

INSTITUCIÓN UNIVERSITARIA POLITÉCNICO GRANCOLOMBIANO
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS
PROGRAMA DE INGENIERÍA DE SISTEMAS

DESARROLLO DE UN INSTRUMENTO DE ANÁLISIS DE REDES PARA EL ESTUDIO DE
SISTEMAS SOCIALES COMPLEJOS

JULIÁN ANDRÉS CASTRO RUGE
DAVID YEPES BUITRAGO

BOGOTÁ
2017

INSTITUCIÓN UNIVERSITARIA POLITÉCNICO GRANCOLOMBIANO
Facultad de Ingeniería y Ciencias Básicas
Programa de Ingeniería de Sistemas

Desarrollo de un Instrumento de Análisis de Redes para el Estudio de Sistemas
Sociales Complejos
Julián Andrés Castro Ruge
David Yepes Buitrago

2017
Bogotá

INSTITUCIÓN UNIVERSITARIA POLITÉCNICO GRANCOLOMBIANO
Facultad de Ingeniería y Ciencias Básicas
Programa de Ingeniería de Sistemas

Julián Andrés Castro Ruge
David Yepes Buitrago

Trabajo de grado para optar al título de
Ingeniero de Sistemas

Asesor
Javier Fernando Niño Velásquez
Co – Asesor
Edwin Andrés Niño Velásquez

INSTITUCIÓN UNIVERSITARIA POLITÉCNICO GRANCOLOMBIANO
FACULTAD DE INGENIERÍA
PROYECTO CURRICULAR DE INGENIERÍA DE SISTEMAS
BOGOTÁ D.C.
2017

Nota de aceptación

Firma del asesor

Firma del Co – Asesor

Firma del Jurado

Firma del Jurado

Bogotá D.C. Diciembre 2017

Agradecimientos

El agradecimiento es la memoria del corazón

Lao-Tsé

La razón de lo que eres y de tus convicciones es producto de las personas y las vicisitudes que atraviesan por tu vida, por esto siempre hay que agradecer lo que te rodea.

Comenzamos agradeciendo a nuestros padres y hermanos por su apoyo, amor y paciencia, sin los cuales no seríamos las personas que hoy en día somos; un agradecimiento especial a la persona que nos ha apoyado e introducido a las maratones de programación competitivas, sin el cual esta tesis y nuestro estudio no podría ser felizmente culminado, el decano de la facultad de ingeniería y ciencias básicas del Politécnico Grancolombiano Rafael Armando García Gómez; damos todas las gracias a nuestros directores de tesis Javier Fernando Niño Velásquez y Edwin Andres Niño Velásquez por sus conocimientos, apoyo y paciencia en la realización de esta tesis, finalmente agradecemos al Politécnico Grancolombiano por su excelente labor como institución universitaria ofreciendo profesores, becas y herramientas para el completo desarrollo de nuestros conocimientos.

Tabla de contenido

1	Introducción.....	11
2	Objetivos.....	12
2.1	Objetivo General.....	12
2.2	Objetivos Específicos.....	12
3	Planteamiento del problema.....	13
4	Alcance y restricciones.....	16
4.1	Alcance.....	16
4.2	Restricciones.....	16
5	Redes Sociales.....	17
5.1	Origen de las Redes Sociales.....	17
5.2	Análisis de Redes Sociales.....	17
5.3	Propiedades de las Redes Sociales.....	18
5.3.1	Distancias en las redes.....	18
5.3.2	Tipos de interacción.....	18
5.3.3	Coeficiente de Agrupamiento.....	18
5.3.4	Cliques.....	19
5.4	Representación de las Redes Sociales.....	20
5.5	Centralidad en las Redes Sociales.....	21
5.5.1	Centralidad de Grado.....	21
5.5.2	Centralidad de Cercanía.....	23
5.5.3	Centralidad de Intermediación.....	23
5.6	Detección de comunidades en Redes Sociales.....	24
5.6.1	Comunidades en Redes Sociales.....	25
5.6.2	Similitud de vértices.....	25
5.6.3	Medición de densidad.....	25
5.6.4	Agrupamiento de Grafos.....	25
5.6.5	Detección de comunidades.....	26
6	Metodología.....	31
6.1	Metodología general.....	31
6.2	Estudio preliminar.....	32
6.3	Definición de producto y planificación.....	32

6.3.1	Herramientas a utilizar.....	33
6.3.2	Historias de usuario	33
6.3.3	Entradas del sistema	34
6.3.4	Salidas del sistema	35
6.3.5	Peticiones de usuario	35
6.3.6	Archivos internos	36
6.3.7	Estimación por puntos de función	36
6.4	Plan de proyecto	38
6.4.1	Pila de producto	38
6.4.2	Definición de prioridades.....	39
6.5	Diagramas de caso de uso.....	40
6.5.1	Importar y exportar red	40
6.5.2	Ejecutar Algoritmos.....	40
6.5.3	Editar nodos de una red.....	41
6.5.4	Editar bordes de la red.....	41
6.6	Fase de desarrollo	41
7	Modelo.....	44
7.1	Diagramas de clases por paquetes	44
7.1.1	Algoritmos.....	44
7.1.2	Colores	45
7.1.3	Frame	45
7.1.4	Panel.....	46
7.2	Funcionalidades	47
7.2.1	Importar Red	47
7.2.2	Exportar Red	48
7.2.3	Vista Gráfica	49
7.2.4	Vista Nodos	50
7.2.5	Vista de bordes	51
7.2.6	Estadísticas básicas de la red	52
8	Conclusiones y recomendaciones.....	53
9	Bibliografía	54

Tabla de figuras

<i>Figura 1: Porcentaje de participación por edad en el uso de redes sociales (Llano, 2017).....</i>	<i>13</i>
<i>Figura 2: Porcentaje de participación de mujeres vs hombre en las redes sociales (Llano, 2017).....</i>	<i>14</i>
<i>Figura 3: Ejemplo red con alto coeficiente de agrupamiento global. Fuente: Elaboración propia</i>	<i>19</i>
<i>Figura 4: Ejemplo clique de tamaño 4. Fuente: Elaboración propia</i>	<i>19</i>
<i>Figura 5: Representación de una red social pequeña a través de una estructura de grafo8. Fuente: Elaboración propia</i>	<i>20</i>
<i>Figura 6: Grafo de ejemplo red. Fuente: Elaboración propia</i>	<i>22</i>
<i>Figura 7: (a) Visión tradicional de las comunidades como conjuntos de nodos con más y / o mejor (b) Se ilustra un grafo con su cuello de botella de conductancia mínima (Community Structure in Large Networks: Applications, 2008).....</i>	<i>24</i>
<i>Figura 8: Ejemplo de clustering en una red social Fuente: Elaboración propia</i>	<i>26</i>
<i>Figura 9 Caso de uso importación y exportación de red. Fuente: Elaboración propia.....</i>	<i>40</i>
<i>Figura 10 Caso de uso ejecución de algoritmos. Fuente: Elaboración propia.....</i>	<i>40</i>
<i>Figura 11: : Caso de uso edición de nodos sobre una red. Fuente: Elaboración propia</i>	<i>41</i>
<i>Figura 12: Caso de uso edición de bordes sobre una red. Fuente: Elaboración propia.....</i>	<i>41</i>
<i>Figura 13: Formato JSON de un grafo sobre el software. Fuente: Elaboración propia</i>	<i>42</i>
<i>Figura 14 : Diagrama de clases (algoritmos). Fuente: Elaboración propia</i>	<i>44</i>
<i>Figura 15: Diagrama de clases (colores). Fuente: Elaboración propia</i>	<i>45</i>
<i>Figura 16: Diagrama de clases (Frame). Fuente: Elaboración propia</i>	<i>45</i>
<i>Figura 17: Diagrama de clases (Panels). Fuente: Elaboración propia</i>	<i>46</i>
<i>Figura 18: Funcionalidad de importación de una red. Fuente: Elaboración propia</i>	<i>47</i>
<i>Figura 19: Funcionalidades de exportación de una red, Fuente: Elaboración propia</i>	<i>48</i>
<i>Figura 20 : Vista de la pestaña Graph, Fuente: Elaboración propia.....</i>	<i>49</i>
<i>Figura 21: Vista de la pestaña Nodes, Fuente: Elaboración propia.....</i>	<i>50</i>
<i>Figura 22: Vista de la pestaña Edge, Fuente: Elaboración propia</i>	<i>51</i>
<i>Figura 23: Estadística de edad en un gráfico de torta, Fuente: Elaboración propia</i>	<i>52</i>
<i>Figura 24: Estadística de edad en un gráfico de barras, Fuente: Elaboración propia</i>	<i>52</i>

Tabla de tablas

<i>Tabla 1: Representación de una red social pequeña a través de una estructura de matriz</i>	<i>21</i>
<i>Tabla 2: Descripción en la cual se muestra el grado de cada uno de los nodos descritos en la red de la Figura 6 ..</i>	<i>22</i>
<i>Tabla 3: Historias de usuario</i>	<i>34</i>
<i>Tabla 4: Entradas del sistema</i>	<i>34</i>
<i>Tabla 5: Salidas del sistema</i>	<i>35</i>
<i>Tabla 6: Peticiones de usuario</i>	<i>35</i>
<i>Tabla 7: Archivos internos</i>	<i>36</i>
<i>Tabla 8: PFSA</i>	<i>36</i>
<i>Tabla 9: FC</i>	<i>37</i>
<i>Tabla 10: PFA</i>	<i>37</i>
<i>Tabla 11: Pila de producto</i>	<i>38</i>
<i>Tabla 12: Prioridades del software</i>	<i>39</i>

Tabla de ecuaciones

<i>Ecuación 1: Cálculo de grado de un nodo</i>	<i>22</i>
<i>Ecuación 2: Cálculo de centralidad de cercanía</i>	<i>23</i>
<i>Ecuación 3: Cálculo de centralidad de intermediación sobre nodos</i>	<i>23</i>
<i>Ecuación 4: Cálculo de centralidad de intermediación sobre aristas.....</i>	<i>24</i>
<i>Ecuación 5: Función de modularidad de Newman.....</i>	<i>29</i>

1 Introducción

Cuando hablamos de redes sociales encontramos que este concepto se asocia al uso masivo de herramientas de internet como Facebook, Instagram, Twitter; como es evidente, el impacto de estas redes sociales sobre la vida cotidiana y el nivel de influencia en la comunicación entre millones de individuos, ha provocado su crecimiento de usuarios de forma exponencial en el transcurso de los últimos 10 años. Cuando el término red social es nombrado en el ámbito científico se hace referencia a un conjunto finito de actores y las relaciones que los vinculan, estos a su vez poseen un conjunto finito de cualidades o características propias de cada uno, como lo puede ser la edad, el género, ciudad de residencia, entre otros; las cuales a través de rigurosos estudios pueden otorgar información que nos permita llegar a conclusiones o dar respuesta a diferentes preguntas como por qué los actores se relacionan entre ellos, también cuestiones como quién podría llegar a ser la persona con mayor influencia dentro de una red social.

Teniendo en cuenta lo anterior se plantean ofrecer las herramientas necesarias para poder realizar diferentes estudios sobre redes sociales y con base en el ARS (Análisis de Redes Sociales), la teoría de grafos, cálculos matemáticos y probabilísticos; este proyecto plantea el desarrollo de una herramienta que apoye o sustente el modelamiento de sociedades a través de estructuras en red.

En el cuarto capítulo se encontrará una breve descripción a cerca del alcance planteado para este proyecto, esto basado en los objetivos tanto generales como específicos con la finalidad de llevarlos a un buen término, además, se plantean unas posibles limitaciones que pueden ocurrir en el desarrollo de este proyecto.

A lo largo del quinto capítulo se realiza una contextualización donde se presentan conceptos de redes sociales, Análisis de Redes Sociales, propiedades, representaciones de redes, medidas relevantes utilizadas en el ARS, por último, se presenta una definición y unos métodos para el análisis y posible reconocimiento de las comunidades encontradas en una red social, esto basado en un criterio de similaridad previamente establecido.

En el capítulo seis se realiza una explicación detallada sobre la metodología implementada para el desarrollo de este proyecto, describiendo principalmente la metodología general debido a que para el óptimo desarrollo de la misma se tuvo que replantear una metodología específica y parcialmente adaptada a nuestras necesidades. Se realiza una descripción sobre la descripción del producto o software y la planeación, posteriormente se describen las herramientas utilizadas y los entregables necesarios para el correcto desarrollo de este software.

El capítulo siete consiste de una muestra tanto interna como externa de la primera versión del software, interna partiendo con el diagrama de clases donde describe la estructura del proyecto mediante UML, finalizando con la muestra externa del software, es decir, funcionalidades, vistas y graficas.

El capítulo ocho, finalmente, presenta las conclusiones generadas al finalizar el proceso de realización de este proyecto, comprobando el cumplimiento de los objetivos planteados inicialmente.

2 Objetivos

2.1 Objetivo General

Proponer una solución que apoye procesos de toma de decisiones que afecten sistemas susceptibles de ser modelados como redes sociales, por medio del cálculo y despliegue de medidas de sus propiedades.

2.2 Objetivos Específicos

- Establecer un estado del arte del Análisis de Redes Sociales e identificar propiedades de estructuras de redes sociales relevantes en procesos de toma de decisiones.
- Construir un instrumento que facilite la medición y visualización de propiedades relevantes de redes sociales para permitir al usuario analizarlas y tomar decisiones.
- Probar las diferentes funcionalidades del instrumento propuesto en el análisis de un caso real.

3 Planteamiento del problema

En la actualidad encontramos una gran tendencia de comunicación y relación de las personas a través de herramientas que facilitan la interacción, las redes sociales es uno de los métodos favoritos en cuanto a comunicación informal hablamos, esto puede ser demostrado por el último balance de usuarios donde muestra que existen 3000 millones activos actualmente (Tiempo, 2017), donde también muestra que casi un cuarto de la población mundial (un 22,9%) usa Facebook, de estos usuarios podemos resaltar que el 53% son mujeres que se encuentran entre los 24 y los 65 años, con mayor representación de ellas en edades de entre los 25 a 35 años, también en cuanto a los usuarios masculinos cuentan con un 47% de los usuarios totales, entre los cuales tienen mayor participación los hombres con edades de entre los 20 y los 34 años. En la red social Facebook se generan más de 240 millones de “likes” y 14.6 billones de fotos son compartidas por hora, en redes sociales como Twitter son generados más de 20 millones de tweets y en Instagram son subidas 3.5 millones de fotos todo esto cada hora todos los días (Llano, 2017).

