | 4 | 0.003646 | 3 | 8 | 0.229216 | 0.264835 |
| 87 | 4 | 0.005749 | 6 | 9 | 0.206197 | 0.247514 |
| 88 | 2 | 0.002713 | 6 | 8 | 0.230585 | 0.274957 |
| 89 | 3 | 0.00122 | 6 | 8 | 0.222094 | 0.27513 |
| 90 | 6 | 0.025762 | 6 | 10 | 0.182075 | 0.232266 |
| 92 | 2 | 0.013934 | 3 | 9 | 0.21024 | 0.244422 |

| Continuación de la tabla D.1 | | | | | | |
|------------------------------|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
| 93 | 2 | 0.013934 | 3 | 9 | 0.21024 | 0.244422 |
| 94 | 2 | 0.041439 | 8 | 9 | 0.231971 | 0.291658 |
| 95 | 2 | 0.032127 | 4 | 9 | 0.224942 | 0.274387 |
| 96 | 5 | 0.034088 | 8 | 9 | 0.226526 | 0.286477 |
| 97 | 3 | 0.000456 | 2 | 8 | 0.212321 | 0.244041 |
| 98 | 3 | 0.001499 | 7 | 7 | 0.23623 | 0.288687 |
| 105 | 0 | 0.000001 | 1 | 10 | 0.141185 | 0.163486 |
| 106 | 4 | 0.000796 | 3 | 10 | 0.171403 | 0.197074 |
| 107 | 4 | 0.000796 | 3 | 10 | 0.171403 | 0.197074 |
| 108 | 4 | 0.000796 | 3 | 10 | 0.171403 | 0.197074 |
| 109 | 6 | 0.011388 | 1 | 10 | 0.181221 | 0.219313 |
| 111 | 1 | 0.003489 | 9 | 7 | 0.260108 | 0.311547 |
| 112 | 3 | 0.001686 | 10 | 7 | 0.242462 | 0.30924 |
| 113 | 3 | 0.000183 | 1 | 8 | 0.195344 | 0.225037 |
| 114 | 0 | 0 | 2 | 11 | 0.124276 | 0.144345 |
| 115 | 0 | 0 | 2 | 11 | 0.124276 | 0.144345 |
| 116 | 1 | 0.000424 | 2 | 8 | 0.20686 | 0.23181 |
| 117 | 1 | 0.000424 | 2 | 8 | 0.20686 | 0.23181 |
| 119 | 3 | 0.000015 | 1 | 10 | 0.155771 | 0.178279 |
| 120 | 3 | 0.000142 | 5 | 9 | 0.184336 | 0.227626 |
| 122 | 3 | 0.00038 | 1 | 8 | 0.21116 | 0.23886 |
| 123 | 1 | 0.000243 | 5 | 8 | 0.217833 | 0.265291 |
| 126 | 3 | 0.000511 | 3 | 7 | 0.225731 | 0.260622 |
| 127 | 3 | 0.000505 | 2 | 7 | 0.225467 | 0.258031 |
| 128 | 3 | 0.000016 | 2 | 10 | 0.155645 | 0.179748 |
| 129 | 3 | 0.000016 | 2 | 10 | 0.155645 | 0.179748 |
| 131 | 3 | 0.000183 | 1 | 8 | 0.195344 | 0.225037 |
| 132 | 3 | 0.000183 | 1 | 8 | 0.195344 | 0.225037 |
| 133 | 6 | 0.001402 | 1 | 11 | 0.153907 | 0.17792 |

