

**Universidad de las Ciencias Informáticas  
Facultad 3**



**Título: Algoritmo para la inferencia de redes sociales a partir de las actividades realizadas por los usuarios en diferentes contextos de interacción.**

**Trabajo de Diploma para optar por el título de  
Ingeniero en Ciencias Informáticas**

**Autores:** Talía Sánchez Viera  
Rancel Ruana Capote

**Tutores:** Dr. Raykenler Yzquierdo Herrera  
Lic. Raynel Batista Téllez

**Co-Tutor:** Ing. Héctor David Peguero Alvarez

**La Habana, Cuba**

**2014**

**DECLARACIÓN DE AUTORÍA**

MINISTERIO DE EDUCACIÓN SUPERIOR

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Declaramos ser autores del presente trabajo de diploma y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales del mismo, con carácter exclusivo. Autorizamos a dicho centro para que haga el uso que estime pertinente con este trabajo.

Para que así conste firmamos la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_.

\_\_\_\_\_  
Talía Sánchez Viera

\_\_\_\_\_  
Rancel Ruana Capote

\_\_\_\_\_  
Raynel Batista Téllez

\_\_\_\_\_  
Raykenler Yzquierdo Herrera

**DEDICATORIA**

***De Talía***

***A mi mamá, por querernos tanto y estar siempre para mí.  
A mis abuelos por apoyarme tanto y siempre animarme a ser alguien en la vida.  
A mi hermano, por quererme mucho.  
A mis hermanitas, por ser mi alegría.  
A mi novio y compañero, por amarnos tanto.***

***De Rancel***

***A mi madre.***

## AGRADECIMIENTOS

### **De Talía:**

*Quisiera agradecerle de todo corazón a mi mamá por querernos tanto, por estar siempre para mí, por ser mi guía, por apoyarme y motivarme, por confiar tanto en mí, por cuidarme siempre, escucharme y quererme, por ser más que una madre. Te agradezco todo lo que soy. Te quiero con todo mi corazón.*

*A Tata y a Chichi, por quererme tanto, por preocuparse tanto por mí, por motivarme a ser alguien en la vida, por darme ánimos, por mimarme tanto, por la alegría de saber que siempre están conmigo y siempre piensan en mí, los adoro.*

*A mi hermano porque aunque me moleste tanto y me des tantas preocupaciones, siempre estás ahí para abrazarme y darme mucho cariño, porque sé que a la hora de las buenas y las malas eres incondicional a mí y me cuidas y te preocupas como mismo lo hago por ti, porque sé que siempre estaremos juntos y somos y seremos inseparables.*

*A mi hermana Yahima por quererme y ayudarme tanto.*

*A mis perris, Luna, Circe y Paloma, la alegría de la casa, por darme tanto cariño cuando me ven, por ser tan agradecidas, por ladrar tanto y por cuidar todos los días a mi mamá mientras estoy en la escuela.*

*A mi papá porque sé que aunque estés lejos estas contento de que me gradúe. A mi tío Lester por ser un guía, por preocuparse tanto por mí, por ser tan inteligente, por ser más que mi tío, por ser como un padre, te quiero y te admiro mucho. A mi tía Ania y a Fabita por preocuparse tanto y pensar siempre en mí, a Fabita por ser como una hermanita, las adoro.*

*A mi querido novio Rancel, por aguantarme estos 5 años, por estar en las buenas y en las malas conmigo, gracias por luchar juntos por esta carrera, por quererme tanto, por ser mío, desearía estar contigo una vida entera, Te amo.*

*A mi suegri por siempre tener esa sonrisa que te caracteriza, por cuidarnos tanto, por siempre ser positiva, por haberle dado la vida a una de las personas que más quiero en mi vida, te quiero.*

*A Tami por cuidarnos tanto, y aguantarnos tanto, por estar siempre ahí. A mi nueva familia, a Martín, a Clara y a Anett los quiero mucho, y espero que estemos mucho tiempo junto.*

*A los compañeros de aula por compartir tantos años juntos. A Héctor por esas largas noches, por ser tan positivo, por ayudarnos tanto y aguantarnos, por estar siempre para nosotros sin importar la hora o el día. A Katia que aunque es muy cayada es una gran persona, gracias por acogernos en su casa y ayudarnos tanto con la tesis. A Rayner por apoyarnos y ayudarnos, porque a pesar de todas las complicaciones y reuniones siempre sacabas un tiempito para nosotros. A Raykenler por su ayuda y consejos. A*

## **AGRADECIMIENTOS**

*todos los que de alguna manera han compartido conmigo esta parte tan especial de mi vida, gracias*

### **De Rancel**

#### **Quisiera agradecer:**

*A la comunidad PHP y especialmente a Yasel, ya que sin su apoyo esta tesis hubiera sido mucho más complicada.*

*A los compañeros del aula.*

*A Leandro, por las noches de primer año sin dormir.*

*A Rafael y Alejandro porque junto a ellos comencé en la programación competitiva y fue de las etapas en la que más he aprendido.*

*Al Yosbe y al Guille, porque juntos, logramos llegar a una regional.*

*A Hector, por las madrugadas en tu casa siempre acompañadas de una natilla o algo más de comer y porque pasaste de tutor a amigo sin darme cuenta.*

*A Raynel, porque desde hace mucho se convirtió en un amigo, una fuente de consejos en los momentos más difíciles. Juntos hemos creado un par de sueños.*

*A los amigos y más que amigos hermanos, Javier, Oscar y Arturo.*

*A Osky, porque juntos intentamos cumplir nuestros sueños y atravesamos dificultades que son mejor no mencionar.*

*A Javi, que es una de las personas con las que me hubiese encantado compartir este momento y aunque no está lo sigo queriendo como un hermano.*

*A Arturo, que lleva dando tumbos conmigo desde la primaria.*

*A mi suegra, Ene, porque me ha acogido como un hijo y por sus consejos, que siempre me han dejado pensando.*

*A mi pareja de tesis y de vida durante estos 5 años Talía, porque me ha impulsado en los momentos de menor motivación, por tu amor, tu cariño y tus consejos, los cuales tienden a sacarme siempre de algún aprieto. Te Amo.*

*A mi madre, porque más que madre es una amiga y todo cuanto he hecho ha sido por ella, espero que te sientas orgullosa de mi, este es solo el comienzo y prometo darte mucho más de que alegrarte.*

**RESUMEN**

En la actualidad se han incrementado considerablemente los datos electrónicos que registran las aplicaciones. La Minería de Procesos ofrece diversas técnicas y herramientas que permiten realizar análisis de estos datos mediante el control de los datos organizacionales, el descubrimiento de procesos y de las estructuras sociales. Teniendo como punto de partida los registros de eventos que guardan los sistemas, se confiere una oportunidad para inferir interacciones sociales, ya que los eventos contienen la información sobre las actividades que ejecutan los usuarios del sistema. Asimismo, estas relaciones pueden ser representadas mediante grafos o matrices empleando el enfoque de la sociometría o sociografía. La Minería de Procesos permite inferir redes de interacción social, pero en la actualidad las métricas definidas para la inferencia de redes sociales solo operan con los datos de actividad de los usuarios contenidos en los registros de eventos. Estas métricas descartan otros datos personales que pudieran ser útiles e incluso determinantes en las interacciones según varias teorías interdisciplinarias de la Sociología, la Psicología, la Economía, y la Antropología. Por tal razón se plantea como objetivo general de la investigación desarrollar un algoritmo de Minería de Procesos para la inferencia de redes sociales a partir de las actividades realizadas por los usuarios, sus datos personales e institucionales.

**Palabras Claves:** Análisis de redes sociales, eventos, métrica, Minería de Procesos, redes de interacción social, sociometría.

**TABLA DE CONTENIDO**

<b>RESUMEN.....</b>	<b>V</b>
<b>INTRODUCCIÓN.....</b>	<b>1</b>
<b>CAPÍTULO 1. Fundamentación teórica.....</b>	<b>5</b>
<b>1.1 Las redes sociales y la sociometría. Análisis conceptual .....</b>	<b>5</b>
<b>1.2 Patrones de comportamiento.....</b>	<b>5</b>
<b>1.3 Análisis de redes sociales.....</b>	<b>7</b>
<b>1.4 Representación de las redes sociales.....</b>	<b>8</b>
<b>1.5 Clasificación de las redes sociales.....</b>	<b>10</b>
<b>1.6 Formas de inferir redes sociales .....</b>	<b>11</b>
<b>1.6.1 Estudios basados en los correos electrónicos.....</b>	<b>11</b>
<b>1.6.2 Estudios basados en patrones de movilidad .....</b>	<b>12</b>
<b>1.6.3 Estudios basados en Minería de Procesos .....</b>	<b>13</b>
<b>1.6.3.1 Minería de Procesos.....</b>	<b>13</b>
<b>1.6.3.2 Etapas de la Minería de Procesos .....</b>	<b>15</b>
<b>1.6.3.3 Minería de Procesos para el análisis de redes sociales .....</b>	<b>17</b>
<b>1.6.3.4 Métricas.....</b>	<b>18</b>
<b>1.6.3.4.1 Métricas basada en causalidad .....</b>	<b>18</b>
<b>1.6.3.4.2 Métricas basadas en casos en común .....</b>	<b>18</b>
<b>1.6.3.4.3 Métricas basadas en actividades en común.....</b>	<b>18</b>
<b>1.6.3.4.4 Métricas basada en tipos de eventos especiales .....</b>	<b>19</b>
<b>1.7 Herramientas para el análisis de redes sociales .....</b>	<b>19</b>
<b>1.7.1 R .....</b>	<b>20</b>
<b>1.7.2 SoNIA .....</b>	<b>20</b>
<b>1.7.3 ProM.....</b>	<b>21</b>
<b>1.7.3.1 Complementos de ProM para la Minería de redes sociales.....</b>	<b>22</b>
<b>1.7.3.1.1 Opciones de visualización .....</b>	<b>24</b>
<b>1.8 Herramientas y tecnologías a utilizar .....</b>	<b>31</b>
<b>1.8.1 Herramientas .....</b>	<b>31</b>
<b>1.8.2 Tecnologías .....</b>	<b>31</b>
<b>1.9 Conclusiones del capítulo.....</b>	<b>33</b>
<b>Capítulo 2: Propuesta de solución .....</b>	<b>34</b>
<b>2.1. Introducción .....</b>	<b>34</b>
<b>2.2 Presentación de solución.....</b>	<b>34</b>