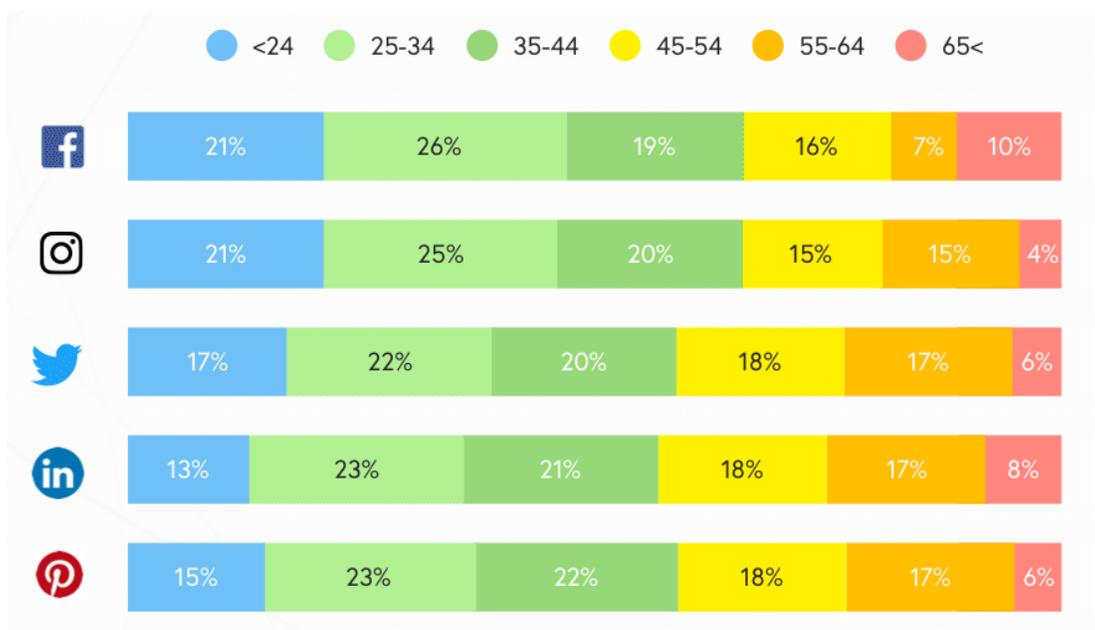


Figura 1: Porcentaje de participación por edad en el uso de redes sociales (Llano, 2017)

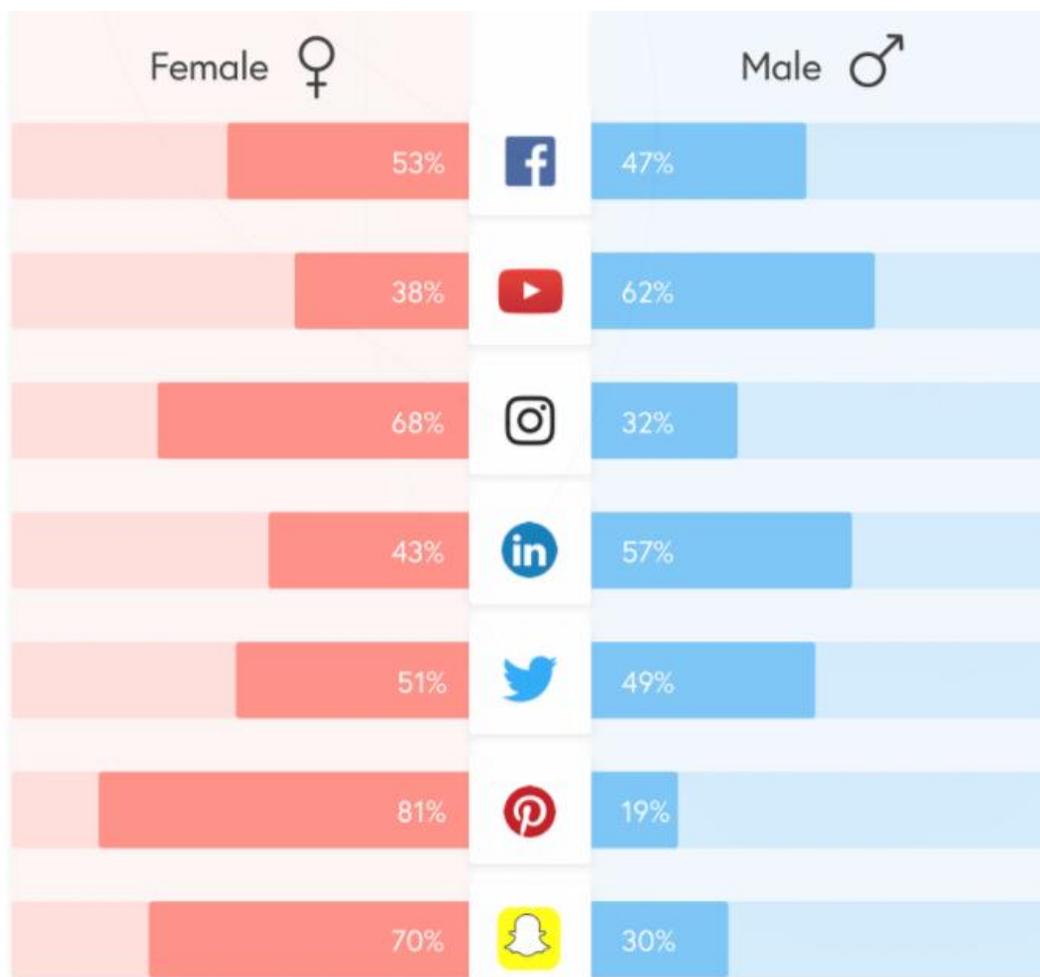


Figura 2: Porcentaje de participación de mujeres vs hombre en las redes sociales (Llano, 2017)

Todo lo anterior demuestra varios puntos importantes, entre los principales, se encuentra la concurrencia diaria de los usuarios en las redes sociales de forma masiva. Podemos destacar también que gran parte de su vida y su información es difundida a través de la red, entre los cuales pueden estar sus gustos, lugares de preferencia, actividades diarias, personas con las que se relaciona frecuentemente, música favorita, metas, entre otras cosas tal como sucede en Facebook y su nueva actualización donde permite a los usuarios compartir su estado de ánimo en una palabra clave (feliz, triste, enfermo, etc.).

La cantidad de personas e información que generan estas herramientas nos abre una brecha grande para un campo de estudio de las ciencias sociales como lo es el Análisis de Redes Sociales (ARS), este campo tiene la materia prima necesaria para alimentar herramientas que permitan tomar esos datos y poder aplicar diferentes procesos que generen métricas o medidas de importancia, sobre diferentes tipos de redes como lo pueden ser redes ya sea completas donde los lazos son las

relaciones específicas en una población definida, o bien en redes personales o redes egocéntricas donde se estudian comunidades personales.

Es por lo mencionado anteriormente que se decidió desarrollar un proyecto cuyo enfoque será generar una herramienta informática, que, en conjunto con la teoría de grafos, distribuciones matemáticas y probabilísticas permita realizar un análisis completo de redes egocéntricas y personales. La cual a su vez podrá generar métricas o medidas que permitan llegar a conclusiones precisas sobre una red en particular, como por ejemplo encontrar las clases y tipos de comunidades, poder conocer los tipos de centralidad que ocurren en la misma, como también identificar grados de homofilia (amor a los iguales) o por el contrario grados de heterofilia, etc.

4 Alcance y restricciones

4.1 Alcance

El presente proyecto tiene como idea principal presentar conceptos básicos sobre las redes sociales, ofreciendo conceptos que parten desde algoritmos, representaciones matemáticas, representaciones gráficas, análisis de datos, medidas relevantes sobre estas (centralidad de grado, grado de intermediación, etc.), enfocándose en el desarrollo de la primera versión de un software de análisis de redes sociales.

El desarrollo de este proyecto contempla un software orientado al ARS estará compuesto por unas funcionalidades básicas pre-establecidas las cuales se describen a continuación:

- Representación gráfica de una red.
- Diámetro de una red.
- Menor camino entre cualquier par de nodos en una red.
- Detección de Comunidades.
- Representación del grado de centralidad de un conjunto de nodos en una red.
- Representación del grado de intermediación de un conjunto de nodos en una red.
- Cálculo del grado de cercanía de un conjunto de nodos en una red.

4.2 Restricciones

- Por políticas de privacidad que se aplican en la mayoría de las redes sociales, el obtener datos reales y/o relevantes para ser usados en estudios que se acercasen más a la realidad, causa un conflicto de veracidad sobre el desarrollo
- Ineficiencia a nivel computacional para cálculos sobre redes, esto debido a que algoritmos como el de detección de comunidades de Girvan Newman sobre redes con conjunto de actores muy grande se comporta de una forma en exceso costosa a nivel tiempo.

5 Redes Sociales

Se puede definir una red social como un conjunto finito de actores (organizaciones, comunidades, sociedades, individuos grupos, etc.) que pueden ser relacionados o vinculados a otros a través de una relación o un conjunto de relaciones sociales (Danah m. Boyd, 2007). Estas redes sociales basan su apoyo en el Análisis de Redes Sociales (ARS) el cual se encarga en tomar las relaciones entre actores como herramientas principales para construir y organizar el comportamiento social de los actores. El punto de análisis es utilizar las relaciones y proporcionar un conjunto de métodos y técnicas para el estudio formal de las relaciones entre actores.

Las secciones que vienen a continuación presentan una descripción detallada del concepto de redes sociales, los tipos de modelos de redes, las propiedades con las que cuenta, su clasificación en base a su estructura, los sistemas que permiten realizar un análisis de estas estructuras y las técnicas de medición utilizadas.

5.1 Origen de las Redes Sociales

Estructuralmente una red social está conformada por grupos de individuos, a los cuales se les llama nodos y los cuales están conectados mediante algún tipo de relación como puede ser una amistad, parentesco, negocios entre muchas otras. Este modelo estructural es frecuentemente utilizado para modelar una situación social. Desde 1930 aparece la sociometría como una manera de formalizar las ciencias sociales, en esta se utiliza la estadística para estudiar poblaciones y la teoría de grafos como herramienta para modelar la relación entre personas.

Debido al alto número de sistemas en la Web que permiten generar sociedades o comunidades virtuales que pueden ser representadas como una estructura de red, la teoría basada en la estructura de las redes sociales fue retomada.

5.2 Análisis de Redes Sociales

El Análisis de Redes Sociales (ARS) nació en los años 70 con la fundación de la International Network for Social Network Analysis o INSNA (Red Internacional para el Análisis de redes sociales), la cual es una asociación académica profesional de investigadores y profesionales del análisis de redes sociales, cuyos miembros tienen intereses en estas redes como un nuevo paradigma teórico. El análisis de redes es una aproximación intelectual amplia para identificar las estructuras sociales que emergen de las diversas formas de relación, así como también es un conjunto específico de métodos y técnicas. El ARS también puede definirse como una herramienta de medición y análisis de las estructuras sociales que surgen de las relaciones entre actores sociales diversos.

5.3 Propiedades de las Redes Sociales

Las redes sociales son representadas mediante grafos, además se utiliza técnicas de la teoría de grafos para el estudio de la estructura en las redes sociales. Sin embargo, existen diferencias en la forma en cómo se aplican los distintos conceptos de la teoría de grafos sobre el análisis de redes sociales. En esta sección se estudiará las diferentes propiedades que presentan las redes sociales en base a su forma, similitud entre los conjuntos de nodos y relaciones que existen en la red y su distribución.

5.3.1 Distancias en las redes

Cuando hablamos de distancia se hace referencia a una medición de la teoría de grafos que nos permite definir propiedades más complejas sobre la posición de los nodos o individuos y sobre la estructura de la red. En ARS una camino o trayectoria es el número de enlaces o adyacencias que existen entre dos nodos en el grafo (donde los nodos no son iguales). Por otro lado, la distancia geodésica es el número mínimo de enlaces que llevan o que son necesarios para llegar de un nodo a otro es decir trayectoria mínima, esta medida en ARS es fundamental ya que nos permite obtener el camino con mayor eficiencia en cuestión de enlaces para llegar desde y hasta cualquier par de actores en una red.

5.3.2 Tipos de interacción

Dentro de una estructura de grafo podemos establecer un nivel de conexión entre nodos y sus relaciones mediante el análisis de cómo interactúan los agentes o nodos dentro de esta red, esto quiere decir que, si es una red social, las relaciones de los agentes puede ser de amistad y normalmente es doblemente dirigida por lo que el usuario A es amigo del usuario B lo cual sucede también de la relación entre B y A. Esto se denomina conectividad lo cual permite hacer una reducción entre dos nodos, dicha conectividad al encontrarse de una forma concurrente en un grafo permite trayectorias más cortas entre todas las parejas de nodos, lo cual a su vez permite la reducción en el diámetro de una red.

5.3.3 Coeficiente de Agrupamiento

En teoría de grafos el coeficiente de agrupamiento es una métrica de similitud, la cual se puede caracterizar de dos formas, la primera es el coeficiente de agrupamiento local el cual mide el nivel de interconexión de un vértice con sus vecinos, donde el agrupamiento o vecindario de un nodo A está compuesto por el número de nodos que se encuentran adyacentes a él, la segunda es el coeficiente de agrupamiento global la cual indica el nivel de agrupamiento de cada uno de los nodos con respecto a toda la red. En hacen una gran propuesta sobre la importancia del cálculo del coeficiente de agrupamiento en estudios de ARS (Strogatz, Noviembre 1997). En la Figura 1 se puede notar como el grafo tiene 4 zonas donde existe un alto nivel de agrupamiento global.

