



**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL
INSTITUTO POLITÉCNICO NACIONAL**

Unidad Mérida

DEPARTAMENTO DE FÍSICA APLICADA

**Modularidad, balance y frustración en redes sociales escolares:
el rol de las relaciones negativas en las comunidades**

TESIS

Que presenta

José Alejandro Brito Montes

Para obtener el grado de

Doctor en Ciencias

en la especialidad de

Física Teórica

Director de Tesis:

Dr. Rodrigo Huerta Quintanilla

Mérida, Yucatán, México

Julio de 2023

Agradecimientos

Agradezco a mi madre María que con su demostración de una madre ejemplar me ha enseñado a no rendirme ante nada y siempre perseverar a través de sus consejos.

A mis hermanos y demás familia, porque me han brindado su apoyo incondicional y por compartir conmigo buenos y malos momentos.

A mis amigos y compañeros por hacer más ameno todo este tiempo.

Agradezco infinitamente al Dr. Rodrigo Huerta, por su tiempo, paciencia, guía y apoyo durante todo el proyecto, y por compartir sus conocimientos y experiencias.

Al M. en C. Efraín Canto Lugo por sus enseñanzas y apoyo a lo largo del Doctorado y específicamente en la revisión de tesis y elaboración de software.

A la Dra. Ana María Hernández quien aportó los datos con las que se realizó la tesis, así como su apoyo en el Doctorado.

A mis sinodales Dra. María Vargas, Dr. Gabriel Pérez, Dr. Raúl Hernández y Dr. Román Castro, por sus comentarios y su valiosa ayuda en el trayecto final del trabajo.

Agradezco al CONAHCYT por el apoyo económico durante los estudios de Doctorado.

Gracias a todas las personas que ayudaron directamente e indirectamente en la realización de este proyecto.

Índice general

Índice de figuras	5
Índice de tablas	9
Lista de abreviaturas	10
Resumen	11
Abstract	12
Introducción	13
1. Teoría de redes	19
1.1. Representación de una red: Matriz de adyacencia	19
1.2. Propiedades de las redes	21
1.2.1. Grado medio	21
1.2.2. Distribución de grado	21
1.2.3. Densidad	21
1.2.4. Caminos.	22
1.2.5. Intermediación	23
1.3. Comunidades	24

<i>ÍNDICE GENERAL</i>	3
1.3.1. Partición	24
1.3.2. Introducción a la modularidad	25
1.3.2.1. Modularidad: Una perspectiva con la Física	25
1.3.3. Algoritmos para la detección de comunidades	27
1.3.3.1. Algoritmo de Newman	27
1.3.3.2. Algoritmo de Louvain	28
1.3.3.3. Algoritmo de Girvan-Newman	30
1.3.4. Club de Karate de Zachary	31
1.3.5. Distribución de comunidades	32
2. Redes sociales	34
2.1. Teoría de Dunbar: Proporción de Dunbar	35
2.2. Redes signadas.	36
2.2.1. Propiedades de los enlaces negativos en las redes	37
2.2.1.1. Reciprocidad	37
2.2.1.2. Transitividad	37
2.3. Comunidades signadas	38
2.4. Modularidad para redes signadas	38
2.5. Porcentaje de Perturbación Comunitario	39
2.6. Teoría del Equilibrio Social: Balance y frustración.	39
2.6.1. Balance	39
2.6.2. Frustración	40
3. Modelo de formación de redes signadas en escuelas	42
3.1. Recopilación de datos	43

<i>ÍNDICE GENERAL</i>	4
3.2. Construcción de las matrices de adyacencia	44
3.3. Proporción de Dunbar: Datos empíricos	45
3.4. Missing links	47
3.5. Calibración del Método de Louvain	49
3.6. Construcción de las comunidades signadas	50
4. Resultados	51
4.1. Proporción de Dunbar para los enlaces negativos	52
4.2. Análisis de las comunidades signadas	53
4.2.1. Distribución de comunidades	55
4.2.2. Subcomunidades antagónicas	56
4.2.3. Modularidad, Balance y Frustración	57
5. Discusión	61
6. Conclusiones y trabajos a futuro	65
Referencias	67
A. Información de soporte	74
A.1. Cuestionarios: Universidad	74
A.2. Balance Social de las redes signadas	81
A.3. Boxplot para el método de Louvain	81
A.4. Programas Importantes	84

Índice de figuras

1.	En (A) se indican los 7 puentes de la ciudad de Königsberg. Cada región es representada por las letras a,b,c,d. En (B) es la representación de la ciudad en una red, donde cada punto representa una región diferente y cada puente es representado por enlace que las une. Imagen de dominio público	14
1.1.	Ejemplo de una red con 11 nodos. En (A), la red es dirigida con 20 enlaces, mientras que en (B) el caso no dirigido tiene 16 enlaces.	20
1.2.	Matrices de adyacencia para las redes de la Figura 1.1. En (A) es el caso dirigido, y en (B) el caso no dirigido.	20
1.3.	Caminos. Red de 6 nodos y 7 enlaces. Una secuencia de nodos conectados uno con otro por un enlace, es un camino. La longitud del camino, es el número de enlaces que lo componen. En este ejemplo, todos los nodos de la red están conectados por un camino, a esto se le conoce como una red conectada.	22
1.4.	Intermediación de enlaces. En la red, el enlace entre los nodos 3 y 4, es el que tiene el valor de intermediación más alto de todos los enlaces.	23
1.5.	Comunidades de la red de la Figura 1.1 (B). Esta imagen muestra una estructura de tres comunidades, los nodos de cada comunidad están densamente conectados internamente y conexiones escasas entre cada comunidad.	24

1.6. **Algoritmo de Louvain.** La red inicial contiene 16 nodos y 28 enlaces. El algoritmo parte desde una partición trivial donde cada nodo está en una comunidad. Inmediatamente, el algoritmo mueve nodos individuales de una comunidad a otra, hasta llegar a una partición local (A). Luego, a partir de esta partición, se construye una nueva red pesada (B), cuyos nodos corresponden a las comunidades encontradas en (A), se generan los autoenlaces que son el grado de los nodos en cada comunidad (2,4,14,16) y los enlaces ponderados son el número de enlaces entre las comunidades. A continuación, se vuelven agrupar los nodos al azar de la red pesada (C). Estos pasos se repiten hasta que la modularidad no pueda mejorar, y el algoritmo se detiene (D). 29

1.7. **Club de Karate de Zachary.** Comunidades encontradas por los algoritmos: (A) Método de Girvan-Newman, (B) Método de Louvain y (C) Método de Newman. Las comunidades resultantes se indican por colores. 32

2.1. **Círculos de Dunbar.** El punto representa a la persona cuya red social está ilustrada. Los círculos concéntricos representan el número e intensidad de las relaciones que conforman la red. Imagen modificada de [54] . 35

2.2. **Red signada.** Esta red tiene 11 nodos con 16 enlaces, de los cuales 11 son enlaces positivos (negro) y 5 negativos(rojo). 37

2.3. **Balance social en redes signadas.** Cuatro estados posibles en redes signadas. (A) y (B) son estados balanceados, mientras que (C) y (D) son estados desbalanceados. Los símbolos (+) indica relación positiva (amistad), y (-) indica relación negativa (enemistad). 40

2.4. **Comunidades signadas.** Esta red esta dividida en 3 comunidades signadas (azul, roja y verde), tiene 12 nodos, 24 enlaces, de los cuales 17 son enlaces positivos y 7 negativos. Notar que esta estructura comunitaria, coincide con la definición de comunidad signada. Los enlaces azules indican relaciones positivos, mientras que los rojos son negativos. . . . 41

3.1. **Mapa del Estado de Yucatán.** El mapa muestra las ubicaciones donde se aplicaron las encuestas. Dos de las tres escuelas primarias se encuentran en Abalá: la epSC en Temozón Sur y la epRRC en el centro del municipio. La primaria epIZ está en Mérida. Las secundarias están localizadas como sigue: esRDC en Umán, esTN2 en Mérida y esJLBG en Hunucmá. De las dos preparatorias, la eprHUN se ubica en Hunucmá y la eprCCP en Mérida. Por último, la universidad unTRS está en Tekax. Los municipios de Abalá, Hunucmá, Tekax son zonas rurales, mientras que Mérida y Umán no. 43

3.2. **Proporción de Dunbar para los datos empíricos.** A) valores obtenidos de la razón de relaciones de amigos entre mejores amigos. B) valores obtenidos de la razón de relaciones de enemigos entre peores enemigos. La línea discontinua que pasa por tres, indica el valor de referencia de Dunbar. 46

3.3. **Heurística para agregar los missing links.** Enlaces azules y rojos representan los enlaces recíprocos de amistad (“mejor amigos”) y enemistad (“peores enemigos”), respectivamente. La línea roja punteada indica un enlace negativo unidireccional, esto es, solamente si el estudiante i dijo ser enemigo de j , o viceversa. La figura muestra, que si este enlace fuera recíproco, habría tres triadas balanceadas y una desbalanceada. En este caso, el enlace negativo unidireccional es agregado como un enlace negativo recíproco. Imagen tomada y modificada de [74] 48

3.4. **Calibración del Algoritmo de Louvain.** En la figura graficamos el número promedio de comunidades y su desviación estándar en función del parámetro \mathbf{R} en el intervalo $[0.25,1.75]$. (A) calibración de la escuela primaria epIZ, el valor óptimo de \mathbf{R} es 0.85. (B) calibración de la secundaria esRDC, el valor óptimo de \mathbf{R} es 1.20. C) calibración de la universidad unTRS, el valor óptimo de \mathbf{R} es 1.30. Cada valor fue obtenido de 300 simulaciones. 50

4.1. **Proporción Dunbar para los enlaces negativos.** Los valores fueron obtenidos de dividir las relaciones de enemigos entre las relaciones de peores enemigos. Los cálculos para las secundarias, las preparatorias y la universidad incluyen los missing links. La línea discontinua que pasa por 3, indica la referencia de la proporción de Dunbar. 52

4.2. **Red signada de la escuela primaria epIZ.** Esta red tiene 417 estudiantes y 1995 relaciones, de los cuales el 29.6 % son de enemistad. Cada estudiante tiene en promedio 6.7 y 2.8 relaciones de amistad y enemistad, respectivamente. 53

4.3. **Red signada de la secundaria esRDC.** Esta red tiene 473 estudiantes y 784 relaciones, de los cuales el 25.6 % son de enemistad. Cada estudiante tiene 2.4 y 0.8 relaciones de amistad y enemistad, respectivamente. 54

4.4. **Distribución de comunidades.** En las imágenes de la a) a I) se muestran las frecuencias de comunidades calculadas en cada escuela. La barras grises son para el método de Newman, y las líneas discotinuas son para el greedy. Cada valor del método greedy fue obtenido de 300 simulaciones. 56

4.5.	Comunidades signadas. A) 4 comunidades de la primaria epRRC (método Newman). Se observan comunidades con nodos negativos y subgrupos antagónicos. B) 4 comunidades de la escuela esTN2 (método greedy). Los mismos fenómenos se observan pero en menor medida.	57
4.6.	Valores de la modularidad. (A) Valores de la modularidad (Q^+) para el método Newman (círculos) y para el método greedy (cuadros). (B) valores de la modularidad signada (Q^s) para los métodos de Newman (círculos) y greedy (cuadros). (C) Porcentaje de Perturbación para los enlaces negativos en las Comunidades (PPC) para los métodos Newman (círculos) y greedy (cuadros). Cada valor del greedy fue obtenido de 300 simulaciones.	58
4.7.	BSC y frustración. (A) Valores de la frustración para los métodos de Newman (círculos) y greedy (cuadros). (B) Valores del BSC (Balance Social Comunitario) para los métodos de Newman (círculos) y greedy (cuadros). Cada valor del greedy fue obtenido de 300 simulaciones. . . .	59
A.1.	Capturas de la plataforma digital diseñada para aplicar las encuestas a los estudiantes de la universidad. Una plataforma digital similar fue diseñada para las secundarias, y preparatorias.	74

Índice de tablas

3.1. Nombres y nomenclaturas de las escuelas participantes en esta tesis.	42
3.2. Características de las escuelas participantes en esta tesis. . . .	44
3.3. Valores de las relaciones de amistad de las redes escolares. Los datos empíricos fueron obtenidos de las encuestas. La proporción de Dunbar se muestra en la última columna, tal valor es obtenido de la razón del número de relaciones de amigos entre el número de relaciones de mejores amigos.	45
3.4. Valores de las relaciones de enemistad de las redes escolares. Los datos empíricos fueron obtenidos de las encuestas. La proporción de Dunbar se muestra en la última columna, este valor es obtenido dividiendo el número de relaciones de enemigos entre el número de relaciones de peores enemigos.	46
3.5. Enlaces positivos y negativos de las redes signadas.	49
4.1. Valores numéricos al incluir los missing links. la tercera columna indica los porcentajes de enlaces negativos añadidos. La sexta indica los porcentajes de la proporción de Dunbar para las relaciones negativas. .	52
4.2. Enlaces positivos y negativos en las comunidades.	54

Lista de abreviaturas

PPC	Porcentaje de Perturbacion de enlaces negativos en las Comunidades
BSC	Balance Social Comunitario
F	Frustración
epSC	escuela primaria “Silvestre Chi”
epRRC	escuela primaria “Rafael Ramírez Castañeda”
epIZ	escuela primaria “Ignacio Zaragoza”
esRDC	escuela secundaria “Raquel Dzib Cicero”
esTN2	escuela secundaria “Técnica Número 2”
esJLBG	escuela secundaria “José León Bojorquez García”
eprCCP	escuela preparatoria “Carlos Castillo Peraza”
eprHUN	escuela preparatoria “Hunucmá-UADY”
unTRS	universidad “Tecnológica Regional del Sur”
R	Resolución
LDU	matriz triangular inferior L, matriz diagonal D y matriz triangular superior U

Resumen

Las redes signadas proporcionan información para estudiar la estructura y composición de las relaciones entre individuos que componen un sistema complejo. Los individuos a través de diferentes criterios, forman grupos u organizaciones denominadas comunidades. Las estructuras de comunidades son una de las propiedades importantes de las redes sociales. En este escrito, analizamos la perturbación causada por las relaciones negativas en las comunidades. Desarrollamos una metodología para obtener y analizar particiones óptimas de comunidades en nueve redes escolares del Estado de Yucatán, México. Implementamos una técnica basada en la teoría del equilibrio social para predecir enlaces negativos no obtenidos, y además aplicamos dos métodos de detección de comunidades: los algoritmos de Newman y Louvain. Obtuvimos valores cercanos a la proporción de Dunbar para ambos tipos de relaciones, positivas y negativas. Se analizaron los conceptos de balance y frustración, y se utilizó la modularidad para cuantificar la perturbación causada por las relaciones negativas en las comunidades. Observamos diferencias entre las comunidades de los distintos niveles educativos. Las comunidades en el nivel primaria son inestables, es decir, significativamente perturbadas por las relaciones negativas. En el nivel de secundaria, las comunidades son semiestables, y en el nivel de preparatoria y universidad las comunidades son estables. El análisis indica que un número mayor de enlaces negativos no necesariamente implica mayor inestabilidad en las comunidades, sino que también intervienen otros factores sociales.

Abstract

Signed networks provide information to study the structure and composition of relationships (positive and negative) among individuals in a complex system. Individuals, through different criteria, form groups or organizations called communities. Community structures are one of the important properties of social networks. In this work, we aim to analyze the perturbation produced by negative relationships in communities. We developed a methodology to obtain and analyze the optimal community partitions in nine school networks in the state of Yucatán, México. We implemented a technique based on the social balance theory in signed networks to complete negative missing links and further applied two methods of community detection: Newman's and Louvain's algorithms. We obtain values close to Dunbar's ratio for both types of relationships, positive and negative. The concepts of balance and frustration were analyzed, and modularity was used to measure the perturbation of negative relationships in communities. We observe differences among communities of different academic degrees. Elementary school communities are unstable, i.e. significantly perturbed by negative relationships, in secondary school communities are semi-stable, and in high school and the university the communities are stable. The analyzes indicate that a greater number of negative links in the networks does not necessarily imply higher instability in the communities, but other social factors are also involved.

Introducción y motivación

El estudio de las interacciones de los componentes de un sistema complejo es un enfoque importante para entender su comportamiento [1,2]. Estas interacciones se pueden abordar desde la perspectiva de las redes [3]. Las redes son herramientas matemáticas que permiten estudiar de manera unificada la dinámica de sistemas complejos de distinta naturaleza, y con diferentes topologías de interacción las cuales dan lugar a fenómenos emergentes. Puntualmente, podemos mencionar algunos ejemplos de redes de sistemas complejos que se han estudiado: La red de internet [4], redes de comunicaciones [5], redes económicas [4, 6], redes sociales [4, 7], redes neuronales [8], red de proteínas [9], etcétera. Una característica interesante, es que a pesar de la diversidad y gran cantidad de sistemas complejos que pueden ser representados por redes, éstos comparten algunas propiedades comunes que poseen y las caracterizan, tales como su distribución de grado o su estructura comunitaria. Propiedades que se definirán más adelante.

Es cierto que el concepto de redes, es muy intuitivo en la sociedad moderna. Al escuchar la palabra red, inmediatamente pensamos en un grupo de elementos o partes conectados. Los inicios del estudio de las redes se remontan en 1736, cuando el matemático y físico suizo Leonard Euler (1707-1783) publicó la solución al problema de los puentes de la ciudad Königsberg (ver Figura 1), este problema consistía en dar un paseo de ida y vuelta atravesando los puentes de Königsberg sólo una vez [10, 11]. Esto marcó el inicio de la teoría de redes, y su estudio ha estado creciendo constantemente en diversas áreas de la ciencia. En las últimas décadas, como resultado de las nuevas aplicaciones e investigaciones sobre redes, se ha descubierto que los sistemas en la naturaleza no siguen una estructura totalmente regular o totalmente aleatoria, sino una combinación de ambas, esto permitió aseverar que los sistemas reales son bien representados por redes complejas, es decir, se le atribuyen redes con topologías no triviales que evolucionan con el tiempo. En otras palabras, una red compleja es una representación tangible de un sistema complejo [3, 12], aún cuando una red puede representar diferentes sistemas podemos generalizar el concepto de red de una manera abstracta:

Una red (o grafo) es una representación esquemática de un sistema. Consiste en un conjunto de nodos(vértices) que representan los elementos del sistema. Pares de nodos están unidos por enlaces (o vínculos), que representan en particular el tipo de relación

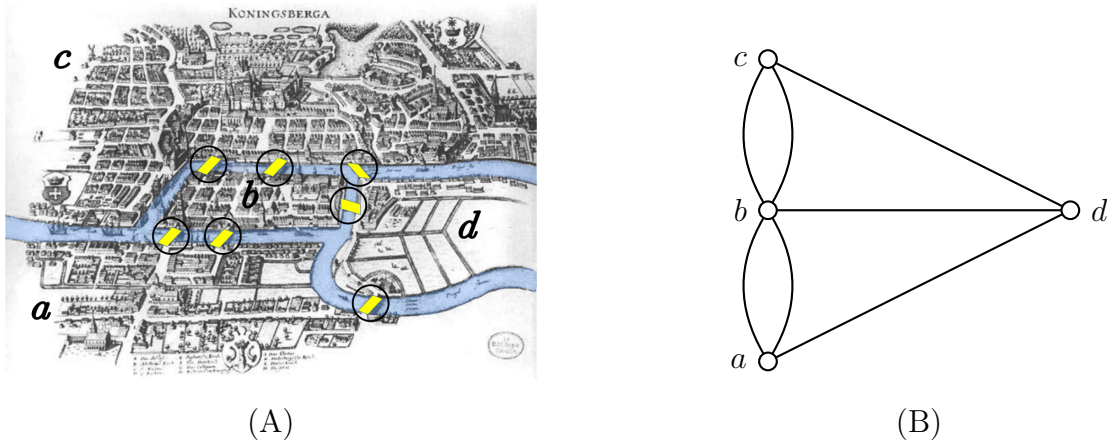


Figura 1: En (A) se indican los 7 puentes de la ciudad de Königsberg. Cada región es representada por las letras a,b,c,d. En (B) es la representación de la ciudad en una red, donde cada punto representa una región diferente y cada puente es representado por enlace que las une. Imagen de dominio público

entre sus elementos.

Recientemente, se han descubierto propiedades importantes de las redes complejas. Destacan entre ellos, los trabajos realizados por Watts y Strogatz en 1998, en ellos establecen una propiedad llamada efecto de “mundo pequeño”, este efecto consiste en que la distancia geodésica promedio (el camino más corto) entre los nodos crece lentamente en función con el número de nodos [13]. Esta propiedad se encuentra en muchos sistemas que nos rodean, incluidos la internet, las redes sociales, entre muchas otras. Albert Barabási [4] ha centrado sus estudios de redes desde de la perspectiva de la mecánica estadística, introdujo el concepto de red libre de escala, que es una propiedad donde la distribución de grado de los nodos $p(k)$ se aproxima bien a:

$$p(k) \sim k^{-\gamma} \quad (1)$$

, La Ec. 1 se le llama ley de potencia, y γ es el exponente de la potencia. Muchos sistemas reales siguen esta topología.

Las aplicaciones y avances de la teoría de redes, se han extendido en varios campos de estudio. En particular, en la sociometría, rama de la sociología donde las relaciones interpersonales son modeladas y estudiadas mediante el análisis de las redes sociales. Los inicios del análisis de las redes sociales se remontan a los principios del siglo XX, y se enfocan en la importancia de las relaciones entre entidades sociales, como lo son las relaciones interpersonales de un grupo de personas (en la escuela, en el trabajo, en clubes sociales, etc), las relaciones de comercio internacional, la red de telefonía móvil, y actualmente grupos sociales en línea (facebook, twitter, instagram, etc.).

Tal análisis de este tipo de relaciones permiten explicar los mecanismos sociales

implícitos que rigen la formación de comunidades sociales, o simplemente comunidades. A pesar de que no existe una definición precisa de comunidad, podemos definirla de una forma general como un conjunto de nodos que internamente está densamente conectados y escasamente conectados con nodos de otros conjuntos. En el contexto de las redes, a la división de la red en comunidades no superpuestas, se le conoce como su estructura comunitaria.

La estructura comunitaria es otra propiedad importante de las redes complejas. Un postulado del análisis de redes es que los elementos de los sistemas complejos crean una estructura comunitaria natural. Por ejemplo, en el contexto social, las comunidades se conforman de vínculos personales (familiares, amigos, colegas, etc.), pasatiempos, gustos u opiniones compartidas, mientras que las comunidades en redes ecológicas describen interacciones entre especies, como pueden ser una red alimentaria [14]. En consecuencia, encontrar comunidades, es útil para obtener una visualización macroscópica y simplificación de la red. Comúnmente las comunidades son definidas en términos de la partición del conjunto de nodos. La mayoría de los métodos de detección de comunidades encuentran este tipo de estructura. Aún cuando, la representación comunitaria obtenida por los métodos de detección podrían ser diferentes, éstas deberían mostrar fenómenos y comportamientos similares del sistema real. Determinar qué representación es mejor, no es tarea fácil e implica un análisis profundo. Existen trabajos que abordan esta cuestión, sin embargo, esta aun discutida [15].

Desde sus inicios, el análisis de redes se ha enfocado casi en su totalidad en redes no signadas (esto es redes con sólo enlaces positivos), no obstante, existen muchos sistemas reales que exhiben dos tipos de enlaces: uno positivo y uno negativo. Entonces, para representar un sistema que mezcla ambos tipos de enlace, se utilizan las redes signadas. Las redes signadas son una extensión de las redes, incluyen información adicional a los enlaces: positiva o negativa. Usualmente, muchas de las redes sociales están compuestas de ambos tipos de enlaces, desde luego, hay otras clases de redes que se componen también de estos tipos de enlaces, por ejemplo las redes ecológicas o de proteínas, entre muchas otras.

El análisis de redes sociales de datos empíricos sobre relaciones negativas son raros pero recientemente han atraído atención, con temas como las relaciones perniciosas en el centro de trabajo [16, 17]. Fuera del contexto laboral, algunas investigaciones estudian las relaciones negativas de redes de estudiantes donde coexisten enlaces positivos y negativos, que pueden ser nuestros amigos y enemigos (estudiantes hostiles/agresivos), respectivamente. La rareza de los estudios de redes sobre relaciones negativas, se puede explicar debido a que este tipo de relación es difícil de obtener, y por lo tanto hay menos información para modelar la estructura de la red.

La presencia de enlaces negativos en las redes signadas desafía conceptos y propiedades de redes no signadas. Por ejemplo, las comunidades en estas redes son diferentes a las de redes no signadas, y se conocen como “comunidades signadas”. Detectar co-

comunidades en estas redes es una tarea compleja. Hay muchas investigaciones que han propuesto métodos para estudiarlos todos con diferente tasa de éxito [18, 19]. Los más populares son los basados en la optimización de la función de calidad llamada Modularidad (propuesta por Newman [20]). La extensión de la modularidad a las redes signadas fue el primer paso para el análisis comunitario en esta clase de redes. Esto implica que la definición de comunidades signadas es una extensión de la definición de comunidades, al agregar la menor cantidad posible de enlaces negativos internos, y que estas sean mutuamente antagónicas (es decir, los enlaces entre las comunidades sean en su mayoría negativos). La información proporcionada por las redes signadas, y que no es percibida por las redes no signadas, puede ser muy importante para comprender la profundidad de sus interacciones y su evolución a un estado de estabilidad social o de mínima tensión. Lo expresado con anterioridad, ha llevado al desarrollo de métricas para medir que tan estable es una red [21, 22]. Explícitamente está ligado a los conceptos de balance social y frustración. El concepto de balance social, se originó en 1940 por varios estudios desarrollados por Heider [23] en la psicología social, y generalizado al lenguaje de la Teoría de redes por Cartwright y Harary [24]. El principio básico de la teoría afirma que las tríadas (tres nodos conectados) en las que las relaciones amigo-enemigo probablemente converjan a dos estados balanceados: “el amigo de mi amigo es mi amigo” y el “enemigo de mi enemigo es mi amigo”, en caso contrario habría tensión entre los integrantes de la tríada [25]. Esta teoría ha sido de mucha importancia e interés para abordar diversos temas, tales como el intercambio de opiniones, influencia social, balance social en redes signadas y balance social en comunidades signadas. Mientras que en estudios de comunidades signadas, la frustración es una medida que indica cuán lejos está una partición de la definición de comunidad. Si las comunidades tienen muchos enlaces negativos internos o no son antagónicas, la frustración será de alto valor. Un valor de frustración cero indica total estabilidad social dentro de las comunidades (todas los enlaces internos son positivos) y son totalmente antagónicas. Intuitivamente, la definición de la frustración muestra en qué grado los enlaces negativos mantienen alejada a la red de una total estabilidad, incluso puede ser usada en otros contextos, por ejemplo en la optimización de comunidades.

De particular interés resulta estudiar las propiedades de las comunidades en redes signadas, debido a que hay pocas investigaciones que han estudiado empíricamente el rol que juegan las relaciones negativas en la formación de las comunidades, particularmente en el ámbito escolar. Por todo esto, este proyecto está enfocado en realizar un análisis estadístico y cuantitativo de comunidades signadas en redes escolares, para encontrar sus particiones óptimas y cuantificar la perturbación de las relaciones de enemistad entre los estudiantes y así contribuir a una determinación más fiable a las comunidades del sistema real.

Durante las investigaciones anteriores que han dado lugar a este trabajo, emergieron una serie de preguntas que sirvieron de motivación para continuar el desarrollo e investigación en este campo. Entre todas, destaca los trabajos de Mankirat and Sarb-

jeet [26] que analizaron las relaciones negativas y demostraron que en un entorno social, aplicar conceptos estándares de las redes sociales a sólo los enlaces negativos podrían llevar a conclusiones erróneas. Entonces, los autores concluyeron que para entender las relaciones negativas se deben considerar las relaciones positivas. Inmediatamente, nos surgió la pregunta: ¿Tiene sentido analizar los enlaces positivos sin tomar en cuenta a los enlaces negativos? Hay muchos estudios de comunidades que sólo consideran enlaces positivos [27, 28]; sin embargo, al no considerar la perturbación causada por los enlaces negativos la estabilidad de las comunidades es afectada, dando lugar a posibles comunidades con un número significativo de enlaces negativos en ellas. Tales comunidades se alejan de lo que entiende tradicionalmente por comunidad. Por lo que los resultados, y conclusiones derivadas de comunidades de sólo enlaces positivos podrían no ser del todo correctas, debido a que no se acercan a las de un sistema real.