Continuación de la tabla D.1

| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
|-----|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| 134 | 6 | 0.012928 | 3 | 10 | 0.181733 | 0.225358 |
| 135 | 3 | 0.000015 | 1 | 10 | 0.15552 | 0.177157 |
| 137 | 6 | 0.012791 | 2 | 10 | 0.181562 | 0.222767 |
| 138 | 6 | 0.002775 | 1 | 10 | 0.175614 | 0.199106 |
| 139 | 3 | 0.000227 | 3 | 8 | 0.190711 | 0.211325 |
| 140 | 3 | 0.000227 | 3 | 8 | 0.190711 | 0.211325 |
| 141 | 3 | 0.000227 | 3 | 8 | 0.190711 | 0.211325 |
| 144 | 2 | 0.000294 | 1 | 9 | 0.187561 | 0.208991 |
| 145 | 4 | 0.002644 | 1 | 9 | 0.204449 | 0.23197 |
| 147 | 6 | 0.02531 | 2 | 9 | 0.212555 | 0.254408 |
| 148 | 0 | 0.000233 | 8 | 8 | 0.22235 | 0.280144 |
| 155 | 3 | 0.000229 | 1 | 8 | 0.196738 | 0.229355 |
| 161 | 3 | 0.000002 | 1 | 11 | 0.135059 | 0.150922 |
| 162 | 3 | 0.000016 | 2 | 10 | 0.156023 | 0.181734 |
| 163 | 3 | 0.000126 | 2 | 9 | 0.183635 | 0.21899 |
| 164 | 5 | 0.001647 | 3 | 11 | 0.157294 | 0.183794 |
| 165 | 5 | 0.005133 | 5 | 10 | 0.185934 | 0.225156 |
| 166 | 5 | 0.005571 | 6 | 10 | 0.186114 | 0.227747 |
| 167 | 5 | 0.00448 | 3 | 10 | 0.185577 | 0.219975 |
| 168 | 3 | 0.000126 | 2 | 9 | 0.183635 | 0.21899 |
| 169 | 1 | 0.000026 | 1 | 9 | 0.179035 | 0.202716 |
| 170 | 1 | 0.000582 | 3 | 7 | 0.23253 | 0.269196 |
| 171 | 6 | 0.003159 | 1 | 10 | 0.175774 | 0.199969 |
| 172 | 5 | 0.001161 | 2 | 11 | 0.157166 | 0.181204 |
| 173 | 0 | 0.000025 | 1 | 9 | 0.182075 | 0.209267 |
| 175 | 1 | 0.000452 | 2 | 8 | 0.208874 | 0.237855 |
| 176 | 3 | 0.002416 | 5 | 6 | 0.264384 | 0.303886 |
| 177 | 3 | 0.002548 | 5 | 6 | 0.261872 | 0.302159 |
| 178 | 3 | 0.002383 | 4 | 6 | 0.254617 | 0.289465 |

| Continuación de la tabla D.1 | | | | | | |
|------------------------------|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
| 179 | 3 | 0.000762 | 3 | 7 | 0.231971 | 0.272021 |
| 180 | 3 | 0.000762 | 3 | 7 | 0.231971 | 0.272021 |
| 181 | 1 | 0.0001 | 2 | 8 | 0.196338 | 0.22487 |
| 182 | 6 | 0.043321 | 4 | 9 | 0.216611 | 0.271248 |
| 183 | 6 | 0.043321 | 4 | 9 | 0.216611 | 0.271248 |
| 184 | 6 | 0.043321 | 4 | 9 | 0.216611 | 0.271248 |
| 185 | 0 | 0.000088 | 4 | 8 | 0.19938 | 0.240513 |
| 187 | 1 | 0.000077 | 4 | 8 | 0.191849 | 0.224895 |
| 190 | 6 | 0.071227 | 4 | 9 | 0.221077 | 0.279279 |
| 191 | 6 | 0.071227 | 4 | 9 | 0.221077 | 0.279279 |
| 192 | 5 | 0.004858 | 3 | 10 | 0.185577 | 0.219975 |
| 193 | 3 | 0.000126 | 2 | 9 | 0.183286 | 0.217695 |
| 194 | 3 | 0.000119 | 2 | 9 | 0.179035 | 0.205 |
| 195 | 3 | 0.000119 | 2 | 9 | 0.179035 | 0.205 |
| 196 | 0 | 0.000332 | 7 | 7 | 0.231415 | 0.27793 |
| 197 | 3 | 0.000111 | 1 | 9 | 0.182938 | 0.21424 |
| 200 | 3 | 0.000273 | 2 | 8 | 0.205757 | 0.243801 |
| 201 | 1 | 0.00058 | 2 | 7 | 0.23681 | 0.276425 |
| 205 | 0 | 0.000201 | 4 | 8 | 0.219568 | 0.263737 |
| 206 | 2 | 0.00033 | 2 | 9 | 0.187743 | 0.211582 |
| 207 | 2 | 0.00033 | 2 | 9 | 0.187743 | 0.211582 |
| 208 | 4 | 0.002966 | 2 | 9 | 0.204666 | 0.234561 |
| 209 | 4 | 0.002966 | 2 | 9 | 0.204666 | 0.234561 |
| 210 | 0 | 0.000049 | 7 | 8 | 0.193193 | 0.23902 |
| 211 | 0 | 0.000007 | 3 | 9 | 0.163283 | 0.194831 |
| 212 | 0 | 0.000007 | 3 | 9 | 0.163283 | 0.194831 |
| 213 | 3 | 0.000132 | 1 | 9 | 0.181904 | 0.207505 |
| 215 | 2 | 0.018281 | 5 | 9 | 0.228673 | 0.271815 |
| 216 | 2 | 0.002277 | 2 | 9 | 0.205101 | 0.232116 |