2.3 Propuesta de solución.....	35
2.3.1 Análisis de varios registros de eventos .....	35
2.3.2 Métrica para detección de autoridad .....	35
2.3.2.1 Definición de métrica para detección de autoridad .....	35
2.3.2.2 Pseudocódigo del algoritmo para el cálculo de $\beta_{xzq}$ .....	36
2.3.2.3 Seudocódigo del algoritmo para el cálculo de $R_{zq}$ .....	38
2.3.3 Índice de Autoridad.....	40
2.3.3.1 Pseudocódigo del algoritmo para el cálculo de $M_i$ .....	41
2.4 La graficación en el complemento.....	43
2.5 Opciones y paneles de visualización .....	44
2.5.1 Filtrado de nodos .....	45
2.5.2 Leyenda .....	46
2.5.3 Resumen estadístico .....	47
2.6 Conclusiones del capítulo.....	48
<b>CAPÍTULO 3. Validación de la solución .....</b>	<b>49</b>
3.1 Introducción .....	49
3.2 Aplicación en un entorno real.....	49
3.3 Extracción de los registros de eventos.....	49
3.3.1 Calidad de los registros de eventos .....	50
3.3.2 Estado de la base de datos de la comunidad UCI de PHP y extracción de registro de eventos .....	52
3.4 Utilización de un sociograma para la validación del componente desarrollado .....	53
3.5 Aplicación del sociograma.....	55
3.6 Correlación entre los resultados del sociograma y los resultados del complemento desarrollado .....	55
3.6.1 Coeficiente de correlación de Spearman.....	55
3.6.2 Coeficiente de correlación lineal o Coeficiente de Pearson.....	56
3.6.3 Diagrama de dispersión.....	57
3.7 Aplicación del coeficiente de correlación de Spearman y coeficiente de Pearson .....	58
3.8 Conclusiones del capítulo.....	62
Conclusiones generales.....	64
Bibliografía.....	65
Anexos.....	68



**ÍNDICE DE FIGURAS**

**Figura 1.** Ejemplo de una red social como un grafo ..... 9

**Figura 2.** Escenario experimental del estudio sobre patrones de movilidad ..... 13

**Figura 3.** Posicionamiento de los tres tipos principales de Minería de Procesos ..... 15

**Figura 4.** Etapas de la Minería de Procesos..... 17

**Figura 5.** Ejemplo del algoritmo Kamada- Kawai para representación gráfica..... 25

**Figura 6.** Ejemplo del algoritmo Circle para representación gráfica..... 26

**Figura 7.** Ejemplo del algoritmo Fruchterman-Reingold para representación ..... 26

**Figura 8.** Ejemplo de una configuración inicial del algoritmo de resorte para representación gráfica..... 27

**Figura 9.** Ejemplo de una configuración final del algoritmo de resorte para representación gráfica..... 27

**Figura 10.** Ejemplo del algoritmo SpringLayout para representación gráfica ..... 27

**Figura 11.** Ejemplo del algoritmo ISOMLayout para representación gráfica ..... 28

**Figura 12.** Ejemplo del algoritmo Ranking View para representación gráfica ..... 29

**Figura 13.** Centralidad de grado..... 30

**Figura 14.** Filtrado de nodos por autoridad..... 46

**Figura 15.** Leyenda por color y filtrado de nodo. .... 46

**Figura 16.** Leyenda por forma. .... 47

**Figura 17.** Resumen estadístico..... 47

**Figura 18.** Contextos identificados en la Comunidad UCI de PHP. .... 50

**Figura 19.** Diagrama entidad relación de la base de datos de transición..... 53

**Figura 20.** Diagrama de dispersión. .... 58

**Figura 21.** Comparación de los resultados obtenidos de correlación..... 61

**Figura 22.** Diagrama de dispersión sociograma / Complemento. .... 62

**ÍNDICE DE TABLAS**

<b>Tabla 1.</b> Clasificación de las redes sociales .....	10
<b>Tabla 2.</b> Herramientas para el análisis y representación de redes sociales. ....	19
<b>Tabla 3.</b> Niveles de madurez de los registros de eventos .....	51
<b>Tabla 4.</b> Relación id-contexto. ....	52
<b>Tabla 5.</b> Ubicación de registros de actividades en la base de datos. ....	52
<b>Tabla 6.</b> Forma de almacenamiento de las trazas en la tabla phpv2_gtp_guests. ....	53
<b>Tabla 7.</b> Coeficiente de correlación del registro de eventos Artículos. ....	59
<b>Tabla 8.</b> Coeficiente de correlación del registro de eventos Descargas. ....	60
<b>Tabla 9.</b> Coeficiente de correlación del registro de eventos Noticias. ....	60
<b>Tabla 10.</b> Coeficiente de correlación de la unión de todos los registros de eventos ...	61

**ÍNDICE DE FÓRMULAS**

<b>Fórmula 1.</b> Baricentro o centroide. ....	30
<b>Fórmula 2.</b> Autoridad del individuo $z$ sobre el individuo $q$ . ....	35
<b>Fórmula 3.</b> Componente de autoridad asociado a los individuos $z$ y $q$ . ....	36
<b>Fórmula 4.</b> Valor neto del usuario $q$ . ....	36
<b>Fórmula 5.</b> Evaluación del índice de autoridad del individuo $i$ . ....	40
<b>Fórmula 6.</b> Valor neto de autoridad del usuario $i$ . ....	40
<b>Fórmula 7.</b> Índice de ganancia o pérdida de autoridad entre $j$ e $i$ . ....	40
<b>Fórmula 8.</b> Coeficiente de correlación de rangos de Spearman. ....	56
<b>Fórmula 9.</b> Coeficiente de correlación de Pearson. ....	56

### **INTRODUCCIÓN**

Con el avance de las Tecnologías de la Informática y las Comunicaciones (TIC) en todas las esferas de la vida social se han incrementado considerablemente los datos electrónicos que registran las aplicaciones y sistemas de gestión de flujos. Desafortunadamente, la información contenida en estos registros no siempre es empleada adecuadamente para analizar procesos subyacentes en los sistemas. Sin embargo, la Minería de Procesos ofrece diversas técnicas y herramientas que permiten mejorar estos análisis mediante el control de los datos organizacionales, el descubrimiento de procesos y de las estructuras sociales a partir de los registros de eventos.

Los registros de eventos revelan cómo los usuarios ejecutan las actividades almacenadas en las trazas del sistema (van der Aalst, y otros, 2004). Asimismo, estas relaciones pueden ser representadas mediante grafos o matrices empleando el enfoque de la sociometría o sociografía. La cual es una técnica de la Sociología surgida en los alrededores del 1930 que permite representar las relaciones interpersonales (van der Aalst, 2011), sentando las bases para el análisis de las redes sociales.

El estudio mediante métodos formales de la manera en que las personas construyen sus relaciones y cómo estas interacciones permiten descubrir diversas potencialidades y patrones, puede contribuir no solo a explicar la ocurrencia de ciertos fenómenos sociales, sino además a emplear los análisis en favor de detectar, estimular, predecir o controlar la magnitud de los comportamientos humanos en diversos contextos y escenarios. En la medida que las personas despliegan sus actividades en la web y las organizaciones emplean aplicaciones para gestionar las actividades de sus negocios, se confiere una oportunidad para inferir interacciones sociales a partir de los registros de eventos que guardan los sistemas, teniendo en cuenta que los eventos contienen la información sobre las actividades que ejecutan los usuarios del sistema.

El ambiente competitivo de los negocios exige una administración eficaz de los recursos humanos y la información como activos fundamentales en la empresa. Estudios revelan que algunos patrones de relaciones encontrados en la empresa no aparecen en los organigramas y se desarrollan a partir de la interacción impuesta y determinada por la organización formal del trabajo (Chiavenato, 2004). Los resultados del estudio demostraron que no siempre la organización social de una empresa se corresponde

## INTRODUCCIÓN

exactamente a su estructura organizativa y que existen otros procesos espontáneos de evolución social ignorados con frecuencia en la administración que también participan directamente en la realización del objetivo de la empresa. El análisis de los patrones informales de relación que se establecen en el tejido social empresarial complementa la explicación del comportamiento de los individuos en el trabajo y fortalece el proceso de toma de decisiones en la empresa.

La Minería de Procesos, ofrece enfoques y herramientas que combinan la gestión de flujos con los conceptos del análisis de redes sociales y permiten inferir redes de interacción social. Sin embargo, las métricas definidas solo operan con los datos de actividad de los usuarios contenidos en un registro específico. Se descartan otros contextos personales que pudieran ser útiles e incluso determinantes en la formación de las interacciones teniendo en cuenta varias teorías interdisciplinarias de la Sociología, la Psicología, la Economía, y la Antropología (van der Aalst, y otros, 2004).

Por lo antes expuesto se plantea el siguiente **problema a resolver**: las técnicas de Minería de Procesos que infieren redes de interacción social a partir de la actividad de los usuarios, no permiten identificar con precisión el comportamiento de los individuos y las relaciones que establecen en sus contextos de interacción, limitando la fiabilidad de los resultados que se derivan de los análisis.

Se delimita como **objeto de estudio**: minería de redes sociales, específicamente en el **campo de acción**: inferencia de redes sociales.

Para solucionar el problema detectado se plantea como **objetivo general**: desarrollar un algoritmo de Minería de Procesos para la inferencia de redes de interacción social que permita elevar la precisión de los análisis a partir de las actividades e información interpersonal en registros de eventos.

Para satisfacer el objetivo general se definieron los siguientes **objetivos específicos**:

- Analizar los enfoques utilizados por los algoritmos de inferencia de redes sociales en la Minería de Procesos para la identificación de ventajas e inconvenientes.
- Diseñar un algoritmo que posibilite la inferencia de redes sociales a partir de las actividades realizadas por los usuarios y sus interacciones en otros contextos.

## INTRODUCCIÓN

- Implementar un complemento a partir del algoritmo diseñado para la inferencia de redes sociales a partir de las actividades realizadas por los usuarios y sus interacciones en otros contextos.
- Evaluar el comportamiento del algoritmo implementado a partir de su utilización en un entorno real.

**Idea a defender:** si se desarrolla un algoritmo de Minería de Procesos para la inferencia de redes de interacción social a partir de las actividades realizadas por los usuarios, sus datos personales e institucionales entonces, se elevará la precisión de los análisis del comportamiento de los individuos en sus contextos de interacción.

Durante el desarrollo del presente trabajo se definen las siguientes **tareas específicas de la investigación:**

- Análisis de los principales conceptos y trabajos relacionados con el análisis de redes sociales en la Minería de Procesos.
- Caracterización de las principales métricas relacionadas con el análisis de redes sociales en la Minería de Procesos.
- Análisis del comportamiento de los algoritmos existentes para la inferencia de redes sociales en la Minería de Procesos.
- Análisis de marcos de trabajo existentes que permiten la inferencia de redes sociales en la Minería de Procesos.
- Diseño de un algoritmo para la inferencia de redes sociales a partir de las actividades realizadas por los usuarios, sus datos personales e institucionales.
- Implementación del algoritmo diseñado permitiendo su integración a un marco de trabajo existente para la Minería de Procesos.
- Generación de los datos necesarios para validar la propuesta.
- Aplicación de las pruebas necesarias y análisis de los resultados obtenidos en la validación.

Para el desarrollo del presente trabajo se emplean los siguientes **métodos científicos:**  
**Analítico-sintético:** permitió analizar individualmente los principales conceptos relacionados con el área de estudio a tratar, posibilitando un análisis profundo de cada uno, para luego llevar a cabo el estudio de las relaciones que se establecen entre ellos.

**Hipotético-deductivo:** se utilizó para guiar la investigación desde el planteamiento del problema hasta la verificación de la solución a partir de las validaciones, orientando la secuencia lógica de las tareas que se realizaron.

**Histórico-lógico:** se empleó para investigar las aplicaciones existentes que permiten la detección y/o inferencia de redes de interacción social y la utilización de estas de los datos interpersonales.