Creemos que las comunidades que no son perturbadas por las relaciones negativas coinciden mejor con las de un sistema real. Paralelamente, ante la falta de conocimiento sobre el impacto que tienen los enlaces negativos en la estructura comunitaria, surgen otras preguntas: ¿Cómo las particiones comunitarias de nuestras redes reflejan las particiones reales?, ¿El número de enlaces negativos implica mayor inestabilidad en las comunidades? y ¿Qué otros factores influyen en la estabilidad de las comunidades? Para responder a estas preguntas, nuestra metodología consistió en una combinación del análisis clásico de comunidades en redes no signadas con la teoría de redes signadas. La técnica utilizada para la recolección de datos fueron encuestas aplicadas a los estudiantes de 9 colegios. Analizaremos las particiones de comunidades extraídas de dos algoritmos diferentes muy aceptados y utilizados en estudios de estructuras comunitarias: El algoritmo de Newman y el Louvain. Para el análisis cuantitativo, usaremos tres medidas: modularidad, balance social y frustración. La modularidad fue utilizada como parámetro que indica la perturbación causada por las relaciones negativas en las comunidades.

De igual forma, considerando que actualmente los estudios sobre interacciones sociales entre escuelas ha tomado gran importancia. Por ejemplo, en la investigación [29], los autores estudian la evolución temporal de la red de contactos cara a cara utilizando sensores de proximidad. Las redes de amistad también son usadas para estudiar fenómenos como liderazgo, la competencia y desempeño académico [30]. En referencia a esto, la metodología propuesta en este trabajo genera comunidades signadas con ciertas propiedades. De este modo, dicha metodología procura revelar estructuras comunitarias e identificar la relación entre la estabilidad de las comunidades y la percepción que tienen los estudiantes de diferentes niveles educativos sobre los conceptos de amistad y enemistad.

En el curso del desarrollo de este trabajo se ha avanzado en las respuestas a las preguntas anteriores. Sin embargo, han surgido nuevas interrogantes que plantean líneas de trabajo a futuro. El objetivo final de esta investigación es, una nueva metodología para

analizar y cuantificar la perturbación causada por las relaciones negativas en particiones óptimas de comunidades.

La investigación contenida en esta tesis fue revisada por pares y publicada en la revista de Plos One [31]. El trabajo de esta tesis se organiza en los siguientes capítulos:

En el primer capítulo, se enuncian los principales conceptos y propiedades de la teoría de redes. Asimismo, en este mismo capítulo se explica la definición de comunidad, además se describen los métodos que utilizamos para detectar comunidades y se introduce el concepto de modularidad y su relación con la mecánica estadística. En el segundo capítulo, se hace una corta descripción de la teoría de redes signadas, agregando resultados importantes para analizar tales redes que nos permitirán el desarrollo de este trabajo. En el tercer capítulo se expone la metodología desarrollada para extraer y analizar particiones óptimas de comunidades. En el capítulo 4 se presentan los resultados del análisis de las redes estudiadas. En seguida, en el capítulo 5 se presenta una discusión general a partir de los hallazgos encontrados. Por último, en el capítulo 6 se muestran las conclusiones y algunas sugerencias de líneas de investigación futuras.

Capítulo 1

Teoría de redes

La teoría de redes es el marco de estudio de las redes complejas. Naturalmente, una red compleja puede ser representada por una red o grafo. Formalmente, se define de la siguiente manera:

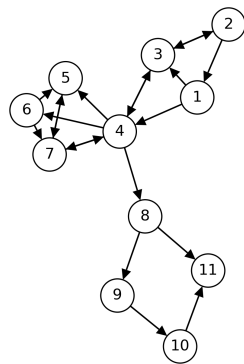
Definición. *Una red o grafo, es un par de conjuntos $G = \{V, L\}$, donde $V = \{v_1, v_2, \dots, v_{N-1}, v_N\}$ es el conjunto de N nodos y L es el conjunto de enlaces que conectan dos elementos de V .*

Es preciso mencionar que la dirección de un enlace entre dos nodos puede ser importante y representar una dinámica diferente cuando se invierte el sentido. En este aspecto, pueden ser no dirigidas o dirigidas. Una red es no dirigida o simple, si todos sus enlaces son recíprocos, i.e si los nodos $i \rightarrow j$ están enlazados, también $j \rightarrow i$ lo estarán. En el caso de las redes dirigidas los enlaces son asimétricos, esto es, al menos existe un par de nodos conectados en los que una dirección del enlace no es correspondido, i.e si los nodos $i \rightarrow j$, no necesariamente $j \rightarrow i$. En la Figura 1.1, podemos observar estas diferencias en la dirección de los enlaces.

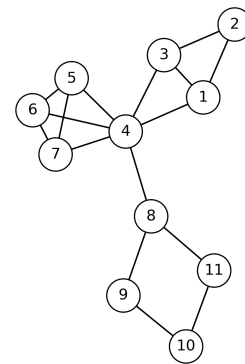
Cabe destacar que no considerar las direcciones de los enlaces ayuda a simplificar los cálculos, además las herramientas para analizarlas están mejor desarrolladas y entendidas para redes no dirigidas que en redes dirigidas.

1.1. Representación de una red: Matriz de adyacencia

Además de la visualización gráfica de una red, tal como muestra la Figura 1.1. Cualquier red también puede ser representada de forma algebraica mediante una matriz



(A) Red dirigida



(B) Red no dirigida (simple)

Figura 1.1: Ejemplo de una red con 11 nodos. En (A), la red es dirigida con 20 enlaces, mientras que en (B) el caso no dirigido tiene 16 enlaces.

$W \in M_{N \times N}(\mathbb{R})$ que indica si existe relación o no entre cada par de nodos, donde w_{ij} es el peso del enlace. A esta matriz de representación se le conoce como matriz de adyacencia. Por ejemplo, en las redes de amistad

$$w_{ij} = \begin{cases} 1 & \text{si son amigos} \\ 0 & \text{en otro caso} \end{cases}$$

En caso, de que la red sea no dirigida la matriz W resultaría simétrica, en otro caso sería asimétrica. Ahora, si la red tiene autoenlaces la $diag(W) \neq 0$, en caso contrario la diagonal sería de ceros. Por ejemplo, las matrices de adyacencia de la Figura 1.1, se muestran en la Figura 1.2. Notar que en las mismas matrices de la Figura 1.2 se observan las diferencias descritas previamente: ambas tienen la diagonal de ceros porque no hay

Nodos	1	2	3	4	5	6	7	8	9	10	11
1	0	0	1	1	0	0	0	0	0	0	0
2	1	0	1	0	0	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0	0
4	0	0	1	0	1	1	1	1	0	0	0
5	0	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	1	0	1	0	0	0	0
7	0	0	0	1	1	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	1	0	1
9	0	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	0	1
11	0	0	0	0	0	0	0	0	0	0	0

(A) Red dirigida

	1	2	3	4	5	6	7	8	9	10	11
1	0	1	1	1	0	0	0	0	0	0	0
2	1	0	1	0	0	0	0	0	0	0	0
3	1	1	0	1	0	0	0	0	0	0	0
4	1	0	1	0	1	1	1	1	0	0	0
5	0	0	0	1	0	1	1	0	0	0	0
6	0	0	0	1	1	0	1	0	0	0	0
7	0	0	0	1	1	1	0	0	0	0	0
8	0	0	0	1	0	0	0	0	1	0	1
9	0	0	0	0	0	0	0	1	0	1	0
10	0	0	0	0	0	0	0	0	1	0	1
11	0	0	0	0	0	0	0	1	0	1	0

(B) Red no dirigida

Figura 1.2: Matrices de adyacencia para las redes de la Figura 1.1. En (A) es el caso dirigido, y en (B) el caso no dirigido.

autoenlaces, mientras que la matriz de la 1.2 (A) es asimétrica y la matriz de la 1.2 (B) es simétrica.

1.2. Propiedades de las redes

La teoría de redes complejas ha definido diversas propiedades para analizar la estructura de cada red. Sin embargo, sólo mencionaremos las más importantes para esta tesis: el grado medio, distribución de grado, la densidad, los caminos, la intermediación de enlaces y las comunidades. Estas descripciones serán para redes no dirigidas.

1.2.1. Grado medio

Una propiedad individual de un nodo, es su grado. El grado k_i del nodo i , se define como el número de enlaces conectados con i . El promedio de los grados de los nodos, se le llama grado medio y se denota por $\langle k \rangle$. Matemáticamente, $\langle k \rangle$ está dada por:

$$\langle k \rangle = \frac{1}{N} \sum_i^N k_i \quad (1.1)$$

1.2.2. Distribución de grado

La distribución de grado de una red, se especifica por la función de distribución $p(k)$, que es la probabilidad de que un nodo seleccionado al azar tenga grado k , esto es:

$$p(k) = \frac{n(k)}{N} \quad (1.2)$$

donde $n(k)$ es el número de nodos con grado k . Así, en función de $p(k)$, el grado medio se puede escribir como:

$$\langle k \rangle = \sum_{k=0}^{\infty} kp(k) \quad (1.3)$$

1.2.3. Densidad

La densidad, D , es la propiedad de la red que cuantifica la proporción del número L de enlaces existentes en ella entre el número de enlaces posibles. El número de enlaces posi-

bles de un red de N nodos es determinado por el coeficiente binomial $C_N^2 = \frac{N(N-1)}{2}$, dando lugar a la siguiente ec. 1.4:

$$D = \frac{2L}{N(N-1)} \quad (1.4)$$

La densidad de la red es cero cuando no existen enlaces entre los nodos y uno cuando todos están relacionado entre sí.

1.2.4. Caminos.

En redes, un camino es una ruta que recorre los enlaces de la red. La longitud del camino, es el número de enlaces que contiene el camino. La geodésica entre dos nodos es la longitud del camino más corto que los conecta. En otras palabras, la distancia es un concepto que indica los saltos mínimos que se deben hacer para llegar de un nodo a otro, y se denota por $d_{i,j}$. El camino más corto entre dos nodos puede no ser único. Ahora, al aplicarlo a cada uno de los nodos, el promedio de las distancias, se obtiene la distancia geodésica de la red ($\langle d \rangle$). Para aclarar este concepto se calcula la distancia de algunos nodos de la Figura 1.3. En ésta se puede observar que la distancia del nodo 1 al 5, es $d_{15} = 3$, y también para ir del nodo 1 al 5 hay dos caminos más cortos con la misma longitud. De la misma forma, $d_{1,4} = 2$. Si se promedian todas las distancias $d_{i,j}$ de los nodos, la distancia geodésica de la red es 1.6.

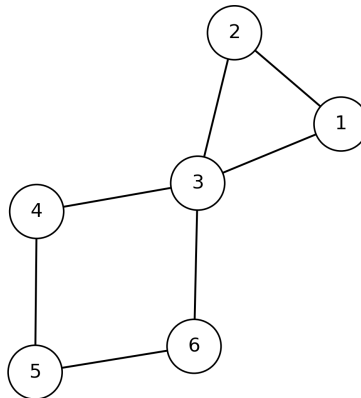


Figura 1.3: Caminos. Red de 6 nodos y 7 enlaces. Un secuencia de nodos conectados uno con otro por un enlace, es un camino. La longitud del camino, es el número de enlaces que la componen. En este ejemplo, todos los nodos de la red están conectados por un camino, a esto se le conoce como una red conectada.

Hay que mencionar, que en el caso de redes no dirigidas, $d_{i,j} = d_{j,i}$, es decir, la distancia entre los nodos i y j es igual a la distancia entre j e i . Sin embargo, para las redes dirigidas, en general, $d_{i,j} \neq d_{j,i}$, inclusive puede no existir un camino entre los nodos i

y j . Por ejemplo, en la Figura 1.1 (A) $d_{9,10} = 1$, sin embargo no existe un camino de regreso.

1.2.5. Intermediación

Esta medida de centralidad de la red, caracteriza la importancia de los enlaces para las propiedades de difusión, y también para identificar la estructura comunitaria de la red. En otras palabras, un enlace con un valor de intermediación de enlace alto, representa un conector similar a un puente entre dos partes de la red, y la eliminación de este enlace puede afectar la comunicación entre los nodos que se conectan a través de los caminos más cortos que están a lo largo de este enlace.

A continuación definiremos esta medida. Dada una red de N nodos conectados por L enlaces. Sea h el enlace que conecta cualesquiera pares de nodos ordenados i y j . La intermediación del enlace h , está dado por:

$$C_I(h) = \frac{2}{N(N-1)} \sum_{a>b} \frac{\sigma_{ab}(h)}{\sigma_{ab}} \quad (1.5)$$

donde σ_{ab} es número de caminos más cortos que conectan los pares de nodos a y b , y $\sigma_{ab}(h)$ es el número de todos esos caminos que pasan por h . En la Figura 1.4 se muestra un ejemplo de una red de 6 nodos, y el enlace entre los nodos 3- 4 tiene el valor de intermediación de enlaces alto de 0.6. La eliminación de este enlace dará como resultado dos subredes densamente conectadas.

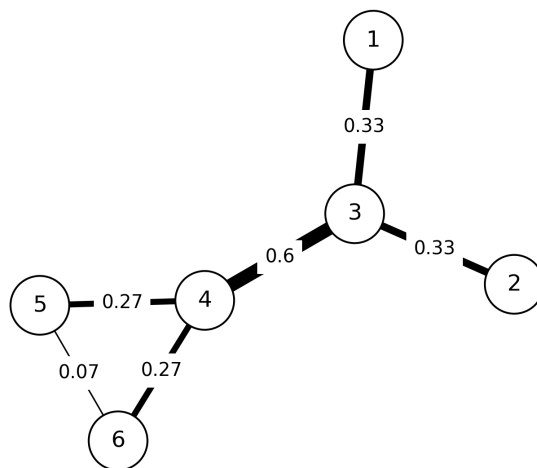


Figura 1.4: Intermediación de enlaces. En la red, el enlace entre los nodos 3 y 4, es el que tiene el valor de intermediación más alto de todos los enlaces.

1.3. Comunidades

Como se ha mencionado previamente el concepto de comunidad es una propiedad importante de las redes. El termino de “comunidad” intuitivamente quiere decir una subred cuyos nodos están densamente conectados pero escasamente conectados con otros nodos fuera de ella [32]. La importancia de esta idea radica en que los nodos de una misma comunidad compartan ideas, aficiones, opiniones, cualidades, características comunes, etc [33]. No obstante, no existe una definición exacta, esto ocasiona que una red pueda ser dividida de distintas maneras, a lo que se le llama partición. La Figura 1.5 muestra una partición de una red en comunidades.

1.3.1. Partición

Una partición es la división de una red en grupo de nodos o comunidades, de modo que cada nodo pertenece a una comunidad. Por lo general, las particiones son representadas a través de un diagrama de árbol o dendograma. Determinar una partición de comunidades en una red es una tarea compleja, debido a que existen muchas maneras de dividir la red, y no todas exhiben una estructura comunitaria real. Por ejemplo, en las redes sociales, detectar comunidades además de brindarnos una visualización total de la red, las relaciones de sus comunidades aclaran el flujo de información, y permite predecir su crecimiento en un futuro [34]. En consecuencia, la detección de comunidades ha atraído interés de muchos investigadores, que han propuesto diversos métodos para dividir y evaluar que tan buena es esta partición. Los más populares, son los basados

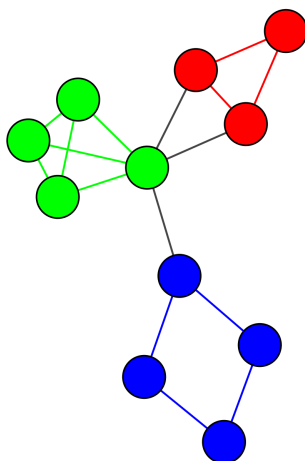


Figura 1.5: Comunidades de la red de la Figura 1.1 (B). Esta imagen muestra una estructura de tres comunidades, los nodos de cada comunidad están densamente conectados internamente y conexiones escasas entre cada comunidad.

en la optimización de la modularidad [20] y en la siguiente sección, explicaremos este concepto. Otros utilizan la matriz de adyacencia para extraer información por métodos espectrales [19,35]. A continuación algunos ejemplos de métodos desarrollados para extraer comunidades.

1.3.2. Introducción a la modularidad

La modularidad es una función de calidad que mide que tan óptima es una partición de comunidades. La modularidad cuantifica la fuerza de división de una red en módulos. Compara la fracción de enlaces positivos internos de las comunidades en la red observada y una red aleatoria la misma distribución de grado. La modularidad se define como:

$$Q^+ = \frac{1}{2w^+} \sum_i \sum_j \left(w_{ij}^+ - \frac{w_i^+ w_j^+}{2w^+} \right) \delta(C_i, C_j) \quad (1.6)$$

donde $\delta(C_i, C_j) = 1$ si los nodos i y j pertenecen a la misma comunidad y $\delta(C_i, C_j) = 0$ en otro caso. El símbolo w^+ representa el número de enlaces positivos. Generalmente, el valor de la modularidad se encuentra en el intervalo $[0, 1]$ aunque en algunos casos se encuentra Q^+ en $[-1,0)$. Un valor positivo de Q^+ indica que tan buena es una partición mientras que un valor negativo indica grupos de nodos con baja densidad de conexiones internas y muchas conexiones con el resto de la red. Un valor cercano a 1 indica comunidades bien definidas. Por otra parte, la modularidad es cuestionable, debido a que en 2007 los trabajos realizados por Fortunato y Bathélemy [36] demostraron que la optimización de la modularidad puede fallar al detectar comunidades más pequeñas que una escala determinada por su número de enlaces y el grado de interconexión de sus módulos, esto es, este límite de resolución va depender del tamaño de la red, sin embargo, esto no fue un problema para nuestro estudio, debido a que las redes estudiadas son relativamente pequeñas.

1.3.2.1. Modularidad: Una perspectiva con la Física

Las redes recientemente han asumido un papel práctico como una herramienta para el estudio de sistemas complejos, donde las redes ofrecen una representación útil para las interacciones de sus componentes abstraídos por enlaces y nodos. El análisis riguroso y el modelado de redes, han tenido un gran éxito en la comprensión de los sistemas complejos. En efecto, el objetivo principal de las investigaciones en redes es descubrir propiedades emergentes no visibles de los nodos a través del análisis de su estructura. Una peculiaridad de los sistemas complejos es que exhiben estructuras a diferentes escalas. Por lo tanto, las redes complejas combinan características estructurales microscópicas resultantes de las propiedades individuales de los nodos con las características mesoscópicas debido a las propiedades de comunidades [37]. Estudiar las dos

escalas por separado es esencial porque las características y propiedades de los nodos pueden ser diferentes a las propiedades de sus comunidades. En la escala mesoscópica, las comunidades son las unidades elementales de estudio y no los nodos. De este modo, el término mesoscópico se utiliza, debido a que su escala relevante se encuentra en un nivel intermedio entre los nodos y la red total.

En los últimos años, el interés de los físicos por las redes ha crecido. En especial, la física estadística ha desempeñado un papel clave para comprender las redes complejas [38]. El enfoque adoptado por la física, es diferente a la de otras ciencias, como las matemáticas o las sociales, debido a que se basa en estudios empíricos de redes en nuestro alrededor, y se enfoca en las propiedades estadísticas, desde una escala microscópica o mesoscópica. Por ejemplo, el grado medio de las conexiones de los nodos, la distribución de grado de los nodos o las distribuciones de sus comunidades. Hay estudios previos, que han evidenciado que al observar una red compleja, el diagrama de sus conexiones no sigue un patrón regular, tampoco es aleatorio. Esto es, la red tiene una estructura intrínseca derivada del carácter estocástico de sus mecanismos de organización. Algunas redes pueden presentar formaciones tipo estrella (formaciones de nodos conectados a un nodo central), aumentando la probabilidad de que otros nodos se conecten a ellas, y tienden a formar nodos altamente conectados. Las preguntas en las que se interesa la física es en cuantificar estos patrones y revelar características importantes del sistema que representa la red, así como la interacción entre su estructura y funcionamiento. En esta tesis se presenta una descripción mesoscópica basados en modelos probabilísticos y estadísticos para la inferencia de propiedades y características latentes de los nodos, a nivel individual y en sus comunidades. A continuación, explicaremos algunos conceptos e ideas de la mecánica estadística usados en la detección de comunidades.

El problema de particionar una red es de interés para la física. Esto dio lugar a métodos basados en los principios de mecánica estadística o en modelos físicos. Muchos de estos métodos, se basan a partir del trabajo de Newman [20] desarrollando algoritmos con enfoques diferentes. Entre todos ellos, destaca el propuesto por Reichardt y Bornholdt basado en el Modelo de Potts [39]. Este modelo reformula la modularidad, y muestra como la detección de comunidades puede ser interpretada como la búsqueda del estado base de un cristal de espín de rango infinito.

Inicialmente, desde la conexión con la modularidad, el modelo de Reichardt y Bornholdt así como otros métodos basados en inferencia estadística, por ejemplo los Modelos de Bloques Estocásticos, pueden ofrecer una mayor información estadística, sin embargo, para este trabajo optamos un enfoque más descriptivo, de acuerdo a los objetivos de la tesis.

1.3.3. Algoritmos para la detección de comunidades

En la actualidad, existen muchos algoritmos que son capaces de maximizar la modularidad con tiempos de ejecución razonables y recursos computacionales bajos. Para esta tesis, elegimos los algoritmos de Newmann y Louvain (greedy). En ambos métodos el número y tamaño de las comunidades no se especifican, sino que son determinados por la misma organización de la red. El objetivo es crear una partición comunitaria con una modularidad máxima inicial, y dividirla en caso que exista una buena subdivisión, y dejar de dividir en caso contrario. A continuación describiremos de manera general ambos métodos.

1.3.3.1. Algoritmo de Newman

El método de Newman es un algoritmo determinístico. El procedimiento utiliza una matriz B (llamada matriz de modularidad). B es una matriz simétrica de $N \times N$, con una fila y columna para cada nodo. La matriz de modularidad B se define como:

$$B_{ij} = A_{ij} - k_i k_j / 2m \quad (1.7)$$

donde A_{ij} es el elemento de la matriz de adyacencia, k_i es el grado de cada nodo y m es el número de enlaces en la red. Luego, se diagonaliza B usando el método de descomposición LDU [40], para obtener sus eigenvalores y eigenvectores. Luego, para encontrar una modularidad inicial, se construye el vector de asignación s en función de los signos del eigenvector que corresponde al eigenvalor más grande. El valor $s_i = +1$, si la i -ésima entrada del eigenvector es positivo. Es $s_i = -1$ si la i -ésima componente del eigenvector es negativo o cero. Como resultado, la red se divide en dos comunidades. Si $s_i = +1$, el nodo i pertenece a la primera comunidad, mientras que para $s_i = -1$, el nodo pertenece a la segunda comunidad. Observamos que, la operación binaria $((s_i s_j + 1)/2)$ es 1, si i y j están en la misma comunidad, y 0 en otro caso. La modularidad de la Ec. 1.6, se puede reescribir, como:

$$\begin{aligned} Q^+ &= \frac{1}{4m} \sum_{ij} B_{ij} (s_i s_j + 1) \\ &= \frac{1}{4m} \sum_{ij} B_{ij} s_i s_j + \sum_{ij} B_{ij} \\ &= \frac{1}{4m} \sum_{ij} B_{ij} s_i s_j \end{aligned} \quad (1.8)$$

donde la última igualdad se sigue desde que $\sum_{ij} A_{ij} = \sum_i k_i = 2m$ y $\sum_{ij} B_{ij} = 0$. De forma más conveniente la Ec. 1.8 se puede escribir como un producto de matrices

$$Q^+ = \frac{1}{4m} \mathbf{s}^t \mathbf{B} \mathbf{s} \quad (1.9)$$

donde \mathbf{B} , es la matriz de modularidad (ec. 1.7). Esta última expresión es similar al Hamiltoniano del modelo de Ising [41], una conexión con la física que ha servido para crear algoritmos computacionales para maximizar la modularidad.

Lo anterior, describe un método basado en matrices para encontrar una buena división de una red en dos partes. Sin embargo, muchas redes contienen más de dos comunidades, por lo que se puede usar este método repetidamente para encontrar buenas divisiones de redes en mayor número de partes. Esto consiste en ir dividiendo cada comunidad en dos, a manera de árbol binario. En cada división, se calcula la diferencia de las modularidades nueva con respecto a la anterior (ΔQ^+). Si $\Delta Q^+ \leq 0$, si no hay mejora en la modularidad y la comunidad no se puede seguir dividiendo, está permanece intacta, en caso contrario, se dividiría iterativamente, hasta que ninguna comunidad pueda dividirse. Finalmente, se obtendría una partición óptima que maximiza Q^+ (para más detalles ver [19, 42]). En el Apéndice A.4 se muestra la implementación del algoritmo de Newman en redes signadas.

1.3.3.2. Algoritmo de Louvain

El algoritmo de Louvain, es uno de los métodos populares para maximizar la modularidad. Fue desarrollado por Blondel et al [43] de la Universidad de Louvaina (origen del nombre de este algoritmo). Este método es simple y elegante. El método es un algoritmo codicioso (greedy) capaz de extraer comunidades de grandes redes con tiempos de ejecución $O(n \cdot \log n)$, donde n es el número de nodos de la red. El método optimiza Q^+ dando como resultado la mejor agrupación posible de los nodos de una red determinada. Pero explorar todas las posibles agrupaciones es ineficiente, por lo que se utilizan métodos heurísticos, i.e un procedimiento para optimizar Q^+ mediante una aproximación iterativa e intuitiva para obtener una buena solución. Este método sigue esta lógica. De esta manera, el proceso de optimización de la modularidad de este algoritmo consta en dos fases elementales. Normalmente, la primera fase se centra en encontrar una partición local. Esta fase consiste en:

- Asignar a cada nodo a una comunidad, esto es inicialmente son N comunidades de un sólo nodo.
- Inspecciona y une a cada par de comunidades conectadas al menos por un enlace. Se calcula ΔQ a antes y después de la unión. Si hay un aumento de ΔQ , la comunidades se dejan unidas, de lo contrario regresan a su comunidad original. Este proceso se aplica iterativamente hasta alcanzar un máximo local de modularidad, la primera fase del algoritmo ha finalizado.

En la segunda fase, se agrupan los nodos de una misma comunidad formando un supernodo. Se construye una red pesada, con estos supernodos. Todos los enlaces entre

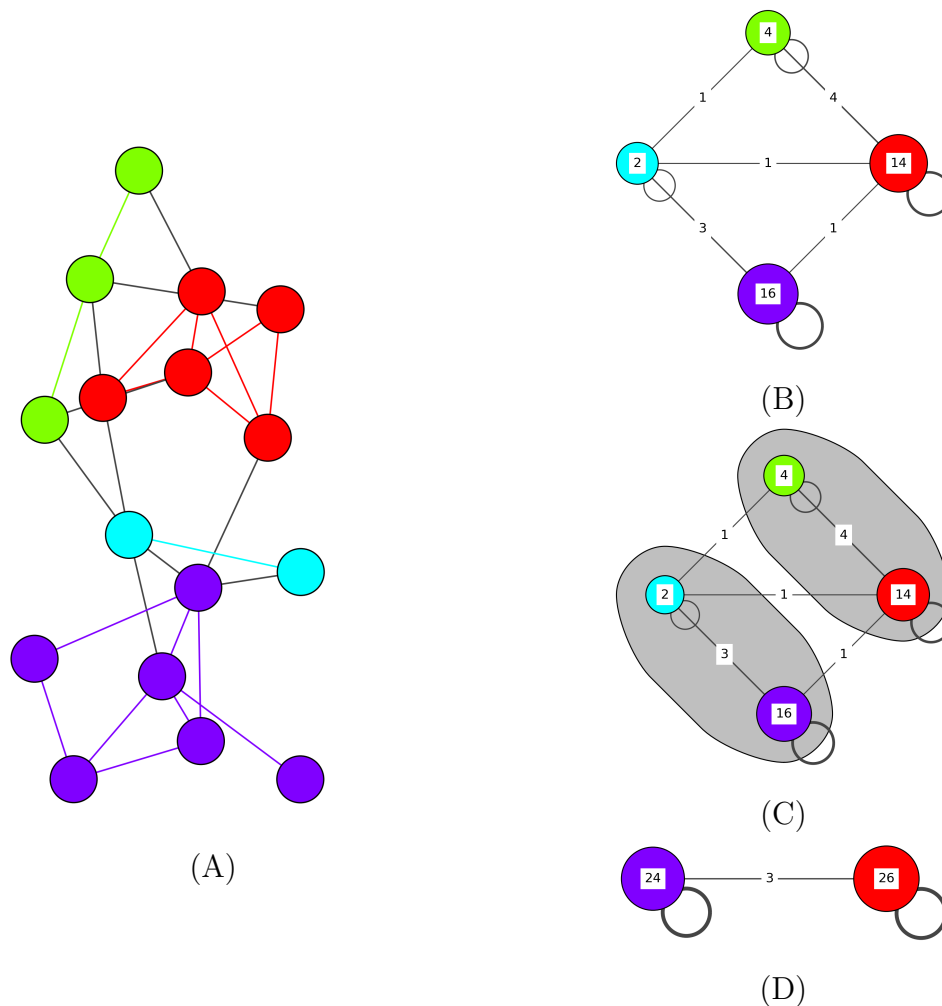


Figura 1.6: Algoritmo de Louvain. La red inicial contiene 16 nodos y 28 enlaces. El algoritmo parte desde una partición trivial donde cada nodo está en una comunidad. Inmediatamente, el algoritmo mueve nodos individuales de una comunidad a otra, hasta llegar a una partición local (A). Luego, a partir de esta partición, se construye una nueva red pesada (B), cuyos nodos corresponden a las comunidades encontradas en (A), se generan los autoenlaces que son el grado de los nodos en cada comunidad (2,4,14,16) y los enlaces ponderados son el número de enlaces entre las comunidades. A continuación, se vuelven agrupar los nodos al azar de la red pesada (C). Estos pasos se repiten hasta que la modularidad no pueda mejorar, y el algoritmo se detiene (D).