Continuación de la tabla D.1

| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
|-----|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| 217 | 0 | 0.000005 | 1 | 9 | 0.162049 | 0.184468 |
| 218 | 1 | 0.000074 | 1 | 8 | 0.190711 | 0.214532 |
| 230 | 0 | 0.000001 | 2 | 10 | 0.141496 | 0.167804 |
| 232 | 4 | 0.002644 | 1 | 9 | 0.204449 | 0.23197 |
| 233 | 3 | 0.000687 | 3 | 7 | 0.226792 | 0.262781 |
| 234 | 3 | 0.000687 | 3 | 7 | 0.226792 | 0.262781 |
| 235 | 3 | 0.000055 | 1 | 8 | 0.184336 | 0.202461 |
| 237 | 2 | 0.024413 | 5 | 9 | 0.211623 | 0.258726 |
| 238 | 2 | 0.023894 | 4 | 9 | 0.211391 | 0.256135 |
| 239 | 2 | 0.024413 | 5 | 9 | 0.211623 | 0.258726 |
| 240 | 2 | 0.023894 | 4 | 9 | 0.211391 | 0.256135 |
| 241 | 6 | 0.034248 | 3 | 10 | 0.184512 | 0.234857 |
| 243 | 7 | 0.022507 | 2 | 11 | 0.168706 | 0.209368 |
| 244 | 7 | 0.022507 | 2 | 11 | 0.168706 | 0.209368 |
| 249 | 3 | 0.000077 | 1 | 9 | 0.176417 | 0.197105 |
| 250 | 1 | 0.000011 | 3 | 9 | 0.161371 | 0.183866 |
| 251 | 1 | 0.000011 | 3 | 9 | 0.161371 | 0.183866 |
| 252 | 1 | 0.000011 | 3 | 9 | 0.161371 | 0.183866 |
| 253 | 2 | 0.005296 | 2 | 10 | 0.174977 | 0.201289 |
| 254 | 2 | 0.042075 | 4 | 9 | 0.22133 | 0.265838 |
| 255 | 2 | 0.042075 | 4 | 9 | 0.22133 | 0.265838 |
| 256 | 3 | 0.000294 | 2 | 8 | 0.198969 | 0.239113 |
| 257 | 0 | 0.000036 | 1 | 8 | 0.188109 | 0.210005 |
| 258 | 1 | 0.000074 | 1 | 8 | 0.190711 | 0.214532 |
| 259 | 1 | 0.000378 | 1 | 8 | 0.206638 | 0.229219 |
| 260 | 7 | 0.229623 | 8 | 10 | 0.182247 | 0.237189 |
| 261 | 7 | 0.229623 | 8 | 10 | 0.182247 | 0.237189 |
| 262 | 7 | 0.229623 | 8 | 10 | 0.182247 | 0.237189 |
| 263 | 7 | 0.229623 | 8 | 10 | 0.182247 | 0.237189 |

| Continuación de la tabla D.1 | | | | | | |
|------------------------------|---------------|--------------------|-------|---------------|-------------------|----------------------------|
| Id | Clase modular | Peso de <i>hub</i> | Grado | Excentricidad | Grado de cercanía | Grado de cercanía armónico |
| 264 | 7 | 0.229623 | 8 | 10 | 0.182247 | 0.237189 |
| 265 | 7 | 0.229623 | 8 | 10 | 0.182247 | 0.237189 |
| 266 | 7 | 0.229623 | 8 | 10 | 0.182247 | 0.237189 |
| 267 | 0 | 0.000036 | 1 | 8 | 0.188109 | 0.210005 |
| Fin de la tabla D.1 | | | | | | |

