

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS
FACULTAD 3



Análisis de influencia social en el contexto universitario

Tesis presentada para optar por el título Ingeniero en Ciencias Informáticas

Autores: Anel Wilfredo Fernández Rodríguez
Jorge Jesús Hidalgo Ruiz

Tutor: Dr. C. Rafael Rodríguez Puente

La Habana, 24 de junio de 2014

Creer en los jóvenes es ver en ellos además de entusiasmo, capacidad; además de energía, responsabilidad; además de juventud, ¡pureza, heroísmo, carácter, voluntad, amor a la patria, fe en la patria! ¡amor a la Revolución, fe en la Revolución, confianza en sí mismos!, convicción profunda de que la juventud puede, de que la juventud es capaz, convicción profunda de que sobre los hombros de la juventud se pueden depositar grandes tareas.

Fidel Castro Ruz



DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste se firma la presente a los 24 días del mes de junio del año 2014.

Anel Wilfredo Fernández Rodríguez

Jorge Jesús Hidalgo Ruiz

Dr. C. Rafael Rodríguez Puente

AGRADECIMIENTOS

*Agradezco a la Revolución Cubana, por brindarme la posibilidad de estudiar y superarme.
Al Comandante en Jefe y líder histórico de la Revolución Fidel Castro Ruz, por haber soñado y creado el hermoso y magnífico proyecto que es hoy la Universidad de las Ciencias Informáticas.
A la FEU, por ser la escuela que nunca termina, a ella le debo gran parte de lo que hoy soy.
A mi mamá y mi papá, los mejores padres del mundo. Sin ellos el camino que he transitado hubiese sido bien difícil, con su eterna compañía y su continuo apoyo todo resultó muy fácil.
A mi bella y amada novia, Martha, mi futura esposa, por todos los lindos momentos vividos.
A mis queridos hermanos, en especial a Tata que siempre ha sido mi ejemplo a seguir por su brillante inteligencia y dedicación al estudio.
A mis queridísimas abuelas, tías, tíos y demás familiares por la ayuda brindada en todo momento, en especial a mi súper sobrino, el Johny, por confiar siempre en su tío.
Al Dr. C. Rafael Rodríguez Puente, mi tutor, por su excelente, impecable y certera guía en la realización de este trabajo.
A Jorge, mi compañero de tesis, por todas las travesuras que vivimos juntos y por haber sido tan comprensivo conmigo en las buenas y en las malas.
A los profesores que ayudaron de una forma u otra en mi formación como profesional, en especial al Ing. Yadian G. Pérez Betancourt, por haberme guiado durante estos fabulosos 5 años.
A los colegas del grupo, que en muchísimas ocasiones necesité de ellos y siempre estuvieron ahí, a Jeandy, el amigo que jamás estuvo ausente cuando lo necesitaba.
A la brigada 3403 de la Universidad de las Ciencias Informáticas por su incondicional ayuda en la realización de este trabajo.
A Rubensín, por haber sembrado en mí el maravilloso mundo de las Matemáticas y la Informática.
A mi hermano e inseparable, a pesar de la gran distancia, amigo Oscarito.*

Anel Wilfredo Fernández Rodríguez

AGRADECIMIENTOS

A la Revolución Cubana por darme la oportunidad de formarme en esta inmensa obra.

A mi mamá Lilian, mi papá Jorge y mi hermana Liliannys por ser los que me motivan en la vida a continuar, por quererme y apoyarme, por siempre estar ahí pues le debo lo que soy sin temor a equivocarme.

A mis abuelos Ricardo, Lucila, Antonio y Gladis por cuidarme desde niño y siempre apoyarme en todo.

A tío Riqui y a tía Yudit por quererme como un hijo y siempre estar al tanto de mí.

A mis tíos paternos, Tomas, Aleida, Alejandro, Iberio, Cuso, Carlos, Julio.

A mis primos.

En general a toda mi familia, pues siempre me apoyaron en todo.

A mi hermano y compañero de tesis Anel, por tantos buenos y malos momentos juntos.

A mis amigos de mil batallas Yunior y Victor.

A la gente de mi grupo Carmen, Yadier, Ariel.

A mi tropa de la FEU que es otra familia más.

A mi tutor por la confianza y todos los consejos.

A mis profesores.

Jorge Jesús Hidalgo Ruiz

DEDICATORIA

A la Revolución Cubana y al Comandante en Jefe Fidel Castro Ruz.

A mis padres.

A mi novia Martha.

A mi sobrino Jonathan.

A mi amigo y hermano Oscarito.

Anel Wilfredo Fernández Rodríguez

A mi mamá.

A mi papá.

A mi hermana.

Jorge Jesús Hidalgo Ruiz

RESUMEN

El análisis de redes sociales es un área de investigación que estudia las redes sociales como grafos. Históricamente aparece como una de las primeras disciplinas en usar la teoría de grafos para hacer ciencia fuera de las matemáticas, conformando una metodología que busca predecir el comportamiento de una red social.

Recientemente la influencia social ha comenzado a atraer más la atención debido a sus importantes aplicaciones. La mayoría de los trabajos sobre esta materia presentan hallazgos cualitativos sobre las influencias sociales.

La presente investigación tiene como objetivo principal, diseñar un modelo para la determinación de influencia social en el contexto universitario a partir de la formalización de patrones de comportamiento de los actores de dicho contexto, que contribuyan a obtener análisis de las interacciones existentes entre los mismos.

Los principales aportes de esta investigación son: la formalización de patrones de comportamiento en el contexto universitario, el diseño de algoritmos que permitan calcular el valor de estos patrones para realizar análisis de influencia social, un modelo extensible, flexible y genérico que contribuya al cálculo de influencia social en dicho contexto; una aplicación informática que permite calcular influencia social, identificar grupos sociales y líderes en temas específicos.

Palabras clave: análisis de redes sociales, contexto universitario, influencia social, patrones de comportamiento.

ÍNDICE GENERAL

INTRODUCCIÓN	1
1 ANÁLISIS DE INFLUENCIA EN REDES SOCIALES: BASES CONCEPTUALES	6
1.1 Marco teórico	6
1.1.1 Teoría de grafos	6
1.1.2 Conceptos básicos de análisis de redes sociales	9
1.2 Herramientas para el análisis de las redes sociales	9
1.3 Análisis crítico de las soluciones existentes	13
1.4 Conclusiones del capítulo	18
2 MODELO PARA EL CÁLCULO DE INFLUENCIA SOCIAL EN EL CONTEXTO UNIVERSITARIO	20
2.1 Base conceptual	21
2.1.1 Análisis de complejidad	24
2.2 Cálculo del valor de los patrones de comportamiento	25
2.2.1 Análisis de complejidad	27
2.2.2 Cálculo de la distribución de temas	28
2.3 Cálculo de influencia social	29
2.4 Conclusiones del capítulo	33
3 VALIDACIÓN DE LA PROPUESTA	35
3.1 Realización computacional del modelo propuesto	35
3.2 Caso de estudio	40
3.3 Aplicación de un sociograma	45
3.4 Comparación del caso de estudio con el sociograma	47
3.4.1 Análisis para el patrón Síndrome del Jabber	50
3.4.2 Análisis para el patrón Participación en Actividades Extracurriculares	51
3.5 Conclusiones del capítulo	52
CONCLUSIONES	54
RECOMENDACIONES	55
REFERENCIAS BIBLIOGRÁFICAS	56
GLOSARIO DE TÉRMINOS	63
ANEXOS	64

ÍNDICE DE FIGURAS

1	Estructura de la tesis	4
2.1	Componentes del modelo propuesto	20
2.2	Ejemplo del uso del enfoque Grafo de Factor Temático	30
3.1	Relaciones sociales existentes entre los actores de la red de entrada	38
3.2	Grafo de influencia social obtenido a partir de la aplicación del modelo para el tema jabber	39
3.3	Red que representa las relaciones sociales existentes en la brigada seleccionada	41
3.4	Grafo de factor temático para el patrón Síndrome del Jabber	42
3.5	Grafo de factor temático para el patrón Participación en Actividades Extracurriculares	43
3.6	Media de las medidas de centralidad de los actores en el tema jabber	45
3.7	Media de las medidas de centralidad de los actores para el tema en actividades extracurriculares	46
3.8	Diagrama de dispersión de las variables X e Y en el patrón Síndrome del Jabber	51
3.9	Diagrama de dispersión de las variables X e Y en el patrón Participación en Actividades Extracurriculares	52

ÍNDICE DE TABLAS

1.1	Evaluación de herramientas de análisis de redes sociales	10
2.1	Notaciones	29
3.1	Historia de Usuario “Construir un grafo a partir de un fichero de texto plano”	36
3.2	Historia de Usuario “Visualizar grafo”	36
3.3	Historia de Usuario “Calcular influencia social”	37
3.4	Tarjeta CRC “ConstruirGrafo”	37
3.5	Tarjeta CRC “VisualizarGrafo”	37
3.6	Tarjeta CRC “CalcularInfluenciaSocial”	38
3.7	Datos de la brigada seleccionada	40
3.8	Datos del caso de estudio utilizado	41
3.9	Valores normalizados de las medidas de centralidad de los actores de la red	44
3.10	Media de las medidas de centralidad para los actores en el tema jabber	44
3.11	Media de las medidas de centralidad para los actores en el tema actividades extracurriculares	45
3.12	Resultado de la aplicación del sociograma	47
3.13	Resultado del sociograma para el patrón Síndrome del Jabber	48
3.14	Resultado del sociograma aplicado en la brigada seleccionada	49
3.15	Valores obtenidos con la aplicación del modelo para del patrón Síndrome del Jabber y el sociograma para dicho patrón	50
3.16	Valores obtenidos con la aplicación del modelo para del patrón Participación en Actividades Extracurriculares y el sociograma para dicho patrón	52

ÍNDICE DE ALGORITMOS

1	<i>CalcularCentralidad</i>	22
2	<i>CalcularCercanía</i>	22
3	<i>CalcularIntermediación</i>	23
4	<i>SíndromeJabber</i>	26
5	<i>SíndromeCorreo</i>	26
6	<i>ParticipaciónActividadesExtracurriculares</i>	27
7	<i>AlgoritmoAprendizajeTAP</i>	32

INTRODUCCIÓN

La creación de asociaciones, sociedades y comunidades es algo constitutivo del ser humano y compartir su conocimiento constituye una característica principal [1]. Hoy en día, Internet es la columna vertebral de la sociedad de la información, ofreciendo innegables posibilidades de intercambio de conocimiento [2].

La aparición de la Web 2.0 permitió la creación de redes sociales interactivas donde las personas o instituciones construyen la información [3]. Las redes sociales se perfilan como el modelo de mayor crecimiento en la actualidad y su estructura ha evolucionado a lo largo de la historia adquiriendo nuevos paradigmas y tipologías [4].

Una red social se define como un conjunto de puntos (actores) vinculados por una serie de relaciones que cumplen determinadas propiedades [5], debiendo poseer o estar vinculadas a ciertas particularidades o rasgos comunes con el fin de poder interactuar entre sí [6]. Para la representación de las redes sociales generalmente se utilizan métodos formales matemáticos, ya que permiten representar las descripciones de las redes de forma compacta y sistemática [7]. Igualmente, las estructuras más conocidas para la representación de sus relaciones, son los grafos y las matrices [8]. La teoría de grafos ha sido aceptada como herramienta de análisis de redes y los algoritmos que se utilizan tienen varios años de uso y aceptación.

En la actualidad, las personas usualmente despliegan sus actividades en la web y las organizaciones emplean aplicaciones informáticas para gestionar las actividades de sus negocios, conformando un ecosistema de relaciones contextuales [9]. Por otra parte, el conocimiento del entorno es algo esencial para percatarse dónde te encuentras, con quién te relacionas, las tendencias, las predicciones, entre otros. Es por esto que el Análisis de Redes Sociales (ARS) es una actividad que está surgiendo como necesaria en los procesos de toma de decisiones a diferentes niveles.

El ARS comienza prestando especial atención al estudio de las estructuras sociales insistiendo menos en por qué los actores hacen lo que hacen y más en la comprensión de las condicionantes estructurales de sus acciones [10]. El ARS estudia la conducta de los individuos a nivel micro, los patrones de

relaciones (la estructura de la red) a nivel macro y las interacciones entre los dos niveles [10].

El análisis de dichas relaciones refleja, en principio, la capacidad de algunas personas de influir en el comportamiento de las demás. El estudio de la influencia no solo resulta difícil de medir, sino que los criterios de influencia varían según el contexto y objetivo de la medición. No obstante, la intensidad de una relación permitiría analizar la influencia que tiene dicha relación sobre la conducta del actor. El análisis de influencia social recae asimismo en la definición de factores y/o conceptos que caracterizan el sentido, la pertinencia, la intensidad, la durabilidad, entre otras de las interacciones.

A partir de un análisis de carácter matemático se han definido varias métricas para representar las interacciones de las redes y sus propiedades (centralidad, intermediación, cercanía, entre otras). Sin embargo, los algoritmos para detectar patrones y otros atributos de las redes no escapan al ruido que provoca el dinamismo y la emergencia de las interacciones, teniendo en cuenta que las personas incursionan y se involucran simultáneamente en varios contextos de interacción pertenecientes a uno o varios dominios.

Las interacciones entre los distintos actores de una red social también se ven reflejadas en el ámbito universitario, a partir de estas interacciones es posible extraer información sobre el comportamiento de los individuos en temáticas específicas. En la actualidad es necesario un marco teórico formal que permita obtener el tipo de información mencionado anteriormente. Esto contribuirá a calcular la influencia que ejercen los distintos actores del ámbito universitario. El análisis de influencia social en este contexto constituye un criterio que aporta información relevante para el proceso de toma de decisiones en la institución.

A partir de la situación descrita se define el siguiente **problema a resolver**:

- La carencia de un marco teórico formal de análisis del modo en que los diversos actores de la vida universitaria construyen sus relaciones e interaccionan entre ellos, dificulta la posibilidad de predecir la influencia que ejercen entre los mismos.

El problema definido se enmarca en el siguiente **objeto de estudio**: indicadores para el análisis de las redes sociales.

El **objetivo general** de la investigación es diseñar un modelo para la determinación de influencia social en el contexto universitario a partir de la formalización de patrones de comportamiento de los actores de dicho contexto, que contribuyan a obtener análisis de las interacciones existentes entre los mismos.

El **campo de acción** es la determinación de influencia social en el contexto universitario.

A partir de un análisis preliminar se enuncia la siguiente **hipótesis**: si se diseña un modelo para la determinación de influencia social en el contexto universitario a partir de la formalización de patrones

de comportamiento de los entes de dicho contexto, entonces se podrá disponer de un marco teórico formal de análisis del modo en que los diversos actores de la vida universitaria interaccionan entre ellos que posibilite predecir la influencia que ejercen los mismos.

El objetivo general se desglosa en los siguientes **objetivos específicos**:

- Definir el marco teórico conceptual a utilizar en la presente investigación.
- Realizar un estudio de las principales tendencias a nivel nacional e internacional sobre el análisis de las interacciones e influencia social en el ámbito universitario.
- Formalizar matemáticamente patrones de comportamiento en el contexto universitario.
- Diseñar los algoritmos necesarios para calcular el valor de estos patrones y para realizar análisis de influencia.
- Implementar un software que calcule el valor de los patrones definidos y que realice análisis de influencia entre los diferentes actores seleccionados para el estudio.
- Validar el software implementado a partir de un caso de estudio.
- Contrastar los resultados obtenidos a partir de la aplicación de la propuesta y de un sociograma a la muestra seleccionada en el caso de estudio.

Para darle cumplimiento a los objetivos trazados se utilizaron los siguientes **métodos de investigación**:

- Métodos teóricos:
 - Analítico-sintético: para descomponer el problema de investigación en elementos por separado, de esta forma se puede profundizar en el estudio de cada elemento, para luego sintetizarlos en la solución de la propuesta.
 - Hipotético-deductivo: para elaborar la hipótesis de investigación y proponer líneas de trabajo a partir de resultados parciales.
 - Histórico-lógico: para llevar a cabo el estudio crítico de las principales tendencias del análisis de interacciones e influencia en las redes sociales, patrones de comportamiento en el ámbito universitario, así como de los algoritmos para calcular el valor de dichos patrones.
 - Modelado: para la implementación de la herramienta computacional y de los algoritmos propuestos.
- Métodos empíricos:

- Criterio de expertos: para identificar los patrones de comportamiento formalizados y diseñar el sociograma aplicado.

La **estructura de la tesis** se expone haciendo uso de una perspectiva de anillo como se puede apreciar en la Figura 1. La misma está concebida a partir del objeto de estudio: Indicadores para el ARS, garantizando de esta forma la consistencia estructural de la tesis que se presenta.

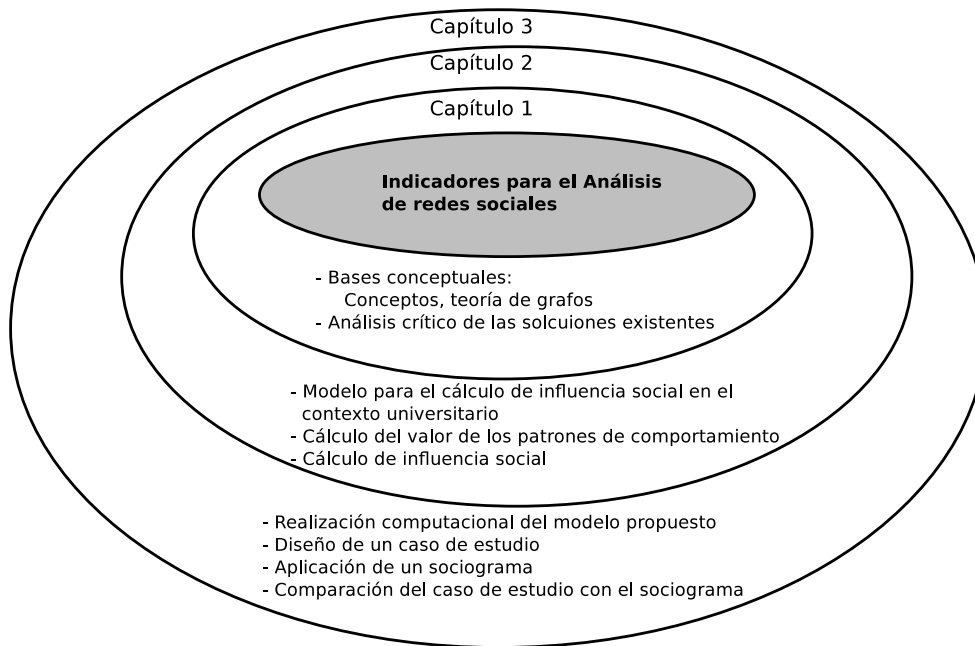


Figura 1: Estructura de la tesis.