Los nodos de una misma comunidad se suman, y se genera el autoenlace del nuevo supernodo. Los enlaces ponderados que conectan los supernodos corresponden al número de enlaces de los pares de nodos conectados en comunidades diferentes. Una vez creada la nueva red, la segunda fase ha finalizado y es posible volver aplicar la primera fase (más detalles en [19, 43]). El esquema del algoritmo de Louvain se muestra en la Figura 1.6.

En principio, el algoritmo de Louvain parte de una partición local y agrupa nodos escogiéndolos aleatoriamente, sin embargo, es posible iniciar el método desde una partición diferente. Esta característica, hace que el algoritmo de Louvain sea estocástico. En consecuencia, el algoritmo genera secuencias de particiones óptimas a diferentes escalas. En particular, los resultados obtenidos por este algoritmo deben analizarse con valores promedio. Sin embargo, se espera que sólo un pequeño número de tales particiones sean significativas, esto nos lleva a plantearnos la siguiente pregunta ¿Cómo podemos elegir particiones significativas y equivalentes? Para abordar este problema, este método tiene un parámetro (llamado Resolución) que debe ser calibrado. En la sección 3.5 de resultados se explica como fue calibrado el algoritmo de Louvain para nuestras redes. En el Apéndice A.4 se muestra la implementación del algoritmo de Louvain en redes signadas utilizando el módulo Python-Louvain.

1.3.3.3. Algoritmo de Girvan-Newman

Este algoritmo desarrollado por Michelle Girvan and Mark Newman [44], es uno de los más populares en la detección de comunidades en redes (al igual que los métodos anteriores) que marcaron el comienzo de una nueva era en este campo, y dio lugar al interés de los físicos. Este algoritmo es un método jerárquico que se encarga de detectar comunidades eliminando progresivamente los enlaces de la red original. Las componentes desconectadas resultantes de este proceso son la comunidades.

El método de Girvan-Newman, de manera general, se enfoca en la medida de centralidad de intermediación (edge betweenness centrality), que es una medida que expresa la frecuencia en la que un enlace participa como puente entre par de nodos logrando la menor cantidad de saltos entre ellos.

Aquí, los enlaces se eligen según los valores de intermediación de los enlaces, estimando la importancia de los enlaces para las comunidades, y muestra aquellos enlaces que poseen una mayor “trafico” o nivel conexión entre los nodos de red que pertenecen a comunidades diferentes.

Los pasos generales del algoritmo son:

1. Se calcula las medidas de intermediación de los enlaces de la red.
2. Se elimina el enlace con mayor valor de intermediación. En caso de empate con otros enlaces, se elige uno de manera aleatoria.
3. Se recalcula de nuevo las medidas de intermediación en todos los enlaces.
4. Se repiten los pasos 2 y 3, hasta que que la red original se divide en al menos 2 subredes separadas.

Lo anterior, describe un método para dividir una red en al menos 2 partes. Sin embargo, comúnmente las redes contienen más comunidades, por lo que es posible usar este método para tener una buena división en varias partes. El proceso de división consiste en dividir cada subred hasta separarla en subredes aisladas, generando un árbol binario. Las subredes resultantes de la división anterior, se dividen recursivamente de la misma manera hasta que comprenden solamente un nodo (para más detalles ver [45]).

En redes con una estructura comunitaria, los enlaces que conectan diferentes comunidades tienen valores altos de intermediación, debido a que todas las rutas más cortas que conectan los nodos de diferentes comunidades respectivas cruzan esos enlaces. Por lo tanto, la eliminación de estos enlaces, los grupos se separan entre sí y la estructura comunitaria se revela.

Desde el punto de vista computacional, el tiempo de ejecución varía dependiendo del método usado para evaluar la intermediación [44]. Por lo general, este tiempo es alrededor de $O(mn^2)$, donde m son los enlaces y n los nodos [18, 44]. Así, el cálculo de la intermediación de cada enlace se vuelve excesivamente costoso para redes con alta cantidad de nodos y enlaces.

1.3.4. Club de Karate de Zachary

El club de Karate de Zachary es una red social formada por 34 miembros de un Club de Karate en EE.UU. Los datos fueron recopilados por Wayne Zachary, durante los años de 1970-1972. La red contiene 78 enlaces que representan las interacciones de los miembros fuera del club. Durante el estudio surgió una disputa entre el administrador y el instructor, causando la división en dos grupos. La mitad de los miembros formó un grupo en torno al instructor, y la otra mitad formó otro grupo con el administrador. Zachary asignó correctamente a los miembros en dos grupos (salvo a uno) después de la división (más detalles ver [46]). La red del Club de Karate de Zachary, es ampliamente utilizado como referencia para testear algoritmos de detección de comunidades. Muchos de estos algoritmos han encontrado con éxito la partición correcta de la red, que corresponde en las dos comunidades principales después de la división. Como ejemplo, en la Figura 1.7, se muestran las comunidades encontradas por los métodos de Girvan-Newman, Louvain y Newman. Cada método generó una partición óptima diferente. Usando el método de Girvan-Newman se encontró una partición de dos comunidades con $Q^+ = 0.381$ (Figura 1.7 (A)). Para el método de Louvain, la red se divide en 4 comunidades con una modularidad $Q^+ = 0.419$ (Figura 1.7 (B)). Por último, aplicando el método de Newman, se obtienen una partición de 5 comunidades con un valor menor de $Q^+ = 0.378$ (Figura 1.7 (C)). Observamos, que la partición obtenida por el Método de Girvan-Newman coincide casi perfectamente con la división después de la disputa entre sus miembros, excepto por un nodo (más detalles en [47]). Sin embargo, estudios

subsecuentes sobre esta red, como los realizados por Duch y Arenas [48], mostraron que si la condición de dos comunidades se relaja, y buscamos particiones con más de dos comunidades, entonces podríamos encontrar una mejor partición de la red con un valor alto de Q^+ , como la partición que muestra el método de Louvain. En particular, esta partición resulta de la división de las dos comunidades resultantes del método de Girvan-Newman. Hay estudios que indican que la partición en 4 comunidades de la Figura 1.7 (B), es una partición significativa para la división real del Club de Karate de Zachary. El método de Newman, es interesante, debido a que genera una partición de 5 comunidades (cercana a la del Louvain) con una modularidad similar al del método de Girvan-Newman. En la literatura, hay muchos métodos, como por ejemplo los algoritmos genéticos [49], que ofrecen otras particiones óptimas a la red de karate con modularidades mayores o similares. Lo anterior, nos indica que no sólo hay que tener en cuenta a las particiones que ofrecen la mejor modularidad, sino que también hay que considerar otras particiones que a pesar de tener una modularidad menor siguen siendo significativas [49].

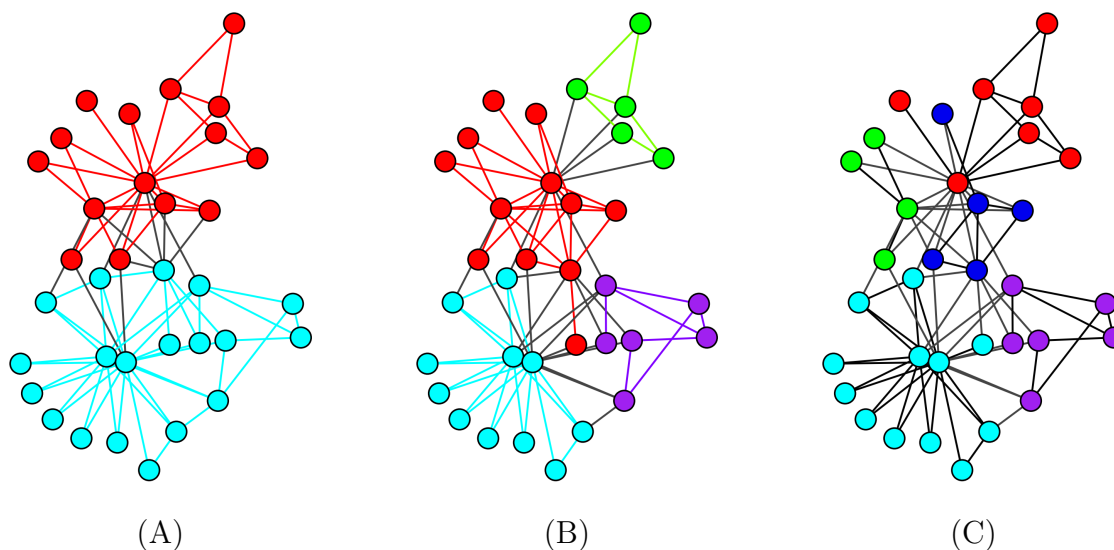


Figura 1.7: Club de Karate de Zachary. Comunidades encontradas por los algoritmos: (A) Método de Girvan-Newman, (B) Método de Louvain y (C) Método de Newman. Las comunidades resultantes se indican por colores.

1.3.5. Distribución de comunidades

Recientemente, se ha desarrollado una amplia variedad de algoritmos para detectar comunidades. Todos estos métodos tienen como objetivo encontrar comunidades inmersas en la red. De acuerdo a la literatura el número y tamaño de las comunidades en

una red están determinados por el diagrama de la red. Esto, nos lleva a una pregunta: ¿Cuál es la distribución de estas comunidades?

Muchos estudios reportan que en una variedad de redes, las distribuciones de sus comunidades en general exhiben muchas comunidades pequeñas que coexisten con unas pocas comunidades grandes, independientemente del método de detección aplicado [19] y que el patrón de sus distribuciones se ajusta bien a una ley de potencia ($\sim k^{-\lambda}$). Sin embargo, esto sugiere que la distribución del tamaño de la comunidad no es resultado del algoritmo en particular, sino más bien de las propiedades intrínsecas de la red. En conclusión, podemos decir que la caracterización de las comunidades es independiente al método aplicado para encontrar las comunidades [50].

Capítulo 2

Redes sociales

A finales del siglo XX, una clase de estudio basada en la teoría de redes con alto impacto en la psicología, sociología y ciencias políticas fue el de las redes sociales. El análisis de las redes sociales proporciona una representación de las relaciones interpersonales (grupos de personas, empresas, etc) basada en redes o grafos que llevan perspectivas sugerentes de patrones de agrupamiento que ocultan los datos cuantitativos. Los principales avances se pueden apreciar desde 1930. Los trabajos que destacan, son realizados por Jacob L. Moreno en la psicología que marcó el inicio de la sociometría. En la antropología, las investigaciones hechas por Bronislaw Malinowski, Alfred Radcliffe-Brown y Claude Lévi-Strauss formaron las bases de la teoría de las redes sociales [51, 52]. Más tarde, en 1956 sosteniéndose en la teoría del cambio de actitud de Friz Heider [23], el trabajo de Cartwright and Harary [24] proporciona un fuerte avance para analizar las relaciones personales con la teoría del equilibrio social; más adelante se explicará en que consiste esta teoría y su importancia para este trabajo.

En la década de 1990 un número creciente de investigadores trabajaron para combinar diferentes líneas de investigación. Entre ellos, destaca el antropólogo Robin Dunbar. Dunbar realizó estudios en ecología investigando a grupos de primates [53]. A partir de este estudio, él mostró que el tamaño del grupo y sus patrones de agrupación están limitados por factores cognitivos. A esto último se le conoce como “La hipótesis del cerebro social”; según esta hipótesis existe una relación entre la evolución del cerebro de los primates y el tamaño de sus grupos sociales derivado de la necesidad de manejar y gestionar grupos sociales más grandes. Esto es, la capacidad cerebral impone en la especie un límite específico al tamaño del grupo. En 1992, Dunbar extrapuló estos resultados en primates a las relaciones humanas. Esta teoría, predijo la proporción teórica del grupo social de los humanos. A este factor de escala, se le conoce como “Proporción de Dunbar’ y se describirá en la siguiente sección.

2.1. Teoría de Dunbar: Proporción de Dunbar

Como se ha mencionado previamente, la hipótesis del cerebro social de primates aplicado a los humanos, predice las relaciones de nuestras amistades en círculos sociales, y éstas pueden ser estructuradas en un sociograma de círculos concéntricos, conocido como “Círculos de Dunbar”. El tamaño inclusivo de cada círculo, sigue una progresión casi geométrica con una proporción de 3: 5,15,50,150. La evidencia para apoyar esta predicción proviene de datos de varios grupos etnográficos y sociológicos [54–58]. En la

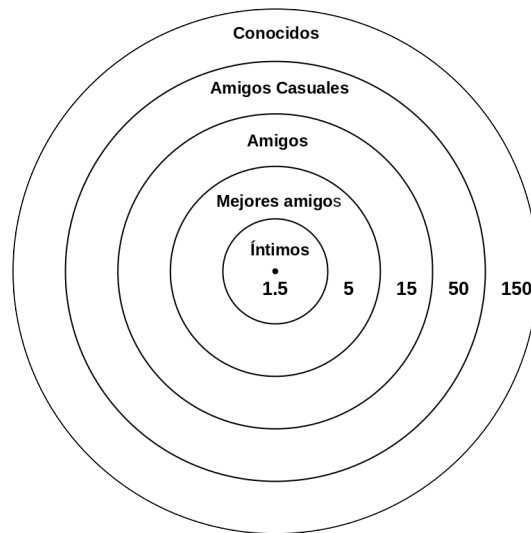


Figura 2.1: Círculos de Dunbar. El punto representa a la persona cuya red social está ilustrada. Los círculos concéntricos representan el número e intensidad de las relaciones que conforman la red. Imagen modificada de [54]

Figura 2.1 se representa esto. Estos círculos presentan diferencias, aunque el número de personas va aumentando, la intensidad de las relaciones afectivas va disminuyendo.

- **Primer círculo.** Parte central, es para las relaciones íntimas de pareja o sentimentales.
- **Segundo círculo.** Es el de los mejores amigos, son los que brindan ayuda y apoyo emocional.
- **Tercer círculo.** Es el de los amigos cercanos o simplemente amigos, son el grupo de simpatía, que contienen las relaciones importantes con contacto frecuente [58–60]
- **Cuarto círculo.** Es el de los amigos casuales o grupo de afinidad [54]. En este punto, las relaciones son menos importantes, esto es, se vuelven menos sobre apoyo

emocional y afección, y más sobre información útil e intereses específicos.

- **Quinto círculo.** Es el de los conocidos o red activa, son las que contienen las relaciones sociales periféricas [59, 60]. Esto es, se refiere a las relaciones en que las personas hacen un menor esfuerzo en mantener contacto [61].

Aunque el marco de trabajo de la Teoría de Dunbar, ha demostrado ser precisa y útil, sólo se ha aplicado relaciones positivas; actualmente no se conocen estudios de la proporción de Dunbar en relaciones negativas. Sin embargo hay investigaciones que estudian y analizan los diferentes niveles de agresividad y hostilidad en redes escolares [62–64], ninguno de ellos reportan alguna proporción entre los distintos niveles.

En general, el campo de estudio de las redes sociales se ha centrado en gran medida a las redes no signadas (redes con sólo enlaces positivos). Pero en la realidad las relaciones sociales pueden ser de dos tipos: positivas o negativas. Para representar tales sistemas, la herramienta más apropiada de la teoría de redes son las redes signadas. En el contexto social, las redes signadas representan relaciones entre las personas, los enlaces positivos indican amigos, confianza, y agrado, mientras que los enlaces negativos indican lo opuesto. Este trabajo se centra principalmente, en el impacto de las relaciones negativas en la estructura comunitaria calculada por métodos de detección de comunidades. A continuación describiremos algunos conceptos y propiedades de las redes signadas. También revisamos las medidas, dinámicas y las distintas propiedades globales que los enlaces positivos y negativos introducen las teorías sociales.

2.2. Redes signadas.

Una red signada G consiste en un conjunto de N nodos $\{v_1, v_2, v_3, \dots, v_N\}$, un conjunto de enlaces positivos L^+ , y un conjunto de enlaces negativos L^- . Una manera útil de representar una red signada es mediante su matriz de adyacencia $W \in \mathbb{R}^{N \times N}$. En particular, para redes signadas que involucran relaciones interpersonales (amistad/enemistad) entre personas, y son de interés para este estudio:

$$w_{ij} = \begin{cases} 1 & \text{si } i \text{ tiene un enlace positivo (amistad) con } j \\ -1 & \text{si } i \text{ tiene un enlace negativo (enesmitad) con } j \\ 0 & \text{otro caso} \end{cases} \quad (2.1)$$

La Figura 2.2 se observa un ejemplo de una red signada.

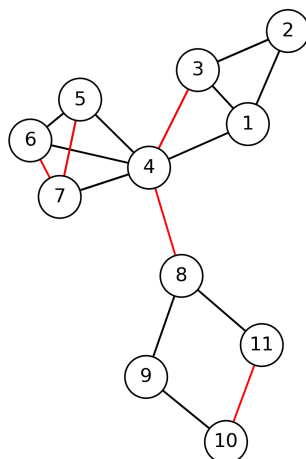


Figura 2.2: Red signada. Esta red tiene 11 nodos con 16 enlaces, de los cuales 11 son enlaces positivos (negro) y 5 negativos(rojo).

2.2.1. Propiedades de los enlaces negativos en las redes

Aún cuando las propiedades de los enlaces positivos están bien desarrolladas y son conocidas, tales como la distribución de grado, la alta reciprocidad y transitividad, no es fácil extender estas propiedades de los enlaces positivos, a los enlaces negativos. En consecuencia, explicaremos dos propiedades esperadas de los enlaces negativos.

2.2.1.1. Reciprocidad

La reciprocidad, según Falk and Fischbacher [65], significa que las personas recompensan las acciones amables, y rechazan las acciones perjudiciales. Estudios han mostrado que los enlaces positivos muestran una mayor reciprocidad [66]. Esto es, en redes con enlaces positivos es muy probable que los nodos se enlacen en ambas direcciones, mientras que los enlaces negativos muestran una reciprocidad menor, inclusive esta propiedad se ha presentado en relaciones personales virtuales [67].

2.2.1.2. Transitividad

La transitividad se da cuando un conjunto de tres nodos se conectan (lo denotaremos como tríada), es decir, si los nodos $i \leftrightarrow j$ y $j \leftrightarrow k$ se conectan entonces también se conectan $i \leftrightarrow k$. Los enlaces positivos muestran una fuerte transitividad, esto se explica desde que las relaciones de “los amigos de mis amigos son mis amigos” suceden con frecuencia [66]. En su contraparte, para los enlaces negativos, los autores Tang et al [66], muestran que este tipo de enlaces presentan una transitividad débil, debido a que en

las redes que estudiaron observaron por igual que “los enemigos de los enemigos son mis amigos” y “los enemigos de mis enemigos son mis enemigos”.

2.3. Comunidades signadas

Ha surgido un creciente interés en las comunidades de sistemas reales. De hecho, se ha concluido que agrupar una red en comunidades con conexiones relativamente densas entre sus nodos proporciona nuevas nociones sobre los rasgos distintivos de la red. Sin embargo, la mayoría de las investigaciones sobre redes, y por ende en comunidades, se han enfocado en las relaciones positivas. Actualmente, ha habido intentos de desarrollar métodos para detectar comunidades en redes signadas, dando lugar a la definición de comunidad signada. Esta definición es una extensión misma de la definición de comunidad en redes no signadas. Tal extensión es sencilla al añadir el criterio “el menor número posible de enlaces negativos internos” [68]. En este sentido, el objetivo es encontrar comunidades con nodos densamente conectados con enlaces positivos internos, y externamente conectados con enlaces negativos a otros nodos de comunidades diferentes (ver Figura 2.4), sin embargo en la literatura hay pocos métodos diseñados para redes signadas. Uno de los primeros pasos, es la extensión de la función de modularidad (Ec. 1.6) para redes signadas.

2.4. Modularidad para redes signadas

Al añadir dos tipos de enlaces, positivos y negativos, la modularidad signada (Q^s) debe compensar la tendencia de los enlaces positivos a formar comunidades, y los enlaces negativos en destruirlas [69]. Implica que es necesario introducir dos términos:

El primero es Q^- que se define de manera análoga a Q^+ para los enlaces negativos:

$$Q^- = \frac{1}{2w^-} \sum_i \sum_j \left(w_{ij}^- - \frac{w_i^- w_j^-}{2w^-} \right) \delta(C_i, C_j) \quad (2.2)$$

donde $\delta(C_i, C_j) = 1$ si los nodos i y j pertenecen a la misma comunidad y $\delta(C_i, C_j) = 0$ en otro caso. El símbolo w^- representa el número de enlaces negativos.

El segundo termino, es Q^+ definida en la Ec.1.6.

Así, Q^s , es una combinación lineal de las contribuciones Q^+ y Q^- , correspondientes a sus fuerzas de división positivas y de negativas. Matemáticamente, Q^s se escribe, como:

$$Q^s = \frac{2w^+}{2w^+ + 2w^-} Q^+ - \frac{2w^-}{2w^+ + 2w^-} Q^- \quad (2.3)$$

Nótese que en ausencia de enlaces negativos $Q^s = Q^+$, esto es se recupera la modularidad original.

2.5. Porcentaje de Perturbación Comunitario

Para entender mejor el rol de los enlaces negativos, introduciremos el Porcentaje de Perturbación causado por enlaces negativos en las Comunidades (PPC). Este concepto usa la modularidad para medir tal perturbación:

$$\text{PPC} = \frac{Q^+ - Q^s}{Q^+} \% \quad (2.4)$$

Observamos que en la ec. 2.3, Q^s es la diferencia de Q^+ y Q^- . En consecuencia, si existe una buena partición, el valor de Q^+ es grande mientras que el valor de Q^- es pequeño, concretamente, Q^- es el valor que mide la perturbación debido a los enlaces negativos en las comunidades.

2.6. Teoría del Equilibrio Social: Balance y frustración.

Un análisis que guarda relación entre la teoría de redes y otras disciplinas tanto sociales como naturales, es la teoría del equilibrio social. Esta teoría ha brindado herramientas esenciales para entender el impacto de las interacciones individuales en la estructura total de las redes signadas. Entre muchas medidas desarrolladas para cuantificar tal impacto, mencionamos dos que conciernen a la tesis: el balance y la frustración.

2.6.1. Balance

En teoría de redes, el concepto de balance, se basa en todas las formas posibles en la que se pueden signar una tríada (ver Figura 2.3), y postula que tríadas con tres enlaces positivos, y los que tienen un sólo enlace positivo y dos enlaces negativos, son socialmente balanceados, y son más frecuentes en redes signadas. Entonces, las propiedades como “el amigo de mi amigo, es mi amigo”, y “el enemigo de mi enemigo, es amigo” representan estos estados balanceados [24]. A partir de ahí, se derivan una cantidad de significados y utilidad que se le puede dar al concepto de balanceado [70]. En nuestro caso, expandimos este concepto para determinar si las comunidades que obtenemos son socialmente balanceadas. Formalmente, definimos el Balance Social Comunitario (BSC)

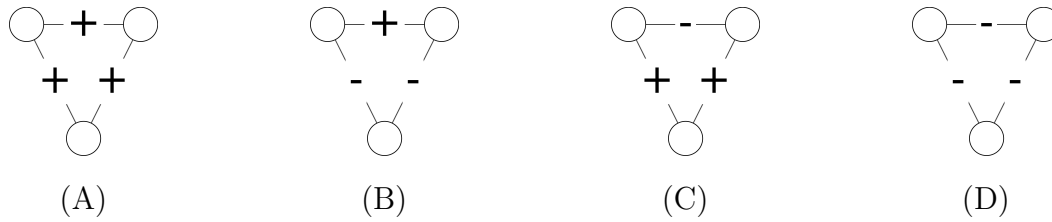


Figura 2.3: Balance social en redes signadas. Cuatro estados posibles en redes signadas. (A) y (B) son estados balanceados, mientras que (C) y (D) son estados desbalanceados. Los símbolos (+) indica relación positiva (amistad), y (-) indica relación negativa (enemistad).

como la razón de triángulos balanceados (Δ_c^b) en las comunidades entre el número total de triángulos dentro de ellas (Δ_c):

$$\text{BSC} = \frac{\Delta_c^b}{\Delta_c} \quad (2.5)$$

2.6.2. Frustración

La medida de frustración se basa en dos ideas. En primer lugar, la partición de comunidades signadas deben tener densas conexiones internas positivas y tantas conexiones externas negativas, como sea posible. En segundo lugar, deben tener el menor número posible de conexiones internas negativas, y escasas conexiones externas positivas. La frustración se opone a estas ideas, y se define la frustración como la suma de conexiones internas negativas y conexiones externas positivas.

$$F(C_1, C_2, \dots, C_k) = \frac{1}{w^+ + w^-} \sum_{ij} w_{ij}^- \delta(C_i, C_j) + w_{ij}^+ (1 - \delta(C_i, C_j)) \quad (2.6)$$

donde $\{C_1, C_2, \dots, C_k\}$ es una partición en k comunidades de la red signada, w^+ es el número de enlaces positivos, w^- es el número de enlaces negativos, w_{ij}^+ y w_{ij}^- representan los enlaces positivos y negativos entre las comunidades i y j , respectivamente. $\delta(C_i, C_j) = 1$ si $i = j$ y cero en otro caso [71]. Normalizamos la frustración con respecto al número total de enlaces en la red. Entonces, esto clarifica la definición de frustración como el número de enlaces que alejan a la red a un estado de estabilidad social.

Para aclarar un poco los conceptos de balance comunitario y frustración, usaremos como ejemplo la red de la la Figura 2.4. Está red esta dividida en tres comunidades. Cada comunidad es representada por los colores azul, verde y rojo, respectivamente. Los enlaces positivos se indican en color azul, y los enlaces negativos en rojo. Inmediatamente, es posible calcular su balance y frustración de manera sencilla. Para esto, contamos el número de triángulos en cada comunidad, para este caso son 1 en la comunidad

azul, 4 en las comunidades rojas y verdes. En total son 9 triadas: 7 triadas balanceadas y 2 desbalanceadas. Por lo tanto, de la ec. 2.5, el BSC $7/9 = 0.78$. Por otra parte, para la frustración vemos que la red tiene 24 enlaces, de los cuales 3 cumplen la ec. 2.6, así $F = 3/24 = 0.13$.