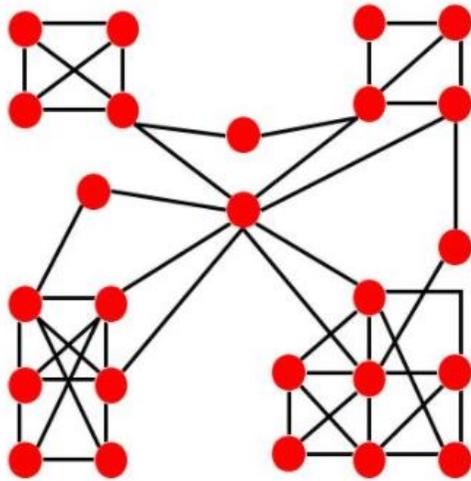


Figura 3: Ejemplo red con alto coeficiente de agrupamiento global. Fuente: Elaboración propia

5.3.4 Cliques

Formalmente un clique lo podemos definir como un conjunto de vértices o nodos en el cual para cada vértice existe una arista que los conecta. En esencia un clique es un subgrafo el cual cumple que para todo nodo existe una adyacencia hacia cualquier otro nodo que este contenido en ese subgrafo, es decir, podemos definirlo como un grafo completo. La Figura 2 muestra un grafo el cual está compuesto por un subgrafo que a su vez es una clique de tamaño 3.

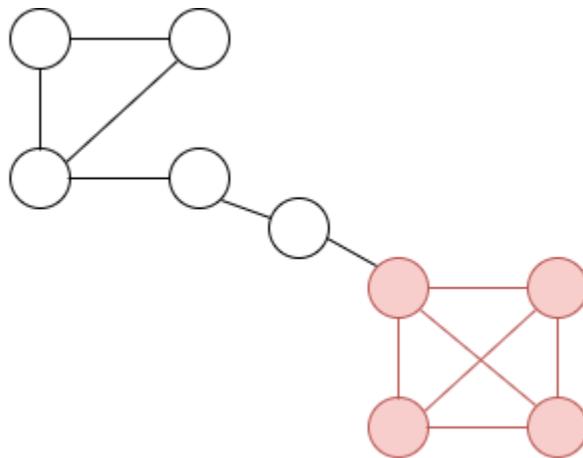


Figura 4: Ejemplo clique de tamaño 4. Fuente: Elaboración propia

5.4 Representación de las Redes Sociales

Sobre una red social podemos definir dos conjuntos importantes de componentes, el primer conjunto se trata de los nodos, también llamados actores o vértices lo cual representa un individuo en esta red social, el segundo conjunto son las aristas que también pueden ser llamadas relaciones o vínculos la cual se encarga de representar la unión o enlace entre dos nodos en una red social. En el ARS se usan dos herramientas fundamentales de las matemáticas para representar la información de las relaciones entre actores, las cuales son los grafos y las matrices, esto con el fin de facilitar su representación al momento de su análisis.

Los grafos los podemos definir también como un conjunto de puntos o figuras que representan a los actores que componen a una red social los cuales son unidos a través de adyacencias o líneas que pueden ser dirigidas o no dirigidas, las cuales representan la unión entre ese par de nodos. En la Figura 3 se observa un ejemplo de la representación de una red social pequeña en estructura de grafo.

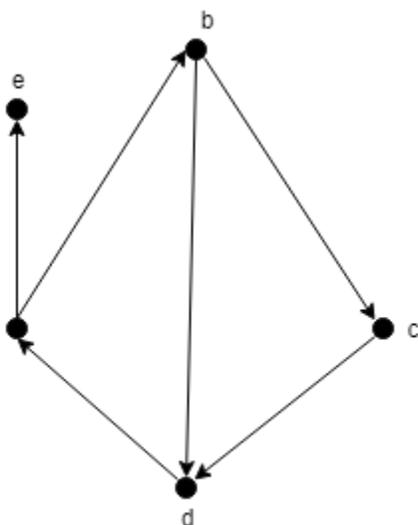


Figura 5: Representación de una red social pequeña a través de una estructura de grafo⁸. Fuente: Elaboración propia

Las matrices por otro lado es una representación más formal ya que es un arreglo bidimensional de números, la cual tiene un tamaño de $N \times N$ es decir es una matriz cuadrada donde N es el número de actores o nodos presentes en la red. Los valores contenidos dentro de dicha matriz representan las relaciones entre nodos, es decir, el valor contenido en la fila i columna j representa el valor de la relación entre el nodo i y el nodo j .

	a	b	c	d	e
a	0	1	0	0	1
b	0	0	1	1	0
c	0	0	0	1	0
d	1	0	0	0	0
e	0	0	0	0	0

Tabla 1: Representación de una red social pequeña a través de una estructura de matriz.

5.5 Centralidad en las Redes Sociales

A partir de la necesidad por entender el comportamiento de los actores en una red social se empieza a ver la importancia del concepto de centralidad, con la cual puede resolverse importantes preguntas como el determinar que tanto influye un vértice en comparación a los demás en una red. La idea de centralidad aplicada en las redes humanas fue introducida por Bavelas en 1948, quien dirigió el primer trabajo de investigación sobre centralidad para el Group Networks Laboratory del M.I.T en la década de los 40 (Freeman, 1979). El uso de matemáticas y los conceptos de teoría de grafos son la base para los cálculos de centralidad esto debido a que ofrecen demostraciones y teorías que sustentan el muestreo y los valores de centralidad identificados. Podemos encontrar en la actualidad muchos estudios que sustentan la importancia de la medida de centralidad sobre el análisis de redes sociales (Elizabeth Costenbader, 2003).

En esta tesis usaremos las tres principales mediciones de centralidad basadas en redes sociales las cuales son:

- Centralidad de grado (Degree)
- Centralidad de cercanía (Closeness)
- Centralidad de intermediación (Betweenness)

5.5.1 Centralidad de Grado

En términos generales se puede definir el grado de un nodo como el número de enlaces en los que él se encuentra relacionado, es decir el número de enlaces que conectan un nodo con cualquier otro nodo en una red. En los grafos dirigidos el grado puede ser dividido en dos formas: el grado de salida (out-degree) el cual se puede definir como el número de enlaces que posee el nodo A y que tienen una dirección hacia otro nodo cualquiera, el grado de entrada (in-degree) por el contrario podemos definirlo como los enlaces que parten de cualquier nodo de la red y tienen una dirección hacia A.

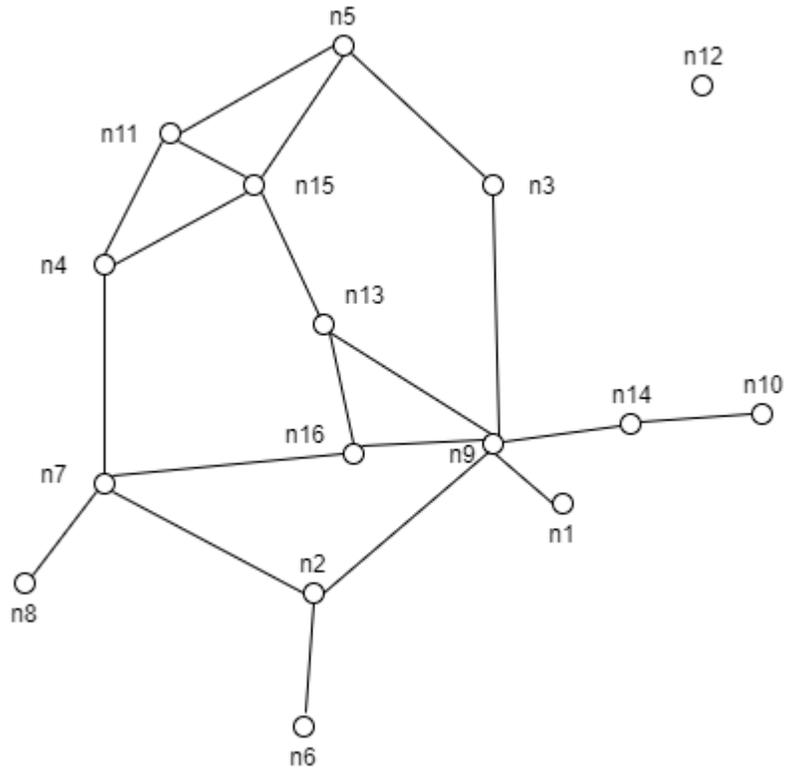


Figura 6: Grafo de ejemplo red. Fuente: Elaboración propia

nodo = n_i	grado = $d(n_i)$	nodo = n_i	grado = $d(n_i)$
n_1	1	n_9	6
n_2	3	n_{10}	1
n_3	2	n_{11}	3
n_4	3	n_{12}	0
n_5	3	n_{13}	3
n_6	1	n_{14}	2
n_7	4	n_{15}	4
n_8	1	n_{16}	3

Tabla 2: Descripción en la cual se muestra el grado de cada uno de los nodos descritos en la red de la Figura 6

Formalmente podemos definir el grado G_i de un nodo de la siguiente forma:

$$G_i = \sum_j^N a_{ij}$$

Donde N es el número de adyacencias del nodo, a_{ij} es la adyacencia entre el nodo a_i y a_j

Ecuación 1: Cálculo de grado de un nodo

5.5.2 Centralidad de Cercanía

Se define la centralidad de cercanía como la menor cantidad de nodos promedio por los que debe cruzar un nodo definido para llegar a cualquier otro (Stanley Wasserman, 1994). Formalmente lo podemos definir como la inversa de la longitud promedio de los caminos más cortos con los demás vértices de la red. Además, define la siguiente ecuación para el cálculo de centralidad de un nodo en grafos dirigidos:

$$G_C(n_i) = \frac{1}{\sum_{j=1}^N d(n_i, n_j)} = \frac{1}{MED_{i \neq j}(d(n_i, n_j))}$$

Ecuación 2: Cálculo de centralidad de cercanía

Donde:

- MED es la medida de las distancias.
- $d(n_i, n_j)$ es la menor distancia entre el nodo n_i y el nodo n_j
- N es el número de nodos en la red.

5.5.3 Centralidad de Intermediación

En esta tesis tendremos en cuenta dos grados de centralidad por intermediación:

- **Centralidad de intermediación sobre nodos:**

Se puede definir como la frecuencia en la que un nodo aparece en los caminos más cortos que se comprenden sobre todo par de nodos en el grafo, es decir, entre más veces aparece un nodo en los caminos más cortos de la red mucho mayor es su grado de intermediación. En la siguiente ecuación se define formalmente el cálculo del grado de intermediación de un nodo i :

$$C_{BET}(i) = \sum_{j,k} \frac{b_{jik}}{b_{jk}}$$

Ecuación 3: Cálculo de centralidad de intermediación sobre nodos

Donde:

- b_{jk} es el número de caminos más cortos desde el nodo j hasta el nodo k .
 - b_{jik} es el número de caminos más cortos desde j hasta k que pasan a través del nodo i .
- **Centralidad de intermediación sobre adyacencias:**
Esta medida representa la importancia de una adyacencia con respecto a los caminos más cortos de una red, es decir, entre más veces aparece una adyacencia en los caminos más

5.6.1 Comunidades en Redes Sociales

Se puede definir una comunidad o agrupamiento como un grupo de nodos y adyacencias las cuales comparten ciertas propiedades en común e influyen de manera similar. Sobre una red social podemos encontrar un gran número de posibles comunidades como lo pueden ser un círculo de amigos de la universidad, familiares, amistades de trabajo, ciudades, entre otras.

Existen dos formas para identificar un buen agrupamiento. La primera, analiza ciertos valores para los vértices del grafo para posteriormente agruparlos con respecto al valor obtenido. La segunda, se basa en la calidad de un agrupamiento dado.

5.6.2 Similitud de vértices

En este tipo de análisis se busca agrupar a los nodos que no están bien conectados pero que tienen alguna similitud compartida entre ellos. Para esto se debe partir de varios criterios de similitud los cuales en esta tesis solo serán mencionados y son los siguientes:

- Similitud basada en distancia y mediciones
- Similitud basada en la adyacencia
- Similitud basada en la conectividad

5.6.3 Medición de densidad

Esta medida puede ser tomada de las siguientes dos formas: La densidad es el promedio de los grados de incidencia de los vértices (Bijit, 2010). Si se suman los grados de incidencia se tendrá el valor $2E$ por lo cual la densidad resulta $\frac{2E}{V}$ donde E es el número de adyacencias y V el número de vértices. Se considera que un grafo es denso si E es proporcional a V^2 de lo contrario se considera disperso. La segunda forma de evaluar la densidad es comparar la densidad de un subgrafo con respecto a una medición de densidad dada (umbral). Es decir, definimos dos normas que debe cumplir este subgrafo, la primera es escoger un entero K el cual será comparado con la cardinalidad del subgrafo y debe cumplir que sean iguales, la segunda es escoger un número racional α el cual debe cumplir ser menor que la densidad del subgrafo.

5.6.4 Agrupamiento de Grafos

También definido como clustering se puede definir el agrupamiento como la clasificación de objetos en diferentes subconjuntos donde los elementos de cada subconjunto comparten una o varias características (Schaeffer, 2007). Esto se hace con el fin de dividir un conjunto de datos en

grupos más pequeños a los cuales se les puede denominar como similares o que tienen alguna relación en algún sentido.

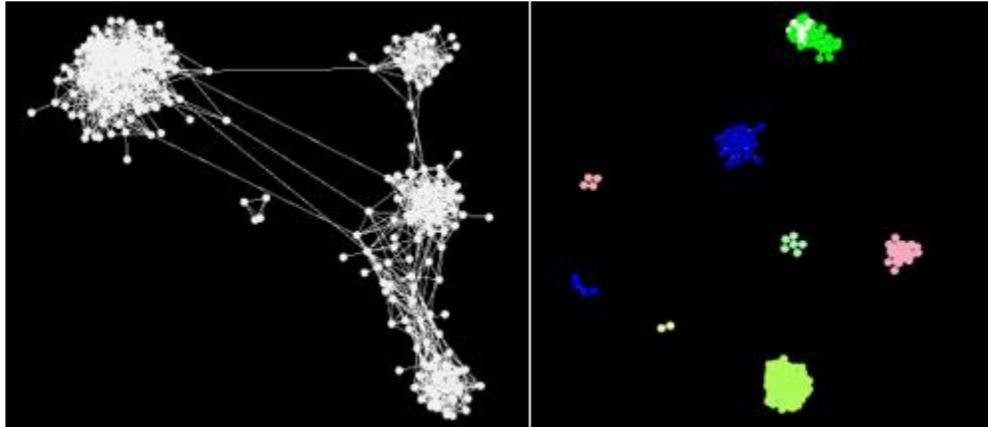


Figura 8: Ejemplo de clustering en una red social Fuente: Elaboración propia

5.6.4.1 Agrupamiento global de grafos

Se define este agrupamiento como una identificación dentro del grafo total, sobre el cual se realiza una partición de los datos a subgrupos de tal forma que los datos que comparten el mismo grupo se encuentren lo más estrechamente relacionados posible, esto mediante alguna medida de similitud definida. Se puede definir formalmente como: G_1, \dots, G_k agrupaciones, tal que $G_i \cap G_j = \emptyset$ cuando $i \neq j$ (Ávalos Gaytán Vanesa, 2009).