**Experimentación:** se empleó para verificar la correspondencia entre los resultados obtenidos con la aplicación del algoritmo propuesto y los resultados de la aplicación de técnicas tradicionales de la Sociología.

El presente documento está estructurado en tres capítulos cuyos contenidos son:

- **Capítulo 1:** se exponen los principales conceptos relacionados con el análisis de redes sociales, la Minería de Procesos y los algoritmos para inferencia de redes sociales. Se realiza un análisis de las métricas existentes para inferir redes de interacción social, exponiendo sus principales potencialidades y dificultades.
- **Capítulo 2:** se presenta la propuesta de solución. Se describen los principales algoritmos diseñados: detección de autoridad y unificación de más de una red social de varios registros de evento. Se presenta la implementación como un componente para el marco de trabajo ProM. Se exponen las opciones de visualización de redes sociales presentes en el componente, haciendo énfasis en las novedades que ofrece el complemento.
- **Capítulo 3:** se presenta el procedimiento realizado para la validación de la solución mediante métodos tradicionales de la Sociología y la Psicología, los cuales se correlacionaron con los resultados obtenidos mediante la aplicación del complemento realizado. Se realiza un análisis de los resultados obtenidos.

### **CAPÍTULO 1. Fundamentación teórica**

#### **1.1 Las redes sociales y la sociometría. Análisis conceptual**

En términos analíticos, una red social es una estructura social compuesta por un conjunto finito de actores y configurada en torno a una serie de relaciones entre ellos (Leonidas Aguirre, 2011).

Una forma de representar una red social es a través de un grafo donde los actores o nodos pueden ser personas o grupos de éstas (por ejemplo: empresas, comunidades, organizaciones, países, ciudades, etcétera). Los vínculos o aristas son las relaciones que se manifiesten entre los actores (por ejemplo: amor, poder, alianzas, amistad, parentesco familiar, contacto por correo electrónico, creencias religiosas comunes, rivalidad, etc.).

En términos teóricos, las redes sociales configuran contextos de comunicación e intercambio entre actores, marcan pautas operativas, normas y valores que condicionan la conducta de los actores en ellas, y posiciones funcionalmente diferenciables que son claves para entender el comportamiento de los nodos dentro de cada red y el desempeño de la red en su conjunto (Leonidas Aguirre, 2011).

La sociometría, también conocida por sociografía se refiere a métodos de representar datos en las relaciones interpersonales en forma de gráfico o matriz (van der Aalst, y otros, 2005). El término sociometría fue acuñado por Jacob Levy Moreno quien llevó a cabo el primer estudio sociométrico en la década de 1930 en una universidad de Nueva York (Estados Unidos). Jacob L. Moreno decidió realizar una radiografía grupal, para mostrar las distintas relaciones que se establecen entre los sujetos que conforman el grupo, exponiendo los lazos de influencia, preferencia y las fuerzas dinámicas que mantenían en movimiento al grupo (van der Aalst, y otros, 2005).

#### **1.2 Patrones de comportamiento**

El comportamiento es la manera de proceder de un individuo con relación a su entorno, su reacción a los estímulos y su respuesta a un tejido de interacciones. Para caracterizar la influencia en una red social, se aprovechan algunos patrones de conducta interiorizados de manera colectiva en la psiquis humana. Estos patrones están enunciados en seis principios definidos por (Cialdini, 2009).

### **Reciprocidad**

La reciprocidad se refiere a la necesidad que existe en las relaciones sociales de restaurar el equilibrio. Por ejemplo, cuando se recibe algo, se siente la necesidad de devolver algo más a cambio. La reciprocidad se fundamenta en que dos personas se tratan de acuerdo a los comportamientos anteriores que haya tenido una con respecto a la otra. Las normas sociales impelen a corresponder al “benefactor”, cuando esa persona ha ayudado, obsequiado o hecho cualquier tipo de favor. A la persona que no se comporta de acuerdo con esta norma se le considera habitualmente como ingrata o aprovechada.

La norma de reciprocidad es uno de los pilares para el funcionamiento de las sociedades. Es importante porque beneficia a los individuos y al grupo en su conjunto. Garantiza que cualquier recurso que uno comparta en un momento determinado le será devuelto por los demás cuando lo necesite.

### **Coherencia**

También llamado del compromiso o de la consistencia. El ser humano necesita ser consecuente en lo que hace, lo que compra y en general en su comportamiento. De una forma obsesiva, se tiene el deseo de parecer coherente con respecto a los demás. A la hora de tomar cualquier decisión, se notará la “presión” de comportarse de acuerdo a los principios propios.

La coherencia se basa en el deseo de ser y parecer una persona de actitudes y comportamientos consecuentes a lo largo del tiempo. Los individuos tratan de ser coherentes ante los demás, pero también de cara a ellos mismos. La coherencia es un rasgo de personalidad muy valorado socialmente, que se asocia a otros rasgos como el ser lógico, racional, estable y honrado. Por el contrario, a una persona poco coherente se la considera superficial, poco inteligente, indecisa y débil.

### **Aprobación social**

Es un principio irreprochable en psicología, postula que el individuo actúa de la misma manera que lo hace la sociedad (o subgrupos sociales) para lograr la aceptación de la misma, a menudo aunque considere que la sociedad esté equivocada. Según el economista John Maynard Keynes “Es más fácil equivocarse con la multitud que enfrentarse a la multitud y decir la verdad”.



### **Simpatía**

La simpatía se divide en cuatro instrumentos básicos: los elogios y halagos, el atractivo físico, la semejanza y la familiaridad.

Raramente un individuo realizará un regalo a otro que no es de su agrado. La simpatía es clave para vender, aunque la simpatía en extremo produce un efecto de rechazo. En pocos campos esto se hace tan evidente como en la política. Frecuentemente, los políticos aparecen rodeados de actores y famosos para apropiarse de parte de la simpatía que estas personas provocan.

Un factor que aumenta la simpatía es la familiaridad. La exposición repetida a un estímulo aumenta la atracción hacia él. La tendencia a evaluar de manera más positiva los objetos que nos son familiares que los nuevos se producen incluso cuando la persona no es consciente de haber visto ese estímulo.

### **Autoridad**

La autoridad tiene muchas manifestaciones diferentes, no siempre relacionadas con el poder directo sino también con la credibilidad. En el principio de autoridad entran en juego dos elementos, jerarquía y símbolos. La jerarquía se basa en la creencia de que las personas que llegan a puestos superiores tienen más conocimiento y experiencia que el resto. Los símbolos aportan credibilidad y pueden ser el uniforme de un policía, la bata de un médico, los títulos que posea un académico, entre otros.

### **Escasez**

Las oportunidades parecen más valiosas cuanto más difíciles resultan conseguirlas. Si un cliente percibe una baja oferta, o una elevada demanda de un bien, inmediatamente se mostrará interesado y estará dispuesto a pagar un precio más alto.

### **1.3 Análisis de redes sociales**

El análisis de redes sociales (ARS) es un acercamiento al estudio de las interacciones sociales humanas. El ARS brinda una colección de métodos, técnicas y herramientas de propósitos sociométricos para analizar redes sociales (van der Aalst, y otros, 2005). El objeto de estudio del ARS son las redes sociales, su morfología, sus patrones de comportamiento general y su relación con los individuos que las componen (Leonidas

## **CAPÍTULO 1. Fundamentación teórica**

Aguirre, 2011). El análisis de redes sociales puede ser utilizado para investigar patrones de afinidad, estructuras de comunidades, o de organización de redes sociales formales e informales. Estas redes sociales pueden estar asociadas a corporaciones, grupos familiares, grupos filiales, estructuras de orden y control, u organizaciones encubiertas (Cook, et al., 2007).

El ARS detecta posiciones dentro de las redes de relaciones y las características funcionales de éstas en relación a toda la red. Permite diferenciar posiciones, estrategias, flujos de transacciones y distribución de poder entre un conjunto de actores en un espacio y tiempo determinados; además de dar cuenta de cómo las redes asisten en configurar identidades, normas y valores compartidos en el ámbito que se esté aplicando (Leonidas Aguirre, 2011).

Uno de los problemas más importantes en el análisis de redes sociales es determinar los roles funcionales de los individuos en la red. Diferentes individuos en la red social a menudo desempeñan diferentes roles. Ejemplo de los roles intuitivos incluyen a los líderes, seguidores, reguladores, personas populares, entre otros (Cook, et al., 2007). Los resultados del ARS pueden también identificar otros roles menos intuitivos aunque estos sigan siendo comunes, como son por ejemplo los nodos puentes los cuales conectan sub-grupos en la red.

En sus inicios la Teoría de Redes solo contaba con una base metodológica pues sus aportes prácticos eran locales y limitados. La incapacidad de explicar la complejidad en las relaciones a gran escala solo pudo ser resuelta en la década de 1970 con la llegada de los algoritmos de computación, orientando la formalización y medición de los datos con mayor precisión (Leonidas Aguirre, 2011).

### **1.4 Representación de las redes sociales**

Una red social se puede representar como un grafo, donde cada nodo, también llamado actor o vértice, puede ser representado como un individuo o grupo de individuos. Una arista, también llamada relación o vínculo, conecta a dos nodos y representa el enlace entre dos individuos en una red social. El grafo puede ser dirigido o no dirigido, por ejemplo: a una persona A le gusta la persona B pero no viceversa (Mejias Olivares, 2010). Además, las relaciones pueden ser binarias (existe una relación o no) o con peso. El peso de las aristas es usado para cuantificar la relación en dependencia de lo que se esté midiendo. El grafo resultante es nombrado sociograma (Van der Aalst, y otros, 2004).



**1.5 Clasificación de las redes sociales**

Una red social puede ser clasificada según las características que presenta.

**Tabla 1.** Clasificación de las redes sociales (Mejias Olivares, 2010).

<b>Clasificación</b>	<b>Argumentación</b>	<b>Sub-categoría</b>
<b>Redes basadas en su tamaño</b>	Este tipo de redes depende del diámetro de la red, es decir, la distancia mayor entre dos actores en la red. No existe una medida exacta para poder determinar cuándo una red social es grande o pequeña, por lo que dicho valor dependerá de lo que se quiera representar con la información.	-Redes a pequeña escala. -Redes a gran escala.
<b>Redes basadas en la evolución</b>	Este tipo de redes depende de los cambios que sufre la red a través del tiempo. Estas redes pueden ser de cualquier tamaño y tener distintas formas.	-Redes Estáticas. -Redes Dinámicas.
<b>Redes basadas en su topología</b>	Este tipo de redes depende de la complejidad de la red.	-Redes Simples. -Redes Complejas.
<b>Redes basadas en su origen</b>	Este tipo de redes depende de su fuente de datos de origen. Muchas de estas redes pueden representar comunidades virtuales y/o del mundo real	-Redes fuera de línea (Off-line por sus siglas en inglés). -Redes en línea (On-line por sus siglas en inglés).

En las redes fuera de línea las relaciones sociales son establecidas sin la intervención de un medio electrónico. La administración y conocimiento de las relaciones recae exclusivamente en el conocimiento del individuo sin ayuda de un sistema informático que le permita llevar la gestión de contactos (Mejias Olivares, 2010).

Las redes en línea son redes que dependen altamente de medios electrónicos y se mantienen ligadas a los cambios en la tecnología de los sistemas (Mejias Olivares, 2010).