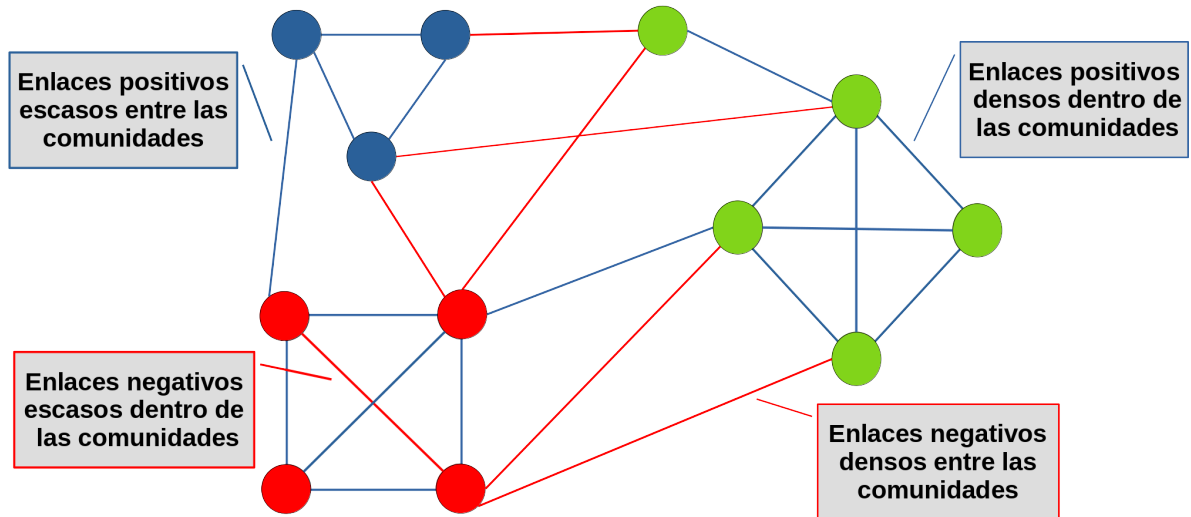


Figura 2.4: Comunidades signadas. Esta red está dividida en 3 comunidades signadas (azul, roja y verde), tiene 12 nodos, 24 enlaces, de los cuales 17 son enlaces positivos y 7 negativos. Notar que esta estructura comunitaria, coincide con la definición de comunidad signada. Los enlaces azules indican relaciones positivas, mientras que los rojos son negativos.

Hoy en día, la importancia de la teoría del equilibrio social en el ámbito del estudio de las comunidades constituyen la columna vertebral para el análisis de redes signadas.

Capítulo 3

Modelo de formación de redes signadas en escuelas

En este capítulo se describe la metodología propuesta en esta tesis para construir las matrices de adyacencia y comunidades a partir de encuestas aplicadas entre mayo a octubre del 2019 a estudiantes de 9 escuelas de diferentes niveles educativos. En la tabla 3.1, se enlistan las 9 escuelas participantes, y la nomenclatura utilizada para nuestro estudio.

Tabla 3.1: Nombres y nomenclaturas de las escuelas participantes en esta tesis.

Nombre de la escuela	Nivel Educativo	Nomenclatura
“Silvestre Chi”	Primaria	epSC
“Rafael Ramirez Castañeda”	Primaria	epRRC
“Ignacio Zaragoza”	Primaria	epIZ
“Raquel Dzib Cicero”	Secundaria	esRDC
“Técnica Número 2”	Secundaria	esTN2
“José León Bojorquez García”	Secundaria	esJLBG
“Carlos Castillo Peraza”	Preparatoria	eprCCP
“Hunucmá-UADY”	Preparatoria	eprHUN
“Tecnológica del Sur”	Universidad	unTRS

Las escuelas se ubican en los municipios de Mérida, Umán, Abalá, Hunucmá y Tekax en el Estado de Yucatán (ver Figura 3.1). En las encuestas, los estudiantes indicaron sus amistades y enemistades. En este proceso, se desarrolló un método para completar enlaces negativos faltantes en algunas escuelas, y que más adelante se explicará a detalle.

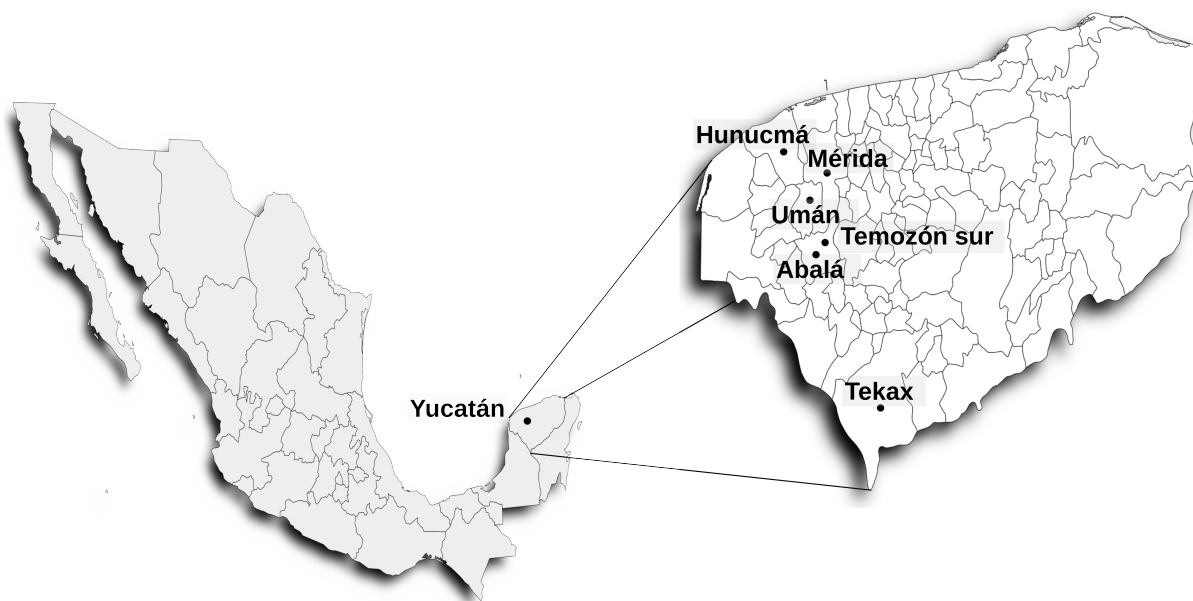


Figura 3.1: Mapa del Estado de Yucatán. El mapa muestra las ubicaciones donde se aplicaron las encuestas. Dos de las tres escuelas primarias se encuentran en Abalá: la epSC en Temozón Sur y la epRRC en el centro del municipio. La primaria epIZ está en Mérida. Las secundarias están localizadas como sigue: esRDC en Umán, esTN2 en Mérida y esJLBG en Hunucmá. De las dos preparatorias, la eprHUN se ubica en Hunucmá y la eprCCP en Mérida. Por último, la universidad unTRS está en Tekax. Los municipios de Abalá, Hunucmá, Tekax son zonas rurales, mientras que Mérida y Umán no.

3.1. Recopilación de datos

Para llevar a cabo nuestros objetivos, se utilizarán los datos recopilados del Proyecto de CONACYT titulado “Estudio de la topología y generación de modelos de difusión en redes sociales en escuelas a diferentes grados académicos ” con folio 282692 CB2016-01. El estudio también fue aprobado por el Comité de Bioética para la Investigación en Seres Humanos del Cinvestav, y se obtuvo el permiso por escrito de los directores de cada escuela para encuestar a sus estudiantes. En las primarias, además se obtuvo la autorización por escrito de los padres/tutores para permitir la participación de sus

hijos en las encuestas. En las secundarias y preparatorias, las autoridades escolares informaron a los padres/tutores sobre la participación de sus hijos, quienes dieron su consentimiento oralmente. En el mismo día de la encuesta, a cada estudiante se les explicó una cláusula de consentimiento que podían aceptar o declinar el uso de sus respuestas. En caso, que el estudiante aceptará, firmaba electrónicamente con su nombre al final de la encuesta.

Las encuestas, aunque similares, se aplicaron de dos maneras distintas. En las tres escuelas primarias (epSC, epRRC, epIZ), la aplicación fue de manera escrita en papel; esta manera fue elegida debido a la inviabilidad de uso de dispositivos electrónicos. En las escuelas restantes la aplicación fue de manera electrónica vía internet a través de una plataforma digital diseñada para este propósito. La base de datos fue guardada en un servidor del Cinvestav-Mérida. En el Apéndice A.1 se anexan capturas de esta plataforma. La recolección fue supervisada por personal profesional en el área, luego de una breve descripción de la información necesaria para realizar la investigación. Los datos se recopilaron de manera confidencial, y se analizaron de forma anónima. En la Tabla 3.2 se muestran la información y características generales de cada escuela.

Tabla 3.2: Características de las escuelas participantes en esta tesis.

Escuela	Edad(años)	Localización	Estudiantes
epSC	6 a 12	Rural	108
epRRC	6 a 12	Rural	226
epIZ	6 a 12	Urbana	419
esRDC	12 a 15	Urbana	613
esTN2	12 a 15	Urbana	457
esJLBG	12 a 15	Rural	270
eprCCP	15 a 18	Urbana	1497
eprHUN	15 a 18	Rural	74
unTRS	18 a 22	Rural	664

La encuesta se dividió en dos partes. Las primeras 6 preguntas son para la información general del estudiante. La segunda parte con 7 preguntas, recogen la información sobre las relaciones de amistad, enemistad entre los estudiantes de la misma escuela. También, hicimos dos preguntas donde los estudiantes definieran como perciben la amistad y enemistad.

3.2. Construcción de las matrices de adyacencia

Con la información obtenida de las encuestas, se construyeron las 9 correspondientes matrices de adyacencia (A). Las redes originalmente son redes dirigidas signadas.

Naturalmente, las matrices resultantes son:

$$A_{ij} = \begin{cases} 1 & \text{Si } i \text{ y } j \text{ si tienen una relación de amistad} \\ -1 & \text{Si } i \text{ y } j \text{ si tiene una relación de enemistad} \\ 0 & \text{otro caso} \end{cases} \quad (3.1)$$

Clasificamos las relaciones positivas y negativas en dos niveles de intensidad. Nos referimos, para una relación recíproca positiva ($A[i, j] = A[j, i] = 1$) como “mejor amigo” entre los pares de nodos i y j . Una relación positiva unidireccional indica una relación de “amigo”. Análogamente, definimos las relaciones de “peor enemigo” y “enemigo” para las relaciones negativas.

3.3. Proporción de Dunbar: Datos empíricos

En este apartado, ordenamos y cuantificamos los valores de las relaciones de amistad y enemistad de los datos empíricos. En las Tablas 3.3 y 3.4 se presentan los resultados:

Tabla 3.3: Valores de las relaciones de amistad de las redes escolares. Los datos empíricos fueron obtenidos de las encuestas. La proporción de Dunbar se muestra en la última columna, tal valor es obtenido de la razón del número de relaciones de amigos entre el número de relaciones de mejores amigos.

Escuelas	Nodos	Amigos	Mejores amigos	Promedio de amigos	Promedio de mejores amigos	Proporción de Dunbar
epSC	108	908	312	16.8	5.7	2.9
epRRC	222	1844	553	16.6	4.9	3.3
epIZ	417	3974	1403	19.0	6.7	2.8
esRDC	473	2390	583	10.1	2.4	4.1
esTN2	390	2211	629	11.3	3.2	3.5
esJLBG	237	1118	387	9.4	3.2	2.9
eprCCP	1429	8535	3025	11.9	4.2	2.8
eprHUN	62	270	95	8.7	3.0	2.9
unTRS	561	2711	834	9.6	2.9	3.3

En las dos tablas anteriores se observa que el número de nodos son diferentes a los de la Tabla 3.2, tal diferencia es debida a que algunos alumnos por diversas razones no participaron. Adicionalmente, en la Figura 3.2 se presenta una gráfica de las proporciones de Dunbar de cada una de las escuelas.

Tabla 3.4: Valores de las relaciones de enemistad de las redes escolares. Los datos empíricos fueron obtenidos de las encuestas. La proporción de Dunbar se muestra en la última columna, este valor es obtenido dividiendo el número de relaciones de enemigos entre el número de relaciones de peores enemigos.

Escuelas	Nodos	Enemigos	Peores enemigos	Promedio de enemigos	Promedio de peores enemigos	Proporción de Dunbar
epSC	108	466	141	8.62	2.61	3.3
epRRC	222	877	248	7.9	2.23	3.53
epIZ	417	1771	592	8.49	2.83	2.99
esRDC	473	782	36	3.30	0.15	21.72
esTN2	390	675	39	3.46	0.2	17.3
esJLBG	237	405	33	3.41	0.27	12.27
eprCCP	1429	1240	54	1.73	0.07	22.96
eprHUN	62	90	3	2.90	0.09	30.00
unTRS	561	525	17	1.87	0.06	30.88

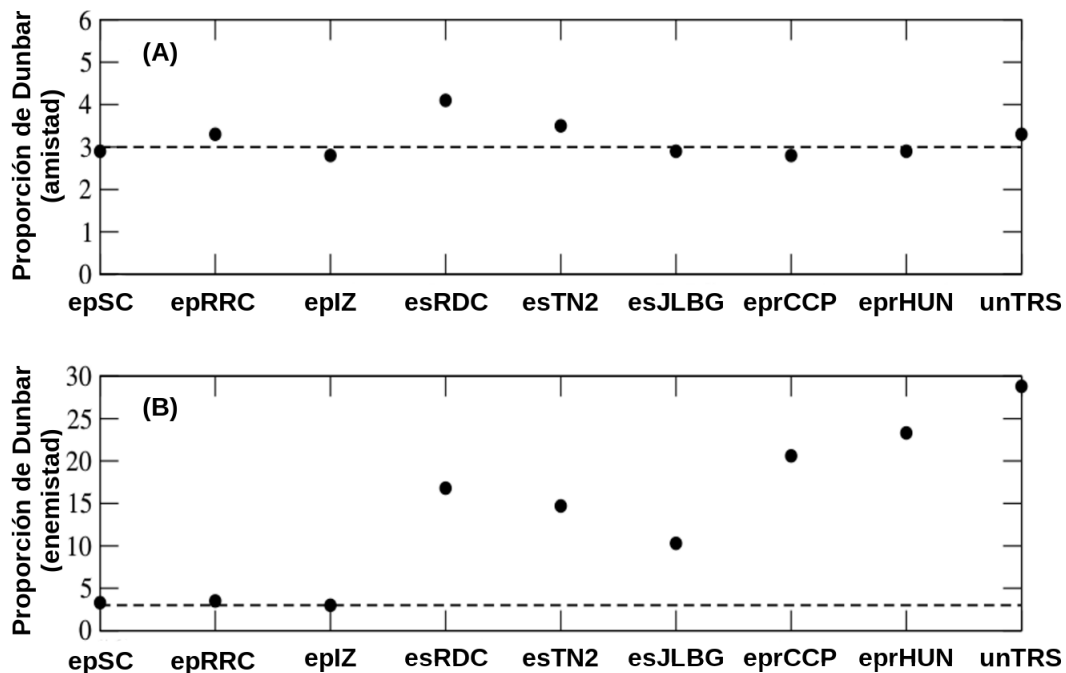


Figura 3.2: Proporción de Dunbar para los datos empíricos. A) valores obtenidos de la razón de relaciones de amigos entre mejores amigos. B) valores obtenidos de la razón de relaciones de enemigos entre peores enemigos. La línea discontinua que pasa por tres, indica el valor de referencia de Dunbar.

Asimismo, se observa que las relaciones de amistad se acercan al valor teórico de la proporción de Dunbar. Con respecto a las relaciones negativas, en las escuelas primarias los resultados coinciden con dicha proporción, en cambio, para el resto de las escuelas dicha proporción es mucho mayor.

3.4. Missing links

En la práctica, el método más utilizado para obtener las muestras de enlaces de una red social, son las respuestas a encuestas. No obstante, en las muestras hay enlaces que no se pueden obtener, debido, por ejemplo, a mala memoria, malentendidos en las preguntas de los encuestados, errores al introducir los datos o encuestados que no quieren proporcionar su información. A los enlaces que no se pueden obtener en la muestra, se les conoce como “missing links”. Los missing links plantean problemas metodológicos importantes a los investigadores debido a la falta de información para modelar la red. Predecir estos enlaces es un reto. Existen técnicas para predecir los missing links [72], pero las que nos interesan son para los enlaces negativos.

En nuestros datos, observamos que en las redes de las secundarias, preparatorias y de la universidad, hay missing links. Lo anterior ha sido observado en otros trabajos [73]. Este fenómeno no se observó en las escuelas primarias. Creemos que la carencia de enlaces negativos hace que la red no represente al sistema real, que tenemos datos incompletos y que estos enlaces se pueden predecir [66]. Hay muchos impedimentos para que los estudiantes brinden esta información sensible de sus relaciones de enemistad; a pesar que se les explicó que su información es confidencial, los participantes se sentían temerosos que su información sea accedida indebidamente, y puedan ser víctimas de represalias. Esto, a diferencia de los alumnos de nivel básico, que proporcionaban sus relaciones de enemistad con más facilidad. El propósito de esta sección es presentar una heurística (una técnica para predecir los missing links) para completar los missing links negativos en las redes de las secundarias, preparatorias y universidad. Para llevar a cabo esto se utilizó la teoría del equilibrio social (ver sección 2.6). Principalmente, se basa en la idea que “el enemigo de mis amigos es probablemente mi enemigo”. Para reducir la probabilidad de agregar un enlace negativo que no sea del sistema real, basta elegir los enlaces unidireccionales donde el estudiante i dijo ser enemigo de j , es decir, si $A_{ij} = -1$ o $A_{ji} = -1$. En la Figura 3.3 se ilustra el método que se usó para decidir si se agrega un enlace de enemistad unidireccional (línea roja discontinua) entre los nodos i y j como recíproco o no. En el Apéndice A.4 se anexa el código del programa en python que implementa la heurística para completar los enlaces negativos.

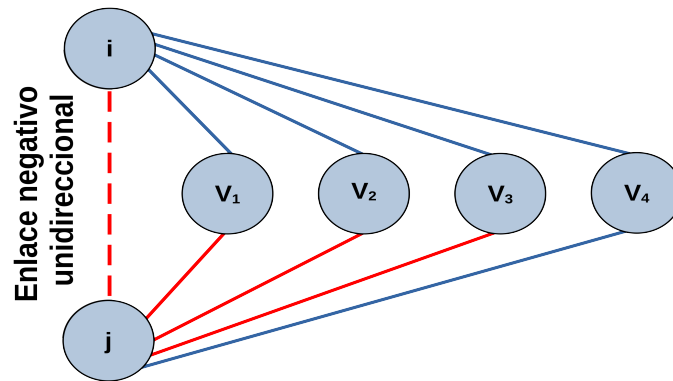


Figura 3.3: Heurística para agregar los missing links. Enlaces azules y rojos representan los enlaces recíprocos de amistad (“mejor amigos”) y enemistad (“peores enemigos”), respectivamente. La línea roja punteada indica un enlace negativo unidireccional, esto es, solamente si el estudiante i dijo ser enemigo de j , o viceversa. La figura muestra, que si este enlace fuera recíproco, habría tres triadas balanceadas y una desbalanceada. En este caso, el enlace negativo unidireccional es agregado como un enlace negativo recíproco. Imagen tomada y modificada de [74]

En resumen, en la Figura 3.3 vemos que si existiera este enlace, se formarían 4 triadas: tres triadas balanceadas y una triada desbalanceada. Dado que el número de triadas balanceadas supera al número de triadas desbalanceadas, el enlace entre i y j , se convierte en recíproco y se agrega a los observados. Esta técnica también es utilizada por los autores Lo et al [74] que se encontraron con el mismo problema.

En la Tabla 3.5 muestra el número de enlaces positivos recíprocos (w^+) y el número de enlaces negativos recíprocos (w^-) después de aplicar la metodología de la sección 3.4. Los símbolos $\langle k^+ \rangle$ y $\langle k^- \rangle$ representan los grados medios de mejores amigos y de peores enemigos, respectivamente. Además, en la columna w^- muestra el porcentaje de enlaces negativos recíprocos en las redes. Antes de agregar los missing links a las escuelas de nivel secundaria, preparatoria y universidad, el porcentaje de enlaces negativos recíprocos en las redes eran menores al 9%. Después de aplicar la metodología, observamos que este porcentaje es similar a los de nivel primaria, mientras que el resto de las escuelas los porcentajes fueron menores (excepto la eprHUN).

Tabla 3.5: Enlaces positivos y negativos de las redes signadas.

Escuelas	Nodos	w^+	w^- (%)	$\langle k \rangle^+$	$\langle k \rangle^-$
epSC	108	312	141 (31.1 %)	5.7	2.6
epRRC	222	553	248 (30.9 %)	4.9	2.2
epIZ	417	1403	592 (29.6 %)	6.7	2.8
esRDC	473	583	201 (25.6 %)	2.4	0.8
esTN2	390	629	273 (30.2 %)	3.2	1.4
esJLBG	237	387	183 (32.1 %)	3.2	1.5
eprCCP	1429	3025	410 (11.9 %)	4.2	0.5
eprHUN	62	95	32 (25.1 %)	3.0	1.0
unTRS	561	834	169 (16.8 %)	2.9	0.6

Las columnas w^+ y w^- representan los enlaces positivos y negativos, respectivamente. $\langle k \rangle^+$ y $\langle k \rangle^-$ son los amigos y enemigos por nodo. El símbolo (%) de la columna w^- indica el porcentaje de enlaces negativos.

3.5. Calibración del Método de Louvain

Previamente, se ha mencionado que el método de Louvain es de carácter estocástico. En este apartado se presentan los parámetros de resolución calculados para las 9 redes académicas al generar particiones óptimas en cada ejecución. El método tiene un parámetro llamado Resolución (\mathbf{R}) que controla la robustez del algoritmo (para más detalles ver [75]). Una buena calibración asegura particiones robustas y estadísticamente equivalentes. Aquí la robustez, se enfoca en la idea que pequeñas modificaciones en el algoritmo no deben alterar significativamente las particiones óptimas (es decir, que una partición persista durante la búsqueda del valor de óptimo de Q^+). En la Figura 3.4, se muestran los valores de \mathbf{R} con respecto al promedio del número de comunidades detectadas de tres de las 9 escuelas (epSC, esTN2, unTRS). Para seleccionar un valor de \mathbf{R} se consideraron tres criterios: el primero es elegir un intervalo alrededor de \mathbf{R} donde sea casi constante, lo que evita que una pequeña perturbación de \mathbf{R} pueda causar una gran variación en el número de comunidades. El segundo criterio, es tener una varianza pequeña para reducir el error con respecto al valor medio, en otras palabras obtener comunidades lo más homogéneas posibles. Y el último criterio es que el valor de \mathbf{R} sea lo más cercano a 1 que corresponde a la escala original en la que el algoritmo de Louvain maximizó Q^+ . Los tres criterios ayudan a reducir el error y el número de simulaciones. Se observó que aproximadamente 150 simulaciones son suficientes para obtener buenos resultados, sin embargo, se hicieron 300 simulaciones para obtener una mejor precisión. En el Apéndice A.4 se anexa el código para la calibración de este método. En [75] explican a más detalle este proceso de calibración del método de Louvain.

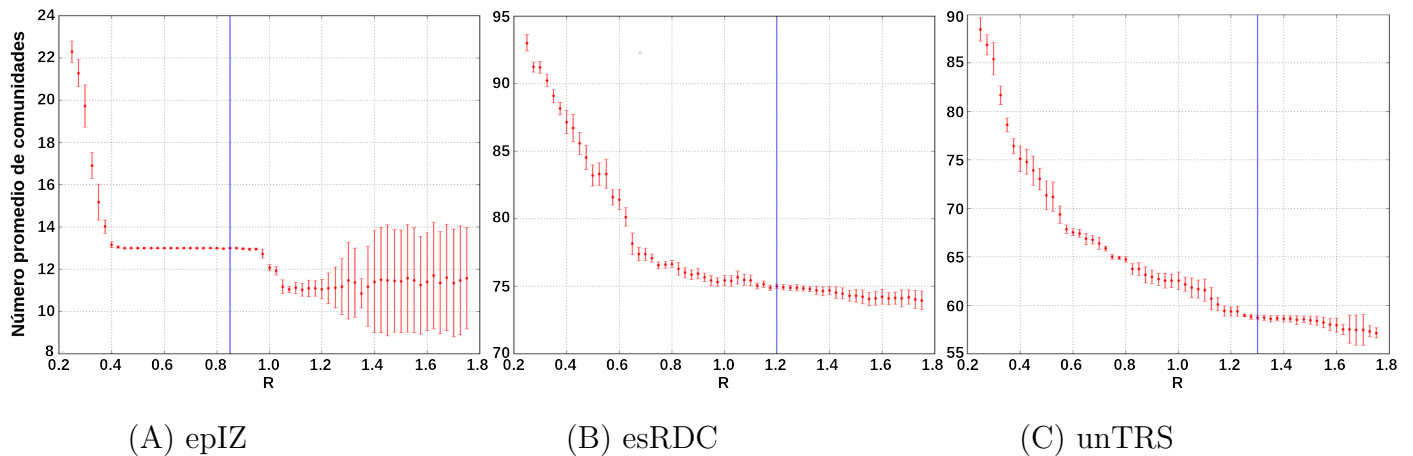


Figura 3.4: Calibración del Algoritmo de Louvain. En la figura graficamos el número promedio de comunidades y su desviación estándar en función del parámetro \mathbf{R} en el intervalo $[0.25, 1.75]$. (A) calibración de la escuela primaria epIZ, el valor óptimo de \mathbf{R} es 0.85. (B) calibración de la secundaria esRDC, el valor óptimo de \mathbf{R} es 1.20. (C) calibración de la universidad unTRS, el valor óptimo de \mathbf{R} es 1.30. Cada valor fue obtenido de 300 simulaciones.

3.6. Construcción de las comunidades signadas

Para construir las comunidades signadas, se efectuó el siguiente procedimiento. En cada una de las 9 escuelas se construyó su red signada a partir de su matriz de adyacencia. Primero, se extrajo la subred de enlaces positivos y se aplicaron los dos algoritmos de detección de comunidades: el Newman y Louvain, para obtener particiones óptimas de comunidades. Posteriormente, a cada partición se le agregaron las relaciones negativas entre sus nodos. De esta manera, obtuvimos particiones óptimas creadas a partir de relaciones positivas que contienen relaciones negativas que perturban su estabilidad.

Capítulo 4

Resultados

La investigación realizada en esta tesis fueron publicados en la revista de Plos One en el artículo titulado Modularity, balance, and frustration in student social networks: The role of negative relationships in communities en Diciembre del 2022 [31].

En el Apéndice A.4 se anexan los programas importantes para los cálculos de las propiedades generales de las redes signadas, el cálculo de las particiones usando los métodos de detección de comunidades mencionados en la tesis y en los mismos programas se calculan las medidas de modularidad, balance y frustración. Todos los programas se escribieron en el lenguaje Python.

4.1. Proporción de Dunbar para los enlaces negativos

Después de incluir los missing links en la sección 3.4 (datos en la Tabla 3.5), se calcularon las proporciones de Dunbar para los missing links. En la Tabla 4.1 se presentan estos valores. Seguidamente, se graficaron estas nuevas proporciones de Dunbar

Tabla 4.1: Valores numéricos al incluir los missing links. la tercera columna indica los porcentajes de enlaces negativos añadidos. La sexta indica los porcentajes de la proporción de Dunbar para las relaciones negativas.

Escuelas	Nodos	% enlaces negativos añadidos	Promedio de enemigos	Promedio de peores enemigos	Proporción de Dunbar
esRDC	473	18	3.30	0.84	3.89
esTN2	390	31	3.46	1.40	2.47
esJLBG	237	33	3.41	1.54	2.21
eprCCP	1429	27	1.73	0.57	3.02
eprHUN	62	30	2.90	1.03	2.81
unTRS	561	28	1.87	0.60	3.10

en la Figura 4.1. En la misma figura, notamos que todos los valores son cercanos a la proporción teórica de Dunbar.

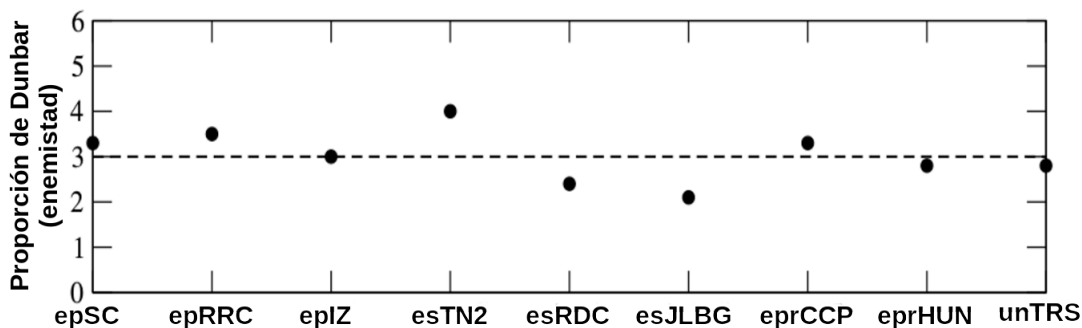


Figura 4.1: Proporción Dunbar para los enlaces negativos. Los valores fueron obtenidos de dividir las relaciones de enemigos entre las relaciones de peores enemigos. Los cálculos para las secundarias, las preparatorias y la universidad incluyen los missing links. La línea discontinua que pasa por 3, indica la referencia de la proporción de Dunbar.