5.6.4.2 Agrupamiento local de grafos

En agrupamiento local los métodos de búsqueda son basados en algoritmos heurísticos y probabilísticos diseñados para encontrar soluciones óptimas. Mediante decisiones probabilísticas establecen una región que contenga a las mejores soluciones. Las soluciones posibles son representadas por un estado el cual es llamado espacio de estado.

5.6.5 Detección de comunidades

Se pueden encontrar muchos algoritmos orientados a encontrar la estructura de comunidades sobre redes (Gregory S. , 2010). Estos algoritmos son basados en dos problemas:

- **Detección de comunidades con traslapamiento:** El problema consiste en el caso en que un actor de una red pertenece o se clasifica en más de una comunidad. Por ejemplo el individuo puede tener en común una actividad laboral lo cual hace que pertenezca a una comunidad de trabajo, pero, a su vez puede realizar actividades estudiantiles lo cual también lo involucra en una comunidad estudiantil.

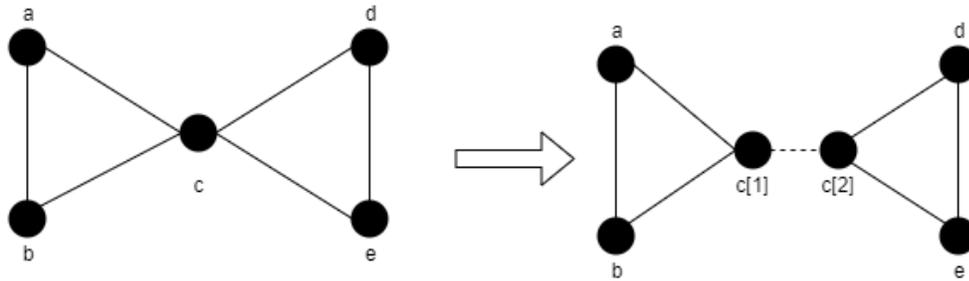


Figura 9: Ejemplo de descomposición de un nodo en una red con traslapamiento. Fuente: Elaboración propia

- **Detección de comunidades disjuntas:** Como se menciona en (Gregory B. Y., Detecting communities in networks by merging cliques, 2009) estos algoritmos son más estrictos en cuanto a la selección de comunidad, debido a que definen n comunidades sobre las cuales un nodo puede pertenecer a únicamente una.

5.6.5.1 Algoritmo basado en la intermediación

Girvan y Newman proponen un algoritmo basado en la intermediación, es decir, se encarga de identificar las agrupaciones mediante la identificación de aristas relevantes sobre una red para su posterior eliminación y reevaluación (Girvan., 2004).

Este algoritmo de Girvan y Newman para la detección de comunidades basado en la intermediación de adyacencias es el siguiente:

1. Calcular la intermediación para todas las aristas en la red

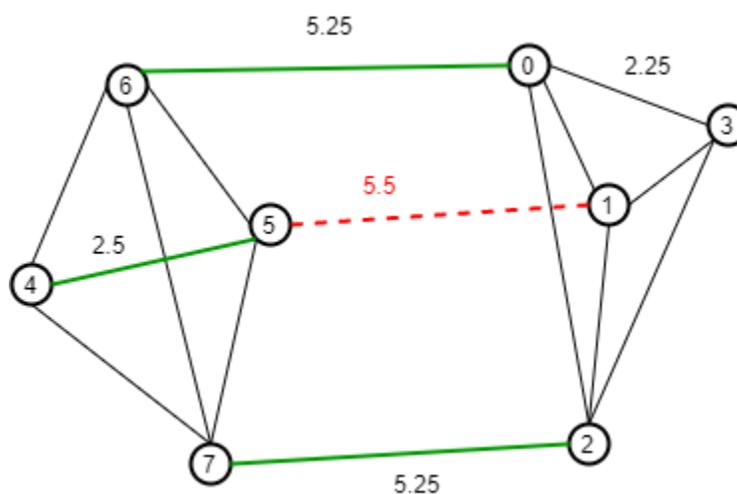


Figura 10: Cálculo de intermediación sobre aristas e identificación de la arista con mayor valor. Fuente: Elaboración propia

2. Remove la arista de mayor valor de intermediación, si se encuentra uno o más empates, posteriormente generar el nuevo grafo.

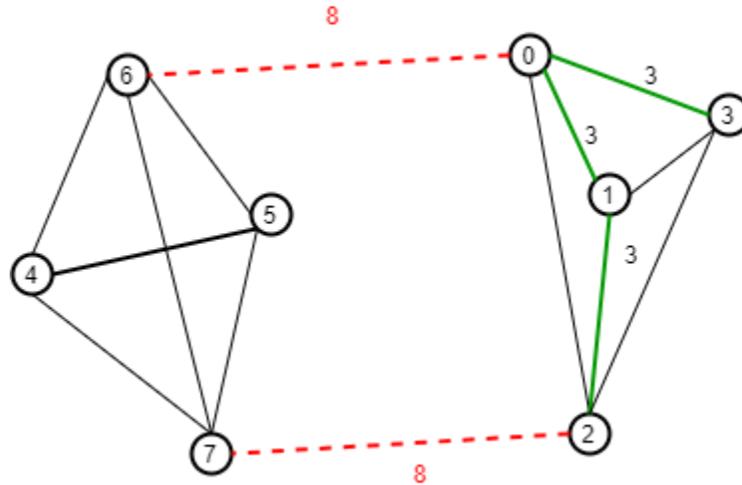


Figura 11: Identificación de aristas empatadas para ser eliminadas. Fuente: Elaboración propia

3. Se realiza la repetición del paso 1 hasta que los valores de intermediación tengan un nivel bajo

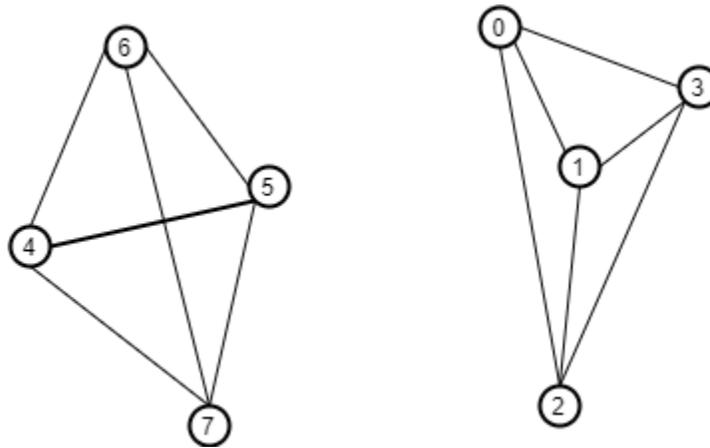


Figura 12: Identificación de comunidades por bajo nivel de intermediación en sus adyacencias. Fuente: Elaboración propia

Después de generar las n comunidades, se deberá aplicar otro algoritmo llamado la función de modularidad la cual mide la calidad de las particiones con respecto a las demás particiones, es decir, evalúa la calidad las particiones y decide si un par de comunidades deben estar unidas o por el contrario se realizó un corte correctamente. La función de modularidad planteada por Newman (Girvan., 2004) está dada por:

$$M = \sum_i (e_{ii} - a_i^2)$$

Ecuación 5: Función de modularidad de Newman

Donde:

- e_{ii} es el número de aristas en la red que conecta vértices de la comunidad i al grupo j
- $a_i = \sum_j e_{ij}$

5.6.5.2 Algoritmo basado en cliques

Podemos definir una comunidad o agrupamiento como un conjunto de actores que comparten un atributo sobre el cual puede definirse que tienen alguna similitud, una clique a su vez es un conjunto de actores que se encuentran estrechamente relacionados entre Sí, por lo cual uno de los principales métodos de identificación de comunidades es basado en la búsqueda de cliques sobre una red. Bowen propone un algoritmo de detección de comunidades a través de cliques sobre una red, cada clique es tomada como una comunidad inicialmente y sobre estas se unen pares de comunidades a través de la función de modularidad propuesta por Newman (Gregory B. Y., Detecting communities in networks by merging cliques, 2009).

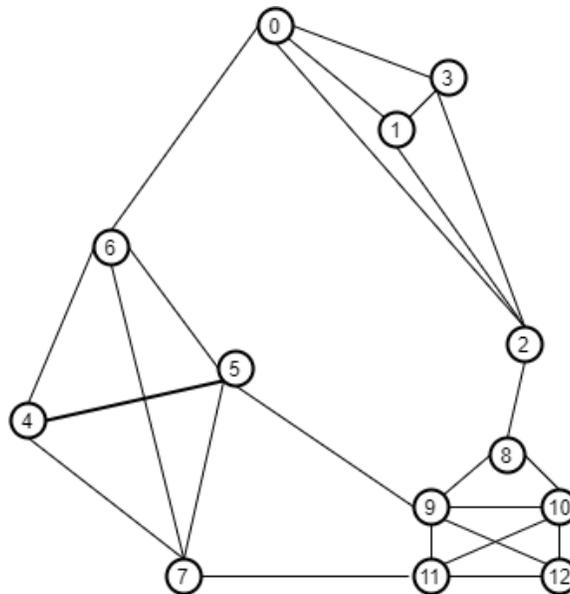


Figura 13: Grafo Original. Fuente: Elaboración propia

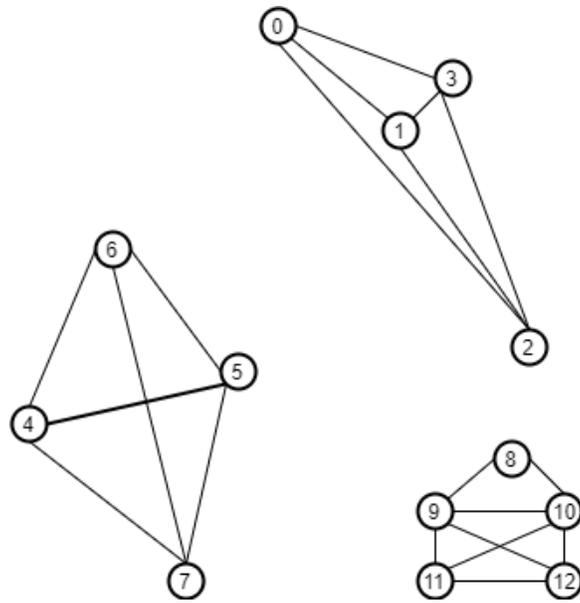


Figura 14: Cliques del grafo original (Figura 11). Fuente: Elaboración propia

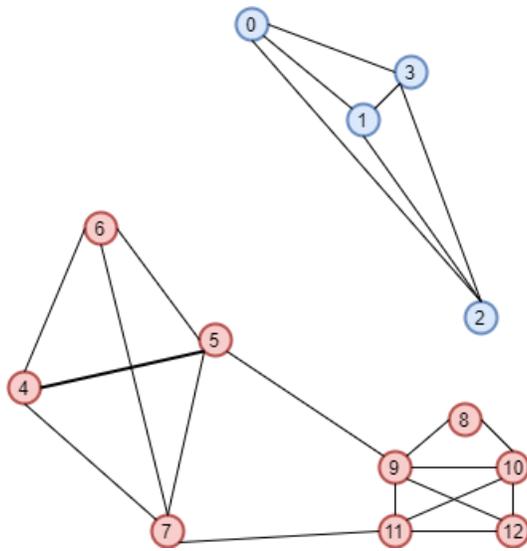


Figura 15: Comunidades detectadas. Fuente: Elaboración propia

6 Metodología

6.1 Metodología general

Primero se realizó una investigación sobre cual metodología utilizar para el planteamiento y el desarrollo del problema, en esta investigación se tuvieron en cuenta metodologías tradicionales o clásicas que cuentan con una etapa de análisis cuidadoso antes de construir el producto, y metodologías ágiles de desarrollo (Patricio Letelier Torre, 2003); dado a que no se encontró un procedimiento que se ajuste completamente al proyecto, se decidió formular una serie de requisitos los cuales darían paso al planteamiento de un nuevo proceso el cual nos facilitará realizar el proyecto de la mejor manera.

En este proyecto se generó una metodología donde primero identificamos el problema, investigamos y solucionamos el problema proponiendo una serie de alternativas y decisiones a tomar en cada paso o fase del proyecto. Dado que el proyecto requería una investigación previa, para tener un producto más consistente, se consultó en varias fuentes bibliográficas para tener un concepto más centrado sobre el problema y cómo solucionarlo.

Una de las fases más importantes a llevar a cabo en el proyecto es aquella donde se reúnen los requisitos más importantes y de más alto impacto, sobre los cuáles se define la planificación y el esquema del proyecto. Después de la etapa donde definimos la solución, se eligieron las herramientas a utilizar; esto definirá gran parte del tiempo del desarrollo, decidirá qué tanto tiempo se gastará en la fase de desarrollo y de pruebas.

El siguiente paso a tomar fue analizar los requisitos y desarrollar un prototipo del producto, incluyendo las partes más críticas y relevantes del proyecto, en esta misma fase también se incluyeron módulos de revisión y aceptación, dado que sin este prototipo no se tendría un esquema general del producto. La fase más crítica del proyecto dado al ciclo de vida y a los riesgos que existían en esta, es el periodo de desarrollo del proyecto; ya con la planeación previa y el análisis de requisitos realizado el siguiente paso es ejecutar el plan de desarrollo del producto, esta fase reúne a varias fases como lo es la de investigación, análisis de requisitos y planificación; en esta fase fuimos muy cautelosos con forme se iba desarrollando.

Finalizando la fase anterior se encuentra la fase donde se realizan las pruebas necesarias para entregar el producto, es una de las partes más importantes del proyecto, es donde se elige la metodología y las herramientas necesarias para hacer las pruebas; se harán pruebas a los módulos realizados para la aplicación y estas pruebas nos dirán cuál es el porcentaje del producto completado y cuáles son los fallos presentados, conforme a estas repuestas decidiremos si el producto necesita más plazo de entrega o si ya es posible entregar el producto.