## **CAPÍTULO 1. Fundamentación teórica**

En sus inicios la mayoría de los estudios sobre redes sociales estaban basados en redes a pequeña escala. En años recientes se han incrementado los estudios basados en la estructura de las redes sociales. Esto se debe al alto número de sistemas en la web que permiten generar comunidades virtuales que pueden ser representadas como una estructura de red social. Estos sistemas web brindan una nueva fuente de información para poder estudiar las propiedades de las redes sociales del mundo real dada la cantidad de información que poseen y la cantidad de usuarios que interactúan (Mejias Olivares, 2010).

El avance de las tecnologías de la informática y las comunicaciones ha permitido acercar a las personas cada vez más, permitiendo la comunicación e interacción entre las mismas desde cualquier parte del mundo. Por lo general las redes sociales en línea involucran un gran número de nodos y poseen propiedades que no pueden ser obtenidas con un simple escalamiento de la red.

### **1.6 Formas de inferir redes sociales**

El campo de las ciencias informáticas y el avance de las tecnologías ha dado inicio a un gran número de aplicaciones que permiten y favorecen el desarrollo de las relaciones interpersonales y el entendimiento de estas (Carrasco Jiménez, y otros, 2012). La disponibilidad de los datos electrónicos que estas aplicaciones generan ha incrementado las formas de recolección de la información (van der Aalst, y otros, 2005). Introduciendo nuevas formas de inferir redes de interacción social.

#### **1.6.1 Estudios basados en los correos electrónicos**

Uno de los campos en los que se ha trabajado el descubrimiento de redes de interacción social es la mensajería electrónica, específicamente en los correos electrónicos. Es posible analizar las interacciones entre los usuarios de correos electrónicos, una de las primeras herramientas realizadas para este propósito es ContactMap (Nardi et al., 2002). Aunque también BuddyGraph (<http://www.buddygraph.com>) y MetaSight (<http://www.metasight.co.uk>) tienen el mismo fin. Estas herramientas se basan en los registros de tráfico de correo como punto de inicio para inferir redes de interacción social. Aunque estas solo monitorean mensajes no estructurados y no pueden distinguir entre diferentes actividades (por ejemplo, discernir interacciones de relaciones de trabajo sobre interacciones sociales). Además de que resulta difícil pero indispensable distinguir entre mensajes electrónicos para actividades particulares dentro del proceso de negocio (por ejemplo, la decisión con respecto a una petición de un préstamo) y

## ***CAPÍTULO 1. Fundamentación teórica***

mensajes electrónicos representando detalles operacionales menos relevantes (por ejemplo, organización de una reunión) (van der Aalst, y otros, 2005).

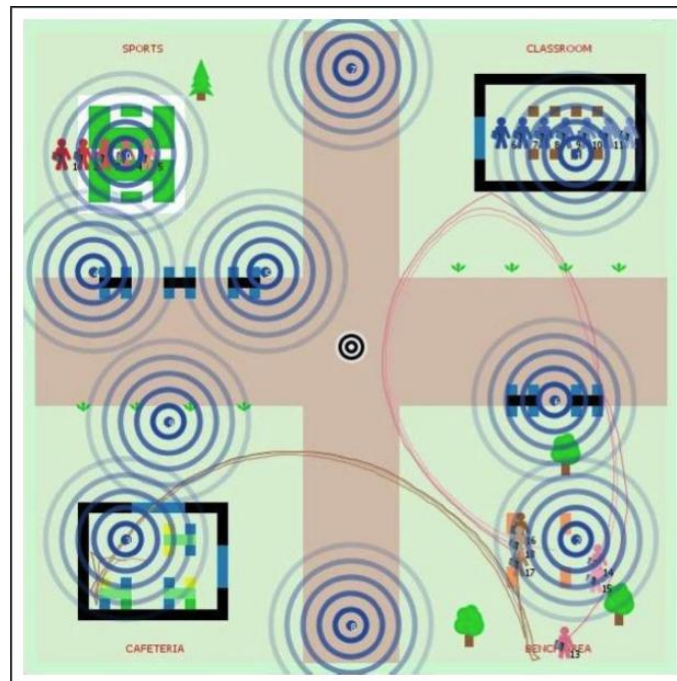
### **1.6.2 Estudios basados en patrones de movilidad**

Los estudios sobre patrones de movilidad buscan inferir interacciones sociales a partir de los móviles que emplean las personas. Un equipo del Instituto Tecnológico y de Estudios Superiores de Monterrey (México) ha desarrollado agentes inteligentes para inferir interacciones a partir de patrones de movilidad, proximidad y acceso inalámbrico. En este estudio la interacción social está definida como el proceso mediante el cual los comportamientos de movilidad de las personas están influidos por la proximidad entre ellas (Carrasco Jiménez, y otros, 2012).

Las redes sociales se construyen al identificar patrones similares descubiertos tras aplicar análisis de grupo en la información registrada en los dispositivos que acceden a la red vía inalámbrica.

Similar a este estudio se han demostrado que es posible inferir relaciones interpersonales basándose en otras fuentes de información, como las trazas de Bluetooth, Sistema de posicionamiento global (GPS, por sus siglas en inglés), informaciones de teléfonos móviles (por ejemplo mensajes y registros de llamada) (Carrasco Jiménez, y otros, 2012).

En la figura 2 se observan diez puntos de acceso inalámbrico y cuatro áreas principales, un área de deporte, una cafetería, un aula y un área de bancos. En este estudio se trabaja con la información recolectada de los puntos de acceso inalámbrico. La información recolectada contiene cuatro tipos de datos con la siguiente estructura: {User, Timestamp, MacAddress, RSS}.



**Figura 2.** Escenario experimental del estudio sobre patrones de movilidad (Carrasco Jiménez, y otros, 2012).

Esta estructura contiene información acerca del individuo del cual se recogen los datos (User), el tiempo en el cual fueron capturados (Timestamp), un ID único del punto de acceso al cual estaba conectado (MacAddress) y la fuerza de la señal del punto de acceso conectado (RSS) (Carrasco Jiménez, y otros, 2012).

Intuitivamente se asume que si dos usuarios están cercanos el uno al otro al mismo tiempo (mismo Timestamp), los cuales se observan en el mismo punto de acceso con la misma dirección Mac (MacAddress) y una similar fuerza de la señal (RSS), se puede asumir que existe algún tipo de relación entre ellos (Carrasco Jiménez, y otros, 2012).

### 1.6.3 Estudios basados en Minería de Procesos

Para una mejor comprensión de los estudios basados en Minería de Procesos se presentan los siguientes conceptos fundamentales.

#### 1.6.3.1 Minería de Procesos

La Minería de Procesos es una disciplina de investigación que estudia la forma de descubrir, monitorear y mejorar procesos reales a través de conocimiento extraído de los registros de eventos. Estos registros se encuentran ampliamente disponibles en los actuales sistemas de información (van der Aalst, 2012).

## ***CAPÍTULO 1. Fundamentación teórica***

La Minería de Procesos es una disciplina de investigación relativamente joven, que se ubica por una parte entre la inteligencia computacional y la Minería de Datos y por otro lado entre la modelación y el análisis de procesos. La Minería de Procesos incluye el descubrimiento de procesos (por ejemplo, extraer modelos de procesos a partir de un registro de eventos) y la verificación de conformidad (por ejemplo, monitorear desviaciones al comparar el modelo y el registro de eventos). Incluye también la minería de redes sociales u organizacionales, la construcción automática, simulación, extensión y reparación de modelos, así como la predicción de casos y las recomendaciones basadas en historia (van der Aalst, 2012).

Se considera como proceso a un conjunto de actividades mutuamente relacionadas o que interactúan. Estas actividades transforman elementos de entrada en resultados (ISO 2000). Una actividad es un paso bien definido en el proceso (van der Aalst, 2012) y está compuesta por tareas que se entienden como una actividad atómica.

Un evento es una acción almacenada en un registro de eventos. Se pueden referir al inicio, completamiento, cancelación, etc. de una actividad para una instancia de proceso específica. A su vez una instancia de proceso hace referencia a la entidad siendo ejecutada por el proceso que es analizado. Los registros de eventos son una colección de eventos que sirven como entrada para la Minería de Procesos.

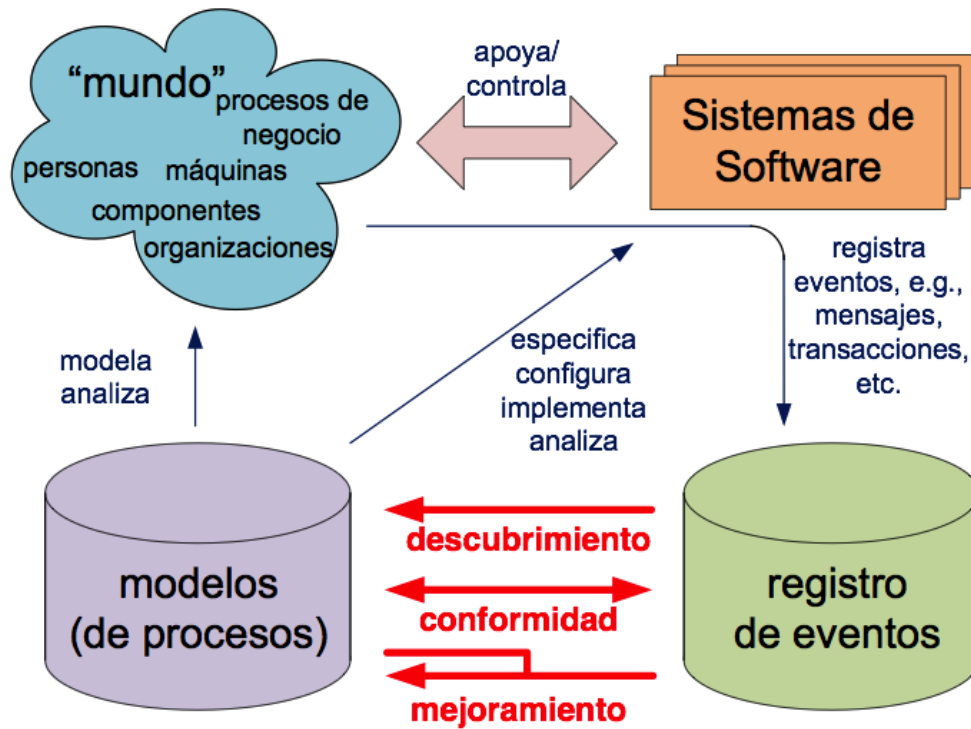
Los registros de eventos pueden ser utilizados para realizar tres tipos de Minería de Procesos. El primer tipo es el descubrimiento, esta técnica toma un registro de eventos y produce un modelo sin usar ninguna información a-priori. El descubrimiento de procesos es la técnica más destacada. El segundo es la verificación de conformidad, en la cual se compara un modelo de proceso existente con un registro de eventos del mismo proceso. La verificación de conformidad puede ser usada para chequear si la realidad, tal como está almacenada en el registro de eventos, es equivalente al modelo y viceversa (van der Aalst, 2012).

Existen distintos tipos de modelos que pueden ser considerados: la verificación de conformidad (la cual puede ser aplicada a modelos procedurales), modelos organizacionales, modelos de procesos declarativos, políticas o reglas de negocio, regulaciones, etcétera (van der Aalst, 2012).