Tabla D.2: Segunda tabla de medidas

| Inicio de la tabla | | | | | |
|--------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 0 | 375.5 | 0.007414 | 0.533333 | 8 | 0.060006 |
| 1 | 0 | 0.002911 | 1 | 1 | 0.023785 |
| 2 | 567.5 | 0.008958 | 0.380952 | 8 | 0.062522 |
| 3 | 3891.25386 | 0.014979 | 0.136364 | 9 | 0.192902 |
| 4 | 0 | 0.00319 | 1 | 1 | 0.035976 |
| 5 | 0 | 0.00319 | 1 | 1 | 0.035976 |
| 6 | 0 | 0.005062 | 1 | 6 | 0.040167 |
| 7 | 0 | 0.003673 | 1 | 3 | 0.122619 |
| 8 | 0 | 0.003673 | 1 | 3 | 0.122619 |
| 9 | 5535.74055 | 0.015946 | 0.176471 | 27 | 0.822095 |
| 10 | 192 | 0.005205 | 0.5 | 3 | 0.125138 |
| 17 | 0 | 0.0026 | 0 | 0 | 0.003731 |
| 18 | 192 | 0.004298 | 0 | 0 | 0.013205 |
| 19 | 192 | 0.004434 | 0.333333 | 1 | 0.107595 |
| 20 | 1003.42236 | 0.011442 | 0.375 | 51 | 1 |
| 21 | 1003.42236 | 0.011442 | 0.375 | 51 | 1 |

| Continuación de la tabla D.2 | | | | | |
|------------------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 22 | 0 | 0.004563 | 1 | 15 | 0.46436 |
| 23 | 88.271429 | 0.006157 | 0.642857 | 18 | 0.535132 |
| 24 | 1223.87924 | 0.009018 | 0.424242 | 28 | 0.788041 |
| 25 | 522.499784 | 0.009151 | 0.363636 | 20 | 0.566429 |
| 26 | 0 | 0.004563 | 1 | 15 | 0.46436 |
| 27 | 197.455556 | 0.004371 | 0.666667 | 4 | 0.160525 |
| 28 | 2150.10937 | 0.008627 | 0.422222 | 19 | 0.534451 |
| 29 | 1705.86255 | 0.012395 | 0.296703 | 27 | 0.503355 |
| 30 | 3077.17313 | 0.012346 | 0.155556 | 7 | 0.203162 |
| 31 | 0.2 | 0.005411 | 0.933333 | 14 | 0.167717 |
| 32 | 2162.43297 | 0.015542 | 0.191176 | 26 | 0.479824 |
| 33 | 0.2 | 0.005411 | 0.933333 | 14 | 0.167717 |
| 34 | 0 | 0.004617 | 1 | 10 | 0.148248 |
| 35 | 0.2 | 0.005411 | 0.933333 | 14 | 0.167717 |
| 36 | 0.2 | 0.005411 | 0.933333 | 14 | 0.167717 |
| 37 | 0 | 0.003717 | 1 | 1 | 0.014548 |
| 38 | 1316 | 0.012833 | 0.178571 | 5 | 0.045126 |
| 39 | 0 | 0.003717 | 1 | 1 | 0.014548 |
| 40 | 126.865152 | 0.006206 | 0.666667 | 14 | 0.329895 |
| 41 | 0 | 0.004476 | 1 | 10 | 0.280426 |
| 42 | 2303.83462 | 0.016436 | 0.226316 | 43 | 0.796236 |
| 43 | 5964.04202 | 0.00868 | 0.4 | 18 | 0.507477 |
| 44 | 0 | 0.004476 | 1 | 10 | 0.280426 |
| 45 | 0 | 0.003333 | 1 | 6 | 0.26904 |
| 46 | 0 | 0.003333 | 1 | 6 | 0.26904 |
| 47 | 0 | 0.003333 | 1 | 6 | 0.26904 |
| 48 | 0 | 0.002447 | 1 | 1 | 0.097244 |
| 49 | 6695.47182 | 0.01212 | 0.2 | 11 | 0.301261 |
| 50 | 0 | 0.00171 | 0 | 0 | 0.039153 |