4.2. Análisis de las comunidades signadas

En la Tabla 3.5 observamos que la proporción (%) de relaciones negativas (w^-) son similares en las escuelas primarias y secundarias, mientras que en las preparatorias y la universidad esta proporción es baja. También, podemos notar que en las escuelas primarias los valores de $\langle k \rangle^+$ y $\langle k \rangle^-$ son grandes, esto indica que en estas escuelas tienen más relaciones de amistad y enemistad que en el resto de las escuelas. Estas diferencias pueden ser explicadas desde la perspectiva de la psicología, podría deberse que a edad temprana (6-12 años) los estudiantes aún no tienen bien definido su criterio para elegir sus amigos, mientras que los adolescentes y adultos jóvenes (desde los 12 años), son más maduros para elegir mejor a las amistades que forman su círculo social. Análogamente, este fenómeno ocurre en las enemistades. Las redes signadas de las escuelas primarias presentaron valores $\langle k \rangle^+$ y $\langle k \rangle^-$ grandes, como resultado permite la formación de grupos visibles. En la Figura 4.2 se observa grupos de nodos de la escuela epIZ con ambos enlaces positivos y negativos, estos grupos están relacionados con la distribución espacial de los alumnos en las escuelas. Esta distribución espacial es muy importante cuando se forman comunidades, debido a que las comunidades probablemente sean conformadas por estudiantes del mismo salón. Este fenómeno no es evidente en el resto de las escuelas, como se observa en la Figura 4.3 que muestra las comunidades de la secundaria esRDC, aquí en gran parte las comunidades no corresponden a los salones. En otras palabras, un salón puede estar conformado de varias comunidades.

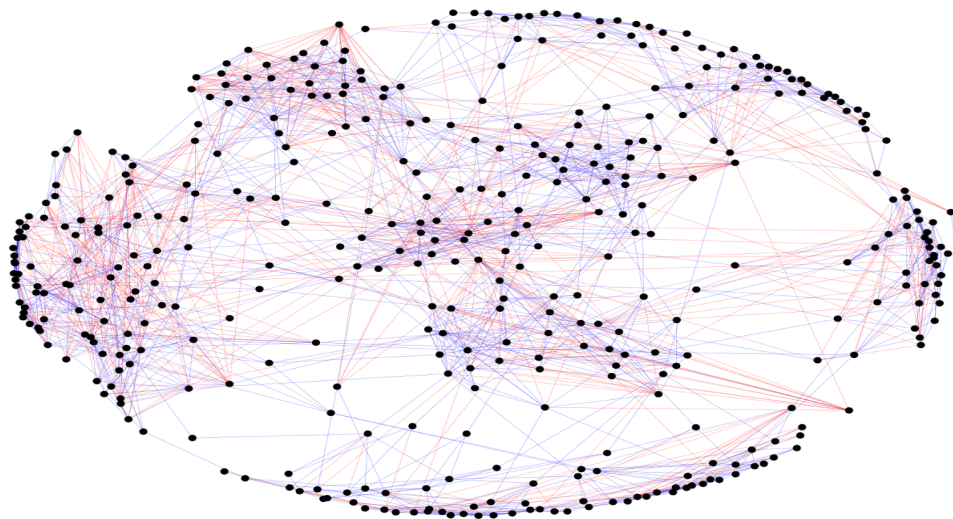


Figura 4.2: Red signada de la escuela primaria epIZ. Esta red tiene 417 estudiantes y 1995 relaciones, de los cuales el 29.6% son de enemistad. Cada estudiante tiene en promedio 6.7 y 2.8 relaciones de amistad y enemistad, respectivamente.

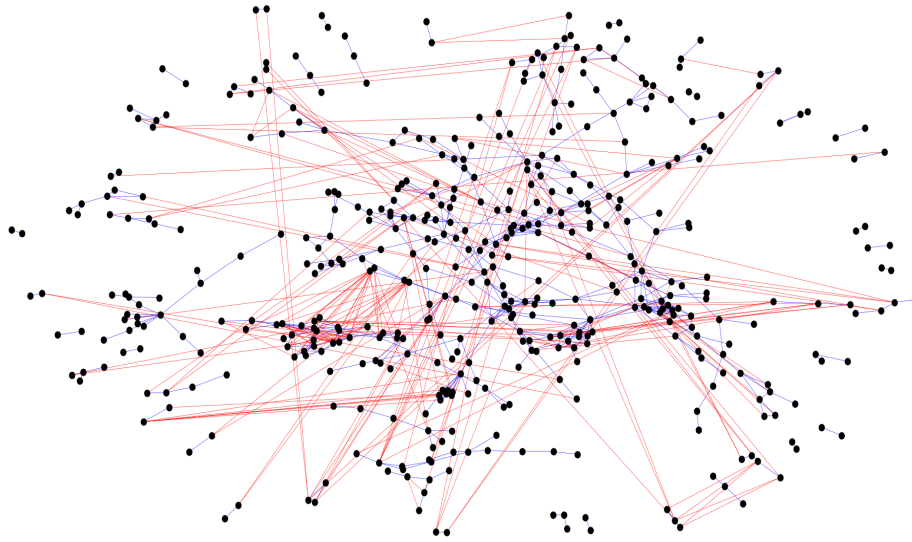


Figura 4.3: Red signada de la secundaria esRDC. Esta red tiene 473 estudiantes y 784 relaciones, de los cuales el 25.6 % son de enemistad. Cada estudiante tiene 2.4 y 0.8 relaciones de amistad y enemistad, respectivamente.

En cada escuela se detectaron comunidades aplicando los métodos de Newman y Greedy, en la Tabla 4.2 se muestran los valores de algunos parámetros obtenidos. No

Tabla 4.2: Enlaces positivos y negativos en las comunidades.

Red	Método Newman			Método Greedy		
	Comu	w_c^+	w_c^- (%)	Comu	w_c^+	w_c^- (%)
epSC	9	215	23 (9.6 %)	6.5	258.3	42.4 (14.1 %)
epRRC	16	414	112 (21.2 %)	7.8	493.9	166.8 (25.2 %)
epIZ	19	1074	325 (23.2 %)	13.0	1255.8	462.4 (26.9 %)
esRDC	85	499	19 (3.6 %)	74.9	545.8	21.0 (3.7 %)
esTN2	55	480	40 (7.6 %)	38.8	574.9	97.0 (14.4 %)
esJLBG	33	338	28 (7.6 %)	24.8	363.9	34.7 (8.7 %)
eprCCP	80	2089	33 (1.5 %)	30.6	2565.7	116.0 (4.3 %)
eprHUN	13	71	1 (1.3 %)	10.9	81.4	2.8 (3.3 %)
unTRS	81	695	17 (2.3 %)	58.8	784.8	46.8 (5.6 %)

Resultados de los métodos de Newman y Greedy. Columnas w_c^+ and w_c^- representan los enlaces internos en las comunidades, respectivamente. Cada valor del Greedy fue obtenido de 300 simulaciones.

se encontró correlación entre el número de comunidades y el grado académico de cada escuela. Podemos observar que el número de comunidades detectadas por el algoritmo greedy es menor a comparación del método de Newman, esto se debe al diseño propio de

los métodos. Hay escuelas (epSC, epIZ y eprHUN) donde las diferencias en el número de comunidades halladas por ambos métodos no son muy grandes, mientras que en otras escuelas (eprCCP, unTRS y epRRC) el número de comunidades calculadas por el método Newman pueden ser el doble con respecto al número de comunidades detectadas por el método greedy. Los valores w_c^+ y w_c^- representan el número de enlaces positivos y negativos dentro de las comunidades, respectivamente. En el método greedy, los valores de w_c^+ y w_c^- son más altos, como es de esperarse por el tamaño de las comunidades. Sin embargo, contrariamente a lo esperado en las escuelas eprCCP, unTRS y epRRC al tener más comunidades por el método Newman que con el método greedy, los valores de w_c^+ no son muy diferentes. Por otra parte, los valores w_c^- parecen ser más consistentes con el número de comunidades de cada método. También, los porcentajes de w_c^- son diferentes a los w^- de la Tabla 3.5: en este caso, las escuelas primarias tienen valores más altos que el resto de las escuelas.

4.2.1. Distribución de comunidades

En la Figura 4.4 se representa la distribución de comunidades para cada escuela estudiada, donde las barras grises muestran la distribución para el método Newman y la línea discontinua para el método greedy. Aquí observamos que en general las comunidades generadas por el método greedy contienen un mayor número de estudiantes. Por ejemplo, en la esRDC, el rango del tamaño de las comunidades con el método Newman, abarca de 2 a 18 estudiantes, mientras que para el greedy, es de 2 a 36 estudiantes. Además, observamos la distribución de comunidades para todas las escuelas primarias, la esJLBG y la eprHUN no presentan ningún patrón. Por el contrario, para las escuelas esRDC, esTN2, y unTRS, sus distribuciones muestran un comportamiento exponencial. Las distribuciones de las tres escuelas se compararon con la distribución acumulativa exponencial complementaria $e^{-\frac{1}{\lambda}}$. En todos los casos, los ajustes fueron mayor al 90%, con $\lambda = 5.51, 6.90$ y 6.82 que representa el número promedio de estudiantes en las comunidades de esRDC, esTN2 y unTRS, respectivamente. Análogamente, $\lambda = 6.31, 10.0$ y 9.55 para el método greedy.

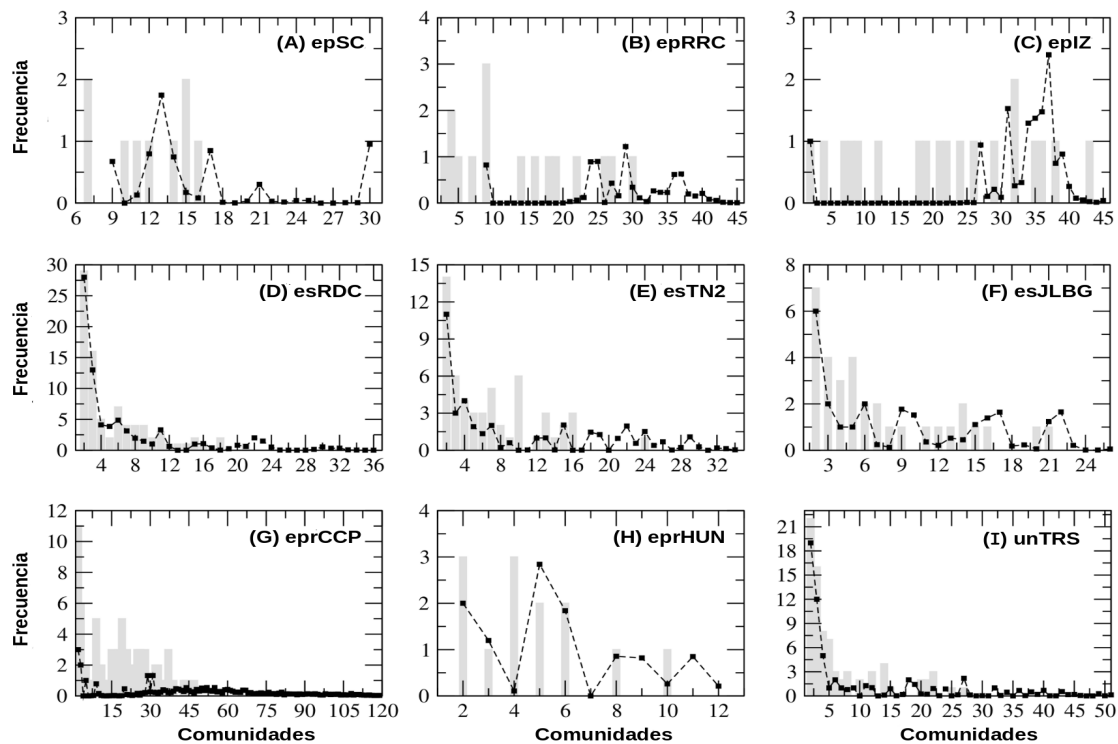


Figura 4.4: Distribución de comunidades. En las imágenes de la a) a I) se muestran las frecuencias de comunidades calculadas en cada escuela. Las barras grises son para el método de Newman, y las líneas discontinuas son para el greedy. Cada valor del método greedy fue obtenido de 300 simulaciones.

4.2.2. Subcomunidades antagónicas

Analizamos las comunidades signadas de todas las escuelas. En la Figura 4.5 mostramos algunas comunidades: 4 comunidades de la primaria epRRC detectadas por el método Newman, y 4 comunidades de la secundaria esTN2 usando el método greedy. En ambos casos, se observaron enlaces negativos. Se observaron dos fenómenos. En primer lugar, la presencia de nodos que tienen muchos enlaces negativos con otros nodos en la misma comunidad (nodo negativo). Sin embargo, en el sistema real, es muy probable que estos nodos no formen parte en las comunidades en los que los métodos de Newman y Louvain los asignaron. El segundo fenómeno, es que en las comunidades coexisten dos subgrupos de nodos antagónicos. Esto, probablemente representa dos subcomunidades antagónicas independientes en el sistema real. Estos dos fenómenos fueron observados mayormente en las escuelas de educación básica, y decrece en las secundarias. En las preparatorias y universidad, el número de relaciones negativas en sus comunidades es menor, incluso existen algunas comunidades que sólo contienen relaciones positivas con perturbación igual cero.

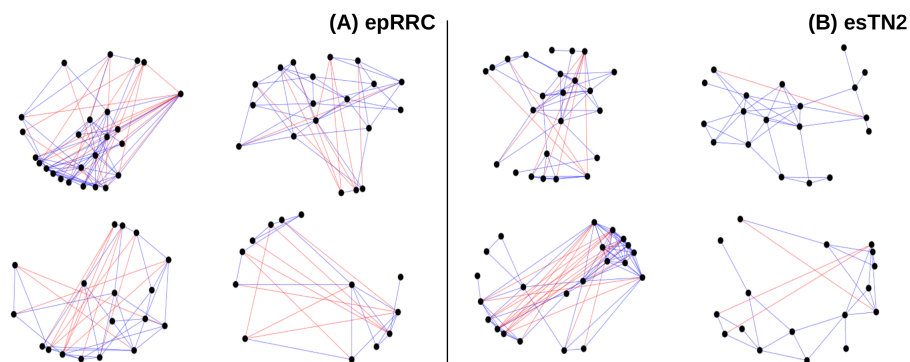


Figura 4.5: Comunidades signadas. A) 4 comunidades de la primaria epRRC (método Newman). Se observan comunidades con nodos negativos y subgrupos antagónicos. B) 4 comunidades de la escuela esTN2 (método greedy). Los mismos fenómenos se observan pero en menor medida.

4.2.3. Modularidad, Balance y Frustración

En la Figura 4.6 (A), se muestran los valores Q^+ para los métodos de Newman y greedy. Observamos, que todas las escuelas tienen altos valores de modularidad comparados con otros estudios. Por ejemplo, las escuelas esRDC, esJLBG y unTRS los valores de Q^+ son cercanos a 0.8 para el método Newman y alrededor de 0.9 para el greedy. Esto último indica que las comunidades están bien formadas por las relaciones positivas, es decir, sus nodos están densamente conectados dentro de sus comunidades y escasamente conectados con otras comunidades. También, observamos en ella que el método greedy optimiza mejor sus comunidades.

En la Figura 4.6 (B) se muestran los valores de la modularidad signada (Q^s) ec. 2.3. En la sección 2.4, Q^s se definió como la diferencia entre Q^+ y Q^- . Si existe una buena partición de comunidades signadas, es inmediato concluir que el valor de Q^+ es mucho mayor que Q^- . De esta misma figura, es muy claro que algunas escuelas la perturbación debida por las relaciones negativas son más notables (epSC, epRRC, epIZ) que otras. La Figura 4.6 (C) muestra los valores del porcentaje de perturbación Q^- en las comunidades (ver Ec. 2.4). Como consecuencia de este último resultado, se observan otros fenómenos interesantes: uno de ellos, es que algunas escuelas (esRDC, esJLBG y unTRS), a pesar de ser comunidades muy bien formadas ($Q^+ \approx 0.9$), tienen cierto grado de perturbación (entre 20 % al 35 %). El segundo, es que las comunidades de las escuelas primarias son inestables (i.e son perturbadas significativamente por las relaciones negativas); para las secundarias, sus comunidades son semi-estables, mientras que para las preparatorias y la universidad son estables. Deducimos, que los criterios para elegir sus relaciones y comunidades sociales de los estudiantes se definen mejor con transcurso del tiempo.

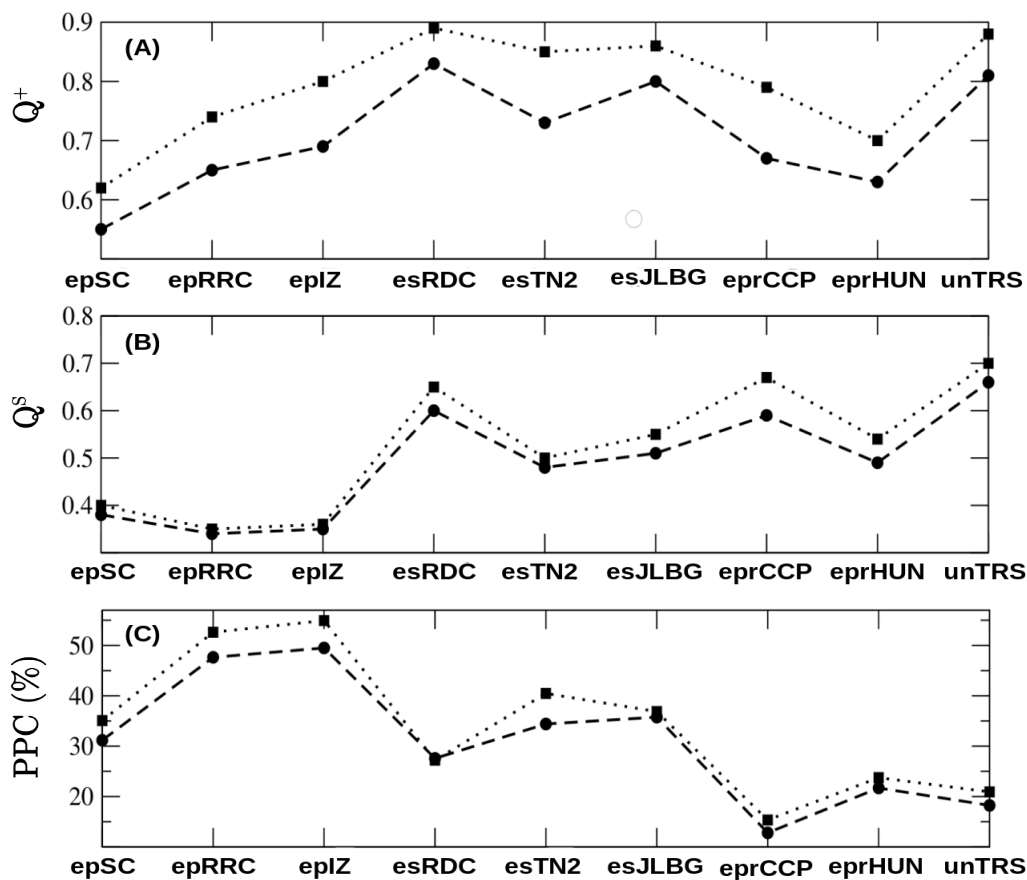


Figura 4.6: Valores de la modularidad. (A) Valores de la modularidad (Q^+) para el método Newman (círculos) y para el método greedy (cuadros). (B) valores de la modularidad signada (Q^s) para los métodos de Newman (círculos) y greedy (cuadros). (C) Porcentaje de Perturbación para los enlaces negativos en las Comunidades (PPC) para los métodos Newman (círculos) y greedy (cuadros). Cada valor del greedy fue obtenido de 300 simulaciones.

En los sistemas sociales, las comunidades perturbadas son fuentes de tensión social o de conflicto. Se podría esperar que para reducir las comunidades perturbadas, bastaría con modificar los algoritmos de detección de comunidades. Sin embargo, esto es difícil, ya que eliminar nodos o dividir una comunidad con mucha frecuencia reduce la modularidad. Actualmente, se han propuesto varios modelos para tratar y reducir las comunidades perturbadas mediante algoritmos genéticos y evolutivos [76].

Por otra parte, de la Ec. 2.6, se sigue que la frustración es una función que cuantifica la calidad de las particiones de las comunidades signadas. En la Figura 4.7 (A) se muestran los valores de la frustración para cada una de escuelas. Observamos, que las escuelas primarias los valores de la frustración son altos para ambos métodos, esto es debido mayormente al número de enlaces negativos en sus comunidades.

Para el resto de las escuelas, el valor de la frustración son menores. Las comunidades obtenidas por el método greedy tienen un mayor porcentaje (w_c^-) de enlaces negativos (ver Tabla 4.2), pero tienen menos frustración; esto se debe a que el método greedy optimiza mejor el número de relaciones positivas entre comunidades. En otras palabras, el método greedy genera comunidades menos amistosas entre ellas, sin embargo, tienen una mayor perturbación de relaciones de enemistad (socialmente más inestables). Con el método Newman, sucede todo lo contrario. La diferencia en los valores de la frustración del método Newman y greedy es casi constante (con excepción de la eprCCP). Esto se debe, a que en principio, los estudiantes tienen desde el inicio un número inicial de relaciones positivas y negativas, por lo tanto, las relaciones serán las mismas, independientemente de la partición óptima de las comunidades. Este fenómeno puede indicar que existe un factor de escala relacionando la frustración con el número de comunidades en que se divide la escuela, o más concretamente con su tamaño medio.

El Balance Social Comunitario (BSC) se muestra en la Figura 4.7 (B), en ella indica la proporción de tríadas balanceadas entre las tríadas totales dentro de las comunidades (ec. 2.5). En las escuelas primarias, el valor de BSC es menor, esto indica comunidades con mayor tensión social. En el resto de las escuelas, las comunidades alcanzan valores de BSC mayores 0.9, lo que representa comunidades sin o con mínima tensión social.

También, puede observarse que los valores BSC para ambos métodos son similares (excepto para la eprHUN), lo que indica que el número de tríadas balanceadas en las comunidades es casi independiente a su partición óptima. Este fenómeno es importante,

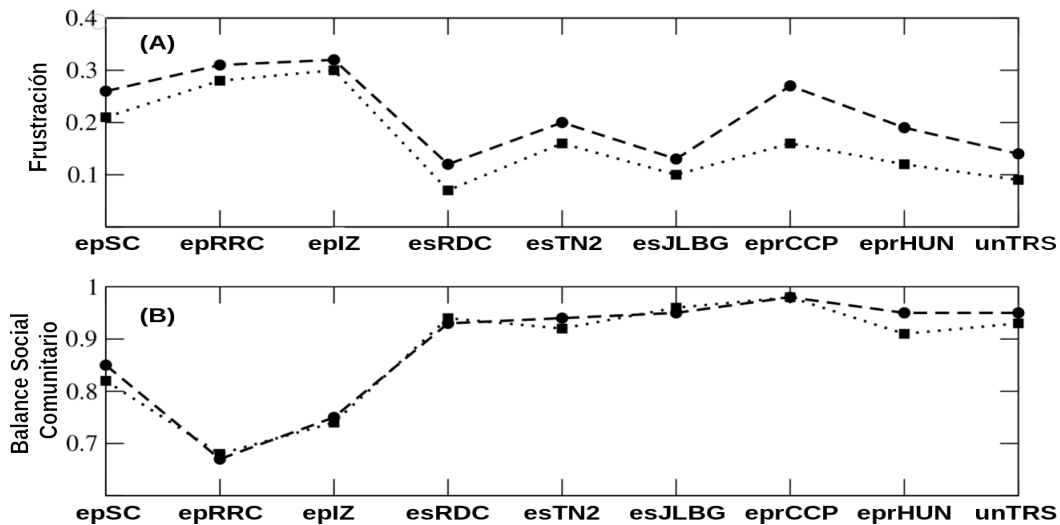


Figura 4.7: BSC y frustración. (A) Valores de la frustración para los métodos de Newman (círculos) y greedy (cuadros). (B) Valores del BSC (Balance Social Comunitario) para los métodos de Newman (círculos) y greedy (cuadros). Cada valor del greedy fue obtenido de 300 simulaciones.

debido a que uno espera que el valor del BSC varia en función a particiones diferentes, pero no es así. Esto se debe a que las redes signadas tienen un factor de balance social intrínseco que creemos afecta al valor de BSC. En otras palabras, el número de tríadas balanceadas en la red signada afecta significativamente el valor de BSC de cualquier partición óptima (en el Apéndice A.2 se adjunta una gráfica del balance de todas las redes signadas). En conclusión, el factor de balance social en sistemas como los estudiados en esta tesis, aporta información importante para saber sobre que tan balanceadas están sus comunidades.

Todos los resultados anteriores indican que las comunidades en las primarias se forman con criterios débiles (o incluso aleatorios). A partir de las secundarias empiezan a definirse mejor y recuperan cierto grado de estabilidad. En particular, en las escuelas primarias, las comunidades del sistema real podrían ser diferentes a las comunidades (inestables) generadas por los algoritmos que solamente tienen en cuenta los enlaces positivos. Por el contrario, a las escuelas secundarias, las comunidades (semiestables) se van acercando a las comunidades del sistema real, mientras que, en las de educación media superior y superior, las comunidades (estables) son una muy buena aproximación a las comunidades reales de los estudiantes.

Capítulo 5

Discusión

En esta sección se analizan y se discuten los resultados obtenidos de la sección anterior, así como algunas consideraciones que surgieron durante el desarrollo del trabajo.

Inicialmente, en los cálculos de los resultados, consideramos que el tamaño más pequeño de las comunidades es de 2 nodos. Sin embargo, esta consideración nos generó una incertidumbre porque en algunas redes hay demasiadas comunidades de 2 nodos (ver Figura 4.4) y es imposible formar tríadas. Dichas comunidades podrían sesgar nuestros resultados. Para descartar un posible sesgo, probamos eliminar las comunidades de 2 nodos y los resultados obtenidos fueron cualitativamente los mismos.

Por otro lado, para entender mejor las relaciones de amistad y enemistad, en las encuestas aplicadas se preguntó a los estudiantes acerca del significado de estos conceptos. La mayoría de las respuestas mostraron que ambos conceptos tienen significados muy extensos que incluyen diferentes elementos de percepción. A pesar de esto, los adolescentes y adultos jóvenes reportaron una gran coherencia en la percepción de las relaciones. Autores como Whetnall [77] and Burk [78] reportaron los mismos hallazgos. La percepción mejor definida de ambos conceptos ayudan a que los alumnos elijan mejor a sus amigos, y a evitar a las personas que les desagradan, dando lugar a comunidades con mayor estabilidad.

Al principio de la investigación, se esperaba que el valor de PPC estuviera directamente relacionada con el número (porcentaje) de enlaces negativos en las redes, sin embargo, no es así. En las secundarias se observa que el porcentaje de enlaces negativos es alto (ver Tabla 3.5) pero en sus comunidades el porcentaje (w_c^-) de enlaces negativos (ver Tabla 4.2) es bajo. Tal resultado indica dos fenómenos en estas escuelas: El primero, indica gran selectividad en la formación de sus comunidades, y el segundo, una transición conductual desde de la educación básica a la educación superior. Estos parecen coincidir con los resultados de Labianca et al [79]. Además, Hallinan and Smith [80] indican que los adolescentes se vuelven más selectivos y conforme crecen

exigen condiciones similares (lealtad, confianza, solidaridad, etc) como requisito para iniciar una amistad. Esto último explica porque, aún cuando las primarias y secundarias tienen porcentajes de relaciones negativas similares en sus redes (entre 25 % y 31 %), los valores de PPC son diferentes. Por lo tanto, podemos concluir que un mayor número de enlaces negativos en la redes no necesariamente implica una mayor inestabilidad en sus comunidades, sino que también otros factores sociales están involucrados. Lo mismo sucede para el BSC y la frustración (ver Figura 4.7).

En la Figura 4.6, se observa que el valor de PPC, muestra una tendencia decreciente según sea el nivel académico: con excepción de la escuela primaria epSC, cuyo valor es más bajo en comparación con las otras primarias. La primaria epSC es una escuela rural con una particularidad: la mayoría de sus alumnos son familiares entre sí (es decir, son hermanos o primos). Generalmente, la relación de parentesco conlleva una relación de amistad (esta información se deriva de las encuestas aplicadas). Así, cuando los alumnos de la escuela epSC forman sus comunidades, integran a sus familiares a sus grupos causando que el número de tríadas balanceadas del tipo (+++) Figura 2.3 (a) aumente y el número de enlaces negativos sea menor. Como consecuencia, el valor del BSC de la escuela epSC es mayor a comparación de las otras primarias (ver Figura 4.7 (B)). Esta particularidad hace que la escuela epSC sea una excepción a la tendencia del valor de PPC en relación al nivel académico. Está claro que la epSC confirma que el número de relaciones negativas en las red no está asociado al valor de PPC.