El tercer tipo de Minería de Procesos es el mejoramiento, el cual extiende o mejora un modelo de procesos existente usando la información acerca del proceso real almacenado en algún registro de eventos. Mientras la verificación de conformidad mide



el alineamiento entre el modelo y la realidad, este busca cambiar o extender el modelo a-priori. Por ejemplo, al usar marcas de tiempo en el registro de eventos, se puede extender el modelo para mostrar cuellos de botella, niveles de servicio, tiempos de procesamiento, y frecuencias (van der Aalst, 2012).



**Figura 3.** Posicionamiento de los tres tipos principales de Minería de Procesos: descubrimiento, verificación de conformidad, y mejoramiento (van der Aalst, 2012).

### 1.6.3.2 Etapas de la Minería de Procesos

Cualquier proyecto de Minería de Procesos comienza con una planificación y una justificación para esta planificación (Etapa 0). Después de iniciado el proyecto se necesita extraer los datos de los eventos, modelos, los objetivos y preguntas a partir de los sistemas, expertos del dominio, y la gestión (Etapa 1). Esto requiere un entendimiento de los datos disponibles (“¿Qué puede ser usado para el análisis?”) y un entendimiento del dominio (“¿Cuáles son las preguntas importantes?”) y tiene como resultado por ejemplo los artefactos: datos históricos, modelos hechos a mano, objetivos y preguntas.

En la Etapa 2, se construye el modelo de control de flujo y se le relaciona con el registro de eventos. Aquí, se pueden utilizar técnicas automáticas de descubrimiento de

## ***CAPÍTULO 1. Fundamentación teórica***

procesos. El modelo de procesos descubierto ya podría proveer respuestas a algunas de las preguntas y gatillar acciones de rediseño o ajuste. Además, se podría filtrar o adaptar el registro de eventos usando el modelo (eliminando actividades poco frecuentes o insertando eventos faltantes) (van der Aalst, 2012).

Algunas veces se necesitan significativos esfuerzos para correlacionar eventos que pertenecen a la misma instancia de un proceso. Los eventos restantes están relacionados con entidades del modelo de proceso. Cuando el proceso es relativamente estructurado, el modelo de control de flujo podría ser extendido con otras perspectivas (por ejemplo, datos, tiempo y recursos) durante la Etapa 3. La relación entre el registro de eventos y el modelo establecido en la Etapa 2 se utiliza para extender el modelo (por ejemplo, se utilizan las marcas de tiempo de los eventos asociados para estimar los tiempos de espera para las actividades). Esto podría utilizarse para responder preguntas adicionales y podría gatillar acciones adicionales. En última instancia, los modelos construidos en la Etapa 3 podrían ser utilizados para apoyar las operaciones (Etapa 4). El conocimiento extraído de los datos de eventos históricos se combina con la información acerca de los casos en ejecución. Esto podría utilizarse para intervenir, predecir, y recomendar. Las Etapas 3 y 4 solamente se pueden alcanzar si el proceso es suficientemente estable y estructurado (van der Aalst, 2012).

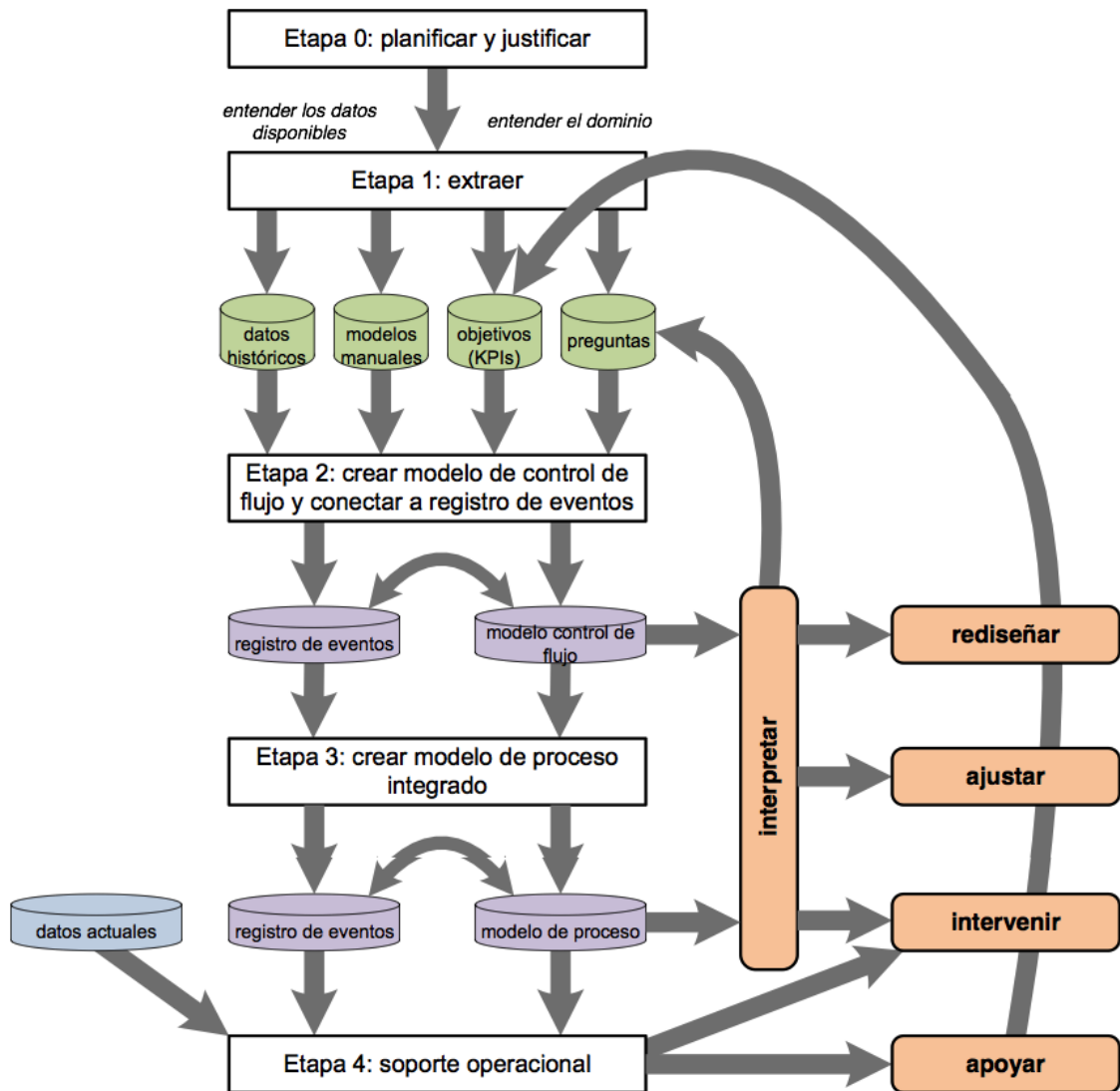


Figura 4. Etapas de la Minería de Procesos (van der Aalst, 2012).

### 1.6.3.3 Minería de Procesos para el análisis de redes sociales

Para realizar análisis de redes sociales a través de la Minería de Procesos es necesario que los registros de eventos contengan información acerca de la persona que inicia o ejecuta el evento, así como el instante de tiempo en el que el evento fue registrado. Si se define un evento “x” como (p, a, r) donde “p” representa la instancia del proceso, “a” la actividad, “r” el recurso (individuo, persona, grupo) y tenemos los eventos completamente ordenados por su marca de tiempo, entonces se puede inferir relaciones sociales entre los recursos. Por ejemplo en caso de existir un evento (p, a1, r1) que es predecesor de (p, a2, r2) en el registro de eventos, entonces existe alguna transferencia de trabajo del recurso r1 a r2. Adicionalmente, si este patrón es común, pero no existe

## **CAPÍTULO 1. Fundamentación teórica**

una transferencia de trabajo de r1 a r3 aunque r2 y r3 tengan los mismos roles en la organización, la relación inferida entre r1 y r2 es más fuerte que entre r1 y r3. Usando esta información es posible elaborar una red social representada a través de un grafo o matriz.

### **1.6.3.4 Métricas**

Para inferir relaciones entre los recursos a partir de los registros de eventos se hace necesario la utilización de alguna forma de medición que permita definir la fortaleza de las relaciones que se establecen entre los individuos. En esta sección se analizan las principales métricas definidas.

#### **1.6.3.4.1 Métricas basada en causalidad**

Las métricas basadas en causalidad analizan para casos individuales, como el trabajo se mueve entre los recursos. Uno de los ejemplos de este tipo de métrica es la transferencia de trabajo. Existe una transferencia de trabajo de un individuo "i" a uno "j" si existen dos actividades subsecuentes donde la primera es terminada por el sujeto "i" y la segunda por "j". La idea es que los recursos estén relacionados si las tareas son transferidas de uno a otro. Otra de las métricas relacionadas a esta categoría es la subcontratación, donde se cuentan el número de veces que un recurso "j" ejecuta una actividad entre dos actividades ejecutadas por "i", esto indica que "i" subcontrata este trabajo a "j" (van der Aalst, y otros, 2004).

#### **1.6.3.4.2 Métricas basadas en casos en común**

Las métricas basadas en unión de casos (instancia de proceso) ignoran las dependencias casuales, esta cuenta cuan frecuente dos recursos realizan actividades para el mismo caso. En estas métricas se asume que si los individuos trabajan juntos en los mismos casos entonces tendrán una relación más fuerte que los individuos que rara vez trabajan juntos (van der Aalst, y otros, 2004).

#### **1.6.3.4.3 Métricas basadas en actividades en común**

Las métricas basadas en actividades no consideran cómo los individuos trabajan juntos en casos compartidos sino que se enfocan en actividades que los mismos realizan. En este tipo de métrica se asume que los individuos que realizan actividades similares tienen una conexión más fuerte que las personas que hacen cosas completamente diferentes. Cada individuo tiene un perfil que se basa en que tan frecuente realizan una

## CAPÍTULO 1. Fundamentación teórica

actividad específica, por lo que la distancia entre un perfil y otro determina el grado de fortaleza de la relación (van der Aalst, y otros, 2004).

### 1.6.3.4.4 Métricas basada en tipos de eventos especiales

Las métricas basadas en tipos de eventos especiales consideran que no todos los eventos son iguales. En las métricas anteriores se consideraba que los eventos correspondían a la ejecución de actividades, sin embargo, existen eventos tales como la reasignación de una actividad de un recurso a otro. Por ejemplo si el sujeto “i” usualmente delega trabajo a “j” pero no viceversa, es probable que “i” tenga una relación de jerarquía con “j”. Desde el punto de vista del análisis de redes sociales estas observaciones son interesantes ya que representan una relación de poder explícita (van der Aalst, y otros, 2004).

### 1.7 Herramientas para el análisis de redes sociales

Dado los programas más empleados en los trabajos de investigación sobre la visualización y el análisis de redes sociales según (Mejias Olivares, 2010) se hizo una selección de los que actualmente se mantienen activos los cuales se muestran en la tabla 2.