| Continuación de la tabla D.2 | | | | | |
|------------------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 51 | 0 | 0.001861 | 0 | 0 | 0.012222 |
| 52 | 4003.09822 | 0.020214 | 0.156863 | 24 | 0.427605 |
| 53 | 1716.45503 | 0.013577 | 0.257576 | 17 | 0.3464 |
| 54 | 2246.98282 | 0.009727 | 0.472222 | 17 | 0.325911 |
| 55 | 319.591481 | 0.005277 | 0.333333 | 2 | 0.09208 |
| 56 | 0 | 0.002956 | 1 | 1 | 0.043197 |
| 61 | 0 | 0.001728 | 0 | 0 | 0.05698 |
| 62 | 1379.12381 | 0.012717 | 0.166667 | 6 | 0.099418 |
| 63 | 0 | 0.002911 | 1 | 3 | 0.178141 |
| 64 | 166.069791 | 0.005911 | 0.666667 | 14 | 0.438374 |
| 65 | 17.447222 | 0.006754 | 0.666667 | 10 | 0.218663 |
| 66 | 1095.27532 | 0.014829 | 0.25641 | 20 | 0.350699 |
| 67 | 0 | 0.002841 | 1 | 1 | 0.04966 |
| 68 | 570 | 0.006653 | 0.4 | 4 | 0.061923 |
| 69 | 0 | 0.005062 | 1 | 6 | 0.040167 |
| 70 | 0 | 0.005062 | 1 | 6 | 0.040167 |
| 71 | 382 | 0.007035 | 0.333333 | 2 | 0.021179 |
| 72 | 193.5 | 0.006638 | 0.333333 | 2 | 0.04999 |
| 74 | 0 | 0.001954 | 0 | 0 | 0.03173 |
| 75 | 2598.88889 | 0.018049 | 0.115385 | 9 | 0.218642 |
| 76 | 399.307143 | 0.008308 | 0.285714 | 6 | 0.171035 |
| 77 | 201.341667 | 0.006498 | 0.3 | 3 | 0.109019 |
| 82 | 382 | 0.006575 | 0.333333 | 2 | 0.04756 |
| 86 | 168.544444 | 0.004042 | 0.666667 | 2 | 0.059167 |
| 87 | 570.5 | 0.0075 | 0.333333 | 5 | 0.073674 |
| 88 | 856.971429 | 0.008526 | 0.133333 | 2 | 0.082145 |
| 89 | 235.928571 | 0.007296 | 0.533333 | 8 | 0.198807 |
| 90 | 2 | 0.005477 | 0.733333 | 11 | 0.165798 |
| 92 | 0 | 0.003224 | 1 | 3 | 0.109944 |

| Continuación de la tabla D.2 | | | | | |
|------------------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 93 | 0 | 0.003224 | 1 | 3 | 0.109944 |
| 94 | 172.376768 | 0.007391 | 0.464286 | 13 | 0.295934 |
| 95 | 0 | 0.003842 | 1 | 6 | 0.219015 |
| 96 | 1122.83333 | 0.007833 | 0.357143 | 10 | 0.252762 |
| 97 | 0 | 0.002839 | 1 | 1 | 0.061022 |
| 98 | 144.441667 | 0.00784 | 0.428571 | 9 | 0.210837 |
| 105 | 0 | 0.002137 | 0 | 0 | 0.010534 |
| 106 | 0 | 0.004238 | 1 | 3 | 0.02303 |
| 107 | 0 | 0.004238 | 1 | 3 | 0.02303 |
| 108 | 0 | 0.004238 | 1 | 3 | 0.02303 |
| 109 | 0 | 0.00155 | 0 | 0 | 0.058564 |
| 111 | 2994.40647 | 0.012243 | 0.166667 | 6 | 0.11594 |
| 112 | 1092.61576 | 0.01281 | 0.177778 | 8 | 0.258122 |
| 113 | 0 | 0.001862 | 0 | 0 | 0.034797 |
| 114 | 0 | 0.003946 | 1 | 1 | 0.008904 |
| 115 | 0 | 0.003946 | 1 | 1 | 0.008904 |
| 116 | 0 | 0.003356 | 1 | 1 | 0.024153 |
| 117 | 0 | 0.003356 | 1 | 1 | 0.024153 |
| 119 | 0 | 0.002259 | 0 | 0 | 0.00949 |
| 120 | 572.5 | 0.008745 | 0.2 | 2 | 0.051982 |
| 122 | 0 | 0.001735 | 0 | 0 | 0.045524 |
| 123 | 333.516667 | 0.007177 | 0.3 | 3 | 0.073881 |
| 126 | 192 | 0.004883 | 0.333333 | 1 | 0.06937 |
| 127 | 0 | 0.003111 | 1 | 1 | 0.067305 |
| 128 | 0 | 0.003776 | 1 | 1 | 0.012164 |
| 129 | 0 | 0.003776 | 1 | 1 | 0.012164 |
| 131 | 0 | 0.001862 | 0 | 0 | 0.034797 |
| 132 | 0 | 0.001862 | 0 | 0 | 0.034797 |
| 133 | 0 | 0.002045 | 0 | 0 | 0.010149 |