6.2 Estudio preliminar

Se realizó un proceso de investigación bibliográfica, donde se tuvo en cuenta libros, artículos y demás fuentes de investigación que nos pudieran aportar a la investigación y resolución del problema; esta investigación requiere que el equipo de trabajo revise e investigue diferentes fuentes, de esta manera se podrá determinar el alcance y la complejidad del proyecto. Pero en esta fase no solo se investiga sobre el problema principal del proyecto de grado, también hay que determinar limitaciones.

Una de las más grandes limitaciones sobre el proyecto, es la solicitud para mostrar y extraer datos reales sobre las redes sociales, dado a que la mayoría tiene cláusulas de privacidad muy puntuales, y no deja extraer mucha información sobre la red, se investigaron herramientas como Facebook development (la cual es una herramienta enfocada en desarrolladores para la extracción de información de la red social de Facebook), pero el acceso a los datos personales de un nodo en la red está restringido por políticas de privacidad de los desarrolladores de estas redes, que es un gran inconveniente y uno de los más grandes riesgos del proyecto hasta el momento.

Otro punto de investigación importante fue buscar herramientas de monitorización de redes sociales como TweetReach, Reachli, BuzzSumo, Many Eyes, GraphChi y Guess, las cuales sirven para representación de datos sobre una población, pero ninguno de los anteriores nos sirve para solucionar el problema planteado, lo que hacen es solo visualización por ejemplo TweetReach determina el alcance de un tweet. GraphChi es una aplicación para graficar muchos datos de manera eficiente pero ninguna herramienta tiene el alcance que nosotros queremos como determinar las diferentes características sobre una red social como la tendencia a relacionarse con personas de sus mismas tendencias, la persona o nodo más central en una red y demás.

Dado a la investigación previa y los problemas que tenemos que solucionar se requiere una herramienta que nos ayude a solucionar estas incógnitas de la manera más eficiente posible; se realizó una investigación sobre cuáles herramientas o plataformas nos ayudarían a dar solución a nuestro problema, se investigó herramientas de desarrollo web como paperscape, sigma js y árbol js; también se tuvieron en cuenta herramientas para una aplicación de escritorio, se estudiaron la viabilidad de cada API como Visual Library, Java Jung, Graph Stream, Gephi y JgraphT. Esta etapa se concluyó haciendo una investigación sobre la viabilidad de utilizar las distintas API anteriormente comentadas para desarrollar la herramienta de investigación teniendo en cuenta puntos importantes como tiempo de aprendizaje, manejo versátil en el lenguaje, código claro y conciso, funcionalidades y representaciones con cada una de estas.

6.3 Definición de producto y planificación

Finalizada la etapa de investigación se definió el producto el cual va a ser una herramienta en la cual el usuario podrá tener respuesta a diferentes incógnitas como las que se plantearon anteriormente, se requiere que la aplicación sea fácil de utilizar y sirva para el análisis de estas redes.

Se tuvo en cuenta las investigaciones anteriormente realizadas, en cuanto a las herramientas y API's de desarrollo, el tiempo de aprendizaje y otros aspectos y se decidió utilizar Graph Stream API de java, porque es un lenguaje que el equipo de trabajo ya maneja muy bien, el manejo visual de la red con esta herramienta es bastante fácil de programar y genera que el equipo no se gaste tanto tiempo realizando la parte visual si no la parte lógica la cual es la base de la herramienta.

6.3.1 Herramientas a utilizar

Las herramientas a utilizar es el API de Graph Stream el cual es un visualizador y presentador de un grafo el cual se le pueden añadir nodos aristas o bordes, añadir atributos a los vértices, como por ejemplo edad, género, ciudad y demás.

Otra herramienta a utilizar es JSON, la cual nos permite recibir un archivo de texto JSON y nos permite representar y construir, este API fue utilizado para importar una red social para representarla en la herramienta.

También se utilizó JFreeChart la cual es una herramienta para graficar los atributos de las personas en una red, por ejemplo, los nodos tienen atributos de edad de diferentes rangos solo se cuentan por edad el número de nodos con ese valor de atributo y se grafica en un diagrama de barras o un diagrama de pastel.

6.3.2 Historias de usuario

No.		Tipo		Descripción
1.	Yo como	Usuario	Quiero	Poder visualizar la red
2.	Yo como	Usuario	Quiero	Poder visualizar los nodos
3.	Yo como	Usuario	Quiero	Poder visualizar las aristas
4.	Yo como	Usuario	Quiero	Poder modificar los nodos
5.	Yo como	Usuario	Quiero	Poder Modificar las aristas
6.	Yo como	Usuario	Quiero	Poder eliminar nodos
7.	Yo como	Usuario	Quiero	Poder eliminar Aristas
8.	Yo como	Usuario	Quiero	Poder Agregar nodos
9.	Yo como	Usuario	Quiero	Poder agregar aristas
10.	Yo como	Usuario	Quiero	Ver la representación en matriz de adyacencia del grafo

11.	Yo como	Usuario	Quiero	Ver los datos de la red
12.	Yo como	Usuario	Quiero	Tener una lista de algoritmos para ejecutar
13.	Yo como	Usuario	Quiero	Tener una lista de los atributos de todos los nodos
14.	Yo como	Usuario	Quiero	Seleccionar los nodos por atributos
15.	Yo como	Usuario	Quiero	Poder acercar la vista de la red
16.	Yo como	Usuario	Quiero	Poder guardar el estado del grafo

Tabla 3: Historias de usuario

6.3.3 Entradas del sistema

Entradas	
Modificar Nodo	1
Agregar Nodo	1
Eliminar Nodo	1
Agregar Borde	1
Modificar Borde	1
Eliminar Borde	1
Seleccionar Nodos	1
Seleccionar Algoritmo	1
Seleccionar Nodo Por Atributo	1
Guardar Red	1
Importar Red	1

Tabla 4: Entradas del sistema

6.3.4 Salidas del sistema

Salidas	
Visualizar Red	1
Atributos de la red	1
Listado de algoritmos	1
Listado de atributos de nodo	1
Listado de nodos	1
Listado de bordes	1
Matriz de adyacencia	1
Graficas por atributo de nodos	1

Tabla 5: Salidas del sistema

6.3.5 Peticiones de usuario

Peticiones	
Poder importar una red	1
Poder guardar una red	1
Poder seleccionar un algoritmo	1
Poder Seleccionar un atributo	1
Poder modificar una red	1
Poder seleccionar entre gráfica de barras y de pastel	1

Tabla 6: Peticiones de usuario

6.3.6 Archivos internos

Archivos	
Algoritmos	10
Frame	1
Panel	8

Tabla 7: Archivos internos

6.3.7 Estimación por puntos de función

6.3.7.1 Estimación PFSA

	Cuenta	Baja	Media	Alta	Totales
Entradas Del Sistema	11	6 x 3	3 x 4	2 x 6	42
Salidas Del Sistema	8	1 x 4	6 x 5	1 x 7	41
Peticiones De Usuario	6	2 x 3	3 x 4	1 x 6	24
Número de archivos internos	19	9 x 7	7 x 10	3 x 15	178
				Total	285

Tabla 8: PFSA

6.3.7.2 Estimación FC

No.	Factor de complejidad	Valor (0 ... 5)
1	Comunicación de datos	3
2	Proceso distribuido	1
3	Objetivos de rendimiento	4
4	Configuración de la exploración compartida	1
5	Tasa de transacciones	0
6	Entrada de datos en línea	0
7	Eficiencia con el usuario final	3
8	Actualizaciones en línea	0
9	Lógica de proceso interno compleja	3
10	Reusabilidad de código	4
11	Conversión e instalación contempladas	3
12	Facilidad de operación (backups, etc)	2

13	Instalaciones múltiples	0
14	Facilidad de cambios	4
Factor de complejidad total (FC)		28

Tabla 9: FC

6.3.7.3 Estimación PFA

$$PFA = PFSA * (0.65 + (0.01 * FC))$$

	PFSA			FC	Multiplicación	Suma	Resultado PFA
PFA	285	0.65	0.01	28	0.28	0.93	265.05

Tabla 10: PFA

6.3.7.4 Estimación de esfuerzo

- **Líneas de código:** Seguimos la fórmula dada, y dado que Java es un lenguaje de cuarta generación, tomamos 20 como las líneas de código por punto de función.
 - Líneas de código = PFA * líneas de código por punto de función
 - Líneas de código = 265.05 * 20 = 5301
- **Horas/persona:** Dado que Java es un lenguaje de cuarta generación, tomamos 5 como la cantidad de horas por punto de función y teniendo en cuenta nuestras habilidades, seguimos la fórmula dada.
 - Horas/persona = PFA / (1/ horas por punto de función)
 - Horas/persona = 265.05 / (1/5) = 1325 horas/persona

6.3.7.5 Estimación de duración

- **En horas:** Con el resultado del cálculo de esfuerzo, tomamos las horas por persona y las dividimos en la cantidad de desarrolladores.
 - Duración en horas = horas/persona / cantidad de personas
 - Duración en horas = 1325 / 2 = 662.5 horas
- **En meses:** Tomamos el resultado anterior y lo dividimos en las horas que se le dedicarán al mes al proyecto.
 - Duración en meses = duración en horas / horas dedicadas al mes
 - Duración en meses = 662.5 / 120 = 5.520833 meses

6.4 Plan de proyecto

6.4.1 Pila de producto

1. Algoritmos	2. Vista Gráfica de la red	3. Vista De Nodos	4. Vista De Bordos	5. Vista De Matriz De adyacencia	6. Vista De Gráficos	7. Importar Y Exportar Red
1.1 Desarrollo de los algoritmos investigados	2.1 Integrar GraphStream en Java Swing	3.1 Listar los nodos en la red	4.1 Listar los bordes en la red	5.1 Realizar la representación de la red como una matriz	6.1 Realizar el gráfico de barras por característica de un nodo	7.1 Definir el formato del archivo
	2.1.1 Realizar la integración de la visualización del grafo	3.1.1 Sección para crear un nodo	4.1.1 Sección para crear un borde		6.2 Realizar el gráfico de pastel por característica de un nodo	7.2 Realizar la funcionalidad de importar red
	2.2 Crear listas	3.1.2 Sección para editar un nodo	4.1.2 Sección para editar un borde			7.3 Realizar la funcionalidad de exportar red
	2.2.1 Lista de las propiedades de la red	3.1.3 Sección para eliminar un nodo	4.1.2 Sección para eliminar un borde			
	2.2.2 Lista de los algoritmos a utilizar					
	2.2.3 Listar las propiedades de los nodos en la red					
	2.3 Unir la sección de algoritmos con la lista en esta vista					

Tabla 11: Pila de producto

6.4.2 Definición de prioridades

No.	Id Historia	Nombre
1	1.1	Desarrollo de los algoritmos investigados
2	2.2.1	Lista de las propiedades de la red
3	2.2.2	Lista de los algoritmos a utilizar
4	2.2.3	Listar las propiedades de los nodos en la red
5	2.3	Unir la sección de algoritmos con la lista en esta vista
6	2.1	Integrar GraphStream en Java Swing
7	7.1	Definir el formato del archivo
8	7.2	Realizar la funcionalidad de importar red
9	7.3	Realizar la funcionalidad de exportar red
10	3.1.1	Sección para crear un nodo
11	3.1.2	Sección para editar un nodo
12	3.1.3	Sección para eliminar un nodo
13	4.1.1	Sección para crear un borde
14	4.1.2	Sección para editar un borde
15	4.1.2	Sección para eliminar un borde
16	5.1	Realizar la representación de la red como una matriz
17	6.1	Realizar el gráfico de barras por característica de un nodo
18	6.2	Realizar el gráfico de pastel por característica de un nodo

Tabla 12: Prioridades del software

6.5 Diagramas de caso de uso

6.5.1 Importar y exportar red

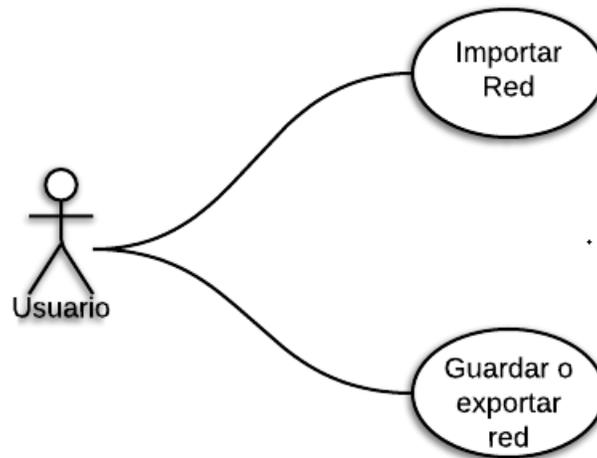


Figura 9 Caso de uso importación y exportación de red. Fuente: Elaboración propia

6.5.2 Ejecutar Algoritmos

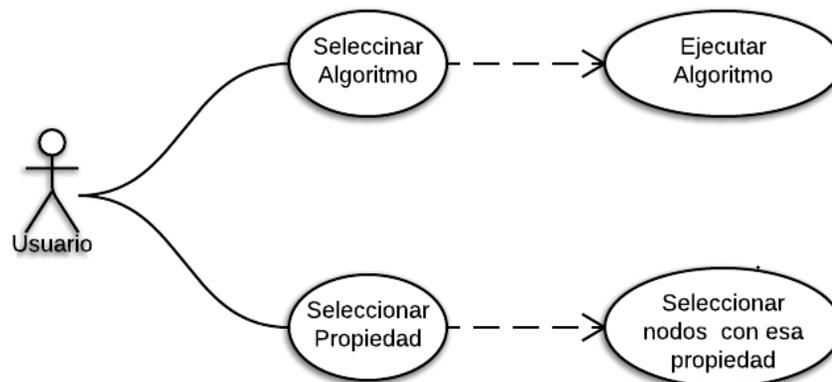


Figura 10 Caso de uso ejecución de algoritmos. Fuente: Elaboración propia

6.5.3 Editar nodos de una red

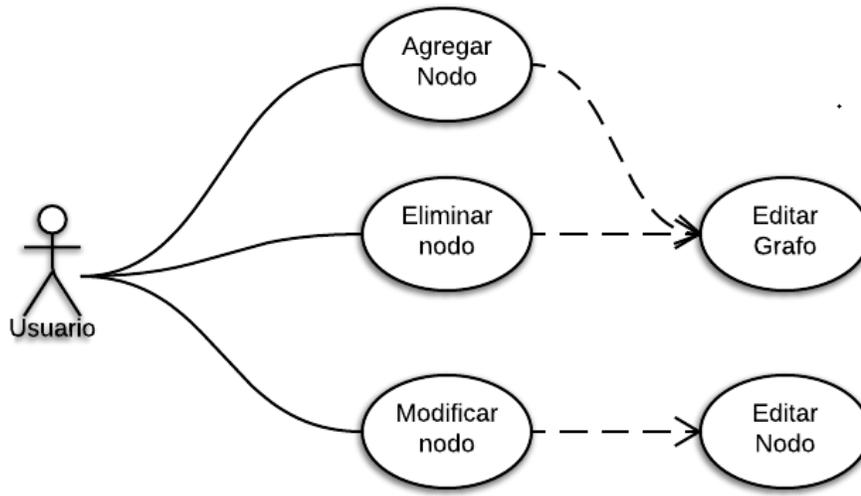


Figura 11: : Caso de uso edición de nodos sobre una red. Fuente: Elaboración propia