Capítulo 1

Análisis de influencia en redes sociales: Bases conceptuales

1. ANÁLISIS DE INFLUENCIA EN REDES SOCIALES: BASES CONCEPTUALES

EN este capítulo se expone el marco conceptual y se realiza un estudio de varios mecanismos que se utilizan en la actualidad para el análisis de influencia en las redes sociales.

1.1. Marco teórico

En la presente sección se exponen conceptos relacionados con la teoría de grafos, el análisis de redes y propiedades básicas de las redes antes mencionadas.

1.1.1. Teoría de grafos

En este epígrafe se presentan definiciones y notaciones utilizadas en los algoritmos que se proponen. Las definiciones que siguen son tomadas de [11].

DEFINICIÓN 1.1.1.1 *Un grafo, o grafo no dirigido, $G = (V, E)$ se define como un conjunto V finito y no vacío de vértices y un multiconjunto E de aristas, donde cada arista $(v_i, v_j) \in E, v_i, v_j \in V$ es un par no ordenado de vértices. Opcionalmente una arista puede tener un valor que la identifique y una lista de atributos. Cuando los elementos de E tienen multiplicidad uno, el grafo se denomina grafo simple.*

La definición de grafo dirigido es similar a la anterior, con la única diferencia que las aristas son pares ordenados, es decir, $E \subseteq V \times V$.

DEFINICIÓN 1.1.1.2 *Sean $G = (V, E)$ y $G' = (V', E')$ dos grafos, si $V' \subseteq V, E' \subseteq E$ se dice que G' es un subgrafo de G y G es un supergrafo de G' .*

DEFINICIÓN 1.1.1.3 *La matriz de adyacencia $A = (a_{ij})_{n \times n}$ del grafo G , donde $n = |V|$, se define de la siguiente forma:*

$$a_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{en otro caso} \end{cases}$$

DEFINICIÓN 1.1.1.4 Un vértice v_j es adyacente (vecino) a otro vértice v_i si $\exists (v_i, v_j) \in E$. Además, se dice que la arista (v_i, v_j) incide en el vértice v_j .

DEFINICIÓN 1.1.1.5 Se denomina grado de un vértice v a la cantidad de aristas que inciden en v y se representa como $g(v)$.

DEFINICIÓN 1.1.1.6 En un grafo dirigido, se denomina grado de entrada de un vértice v a la cantidad de aristas que tienen como vértice final a v y se representa como $g^-(v)$. De forma similar, se denomina grado de salida a la cantidad de aristas que tienen como vértice inicial a v y se representa como $g^+(v)$.

DEFINICIÓN 1.1.1.7 Se define un grafo ponderado como una estructura $G = (V, E, f_c)$, donde:

- V representa el conjunto de vértices del grafo.
- E representa un multiconjunto de aristas del grafo.
- La función $f_c : E \rightarrow \mathbb{R}^+$ le hace corresponder a cada arista un valor real positivo denominado costo.

DEFINICIÓN 1.1.1.8 La matriz de costo $C = (c_{ij})_{n \times n}$ de un grafo ponderado $G = (V, E, f_c)$, donde $n = |V|$, se define como:

$$c_{ij} = \begin{cases} k & \text{si } \exists (v_i, v_j) \in E \text{ y } k \text{ es el costo de ir desde } v_i \text{ hasta } v_j, k > 0, k = f_c(v_i, v_j) \\ 0 & \text{en otro caso} \end{cases}$$

DEFINICIÓN 1.1.1.9 Se denomina camino desde el vértice v_i al vértice v_j en un grafo $G = (V, E)$ a la secuencia de vértices $CA = v_{k_1}, v_{k_2}, \dots, v_{k_t}$ si $\exists (v_{k_1}, v_{k_2}), (v_{k_2}, v_{k_3}), \dots, (v_{k_{t-1}}, v_{k_t}) \in E, v_i = v_{k_1}, v_j = v_{k_t}$. Además, una arista solo puede aparecer una vez.

DEFINICIÓN 1.1.1.10 Se denomina longitud de un camino $CA = v_{k_1}, v_{k_2}, \dots, v_{k_t}, t > 1$, a la suma de los costos de todas las aristas presentes en el mismo y se representa como $|CA| = \sum_{h=1}^{t-1} c_{k_h k_{h+1}}$. En el caso de grafos no ponderados, la longitud del camino se puede calcular como la cantidad de aristas presentes en el mismo, o sea $|CA| = t - 1$.

La centralidad basada en vértices se define con el fin de medir la importancia de un vértice en la red. Esta ha llamado mucho la atención como una herramienta para el estudio de las redes sociales. Un

vértice con alta centralidad suele ser considerado más altamente influenciable que otros vértices de la red [12, 13].

A continuación se muestran varios conceptos definidos en la literatura [14].

DEFINICIÓN 1.1.1.11 *Sea A la matriz de adyacencia de un grafo G , y $g(v_i)$ el grado del vértice v_i , entonces el grado de centralidad c_i^{DEG} del vértice v_i se define:*

$$c_i^{DEG} = g(v_i)$$

Una manera de interpretar el grado de centralidad de un vértice v_i es contar el número de caminos de longitud 1 que terminan en dicho vértice. Una generalización natural de este punto de vista es la centralidad *camino* – K que es el número de caminos de longitud menor o igual que K que terminan en el vértice v_i .

DEFINICIÓN 1.1.1.12 *La medida de cercanía, definida por el matemático Murray Beauchamp en 1965 [15] y luego popularizada por Freeman en 1979 [16], es la más conocida y utilizada de las medidas radiales de longitud. Se basa en calcular el promedio de las distancias más cortas desde un vértice hacia todos los demás. Formalmente, la cercanía se define como:*

$$c_i^{CLO} = e_i^T S$$

donde e_i^T es la transpuesta de un vector columna cuyo i -ésimo elemento es 1 y el resto son 0, S es la matriz cuyos elementos en la posición (i, j) corresponden a la distancia más corta desde el vértice v_i al vértice v_j .

DEFINICIÓN 1.1.1.13 *La intermediación es una medida que cuantifica la frecuencia o el número de veces que un vértice actúa como un puente a lo largo del camino más corto entre otros dos vértices. De manera formal, la intermediación se define como:*

$$c_i^{BET} = \sum_{j,k} \frac{b_{jik}}{b_{jk}}$$

donde b_{jk} es la cantidad de caminos mínimos desde v_j a v_k y b_{jik} la cantidad de caminos mínimos desde v_j a v_k que pasan por v_i .

1.1.2. Conceptos básicos de análisis de redes sociales

Los conceptos que se enuncian a continuación son tomados de [17].

DEFINICIÓN 1.1.2.1 *Se denomina actor a la entidad social que interactúa con otras entidades. Esta puede ser una persona, un grupo de individuos, una organización, una ciudad, una nación, entre otros.*

DEFINICIÓN 1.1.2.2 *Un vínculo o conexión corresponde al tipo de relación que conecta un par de actores distintos. Estos vínculos pueden ser de naturaleza diferente: relaciones transnacionales, de comunicación, instrumentales, sentimentales, de poder o autoridad, de parentesco, entre otros [18].*

DEFINICIÓN 1.1.2.3 *Una díada se compone de un par de actores y la posible relación existente entre ellos. Varios estudios parten de la díada como unidad fundamental, analizando para cada par de actores, si el vínculo es recíproco o no.*

DEFINICIÓN 1.1.2.4 *Se denomina tríada al conjunto de tres actores y las posibles relaciones existentes entre ellos. Una propiedad interesante de una tríada es la existencia o no, de transitividad entre sus miembros. Es decir, comprobar que si el actor i tiene una relación con el actor j y a su vez el actor j tiene una relación con el actor k , entonces el actor i también tiene una relación con el actor k .*

DEFINICIÓN 1.1.2.5 *Se denomina grupo social a un conjunto finito de actores con características o intereses similares.*

DEFINICIÓN 1.1.2.6 *Una red social se define como un conjunto finito de actores que tienen un determinado número de relaciones definidas entre ellos. Una red social es modelada como un grafo $G = (V, E)$. En la presente investigación, el conjunto de vértices V se corresponde a los actores y el conjunto de aristas E a las relaciones sociales.*

1.2. Herramientas para el análisis de las redes sociales

Navarro y Salazar realizaron una evaluación de varias herramientas para el análisis de las redes sociales [19], basándose en los siguientes indicadores:

- **Dis.** Disponibilidad del software: gratis, pagado o shareware. Los valores están en USD.

- **Int.** Interfaz: claridad de la interfaz de usuario.
- **Mét.** Métricas: cantidad y facilidad de uso de la métricas y sus resultados.
- **Imp./Exp.** Importar/Exportar: capacidad para importar o exportar otros formatos de archivo, ya sea de otra herramienta de análisis, Excel u otros.
- **Gráf.** Gráfica: Capacidad para graficar redes.
- **Docs.** Documentación: disponibilidad y claridad de la documentación.
- **Actual.** Frecuencia de actualización del software y nuevas versiones recientes.
- **Total.** Promedio aritmético de los parámetros.

En la Tabla 1.1 se muestra la evaluación realizada en [19] teniendo en cuenta los indicadores anteriores.

Tabla 1.1: *Evaluación de herramientas de análisis de redes sociales. Tomado de [19]*

Nombre	Disp.	Int.	Met.	Imp./Exp.	Gráf.	Docs.	Actual	Total
Anthropac	Desde 30	3	4	2	3	5	3	3.33
Blanche	Gratis	4	3	4	4	5	4	4
FATCAT	Gratis	3	3	2	2	4	2	2.66
Iknow	Gratis	4	3	2	2	4	3	3
Inflow	-	4	4	*	3	*	4	3.75
Krackplot	Gratis	3	3	3	3	5	2	3.16
MultiNet	-	4	3	3	3	4	4	3.5
Negopy	Shareware	1	3	2	1	5	3	2.5
Netdraw	Gratis	5	2	5	5	3	5	4.16
NetMiner	Desde 35	5	5	5	5	5	5	5
NetViz	Gratis	4	4	3	4	4	4	3.83
Pajack	Gratis	4	5	4	4	4	5	4.33
Stocnet	Gratis	4	3	4	2	5	5	3.83
UCINET	Desde 40	5	5	5	5	5	5	5
Visone	Gratis	4	4	4	4	3	4	3.83

A continuación se presenta una breve reseña de varios sistemas informáticos utilizados para la investigación de redes sociales.

UCINET¹: la herramienta está concebida para el Sistema Operativo Windows y es una de las herramientas computacionales con más difusión en el ARS [20]. UCINET es un paquete de

¹UCINET: <http://www.analytictech.com/ucinet/>

herramientas que cumple roles diferentes y complementarios. A continuación se describen los tres módulos principales.

- **UCINET**: módulo central que calcula los indicadores del ARS y que permite acceder a los otros programas. Tiene una amplia gama de rutinas y algoritmos de cálculo y operaciones sobre matrices.
- **Spreadsheet**: módulo que permite capturar los datos mediante matrices de adyacencia o de atributos y cuenta con herramientas para el análisis matricial. En general se utiliza Spreadsheet siempre que se quiera modificar una matriz.
- **NetDraw**²: módulo que asume la función de graficar las redes sociales. Las etapas del ARS a través de estas herramientas informáticas se pueden resumir en las siguientes:
 - Recolección de datos; por ejemplo, a través de encuestas o datos ya existentes.
 - Transformación de esos datos en datos que relacionan a los actores que van a componer la red.
 - Generación de la matriz de adyacencia a partir de los datos.
 - Generación opcional de matrices de atributos para complementar el estudio.
 - Captura de las matrices en Spreadsheet.

Los indicadores del ARS que pueden ser utilizados en dicha aplicación, se encuentran:

- Densidad de la red.
- Matriz de distancias geodésicas.
- Grado de centralidad.
- Grado de intermediación.
- Poder de Bonacich.

Pajek³: desde su creación, a finales de 1996, es utilizado principalmente para ARS, ya que representa una variada gama de opciones para manipular, transformar y analizar grafos de manera fácil y dinámica.

Por otro lado, al comparar su capacidad con UCINET, se debe destacar que si bien PAJECK puede procesar una mayor cantidad de datos, no permite realizar operaciones de escalado multidimensional, lo que produce la diferenciación entre ambas herramientas según la índole de las redes que se deseen estudiar [21].

²NetDraw: <https://sites.google.com/site/netdrawsoftware/>

³Pajek: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

NetMiner⁴: es una herramienta innovadora para análisis de exploración y visualización de datos en la red [22]. Facilita la detección de patrones subyacentes y las estructuras presentes en la red. Puede ser empleada para la investigación en general, la enseñanza y el análisis profesional de las redes sociales. Además, se puede aplicar eficazmente a diversos campos de negocio, donde los factores estructurales de la red pueden influir críticamente en el rendimiento; por ejemplo: intra e inter-organizacional, financiero, web, inteligencia criminal, las telecomunicaciones, las redes de transporte [22].

Egonet [23]: es una herramienta para la recolección de datos y análisis de redes egocéntricas, creado por Christopher McCarthy⁵. Se trata de una herramienta especialmente orientada a la recolección y análisis de redes egocéntricas o personales. Sus principales prestaciones son la generación de cuestionarios para la obtención de datos, el cálculo de mediciones generales pertinentes para el ARS y la generación de grafos. El programa está disponible en código abierto y su uso es gratuito. Egonet se encuentra escrito en Java y está compuesto por los siguientes módulos:

- **EgoNetW**: utiliza el mismo formato de UCINET. Permite crear formatos de cuestionarios para su posterior estudio.
- **EgoNetClientW**: se utiliza para la carga de datos una vez definidas las preguntas pertinentes y la estructura de los cuestionarios [23].

Las herramientas web para la medición de influencia social proponen varios factores a tener en cuenta. Según el índice de Klout⁶ se proponen los siguientes [24]:

- **Alcance real**: número de personas sobre las que se ejerce influencia. Responde a la pregunta ¿en cuántas personas influyes? Para ello analizan cómo las personas con las que un usuario está conectado interactúan con él y con su contenido.
- **Amplificación**: ¿cuánto influyes en la persona sobre la que tienes influencia? Se mide tomando en cuenta cómo los mensajes provocan conversación o son replicados (retuiteados, marcados como “me gusta” en Facebook, merecedores de un +1 en Google+, entre otros). Cada vez que uno de sus seguidores realiza una de estas acciones es considerada como una señal de autoridad y calidad del contenido del usuario.
- **Impacto en la red**: determina la influencia que a su vez tienen los seguidores que están dentro del alcance real del usuario evaluado. Mide, por tanto, la calidad de sus seguidores (¿cuán influyentes son ellos?).

⁴NetMiner: <http://www.netminer.com/index.php>

⁵PhD. Christopher McCarthy. Sitio web: <http://www.bebr.ufl.edu/facultystaff/chrism>

⁶Klout: <http://klout.com/home>

Por otro lado, PeerIndex⁷ busca medir la reputación del usuario en los temas específicos en torno a los que desempeña su actividad en línea:

- **Autoridad:** la medida de la confianza, el modo en que las demás personas confían en las recomendaciones y opiniones del usuario evaluado, de modo general y en temas concretos. Es una posición relativa respecto al resto de las personas en un determinado tema. Se calcula en una escala de 0 a 100 (una cifra superior a 40 coloca al usuario entre las autoridades de un tema).
- **Audiencia:** equivalente a los parámetros de alcance real y amplificación de Klout, puesto que mide el impacto de las acciones del usuario sobre sus seguidores y la manera en que éstos atraen con su contenido.
- **Actividad:** representa la frecuencia de actualización del status del usuario en relación con el resto de la comunidad PeerIndex en torno a un tema. Según sea la naturaleza del tema (y exija una conversación en torno a él), una actividad excesiva podría considerarse spam y penalizar al usuario, o viceversa: que su intervención en el debate sea tan esporádica que llegue a ser irrelevante.

Las herramientas antes mencionadas implementan alrededor de 30 algoritmos como parte de sus principales funcionalidades, de los cuales no es pública su implementación. Por tanto, hay una clara necesidad de métodos y técnicas para analizar y cuantificar influencia social.

1.3. Análisis crítico de las soluciones existentes

En estudios pioneros realizados en [25] y [26], se detectó que en todos los grupos sociales existen individuos particularmente activos, más sensibles que el resto a los intereses del propio grupo y más proclives a expresar abiertamente sus opiniones. Los autores constataban la primacía de la influencia personal por encima de la influencia de la comunicación de masas en diferentes ámbitos de decisiones. Por otro lado, varias teorías de la sociología muestran que el efecto de la influencia social desde varios ángulos (temas) puede ser diferente. En la comunidad de investigación, tales influencias son bien conocidas. La mayoría de los investigadores están influenciados por otros en términos de colaboración y citas. Las informaciones más importantes en la comunidad de investigación son las redes de co-autoría, que capturan la dinámica social de la comunidad y sus publicaciones, lo que implica la distribución de temas (intereses) de los autores [27, 28].

⁷PeerIndex: <http://peerindex.com/>

En estudios realizados en [29] sobre la influencia del entorno socioeconómico en la realización de estudios universitarios en España, se aprecia que tanto el nivel educativo del padre como el de la madre resultan determinantes a la hora de explicar la demanda de educación universitaria de los jóvenes. En este sentido, las estimaciones apuntan a que el tener padres con bajos niveles educativos influye como una barrera cultural en la realización de estudios universitarios de sus hijos.

Según [30] la influencia social en los grupos de interés puede traducirse por el concepto de posición (cuanto más cerca se esté del centro de decisión, mayor efectividad tendrá la estrategia del grupo de interés y mayor éxito se logrará). Mientras que otros autores recurren al concepto de élite y aseguran que se debe descubrir quién, a la vista de los que toman las decisiones, son las personas u organizaciones que más se tienen en cuenta a la hora de aprobar una política.