| Continuación de la tabla D.2 | | | | | |
|------------------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 134 | 192 | 0.004488 | 0.333333 | 1 | 0.070748 |
| 135 | 0 | 0.002184 | 0 | 0 | 0.00889 |
| 137 | 0 | 0.002822 | 1 | 1 | 0.068713 |
| 138 | 0 | 0.002029 | 0 | 0 | 0.013367 |
| 139 | 0 | 0.004395 | 1 | 3 | 0.020566 |
| 140 | 0 | 0.004395 | 1 | 3 | 0.020566 |
| 141 | 0 | 0.004395 | 1 | 3 | 0.020566 |
| 144 | 0 | 0.001981 | 0 | 0 | 0.012961 |
| 145 | 0 | 0.001823 | 0 | 0 | 0.027693 |
| 147 | 0 | 0.002782 | 1 | 1 | 0.105328 |
| 148 | 1858.35743 | 0.010251 | 0.214286 | 6 | 0.105627 |
| 155 | 0 | 0.001743 | 0 | 0 | 0.046452 |
| 161 | 0 | 0.002717 | 0 | 0 | 0.003373 |
| 162 | 192 | 0.004567 | 0 | 0 | 0.011409 |
| 163 | 0 | 0.00344 | 1 | 1 | 0.04122 |
| 164 | 0.333333 | 0.00377 | 0.666667 | 2 | 0.030748 |
| 165 | 158.166667 | 0.005821 | 0.5 | 5 | 0.063564 |
| 166 | 160 | 0.006871 | 0.466667 | 7 | 0.070544 |
| 167 | 62.666667 | 0.003647 | 0.666667 | 2 | 0.050009 |
| 168 | 0 | 0.00344 | 1 | 1 | 0.04122 |
| 169 | 0 | 0.001993 | 0 | 0 | 0.012644 |
| 170 | 0 | 0.004192 | 1 | 3 | 0.066056 |
| 171 | 0 | 0.001879 | 0 | 0 | 0.015645 |
| 172 | 0 | 0.002736 | 1 | 1 | 0.022552 |
| 173 | 0 | 0.001862 | 0 | 0 | 0.018045 |
| 175 | 0 | 0.003131 | 1 | 1 | 0.036471 |
| 176 | 2528.38515 | 0.005808 | 0.6 | 6 | 0.151265 |
| 177 | 229.310289 | 0.005723 | 0.7 | 7 | 0.170254 |
| 178 | 0 | 0.004624 | 1 | 6 | 0.13944 |

| Continuación de la tabla D.2 | | | | | |
|------------------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 179 | 0 | 0.003763 | 1 | 3 | 0.11948 |
| 180 | 0 | 0.003763 | 1 | 3 | 0.11948 |
| 181 | 0 | 0.003195 | 1 | 1 | 0.029977 |
| 182 | 0 | 0.004006 | 1 | 6 | 0.197701 |
| 183 | 0 | 0.004006 | 1 | 6 | 0.197701 |
| 184 | 0 | 0.004006 | 1 | 6 | 0.197701 |
| 185 | 302.866667 | 0.005487 | 0.666667 | 4 | 0.05504 |
| 187 | 570 | 0.00599 | 0.5 | 3 | 0.031032 |
| 190 | 0 | 0.003636 | 1 | 6 | 0.30122 |
| 191 | 0 | 0.003636 | 1 | 6 | 0.30122 |
| 192 | 0 | 0.003569 | 1 | 3 | 0.054478 |
| 193 | 0 | 0.003364 | 1 | 1 | 0.04062 |
| 194 | 0 | 0.003099 | 1 | 1 | 0.028511 |
| 195 | 0 | 0.003099 | 1 | 1 | 0.028511 |
| 196 | 2844.08515 | 0.009787 | 0.190476 | 4 | 0.078519 |
| 197 | 0 | 0.001954 | 0 | 0 | 0.03173 |
| 200 | 0 | 0.002864 | 1 | 1 | 0.060064 |
| 201 | 0 | 0.002789 | 1 | 1 | 0.086564 |
| 205 | 269.09076 | 0.005278 | 0.666667 | 4 | 0.071442 |
| 206 | 0 | 0.003446 | 1 | 1 | 0.016994 |
| 207 | 0 | 0.003446 | 1 | 1 | 0.016994 |
| 208 | 0 | 0.00317 | 1 | 1 | 0.033558 |
| 209 | 0 | 0.00317 | 1 | 1 | 0.033558 |
| 210 | 2372 | 0.010822 | 0.190476 | 4 | 0.053883 |
| 211 | 0 | 0.004816 | 1 | 3 | 0.028135 |
| 212 | 0 | 0.004816 | 1 | 3 | 0.028135 |
| 213 | 0 | 0.001807 | 0 | 0 | 0.026303 |
| 215 | 531.067965 | 0.005513 | 0.4 | 4 | 0.153118 |
| 216 | 0 | 0.002918 | 1 | 1 | 0.032714 |