Los conceptos de balance y estabilidad han sido estudiados desde tiempo atrás, y son importantes en el campo de las redes signadas. Los autores Sun *et. al.* definen que una comunidad es estable si y sólo si todas sus tríadas son balanceadas [22]. En nuestro caso, definimos estabilidad de una comunidad en relación a los enlaces negativos que contiene (independientemente de si los enlaces negativos formen parte de tríadas balanceadas o no).

La Figura 4.7 (B) muestra que las escuelas primarias (comunidades inestables) son las más desbalanceadas. En cambio las escuelas secundarias (comunidades semi-estables) tienen los valores de BSC cercanas a las escuelas con comunidades estables. Esto significa que gran parte de los enlaces negativos en las comunidades semi-estables forman parte de tríadas balanceadas.

En la sección 3.4 se explicó que la heurística utilizada sólo añade enlaces negativos faltantes que tienen una alta probabilidad de existir en el sistema real. Es importante mencionar que la heurística provoca un sesgo al incluir estos enlaces, debido a que se agregan más tríadas balanceadas que tríadas desbalanceadas; por esta razón, los valores de BSC podrían estar sobrestimados.

Con respecto a la sobreestimación del BSC, calculamos sus valores antes de introducir missing links y encontramos que todos estos valores son cercanos a su máximo valor 1. Este hecho se explica por la ausencia de enlaces negativos, dando lugar a que la mayoría de las tríadas balanceadas, están formadas por tres enlaces positivos (ver Figura 2.3 a)).

Dicho de otra manera, al añadir los missing links, los valores BSC son más bajos. Aún así, los valores BSC podrían estar sobrestimados con respecto al valor del sistema real y creemos que probablemente este valor sea entre 0.85 y 0.90 tomando como referencia los valores de BSC de las escuelas de nivel básico (ver Figura 4.7 (B)).

Con respecto a los algoritmos de comunidades, elegimos dos métodos diferentes. Aun cuando el método de Newman entrega más comunidades [81,82] y el método de Louvain comunidades más grandes [36], los resultados obtenidos son cualitativamente y cuantitativamente similares. Como se ha mencionado, los métodos son algorítmicamente diferentes; no obstante, comparten un mismo objetivo que es agrupar nodos densamente conectados. De esta manera, las comunidades obtenidas con los métodos distintos tienen propiedades similares. Adicionalmente, durante el análisis de los cálculos, comparamos los resultados utilizando un tercer algoritmo : el método de Girvan-Newman (ver sección 1.3.3.3), y obtuvimos resultados similares. A partir de esto, es posible conjeturar que independientemente del algoritmo de comunidades utilizado obtendríamos resultados y conclusiones similares.

Ahora, en relación al desempeño de los algoritmos, se ha demostrado que el algoritmo de Louvain es muy rápido para proporcionar particiones óptimas de redes grandes (el orden de complejidad es de $O(n \cdot \log(n))$) [83]. Las redes estudiadas en esta tesis son relativamente pequeñas, por tal motivo la diferencia en los tiempos de ejecución de ambos algoritmos fue mínima. Por otra parte, el método de Louvain produce comunidades con valores de PPC ligeramente mayores, pero las particiones tienen valores de frustración levemente menores. Por último, los valores de BSC son casi iguales para ambos métodos. Por estas razones, en este estudio no hay suficientes argumentos para decidir cual de los dos métodos se ha desempeñado mejor. Debido a que el método de Louvain es estocástico, en el Apéndice A.3, se adjunta como complemento a nuestros cálculos graficas de caja (boxplot) para dicho método, se observa poca variabilidad debida a su calibración.

En la literatura, se especifica que la mayoría de los enlaces negativos en las comunidades signadas, por definición, deben encontrarse afuera conectando a las comunidades y con muy pocos enlaces negativos dentro de ellas. En otras palabras, las comunidades deben ser estables. En este trabajo encontramos comunidades que sólo involucraban enlaces positivos, y obtuvimos comunidades en las que se alejaban de la definición de comunidad signada. Creemos que estas comunidades no representan las comunidades del sistema real. Tampoco, tenemos la intención de decir que los algoritmos de detección de comunidades son erróneos. De hecho, estos métodos han sido validados en varios trabajos que analizaban redes reales (la más conocida es la del Club de Karate de Zachary, [46]), como sintéticas, y han mostrado un nivel alto de confiabilidad; sin embargo, hay estudios que han mostrado , que es posible encontrar redes aleatorias (distribución de Poisson) que presentan valores altos de modularidad [84], a pesar que no tiene estructura de comunidad. Considerando, lo anterior, es posible que los valores altos de modularidad Q^+ en las comunidades de las escuelas primarias (y en algunas

secundarias), sean consecuencia de las relaciones aleatorias. De aquí, surge la pregunta, si sólo tenemos en cuenta las relaciones positivas ¿Cómo sabemos que las comunidades encontradas son una representación confiable del sistema? Conjeturamos que si los estudiantes tienen los conceptos de amistad y enemistad bien definidos, obtendríamos comunidades más estables, y éstas serán una representación confiable a las comunidades del sistema real.

Durante el análisis de las comunidades se encontraron nodos negativos y subcomunidades antagónicas dentro de las comunidades.

Como un plus y aunque esto rebasa nuestro campo de estudio, nuestra metodología posiblemente este detectando casos de acoso escolar. La presencia de nodos negativos en las comunidades se podrían interpretar como un individuo que acosa a un grupo de alumnos, o como un grupo de alumnos que integra mal intencionadamente a un individuo para acosarlo [85]. Mientras que las subcomunidades antagónicas se interpretarían como subgrupos de alumnos que se acosan entre sí. Esto podría servir para brindar ayuda a las víctimas potenciales, reducir la probabilidad de que una persona sufra acoso y evitar posibles enfrentamientos entre grupos de personas, en este caso en el ámbito escolar. Esto es un asunto que actualmente estamos investigando.

Capítulo 6

Conclusiones y trabajos a futuro

En esta tesis realizamos un estudio detallado de comunidades en redes signadas. Las actividades de investigación realizadas, abarcan desde trabajo de campo hasta el desarrollo de software científico para el procesamiento de datos, análisis y visualización de redes. Abordamos varias cuestiones que surgieron a lo largo del desarrollo del trabajo. Podemos resumir nuestras contribuciones de la siguiente manera:

Se realizaron encuestas sobre las relaciones de amistad y enemistad a un total 4328 estudiantes de 9 escuelas de diferentes niveles educativos y se construyeron sus matrices de adyacencia signadas. En la sección 3.4 introdujimos un método para completar los missing links. El método se basa en conceptos de la teoría del equilibrio social para determinar que enlaces negativos añadir.

Extendimos el concepto de la proporción de Dunbar, desarrollado para las relaciones positivas, a las relaciones negativas. Se encontró que para ambos tipos de relaciones las proporciones son cercanas a 3. No fue hasta que se agregaron los missing links negativos que las proporciones se acercaron a su valor esperado. Comprendemos que tal extensión no es trivial, pero es interesante que en las escuelas primarias este valor de proporción tiende al mismo valor de las amistades, y que suceda lo mismo para las otras escuelas después de agregar los missing links. También, creemos que si otros estudios no han encontrado esta proporción en las relaciones negativas, es precisamente porque no las tienen completas. Por lo anterior, es necesario desarrollar un análisis más profundo y detallado para confirmar este resultado.

En el Cap. 3 desarrollamos una metodología para obtener comunidades signadas y cuantificar la perturbación causado por los enlaces negativos en las particiones óptimas de comunidades. Esta cuantificación se obtuvo de los valores de PPC. Cuando se analizó este valor se observó que no hay una correlación proporcional entre el número de enlaces y la inestabilidad de las comunidades. Además, las medidas de BSC y la frustración ayudaron a identificar los distintos niveles de estabilidad en las comunida-

des de las escuelas. Se encontró que en las escuelas primarias las comunidades son más inestables, en las secundarias son semiestables, y en tanto en las preparatorias como en la universidad son estables.

Aunque nuestra metodología obtiene buenos resultados en aproximarse a las comunidades reales, se recomienda el uso de otros algoritmos para optimizar las comunidades signadas. Una opción, es usar algoritmos de machine learning [86] y creemos que pueden ayudar a encontrar una forma para maximizar la estabilidad de las comunidades sin afectar la modularidad. Otra opción es desarrollar algoritmos que usen directamente la Ec. 2.3 para particionar la red en comunidades signadas, y establecer medidas para evaluar la calidad de ellas y las obtenidas en este trabajo [71, 87]. Una última opción es probar nuestra metodología con métodos de inferencia de estadística (Modelos de Bloques estocásticos) para encontrar las particiones óptimas que solo contengan enlaces positivos [88]. Estas opciones nos parecen interesantes y servirán para ampliar y complementar nuestros resultados y las conclusiones expuestas en esta tesis. Muy probablemente abordaremos estos problemas en un futuro.

Con estos hallazgos, concluimos que el nivel académico (edad) está relacionado con los puntos de vistas que tienen los estudiantes acerca de los conceptos de amistad y enemistad, y esto influye en la estabilidad de las comunidades.

Para terminar este escrito, nuestra metodología puede servir de guía a futuras investigaciones en redes sociales que involucren estudiantes. Así como, obtener una primera caracterización de sus comunidades facilitando la comprensión de su organización y que tomen en cuenta los enlaces negativos.

Referencias

- [1] Gezhi W, Upinder SB, Ravi I. Complexity in Biological Signaling Systems. *Science*. 1999;284(5411):92–96.
- [2] Hmelo-Silver CE, Pfeffer MG. Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*. 2004;28(1):127–138.
- [3] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. *Physics Reports*. 2006;424(4):175–308.
- [4] Barabási AL, Pósfai M. *Network Science*. Cambridge; 2016.
- [5] McBride N. Actor-network theory and the adoption of mobile communications. *Geography*. 2003;88(4):266–276.
- [6] Souma W, Fujiwara Y, Aoyama H. Complex networks and economics. *Physica A: Statistical Mechanics and its Applications*. 2003;324(1):396–401.
- [7] Vega Redondo F. *Complex Social Networks*. Cambridge; 2007.
- [8] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*. 2009;20(1):61–80.
- [9] Wuchty S. Scale-Free Behavior in Protein Domain Networks. *Molecular Biology and Evolution*. 2001;18(9):1694–1702.
- [10] Euler L. *Solutio problematis ad geometriam situs pertinentis*. *Commentarii academiae scientiarum Petropolitanae*. 1741; p. 128–140.
- [11] Shields R. Cultural Topology: The Seven Bridges of Königsburg, 1736. *Theory, Culture & Society*. 2012;29(4-5):43–57.
- [12] Mata ASd. Complex networks: A mini-review. *Braz J Phys*. 2020;50(5):658–672.
- [13] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998;393(6684):440–442.

- [14] Hui D. Food Web: Concept and Applications. *Nature Education Knowledge*. 2012;3(12).
- [15] He Z, Tajeuna EG, Wang S, Bouguessa M. A Comparative Study of Different Approaches for Tracking Communities in Evolving Social Networks. In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA); 2017. p. 89–98. Available from: <https://doi.org/10.1109/DSAA.2017.62>.
- [16] Robins G, Pattison P, Wang P. Closure, connectivity and degree distributions: Exponential random graph (p^*) models for directed social networks. *Social Networks*. 2009;31(2):105–117.
- [17] Nguyen LTH, Wada T, Masubuchi I, Asai T, Fujisaki Y. Opinion formation over signed gossip networks. *SICE J Control Meas Syst Integr*. 2017;10(3):266–273.
- [18] Fortunato S, Hric D. Community detection in networks: A user guide. *Phys Rep*. 2016;659:1–44.
- [19] Fortunato S. Community detection in graphs. *Phys Rep*. 2010;486(3-5):75–174.
- [20] Newman MEJ. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*. 2006;103(23):8577–8582.
- [21] Nguyen NP, Alim MA, Dinh TN, Thai MT. A method to detect communities with stability in social networks. *Soc Netw Anal Min*. 2014;4(1).
- [22] Sun R, Chen C, Wang X, Zhang Y, Wang X. Stable Community Detection in Signed Social Networks. *IEEE Transactions on Knowledge and Data Engineering*. 2022;34(10):5051–5055.
- [23] Fritz H. Attitudes and Cognitive Organization. *The Journal of Psychology*. 1946;21(1):107–112.
- [24] Cartwright D, Harary F. Structural balance: a generalization of Heider’s theory. *Psychol Rev*. 1956;63(5):277–293.
- [25] James AD. Clustering and Structural Balance in Graphs. *Human Relations*. 1967;20(2):181–187.
- [26] Mankirat K, Sarbjeet S. Analyzing negative ties in social networks: A survey. *Egypt Inform J*. 2016;17(1):21–43.
- [27] Zafarani R, Abbasi MA, Liu H. *Social Media Mining: An Introduction*. Cambridge University Press; 2014. Available from: <https://doi.org/10.1017/CB09781139088510>.

- [28] Asim Y, Malik AK, Raza B, Shahid AR. A trust model for analysis of trust, influence and their relationship in social network communities. *Telematics and Informatics*. 2019;36:94–116. doi:<https://doi.org/10.1016/j.tele.2018.11.008>.
- [29] Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton JF, et al. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE*. 2011;6(8):e23176.
- [30] Liu M, Chen X. Friendship networks and social, school and psychological adjustment in Chinese junior high school students. *Psychol Schs*. 2003;40:5–17.
- [31] Brito-Montes J, Canto-Lugo E, Huerta-Quintanilla R. Modularity, balance, and frustration in student social networks: The role of negative relationships in communities. *PLOS ONE*. 2022;17(12):1–20.
- [32] Morrison G, Mahadevan L. Discovering communities through friendship. *PLoS ONE*. 2012;7(7):e38704.
- [33] Zhao Z, Feng S, Wang Q, Huang JZ, Williams GJ, Fan J. Topic oriented community detection through social objects and link analysis in social networks. *Knowl Based Syst*. 2012;26:164–173.
- [34] Murata T. In: Furht B, editor. *Detecting Communities in Social Networks*. Boston, MA: Springer US; 2010. p. 269–280. Available from: https://doi.org/10.1007/978-1-4419-7142-5_12.
- [35] Xu Y. A spectral method to detect community structure based on the communicability modularity. *Physica A*. 2020;537:122751.
- [36] Fortunato S, Barthélemy M. Resolution limit in community detection. *Proc Natl Acad Sci*. 2007;104(1):36–41.
- [37] Reichardt J, Alaminó R, Saad D. The Interplay between Microscopic and Mesoscopic Structures in Complex Networks. *PLOS ONE*. 2011;6(8):1–8.
- [38] Newman M. The physics of networks. *Physics Today*. 2008;61(11):33–38.
- [39] Reichardt J, Bornholdt S. Statistical mechanics of community detection. *Phys Rev E*. 2006;74:016110.
- [40] Burden RL, Faires JD. 6. In: *Análisis numérico*. 7th ed. International Thomson Editores; 2002. p. 404.
- [41] Eaton E, Mansbach R. A Spin-Glass Model for Semi-Supervised Community Detection. *Proc Conf AAAI Artif Intell*. 2021;26(1):900–906.

- [42] Hernández-Hernández AM, Viga-de Alva D, Huerta-Quintanilla R, Canto-Lugo E, Laviada-Molina H, Molina-Segui F. Friendship Concept and Community Network Structure among Elementary School and University Students. *PLOS ONE*. 2016;11(10):1–17.
- [43] D Blondel V, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;2008(10).
- [44] Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E*. 2004;69:026113.
- [45] Guimerà R, Danon L, Díaz-Guilera A, Giralt F, Arenas A. Self-similar community structure in a network of human interactions. *Phys Rev E*. 2003;68:065103.
- [46] Zachary WW. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*. 1977;33(4):452–473.
- [47] Newman MEJ. Fast algorithm for detecting community structure in networks. *Phys Rev E*. 2004;69:066133.
- [48] Duch J, Arenas A. Community detection in complex networks using extremal optimization. *Phys Rev E*. 2005;72:027104.
- [49] Nicosia V, Mangioni G, Carchiolo V, Malgeri M. Extending the definition of modularity to directed graphs with overlapping communities. *J Stat Mech*. 2009;2009(03):P03024.
- [50] Lancichinetti A, Kivela M, Saramaki J, Fortunato S. Characterizing the Community Structure of Complex Networks. *PLOS ONE*. 2010;5:1–8.
- [51] Prell C. *Social Network Analysis : History, Theory and Methodology*. Sage; 2011.
- [52] Markham AN. Fieldwork in social media: What would Malinowski do? *Qual Commun Res*. 2013;2(4):434–446.
- [53] Dunbar RIM. Neocortex size as a constraint on group size in primates. *J Hum Evol*. 1992;22(6):469–493.
- [54] Sutcliffe A, Dunbar R, Binder J, Arrow H. Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British Journal of Psychology*. 2012;103(2):149–168.
- [55] Dunbar RIM, Arnaboldi V, Conti M, Passarella A. The structure of online social networks mirrors those in the offline world. *Social Networks*. 2015;43:39–47.
- [56] Tamarit I, Sánchez A, Cuesta JA. Beyond Dunbar circles: a continuous description of social relationships and resource allocation. *Scientific Reports*. 2022;12(1):2287.

- [57] Dunbar RIM. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*. 1993;16(4):681–694.
- [58] Zhou WX, Sornette D, Hill RA, Dunbar RIM. Discrete hierarchical organization of social group sizes. *Proceedings Biological Sciences*. 2005;272(1561):439–444.
- [59] Hill RA, Dunbar RIM. Social network size in humans. *Hum Nat*. 2003;14(1):53–72.
- [60] Kardos P, Leidner B, Pléh C, Soltész P, Unoka Z. Empathic people have more friends: Empathic abilities predict social network size and position in social network predicts empathic efforts. *Social Networks*. 2017;50:1–5.
- [61] G B Roberts S, Dunbar RIM, V Pollet T, Kuppens T. Exploring variation in active network size: Constraints and ego characteristics. *Social Networks*. 2009;31(2):138–146.
- [62] Wani MA, Sankar R, Raghavi R, Chinmaya B. Aggression among Annamalai university students. *Global Journal of Intellectual & Developmental Disabilities*. 2017;1(3).
- [63] Çoban O, Yildirim M. The Comparison of High School Students' Level of Aggression Based on Demographic Features. *Asian Journal of Education and Training*. 2018;4(4):363–370.
- [64] Ballard ME, Rattley KT, Fleming WC, Kidder-Ashley P. School Aggression and Dispositional Aggression among Middle School Boys. *RMLE Online*. 2004;27(1):1–11.
- [65] Falk A, Fischbacher U. A theory of reciprocity. *SSRN Electron J*. 2001;.
- [66] Tang J, Chang Y, Aggarwal C, Liu H. A Survey of Signed Network Mining in Social Media. *ACM Comput Surv*. 2016;49(3).
- [67] Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*. 2010;107(31):13636–13641.
- [68] Esmailian P, Jalili M. Community Detection in Signed Networks: the Role of Negative ties in Different Scales. *Scientific reports*. 2015;5:14339.
- [69] Gómez S, Jensen P, Arenas A. Analysis of community structure in networks of correlated data. *Phys Rev E*. 2009;80:016114.
- [70] Zheng X, Zeng D, Wang FY. Social balance in signed networks. *Inf Syst Front*. 2015;17(5):1077–1095.

- [71] Anchuri P, Magdon-Ismael M. Communities and Balance in Signed Networks: A Spectral Approach. In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 2012. p. 235–242. Available from: <https://doi.org/10.1109/ASONAM.2012.48>.
- [72] Songmuang P, Sirisup C, Suebsriwichai A. Missing Link Prediction Using Non-Overlapped Features and Multiple Sources of Social Networks. *Information*. 2021;12(5).
- [73] Escribano D, Doldán-Martelli V, Lapuente FJ, Cuesta JA, Sánchez A. Evolution of social relationships between first-year students at middle school: from cliques to circles. *Scientific Reports*. 2021;11(1):11694.
- [74] Lo D, Surian D, Prasetyo PK, Zhang K, Lim EP. Mining direct antagonistic communities in signed social networks. *Inf Process Manag*. 2013;49(4):773–791.
- [75] Lambiotte R, Delvenne JC, Barahona M. Laplacian dynamics and multiscale modular structure in networks; 2008. Available from: <https://arxiv.org/abs/0812.1770v3>.
- [76] Girdhar N, Haradwaj KK. Community Detection in Signed Social Networks Using Multiobjective Genetic Algorithm. *J Assoc Inf Sci Technol*. 2019;70(8).
- [77] Whetnall EM. *The Moral Judgment of the Child*. By Jean Piaget. (International Library of Psychology and Philosophy. London: Kegan Paul, Trench, Trübner & Co.1932. Pp. ix 417. Price 12s. 6d.). *Philosophy*. 1933;8(31):373–374.
- [78] Burk WJ, Laursen B. Adolescent perceptions of friendship and their associations with individual adjustment. *Int J Behav Dev*. 2005;29(2):156–164.
- [79] Labianca G, Brass DJ. Exploring the Social Ledger: Negative Relationships and Negative Asymmetry in Social Networks in Organizations. *Acad Manage Rev*. 2006;31(3):596–614.
- [80] Hallinan M, Smith SS. Classroom Characteristics and Student Friendship Cliques. *Social Forces*. 1989;67:898–919.
- [81] Orman GK, Labatut V, Cherifi H. Qualitative Comparison of Community Detection Algorithms. In: Cherifi H, Zain JM, El-Qawasmeh E, editors. *Digital Information and Communication Technology and Its Applications*. Springer Berlin Heidelberg; 2011. p. 265–279. Available from: https://doi.org/10.1007/978-3-642-22027-2_23.
- [82] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. *Phys Rev E*. 2008;78(4):046110.

- [83] Traag V. Faster unfolding of communities: Speeding up the Louvain algorithm. *Phys Rev E*. 2015;92:032801.
- [84] Guimerà R, Sales-Pardo M, Amaral LAN. Modularity from fluctuations in random graphs and complex networks. *Phys Rev E*. 2004;70:025101.
- [85] Wittek M, Kroneberg C, Lämmermann K. Who is fighting with whom? How ethnic origin shapes friendship, dislike, and physical violence relations in German secondary schools. *Social Networks*. 2020;60:34–47.
- [86] Hafez AI, Ghali NI, Hassanien AE, Fahmy AA. Genetic Algorithms for community detection in social networks. In: 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA). IEEE; 2012. p. 460–465. Available from: <https://doi.org/10.1109/ISDA.2012.6416582>.
- [87] Zhang Y, Liu Y, Ma X, Song J. Community detection in signed networks by relaxing modularity optimization with orthogonal and nonnegative constraints. *Neural Comput & Applic*. 2020;32(14):10645–10654.
- [88] Jiang JQ. Stochastic block model and exploratory analysis in signed networks. *Phys Rev E*. 2015;91(6):062805.

Apéndice A

Información de soporte

A.1. Cuestionarios: Universidad

Figura A.1: Capturas de la plataforma digital diseñada para aplicar las encuestas a los estudiantes de la universidad. Una plataforma digital similar fue diseñada para las secundarias, y preparatorias.




ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS



Por favor indique su número de usuario

Una vez introducido tu número de usuario en el espacio de color verde claro, por favor haz click en el botón siguiente.


Siguiente



ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

AVISO IMPORTANTE


Esto no es un examen. Estamos realizando un estudio sobre las relaciones de amistad en los niveles de básica primaria, secundaria, preparatoria y universitario. Nos interesa conocer tu opinión respecto a tus relaciones amigables y de tipo poco cordial. Si tienes alguna duda por favor preguntale al encuestador. Gracias por tu ayuda y colaboración. Por favor responde las siguientes preguntas de manera sincera. Ten en cuenta por favor que las relaciones de amistad deben ser de carácter personal (personas de tu institución educativa con las cuales realices actividades como jugar, comer, estudiar, etc), es decir NO amistades con las cuales tengas únicamente contacto través de redes sociales como Twitter, Facebook, Instagram, etc.



ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

I. DATOS PERSONALES

1. Por favor indica la fecha de hoy (día-mes-año Ejemplo 14-02-2018)
2. Por favor indica tu edad en años. Ejemplo: 8
3. Por favor indica tu sexo
4. Por favor indica tu carrera
5. Por favor indica el semestre que estas cursando
6. Cuántos hermanos y hermanas tienes?. (Indique el número. Ejemplo: 2)
7. Cuántos primos y primas tienes?. (Indique el número. Ejemplo 5)




ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

II. DATOS SOCIALES

Por favor lea las instrucciones y sigalas con cuidado

PREGUNTA 1: HAZ UNA LISTA DE PALABRAS (LAS QUE QUIERAS) QUE INDIQUEN PARA TI QUE ES UN AMIGO(A)

1. <input style="width: 100%;" type="text"/>	2. <input style="width: 100%;" type="text"/>	3. <input style="width: 100%;" type="text"/>
4. <input style="width: 100%;" type="text"/>	5. <input style="width: 100%;" type="text"/>	6. <input style="width: 100%;" type="text"/>
7. <input style="width: 100%;" type="text"/>	8. <input style="width: 100%;" type="text"/>	9. <input style="width: 100%;" type="text"/>




ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

II. DATOS SOCIALES

PREGUNTA 2: HAZ UNA LISTA DE PALABRAS (LAS QUE QUIERAS) QUE INDIQUEN PORQUÉ ALGUIEN **NO ES TU AMIGO(A)**

1. <input style="width: 100%;" type="text"/>	2. <input style="width: 100%;" type="text"/>	3. <input style="width: 100%;" type="text"/>
4. <input style="width: 100%;" type="text"/>	5. <input style="width: 100%;" type="text"/>	6. <input style="width: 100%;" type="text"/>
7. <input style="width: 100%;" type="text"/>	8. <input style="width: 100%;" type="text"/>	9. <input style="width: 100%;" type="text"/>



ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

II. DATOS SOCIALES


PREGUNTA 3: CUALES ESTUDIANTES DE LA UNIVERSIDAD SON TUS AMIGOS ? Selecciona en las siguientes listas desplegables la carrera, el semestre y el nombre de tus amigos(as) de la Universidad. Recuerda por favor que la definición de amigos(as) se refiere a aquellos con los que realizas más actividades cara a cara, como hablar, realizar actividades recreativas, practicar un deporte, salir a lugares de entrenamiento, etc.

Para agregar personas a su lista haz click en el botón "Agregar amigo(a)". No te elijas a ti mismo, es decir no te consideres a ti mismo como amigo(a). Si deseas borrar alguno de los nombres ingresados, elije en el menú de carreras la opción "-Borrar dato-" y en el menú de semestre elije la opción "0". Cuando termines de ingresar tu lista de amigos, por favor da click en el botón "Siguiente".

--Seleccione una carrera-- ▼

-semestre- ▼

Selecciona el nombre de tu amigo(a) ▼



ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

II. DATOS SOCIALES

PREGUNTA 4: CON CUALES ESTUDIANTES TIENES UNA RELACION POCO CORDIAL O UNA MALA RELACION ? . Si deseas borrar alguno de los nombres ingresados, elije en el menú de carreras la opción "-Borrar dato-" y en el menú de semestre elije la opción "0". Cuando termines de ingresar tu lista de amigos, por favor haz click en el botón "Siguiente".

--Seleccione una carrera-- ▼

-semestre- ▼

Selecciona el nombre de la persona ▼

ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

Cinvestav
UNIDAD MERIDA

II. DATOS SOCIALES

PREGUNTA 5: QUIENES DE LOS ESTUDIANTES DE LA UNIVERSIDAD SON TUS HERMANOS? Para agregar personas a tu lista haz click en el botón "Agregar hermano(a)". Si deseas borrar alguno de los nombres ingresados, elije en el menú de carreras la opción "-Borrar dato-" y en el menú de semestre elija la opción "0". Cuando termines de ingresar tu lista de amigos, por favor haz click en el botón "Siguiente".

--Seleccione una carrera-- -semestre- Selecciona el nombre de tu hermano(a)

Agregar hermano(a)

Siguiente

ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

Cinvestav
UNIDAD MERIDA

II. DATOS SOCIALES

PREGUNTA 6: QUIENES DE LOS ESTUDIANTES DE LA UNIVERSIDAD SON TUS PRIMOS? Para agregar personas a tu lista de click en el botón "Agregar amigo/a". Si deseas borrar alguno de los nombres ingresados, elije en el menú de carreras la opción "-Borrar dato-" y en el menú de semestre elije la opción "0". Cuando termines de ingresar tu lista de amigos, por favor de click en el botón "Siguiente".