Con el nacimiento de la web se crean nuevos campos de investigación en los que surge el interés del análisis de influencia. El surgimiento y la rápida difusión de aplicaciones sociales y los medios de comunicación, como la mensajería instantánea (AIM, MSN, Jabber), los sitios de intercambio (Flickr, Picassa, YouTube), blogs (Blogger, WordPress, LiveJournal), enciclopedias en línea (Wikipedia, PBWiki), microblogs (Twitter, Jaiku), redes sociales (MySpace, Facebook), por mencionar algunos, muestran que la influencia social es cada vez más una fuerza predominante, compleja y sutil que rige la dinámica de las redes sociales [14, 31].

Christakis y Fowler estudiaron el efecto de la influencia social en temas relacionados con la salud: como el consumo de alcohol [32], la obesidad [33], el tabaquismo [34], problemas al dormir [35], la soledad [36] y la felicidad [37]. En estos estudios, utilizan datos que abarcan 12 000 personas aproximadamente y se interrelacionan el estado de salud y la estructura de red social en un período de 32 años.

El análisis de Christakis y Fowler argumenta que, incluso teniendo en cuenta los efectos de la selección y las variables de confusión, la evidencia sugiere que el estado de salud puede estar influenciado por el estado de salud de los vecinos.

Esto evidencia que un problema central en el análisis de influencia social es comprender la interacción entre la similitud de las relaciones sociales [38]. En varias investigaciones se ha tratado de identificar la influencia y la correlación en las redes sociales desde diferentes aspectos: semejanza social e influencia [38, 39], comercialización a través de influencia social [40, 41], maximización de influencia [42], modelo de influencia social a través de la conformidad, el cumplimiento y la obediencia [43, 44] e influencia social en las comunidades virtuales [45, 46].

Holme y Newman propusieron un modelo generativo para equilibrar los efectos de la selección y

la influencia. Los resultados de la selección tienden a generar un gran número de pequeños grupos, mientras que la influencia social generará grandes grupos coherentes. Cada actor en el modelo de Holme-Newman en un momento dado solo puede tener una opinión, lo cual puede ser un obstáculo en las redes sociales reales [47].

Para hacer frente a esta limitación, Crandall et al. introdujeron vectores de opinión multidimensionales en modelos de redes sociales complejas. Similar al modelo de Holme-Newman, Crandall et al. también proponen un modelo generativo teniendo en cuenta las muestras de las actividades de una persona en función de su propia historia y la historia de sus vecinos [38].

El modelo de Crandall et al. requiere varios parámetros, por tanto se necesitan más datos para conocer los mismos. Estos modelos son todavía poco concluyentes.

Anagnostopoulos et al. tratan de diferenciar la influencia social de las variables de confusión proponiendo la prueba de reversión de bordes. Dicha prueba se basa en el hecho de que dos amigos a menudo comparten características comunes o se ven afectados por las mismas variables externas y estas son independientes. Por tanto, la reversión de bordes no cambiará la estimación de la correlación social [39].

Por otro lado, la influencia social se extiende en la dirección especificada por los bordes del grafo y por tanto la reversión de bordes debe cambiar de manera intuitiva la estimación de la correlación. Anagnostopoulos et al. prueban sus modelos usando datos de Flickr y valida la influencia social como fuente de la correlación entre las acciones de los individuos con relaciones sociales.

Las pruebas propuestas por [39] asumen una red estática, lo cual es cierto en muchas redes sociales reales. LaFond y Neville proponen una prueba de aleatorización diferente con el uso de un modelo de autoregresión relacional. Específicamente, proponen modelar la red social como un grafo de tiempo-evolución [48].

Aral et al. proponen un modelo de difusión para diferenciar la selección y la influencia social, y desarrollan un marco de estimación de muestra equivalente para distinguir influencia en redes dinámicas [49].

La influencia es generalmente reflejada en cambios en las acciones (comportamiento de usuario) en la red social. Varios trabajos [50, 51] han estudiado el problema de aprender el grado de influencia de las acciones históricas de los usuarios, mientras que otros [52, 53] investigan cómo las acciones sociales evolucionan en el contexto de la red y cómo estas acciones se ven afectadas por factores sociales de influencia.

Goyal et al. estudian el problema de aprendizaje de los grados de influencia (denominadas

probabilidades) de un registro histórico de las acciones del usuario. La principal ventaja de los métodos propuestos por Goyal et al. es que el modelo es escalable y es eficaz para una red social. Una limitación es que no tiene en cuenta la correlación entre las acciones del usuario y los atributos asociados a los mismos [50].

Manavoglu et al. proponen un enfoque basado en el aprendizaje de la conducta individual (acciones) de usuarios de la Web, donde un modelo de comportamiento es un modelo probabilístico que describe las acciones que el usuario va a realizar en el futuro. Primero construyen un modelo de comportamiento global para toda la población y luego personalizan este modelo global para los usuarios existentes mediante la asignación a cada uno de ellos de un valor numérico, denominado peso [54].

Con esto se elimina uno de los problemas de personalización, que es la falta de información sobre cada individuo. Esto se consigue a partir de un modelo global y la optimización de los pesos para cada individuo con respecto a la cantidad de datos disponibles para el mismo.

La influencia puede ser también reflejada por las interacciones entre los usuarios. Por lo general, las comunidades en línea contienen información adicional acerca las interacciones de ellos. Por ejemplo, un usuario de Facebook tiene una página donde sus amigos pueden enviar mensajes. Sobre la base de los mensajes publicados en el muro, se puede inferir cuáles amigos están cerca y quiénes son solo conocidos. Del mismo modo, se puede utilizar en Twitter para inferir la fuerza de una relación. Xiang et al. proponen un modelo de variable latente para inferir la fortaleza de la relación. El modelo representa la causalidad intrínseca de las interacciones sociales a través de las dependencias estadísticas [51].

Varios autores se refieren al problema de maximización de la influencia social por lo general enfocado a las campañas de marketing viral. El primer tratamiento algorítmico del problema fue proporcionado por Domingos y Richardson, que modelaron el proceso de difusión en términos de campos aleatorios de Markov y propusieron soluciones heurísticas para dicho problema [40, 41].

Posteriormente, Kempe et al. estudian la maximización de la influencia como un problema de optimización discreto centrado en dos modelos de propagación fundamentales, denominados Modelo de Cascada Independiente y Modelo de Umbral Lineal. En estos dos modelos, en una fecha y hora determinada, cada vértice puede estar activo o inactivo y la tendencia de cada vértice para convertirse en activo aumenta monótonamente a medida que sus vecinos se activan. Un vértice activo nunca se desactiva nuevamente [42].

Aunque simple, el algoritmo es computacionalmente costoso. Kempe et al. ejecutan simulaciones de Monte Carlo varias veces hasta obtener una estimación precisa de la extensión prevista. Ejecutar

muchas simulaciones de propagación es extremadamente costoso en redes que involucren un elevado número de actores.

Leskovec et al. estudian el problema de propagación desde una perspectiva diferente: ¿cómo seleccionar los actores de una red con el fin de detectar la propagación de un virus lo más rápido posible? Desarrollan un algoritmo eficiente basado en una optimización lazy-forward en la selección de nuevas semillas, aproximándose a ubicaciones óptimas, mejorando el tiempo de ejecución del algoritmo voraz simple [55]. A pesar de esta gran mejora con respecto al algoritmo voraz simple, su método todavía no es escalable como se muestra en [40].

Los métodos anteriores pueden ser utilizados para realizar análisis de influencia social así como maximizarlas. Sin embargo, no distinguen la influencia desde diferentes temas. Esto se evidencia en la observación de los fenómenos sociales desde distintas perspectivas. Por ejemplo, los compañeros de trabajo tienen una fuerte influencia en el oficio de un actor, mientras que sus amigos tienen una fuerte influencia en la vida diaria de dicho actor. Por tanto, existe una dificultad en cuanto a la diferenciación de las influencias sociales basadas en temas.

Varios autores han estudiado el análisis de influencia social basado en temas. Weng et al. analizan la influencia por temas en Twitter por medio de un proceso de dos pasos. En primer lugar, los temas de interés para cada usuario se extraen y luego se construyen redes de relación de temas específicos. Entonces, con el fin de medir la influencia de cada usuario proponen TwitterRank, una extensión del algoritmo PageRank teniendo en cuenta tanto la similitud por temas como la relación social [56].

Tang et al. introducen el problema del análisis de influencia social basado en temas. Dada una red social y una distribución de temas para cada actor, el problema consiste en encontrar las subredes sobre temas específicos y los valores de influencia por cada uno de los temas entre los miembros de las subredes. Para ello proponen el enfoque Grafo de Factor Temático (TFG) [31].

Este enfoque realiza un análisis similar al que se pretende hacer en la presente investigación. Sin embargo, el cálculo de influencia se realiza de manera genérica, por lo que no tiene en cuenta el modo en que los diversos actores de la vida universitaria construyen las relaciones específicas de este contexto e interaccionan entre ellos. Además, no se describe cómo calcular la distribución de temas para los actores de la red.

Es por ello que se propone el uso del enfoque TFG para realizar una parte del análisis que se desea hacer en el presente trabajo, así como incluir los patrones de comportamiento del ámbito universitario para representar los temas.

1.4. Conclusiones del capítulo

Luego de realizar un estudio de los conceptos básicos del análisis de redes, las medidas de centralidad en las redes y las herramientas existentes para el ARS, se concluye lo siguiente:

- El ARS posibilita describir la conducta de los individuos de una red, así como mostrar la capacidad de algunas personas de influir en el comportamiento de los demás, por lo que constituye una actividad efectiva en los procesos de toma de decisiones.
- Las tendencias estudiadas para el ARS no tienen en cuenta las relaciones sociales en el contexto universitario para analizar y cuantificar la influencia social.
- Existe la necesidad de formalizar temas que contribuyan al cálculo de influencia social en el contexto universitario.

Capítulo 2

Modelo para el cálculo de influencia social en el contexto universitario

2. MODELO PARA EL CÁLCULO DE INFLUENCIA SOCIAL EN EL CONTEXTO UNIVERSITARIO

EN este capítulo se presenta un modelo para el cálculo de influencia social en el contexto universitario a partir de patrones de comportamiento. Haciendo uso del mismo se calcula la probabilidad que tiene un actor de influir sobre los demás en un tema determinado.

En la Figura 2.1 se muestra un esquema que representa los componentes del modelo que se propone.

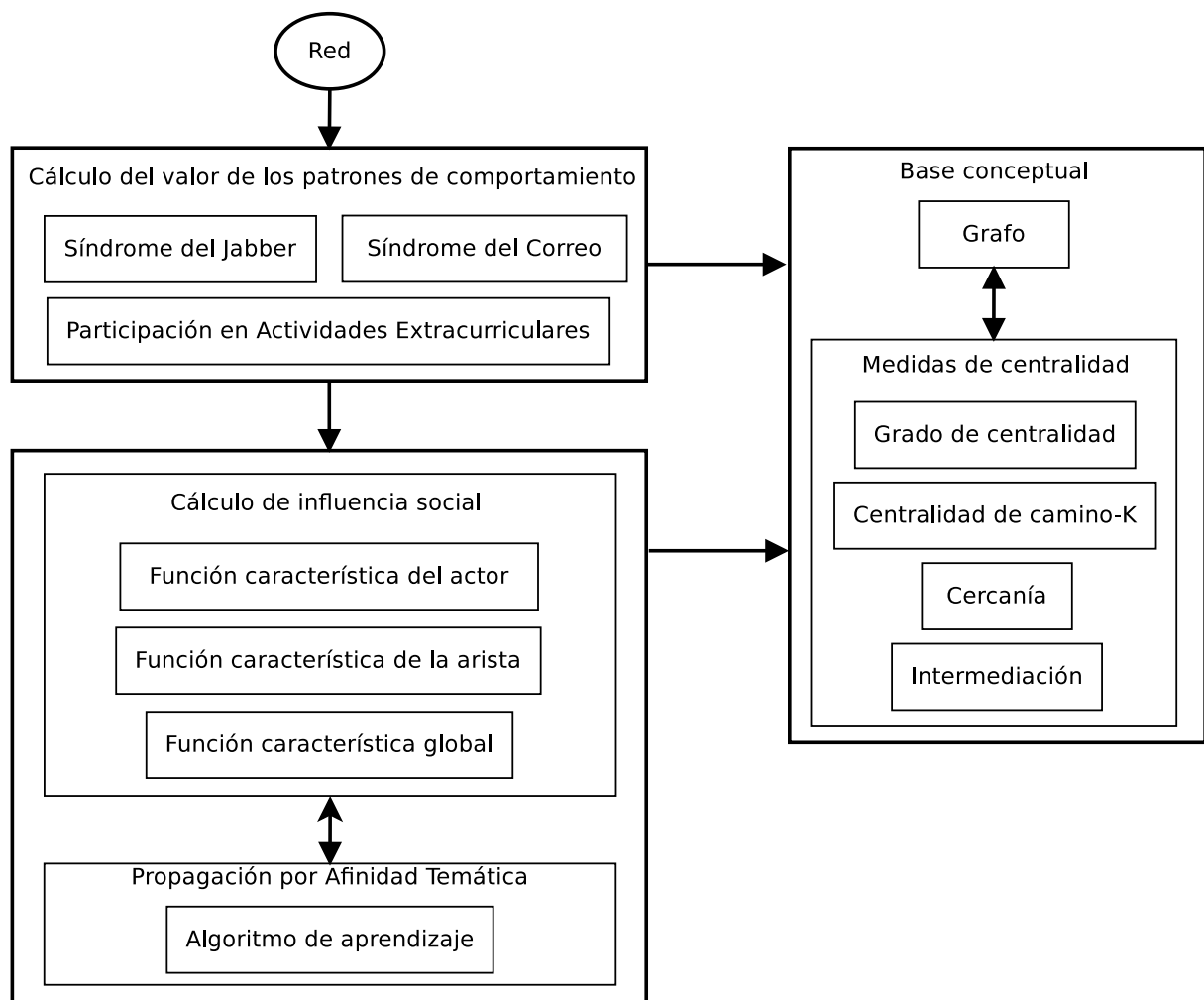


Figura 2.1: Componentes del modelo propuesto.

A continuación se expone en detalle cada uno de los componentes de este modelo así como las relaciones existentes entre los mismos.

El modelo propuesto recibe como entrada una red que representa un conjunto de actores y las relaciones sociales existentes entre ellos. Dicha red será modelada utilizando un grafo ponderado. Luego se procede a calcular el valor de los patrones de comportamiento identificados por los autores de este trabajo: síndrome del jabber, síndrome del correo y participación en actividades extracurriculares; que tienen lugar contexto universitario.

Los patrones de comportamiento son atributos que distinguen las relaciones entre individuos en contextos específicos, en los cuales estos atributos tienden a sistematizarse atendiendo a los escenarios [57]. En el análisis de influencia social que se realizará es necesario caracterizar las relaciones entre los actores de la red, lo cual se garantiza con el uso de patrones de comportamiento como se describió anteriormente.

En la presente investigación no se pretende caracterizar el entorno universitario, sino realizar el análisis de influencia a partir de algunos de los patrones de comportamiento que pueden ser identificados en este contexto.

Estos patrones representarán varios de los temas que pueden ser generados por los actores de la red. Para poder realizar el cálculo de los mismos, los vértices del grafo de entrada deben contener la información referente a las variables que determinan su comportamiento (cantidad de mensajes enviados y recibidos, cantidad de correos enviados y recibidos, tiempo destinado al trabajo con el cliente de mensajería instantánea, tiempo destinado al trabajo con el cliente de correo, cantidad de actividades extracurriculares en las que el actor participó y total de actividades posibles a participar). Posteriormente se calcula la distribución de temas, dicha distribución representa la probabilidad de que un tema determinado sea generado por un actor, lo cual constituye la entrada del componente Cálculo de influencia social.

Finalmente se obtiene como resultado la probabilidad de un actor de influir sobre los demás en cada uno de los temas que se incluyen en el modelo.

2.1. Base conceptual

En este epígrafe se enuncian varios conceptos referidos en la literatura. Se utiliza el concepto de grafo para modelar una red social. Haciendo uso de éste, se describen varias medidas de centralidad definidas en el capítulo anterior, las cuales serán utilizadas en los demás componentes del modelo y se enuncian a continuación:

- Grado de centralidad.
- Centralidad de camino-K.
- Cercanía.
- Intermediación.

Estas medidas de centralidad se utilizarán para contrastar la influencia social obtenida a partir de la aplicación del modelo propuesto.

Para la implementación de dichas medidas de centralidad, se diseñaron los siguientes Algoritmos 1, 2 y 3.

Algoritmo 1 *CalcularCentralidad*

Entrada: Un vértice v_i

Salida: Grado de centralidad C_i^{DET} para el vértice v_i

- 1: $C_i^{DET} \leftarrow -1$
 - 2: **Si** $v_i \in V$ **entonces**
 - 3: $C_i^{DET} \leftarrow \text{grado}(v_i)$
 - 4: **Fin Si**
 - 5: **Retornar** C_i^{DET}
-

Algoritmo 2 *CalcularCercanía*

Entrada: Un vértice v_i , el vector D_n de distancias mínimas de v_i a los demás vértices

Salida: La cercanía C_i^{CLO} para el vértice v_i

- 1: $C_i^{CLO} \leftarrow 0$
 - 2: **Si** $v_i \in V$ **entonces**
 - 3: $sum \leftarrow 0$
 - 4: **Para todo** $w \in V$ **hacer**
 - 5: $sum \leftarrow sum + D_n[w]$
 - 6: **Fin Para**
 - 7: $C_i^{CLO} \leftarrow \frac{sum}{\text{Longitud}(D_n)}$
 - 8: **Fin Si**
 - 9: **Retornar** C_i^{CLO}
-

Sea $P[v]$ el conjunto de predecesores en los caminos mínimos que terminan en el vértice v , $\sigma[v]$ la cantidad de caminos mínimos que terminan en el vértice v y $d[v]$ la longitud de los caminos mínimos que terminan en v .