| Continuación de la tabla D.2 | | | | | |
|------------------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 217 | 0 | 0.002088 | 0 | 0 | 0.011239 |
| 218 | 0 | 0.001974 | 0 | 0 | 0.017333 |
| 230 | 0 | 0.003632 | 1 | 1 | 0.016278 |
| 232 | 0 | 0.001823 | 0 | 0 | 0.027693 |
| 233 | 0 | 0.003739 | 1 | 3 | 0.098617 |
| 234 | 0 | 0.003739 | 1 | 3 | 0.098617 |
| 235 | 0 | 0.002156 | 0 | 0 | 0.010325 |
| 237 | 95.5 | 0.005167 | 0.7 | 7 | 0.154065 |
| 238 | 0 | 0.004099 | 1 | 6 | 0.148862 |
| 239 | 95.5 | 0.005167 | 0.7 | 7 | 0.154065 |
| 240 | 0 | 0.004099 | 1 | 6 | 0.148862 |
| 241 | 5.416667 | 0.003033 | 0.666667 | 2 | 0.14259 |
| 243 | 0 | 0.002574 | 1 | 1 | 0.071005 |
| 244 | 0 | 0.002574 | 1 | 1 | 0.071005 |
| 249 | 0 | 0.001878 | 0 | 0 | 0.015499 |
| 250 | 0 | 0.004723 | 1 | 3 | 0.01612 |
| 251 | 0 | 0.004723 | 1 | 3 | 0.01612 |
| 252 | 0 | 0.004723 | 1 | 3 | 0.01612 |
| 253 | 0 | 0.00253 | 1 | 1 | 0.039038 |
| 254 | 0 | 0.003711 | 1 | 6 | 0.226444 |
| 255 | 0 | 0.003711 | 1 | 6 | 0.226444 |
| 256 | 0 | 0.003042 | 1 | 1 | 0.066528 |
| 257 | 0 | 0.001962 | 0 | 0 | 0.013799 |
| 258 | 0 | 0.001974 | 0 | 0 | 0.017333 |
| 259 | 0 | 0.00193 | 0 | 0 | 0.019138 |
| 260 | 0 | 0.005284 | 1 | 28 | 0.565034 |
| 261 | 0 | 0.005284 | 1 | 28 | 0.565034 |
| 262 | 0 | 0.005284 | 1 | 28 | 0.565034 |
| 263 | 0 | 0.005284 | 1 | 28 | 0.565034 |

| Continuación de la tabla D.2 | | | | | |
|------------------------------|------------------------|-----------------|------------------------------|----------------------|----------------------------|
| Id | Centralidad intermedia | <i>Pagerank</i> | Coefficiente de agrupamiento | Número de triángulos | Centralidad de eigenvector |
| 264 | 0 | 0.005284 | 1 | 28 | 0.565034 |
| 265 | 0 | 0.005284 | 1 | 28 | 0.565034 |
| 266 | 0 | 0.005284 | 1 | 28 | 0.565034 |
| 267 | 0 | 0.001962 | 0 | 0 | 0.013799 |
| Fin de la tabla D.2 | | | | | |

Bibliografía

- [1] Anónimo. *Biblia de Jerusalem*. Descleé de Brouwer, Bilbao, 1967.
- [2] A.-L. Barabási y R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [3] A. Barrat, M. Barthélemy, y A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge University Press, New York, 1 edition, 2008.
- [4] V. D. Blondel, J. L. Guillaume, R. Lambiotte, y E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10:1000, 2008.
- [5] J. A. Bondy y U. S. R. Murty. *Graph Theory*. Springer, Springer-Verlag London, Ltd, London, 1 edition, 2008.
- [6] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.
- [7] S. C. Carlson. Königsberg bridge problem. <https://www.britannica.com/science/Konigsberg-bridge-problem>, 2010. Fecha de consulta: 2019-07-12.
- [8] P. Erdős y A. Rényi. On random graphs. *Publ. Math.*, 6:290–297, 1959.
- [9] P. Erdős y A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–60, 1960.
- [10] P. Erdős y A. Rényi. On the strength of connectedness of random graphs. *Acta. Math. Sci. Hung.*, 12:261–267, 1961.
- [11] M. J. Gage. Woman as an inventor. <https://archive.org/details/jstor-25118273/page/n11>. Fecha de consulta: 2019-06-18.