**Tabla 2.** Herramientas para el análisis y representación de redes sociales.

Nombre	Tipo de herramienta	Empresa	Análisis de redes sociales
<b>GNU R</b>	Código abierto	Lucent Technologies (www.r-project.org)	Si
<b>ProM</b>	Código abierto	Process Mining Group (managed by the AIS group at TU/e) (www.processmining.org)	Si
<b>SoNIA</b>	Código abierto	University of Stanford (www.stanford.edu)	Si

### **1.7.1 R**

“R” es un entorno de trabajo para computación estadística y gráfica. Fue desarrollado por Bell Laboratories (Formalmente AT&T, actualmente Lucent Technologies) por John Chambers y un grupo de desarrolladores. Uno de los puntos fuertes de “R” es la facilidad con la que se puede desarrollar un diseño con calidad, incluyendo símbolos y fórmulas matemáticas. El mismo se ocupa de los valores predeterminados de las opciones de diseño de menor importancia en los gráficos, pero el usuario conserva el control total de estos. Éste está disponible como software libre bajo los términos de la Licencia Pública General de GNU de la Free Software Foundation. Compila y ejecuta en una amplia variedad de plataformas UNIX y sistemas similares (incluyendo FreeBSD y Linux), Windows y MacOS (Chambers , y otros).

“R” es también un lenguaje de programación. Una de sus características más importantes es su extensibilidad. Numerosos investigadores han creado paquetes de “R”, que son básicamente librerías con implementaciones de diversos algoritmos, y que son publicadas de forma pública en la red CRAN (The Comprehensive “R” Archive Network, por su denominación en inglés). El repositorio de paquetes CRAN cuenta actualmente con 3.725 paquetes disponibles. Entre estos paquetes destacan el paquete “network”, el paquete “sna”, el paquete “statnet” y el paquete “igraph”, para análisis de redes sociales. Los dos primeros están disponibles en CRAN, el tercero actualmente disponible en statnet (WebMining Consultores, 2012).

El paquete “sna” versión 2.3-2 fue distribuido en febrero del 2014, el cual presenta una gama de herramientas para el análisis de redes sociales. Entre sus funcionalidades incluye los índices de nodo y de nivel gráfico, distancia estructural y de covarianza, métodos de detección de equivalencia estructural, la regresión de la red, generación de gráficos al azar y la visualización de la red en 2 y 3 dimensiones (Butts, 2014).

El análisis de este entorno permitió identificar que este permite importar archivos XML, pero no adecuadamente un registro de eventos en formato XES. Además, éste solo tiene en cuenta un contexto de interacción cuando realiza la inferencia de redes sociales.

### **1.7.2 SoNIA**

“SoNIA” (Social Network Image Animator) es un proyecto de código abierto en forma de módulos, está basado en Java y concebido para la visualización de datos de una red

## **CAPÍTULO 1. Fundamentación teórica**

social, tanto dinámicos como longitudinales. Por dinámica se refiere además de información sobre las relaciones (Aristas) entre varias entidades (actores, nodos) también hay información sobre cuándo se producen estas relaciones, o por lo menos el orden relativo en que se producen (McFarland, y otros, 2014).

El objetivo de esta herramienta es leer la información de una red dinámica de varios formatos, ayudar al usuario en la construcción de los diseños de la red y exportar los resultados en imágenes o animaciones. Además, proporciona información sobre las técnicas y ajustes de los parámetros utilizados para la construcción de los diseños y algún tipo de estadística que indica la "exactitud" o grado de distorsión presente en el diseño de la red (McFarland, y otros, 2014).

“SoNIA” puede ser utilizada como plataforma de trabajo para desarrollar, probar y comparar varias técnicas de diseños de redes sociales. También se puede concebir como un navegador de datos de la red social basada en el tiempo, ya que tiene opciones para inspeccionar visualmente un conjunto de relaciones en grandes conjuntos de datos a diferentes secciones de tiempo y niveles de agregación. Esta herramienta fue desarrollada por Dan McFarland y Skye Bender-deMoll, originalmente con el apoyo de la Universidad de Stanford (McFarland, y otros, 2014).

La herramienta solo permite la importación de archivos en texto plano, con la extensión “.net” o su versión nativa “.son”. Además, ésta solo tiene en cuenta un contexto de interacción cuando realiza la inferencia de redes sociales.

### **1.7.3 Marco de trabajo ProM**

“ProM” es un marco de trabajo capaz de importar registros de evento en formato XES o MXML además de poder cargar modelos de procesos de diferentes estándares. Algunas de sus principales funcionalidades son: el descubrimiento de la perspectiva de control de flujo de un proceso, la minería de procesos poco estructurados, el análisis de los recursos y la perspectiva de datos de un proceso, la inspección de los parámetros de rendimiento específicos de un proceso y la verificación del cumplimiento de un estándar dado un registro de eventos y un modelo de proceso (Ailenei, 2011).

Adicionalmente la herramienta soporta una gran variedad de técnicas de filtrado las cuales pueden ser usadas en un registro de eventos en el paso de preprocesamiento. Dicha herramienta provee una gran cantidad de formatos de exportación, desde representación visual de los resultados (ejemplo PNG) a diferentes formatos de registros

(ejemplo CSV) (Ailenei, 2011).

“ProM” es un marco de trabajo de código abierto desarrollado para dar soporte a una amplia variedad de técnicas de Minería de Procesos. Diferentes implementaciones de algoritmos al servicio de diferentes funcionalidades están disponibles en forma de complementos, donde investigadores de todo el mundo pueden hacer su aporte (Ailenei, 2011).

Como una plataforma académica este marco de trabajo está en la vanguardia en la investigación de Minería de Procesos. Esta hace fácil el desarrollo y prueba de nuevos algoritmos, aunque se requieren de ciertos niveles de conocimiento previos en Minería de Procesos para usarlo (Ailenei, 2011).

Esta herramienta permite la importación de registros de eventos en formato XES, pero los complementos desarrollados para el análisis de redes sociales solo tienen en cuenta un contexto de interacción. Esto se evidencia mediante la limitación de importar solo un registro de eventos para realizar los análisis.

### 1.7.3.1 Complementos de ProM para la Minería de redes sociales

Entre los diferentes tipos de complementos que existen para “ProM” están los de Minería de redes sociales. Estos complementos crean redes sociales a partir de un criterio determinado (van der Aalst, 2011).

- **Mine for a Handover of Work Social Network** (por su denominación en español, Minería de una red social de transferencia de trabajo).

En un caso (una instancia de proceso) hay una entrega de trabajo del individuo  $i$  al  $j$ , si hay dos actividades consecutivas donde la primera es completada por  $i$  y la segunda por  $j$ . Esta noción se puede refinar de diversas maneras. Por ejemplo, el conocimiento de la estructura del proceso se puede usar para detectar si existe realmente una dependencia causal entre ambas actividades. También es posible tener en cuenta no sólo la sucesión directa, sino también la sucesión indirecta usando un factor de caída de causalidad ( $\beta$ ), es decir, si hay 3 actividades, una en el medio de una actividad completada por  $i$  y una actividad completada por  $j$ , el factor de caída es la causalidad  $\beta^3$  (van der Aalst, y otros, 2004).

- **Mine for a Reassignment Social Network** (por su denominación en español,



## **CAPÍTULO 1. Fundamentación teórica**

Minería de una red social de Reasignación de trabajo).

Este complemento considera el tipo de evento en el que trabajan los actores. Hasta ahora las otras métricas suponen que los eventos corresponden a la ejecución de las actividades, sin embargo, hay también eventos como la reasignación de una actividad de un individuo a otro que no lo supone. Este complemento implementa la métrica basada en tipos de eventos especiales (van der Aalst, y otros, 2004).

- **Mine for a Similar Task Social Network** (por su denominación en español, Minería de una red social de tareas en común).

Este complemento no tiene en cuenta cómo los individuos trabajan juntos en los casos comunes, pero se centra en las actividades que realizan. La suposición es que las personas que hacen cosas similares tienen relaciones más fuertes que las que hace cosas completamente diferentes. Cada individuo tiene un "perfil" sobre la base de la frecuencia con la que llevan a cabo actividades específicas. Hay muchas maneras de medir la "distancia " entre dos perfiles, las cuales son aplicadas en algunas métricas. Hay cuatro tipos de medidas de distancia. La distancia Euclídea es la distancia "ordinaria " entre dos puntos que uno podría medir con una regla (esta solo da resultados correctos si los actores ejecutan comparables volúmenes de trabajo). El coeficiente de correlación de Pearson se utiliza con frecuencia para encontrar la relación entre los casos. El coeficiente de similitud es una estadística que se usa para comparar la similitud y diversidad de conjuntos de muestras. La distancia de Hamming (Hamming distance, por sus siglas en inglés) no tiene en cuenta la frecuencia absoluta, pero si detecta si es 0 o no (van der Aalst, y otros, 2004).

- **Mine for a Subcontracting Social Network** (por su denominación en español, Minería de una red social de subcontratación de tareas).

El complemento se centra en contar el número de veces que el individuo j ejecuta una actividad en el intervalo entre dos actividades ejecutadas por el individuo i. Esto indica que el trabajo fue subcontratado de i a j (van der Aalst, y otros, 2004).

- **Mine for a Working Together Social Network** (por su denominación en español, Minería de una red social de trabajo conjunto).

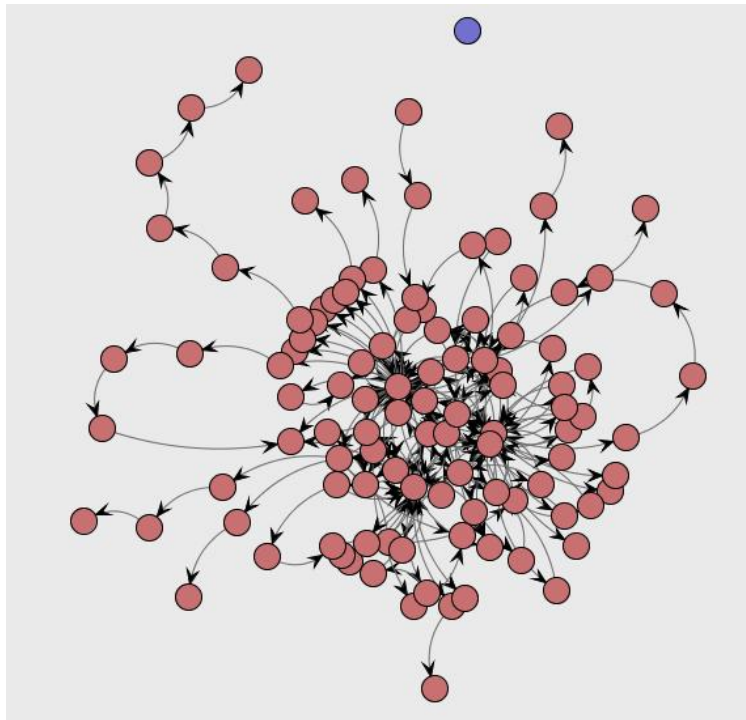
## **CAPÍTULO 1. Fundamentación teórica**

Este complemento ignora las dependencias causales, simplemente cuenta la frecuencia con que dos personas están realizando actividades en un mismo caso. Si los individuos trabajan juntos en los casos, van a tener una relación más fuerte que las personas que rara vez trabajan juntos. Hay tres tipos de métodos para calcular las métricas de trabajo en conjunto (Working together metric, por su denominación en inglés). Uno de ellos es dividir el número de casos conjuntos por el número de casos en los que el individuo  $i$  apareció. Es importante utilizar una notación relativa. Por ejemplo, supongamos que el individuo  $i$  participa en tres de los casos, el individuo  $j$  participa en seis casos, y trabajan juntos en tres ocasiones. En esta situación,  $i$  siempre trabaja junto con  $j$ , pero  $j$  no siempre lo hace. Por lo tanto, el valor de  $i$  a  $j$  tiene que ser mayor que el valor de  $j$  a  $i$ . Otro método para calcular esta métrica puede ser también adicionar la distancia entre las actividades en la cuenta (van der Aalst, y otros, 2004).