Algoritmo 3 *Calcular Intermediación*. Tomado de [58].

Entrada: Un grafo G

Salida: Vector de intermediación C_i^{BET} para los vértices del grafo G

```

1:  $C_i^{BET}[v] \leftarrow 0, v \in V$ 
2: Para todo  $v \in V$  hacer
3:    $P[v] \leftarrow Lista()$ 
4:    $\sigma[v] \leftarrow 0$ 
5:    $d[v] \leftarrow -1$ 
6: Fin Para
7: Para todo  $s \in V$  hacer
8:    $S \leftarrow Pila()$ 
9:    $\sigma[s] \leftarrow 1$ 
10:   $d[s] \leftarrow 0$ 
11:   $Q \leftarrow Cola()$ 
12:   $Encolar(s, Q)$ 
13:  Mientras  $!Vacía(Q)$  hacer
14:     $v \leftarrow Desencolar(Q)$ 
15:     $Adicionar(v, S)$ 
16:    Para todo  $w \in Adyacentes(v)$  hacer
17:      Si  $d[w] < 0$  entonces
18:         $Encolar(w, Q)$ 
19:         $d[w] \leftarrow d[v] + 1$ 
20:      Fin Si
21:      Si  $d[w] = d[v] + 1$  entonces
22:         $\sigma[w] = \sigma[w] + \sigma[v]$ 
23:         $Adicionar(v, P[w])$ 
24:      Fin Si
25:    Fin Para
26:  Fin Mientras
27:  Para todo  $v \in V$  hacer
28:     $\delta[v] \leftarrow 0$ 
29:  Fin Para
30:  Mientras  $!Vacía(S)$  hacer
31:     $w \leftarrow Sacar(S)$ 
32:    Para todo  $v \in P[w]$  hacer
33:       $\delta[v] \leftarrow \delta[v] + \frac{\sigma[v]}{\sigma[w]}(1 + \delta[w])$ 
34:    Fin Para
35:    Si  $w \neq s$  entonces
36:       $C_i^{BET}[w] \leftarrow C_i^{BET}[w] + \delta[w]$ 
37:    Fin Si
38:  Fin Mientras
39: Fin Para
40: Retornar  $C_i^{BET}$ 

```

2.1.1. Análisis de complejidad

En este epígrafe se calcula la complejidad temporal, según el enfoque teórico, de los algoritmos 1, 2 y 3. Para el cálculo de la misma se determina la complejidad de cada paso del algoritmo y se utilizan las reglas de la suma y la multiplicación de la notación asintótica “O grande” (O) para obtener la complejidad de dichos algoritmos.

Análisis del Algoritmo 1: *Calcular Centralidad*

Haciendo uso de la estructura de datos lista de vértice - lista de adyacencia, ambas implementadas haciendo uso de la estructura diccionario, se garantiza que la operación de saber si un vértice pertenece al conjunto de vértices del grafo es $O(1)$ y la operación de obtener el grado de un vértice es $O(1)$. A partir de estos criterios, se muestra la complejidad de cada paso del algoritmo.

- Paso 1 $O(1)$.
- Pasos 2-4 $O(1)$.
 - Paso 2 $O(1)$.
 - Paso 3 $O(1)$.
- Pasos 5 $O(1)$.

Aplicando la regla de la suma se puede concluir que la complejidad del Algoritmo 1 es $O(1)$.

Análisis del Algoritmo 2: *Calcular Cercanía*

Sea D_n una colección unidimensional de longitud n . A continuación se muestra la complejidad de cada paso del algoritmo.

- Paso 1 $O(1)$.
- Pasos 2-8 $O(|V|)$. Note que este tiempo se logra debido a que en la operación de los pasos 4-6 la complejidad es $O(|V|)$, luego la complejidad del paso 7 es $O(1)$, por tanto, por la regla de la suma se puede comprobar que la complejidad es $O(|V|)$.
- Paso 9 $O(1)$.

Se puede concluir que la complejidad del Algoritmo 2 es $O(|V|)$.

Análisis del Algoritmo 3: *Calcular Intermediación*

Según [58], la complejidad del Algoritmo 3 es $O(|V||E| + |V| \log(|V|))$.

2.2. Cálculo del valor de los patrones de comportamiento

El Síndrome del Jabber (SJ) es un patrón que está presente de manera significativa en los jóvenes. Según [59] y [60], este tipo de comunicación es visto como un fenómeno, el cual es fruto del contexto tecnológico actual que ve desarrollarse al máximo el fortalecimiento de las relaciones sociales. La inclinación juvenil hacia la mensajería no resulta un fenómeno extraño. Varios autores [61, 62, 63] hacen énfasis en la relación que existe entre los jóvenes y la tecnología. Según [64] la mensajería instantánea cobra relevancia en sus procesos de comunicación y socialización.

Por otra parte el Síndrome del Correo (SC) se evidencia de manera similar en el contexto de la investigación. De los usuarios que registran sus actividades en Internet, el 92 % ellos tienen el correo electrónico como servicio más utilizado [65].

Lo anteriormente expuesto muestra que la comunicación por mensajería instantánea y vía correo electrónico constituyen dos de las actividades más realizadas mediante el uso de las Tecnologías de la Información y las Comunicaciones (TIC). Por lo que la utilización de estos patrones garantizará una caracterización relevante de las relaciones sociales en el contexto universitario.

Dentro de las variables que juegan un papel importante en la adaptación y la integración del estudiante al contexto universitario se encuentra la participación en actividades extracurriculares [66, 67, 68]. La participación en actividades de este tipo tiende a estar asociada con mayores niveles de desarrollo social, el aprendizaje y la satisfacción académica [69, 70].

Es por ello que el patrón Participación en Actividades Extracurriculares (PAE) presenta especial importancia en el análisis de influencia de las relaciones sociales que se conforman en estos eventos. Por todo lo anterior, los autores de la presente investigación proponen el uso de los patrones de comportamiento SJ, SC y PAE para facilitar el cálculo de influencia social en el contexto universitario. El modelo que se propone es extensible, garantizando así la posibilidad de incluir más patrones a medida que se avance en las diferentes investigaciones.

A continuación se definen los patrones identificados.

Síndrome del Jabber: el SJ es un patrón de comportamiento que se basa en la razón entre la cantidad de mensajes recibidos, cantidad de mensajes enviados sobre el tiempo destinado al cliente de mensajería instantánea. El SJ se define matemáticamente como:

$$SJ_i = \frac{m^-(v_i) + m^+(v_i)}{H(v_i)} \quad (2.1)$$

donde $m^-(v_i)$ es la cantidad de mensajes recibidos del actor v_i , $m^+(v_i)$ es la cantidad de mensajes enviados por v_i , $H(v_i)$ es el tiempo en que v_i permaneció conectado al cliente jabber.

Síndrome del Correo: de manera similar el SC es un patrón de comportamiento de los estudiantes universitarios basado en la razón entre la cantidad de correos recibidos, cantidad de correos enviados sobre el tiempo destinado al cliente de correo. Este patrón es definido matemáticamente como:

$$SC_i = \frac{e^-(v_i) + e^+(v_i)}{W(v_i)} \quad (2.2)$$

donde $e^-(v_i)$ es la cantidad de correos recibidos del actor v_i , $e^+(v_i)$ es la cantidad de correos enviados por v_i , $W(v_i)$ es el tiempo que destinó v_i al trabajo con el cliente de correo.

Participación en Actividades Extracurriculares: el patrón PAE establece en qué medida un estudiante universitario participa en actividades extracurriculares. Se basa en el cálculo de la razón entre la cantidad de actividades en las que dicho estudiante participa y la cantidad total de actividades en las que podía participar. Por lo que su definición matemática es la siguiente:

$$PAE_i = \frac{|A_i|}{|A|} \quad (2.3)$$

donde A_i es el conjunto de actividades en las que participó el actor v_i y A es el conjunto de todas las actividades en las que podía participar dicho actor. Note que $A_i \subseteq A$.

Para el cálculo del valor de dichos patrones se diseñaron los Algoritmos 4, 5 y 6.

Algoritmo 4 *SíndromeJabber*

Entrada: Un vértice v_i

Salida: Síndrome del jabber SJ_i para el vértice v_i

- 1: $SJ_i \leftarrow 0$
 - 2: **Si** $v_i \in V$ **entonces**
 - 3: $SJ_i \leftarrow \frac{MensajesEnviados(v_i) + MensajesRecibidos(v_i)}{TiempoConectadoJabber(v_i)}$
 - 4: **Fin Si**
 - 5: **Retornar** SJ_i
-

Algoritmo 5 *SíndromeCorreo*

Entrada: Un vértice v_i

Salida: Síndrome del correo SC_i para el vértice v_i

- 1: $SC_i \leftarrow 0$
 - 2: **Si** $v_i \in V$ **entonces**
 - 3: $SC_i \leftarrow \frac{CorreosEnviados(v_i) + CorreosRecibidos(v_i)}{TiempoConectadoCorreo(v_i)}$
 - 4: **Fin Si**
 - 5: **Retornar** SC_i
-

Algoritmo 6 *Participación Actividades Extracurriculares*

Entrada: Un vértice v_i

Salida: Participación en actividades extracurriculares PAE_i para el vértice v_i

- 1: $PAE_i \leftarrow 0$
 - 2: **Si** $v_i \in V$ **entonces**
 - 3: $C_i \leftarrow CantidadActividades(v_i)$
 - 4: $T \leftarrow TotalActividadesPosibles()$
 - 5: $PAE_i \leftarrow \frac{C_i}{T}$
 - 6: **Fin Si**
 - 7: **Retornar** PAE_i
-

2.2.1. Análisis de complejidad

En este epígrafe se calcula la complejidad temporal, según el enfoque teórico, de los algoritmos propuestos en la sección anterior.

Análisis del Algoritmo 4: *Síndrome Jabber*

Haciendo uso de la estructura de datos lista de vértice - lista de adyacencia, ambas implementadas haciendo uso de la estructura diccionario, se garantiza que la operación de obtener información de un vértice en el grafo es $O(1)$. A partir de estos criterios, se muestra la complejidad de cada paso del algoritmo.

- Paso 1 $O(1)$.
- Pasos 2-4 $O(1)$.
 - Paso 3 $O(1)$.
- Paso 5 $O(1)$.

Aplicando la regla de la suma se puede concluir que la complejidad del Algoritmo 4 es $O(1)$.

Análisis del Algoritmo 5: *Síndrome Correo*

- Paso 1 $O(1)$.
- Pasos 2-4 $O(1)$.
 - Paso 3 $O(1)$.
- Paso 5 $O(1)$.

Aplicando la regla de la suma se puede concluir que la complejidad del Algoritmo 5 es $O(1)$.

Análisis del Algoritmo 6: *Participación Actividades Extracurriculares*

Sean A_i , A dos listas simplemente enlazadas. A continuación se muestra la complejidad de cada paso del algoritmo.

- Paso 1 $O(1)$.
- Pasos 2-6 $O(1)$. Note que este tiempo se logra debido a que en la operación de los pasos 3 y 4 la complejidad es $O(1)$, luego la complejidad del paso 5 es $O(1)$ debido a que ambas listas tienen como atributo su longitud, por tanto, por la regla de la suma se puede comprobar que la complejidad es $O(1)$.
- Paso 7 $O(1)$.

Se concluye que la complejidad del Algoritmo 6 es $O(1)$.

2.2.2. Cálculo de la distribución de temas

Para calcular la distribución de un tema $z \in \{1, 2, \dots, T\}$ para un actor $v_i \in V$, en lo adelante theta (θ_{zv_i}), se tiene en cuenta que los patrones definidos en el modelo representan los temas que pueden ser generados por los actores. Dicha distribución representa la probabilidad que un tema determinado sea generado por un actor y se define como:

$$\theta_{zv_i} = \frac{PC_z(v_i)}{\sum_{j=1}^T PC_j(v_i)} \quad (2.4)$$

donde $PC_z(v)$ es el valor del patrón de comportamiento z (z puede tomar los valores SJ, SC y PAE) para el actor v_i .

LEMA 2.2.1 $\forall z \in \{1, 2, 3, \dots, T\}, \forall v_i \in V, i \in \{1, 2, \dots, |V|\}$, se cumple que $\theta_{zv_i} = \frac{PC_z(v_i)}{\sum_{j=1}^T PC_j(v_i)}$,

entonces $\sum_{z=1}^T \theta_{zv_i} = 1$, donde T es la cantidad de temas.

DEMOSTRACIÓN:

$$\begin{aligned} \sum_{i=1}^T \theta_{zv_i} &= \frac{PC_1(v_i)}{\sum_{j=1}^T PC_j(v_i)} + \dots + \frac{PC_T(v_i)}{\sum_{j=1}^T PC_j(v_i)} = \frac{PC_1(v_i) + PC_2(v_i) + \dots + PC_T(v_i)}{\sum_{j=1}^T PC_j(v_i)} \\ &= \frac{PC_1(v_i) + PC_2(v) + \dots + PC_T(v_i)}{PC_1(v_i) + PC_2(v_i) + \dots + PC_T(v_i)} \\ &= 1 \end{aligned}$$

$$\therefore \sum_{z=1}^T \theta_{zv_i} = 1 \quad \square$$

A partir de lo anterior se demuestra que la suma de la distribución de los temas para un actor es igual a 1, lo cual constituye la propiedad fundamental de la probabilidad.

Si se desea calcular la distribución del tema jabber (SJ) para el actor v_i y se tienen como restantes temas el correo y las actividades extracurriculares, se realiza de la siguiente manera:

$$\theta_{SJv_i} = \frac{SJ_i}{SJ_i + SC_i + PAE_i} \quad (2.5)$$

De esta forma el valor resultante representa la probabilidad de que el tema jabber sea generado por el actor v_i y además se garantiza la restricción $\sum_{z=1}^T \theta_{zv_i} = 1$, especificada en [31].

2.3. Cálculo de influencia social

Una vez calculado los valores de la distribución de temas para los actores de la red, se procede al cálculo de influencia social. Para ello se propone el uso del enfoque TFG y Propagación por Afinidad Temática (TAP) ambos descritos en [31].

El enfoque TFG tiene los siguientes componentes: el conjunto de actores de la red V y un conjunto de variables $\{y_i \mid i = 1, 2, \dots, |V|\}$, las cuales se corresponden con los $|V|$ actores de la red de entrada. Las notaciones están descritas en la Tabla 2.1.

Tabla 2.1: Notaciones.

Símbolo	Descripción
T	El número de temas.
y_i^z	El actor v_i representativo en el tema z .
y_i	Vector de actores representativos para todos los temas sobre el actor v_i .
θ_{zv_i}	La distribución del tema z para el actor v_i .
e_{ij}	Una arista que conecta al actor v_i con el actor v_j .
w_{ij}^z	La fuerza de interacción de la arista e_{ij} relacionada con el tema z .
μ_{ij}^z	La influencia social del actor v_i sobre el actor v_j en el tema z .

El vector $y_i \in \{1, \dots, |V|\}^T$ modela las influencias de otros actores hacia v_i basadas en temas. Cada elemento y_i^z toma el valor del conjunto $\{1, \dots, |V|\}$, el cual representa el actor que tiene la mayor probabilidad de influir sobre el actor v_i en el tema z .

En la Figura 2.2 se muestra un ejemplo del uso del enfoque TFG. El conjunto de actores está

compuesto por $\{v_1, \dots, v_4\}$, los cuales tienen sus correspondientes vectores $Y = \{y_1, \dots, y_4\}$. Las aristas entre los actores indican las cuatro relaciones sociales en la red.

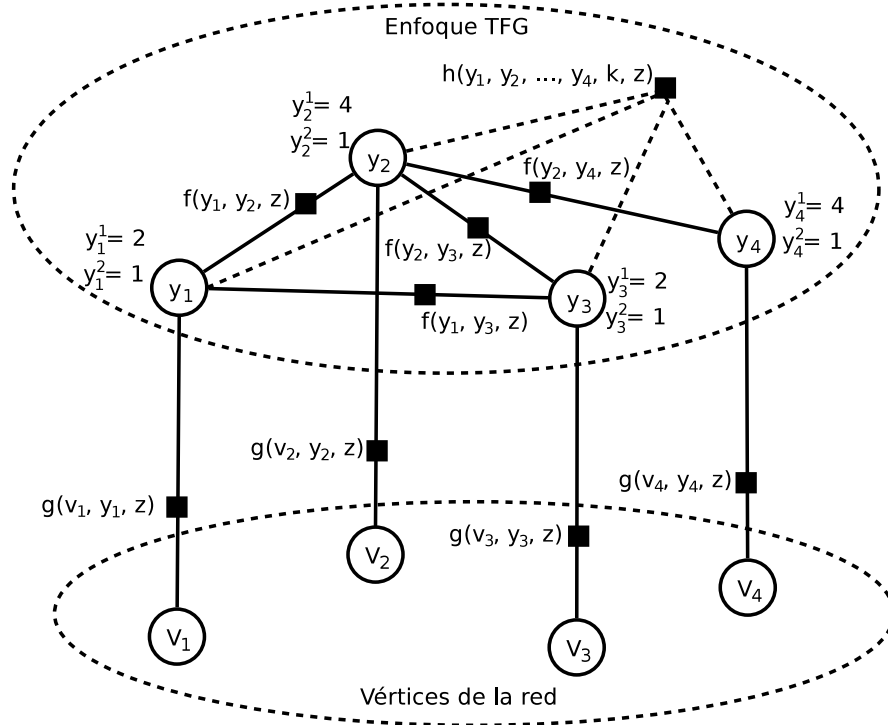


Figura 2.2: Ejemplo del uso del enfoque TFG.

En [31] se definen además tres tipos de funciones características:

- **Función característica del actor** $g(v_i, y_i, z)$: es una función característica definida en cada actor específicamente en el tema z .
- **Función característica de la arista** $f(y_i, y_j, z)$: es una función característica definida en cada arista específicamente en el tema z .
- **Función característica global** $h(y_1, \dots, y_{|V|}, k, z)$: es una función característica definida sobre todos los actores en el tema z .