- [12] R. Ganguli y S. Roy. A study on course timetable scheduling using graph coloring approach. *Int. Journal of Comp. and Appl. Math.*, 12:469–485, 2017.
- [13] E. N. Gilbert. Random graphs. *Ann. Math. Stat.*, 30:1141–1144, 1959.
- [14] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360–1380, 1973.
- [15] M. S. Granovetter. The strength of weak ties: A network theory revisited. In P. Marsden y N. Linn, editores, *Social structure and network analysis*, pages 105–130. Sage, Beverly Hills, Calif., 1982.
- [16] C. Kadushin. *Understanding Social Networks. Theories, Concepts and Findings*. Oxford University Press, 198 Madison Avenue, New York, New York 10016, 2012.
- [17] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the AMC*, 46(5):604–632, 1999.
- [18] M. Latapy. Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theoretical Computer Science (TCS)*, 407(1-3):458–473, 2008.
- [19] E. O. Laumann y Y. Youm. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the united states: a network explanation. *Sexually transmitted diseases*, 26(5):250–261, 1999.
- [20] R. K. Merton. The Matthew effect in science. *Science*, 159:56–63, 1968.
- [21] S. Milgram. The small world problem. *Psychol. Today*, 2:60–67, 1967.
- [22] M. Molloy y B. Reed. A critical point for random graphs with given degree sequence. *Random Struct. Algor.*, 6:161–179, 1995.
- [23] J. L. Moreno. *Who shall survive?* Beacon House, Inc., Beacon, N. Y., 1934.
- [24] M. E. J. Newman. Scientific collaboration networks. I. network construction and fundamental results. *Phys. Rev. E*, 64:016131, 2001.
- [25] M. E. J. Newman. Scientific collaboration networks. II. shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 64:016132, 2001.

Bibliografía

- [26] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. Natl Acad. Sci.(USA)*, 98:404–409, 2001.
- [27] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, Oxford University Press Inc., New York, 2010.
- [28] L. Page, S. Brin, R. Motwani, y T. Winograd. The pagerank citation ranking: Bringing order to the web. Reporte técnico, Stanford University, InfoLab, 1999.
- [29] D. J. de S. Price. Networks of scientific papers. *Science*, 149:510–515, 1965.
- [30] D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform.*, 27:292–306, 1976.
- [31] B. Resnick. Nettie stevens discovered xy sex chromosomes. she didn't get credit because she had two x's. <https://www.vox.com/2016/7/7/12105830/nettie-stevens-genetics-gender-sex-chromosomes>. Fecha de consulta: 2019-06-18.
- [32] F. Sánchez. Pagerank. <https://www.humanlevel.com/diccionario-marketing-online/pagerank-google>. Fecha de consulta: 2019-06-18.
- [33] R. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.
- [34] J. Travers y S. Milgram. An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.
- [35] D. J. Watts y S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
- [36] Página web: Coloquio VNL 2013. <http://xamanek.izt.uam.mx/coloquio/2013/>. Fecha de consulta: 2017-04-14.
- [37] Página web: Coloquio VNL 2014. <http://xamanek.izt.uam.mx/coloquio/2014/>. Fecha de consulta: 2017-04-11.
- [38] Página web: Coloquio VNL 2015. <http://xamanek.izt.uam.mx/coloquio/2015/>. Fecha de consulta: 2017-04-09.

- [39] Página web: Coloquio VNL 2016. <http://xamanek.izt.uam.mx/coloquio/2016/>. Fecha de consulta: 2017-04-08.
- [40] Página web: Coloquio VNL 2017. <http://xamanek.izt.uam.mx/coloquio/2017/>. Fecha de consulta: 2017-04-08.
- [41] Página web: Coloquio VNL 2018. <http://xamanek.izt.uam.mx/coloquio/2018/>. Fecha de consulta: 2018-04-17.
- [42] Página web: Geogebra. <https://www.geogebra.org/m/AahnrwUv>. Fecha de consulta: 2018-09-20.
- [43] Página web: Gephi. <https://gephi.org/>. Fecha de consulta: 2018-09-15.
- [44] Página web: Gephi centralidad del eigenvector. <https://gephi.org/gephi/0.9.1/apidocs/org/gephi/statistics/plugin/EigenvectorCentrality.html>. Fecha de consulta: 2019-02-03.
- [45] Página web: WolframAlpha. <https://www.wolframalpha.com/widgets/view.jsp?id=298f587406c914fad5373bb689300433>. Fecha de consulta: 2019-02-15.