6.5.4 Editar bordes de la red

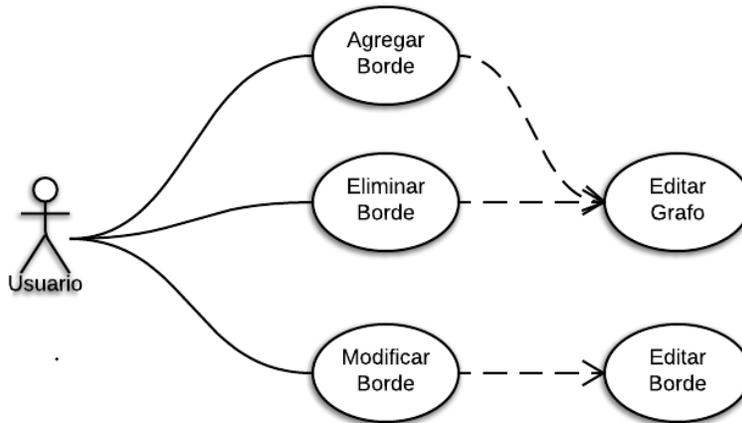


Figura 12: Caso de uso edición de bordes sobre una red. Fuente: Elaboración propia

6.6 Fase de desarrollo

En esta fase se toma el plan de desarrollo y se realizan las diferentes tareas planteadas en la sección anterior, las tareas se van realizando conforme a la lista de prioridades. El punto de partida en el desarrollo del proyecto, fue empezando realizando los algoritmos que se investigaron con anterioridad, el algoritmo inicial fue el de distancia mínima entre dos nodos en la red mostrando el camino de esta distancia mínima, los siguientes realizados fueron todos los algoritmos de

centralidad, componentes fuertemente conexas, puntos de articulación, árbol de expansión mínimo, detección de comunidades por grado de intermediación sobre las aristas, detección de comunidades a partir de cliques y función de modularidad.

Después de realizar los algoritmos, se realizó una interfaz gráfica en Java Swing la cual contendrá las diferentes vistas planteadas en la fase de planificación. En el Frame creado con anterioridad se realiza un apartado el cual contiene las listas de los algoritmos, propiedades de los nodos en la red, y las propiedades de la red; se empezó realizando las propiedades del grafo como número de nodos, número de aristas diámetro del grafo y demás. El siguiente punto a realizar fue la lista de los algoritmos a utilizar. Después se realizó la lista de las propiedades que tienen los nodos en el grafo. En esta iteración se realizó la interfaz la cual representa de forma gráfica la red, en este punto se tuvo un retraso dado a que la librería no se terminaba de adaptar al producto, anteriormente se utilizaba Java Jung, dado a que el código de cómo se manejaba esta librería no se modelaba a las necesidades del producto, se decidió cambiar a un API diferente la cual es Graph Stream, se requirió modificar los algoritmos y las listas dado a que la librería a la cual se hizo el cambio se utiliza de forma diferente.

El siguiente paso en el proyecto es el poder que el usuario pueda importar y guardar el grafo, primero se definió el archivo en cual formato se iba a recibir y guardar, principalmente el formato propuesto fue formato JSON, el formato definido es el siguiente.

```
{"userId":2112,"userName":"userName","nodes":
[{"id":"121221212","userName":"NodeA","name":"Name Node A"},
{"id":"121221213","userName":"NodeB","name":"Name Node B"}],
"links":[{"source":"NodeA","target":"NodeB"}]}
```

Figura 13: Formato JSON de un grafo sobre el software. Fuente: Elaboración propia

Ya se tiene el formato el siguiente paso es implementar los métodos requeridos para leer y guardar la red, esto se realiza con el API gson-2.6.2, con la cual se puede leer un archivo JSON, en el formato ya definido; para escribir un archivo y guardar la red que está en el software se escribe un file input en la ruta especificada por el usuario y se escribe un archivo común y corriente para escribir y guardar la red no fue necesario utilizar ninguna API, el software guarda la red en el mismo formato mencionado.

La siguiente iteración se dirigió a realizar el apartado de la lista de nodos, este apartado es algo complejo dado a la cantidad de funcionalidades que hay en este, está la lista de los nodos que es la función principal de este apartado, pero también está la funcionalidad de crear un nuevo nodo agregando propiedades, también está la funcionalidad de modificar nodo la cual se puede agregar, quitar y modificar propiedades de este y esta la funcionalidad de eliminar nodo. Dado a que esta componente se volvió compleja de programar se decidió separar las componentes por módulos para la creación de este apartado.

Igual que el anterior, la siguiente iteración se concentró en hacer un apartado para los bordes o las aristas, tiene las mismas funcionalidades que la anterior listar las aristas, crear, eliminar y modificar arista, y al igual que los nodos tiene propiedades como peso o costo de esta misma. La

siguiente iteración, aunque pequeña muy importante dado a que una de las representaciones más importantes es, la representación por matriz de adyacencia la cual tiene $N \times N$ espacios, cada espacio se rellena con el peso que tiene la arista del nodo u al nodo v , es una sección con poca funcionalidad, pero muy importante.

La última sección que se realizó fue la de gráficas de barras y pastel, estas gráficas son realizadas con el API JFreeChart la cual nos da una opción muy viable el cual solo le tenemos que pasar el parámetro y la cantidad para cada uno de ellos, es una de las partes menos importantes del proyecto y es una de las partes más fáciles en tiempo.

La última iteración del proyecto se utiliza para corregir errores que se detectaron en la aplicación en todos los módulos, correcciones en el módulo de vista de la red, módulo de nodos y aristas.

7 Modelo

7.1 Diagramas de clases por paquetes

7.1.1 Algoritmos

En el módulo de algoritmos se encuentra las diferentes funciones que podrán ser ejecutadas en la red actual, es decir, si se quiere utilizar alguna de estas implementaciones se debe realizar la instancia de alguna de estas clases, las cuales reciben como parámetro un grafo sobre el cual después de la ejecución del programa generara una serie de cambios, tanto a nivel de datos y a nivel visual.

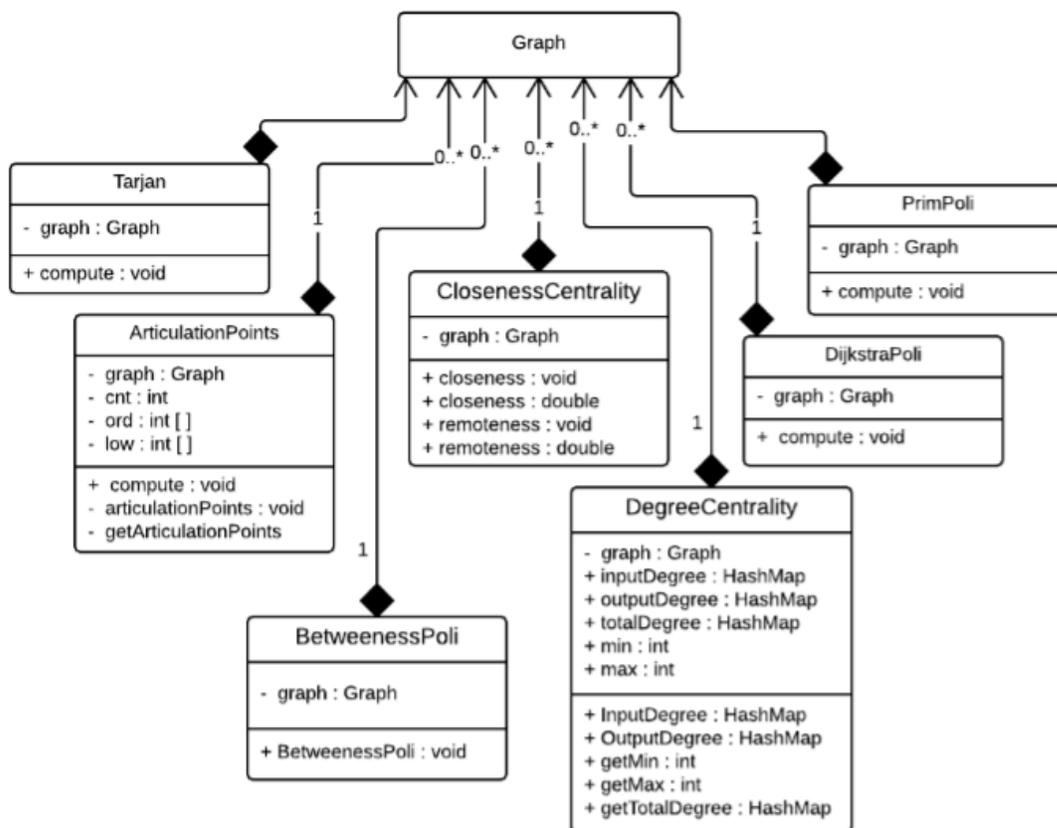


Figura 14 : Diagrama de clases (algoritmos). Fuente: Elaboración propia

7.1.2 Colores

En el módulo de colores nos proporciona una ayuda, a obtener una lista de colores en orden aleatorio, dado a que el producto requería que el usuario diferenciara de alguna manera los algoritmos ejecutados; la primera clase (ColorRandom) tiene como función proporcionar una lista de colores seleccionados dado a un diccionario en orden aleatorio, y la segunda clase (ColorToHex) dado a 3 enteros, los cuales son la representación de un color en RGB, nos proporciona ese mismo color pero en su representación hexadecimal.

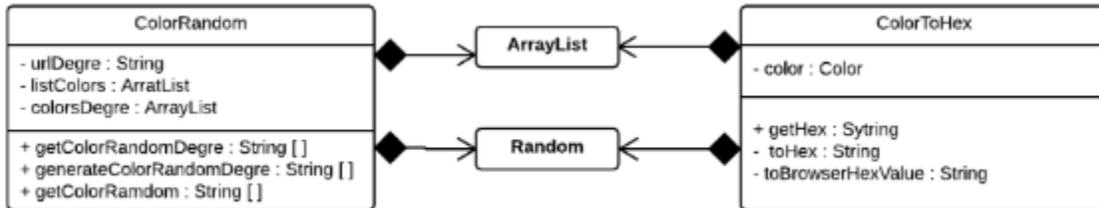


Figura 15: Diagrama de clases (colores). Fuente: Elaboración propia

7.1.3 Frame

Este módulo contiene las principales vistas del producto, una de estas es utilizada como pantalla de carga de la aplicación, cuando carga la aplicación se muestra la siguiente vista (PrincipalFrame), la cual contendrá las diferentes funcionalidades con las que el usuario podrá interactuar.

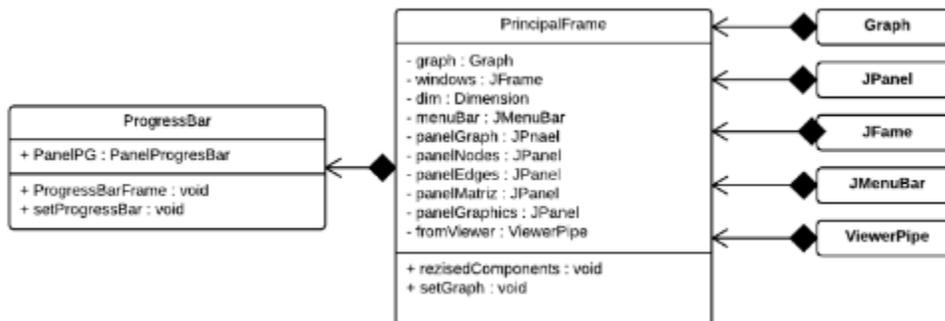


Figura 16: Diagrama de clases (Frame). Fuente: Elaboración propia

7.1.4 Panel

En esta sección del producto, estarán todas las vistas que extienden de JPanel, con las que el usuario puede interactuar, estas clases se encuentran divididas de la siguiente forma:

- **PanelGraph:** En esta clase se hace una contextualización general de la red y los posibles algoritmos a ejecutar sobre esta.
- **PanelNodes:** Esta clase contiene una lista de los nodos de la red y funcionalidades como un modelo CRUD (crear, leer, modificar y eliminar).
- **PanelEdges:** Esta vista contiene un listado de las adyacencias en la red y proporciona las mismas funcionalidades de la vista anterior, orientado a los bordes.
- **PanelMatrix:** En este podremos encontrar una representación con una matriz de adyacencia de la red actual.
- **PanelGraphics:** Esta vista proporciona representaciones graficas sobre los conjuntos de características de los nodos que se encuentran a lo largo de la red.