En la presente investigación se define la función característica $g(v_i, y_i, z)$ de la siguiente manera:

$$g(v_i, y_i, z) = \begin{cases} \frac{w_{iy_i}^z}{\sum_{j \in VEC(i)} (w_{ij}^z + w_{ji}^z)} & y_i^z \neq i \\ \frac{\sum_{j \in VEC(i)} w_{ji}^z}{\sum_{j \in VEC(i)} (w_{ij}^z + w_{ji}^z)} & y_i^z = i \end{cases} \quad (2.6)$$

donde $VEC(i)$ representa los índices de los vecinos del actor v_i ; $w_{ij}^z = \theta_j \alpha_{ij}$ refleja la fuerza de interacción entre el actor v_i y el actor v_j ; α_{ij} denota el peso de la arista e_{ij} .

El valor de α_{ij} puede ser definido de varias formas. Por ejemplo, en redes de coautoría, α_{ij} puede definirse como el número de artículos publicados por v_i y v_j . En la presente investigación α_{ij} se define como:

$$\alpha_{ij} = \begin{cases} 1 & \text{si } e_{ij} \in E \\ 0 & \text{si } e_{ij} \notin E \end{cases}$$

Para realizar el cálculo de influencia social se debe identificar cuál actor tiene la mayor probabilidad de influir sobre otro en un tema específico. Es decir, maximizar la función de probabilidad $P(v_i, Y)$, la cual se define como:

$$P(v_i, Y) = \frac{1}{Z} \prod_{k=1}^{|V|} \prod_{z=1}^T h(y_1, \dots, y_{|V|}, k, z) \prod_{i=1}^{|V|} \prod_{z=1}^T g(v_i, y_i, z) \prod_{e_{kl} \in E} \prod_{z=1}^T f(y_i, y_i, z) \quad (2.7)$$

Para ello, en [31] se propone usar TAP para modelar el aprendizaje, tomando como función objetivo 2.7 para encontrar el parámetro de configuración que maximice dicha función objetivo. Para su implementación se hace uso de las siguientes ecuaciones:

$$b_{ij}^z = \log \frac{g(v_i, y_i, z)|_{y_i^z=j}}{\sum_{k \in VEC(i) \cup \{i\}} g(v_i, y_i, z)|_{y_i^z=k}} \quad (2.8)$$

$$r_{ij}^z = b_{ij}^z - \max_{k \in VEC(j)} \{b_{ik}^z + a_{ik}^z\} \quad (2.9)$$

$$a_{jj}^z = \max_{k \in VEC(j)} \min\{r_{kj}^z, 0\} \quad (2.10)$$

$$a_{ij}^z = \min(\max\{r_{jj}^z, 0\}, -\min\{r_{kj}^z, 0\} - \max_{k \in VEC(j) \setminus \{i\}} \min\{r_{kj}^z, 0\}), i \in VEC(j) \quad (2.11)$$

Finalmente, se define el valor de influencia social usando una función basada en las variables r y a :

$$\mu_{ij}^z = \frac{1}{1 + e^{-(r_{ji}^z + a_{ji}^z)}} \quad (2.12)$$

Para la implementación del cálculo de influencia social, Tang et al. proponen el Algoritmo 7 y a su vez muestran que la complejidad de dicho algoritmo es $O(|E|T)$ [31].

Algoritmo 7 *Algoritmo AprendizajeTAP*

Entrada: Un grafo $G = (V, E)$ y una distribución de temas $\theta_v, v \in V$

Salida: Grafos de influencia social basados en temas $G_z = (V_z, E_z), z = 1, 2, \dots, T$

- 1: Calcular la función característica $g(v_i, y_i, z)$ de acuerdo a 2.6
 - 2: Calcular b_{ij}^z de acuerdo a 2.8
 - 3: $\{r_{ij}^z\} \leftarrow 0$
 - 4: **Repetir**
 - 5: **Para todo** $(e_{ij}, z), e_{ij} \in E, z \in T$ **hacer**
 - 6: Actualizar r_{ij}^z de acuerdo a 2.9
 - 7: **Fin Para**
 - 8: **Para todo** $(v_j, z), v_j \in V, z \in T$ **hacer**
 - 9: Actualizar a_{jj}^z de acuerdo a 2.10
 - 10: **Fin Para**
 - 11: **Para todo** $(e_{ij}, z), e_{ij} \in E, z \in T$ **hacer**
 - 12: Actualizar a_{ij}^z de acuerdo a 2.11
 - 13: **Fin Para**
 - 14: **Hasta** *convergencia*
 - 15: **Para todo** $v_t \in V$ **hacer**
 - 16: **Para todo** $s \in VEC(t) \cup \{t\}$ **hacer**
 - 17: Calcular μ_{st}^z de acuerdo a 2.12
 - 18: **Fin Para**
 - 19: **Fin Para**
 - 20: **Retornar** Generar $G_z = (V_z, E_z)$ para cada tema z de acuerdo a μ_{ij}^z {Por cada actor $v_j \in V$ se identifica el actor $v_i \in V$ de mayor influencia sobre v_j y se crea la arista e_{ij} con un peso que representa μ_{ij}^z }
-

Note que el Algoritmo 7 en el paso 20 realiza una transformación del grafo de entrada a un grafo dirigido y ponderado, donde el peso de cada arista dirigida $e_{ij} \in E$ representa la probabilidad de influencia que ejerce el actor v_i sobre el actor v_j .

Se considera oportuno enfatizar en que la propuesta realizada constituye un modelo extensible, flexible y genérico. A continuación se aborda cada uno de estos aspectos.

Extensible: para enriquecer los análisis de influencia a partir del resultado de la aplicación del modelo, se pueden adicionar nuevos patrones de comportamiento que se logren identificar. Esto no interfiere en la coherencia del modelo, lo cual evidencia que el mismo es extensible.

Flexible: el modelo presentado posibilita que no se tengan en cuenta algunos patrones en caso de no contar con la información necesaria acerca del comportamiento de los actores de la red. Además, los patrones de comportamiento que se identifiquen pueden ser definidos de distintas maneras teniendo en cuenta varios criterios, así como el cálculo de influencia social puede ser realizado a partir de otro enfoque (distinto del enfoque TFG). Esto refleja la flexibilidad que presenta la propuesta.

Genérico: el diseño por componentes del modelo posibilita que pueda ser aplicado en distintos contextos. Para ello es necesario modificar el componente Cálculo del valor de los patrones de comportamiento, identificando los patrones de comportamiento de los usuarios en el contexto deseado y definiendo la forma de calcular estos patrones. De esta manera se muestra la generalidad del modelo.

2.4. Conclusiones del capítulo

En el presente capítulo se ha descrito un modelo para el cálculo de influencia social en el contexto universitario. Como parte del mismo se han enunciado varias definiciones y se ha realizado el diseño y análisis de varios algoritmos. En el modelo propuesto la formalización de patrones de comportamiento en el contexto universitario y su uso en el cálculo de influencia social constituye la principal novedad. Finalmente se concluye lo siguiente:

- El uso de los conceptos relacionados con la teoría de grafos y medidas de centralidad contribuye a definir patrones de comportamiento en el ámbito universitario.
- La formalización matemática de estos patrones y el uso de algoritmos diseñados para calcular el valor de dichos patrones contribuye a calcular influencia social en el contexto de la investigación en los temas jabber, correo y actividades extracurriculares.
- El modelo presentado es extensible, debido a que se pueden adicionar más patrones de comportamiento sin afectar los demás componentes del mismo, enriqueciendo de esta forma los análisis posteriores.
- El modelo propuesto permite realizar el cálculo de influencia social a partir de otro enfoque (distinto al enfoque TFG), así como definir los patrones de comportamiento que se identifiquen de distintas maneras teniendo en cuenta varios criterios, mostrándose la flexibilidad del mismo.
- Se garantiza además la generalidad del modelo. Ello se evidencia en su posible uso en otros contextos, solo bastaría identificar los patrones que tienen lugar en ellos.

Capítulo 3

Validación de la propuesta

3. VALIDACIÓN DE LA PROPUESTA

EN este capítulo, se realiza una descripción de la realización computacional del modelo presentado. Se explica un caso de estudio que se utilizó para validar la propuesta.

Además, se muestran los resultados obtenidos a partir de la aplicación de un sociograma, para contrastar los resultados obtenidos con la aplicación del modelo teniendo en cuenta los criterios de los actores involucrados en el caso de estudio. Finalmente se realiza una discusión de los resultados obtenidos.

3.1. Realización computacional del modelo propuesto

Para la realización computacional del modelo propuesto se utilizó XP como metodología de desarrollo de software, debido a las características y la cantidad de recursos humanos disponibles para realizar la presente investigación.

En XP la producción de código debe realizarse en parejas de programadores. Puesto que este estilo de programación brinda una amplia gama de beneficios, dentro los cuales están: la mayoría de los errores son detectados conforme son introducidos en el código, por consiguiente la tasa de errores del producto final es baja; se posibilita la transferencia de conocimientos de programación entre los miembros del equipo y los programadores se comunican entre ellos, mejorando así el flujo de información y la dinámica del equipo [71].

Haciendo uso de la metodología anterior se desarrolló una aplicación informática que brinda las siguientes funcionalidades:

- Construir un grafo a partir de un fichero de texto plano.
- Visualizar grafos.
- Calcular influencia social en el contexto universitario a partir de los patrones de comportamiento SJ, SC y PAE.

Para describir las funcionalidades anteriores se utilizaron las Historias de Usuario (HU) que se muestran en las tablas 3.1, 3.2 y 3.3.

La herramienta fue desarrollada utilizando Java como lenguaje de programación (versión 7.0.55) y la biblioteca de clases JUNG (versión 2.0.1) para la representación de los grafos.

Tabla 3.1: HU “Construir un grafo a partir de un fichero de texto plano”.

Historia Usuario “Construir un grafo a partir de un fichero de texto plano”	
Número: 1	Nombre HU: Construir un grafo a partir de un fichero de texto plano.
Modificación de HU número:	
Referencia:	
Programadores: Anel Wilfredo Fernández Rodríguez, Jorge Jesús Hidalgo Ruiz.	Iteración asignada: 1
Prioridad en el negocio: Alta.	Puntos estimados (días): 1
Riesgo de desarrollo: Medio.	Puntos reales (días): 1
Descripción: La aplicación debe ser capaz de leer de un fichero en texto plano y construir un grafo a partir de la información descrita en dicho fichero.	
Observaciones:	

Tabla 3.2: HU “Visualizar grafo”.

Historia Usuario “Visualizar grafo”	
Número: 2	Nombre HU: Visualizar grafo.
Modificación de HU número:	
Referencia:	
Programadores: Anel Wilfredo Fernández Rodríguez, Jorge Jesús Hidalgo Ruiz.	Iteración asignada: 1
Prioridad en el negocio: Alta.	Puntos estimados (días): 2
Riesgo de desarrollo: Medio.	Puntos reales (días): 2
Descripción: La aplicación debe ser capaz de visualizar un grafo de acuerdo a la estructura definida.	
Observaciones:	

Tabla 3.3: HU "Calcular influencia social".

Historia Usuario "Calcular influencia social"	
Número: 3	Nombre HU: Calcular influencia social.
Modificación de HU número:	
Referencia:	
Programadores: Anel Wilfredo Fernández Rodríguez, Jorge Jesús Hidalgo Ruiz.	Iteración asignada: 1
Prioridad en el negocio: Alta.	Puntos estimados (días): 3
Riesgo de desarrollo: Alto.	Puntos reales (días): 3
Descripción: La aplicación debe ser capaz de calcular influencia social a partir de un grafo y una distribución de temas.	
Observaciones:	

En las tablas 3.4, 3.5 y 3.6 se muestran las tarjetas Clase, Responsabilidad y Colaboración (CRC), las cuales describen las relaciones de las clases utilizadas para el desarrollo de la aplicación.

Tabla 3.4: CRC "ConstruirGrafo".

CRC "ConstruirGrafo"	
Responsabilidad:	Colaboradores:
Construir un grafo a partir de un fichero de texto plano	MainJFrame, Graph, Vértice, Arista

Tabla 3.5: CRC "VisualizarGrafo".

CRC "VisualizarGrafo"	
Responsabilidad:	Colaboradores:
Visualizar grafo	MainJFrame, JViewerPanel, Graph, Arista, Vértice, Tema, TopicDistributionScore, VisualizationViewer, Layout

Tabla 3.6: CRC “CalcularInfluenciaSocial”.

CRC “CalcularInfluenciaSocial”	
Responsabilidad:	Colaboradores:
Calcular influencia social	TAP, Graph, Arista, Vértice, Tema

En la Figura 3.1 se muestra una imagen de la aplicación desarrollada en la que se visualizan las relaciones sociales existentes entre los actores de la red seleccionada. En el panel de la izquierda se muestran varias pestañas. La primera representa el grafo original que modela dicha red y las restantes representan los grafos resultantes de la aplicación del modelo por cada uno de los temas especificados en la entrada.

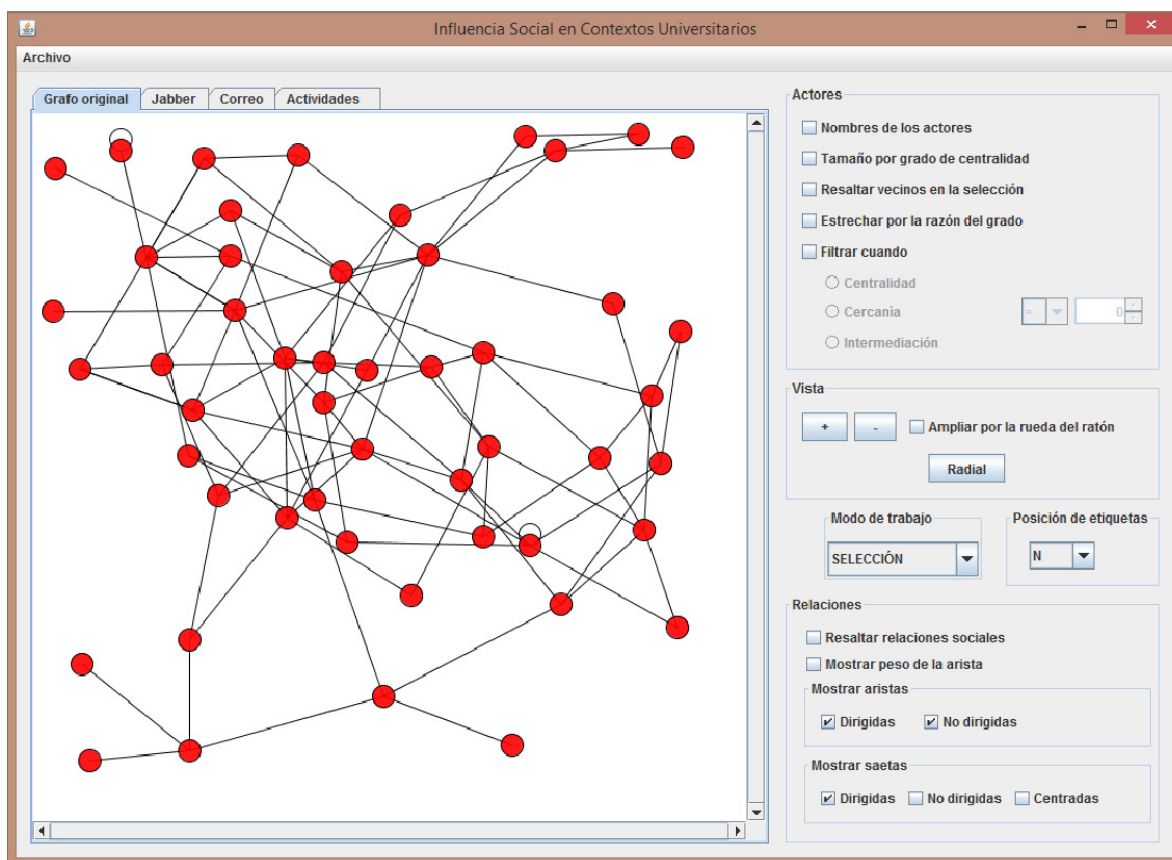


Figura 3.1: Relaciones sociales existentes entre los actores de la red de entrada.

En la parte derecha de la Figura 3.1 se muestra un panel con varias opciones de visualización que permiten enriquecer los análisis que se realicen sobre la red que se esté visualizando.

La Figura 3.2 muestra el grafo de influencia social obtenido a partir de la aplicación del modelo para el tema jabber. Dicho grafo es un grafo dirigido y ponderado $G = (V, E, f_c)$, la función $f_c : E \rightarrow \mathbb{R}^+$ le hace corresponder a cada arista $(v_i, v_j) \in E$, $v_i, v_j \in V$ un valor real positivo denominado costo, cuya interpretación es la probabilidad que tiene el actor v_i de influir sobre el actor v_j . Se puede apreciar también que en la selección de un actor, se muestran aquellos que influyen sobre él.