### **1.7.3.1.1 Opciones de visualización**

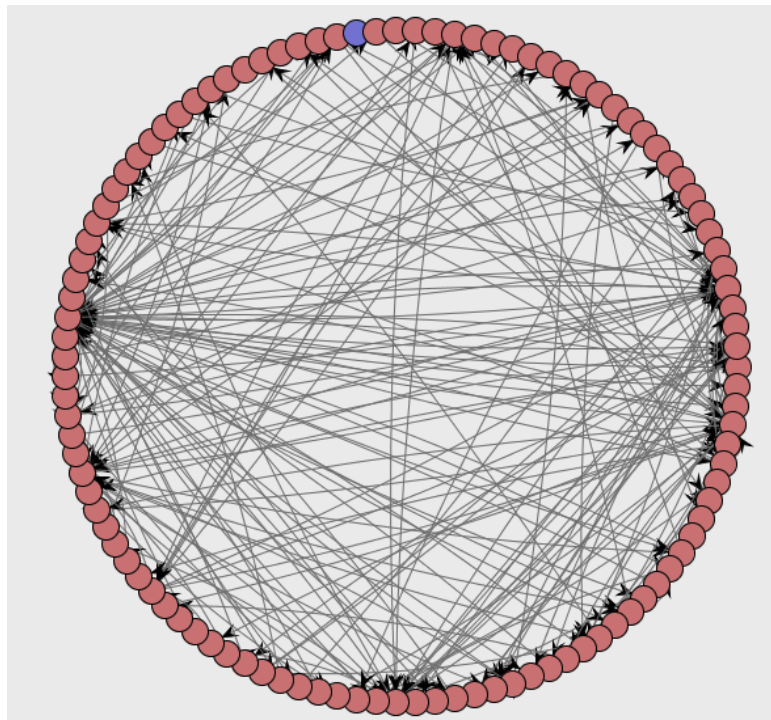
Independientemente de la métrica implementada cada uno de estos complementos comparten el mismo módulo de visualización. Este módulo permite modificar la distribución de los nodos de la red a través de seis algoritmos de representación: KKLayout, Circle Layout, FRLLayout, SpringLayout, ISOMLayout, Ranking View.

**KKLayout:** el Modelo Kamada – Kawai calcula el balance total de la gráfica, como la suma cuadrada de las diferencias entre la distancia ideal y la distancia real para todos los vértices. Se basa en la teoría de que en algunos escenarios la eliminación o reducción del número de ejes que cruzan sus aristas en un grafo no es un buen criterio estético para un algoritmo de diseño. Además, se afirma que el balance total de diseño que está relacionado con las características individuales del grafo es tan importante, o puede considerarse más importante que la reducción de los cruzamientos de vértices (Kamada, et al., 1989).



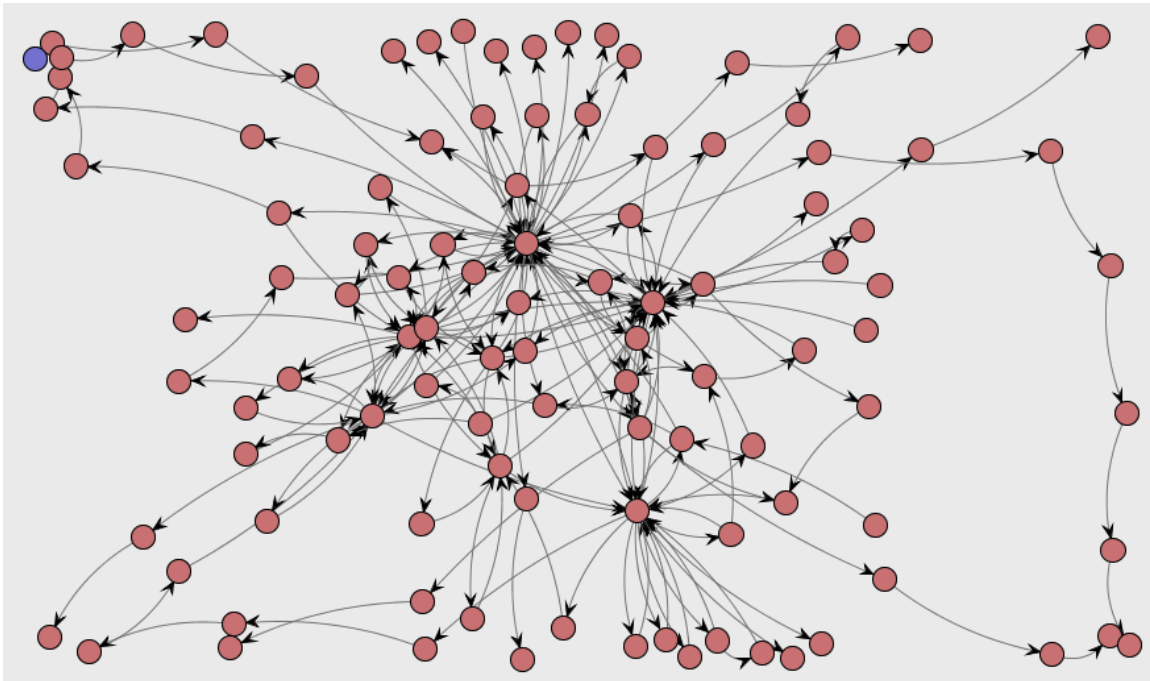
**Figura 5.** Ejemplo del algoritmo Kamada- Kawai para representación gráfica, generado en ProM.

**Circle Layout:** en el diseño en círculo se posicionan todos los nodos en el borde de un círculo quedando las aristas en el interior de este.



**Figura 6.** Ejemplo del algoritmo Circle para representación gráfica, generado en ProM.

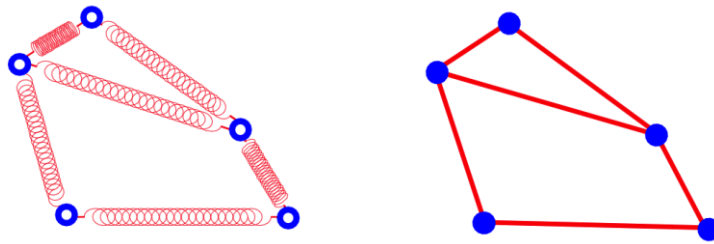
**FRLayout:** el diseño Fruchterman-Reingold se basa en dos principios fundamentales: los vértices conectados por una arista se atraen, pero estos no pueden ser dibujados demasiado cerca el uno al otro. El que tan cercano se dibujen los nodos estará en dependencia del espacio disponible para la representación. Este algoritmo está basado en la física de partículas y en los cuerpos celestiales, los cuales poseen fuerza de atracción y repulsión (Fruchterman, et al., 1991).



**Figura 7.** Ejemplo del algoritmo Fruchterman-Reingold para representación gráfica, generado en ProM.

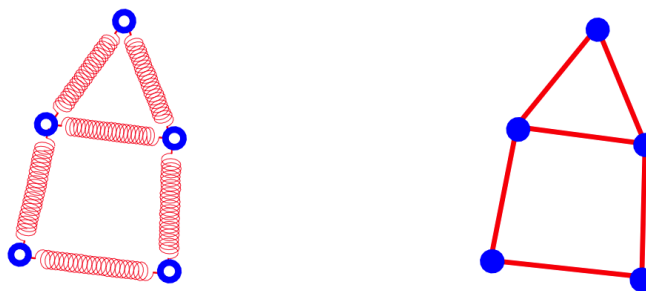
**SpringLayout:** El algoritmo resorte es un algoritmo heurístico basado en un modelo físico. El proceso de dibujo está basado en simular un sistema mecánico donde los vértices son remplazados por anillos y las aristas son remplazadas por muelles. Los muelles atraen a los anillos si estos se encuentran muy separados y los apartan si estos están muy cerca (Eades, 1984). El algoritmo SpringLayout es ampliamente usado y generalmente considerado como uno de los mejores algoritmos para dibujar grafos no dirigidos (Battista, et al., 1999).

### **Configuración inicial**

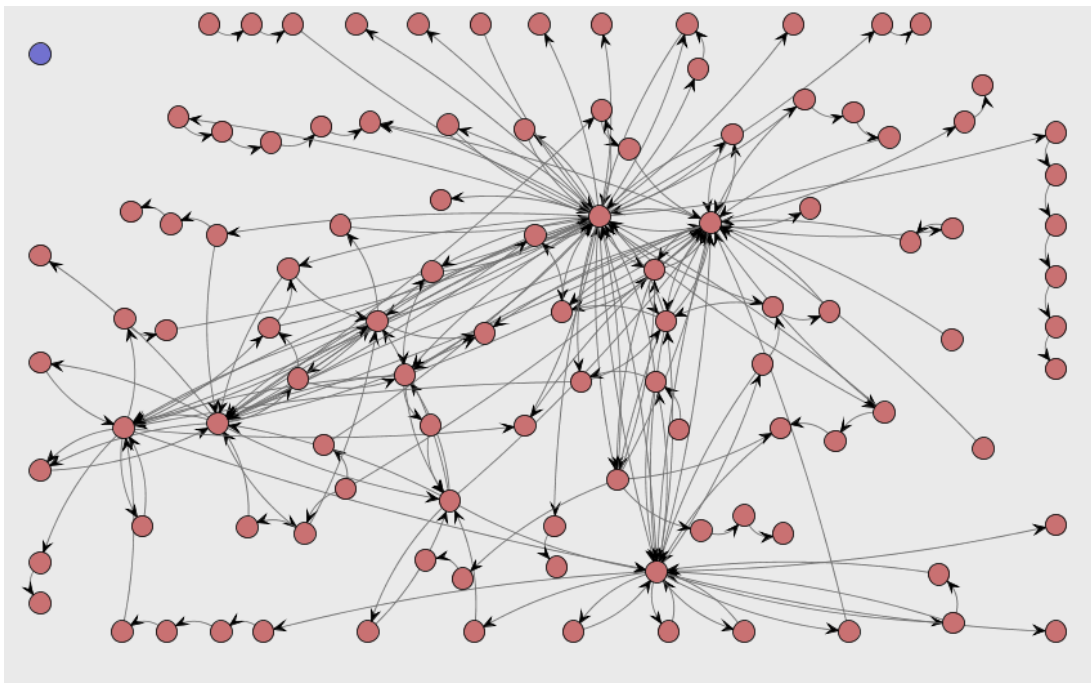


**Figura 8.** Ejemplo de una configuración inicial del algoritmo de resorte (SpringLayout) para representación gráfica (Cruz, et al., 2009).