--Seleccione una carrera-- -semestre- Selecciona el nombre de tu primo(a)

Agregar primo(a)

Siguiente



ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS


II. DATOS SOCIALES

PREGUNTA 7: CON CUALES ESTUDIANTES CONFORMAS GENERALMENTE GRUPOS DE TRABAJO, TE REUNES A ESTUDIAR PARA EXAMENES O REALIZAR TAREAS ? Para agregar personas a tu lista haz click en el botón "Agregar primo(a)". Si deseas borrar alguno de los nombres ingresados, elije en el menú de carreras la opción "-Borrar dato-" y en el menú de semestre elije la opción "0". Cuando termines de ingresar tu lista de amigos, por favor haz click en el botón "Siguiente".

--Seleccione una carrera--

-semestre-

Selecciona el nombre de tu primo(a)



ESTUDIO DE LA TOPOLOGÍA Y GENERACIÓN DE MODELOS DE DIFUSIÓN EN REDES SOCIALES EN ESCUELAS A DIFERENTES GRADOS ACADÉMICOS

Para finalizar la recolección de los datos que tu muy amablemente nos ha suministrado, te pedimos por último escribas tu nombre en el cuadro que aparece en la parte inferior de esta página. Esta, será tu autorización para que guardemos tus datos y los usemos única y exclusivamente para fines de investigación científica. Te pedimos también leas con atención y si estas de acuerdo, escribe tu nombre completo y haz click en el boton Guardar al final de esta página. De lo contrario, si no deseas que hagamos uso de tus datos, por favor haz click en el boton Declinar al final de la página.

Por este medio doy autorización al Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional CINESTAV-IPN Unidad Mérida para recopilar todos mis datos personales suministrados por mí en persona a través de este medio electrónico. Tengo conocimiento de que la institución lleva a cabo un proyecto de investigación científica cuyo objetivo es determinar las redes de interacción entre los estudiantes de mi escuela, para entender como se propaga la información en estas redes. Comprendo que mis datos son necesarios para fines estadísticos. Así mismo, **se me ha informado que el tratamiento de mis datos se hará de manera anónima** y serán resguardados conforme a lo estipulado por el Instituto Federal de Acceso a la Información y Protección de Datos. Doy mi autorización teniendo conocimiento del fin por el cual han sido recabados.

Nombre completo:

**¡ GRACIAS POR SU
COLABORACIÓN !**

A.2. Balance Social de las redes signadas

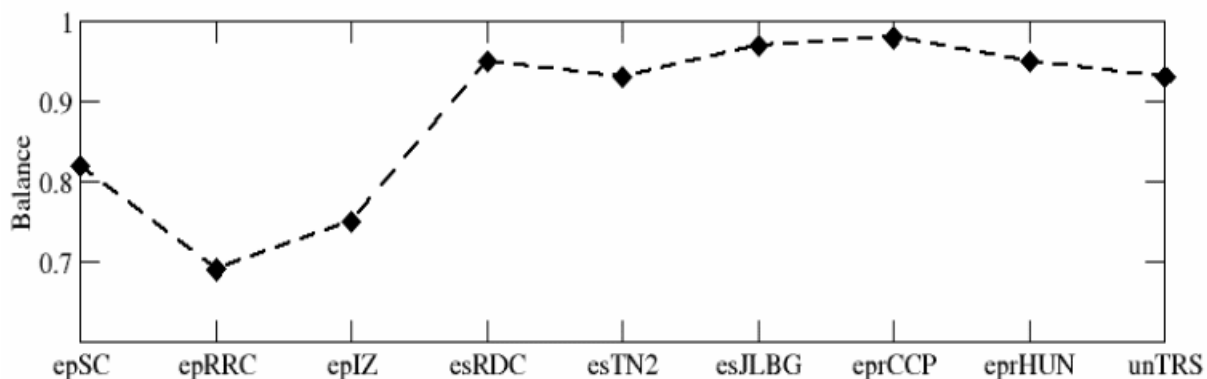


Figura A.2.1 Valores del balance social de las redes signadas. Cada valor fue obtenido de dividir el número de tríadas balanceadas en la red entre el número total de tríadas en ella.

A.3. Boxplot para el método de Louvain

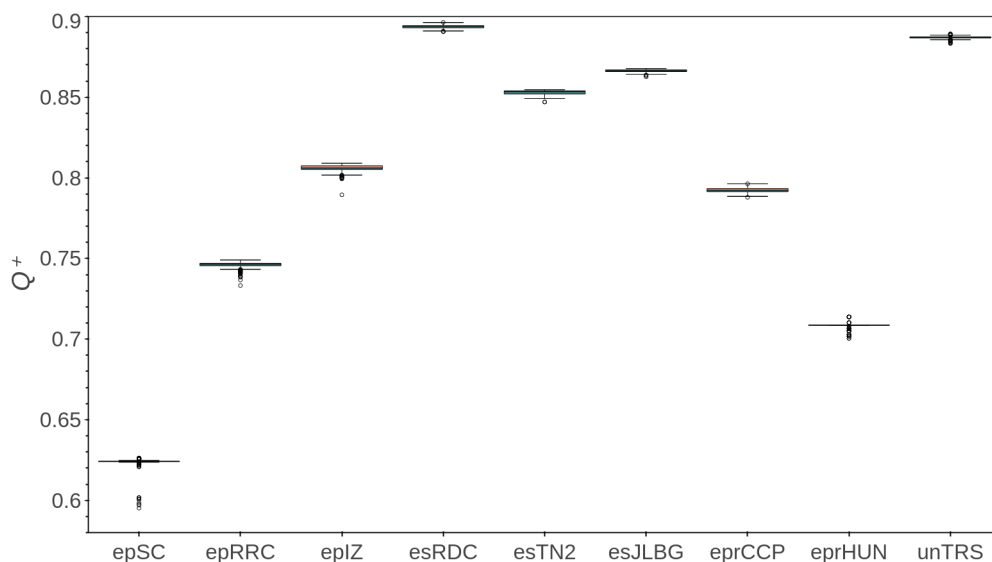


Figura A.3.1: Boxplot para los valores Q^+ (Método de Louvain). La línea central en los recuadros representa la mediana, los recuadros el primer y tercer cuartil, las líneas, el intervalo de confianza del 95 %, y los puntos, los valores atípicos. Cada boxplot se obtuvo de 300 simulaciones.

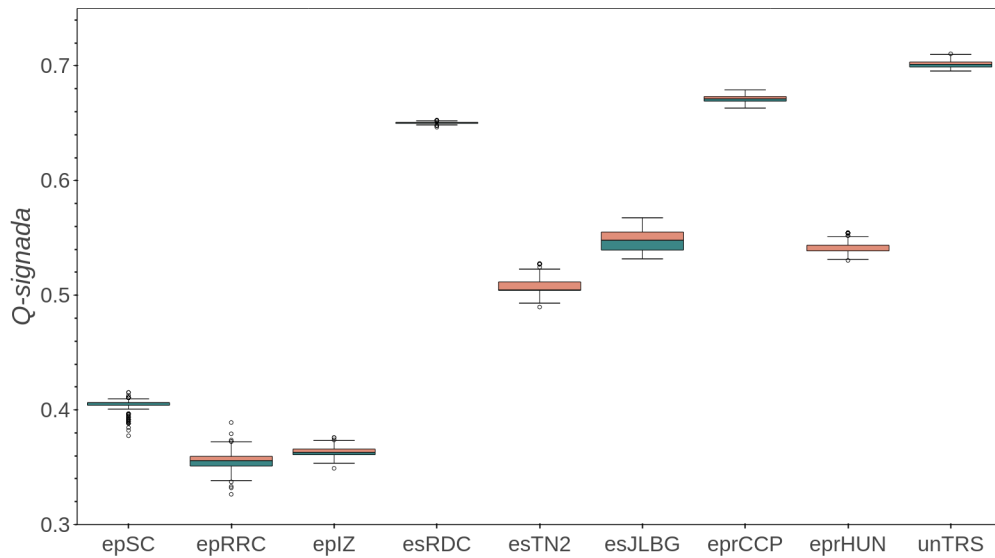


Figura A.3.2: Boxplot para los valores Q-signada (Q^s) (Método de Louvain). La línea central en los recuadros representa la mediana, los recuadros el primer y tercer cuartil, las líneas, el intervalo de confianza del 95 %, y los puntos, los valores atípicos. Cada boxplot se obtuvo de 300 simulaciones.

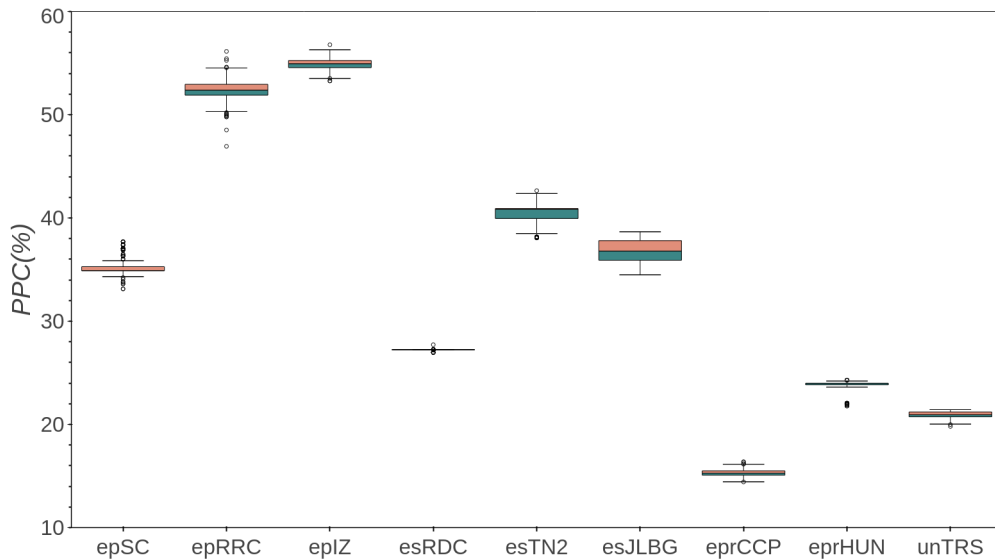


Figura A.3.3: Boxplot para los valores PPC (Método de Louvain). La línea central en los recuadros representa la mediana, los recuadros el primer y tercer cuartil, las líneas, el intervalo de confianza del 95 %, y los puntos, los valores atípicos. Cada Boxplot se obtuvo de 300 simulaciones.

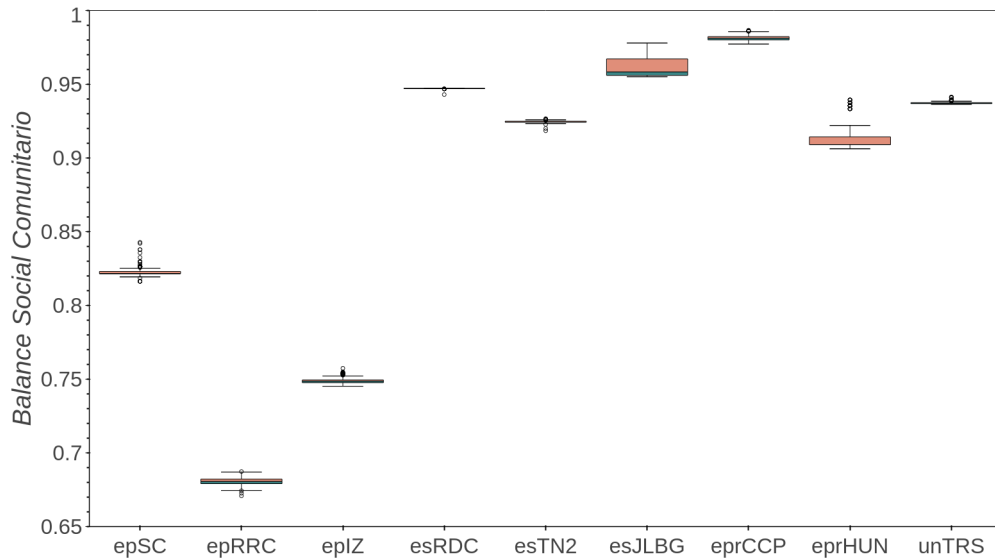


Figura A.3.4: Boxplot para los valores BSC (Método de Louvain). La línea central en los recuadros representa la mediana, los recuadros el primer y tercer cuartil, las líneas, el intervalo de confianza del 95 %, y los puntos, los valores atípicos. Cada Boxplot se obtuvo de 300 simulaciones.

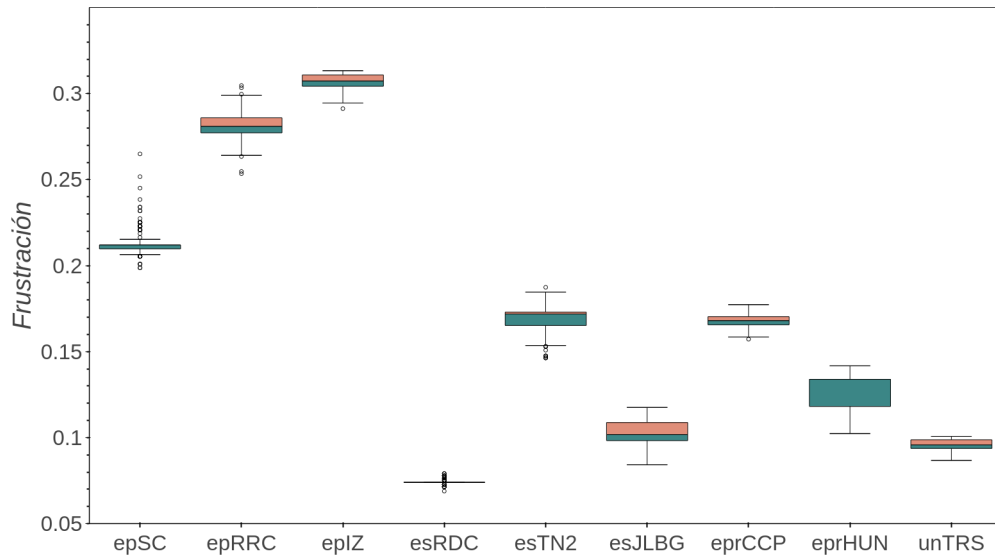


Figura A.3.5: Boxplot para los valores de la Frustración (Método de Louvain). La línea central en los recuadros representa la mediana, los recuadros el primer y tercer cuartil, las líneas, el intervalo de confianza del 95 %, y los puntos, los valores atípicos. Cada Boxplot se obtuvo de 300 simulaciones.

A.4. Programas Importantes

```
#!/usr/bin/python #-*- coding:utf-8 -*-
#####
#####
#Proyecto Conacyt: Estudio de la topología y generación de modelos de
difusión # en redes sociales en escuelas a
d#iferentes grados académicos
#Tesis Doctoral: Estudio de las comunidades en redes signadas:
Modularidad, # balance y frustración en redes
#académicas de Yucatán#
#Autores del código fuente: José Brito Montes, Efrain Canto Lugo
#Fecha: 18 Marzo 2019 #
#Descripción: # Este programa completa los probables enlaces mutuos
negativos perdidos en las redes signadas utilizando#
#la Teoría del Balance y la idea común de "El enemigo de enemigo es
mi amigo"#
#Importando
librerías#####
#####
import numpy as np
import networkx as nx
import matplotlib.pyplot as plt
#Generando
grafos#####
#####
Mf = np.loadtxt('hunucAT.txt', int)
Mnp = np.matrix(Mf)
N=len(Mnp)

GRAFOA=nx.Graph()
GRAFOA.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==1:
            GRAFOA.add_edge(i,j,weight=1)
        elif Mnp[i,j]==2:
            GRAFOA.add_edge(i,j,weight=-1)

GRAFOE=nx.Graph()
GRAFOE.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==2:
            GRAFOE.add_edge(i,j,weight=-1)

edges=list(GRAFOA.edges())
edges_neg=[]
#Agregando atributos a los
enlaces#####
#####
relacion=nx.get_edge_attributes(GRAFOA,'weight')
I=list(nx.isolates(GRAFOA))
GRAFOA.remove_nodes_from(I)
#Lista de enlaces
negativos#####
#####
```

```

for edge in edges:
    edge=sorted(list(edge))
    edge=tuple(edge)
    if relacion[edge]==-1:
        edges_neg.append(edge)
#Triangulos en la
red*****
*****
triangles_total=0
triangulos_balanceados=[]
triangulos_no_balanceados=[]
triangle_balanced=0
triangle_unbalanced=0
enlaces_mutuos_negativos=[]
triangles = [c for c in nx.enumerate_all_cliques(GRAFOA) if
len(c)==3]
for triangle in triangles:
    triangulo=GRAFOA.subgraph(triangle)
    edge_tri=[tuple(sorted(c)) for c in triangulo.edges()]
    balance=1
    for edget in edge_tri:
        balance=balance*relacion[edget]
    if balance==1:
        triangulos_balanceados.append(triangle)
        triangle_balanced+=1
    else:
        triangulos_no_balanceados.append(triangle)
        triangle_unbalanced+=1
print 'Numero de Triangulos', len(triangles)
print 'Numero de Triangulos balanceados', len(triangulos_balanceados)
print 'Numero de Triangulos no
balanceados', len(triangulos_no_balanceados)

for edge_negative in edges_neg:
    cont_balanced=0
    cont_unbalanced=0
    for triangle_B in triangulos_balanceados:
        trianguloAux=GRAFOA.subgraph(triangle_B)
        edges_trianguloAux=[tuple(sorted(c)) for c in
trianguloAux.edges()]
        if edge_negative in edges_trianguloAux:
            cont_balanced+=1
    for triangle_NB in triangulos_no_balanceados:
        trianguloAuxb=GRAFOA.subgraph(triangle_NB)
        edges_trianguloAuxb=[tuple(sorted(c)) for c in
trianguloAuxb.edges()]
        if edge_negative in edges_trianguloAuxb:
            cont_unbalanced+=1
    if cont_balanced!=0 and cont_balanced >= cont_unbalanced:
        enlaces_mutuos_negativos.append(edge_negative)
for c in enlaces_mutuos_negativos:
    edges_neg.remove(c)
enlaces_removidos=edges_neg
print 'Numero de enlaces negativos en la red', GRAFOE.size()
print 'Probable numero de enlaces mutuos negativos en la

```

```
red', len(enlaces_mutuos_negativos)
print 'Enlaces removidos', len(enlaces_removidos)
GRAFOA.remove_edges_from(enlaces_removidos)
A = nx.adjacency_matrix(GRAFOA)
A=A.todense()
#Imprimiendo
datos*****
*****
salida=open('hunucDT.txt','w')
for i in range (GRAFOE.order()):
    for j in range(GRAFOE.order()):
        salida.write(str(A[i,j])+ " ")
        salida.write('\n')
salida.close()
#*****
*****
```

```

#!/usr/bin/python
#-*- coding:utf-8 -*-
#*****
#Proyecto Conacyt: Estudio de la topología y generación de modelos de difusión
# en redes sociales en escuelas a diferentes grados académicos
#
#Tesis Doctoral: Estudio de las comunidades en redes signadas: Modularidad,
# balance y frustración en redes académicas de Yucatán
#
#Autores del código fuente: José Brito Montes, Efrain Canto Lugo
#Fecha: 1 dic 2019
#
#Descripción:
# Este programa calcula las propiedades generales de las signed networks así
# como las componentes de cada una de ellas

# Importar librerías de python *****
import matplotlib.pyplot as plt
import pylab as pl
import numpy as np
import networkx as nx

# Generar grafos *****
Mf = np.loadtxt('utrs.txt', int)
Mnp = np.matrix(Mf)
N=len(Mnp)

GRAFO=nx.Graph()
GRAFO.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==1:
            GRAFO.add_edge(i,j,color='b',weight=1)
        elif Mnp[i,j]==2:
            GRAFO.add_edge(i,j,color='r',weight=-1)

GPOS=nx.Graph()
GPOS.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==1:
            GPOS.add_edge(i,j,color='b',weight=1)

GNEG=nx.Graph()
GNEG.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==2:
            GNEG.add_edge(i,j,color='r',weight=-1)

#Asignar Etiquetas a los nodos *****
etiq = {}
N=GRAFO.order()
for i in range(N):
    etiq[i]=GRAFO.nodes()[i]

#Calcular propiedades *****
print '-----'
print 'Nodos', GRAFO.order()
print 'Nodos+', GPOS.order()
print 'Nodos-', GNEG.order()
print

```

```

print 'Enlaces', GRAFO.size()
print 'Enlaces+', GPOS.size()
print 'Enlaces-', GNEG.size()
print

d=GRAFO.degree()
res=float(sum(d.values())/len(d))
print 'Grado medio', round(res,5)
d=GPOS.degree()
res=float(sum(d.values())/len(d))
print 'Grado medio+', round(res,5)
d=GNEG.degree()
res=float(sum(d.values())/len(d))
print 'Grado medio-', round(res,5)
print

print 'Densidad', round(nx.density(GRAFO),5)
print 'Densidad+', round(nx.density(GPOS),5)
print 'Densidad-', round(nx.density(GNEG),5)
print

print 'Clustering', round(nx.average_clustering(GRAFO),5)
print 'Clustering+', round(nx.average_clustering(GPOS),5)
print 'Clustering-', round(nx.average_clustering(GNEG),5)
print

print 'Red conectada?', nx.is_connected(GRAFO)
print 'Red conectada?+', nx.is_connected(GPOS)
print 'Red conectada?-', nx.is_connected(GNEG)
print

print 'Componentes', nx.number_connected_components(GRAFO)
print 'Componentes+', nx.number_connected_components(GPOS)
print 'Componentes-', nx.number_connected_components(GNEG)
print

print 'Nodos aislados', len(nx.isolates(GRAFO))
print 'Nodos aislados+', len(nx.isolates(GPOS))
print 'Nodos aislados-', len(nx.isolates(GNEG))
print

GPOS.remove_nodos_from(list(nx.isolates(GPOS)))
print 'Componentes+ sin aislados', nx.number_connected_components(GPOS)
listComp=list(nx.connected_components(GPOS))

for comp in listComp:
    comp=list(comp)
    if isinstance(comp, list):
        H=GPOS.subgraph(comp)
        print 'Nodos en componente', H.order()
        print 'Enlaces en componente', H.size()

    etiqc={}
    for i in range(H.order()):
        etiqc[H.nodes()[i]]=H.nodes()[i]
    pos=nx.spring_layout(H)
    nx.draw_networkx_nodes(H, pos, node_color='red', node_size=300)
    nx.draw_networkx_labels(H, pos, labels=etiqc, font_size=10, font_color='black')
    nx.draw_networkx_edges(H, pos, width=0.5)
    plt.draw()
    plt.show()

# Generar el grafo de la red *****

```

```
colors = [GRAFO[u][v]['color'] for u,v in GRAFO.edges()]
pos = nx.spring_layout(GRAFO)
nx.draw_networkx_nodes(GRAFO, pos, node_color = '#000000', node_size=50)
#nx.draw_networkx_labels(GRAFO, pos, labels=etiq, font_size=10.0)
nx.draw_networkx_edges(GRAFO, pos, edge_color=colors,arrows=False, width=0.25)
plt.draw()
plt.show()
```



```

#!/usr/bin/python
#-*- coding:utf-8 -*-
#*****
#Proyecto Conacyt: Estudio de la topología y generación de modelos de difusión
# en redes sociales en escuelas a diferentes grados académicos
#
#Tesis Doctoral: Estudio de las comunidades en redes signadas: Modularidad,
# balance y frustración en redes académicas de Yucatán
#
#Autores del código fuente: José Brito Montes, Efrain Canto Lugo
#Fecha: 5 nov 2019
#
#Descripción:
# Este algoritmo promedia varias ejecuciones del algoritmo Greedy. Dada una resolución,
# obtiene el número de comunidades promedio y su varianza. Esto lo hace para resoluciones
# de 0.25 a 1.75 con incrementos de 0.025 unidades.

import numpy as np
import networkx as nx
import matplotlib.pyplot as plt
import community

#***** Variables globales *****
escNum=9
escUmbral=[312,553,1402,57,571,361,3013,76,43]
escNom=['esc1.txt','esc2.txt','esc3.txt','esrdc.txt','estn2.txt','jlbgt.txt','ccp.txt','hu
nuc.txt','utrs.txt']
escArch=['esc1Opt.txt','esc2Opt.txt','esc3Opt.txt','esrdcOpt.txt','estn2Opt.txt','jlbgtOpt
.txt','ccpOpt.txt','hunucOpt.txt','utrsOpt.txt']

#escNum=1
#escUmbral=[3013]
#escNom=['ccp.txt']
#escArch=['ccpOpt.txt']

#***** Programa principal *****
for esc in range(escNum):

    archOpt=open(escArch[esc],"w")
    Mf = np.loadtxt(escNom[esc], int)
    Mnp = np.matrix(Mf)
    N=len(Mnp)

    GRAFOA=nx.Graph()
    GRAFOA.add_nodes_from(range(N))
    for i in range(N):
        for j in range(i+1,N):
            if Mnp[i,j]==1:
                GRAFOA.add_edge(i,j,weight=1)

    #*****
    print '*****'
    print 'Escuela', escNom[esc]
    print 'Alumnos', N

    #Componentes de la red*****
    I=list(nx.isolates(GRAFOA))
    GRAFOA.remove_nodes_from(I)
    print 'Alumnos red amistad', GRAFOA.order()

    componentes=list(nx.connected_components(GRAFOA))
    comp=[]

```

```

for c in componentes:
    sub_graph=GRAFOA.subgraph(c)
    if sub_graph.size() >= escUmbral[esc]:
        comp.append(c)
comuComp=len(componentes)-len(comp)
print 'Componentes que son comunidades', comuComp

for res in np.arange(0.25,1.75001,0.025):
    print 'Promediando resolucion', res
    comuMed=0.0
    comuVar=0.0
    samplesT=200
    tamComu=[]
    for samples in range(samplesT):
        listComunew=[]
        for c in comp:
            sub_graph=GRAFOA.subgraph(c)
            partition = community.best_partition(sub_graph,resolution=res,randomize=True)
            for com in set(partition.values()):
                listComunew.append([nodes for nodes in partition.keys() if partition[nodes] ==
com])
            tamComu.append(len(listComunew))
        comuMed=comuMed+len(listComunew)
        comuMed=comuMed/float(samplesT)
        for samples in range(samplesT):
            comuVar=comuVar+(tamComu[samples]-comuMed)**2
        comuVar=comuVar/float(samplesT)
        archOpt.write(str(res)+' '+str(comuMed+comuComp)+' '+str(comuVar)+'\n') # se
adicionan las componentes que son comunidades
    archOpt.close()

```



```

#!/usr/bin/python
#-*- coding:utf-8 -*-
#*****
#Proyecto Conacyt: Estudio de la topología y generación de modelos de difusión
# en redes sociales en escuelas a diferentes grados académicos
#
#Tesis Doctoral: Estudio de las comunidades en redes signadas: Modularidad,
# balance y frustración en redes académicas de Yucatán
#
#Autores del código fuente: José Brito Montes, Efraín Canto Lugo
#Fecha: 7 nov 2019
#
#Descripción:
# Este algoritmo Gráfica los datos obtenidos por el programa optGre.py

import numpy as np
import matplotlib.pyplot as plt

#***** Variables globales *****
escNum=9
escTitle=['esc1','esc2','esc3','esrdc','estn2','jlbjg','ccp','hunuc','utrs']
escArch=['esc1Opt.txt','esc2Opt.txt','esc3Opt.txt','esrdcOpt.txt','estn2Opt.txt','jlbjgOpt
.txt','ccpOpt.txt','hunucOpt.txt','utrsOpt.txt']
escFig=['esc1Opt.png','esc2Opt.png','esc3Opt.png','esrdcOpt.png','estn2Opt.png','jlbjgOpt.
png','ccpOpt.png','hunucOpt.png','utrsOpt.png']
escOpt=[1.0, 1.0, 0.85, 1.2, 1.0, 1.025, 0.925, 0.925, 1.3]

#escNum=1
#escTitle=['esc1']
#escArch=['esc1Opt.txt']
#escFig=['esc1Opt.png']
#escOpt=[1.0]

#***** Programa principal *****
for esc in range(escNum):
    filename = escArch[esc]
    data = np.loadtxt(filename)
    plt.figure(esc,figsize=(10,8))

    x=np.arange(0.25,1.75001,0.025)
    y=data[:,1]
    yerr=data[:,2]
    plt.errorbar(x, y, yerr=yerr, xerr=None, fmt='.r')
    plt.xlabel("Resolution")
    plt.ylabel("Average number of communities")
    plt.title("Greedy algorithm calibration "+["+escTitle[esc]+"])
    plt.axvline(escOpt[esc])
    plt.grid()
    plt.savefig(escFig[esc], dpi=100)