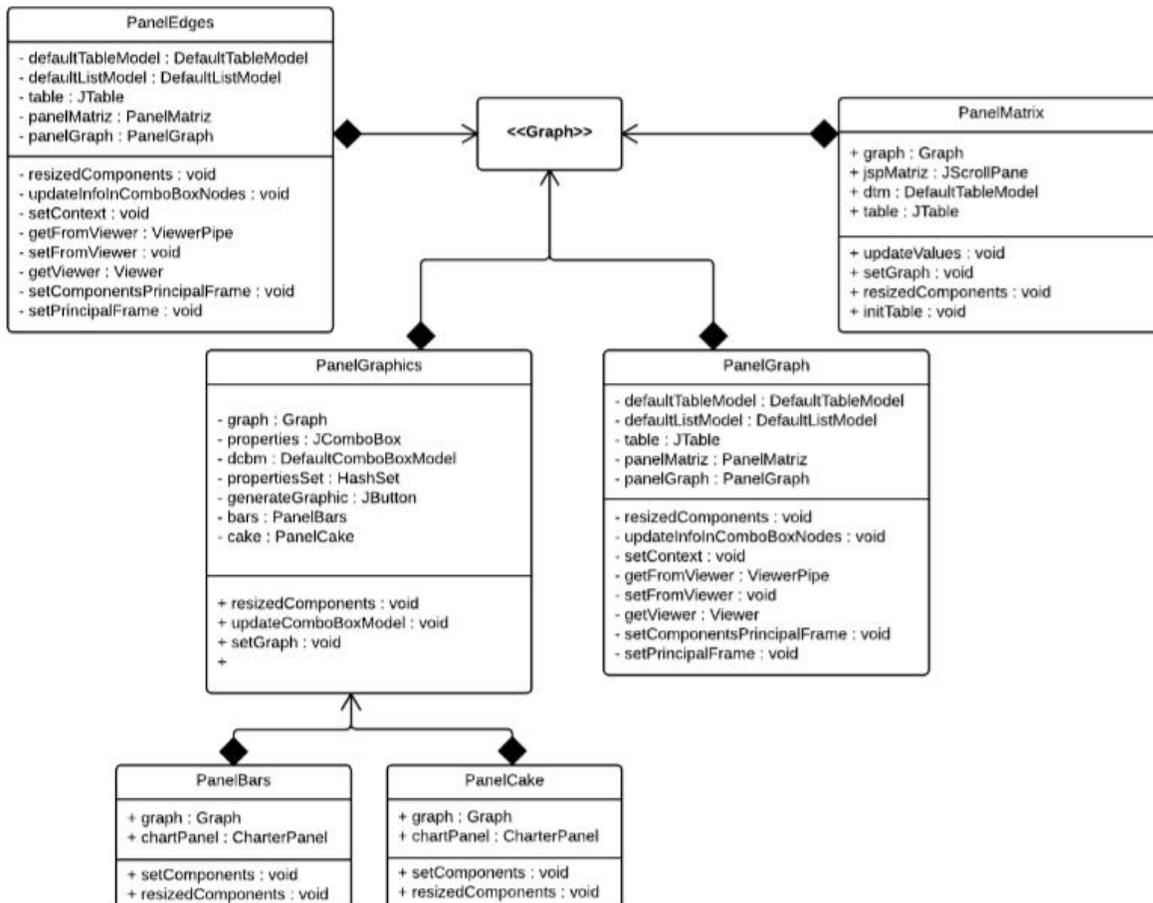


Figura 17: Diagrama de clases (Panels). Fuente: Elaboración propia

7.2 Funcionalidades

7.2.1 Importar Red

La siguiente funcionalidad proporciona al usuario la posibilidad de cambiar la red, seleccionando un archivo en formato JSON con la estructura general definida en la metodología y ejemplificada en la Figura 16.

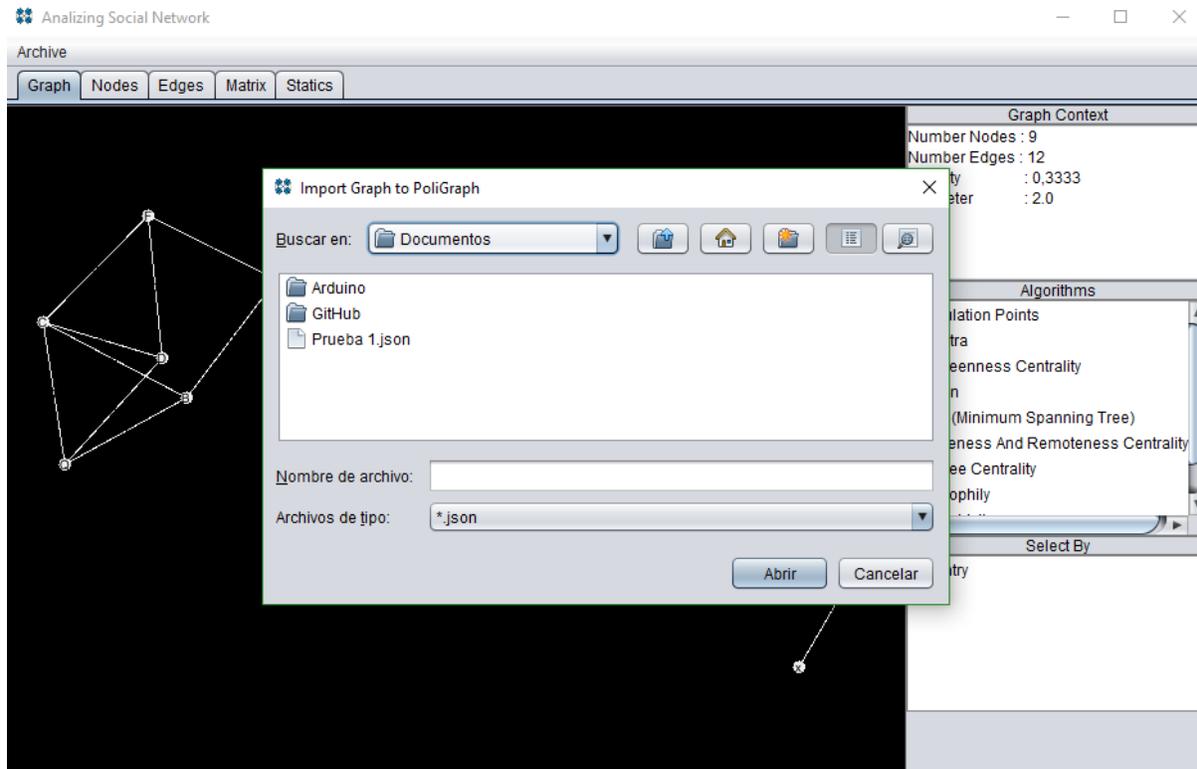


Figura 18: Funcionalidad de importación de una red. Fuente: Elaboración propia

7.2.2 Exportar Red

El usuario también cuenta con la funcionalidad de guardar la red con el formato JSON definido anteriormente, el programa ofrece una ventana la cual ayudara al usuario a guardar este archivo en una ruta especifica.

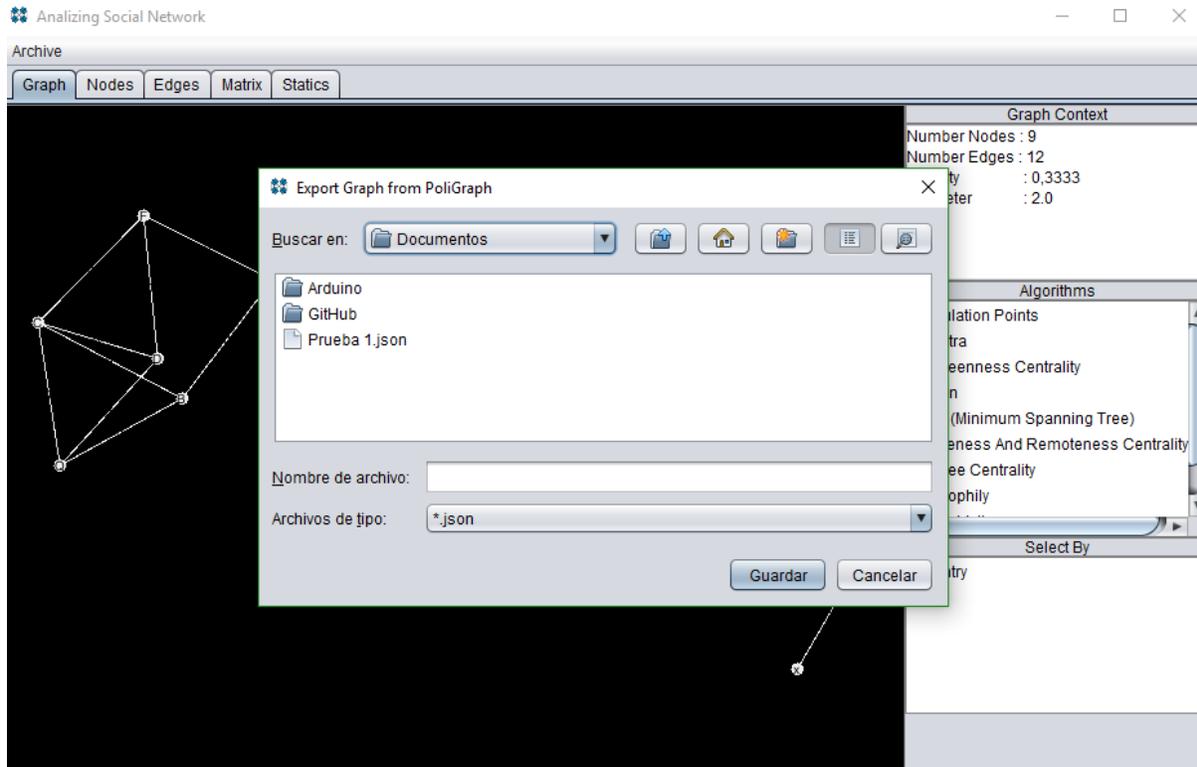


Figura 19: Funcionalidades de exportación de una red, Fuente: Elaboración propia

7.2.3 Vista Gráfica

Esta es la vista principal de la aplicación la cual ofrece una representación gráfica de la red actual, donde los nodos son representados como puntos en un espacio y las adyacencias como líneas entre dos puntos; también ofrece una contextualización de la red que se está visualizando; a continuación, se ofrece una lista con los algoritmos que se podrán ejecutar sobre la red; por último, se ofrece la lista de las características que se encuentran a lo largo de la red como edad, genero, edad etc.

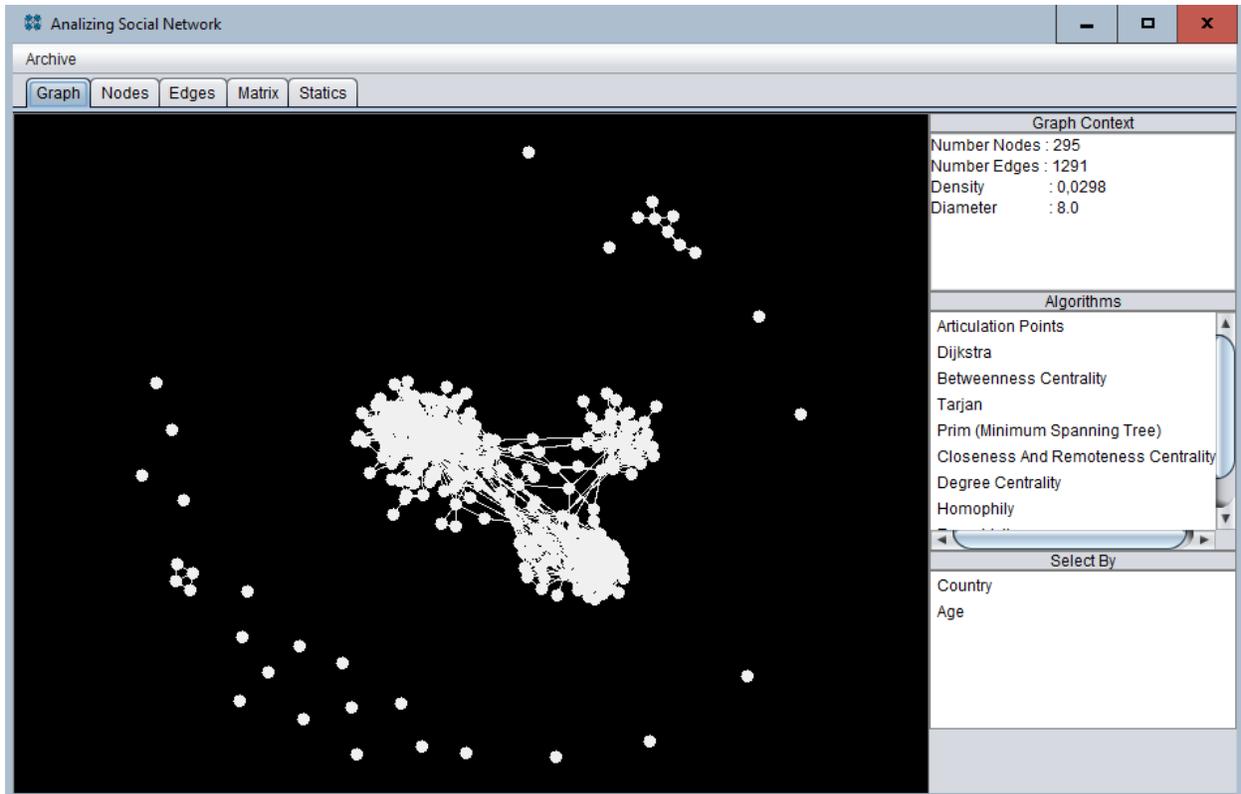


Figura 20 : Vista de la pestaña Graph, Fuente: Elaboración propia

7.2.4 Vista Nodos

Esta vista da un detalle en forma de lista a cerca de los nodos contenidos en una red, ofreciendo información relevante como lo es su grado, edad, país, nombre de usuario, entre otros; esto con el fin de ofrecer mayor información de los datos contenidos. Además, ofrece funcionalidades como, crear, eliminar y editar cualquier nodo en la red.