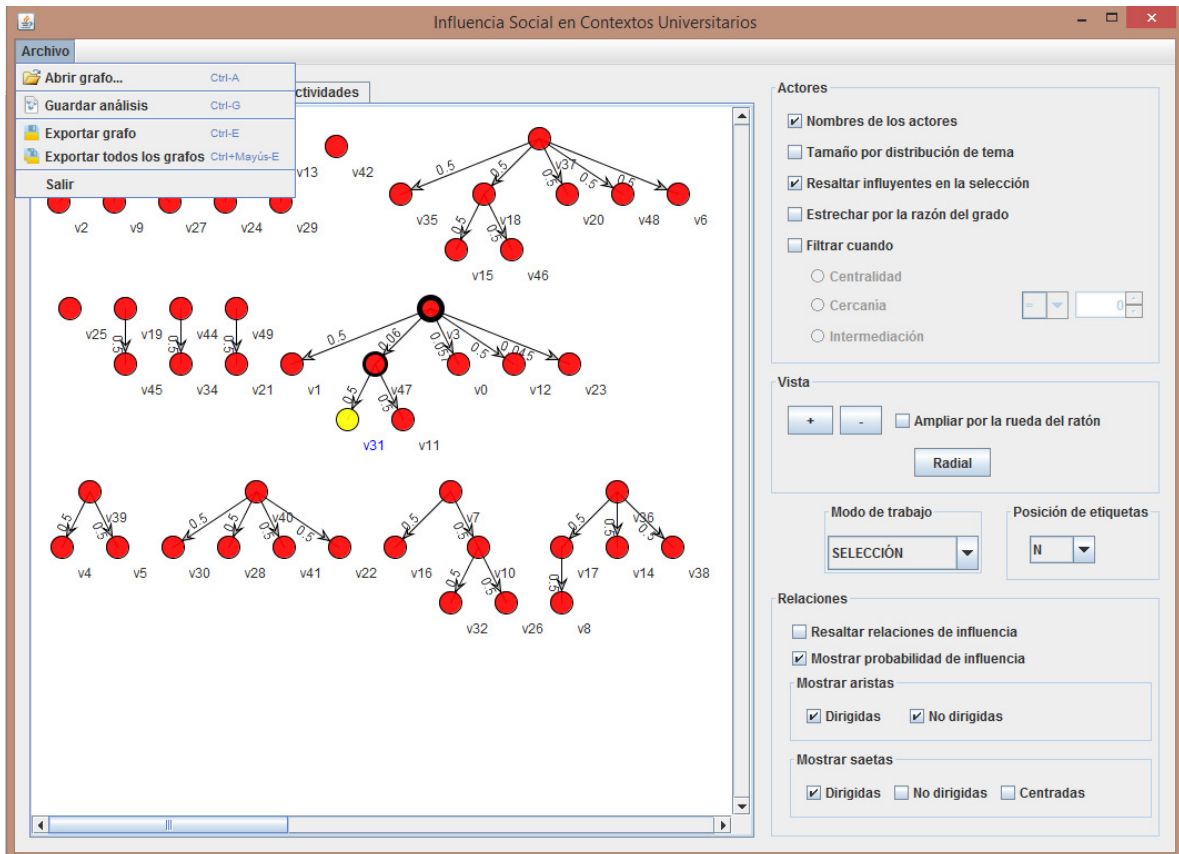


Figura 3.2: Grafo de influencia social obtenido a partir de la aplicación del modelo para el tema jabber.

En la Figura 3.2 se seleccionó el actor v_{31} y a su vez resaltan los actores v_{47} y v_3 , formándose una tríada compuesta por (v_{31}, v_{47}, v_3) , donde el tercer actor (v_3) es el que más influye sobre el segundo (v_{47}) y a su vez este es el que más influye sobre el primero (v_{31}).

Teniendo en cuenta lo anterior, se puede identificar el actor más influyente en cada uno de los grupos sociales que se identifican a partir del análisis realizado en un tema específico. Esto es una información relevante para una posible caracterización de la red en dicho tema.

La herramienta cuenta con funcionalidades que permiten guardar el análisis realizado de influencia social, así como la opción de exportar como imagen el grafo que se esté visualizando.

3.2. Caso de estudio

A continuación se describe el caso de estudio utilizado para aplicar el modelo propuesto, empleándose para ello la aplicación informática expuesta anteriormente.

Al analizar el contexto universitario se determinó que la brigada es el espacio ideal para aplicar la propuesta. Los actores de este ámbito están en constante relación ya que comparten el mismo espacio la mayor parte del tiempo. Es por ello que la mayoría de los análisis que se realizan sobre grupos de estudiantes se hacen en la brigada como estructura. Por lo que se seleccionó una brigada de cuarto año de la Facultad 3 de la Universidad de las Ciencias Informáticas (UCI).

Dicha brigada tiene buena participación en las actividades extracurriculares, ocupa el primer lugar en la Emulación Inter-brigadas de la Federación Estudiantil Universitaria (FEU) y usa regularmente el jabber en un mismo espacio de trabajo. Esta brigada está compuesta por 24 estudiantes.

Para la conformación del caso de estudio se analizaron los historiales de conversación de los clientes de mensajería instantánea de los miembros de la brigada, así como su participación en las actividades convocadas por la universidad. Aquellos datos que presentaban ruido fueron desechados para el caso de estudio, como se muestra en la Tabla 3.7.

Tabla 3.7: Datos de la brigada seleccionada.

Patrones de comportamiento	Datos recolectados		Datos con ruido
	Real	Posible	
PAE	24	24	0
SJ	19	24	8

Es necesario señalar que no se recopiló la información de los actores del uso del correo ya que no se contaba con los datos necesarios para realizar el cálculo del valor del patrón SC.

Sean A_i el conjunto de actividades en las que participó el actor v_i , A el conjunto de todas las actividades en las que podía participar dicho actor, PAE_i el valor del patrón de comportamiento PAE para el actor v_i , $m^-(v_i)$ la cantidad de mensajes recibidos a través del jabber para el actor v_i , $m^+(v_i)$ la cantidad de mensajes enviados a través del jabber para el actor v_i , $H(v_i)$ el tiempo que el actor v_i permaneció conectado al jabber, SJ_i el valor del patrón de comportamiento SJ para el actor v_i , θ_{PAEv_i} la distribución del tema actividades extracurriculares para el actor v_i , θ_{SJv_i} la distribución del tema jabber para el actor v_i .

A partir de lo anterior en la Tabla 3.8 se muestran los datos del caso de estudio utilizado.

Tabla 3.8: Datos del caso de estudio utilizado.

Actor (v_i)	Participación en actividades extracurriculares			Síndrome del jabber				Distribución de temas	
	$ A_i $	$ A $	PAE_i	$m^-(v_i)$	$m^+(v_i)$	$H(v_i)(min)$	SJ_i	θ_{PAEv_i}	θ_{SJv_i}
v_1	24	29	0.83	154	163	1440	0.22	0.79	0.21
v_2	19	29	0.66	26	34	1440	0.04	0.94	0.06
v_3	18	29	0.62	17	15	1440	0.02	0.97	0.03
v_4	17	29	0.59	195	159	1440	0.25	0.70	0.30
v_5	17	29	0.59	70	71	1440	0.10	0.86	0.14
v_6	17	29	0.59	80	63	1440	0.10	0.86	0.14
v_7	16	29	0.55	17	27	1440	0.03	0.95	0.05
v_8	15	29	0.52	169	215	1440	0.27	0.66	0.34
v_9	14	29	0.48	223	168	1440	0.27	0.64	0.36
v_{10}	12	29	0.41	32	24	1440	0.04	0.91	0.09
v_{11}	9	29	0.31	140	172	1440	0.22	0.59	0.41

En la Figura 3.3 se muestran las relaciones sociales existentes en la brigada seleccionada para el caso de estudio.

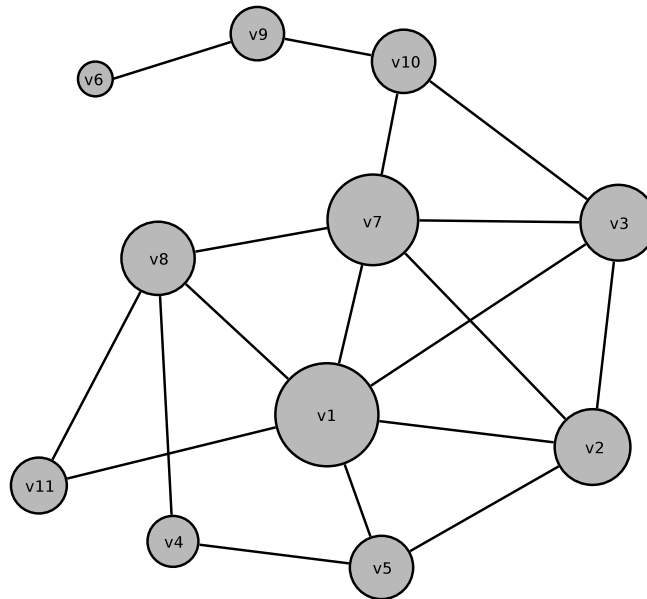


Figura 3.3: Red que representa las relaciones sociales existentes en la brigada seleccionada.

Con la aplicación del modelo al caso de estudio descrito anteriormente se obtuvieron los siguientes resultados.

Síndrome del Jabber

En la Figura 3.4 se muestra el grafo de factor temático obtenido luego de la aplicación del modelo para el patrón SJ. En la misma se puede apreciar las relaciones de influencia existentes entre los distintos actores de la red, donde v_1 es el actor que más influye sobre los actores v_2 , v_3 y v_5 , con una probabilidad de 0,918; 0,963 y 0,816 respectivamente.

Note que el tamaño de los vértices del grafo está dado por el valor del patrón de comportamiento específico que tiene cada actor.

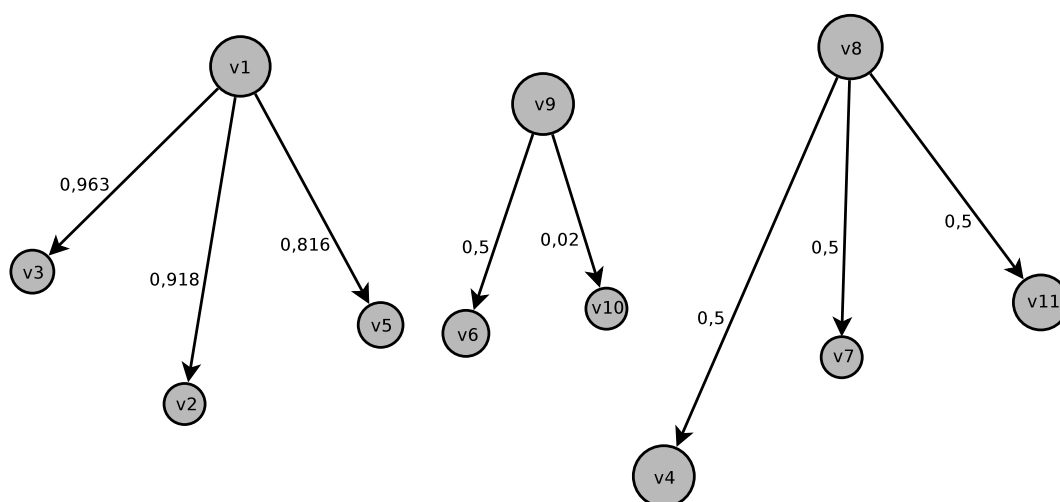


Figura 3.4: Grafo de factor temático para el patrón SJ.

De manera similar ocurre con los actores v_8 y v_9 , los cuales son los más influyentes en sus respectivos grupos sociales. De esta forma se identifica quiénes son los líderes en el tema jabber, los cuales no tienen aristas incidentes ($g^-(v) = 0$) y presentan un mayor valor en el patrón de comportamiento SJ.

Participación en Actividades Extracurriculares

En la Figura 3.5 se muestra el grafo de factor temático obtenido a través de la aplicación del modelo para el patrón PAE. En la misma se pueden apreciar las relaciones de influencia existentes entre los distintos actores de la red, donde v_1 es el actor que más influye sobre los actores v_2 , v_7 , v_8 y v_{11} , con una probabilidad de 0,571; 0,169; 0,699 y 0,921 respectivamente.

De igual forma ocurre con los actores v_3 , v_5 y v_9 , los cuales son los más influyentes en sus respectivos grupos sociales, identificándose como líderes en el tema.

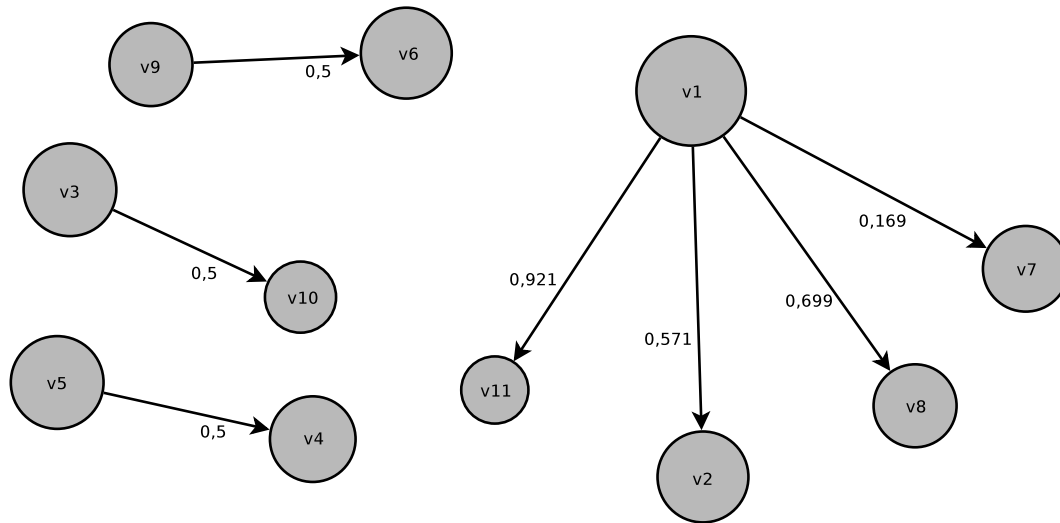


Figura 3.5: Grafo de factor temático para el patrón PAE.

Los análisis anteriores aportan información relevante para el proceso para la toma de decisiones en el ámbito universitario.

A continuación se muestra la relación entre las medidas de centralidad definidas en la base conceptual del modelo presentado y los actores identificados como líderes en los diferentes temas, a manera de complementación del análisis realizado anteriormente.

Para una mejor interpretación de los resultados, los valores del grado de centralidad, cercanía e intermediación fueron normalizados (ver Tabla 3.9). Para ello se utilizó la herramienta Weka¹ aplicando el filtro no supervisado de normalización sobre atributos.

Para mostrar la relación entre dichas medidas y los actores identificados como líderes se realizó una prueba de hipótesis, ya que esta nos permite medir el comportamiento de una variable en una muestra [72].

Síndrome del Jabber

En los análisis realizados anteriormente se identificaron como líderes los actores v_1 , v_8 , v_9 para el tema jabber. En la Tabla 3.10 se muestra la media de los valores del grado de centralidad, cercanía e intermediación de los actores, así como los valores resultantes luego de aplicar la prueba de hipótesis con el estadígrafo $T - Student$ para una varianza desconocida y un nivel de confianza de 95 %.

¹Weka: <http://www.cs.waikato.ac.nz/~ml/weka/>

Tabla 3.9: Valores normalizados de las medidas de centralidad de los actores de la red.

Actor	Grado de centralidad	Cercanía	Intermediación
v_1	1	0.894118	0.625
v_2	0.6	0.715789	0.171875
v_3	0.6	0.8	0.34375
v_4	0.2	0.352	0.03125
v_5	0.4	0.452174	0.125
v_6	0	0	0
v_7	0.8	1	0.859375
v_8	0.6	0.715789	0.46875
v_9	0.2	0.266667	0.5625
v_{10}	0.4	0.64	1
v_{11}	0.3	0.26	0.09

Tabla 3.10: Media de las medidas de centralidad para los actores en el tema jabber.

Medida de centralidad	Líderes (μ_1)	Restantes actores (μ_2)	T	$T_{0,05;9}$
Grado de centralidad	0.6	0.4	0.991	1.833
Cercanía	0.625525	0.544995	0.380	1.833
Intermediación	0.552083	0.316406	0.989	1.833

Asumiendo como hipótesis nula $H_0 : \mu_1 \geq \mu_2$ e hipótesis alternativa $H_1 : \mu_1 < \mu_2$ se puede apreciar que para las tres medidas se cumple que $T > -T_{0,05;9}$, lo cual muestra que no se encuentran en las regiones críticas. Por tanto, se asume la hipótesis H_0 como verdadera para dichas medidas.

Por lo que se puede concluir que los actores identificados como líderes en el tema jabber presentan, de manera general, un mayor grado de centralidad, cercanía e intermediación que los restantes actores.

En la Figura 3.6 se muestra un gráfico de barras donde se visualiza el análisis realizado.

Participación en Actividades Extracurriculares

En el tema actividades extracurriculares los líderes identificados fueron v_1 , v_3 , v_5 y v_9 . De manera similar se realiza en análisis para el tema actividades extracurriculares como se realizó anteriormente. En la Tabla 3.11 se muestra la media de los valores de las medidas de centralidad de dichos líderes y los restantes actores, así como los valores resultantes luego de aplicar la prueba de hipótesis con el estadígrafo $T - Student$ para una varianza desconocida y un nivel de confianza de 95 %.

Asumiendo como $H_0 : \mu_1 \geq \mu_2$ e $H_1 : \mu_1 < \mu_2$ se puede apreciar que para las tres medidas se

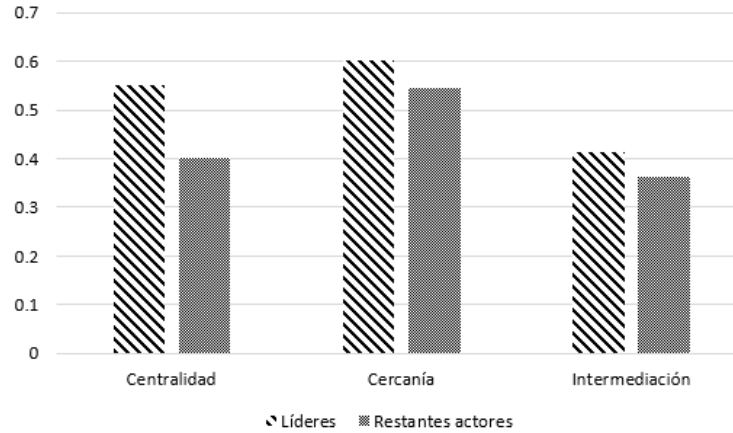


Figura 3.6: Media de las medidas de centralidad de los actores en el tema jabber.

Tabla 3.11: Media de las medidas de centralidad para los actores en el tema actividades extracurriculares.

Medida de centralidad	Líderes (μ_1)	Restantes actores (μ_2)	T	$T_{0,05;9}$
Grado de entralidad	0.55	0.4	1.051	1.833
Cercanía	0.60324	0.546225	0.636	1.833
Intermediación	0.414063	0.361607	0.423	1.833

cumple que $T > -T_{0,05;9}$, lo cual muestra que no se encuentran en las regiones críticas. Por tanto, se asume la hipótesis H_0 como verdadera para las tres medidas, lo cual muestra que los líderes en el tema actividades extracurriculares presentan, de manera general, un mayor valor para dichas medidas. En la Figura 3.7 se observa un gráfico de barras donde se visualiza el análisis realizado anteriormente. De esta forma se evidencia que los líderes identificados a partir de la aplicación del modelo desempeñan un rol de controladores de la información, mientras que los demás actores desempeñan el rol de propagadores de la misma.