```

```

#!/usr/bin/python
#-*- coding:utf-8 -*-
#*****
#Proyecto Conacyt: Estudio de la topología y generación de modelos de difusión
# en redes sociales en escuelas a diferentes grados académicos
#
#Tesis Doctoral: Estudio de las comunidades en redes signadas: Modularidad,
# balance y frustración en redes académicas de Yucatán
#
#Autores del código fuente: José Brito Montes, Efrain Canto Lugo
#Fecha: 20 may 2019
#
#Descripción:
# Este programa descompone una signedNetwork en comunidades utilizando el método Greedy
# esto lo hace utilizando sólo la información de los enlaces positivos, posteriormente
# adhiere enlaces negativos y calcula la Qsigned, Balance y Frustración de las
# comunidades

import numpy as np
import networkx as nx
import matplotlib.pyplot as plt
import community

#***** Variables globales *****
escNum=9
escUmbral=[312,553,1402,57,571,361,3013,76,43]
escNom=['esc1.txt','esc2.txt','esc3.txt','esrdc.txt','estn2.txt','jlbq.txt','ccp.txt','hu
nuc.txt','utrs.txt']
escOpt=[1.0,1.0,0.85,1.2,1.0,1.025,0.925,0.925,1.3]

#escNum=1
#escUmbral=[3013]
#escNom=['ccp.txt']
#escOpt=[0.925]

archdistC=open("distC.txt","w")

#***** Programa principal *****
for esc in range(escNum):

    Mf = np.loadtxt(escNom[esc], int)
    Mnp = np.matrix(Mf)
    N=len(Mnp)

    GRAFO=nx.Graph() #GRAFO es una variable que se crea pero no se utiliza en este
programa
    GRAFO.add_nodes_from(range(N))
    for i in range(N):
        for j in range(i+1,N):
            if Mnp[i,j]==1:
                GRAFO.add_edge(i,j,weight=1)
            elif Mnp[i,j]==2:
                GRAFO.add_edge(i,j,weight=-1)

    GRAFOA=nx.Graph()
    GRAFOA.add_nodes_from(range(N))
    for i in range(N):
        for j in range(i+1,N):
            if Mnp[i,j]==1:
                GRAFOA.add_edge(i,j,weight=1)

    GRAFOE=nx.Graph() #GRAFOE es una variable que se crea pero no se utiliza en este
programa

```

```

GRAFOE.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==2:
            GRAFOE.add_edge(i,j,weight=-1)

#####
print '*****'
print 'Escuela', escNom[esc]
print 'Alumnos', N

#Componentes de la red*****
I=list(nx.isolates(GRAFOA))
GRAFOA.remove_nodes_from(I)
print 'Alumnos red amistad', GRAFOA.order()

componentes=list(nx.connected_components(GRAFOA))
comp=[]

for c in componentes:
    sub_graph=GRAFOA.subgraph(c)
    if sub_graph.size() >= escUmbral[esc]:
        comp.append(c)

listComunew=[]
for c in comp:
    sub_graph=GRAFOA.subgraph(c)
    partition = community.best_partition(sub_graph,resolution=escOpt[esc],randomize=True)
    for com in set(partition.values()):
        listComunew.append([nodes for nodes in partition.keys() if partition[nodes] ==
com])

#Agregando las componentes restantes*****
for h in componentes:
    sub=GRAFOA.subgraph(h)
    if sub.size() < escUmbral[esc]:
        listComunew.append(list(h))

#Eliminar comunidades que son nodos aislados*****
for comu in listComunew:
    if isinstance(comu, list)==False:
        GRAFOA.remove_node(comu)

#Verificar que todas las comunidades son conexas*****
for comu in listComunew:
    if isinstance(comu, list):
        comu_sub=GRAFOA.subgraph(comu)
        if nx.is_connected(comu_sub)==False:
            print 'Desconectada', comu
            break
n_c=len(listComunew)
print 'Numero de comunidades', n_c

#Agregando enlaces negativos en la red*****
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==2 and (GRAFOA.has_node(i) and GRAFOA.has_node(j)):
            GRAFOA.add_edge(i,j,weight=-1)
relacion=nx.get_edge_attributes(GRAFOA,'weight')

#Graficar las comunidades individualmente*****
#for i in range(N):
#    for j in range(i+1,N):

```

```

    #if Mnp[i,j]==1:
        #GRAFOA.add_edge(i,j,color='b')
    #elif Mnp[i,j]==2 and (GRAFOA.has_node(i) and GRAFOA.has_node(j)):
        #GRAFOA.add_edge(i,j,color='r')

for comu in listComunew:
    if isinstance(comu, list):
        H=GRAFOA.subgraph(comu)
        archdistC.write(str(H.order())+'\n')

        #etiqc={}
        #for i in range(H.order()):
            #etiqc[H.nodes()[i]]=H.nodes()[i]
        #colors = [H[u][v]['color'] for u,v in H.edges()]
        #pos = nx.spring_layout(H)
        #nx.draw_networkx_nodes(H, pos, node_color = '#FF0000', node_size=150)
        #nx.draw_networkx_labels(H, pos, labels=etiqc, font_size=10.0)
        #nx.draw_networkx_edges(H, pos, edge_color=colors,arrows=False, width=0.5)
        #plt.draw()
        #plt.show()

#*****
m=GRAFOA.size()
m_pos=0
m_neg=0
for edge in list(GRAFOA.edges()):
    a=list(edge)
    a.sort()
    a=tuple(a)
    if relacion[a]==1:
        m_pos+=1
    else:
        m_neg+=1

print 'Enlaces en la red', m
print 'Enlaces positivos en la red', m_pos
print 'Enlaces negativos en la red', m_neg

enlaces_c=[]
enlaces_cp=[]
enlaces_cn=[]
degree_neg=[]
degree_pos=[]
l_p=0
l_n=0
n_cv=0
for h in listComunew:
    if isinstance(h, list):
        n_cv+=1
        sub=GRAFOA.subgraph(h)
        nodos_sub=list(sub.nodes())
        enlaces_sub=list(sub.edges())
        enlaces=sub.size()
        m_p=0
        m_n=0
        for l in enlaces_sub:
            tupla=list(l)
            tupla.sort()
            tupla=tuple(tupla)
            if relacion[tupla]==1:
                m_p+=1
            else:
                m_n+=1

```

```

enlaces_c.append(enlaces)
enlaces_cp.append(m_p)
enlaces_cn.append(m_n)
l_p=0
l_n=0
for l in h:
    for j in GRAFOA.neighbors(l):
        a=list((l,j))
        a.sort()
        a=tuple(a)
        if relacion[a]==1:
            l_p+=1
        else:
            l_n+=1
degree_pos.append(l_p)
degree_neg.append(l_n)

print 'Comunidades verdaderas', n_cv
print 'Total de enlaces en la comunidad', enlaces_c
print 'Enlaces positivos en la comunidad', enlaces_cp
print 'Enlaces negativos en la comunidad', enlaces_cn

print 'Enlaces positivos que caen en las comunidades', sum(enlaces_cp)
print 'Enlaces negativos que caen en las comunidades', sum(enlaces_cn)
print 'Porc de pos que caen en las comunidades',
sum(enlaces_cp)/float(sum(enlaces_cp)+sum(enlaces_cn))
print 'Porc de neg que caen en las comunidades',
sum(enlaces_cn)/float(sum(enlaces_cp)+sum(enlaces_cn))

print 'Porc de pos que caen en las comunidades / positivos totales',
sum(enlaces_cp)/float(m_pos)
print 'Porc de neg que caen en las comunidades / negativos totales',
sum(enlaces_cn)/float(m_neg)

#Q y Q_signed*****
Qs=0.0
Q=0.0
if m_neg!=0:
    for i in range(n_cv):
        Qs=Qs+(enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos))-(enlaces_cn[i]-
degree_neg[i]**2/(4.0*m_neg))
        Q=Q+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
    else:
        for i in range(n_cv):
            Qs=Qs+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
            Q=Q+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
Qs=Qs/(m_pos+m_neg)
Q=Q/m_pos
print 'Q_signed', Qs
print 'Q', Q
print '(Q-Qs)/Q', ((Q-Qs)/Q)*100.0

#Social Balance Factor*****
triangles_balanced=0.0
triangles_unbalanced=0.0
triangles_total=0.0
for h in listComunew:
    if isinstance(h,list):
        h.sort()
        sub_comu=GRAFOA.subgraph(h)
        triangles = [c for c in nx.enumerate_all_cliques(sub_comu) if len(c)==3]
        if len(triangles)>0:
            triangles_total=triangles_total+len(triangles)

```

```

for triangle in triangles:
    balance=1.0
    tri=nx.subgraph (GRAFOA,triangle)
    edges_tri=list(tri.edges())
    for edge in edges_tri:
        a=list(edge)
        a.sort()
        a=tuple(a)
        balance=balance*relacion[a]
    if balance==1:
        triangles_balanced=triangles_balanced+1.0
    else:
        triangles_unbalanced=triangles_unbalanced+1.0
SBF=triangles_balanced/triangles_total
print 'Triangles balanceados', triangles_balanced
print 'Triangles no balanceados', triangles_unbalanced
print 'Triangles total', triangles_total
print 'Porc Triangles No balanceados que caen en las comunidades ',
triangles_unbalanced/float(triangles_total)
print 'Social Balance Factor', SBF

#Fustracion (coherencia) inicial*****
F=float((m_pos + sum(enlaces_cn) - sum(enlaces_cp)))/m
print 'Frustracion(Incoherencia) de la red', F
archdistC.close()

```

```

#!/usr/bin/python
#-*- coding:utf-8 -*-
#*****
#Proyecto Conacyt: Estudio de la topología y generación de modelos de difusión
# en redes sociales en escuelas a diferentes grados académicos
#
#Tesis Doctoral: Estudio de las comunidades en redes signadas: Modularidad,
# balance y frustración en redes académicas de Yucatán
#
#Autores del código fuente: José Brito Montes, Efrain Canto Lugo
#Fecha: 25 may 2019
#
#Descripción:
# Este programa realiza los mismos cálculos que el programa propGre.py y además
# realiza promedios sobre 100 simulaciones estocásticas.

import numpy as np
import networkx as nx
import matplotlib.pyplot as plt
import community

#***** Variables globales *****
escNum=9
escUmbral=[312,553,1402,57,571,361,3013,76,43]
escNom=['esc1.txt','esc2.txt','esc3.txt','esrdc.txt','estn2.txt','jlbq.txt','ccp.txt','hu
nuc.txt','utrs.txt']
escOpt=[1.0,1.0,0.85,1.2,1.0,1.025,0.925,0.925,1.3]

#escNum=1
#escUmbral=[3013]
#escNom=['ccp.txt']
#escOpt=[0.925]

archdistC=open("distC.txt","w")

#***** Programa principal *****
for esc in range(escNum):

    Mf = np.loadtxt(escNom[esc], int)
    Mnp = np.matrix(Mf)
    N=len(Mnp)

    print '*****'
    print 'Escuela', escNom[esc]
    print 'Alumnos', N

    sn_cv=0.0
    sQs=0.0
    sQ=0.0
    sSBF=0.0
    sF=0.0
    repetT=100

    for repet in range(repetT):

        GRAFO=nx.Graph() #GRAFO es una variable que se crea pero no se utiliza en este
programa
        GRAFO.add_nodes_from(range(N))
        for i in range(N):
            for j in range(i+1,N):
                if Mnp[i,j]==1:
                    GRAFO.add_edge(i,j,weight=1)
                elif Mnp[i,j]==2:

```

```

    GRAFO.add_edge(i, j, weight=-1)

GRAFOA=nx.Graph()
GRAFOA.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1, N):
        if Mnp[i, j]==1:
            GRAFOA.add_edge(i, j, weight=1)

GRAFOE=nx.Graph() #GRAFOE es una variable que se crea pero no se utiliza en este
programa
GRAFOE.add_nodes_from(range(N))
for i in range(N):
    for j in range(i+1, N):
        if Mnp[i, j]==2:
            GRAFOE.add_edge(i, j, weight=-1)

#Componentes de la red*****
I=list(nx.isolates(GRAFOA))
GRAFOA.remove_nodes_from(I)

componentes=list(nx.connected_components(GRAFOA))
comp=[]

for c in componentes:
    sub_graph=GRAFOA.subgraph(c)
    if sub_graph.size() >= escUmbral[esc]:
        comp.append(c)

listComunew=[]
for c in comp:
    sub_graph=GRAFOA.subgraph(c)
    partition = community.best_partition(sub_graph, resolution=escOpt[esc], randomize=True)
    for com in set(partition.values()):
        listComunew.append([nodes for nodes in partition.keys() if partition[nodes] ==
com])

#Agregando las componentes restantes*****
for h in componentes:
    sub=GRAFOA.subgraph(h)
    if sub.size() < escUmbral[esc]:
        listComunew.append(list(h))

#Eliminar comunidades que son nodos aislados*****
for comu in listComunew:
    if isinstance(comu, list)==False:
        GRAFOA.remove_node(comu)

#Verificar que todas las comunidades son conexas*****
for comu in listComunew:
    if isinstance(comu, list):
        comu_sub=GRAFOA.subgraph(comu)
        if nx.is_connected(comu_sub)==False:
            print 'Desconectada', comu
            #break
n_c=len(listComunew)

#Agregando enlaces negativos en la red*****
for i in range(N):
    for j in range(i+1, N):
        if Mnp[i, j]==2 and (GRAFOA.has_node(i) and GRAFOA.has_node(j)):
            GRAFOA.add_edge(i, j, weight=-1)
relacion=nx.get_edge_attributes(GRAFOA, 'weight')

```



```

#Graficar las comunidades individualmente*****
#for i in range(N):
#    #for j in range(i+1,N):
#        #if Mnp[i,j]==1:
#            #GRAFOA.add_edge(i,j,color='b')
#        #elif Mnp[i,j]==2 and (GRAFOA.has_node(i) and GRAFOA.has_node(j)):
#            #GRAFOA.add_edge(i,j,color='r')

for comu in listComunew:
    if isinstance(comu, list):
        H=GRAFOA.subgraph(comu)
        archdistC.write(str(H.order())+'\n')

        #etiqc={}
        #for i in range(H.order()):
        #    #etiqc[H.nodes()[i]]=H.nodes()[i]
        #    #colors = [H[u][v]['color'] for u,v in H.edges()]
        #    #pos = nx.spring_layout(H)
        #    #nx.draw_networkx_nodes(H, pos, node_color = '#FF0000', node_size=150)
        #    #nx.draw_networkx_labels(H, pos, labels=etiqc, font_size=10.0)
        #    #nx.draw_networkx_edges(H, pos, edge_color=colors,arrows=False, width=0.5)
        #    #plt.draw()
        #    #plt.show()

*****
m=GRAFOA.size()
m_pos=0
m_neg=0
for edge in list(GRAFOA.edges()):
    a=list(edge)
    a.sort()
    a=tuple(a)
    if relacion[a]==1:
        m_pos+=1
    else:
        m_neg+=1

enlaces_c=[]
enlaces_cp=[]
enlaces_cn=[]
degree_neg=[]
degree_pos=[]
l_p=0
l_n=0
n_cv=0
for h in listComunew:
    if isinstance(h, list):
        n_cv+=1
        sub=GRAFOA.subgraph(h)
        nodos_sub=list(sub.nodes())
        enlaces_sub=list(sub.edges())
        enlaces=sub.size()
        m_p=0
        m_n=0
        for l in enlaces_sub:
            tupla=list(l)
            tupla.sort()
            tupla=tuple(tupla)
            if relacion[tupla]==1:
                m_p+=1
            else:
                m_n+=1

```

```

enlaces_c.append(enlaces)
enlaces_cp.append(m_p)
enlaces_cn.append(m_n)
l_p=0
l_n=0
for l in h:
    for j in GRAFOA.neighbors(l):
        a=list((l,j))
        a.sort()
        a=tuple(a)
        if relacion[a]==1:
            l_p+=1
        else:
            l_n+=1
    degree_pos.append(l_p)
    degree_neg.append(l_n)
sn_cv=sn_cv+n_cv

#Q y Q_signed*****
Qs=0.0
Q=0.0
if m_neg!=0:
    for i in range(n_cv):
        Qs=Qs+(enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos))-(enlaces_cn[i]-
degree_neg[i]**2/(4.0*m_neg))
        Q=Q+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
    else:
        for i in range(n_cv):
            Qs=Qs+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
            Q=Q+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
Qs=Qs/(m_pos+m_neg)
Q=Q/m_pos
sQs=sQs+Qs
sQ=sQ+Q

#Social Balance Factor*****
triangles_balanced=0.0
triangles_unbalanced=0.0
triangles_total=0.0
for h in listComunew:
    if isinstance(h,list):
        h.sort()
        sub_comu=GRAFOA.subgraph(h)
        triangles = [c for c in nx.enumerate_all_cliques(sub_comu) if len(c)==3]
        if len(triangles)>0:
            triangles_total=triangles_total+len(triangles)
            for triangle in triangles:
                balance=1.0
                tri=nx.subgraph(GRAFOA,triangle)
                edges_tri=list(tri.edges())
                for edge in edges_tri:
                    a=list(edge)
                    a.sort()
                    a=tuple(a)
                    balance=balance*relacion[a]
                if balance==1:
                    triangles_balanced=triangles_balanced+1.0
                else:
                    triangles_unbalanced=triangles_unbalanced+1.0
SBF=triangles_balanced/triangles_total
sSBF=sSBF+SBF

#Frustracion (coherencia) inicial*****

```

```
F=float((m_pos + sum(enlaces_cn) - sum(enlaces_cp))/m
sF=sF+F

print 'Comunidades verd Prom', sn_cv/repetT
print 'Qsig Prom', sQs/repetT
print 'Q Prom', sQ/repetT
print '(Q-Qs)/Q Prom', ((sQ-sQs)/sQ)*100.0
print 'SBF Prom', sSBF/repetT
print 'Frustracion Prom', sF/repetT

archdistC.close()
```

```

#!/usr/bin/python
#-*- coding:utf-8 -*-
#*****
#Proyecto Conacyt: Estudio de la topología y generación de modelos de difusión
# en redes sociales en escuelas a diferentes grados académicos
#
#Tesis Doctoral: Estudio de las comunidades en redes signadas: Modularidad,
# balance y frustración en redes académicas de Yucatán
#
#Autores del código fuente: José Brito Montes, Efrain Canto Lugo
#Fecha: 15 jul 2019
#
#Descripción:
# Este programa descompone una signedNetwork en comunidades utilizando el método de
# Newman esto lo hace utilizando sólo la información de los enlaces positivos,
# posteriormente adhiere enlaces negativos y calcula la Qsigned, Balance y Frustración de
# las comunidades

import numpy as np
import networkx as nx
import matplotlib.pyplot as plt
import newman

#***** Variables globales *****
escNum=9
escUmbral=[312,553,1402,57,571,361,3013,76,43]
escNom=['esc1.txt', 'esc2.txt', 'esc3.txt', 'esrdc.txt', 'estn2.txt', 'jlbg.txt', 'ccp.txt', 'hu
nuc.txt', 'utrs.txt']

#escNum=1
#escUmbral=[312]
#escNom=['esc1.txt']

archdistC=open("distC.txt", "w")

#***** Programa principal *****
for esc in range(escNum):

    Mf = np.loadtxt(escNom[esc], int)
    Mnp = np.matrix(Mf)
    N=len(Mnp)

    GRAFO=nx.Graph()
    GRAFO.add_nodes_from(range(N))
    for i in range(N):
        for j in range(i+1,N):
            if Mnp[i,j]==1:
                GRAFO.add_edge(i,j,weight=1)
            elif Mnp[i,j]==2:
                GRAFO.add_edge(i,j,weight=-1)

    GRAFOA=nx.Graph()
    GRAFOA.add_nodes_from(range(N))
    for i in range(N):
        for j in range(i+1,N):
            if Mnp[i,j]==1:
                GRAFOA.add_edge(i,j,weight=1)

    GRAFOE=nx.Graph() #GRAFOE es una variable que se crea pero no se utiliza en este
programa
    GRAFOE.add_nodes_from(range(N))
    for i in range(N):
        for j in range(i+1,N):

```

```

    if Mnp[i,j]==2:
        GRAFOE.add_edge(i,j,weight=-1)

#*****
print '*****'
print 'Escuela', escNom[esc]
print 'Alumnos', N

#Componentes de la red*****
I=list(nx.isolates(GRAFOA))
GRAFOA.remove_nodes_from(I)
print 'Alumnos red amistad', GRAFOA.order()

componentes=list(nx.connected_components(GRAFOA))
comp=[]

for c in componentes:
    sub_graph=GRAFOA.subgraph(c)
    if sub_graph.size() >= escUmbral[esc]:
        comp.append(c)

listComunew=[]
for c in comp:
    sub_graph=GRAFOA.subgraph(c)
    matriz=nx.adjacency_matrix(sub_graph)

#Se pierden la numeracion de los nodos-----
matriz=matriz.todense()
archivo = open("matriz.txt","w")
for i in range(len(matriz)):
    for j in range(len(matriz)):
        archivo.write(str(matriz[i,j])+" ")
        archivo.write(' \n')
archivo.close()
mady='matriz.txt'

#Se aplica el newman a cada una de las componentes-----
listComu=newman.community(mady)
Mfaux=np.loadtxt(mady,int)
Mnpaux=np.matrix(Mfaux)
GRAFO=nx.from_numpy_matrix(Mnpaux,create_using=nx.Graph())
mapping=[]
nodos=list(sub_graph.nodes())
for i in range(GRAFO.order()):
    mapping.append((i,nodos[i]))
mapping=dict(mapping)
for h in listComu:
    if isinstance(h,list):
        sub=[]
        for i in h:
            sub.append(mapping[i])
        listComunew.append(sub)
        del sub
    else:
        listComunew.append(mapping[h])

#Agregando las componentes restantes*****
for h in componentes:
    sub=GRAFOA.subgraph(h)
    if sub.size() < escUmbral[esc]:
        listComunew.append(list(h))

#Eliminar comunidades que son nodos aislados*****

```

```

for comu in listComunew:
    if isinstance(comu, list)==False:
        GRAFOA.remove_node(comu)

#Verificar que todas las comunidades son conexas*****
for comu in listComunew:
    if isinstance(comu, list):
        comu_sub=GRAFOA.subgraph(comu)
        if nx.is_connected(comu_sub)==False:
            print 'Desconectada', comu
            break
n_c=len(listComunew)
print 'Numero de comunidades', n_c

#Agregando enlaces negativos en la red*****
for i in range(N):
    for j in range(i+1,N):
        if Mnp[i,j]==2 and (GRAFOA.has_node(i) and GRAFOA.has_node(j)):
            GRAFOA.add_edge(i,j,weight=-1)
relacion=nx.get_edge_attributes(GRAFOA,'weight')

#Graficar las comunidades individualmente*****
#for i in range(N):
#    for j in range(i+1,N):
#        #if Mnp[i,j]==1:
#            #GRAFOA.add_edge(i,j,color='b')
#        #elif Mnp[i,j]==2 and (GRAFOA.has_node(i) and GRAFOA.has_node(j)):
#            #GRAFOA.add_edge(i,j,color='r')

for comu in listComunew:
    if isinstance(comu, list):
        H=GRAFOA.subgraph(comu)
        archdistC.write(str(H.order())+'\n')

    #etiqc={}
    #for i in range(H.order()):
    #    #etiqc[H.nodes()[i]]=H.nodes()[i]
    #    #colors = [H[u][v]['color'] for u,v in H.edges()]
    #    #pos = nx.spring_layout(H)
    #    #nx.draw_networkx_nodes(H, pos, node_color = '#FF0000', node_size=150)
    #    #nx.draw_networkx_labels(H, pos, labels=etiqc, font_size=10.0)
    #    #nx.draw_networkx_edges(H, pos, edge_color=colors,arrows=False, width=0.5)
    #    #plt.draw()
    #    #plt.show()

#*****
m=GRAFOA.size()
m_pos=0
m_neg=0
for edge in list(GRAFOA.edges()):
    a=list(edge)
    a.sort()
    a=tuple(a)
    if relacion[a]==1:
        m_pos+=1
    else:
        m_neg+=1

print 'Enlaces en la red', m
print 'Enlaces positivos en la red', m_pos
print 'Enlaces negativos en la red', m_neg

enlaces_c=[]

```

```

enlaces_cp=[]
enlaces_cn=[]
degree_neg=[]
degree_pos=[]
l_p=0
l_n=0
n_cv=0
for h in listComunew:
    if isinstance(h, list):
        n_cv+=1
        sub=GRAFOA.subgraph(h)
        nodos_sub=list(sub.nodes())
        enlaces_sub=list(sub.edges())
        enlaces=sub.size()
        m_p=0
        m_n=0
        for l in enlaces_sub:
            tupla=list(l)
            tupla.sort()
            tupla=tuple(tupla)
            if relacion[tupla]==1:
                m_p+=1
            else:
                m_n+=1
        enlaces_c.append(enlaces)
        enlaces_cp.append(m_p)
        enlaces_cn.append(m_n)
        l_p=0
        l_n=0
        for l in h:
            for j in GRAFOA.neighbors(l):
                a=list((l, j))
                a.sort()
                a=tuple(a)
                if relacion[a]==1:
                    l_p+=1
                else:
                    l_n+=1
        degree_pos.append(l_p)
        degree_neg.append(l_n)

print 'Comunidades verdaderas', n_cv
print 'Total de enlaces en la comunidad', enlaces_c
print 'Enlaces positivos en la comunidad', enlaces_cp
print 'Enlaces negativos en la comunidad', enlaces_cn

print 'Enlaces positivos que caen en las comunidades', sum(enlaces_cp)
print 'Enlaces negativos que caen en las comunidades', sum(enlaces_cn)
print 'Porc de pos que caen en las comunidades',
sum(enlaces_cp)/float(sum(enlaces_cp)+sum(enlaces_cn))
print 'Porc de neg que caen en las comunidades',
sum(enlaces_cn)/float(sum(enlaces_cp)+sum(enlaces_cn))

print 'Porc de pos que caen en las comunidades / positivos totales',
sum(enlaces_cp)/float(m_pos)
print 'Porc de neg que caen en las comunidades / negativos totales',
sum(enlaces_cn)/float(m_neg)

#Q y Q_signed*****
Qs=0.0
Q=0.0
if m_neg!=0:
    for i in range(n_cv):

```

```

    Qs=Qs+(enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos))-(enlaces_cn[i]-
degree_neg[i]**2/(4.0*m_neg))
    Q=Q+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
else:
    for i in range(n_cv):
        Qs=Qs+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
        Q=Q+enlaces_cp[i]-degree_pos[i]**2/(4.0*m_pos)
Qs=Qs/(m_pos+m_neg)
Q=Q/m_pos
print 'Q_signed', Qs
print 'Q', Q
print '(Q-Qs)/Q', ((Q-Qs)/Q)*100.0

#Social Balance Factor*****
triangles_balanced=0.0
triangles_unbalanced=0.0
triangles_total=0.0
for h in listComunew:
    if isinstance(h,list):
        h.sort()
        sub_comu=GRAFOA.subgraph(h)
        triangles = [c for c in nx.enumerate_all_cliques(sub_comu) if len(c)==3]
        if len(triangles)>0:
            triangles_total=triangles_total+len(triangles)
            for triangle in triangles:
                balance=1.0
                tri=nx.subgraph(GRAFOA,triangle)
                edges_tri=list(tri.edges())
                for edge in edges_tri:
                    a=list(edge)
                    a.sort()
                    a=tuple(a)
                    balance=balance*relacion[a]
                if balance==1:
                    triangles_balanced=triangles_balanced+1.0
                else:
                    triangles_unbalanced=triangles_unbalanced+1.0
SBF=triangles_balanced/triangles_total
print 'Triangles balanceados', triangles_balanced
print 'Triangles No balanceados', triangles_unbalanced
print 'Triangles total', triangles_total
print 'Porc Triangles No balanceados que caen en las comunidades ',
triangles_unbalanced/float(triangles_total)
print 'Social Balance Factor', SBF

#Fustracion (coherencia) inicial*****
F=float((m_pos + sum(enlaces_cn) - sum(enlaces_cp))/m)
print 'Frustracion(Incoherencia) de la red', F
archdistC.close()

```