The screenshot shows the 'Analyzing Social Network' application with the 'Nodes' view selected. The main table displays the following data:

Id Node	Label	Degree	Color	Properties
0	0	14	####	userName : juliancastroruge ; profile : https://www.facebook.com/juliancastroruge?fref=pb&hc_location=friends_tab ; name : Julia
1	1	8	####	userName : ANDDRESKMS5 ; profile : https://www.facebook.com/ANDDRESKMS5?fref=pb&hc_location=friends_tab ; name : Carlos And
2	2	13	####	userName : naniita.velasko ; profile : https://www.facebook.com/naniita.velasko?fref=pb&hc_location=friends_tab ; name : Nan
3	3	8	####	userName : evelin.moralez.9 ; profile : https://www.facebook.com/evelin.moralez.9?fref=pb&hc_location=friends_tab ; name : El
4	4	48	####	userName : jenniferthalia ; profile : https://www.facebook.com/jenniferthalia?fref=pb&hc_location=friends_tab ; name : Jennifer P
5	5	22	####	userName : paula.ortiz.1 ; profile : https://www.facebook.com/paula.ortiz.1?fref=pb&hc_location=friends_tab ; name : Cami
6	6	65	####	userName : juandavid.callejas.52 ; profile : https://www.facebook.com/juandavid.callejas.52?fref=pb&hc_location=friends_tab ; name : Juan D
7	7	6	####	userName : flyer.beltran.9 ; profile : https://www.facebook.com/flyer.beltran.9?fref=pb&hc_location=friends_tab ; name : Flyer B
8	8	78	####	userName : anasoraaidab ; profile : https://www.facebook.com/anasoraaidab?fref=pb&hc_location=friends_tab ; name : Sory Bel
9	9	4	####	userName : juankstro.10 ; profile : https://www.facebook.com/juankstro.10?fref=pb&hc_location=friends_tab ; name : Juan
10	10	1	####	userName : nestor.sanchezsuares ; profile : https://www.facebook.com/nestor.sanchezsuares?fref=pb&hc_location=friends_tab ; name : n
11	11	4	####	userName : fernanda.lopez.7547 ; profile : https://www.facebook.com/fernanda.lopez.7547?fref=pb&hc_location=friends_tab ; name : Fe
12	12	39	####	userName : jhonatan.callejasbeltran ; profile : https://www.facebook.com/jhonatan.callejasbeltran?fref=pb&hc_location=friends_tab ; name : Jho
13	13	8	####	userName : nataly.saenzlopez ; profile : https://www.facebook.com/nataly.saenzlopez?fref=pb&hc_location=friends_tab ; name
14	14	3	####	userName : yulipaola.forero ; profile : https://www.facebook.com/yulipaola.forero?fref=pb&hc_location=friends_tab ; name : Yul
15	15	20	####	userName : marianita.vanegasbeltran ; profile : https://www.facebook.com/marianita.vanegasbeltran?fref=pb&hc_location=friends_tab ; name : M
16	16	2	####	userName : jennip2 ; profile : https://www.facebook.com/jennip2?fref=pb&hc_location=friends_tab ; name : Pao Mu
17	17	14	####	userName : pablo.callejas3 ; profile : https://www.facebook.com/pablo.callejas3?fref=pb&hc_location=friends_tab ; name : Pablo
18	18	25	####	userName : johana.beltran.98 ; profile : https://www.facebook.com/johana.beltran.98?fref=pb&hc_location=friends_tab ; name : J
19	19	14	####	userName : paula.munoz.94043 ; profile : https://www.facebook.com/paula.munoz.94043?fref=pb&hc_location=friends_tab ; name : F
20	20	2	####	userName : ibama.delgado ; profile : https://www.facebook.com/ibama.delgado?fref=pb&hc_location=friends_tab ; name : Luc
21	21	25	####	userName : bsandralucero ; profile : https://www.facebook.com/bsandralucero?fref=pb&hc_location=friends_tab ; name : Luc
22	22	5	####	userName : katherine.hernandez.75641297 ; profile : https://www.facebook.com/katherine.hernandez.75641297?fref=pb&hc_location=friends_tab ; r
23	23	3	####	userName : juancarlos.castroalgaio ; profile : https://www.facebook.com/juancarlos.castroalgaio?fref=pb&hc_location=friends_tab ; name : Ju
24	24	19	####	userName : astridjanneth.callejasalbarracon ; profile : https://www.facebook.com/astridjanneth.callejasalbarracon?fref=pb&hc_location=friends_tab ; nan
25	25	14	####	userName : iola.menjura ; profile : https://www.facebook.com/iola.menjura?fref=pb&hc_location=friends_tab ; name : Lola Me
26	26	5	####	userName : milite.fawn ; profile : https://www.facebook.com/milite.fawn?fref=pb&hc_location=friends_tab ; name : Milite F
27	27	3	####	userName : profile.php?id=100011503234861 ; profile : https://www.facebook.com/profile.php?id=100011503234861&fref=pb&hc_location=friends
28	28	2	####	userName : nicolas.castroruge ; profile : https://www.facebook.com/nicolas.castroruge?fref=pb&hc_location=friends_tab ; name : Nicol
29	29	13	####	userName : laura.rozocallejas ; profile : https://www.facebook.com/laura.rozocallejas?fref=pb&hc_location=friends_tab ; name : Laura
30	30	4	####	userName : juan.cortes.75839923 ; profile : https://www.facebook.com/juan.cortes.75839923?fref=pb&hc_location=friends_tab ; name : Juan
31	31	32	####	userName : jhonatan.callejas.7 ; profile : https://www.facebook.com/jhonatan.callejas.7?fref=pb&hc_location=friends_tab ; name : Jho
32	32	7	####	userName : alex2g1228 ; profile : https://www.facebook.com/alex2g1228?fref=pb&hc_location=friends_tab ; name : Alex
33	33	1	####	userName : dianithap.morenthap ; profile : https://www.facebook.com/dianithap.morenthap?fref=pb&hc_location=friends_tab ; name : Sa
34	34	6	####	userName : sandra.c.pinzon.9 ; profile : https://www.facebook.com/sandra.c.pinzon.9?fref=pb&hc_location=friends_tab ; name : Sand
35	35	2	####	userName : lingridbecerraA ; profile : https://www.facebook.com/lingridbecerraA?fref=pb&hc_location=friends_tab ; name : Kath
36	36	18	####	userName : Angelik.cdim0712 ; profile : https://www.facebook.com/Angelik.cdim0712?fref=pb&hc_location=friends_tab ; name : MAría
37	37	30	####	userName : Derlyzu ; profile : https://www.facebook.com/Derlyzu?fref=pb&hc_location=friends_tab ; name : Derly Lesn

The right-hand panel contains controls for node management: 'Add Node' (with fields for Id and Label), 'Select Node' (with a dropdown menu), and 'Remove Node' (with a 'Remove' button).

Figura 21: Vista de la pestaña Nodes, Fuente: Elaboración propia

7.2.5 Vista de bordes

Esta vista ofrece un listado del conjunto de bordes o relaciones que se encuentran involucrados en la red, la cual ofrece información de relevancia como lo es su id, los nodos sobre los cuales existe la relación, el peso o el costo de la adyacencia y un listado de propiedades. También a su vez cuenta con funcionalidades básicas como lo es crear, editar y borrar una adyacencia, esto en la parte izquierda de la vista.

Id Edge	Node A	Node B	Width	Color	Properties
miller.m...	290	48	1.0	#fff	width : 1.0 ;
miller.m...	290	41	1.0	#fff	width : 1.0 ;
miller.m...	290	237	1.0	#fff	width : 1.0 ;
miller.m...	290	216	1.0	#fff	width : 1.0 ;
miller.m...	290	284	1.0	#fff	width : 1.0 ;
miller.m...	290	8	1.0	#fff	width : 1.0 ;
miller.m...	290	182	1.0	#fff	width : 1.0 ;
miller.m...	290	282	1.0	#fff	width : 1.0 ;
fabian.d...	289	157	1.0	#fff	width : 1.0 ;
fabian.d...	289	209	1.0	#fff	width : 1.0 ;
fabian.d...	289	258	1.0	#fff	width : 1.0 ;
fabian.d...	289	246	1.0	#fff	width : 1.0 ;
fabian.d...	289	221	1.0	#fff	width : 1.0 ;
fabian.d...	289	6	1.0	#fff	width : 1.0 ;
fabian.d...	289	251	1.0	#fff	width : 1.0 ;
fabian.d...	289	93	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	231	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	266	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	166	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	267	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	268	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	287	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	286	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	278	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	269	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	276	1.0	#fff	width : 1.0 ;
lizzeth.m...	288	285	1.0	#fff	width : 1.0 ;
kevin.gir...	287	217	1.0	#fff	width : 1.0 ;
kevin.gir...	287	285	1.0	#fff	width : 1.0 ;
kevin.gir...	287	286	1.0	#fff	width : 1.0 ;
kevin.gir...	287	266	1.0	#fff	width : 1.0 ;
kevin.gir...	287	278	1.0	#fff	width : 1.0 ;
carlaviv...	286	66	1.0	#fff	width : 1.0 ;
carlaviv...	286	217	1.0	#fff	width : 1.0 ;
carlaviv...	286	276	1.0	#fff	width : 1.0 ;
carlaviv...	286	178	1.0	#fff	width : 1.0 ;
carlaviv...	286	231	1.0	#fff	width : 1.0 ;
carlaviv...	286	159	1.0	#fff	width : 1.0 ;
carlaviv...	286	166	1.0	#fff	width : 1.0 ;

Figura 22: Vista de la pestaña Edge, Fuente: Elaboración propia

7.2.6 Estadísticas básicas de la red

Esta vista plasma estadísticas relevantes sobre la red, esto partir de las características que en ella se contienen, es decir, plasma en graficas de torta y barras el conteo de una propiedad previamente definida.

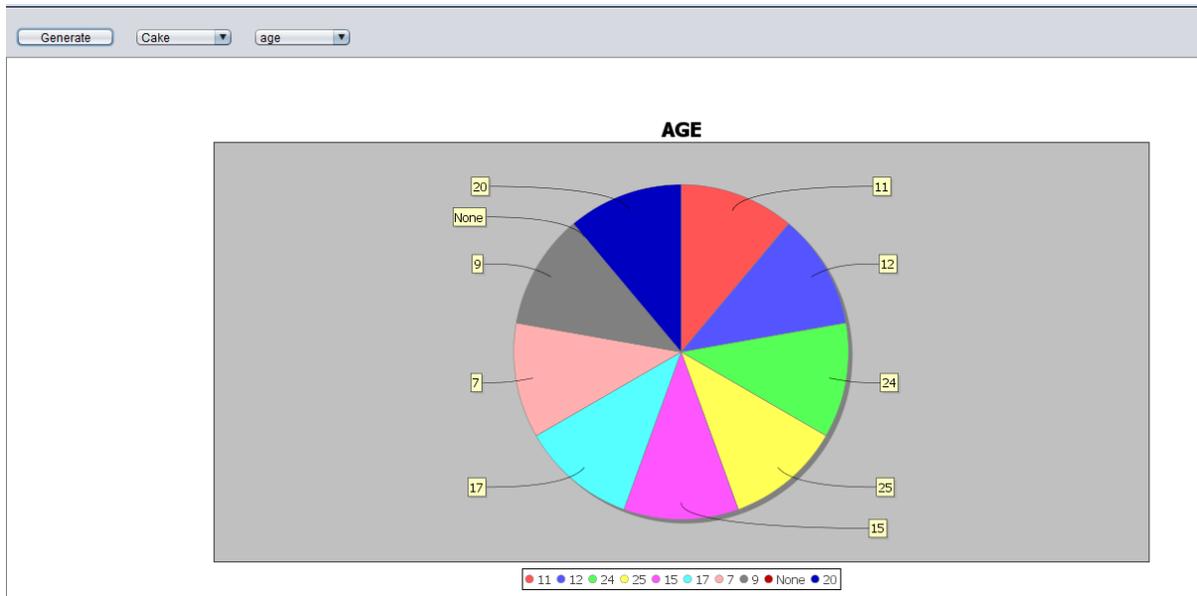


Figura 23: Estadística de edad en un gráfico de torta, Fuente: Elaboración propia



Figura 24: Estadística de edad en un gráfico de barras, Fuente: Elaboración propia

8 Conclusiones y recomendaciones

Como resultado de todo el proceso realizado, comenzando con el planteamiento del problema, pasando por la construcción de una nueva metodología y terminando con el desarrollo de la herramienta a través de la implementación de esta; y además basándonos en los objetivos previamente establecidos se puede llegar a las siguientes conclusiones:

- Se valida que la representación de una red social como una estructura de grafo es acertada y facilita la definición e implementaciones sobre problemas básicos de redes.
- La metodología propuesta probó ser eficiente porque se adaptó a las iteraciones y procesos ejecutados, conduciendo al desarrollo exitoso del proyecto.
- Se recomienda que para un mejor desarrollo de este tipo de herramientas se enfatice en la optimalidad tanto de la representación interna de los datos como de la implementación de los algoritmos, esto debido a la alta complejidad de los sistemas modelados.

9 Bibliografía

- Ávalos Gaytán Vanesa, R. R. (2009). Agrupamiento local en grafos dirigidos. *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*, 8.
- Bijit, P. L. (2010). *Estructuras de Datos y Algoritmos*. Chile.
- Community Structure in Large Networks: Applications, M. a. (2008). *J. Leskovec, K.J. Lang, A. Dasgupta, and M.W. Mahoney*. Berlín, Alemania: Springer Science+Business Media.
- Danah m. Boyd, N. B. (2007). *Social Network Sites: Definition, History, and Scholarship*. Pennsylvania: Journal of Computer-Mediated Communication.
- Elizabeth Costenbader, T. W. (2003). *The stability of centrality measures*. Países Bajos: Elsevier.
- Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 215-239.
- Girvan., M. E. (2004). *Finding and evaluating community structure in networks*. Ann Arbor: Department of Physics and Center for the Study of Complex Systems, University of Michigan.
- Gregory, B. Y. (2009). *Detecting communities in networks by merging cliques*. Australia: University of Bristol.
- Gregory, B. Y. (2009). *Detecting communities in networks by merging cliques*. Bristol: Department of Computer Science, University of Bristol .
- Gregory, S. (2010). Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 27.
- Llano, J. C. (20 de Mayo de 2017). *Juan Carlos Mejía Llano*. Obtenido de Juan Carlos Mejía Llano: <http://www.juancmejia.com/marketing-digital/estadisticas-de-redes-sociales-usuarios-de-facebook-instagram-linkedin-twitter-whatsapp-y-otros-infografia/>
- Patricio Letelier Torre, E. A. (2003). *Metodologías Ágiles en el Desarrollo de Software*. Alicante: JISBD.
- Schaeffer, S. E. (2007). Graph clustering. *Elsevier*, 38.
- Stanley Wasserman, K. F. (1994). *Social Network Analysis: ethods and Applications (Structural Analysis in the Social Sciences)*. Illinois: The Press Syndicate Of The University of Cambridge.
- Strogatz, D. J. (Noviembre 1997). Collective dynamics of 'small-world' networks. *Nature*, 441-442.
- Tiempo, E. (08 de Agosto de 2017). *El Tiempo*. Obtenido de El Tiempo: <http://www.eltiempo.com/tecnosfera/novedades-tecnologia/numero-de-usuarios-de-redes-sociales-en-el-mundo-117556>