3.3. Aplicación de un sociograma

Para contrastar los resultados obtenidos luego de la aplicación el modelo se decide realizar un sociograma como instrumento de medición diseñado por un experto en sociología. El mismo permite representar y caracterizar el conjunto de relaciones que se establecen en la red.

La encuesta presenta un total de 3 preguntas, cada una con 3 incisos correspondientes a los patrones de comportamiento identificados por los autores (ver anexo A).

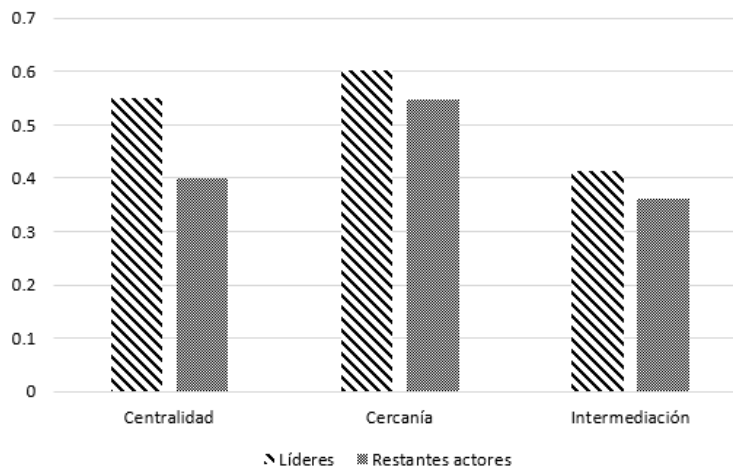


Figura 3.7: Media de las medidas de centralidad de los actores para el tema en actividades extracurriculares.

El sociograma fue aplicado a 24 estudiantes de 24 posibles para un 100 % de la matrícula de la brigada seleccionada. En la Tabla 3.12 se muestran los resultados del mismo.

A partir de un análisis de carácter sociológico, el experto en el tema ponderó cada respuesta. La respuesta de la primera pregunta tiene una ponderación de 1 y la segunda de 0,5. Luego se suma la cantidad de veces que un actor fue seleccionado en la primera posición en todas las preguntas y se multiplica por la ponderación correspondiente.

Similarmente se realiza con las posiciones dos y tres por cada patrón. Posteriormente se promedian los tres valores obtenidos con el valor de la respuesta a la pregunta tres y finalmente se obtiene el valor de un patrón específico según el sociograma utilizado.

En la Tabla 3.13 se muestra en detalles el resultado obtenido para el patrón SJ.

Sea $P_{j,i}^{kl}$ la cantidad de veces en la que el actor i fue seleccionado en la opción j en la pregunta k inciso l , donde $i = 1, 2, \dots, 24$; $j = 1, 2, 3$; $k = 1, 2, 3$ y $l = a, b, c$. Por ejemplo, $P_{1,5}^{2b}$ es la cantidad de veces en que el actor 5 (G) fue seleccionado en la opción 1 de la pregunta dos inciso b.

La columna tres de la Tabla 3.13 representa P_{1i}^{1a} , la columna cuatro representa P_{2i}^{1a} y así sucesivamente hasta la columna ocho.

Sea R_i^l el valor de la pregunta tres inciso l para el actor i , donde $i = 1, 2, \dots, 24$ y $l = a, b, c$. Por ejemplo, R_3^b es el valor de la pregunta tres inciso b para el actor 3.

En la columna nueve se representa R_i^b , luego en la columna diez el valor que se muestra es la suma de la ponderación de las respuestas a las preguntas uno y dos en las primeras opciones, $P_{1i}^{1a} + P_{1i}^{2a}0,5$.

De manera similar se calculan los valores de las columnas 11 y 12, $P_{2i}^{1a} + P_{2i}^{2a}0,5$ y $P_{3i}^{1a} + P_{3i}^{2a}0,5$

Tabla 3.12: Resultado de la aplicación del sociograma.

ID	Pregunta 1									Pregunta 2									Pregunta 3		
	1a			1b			1c			2a			2b			2c					
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
A	W	I	P	W	I	P	W	N	I	H	C	M	I	C	M	I	N	L	5	2	4
B	L	G	I	P	I		G	L	D	O	P	G	I			I	L		3	5	5
Y	J	N		J	N		J	N		G	P	H	F	C	L	L	N	F	1	4	3
T	N	F	G	N	G	F	N	T	G	O	C	N	F	N	C	T	N	L	5	4	5
G	Q	R	C	S	V	R	P	V	F	C	H	P	O	C	R	P	V	F	4	5	5
Q	C	G	V	N	H	O	G	V	P	G	C	S	N	C	H	G	X	I	2	5	3
R	H	C	O	H	C	L	C	L	Q	H	O	C	L	I	C	L	I	N	4	5	3
P	I	G	C	I	G	C	I	F	V	G	F	V	I	X	C	I	K	Z	1	5	1
D	M	K	Z	M	F	C	V	K	Z	P	C	H	I	L	C	I	K	Z	5	3	5
F	V	G	K	V	G	K	V	G	P	C	M	P	I	L	N	F	V	G	3	3	5
W	A	R	N	A	N	M	N	I	D	L	H	N	C	I	L	C	N	I	5	5	3
S	Y	J	K	K	Z	J	K	Z	J	R	C	P	I	C	O	L	N	T	4	5	5
O	C	H	L	C	H	L	C	G	L	L	P	C	C	I	D	K	Z	L	4	5	5
I	O	G	P	L	N	F	P	G	D	C	G	H	L	N	D	P	G	B	5	3	4
N	T	S	X	T	I	W	T	G	X	M	H	P	I	O	C	N	L	I	4	3	5
V	R	C	O	R	C	O	G	Q	O	M	O	I	M	I	O	V	G	P	2	5	5
X	H	P	G	H	P	G	D	H	K	O	C	H	I	O	C	J	L	N	3	4	5
J	S	N	Y	S	N	Y	N	Y	Z	G	H	S	I	O	P	L	K	Z	2	2	3
L	M	C	H	F	Y	I	K	Z	D	H	O	P	I	Y	F	G	P	N	2	5	5
K	C	O	R	I	S	P	D	J	Z	C	O	R	I	N	L	I	Z	L	3	1	4
C	H	L	O	Z	W	X	G	H	O	H	O	C	G	P	L	L	I	Z	4	5	4
H	X	L	R	X	R	C	C	O	R	X	L	C	I	L	G	G	L	P	3	5	3
M	Q	P	V	W	X	C	Q	G	P	S	F	G	I	D	L	G	V	F	3	4	5
Z	K	S	R	R	C	A	K	S	9	S	P	C	C	I	X	K	L	I	1	2	5

respectivamente.

Finalmente la columna 13 muestra el valor del patrón de comportamiento SJ según el sociograma utilizado. Este valor se calcula promediando los valores de las columnas 9, 10, 11 y 12. En la Tabla 3.14 se muestra el resultado final del sociograma aplicado para los patrones SJ y PAE.

3.4. Comparación del caso de estudio con el sociograma

Para la comparación del caso de estudio con el sociograma aplicado se seleccionaron los actores que están presentes en ambos escenarios, por lo que finalmente se cuenta con un total de 11 actores.

Para determinar correlación existente entre los resultados obtenidos con el sociograma aplicado y en

Tabla 3.13: Resultado del sociograma para el patrón SJ.

No.	ID	Pregunta 1			Pregunta 2			Pregunta 3	Índice Jabber			Promedio Jabber
		1a-Jabber	1	2	3	1	2		3	2	1	
1	A	1	0	0	0	0	0	2	1	0	0	0.750
2	B	0	0	0	0	0	0	5	0	0	0	1.250
3	Y	1	0	1	0	0	0	4	1	0	1	1.500
4	T	1	0	0	0	0	0	4	1	0	0	1.250
5	G	0	5	2	4	1	2	5	2	5.5	3	3.875
6	Q	2	0	0	0	0	0	5	2	0	0	1.750
7	R	1	2	3	1	0	1	5	1.5	2	3.5	3.000
8	P	0	2	2	1	4	5	5	0.5	4	4.5	3.500
9	D	0	0	0	0	0	0	3	0	0	0	0.750
10	F	0	1	0	0	2	0	3	0	2	0	1.250
11	W	1	0	0	0	0	0	5	1	0	0	1.500
12	S	1	2	0	2	0	2	5	2	2	1	2.500
13	O	1	1	3	3	5	0	5	2.5	3.5	3	3.500
14	I	1	1	1	0	0	1	3	1	1	1.5	1.625
15	N	1	2	1	0	0	2	3	1	2	2	2.000
16	V	1	0	2	0	0	1	5	1	0	2.5	2.125
17	X	1	0	1	1	0	0	4	1.5	0	1	1.625
18	J	1	1	0	0	0	0	2	1	1	0	1.000
19	L	1	2	1	2	1	0	5	2	2.5	1	2.625
20	K	1	1	2	0	0	0	1	1	1	2	1.250
21	C	3	3	2	4	6	5	5	5	6	4.5	5.125
22	H	3	1	1	4	4	4	5	5	3	3	4.000
23	M	2	0	0	2	1	1	4	3	0.5	0.5	2.000
24	Z	0	0	1	0	0	0	2	0	0	1	0.750

el caso de estudio al que se le aplicó la propuesta realizada en la presente investigación, se utiliza el coeficiente de correlación de Pearson.

Este coeficiente determina el grado de relación entre dos variables cuantitativas y se define como:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (3.1)$$

Tabla 3.14: Resultado del sociograma aplicado en la brigada seleccionada.

ID	Promedio	
	Jabber	Actividades extracurriculares
A	0.750	1.000
B	1.250	1.375
Y	1.500	1.000
T	1.250	2.000
G	3.875	4.625
Q	1.750	1.500
R	3.000	1.000
P	3.500	2.125
D	0.750	2.750
F	1.250	2.250
W	1.500	1.000
S	2.500	1.500
O	3.500	2.000
I	1.625	3.125
N	2.000	3.625
V	2.125	3.000
X	1.625	1.625
J	1.000	1.625
L	2.625	3.750
K	1.250	2.875
C	5.125	1.875
H	4.000	1.250
M	2.000	1.250
Z	0.750	3.250

donde σ_{XY} es la covarianza de (X, Y) , σ_X es la desviación típica de la variable X y σ_Y es la desviación típica de la variable Y . La interpretación del valor de $\rho_{X,Y}$ es la siguiente [73]:

- Si $\rho_{X,Y} = 1$, existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables, denominada relación directa.
- Si $0 < \rho_{X,Y} < 1$, existe una correlación positiva.
- Si $\rho_{X,Y} = 0$, no existe relación lineal entre las dos variables, es decir, la correlación es nula.

- Si $-1 < \rho_{X,Y} < 0$, existe una correlación negativa.
- Si $\rho_{X,Y} = -1$, existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables, denominada relación inversa.

3.4.1. Análisis para el patrón Síndrome del Jabber

En la Tabla 3.15 se muestran los resultados obtenidos en la aplicación del sociograma y el cálculo del valor del patrón SJ a partir de la aplicación del modelo.

Tabla 3.15: Valores obtenidos con la aplicación del modelo para del patrón SJ y el sociograma para dicho patrón.

Aplicación del modelo (X)		Sociograma (Y)	
Actor	SJ	Actor	SJ
v_9	0.272	v_8	5.125
v_8	0.267	v_{11}	4.000
v_4	0.246	v_4	3.500
v_1	0.220	v_1	2.625
v_{11}	0.217	v_9	2.000
v_6	0.099	v_5	2.000
v_5	0.098	v_6	1.250
v_2	0.042	v_7	1.250
v_{10}	0.039	v_{10}	1.000
v_7	0.031	v_2	0.750
v_3	0.022	v_3	0.750

A partir de los resultados mostrados en la Tabla 3.15, se aplica el coeficiente de correlación de Pearson obteniéndose el resultado $\rho_{X,Y} = 0,90818$, lo cual refleja que existe una correlación positiva. En la mayoría de los estudios sociales se considera que existe una correlación significativa entre dos variables cuando dicho coeficiente es mayor a 0,8.

El diagrama de dispersión de datos es una gráfica del tipo $X \rightarrow Y$ cuyo objetivo es analizar la forma en que dos variables numéricas están relacionadas. El diagrama de dispersión se obtiene coleccionando los datos en pares de valores sobre dos variables (x, y) [74].

En la Figura 3.8 se muestra el diagrama de dispersión de las variables X e Y mostradas en la Tabla 3.15.

Como se puede apreciar el valor de Y aumenta a medida que aumenta el valor X , lo cual muestra que

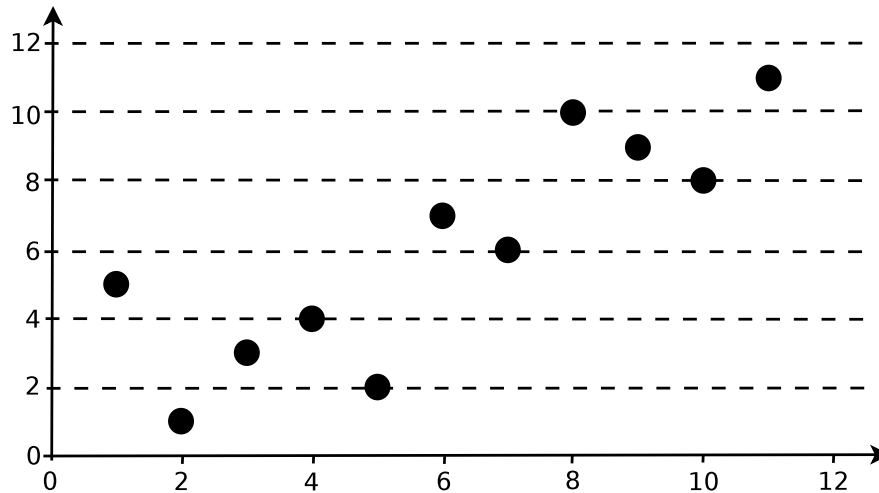


Figura 3.8: Diagrama de dispersión de las variables X e Y mostradas en la Tabla 3.15.

existe una correlación positiva [74].

Teniendo en cuenta lo anterior se evidencia que existe un alto grado de correlación entre los valores obtenidos en el cálculo del valor del patrón de comportamiento SJ a través de la aplicación del modelo propuesto y en la aplicación del sociograma para dicho patrón.

3.4.2. Análisis para el patrón Participación en Actividades Extracurriculares

En la Tabla 3.16 se muestran los resultados obtenidos en la aplicación del sociograma y el cálculo del valor del patrón PAE a partir de la aplicación del modelo.

A partir del resultado anterior, se aplica el coeficiente de correlación de Pearson obteniéndose el resultado $\rho_{X,Y} = 0,89385$, lo cual refleja que existe una correlación positiva.

En la Figura 3.9 se muestra el diagrama de dispersión de los datos anteriores.

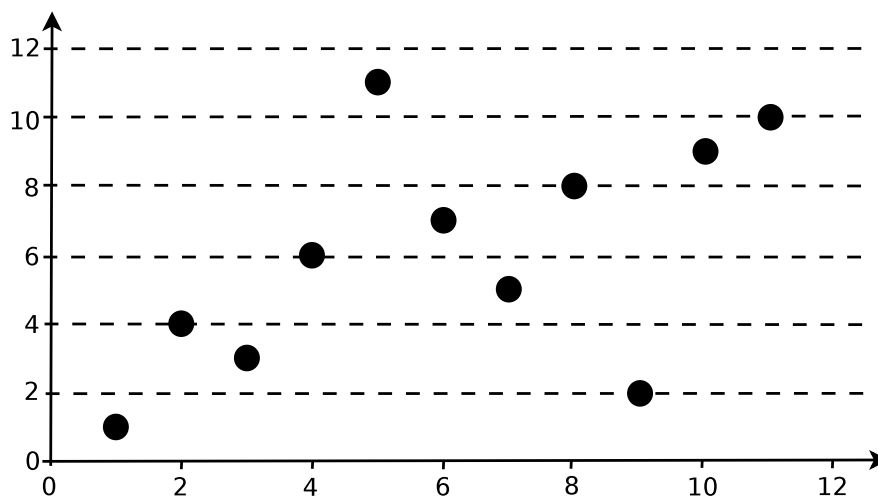
Como se puede apreciar el valor de Y aumenta a medida que aumenta el valor X , lo cual muestra que existe una correlación positiva [74].

Por lo que se puede concluir que existe un alto grado de correlación entre los valores obtenidos en el cálculo del patrón de comportamiento PAE mediante a través de la aplicación del modelo propuesto y en la aplicación del sociograma para dicho patrón.

Por todo lo anterior se puede afirmar que el resultado obtenido al aplicar el modelo propuesto presenta una correlación significativa considerando el resultado del sociograma aplicado a los actores. Lo que permite contrastar la veracidad de los resultados obtenidos a partir del cálculo del valor de los patrones

Tabla 3.16: Valores obtenidos con la aplicación del modelo para del patrón PAE y el sociograma para dicho patrón.

Definición matemática (X)		Sociograma (Y)	
Actor	PAE	Actor	PAE
v_1	0.828	v_1	3.750
v_2	0.655	v_9	3.625
v_3	0.621	v_3	3.250
v_4	0.586	v_2	3.000
v_5	0.586	v_7	2.875
v_6	0.586	v_4	2.125
v_7	0.552	v_6	2.000
v_8	0.517	v_8	1.875
v_9	0.483	v_{10}	1.625
v_{10}	0.414	v_{11}	1.250
v_{11}	0.310	v_5	1.250

**Figura 3.9:** Diagrama de dispersión de las variables X e Y mostradas en la Tabla 3.16.

identificados en el ámbito universitario.

3.5. Conclusiones del capítulo

Luego del trabajo realizado se arribó a las siguientes conclusiones:

- El sociograma utilizado mostró que existe correspondencia entre los valores de los patrones de

comportamiento obtenidos a partir de la aplicación del modelo y los valores arrojados por dicho instrumento, con un alto grado de correlación.

- El alto grado de correlación obtenido justifica el uso de las definiciones matemáticas (que forman parte del modelo propuesto) de los patrones de comportamiento identificados por los autores.
- La realización del caso de estudio evidenció la efectividad del modelo presentado, mostrando que con el mismo se calcula influencia social en el contexto universitario atendiendo a varios temas. Además, el modelo propuesto facilita la identificación de grupos sociales y líderes por cada tema.
- La herramienta desarrollada permitió realizar análisis en la red contrastando la influencia social calculada a partir de la aplicación del modelo con las medidas: grado de centralidad, cercanía e intermediación.

CONCLUSIONES

Como resultado de la presente investigación se obtuvo un modelo para calcular influencia social en el contexto universitario a partir de la definición de patrones de comportamiento, atendiendo a los temas jabber, correo y actividades extracurriculares. El mismo garantiza extensibilidad, flexibilidad y generalidad. En base a los resultados obtenidos se arribó a las siguientes conclusiones:

1. Las herramientas estudiadas para el ARS no tienen características distintivas que posibiliten el análisis de influencia social en el contexto universitario.
2. La extensibilidad del modelo propuesto radica en la posibilidad de adicionar más patrones de comportamiento sin afectar los demás componentes del mismo, enriqueciendo de esta forma los análisis posteriores.
3. El modelo permite realizar el cálculo de influencia social a partir de otro enfoque, así como definir los patrones de comportamiento que se logren identificar en un contexto determinado, mostrándose la flexibilidad del mismo.
4. El modelo presentado es genérico. Ello se evidencia en su posible uso en otros contextos.
5. La realización del caso de estudio evidenció la efectividad del modelo propuesto, mostrando que dicho modelo facilita el cálculo de influencia social en el contexto universitario, atendiendo a varios temas. Además, el modelo propuesto facilita la identificación de grupos sociales y líderes por cada tema.
6. A partir de los resultados obtenidos se justifica el uso del modelo propuesto para calcular influencia social en el contexto universitario.

RECOMENDACIONES

La investigación realizada no cubre todos los aspectos posibles del análisis de redes sociales en el contexto universitario. Por este motivo se recomienda lo siguiente:

1. Incluir a la herramienta desarrollada la funcionalidad de efectuar los cálculos estadísticos necesarios para realizar comparaciones entre los resultados obtenidos.
2. Definir otros patrones de comportamiento en el contexto universitario que resulten de interés institucional.
3. Introducir en los análisis de influencia y comparación de resultados otras variables como: el índice académico, las medidas disciplinarias tecnológicas, entre otras.
4. Extender el modelo con patrones de comportamiento en trabajadores.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Barabasi, A. L. *Linked: The New Science Of Networks*. Basic Books, Mayo 2002. ISBN: 9780786746965.
- [2] Prieto Gutiérrez, J. J. Herramientas para el análisis y monitoreo en Redes Sociales. *IRIE. International Review of Information Ethics*, vol. 16, pp. 33–40, 2011.
- [3] Fumero, A. & Roca, G. *Web 2.0*. Fundación Orange, Madrid, 2007. ISBN: M-15478-2007.
- [4] Hervás, J. M. G. Redes sociales en el móvil. , *Telos: Cuadernos de comunicación e innovación*, no. 83, pp. 8–12, 2010.
- [5] Santos, F. R. El concepto de red social. *Reis*, pp. 137–152, 1989.
- [6] Ros-Martín, M. Evolución de los servicios de redes sociales en internet. vol. 18, *El profesional de la información*, no. 5, pp. 552–558, 2009.
- [7] Manchego, F. E. A. *Sistema de información de detección de plagio en documentos digitales usando el método Document Fingerprinting*. Tesis de doctorado, Pontificia Universidad Católica del Perú. Facultad de Ciencias e Ingeniería. Mención: Ingeniería Informática, 2009.
- [8] Barón, M. A. N. & Salina, S. O. Identificación de relaciones entre los nodos de una red social. vol. 18, *Ingeniería*, no. 1, 2013.
- [9] Celaya, J. La empresa en la Web 2.0. *Barcelona: Gestión*, vol. 2008, pp. 284, 2000.
- [10] Sanz Menéndez, L. Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes, 2003. Publicado en: *Apuntes de Ciencia y Tecnología*, 7: 21-29, junio de 2003. Disponible en: <http://digital.csic.es/handle/10261/1569>. Consultado el: 2013-11-29.

- [11] Tucker, A. B. *Computer Science Handbook, Second Edition*. Chapman & Hall/CRC, 2nd edition, 2004.
- [12] Koutrika, G.; Effendi, F. A.; Gyöngyi, Z.; Heymann, P.; Garcia-Molina, H. et al. Combating spam in tagging systems: An evaluation. vol. 2, *ACM Transactions on the Web (TWEB)*, no. 4, pp. 22, 2008.
- [13] Di Caro, L.; Candan, K. S. & Sapino, M. L. Using tagflake for condensing navigable tag hierarchies from tag clouds. pp. 1069–1072. ACM, 2008. ISBN: 1605581933.
- [14] Aggarwal, C. C. *Social network data analytics*. Springer, 2011. ISBN: 1441984615.
- [15] Beauchamp, M. A. An improved index of centrality. vol. 10, *Behavioral Science*, no. 2, pp. 161–163, 1965.
- [16] Freeman, L. C. Centrality in social networks conceptual clarification. vol. 1, *Social networks*, no. 3, pp. 215–239, 1979.
- [17] Wasserman, S. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Noviembre 1994. ISBN: 9780521387071.
- [18] Quijada, D. F. El análisis de redes sociales aplicado al estudio de la estructura de las industrias culturales. En *I+ C Investigar a comunicación [Recurso electrónico]: Invetigar la comunicación= Invetigar la comunicació= Komunikazio-Ikerketa: Actas y memoria final: Congreso Internacional Fundacional AE-IC, Santiago de Compostela, 30, 31 de enero y 1 de febrero de 2008*, pp. 24. Asociación Española de Investigación de la Comunicación, 2008.
- [19] Navarro Sánchez, L. A. & Salazar Fernández, J. P. Análisis de redes sociales aplicado a redes de investigación en ciencia y tecnología. vol. 3, *Sínt. tecnol*, no. 2, pp. 69–86, 2007.
- [20] Borgatti, S. P.; Everett, M. G. & Freeman, L. C. Ucinet for Windows: Software for Social Network Analysis. *Analytic Technologies*, 2002.
- [21] Molina Jara, F. J. *Construcción de equipos de trabajo mediante análisis de redes sociales e identificación de atributos personales*. Tesis de doctorado, Universidad de Chile, 2012. Disponible en: http://www.cybertesis.uchile.cl/tesis/uchile/2010/cf-molina_fj/html/index.html. Consultado el: 2012-09-14.
- [22] Furht, B. *Handbook of social network technologies and applications*, vol. 1. Springer, 2010.

- [23] Franco, A. EGONET, 2009. Disponible en: <http://escoladeredes.net/profiles/blogs/egonet-1>. Consultado el: 2013-10-24.
- [24] Serrano-Puche, J. Herramientas web para la medición de la influencia digital: análisis de Klout y PeerIndex. vol. 21, *El profesional de la información*, no. 3, pp. 298–303, 2012.
- [25] Lazarsfeld, P. F.; Berelson, B. & Gaudet, H. *The people's choice: How the voter makes up his mind in a presidential election*. Duell, Sloan and Pearce, New York, 1944. ISBN: 978 0231031585.
- [26] Katz, E. & Lazarsfeld, P. F. *Personal Influence, The part played by people in the flow of mass communications*. Transaction Publishers, 1970. ISBN: 1412830702.
- [27] Granovetter, M. S. The strength of weak ties. *American journal of sociology*, pp. 1360–1380, 1973.
- [28] Krackhardt, D. The strength of strong ties: The importance of philos in organizations. *Networks and organizations: Structure, form, and action*, vol. 216, pp. 239, 1992.
- [29] López Díaz, M. El análisis de redes sociales. Una herramienta explicativa para la comprensión de las relaciones internacionales. Abril 2010. Disponible en: <http://www.revistas.unam.mx/index.php/rri/article/view/16309>. Consultado el: 2013-10-21.
- [30] Medina Iborra, I. ¿Cómo medir la influencia de los grupos de interés? (Propuestas desde el pluralismo, el elitismo y el nuevo institucionalismo). , *Instituto de Ciències Polítiques y Sociales, Universidad Autònoma de Barcelona. Working Paper*, no. 279, 2009.
- [31] Tang, J.; Sun, J.; Wang, C. & Yang, Z. Social influence analysis in large-scale networks. pp. 807–816. ACM, 2009. ISBN: 1605584959.
- [32] Rosenquist, J. N.; Murabito, J.; Fowler, J. H. & Christakis, N. A. The spread of alcohol consumption behavior in a large social network. vol. 152, *Annals of Internal Medicine*, no. 7, pp. 426–433, 2010.
- [33] Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. vol. 357, *New England journal of medicine*, no. 4, pp. 370–379, 2007.
- [34] Christakis, N. A. & Fowler, J. H. The collective dynamics of smoking in a large social network. vol. 358, *New England journal of medicine*, no. 21, pp. 249–258, 2008.

- [35] Mednick, S. C.; Christakis, N. A. & Fowler, J. H. The spread of sleep loss influences drug use in adolescent social networks. vol. 5, *PloS one*, no. 3, pp. 775, 2010.
- [36] Cacioppo, J. T.; Fowler, J. H. & Christakis, N. A. Alone in the crowd: the structure and spread of loneliness in a large social network. vol. 97, *Journal of personality and social psychology*, no. 6, pp. 977, 2009.
- [37] Fowler, J. H.; Christakis, N. A.; Steptoe & Roux, D. Dynamic spread of happiness in a large social network: longitudinal analysis of the Framingham Heart Study social network. *BMJ: British medical journal*, pp. 23–27, 2009.
- [38] Crandall, D.; Cosley, D.; Huttenlocher, D.; Kleinberg, J. & Suri, S. Feedback effects between similarity and social influence in online communities. En *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 160–168. ACM, 2008. ISBN: 1605581933.
- [39] Anagnostopoulos, A.; Kumar, R. & Mahdian, M. Influence and correlation in social networks. En *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 7–15. ACM, 2008. ISBN: 1605581933.
- [40] Domingos, P. & Richardson, M. Mining the network value of customers. En *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 57–66. ACM, 2001. ISBN: 158113391X.
- [41] Richardson, M. & Domingos, P. Mining knowledge-sharing sites for viral marketing. En *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 61–70. ACM, 2002. ISBN: 158113567X.
- [42] Kempe, D.; Kleinberg, J. & Tardos, É. Maximizing the spread of influence through a social network. En *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146. ACM, 2003. ISBN: 1581137370.
- [43] Cialdini, R. B. & Goldstein, N. J. Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, vol. 55, pp. 591–621, 2004.
- [44] Elias, S. M. & Pratkanis, A. R. Teaching social influence: Demonstrations and exercises from the discipline of social psychology. vol. 1, *Social Influence*, no. 2, pp. 147–162, 2006.

- [45] Eastwick, P. W. & Gardner, W. L. Is it a game? Evidence for social influence in the virtual world. vol. 4, *Social Influence*, no. 1, pp. 18–32, 2009.
- [46] Bakshy, E.; Karrer, B. & Adamic, L. A. Social influence and the diffusion of user-created content. En *Proceedings of the 10th ACM conference on Electronic commerce*, pp. 325–334. ACM, 2009. ISBN: 1605584584.
- [47] Holme, P. & Newman, M. E. Nonequilibrium phase transition in the coevolution of networks and opinions. vol. 74, *Physical Review E*, no. 5, pp. 056108, 2006.
- [48] La Fond, T. & Neville, J. Randomization tests for distinguishing social influence and homophily effects. En *Proceedings of the 19th international conference on World wide web*, pp. 601–610. ACM, 2010. ISBN: 1605587990.
- [49] Aral, S.; Muchnik, L. & Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. vol. 106, *Proceedings of the National Academy of Sciences*, no. 51, pp. 21544–21549, 2009.
- [50] Goyal, A.; Bonchi, F. & Lakshmanan, L. V. Learning influence probabilities in social networks. En *Proceedings of the third ACM international conference on Web search and data mining*, pp. 241–250. ACM, 2010. ISBN: 160558889X.
- [51] Xiang, R.; Neville, J. & Rogati, M. Modeling relationship strength in online social networks. En *Proceedings of the 19th international conference on World wide web*, pp. 981–990. ACM, 2010. ISBN: 1605587990.
- [52] Scripps, J.; Tan, P.-N. & Esfahanian, A.-H. Measuring the effects of preprocessing decisions and network forces in dynamic network analysis. En *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 747–756. ACM, 2009. ISBN: 1605584959.
- [53] Tang, L. & Liu, H. Relational learning via latent social dimensions. En *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 817–826. ACM, 2009. ISBN: 1605584959.
- [54] Manavoglu, E.; Pavlov, D. & Giles, C. L. Probabilistic user behavior models. En *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, pp. 203–210. IEEE, 2003. ISBN: 0769519784.

- [55] Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J. & Glance, N. Cost-effective outbreak detection in networks. En *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 420–429. ACM, 2007. ISBN: 1595936092.
- [56] Weng, J.; Lim, E.-P.; Jiang, J. & He, Q. Twiterrank: finding topic-sensitive influential twitterers. En *Proceedings of the third ACM international conference on Web search and data mining*, pp. 261–270. ACM, 2010. ISBN: 160558889X.
- [57] Ritzer, G. *Modern Sociological Theory*. The American Sociologist, 5th edition, 2008.
- [58] Brandes, U. A faster algorithm for betweenness centrality*. vol. 25, *Journal of Mathematical Sociology*, no. 2, pp. 163–177, 2001.
- [59] De Laire, F. ¿Identidad juvenil? La insoportable levedad del ser: aportes para renovar el marco teórico de los estudios sobre juventud. , *Revista Mad*, no. 4, 2011.
- [60] Gil, A.; Feliu, J.; Rivero, I. & Gil, E. P. ¿Nuevas tecnologías de la información y la comunicación o nuevas tecnologías de relación? Niños, jóvenes y cultura digital. *Consultado en octubre*, vol. 12, pp. 2007, 2003.
- [61] MacLuhan, M. & Fiore, Q. *El medio es el mensaje: un inventario de efectos*. Paidós, 1969.
- [62] Negroponte, N. & Plaking, D. *Ser digital*. Editorial Atlántida, 1995. ISBN: 9500814730.
- [63] Balardini, S. Dedeejays y ciberchabones: subjetividades juveniles y tecnocultura. *Revista Jóvenes*, vol. 20, 2004.
- [64] Balaguer, R. El chat y el Messenger: instrumentos de entrenamiento en comunicación para tiempos de incertidumbre y baja atención. *Archivo del Observatorio para la CiberSociedad Ponencia presentada en el marco de las VIII Jornadas de AIDEP, The British Schools, Montevideo*, vol. 6, 2005. Disponible en: <http://www.cibersociedad.net/archivo/articulo.php>. Consultado el: 2014-05-23.
- [65] Pearson Education. Most Popular Internet Activities, 2014. Disponible en: <http://www.infoplease.com/ipa/A0921862.html>. Consultado el: 2014-05-23.
- [66] Astin, A. W. *What matters in college? Four critical years revisited*. Jossey-Bass San Francisco, 1993. ISBN: 1555424929.

- [67] Pascarella, E. T. Reassessing the effects of living on campus versus commuting to college: A causal modelling approach. vol. 7, *Review of Higher Education*, no. 3, pp. 247–260, 1984.
- [68] Weidman, J. Undergraduate socialization: A conceptual approach. *Higher education: Handbook of theory and research*, vol. 5, pp. 289–322, 1989.
- [69] Kuh, G. D. In their own words: What students learn outside the classroom. vol. 30, *American Educational Research Journal*, no. 2, pp. 277–304, 1993.
- [70] National Institute of Education (US). *Involvement in learning: realizing the potential of American higher education: final report of the Study Group on the Conditions of Excellence in American Higher Education*. National Institute of Education, US Dept. of Education, 1984.
- [71] Beck, K. & Andres, C. *Extreme programming explained: embrace change*. Addison-Wesley Professional, 2004.
- [72] Taeger, D. & Kuhnt, S. *Statistical hypothesis testing*. Wiley Online Library, 1st edition, 2014.
- [73] Kenney, J. F. *Mathematics of statistics*. Van Nostrand, Princeton, 3rd edition, 1962.
- [74] Jessica, U. M. *Seeing through statistics*. Thomson Brooks/Cole, 2005.

GLOSARIO DE TÉRMINOS

ARS Análisis de Redes Sociales

TFG Grafo de Factor Temático (del inglés Topical Factor Graph)

TAP Propagación por Afinidad Temática (del inglés Topical Afinity Propagation)

TIC Tecnologías de la Información y las Comunicaciones

SJ Síndrome del Jabber

SC Síndrome del Correo

PAE Participación en Actividades Extracurriculares

XP Metodología de desarrollo de software ágil (del inglés Extreme Programing)

HU Historia de Usuario (artefacto generado por la metodología XP)

CRC Clase, Responsabilidad y Colaboración (artefacto generado por la metodología XP, del inglés Class, Responsibility and Collaboration)

UCI Universidad de las Ciencias Informáticas

FEU Federación Estudiantil Universitaria

A. SOCIOGRAMA

La presente encuesta tiene como objetivo validar una propuesta de modelo para el cálculo de influencia social en el contexto universitario, así como caracterizar el contexto de interacción a partir de la información inter-personal. Le rogamos sea lo más fiel posible.

Preguntas:

1. Seleccione de los miembros de su brigada, tres de ellos teniendo en cuenta:
 - a) ¿Con quién prefiere comunicarse más por el jabber?
 - b) ¿Con quién prefiere comunicarse más por el correo?
 - c) ¿Con quién prefiere participar en actividades extracurriculares?

2. Según su criterio, seleccione de los miembros de su brigada tres de ellos teniendo en cuenta:
 - a) ¿Quién usa más el jabber?
 - b) ¿Quién usa más el correo?
 - c) ¿Quién participa más en actividades extracurriculares?

3. Califíquese de 1 (incierto) a 5 (cierto) teniendo en cuenta lo siguiente:
 - a) Me comunico más por correo.
 - b) Me comunico más por jabber.
 - c) Me gusta participar en actividades extracurriculares.