### Configuración final

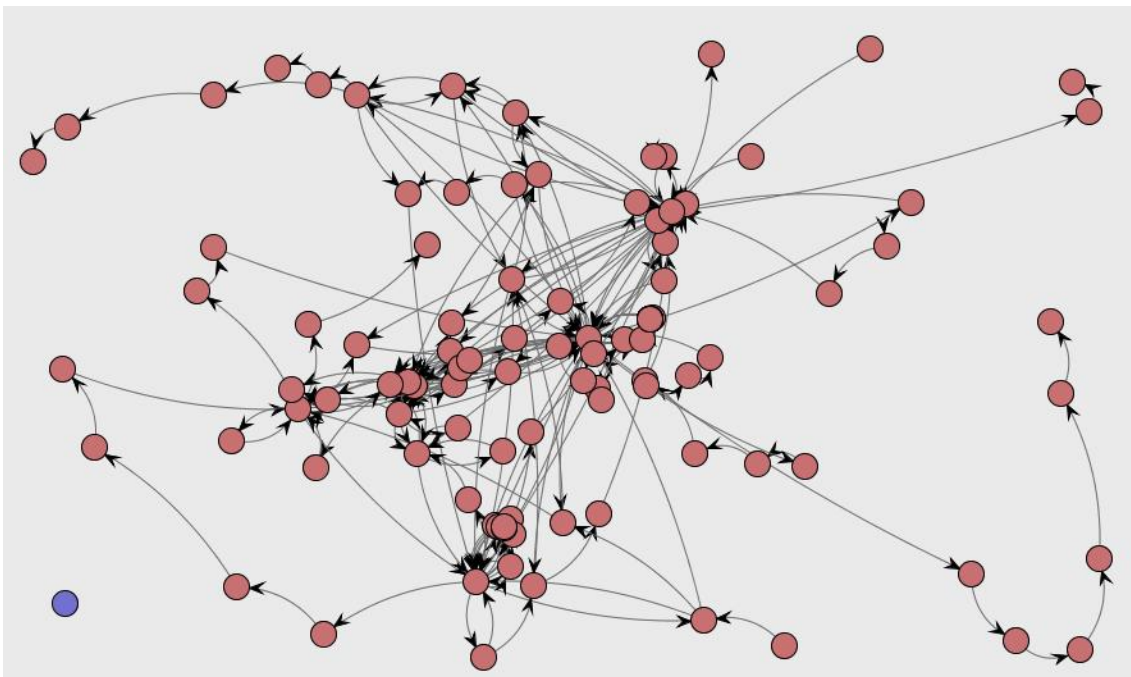


**Figura 9.** Ejemplo de una configuración final del algoritmo de resorte (SpringLayout) para representación gráfica (Cruz, et al., 2009).



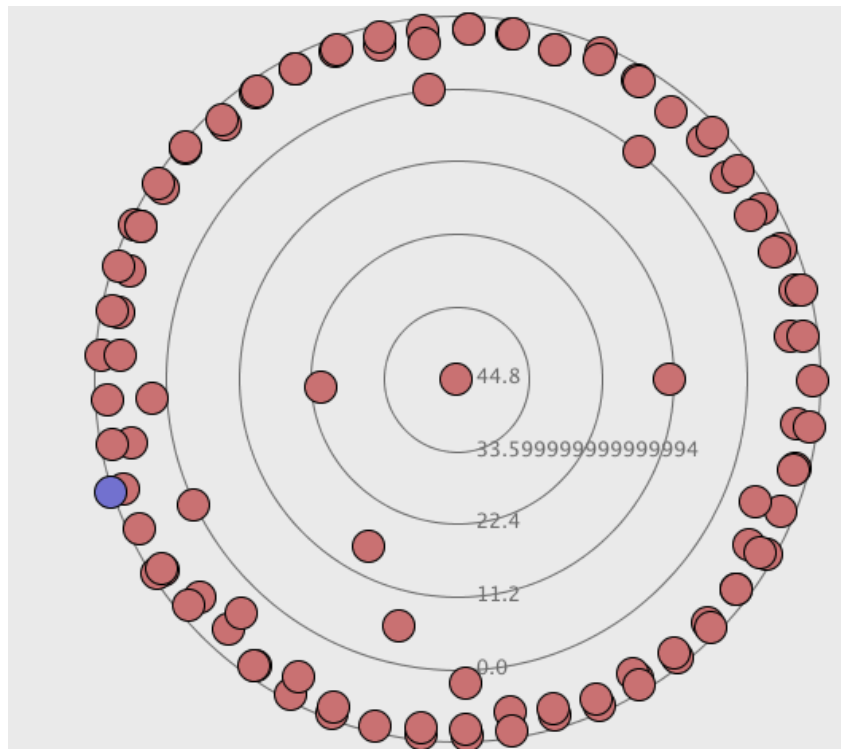
**Figura 10.** Ejemplo del algoritmo SpringLayout para representación gráfica, generado en ProM.

**ISOMLayout:** Los grafos de auto-organización (self-organizing graphs) son un enfoque para el diseño gráfico basado en un algoritmo de aprendizaje competitivo. Este método es una extensión de estrategias de autoorganización conocida de las redes neuronales no supervisadas, nombradas por Kohonen's como mapas de auto-organización (self-organizing map). Se caracteriza por ser muy flexible y adaptable a tipos arbitrarios de espacios de visualización, por ello este modelo es parametrizado de forma explícita por un modelo de métrica de espacios de diseños. Este método consume relativamente pocos recursos computacionales y no necesita ningún procesamiento previo de alta resistencia (Mayer, 1998).



**Figura 11.** Ejemplo del algoritmo ISOMLayout para representación gráfica, generado en ProM.

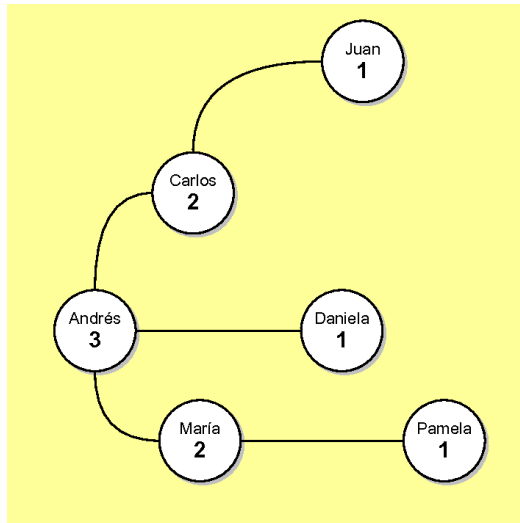
**Ranking View:** Dibuja una diana donde los nodos más cercanos al centro son los que poseen un mayor índice.



**Figura 12.** Ejemplo del algoritmo Ranking View para representación gráfica, generado en ProM.

Para el cálculo de estos índices el módulo de visualización brinda las opciones: Degree, In degree, Out Degree, Closeness, Bary Center, Betweenness.

**Degree:** Centralidad de grado (degree centrality). Un hombre o mujer popular es aquel que tiene muchos amigos o conocidos. Con esta simple intuición, se puede definir la centralidad de grado (*degree centrality*). En términos de grafos, la centralidad de grado de un actor se calcula como su número de vecinos (nodos adyacentes). Si se modela una red social de amistad, la centralidad de cada actor consiste en su número de amigos (Monsalve, 2008).



**Figura 13.** Centralidad de grado. La centralidad de cada actor se calcula como su número de vecinos (Monsalve, 2008).

**In degree, Out degree:** Para los grafos dirigidos el grado del nodo suele ser dividido en dos formas: el grado de entrada (*in-degree*) y el grado de salida (*out-degree*). El grado de entrada del nodo *i* en un grafo dirigido es el número de nodos  $N_V$  que tienen un enlace que se dirigen hacia *i*, el grado de salida del nodo es el número de nodos  $N_V$  que poseen enlaces con él y que son dirigidos desde *i* (Mejias Olivares, 2010).

**Closeness:** centralidad de cercanía (closeness centrality). Esta medida de posición se obtiene de considerar la distancia promedio al resto de la red. El actor que está más cerca de todo otro elemento de la red es el más central (Monsalve, 2008).

Formalmente se puede expresar la centralidad de cercanía como el inverso a la suma de las distancias, o sea,  $c_i = 1 / \sum_j d_{ij}$  donde  $d_{ij}$  es la distancia entre el actor *i* y el actor *j*. En este caso se habla de distancias euclídeas o rutas mínimas (Monsalve, 2008).

**Bary Center:** en el baricentro o centroide la posición de cada vértice *u* es calculada como el promedio de las posiciones de sus vecinos.

$$Prom(u) = \frac{1}{grado(u) * \sum_{v \in U_n} X_1(v)}$$

$$v \in U_n$$

**Fórmula 1.** Baricentro o centroide.

En el caso de que dos vértices tengan el mismo valor **Prom (u)** se les asigna



## **CAPÍTULO 1. Fundamentación teórica**

arbitrariamente algún valor próximo. En este método no se calcula un valor exacto de coordenadas, sino las posiciones relativas de los vértices dentro de cada capa (Martig, y otros, 2000).

**Betweenness:** centralidad de intermediación (betweenness centrality). Existen actores los cuales unen o están en medio de varias sub-redes. Si estos son eliminados, desconectan una red, o interrumpen sustancialmente los flujos que se pudieran producir en ésta. Estos actores, se pueden descubrir mediante la sumatoria de cuántas rutas mínimas pasan por ellos; o sea, por su calidad de intermediarios o puntos intermedios. Por eso, se define la centralidad de intermediación como el número de rutas mínimas en las que el actor participa (Monsalve, 2008).

### **1.8 Herramientas y tecnologías a utilizar**

#### **1.8.1 Herramientas**

##### **Netbeans7.1**

NetBeans es un IDE (traducido al español Entorno de Desarrollo Integrado) libre y de código abierto, que con variados módulos brinda las herramientas necesarias para el desarrollo de aplicaciones profesionales de escritorio, web y aplicaciones móviles con la plataforma Java, así como con PHP y otras (Oracle Corporation and its affiliates, 2012).

##### **XESame**

XESame es un mapeador de XES, el cual proporciona una forma genérica para la extracción de registros de eventos de una fuente de datos. Este está diseñado para ser fácil de usar y no se requieren conocimientos de programación ya que toda la interacción con la fuente de datos de registro de eventos se puede definir a través de una interfaz gráfica de usuario (Buijs, y otros, 2011).

#### **1.8.2 Tecnologías**

##### **Java**

Java es un lenguaje de programación de propósito general basado en clases y orientado a objetos. Está diseñado lo suficientemente simple como para que muchos desarrolladores puedan trabajar con fluidez con el lenguaje. Java está estrechamente

## **CAPÍTULO 1. Fundamentación teórica**

relacionado con los lenguajes C y C++, pero se organiza de forma diferente, con algunos aspectos de C y C++ omitidos y algunas ideas de otros lenguajes incluidos. Este es un lenguaje de programación fuertemente tipado, lo que permite detectar errores de compilación en tiempo de compilación (Gosling, y otros, 2005).

### **MySQL**

MySQL es un sistema gestor de bases de datos (SGBD, DBMS por sus siglas en inglés) muy conocido y ampliamente usado por su simplicidad y notable rendimiento. Es de distribución libre bajo licencia GPL además de contar con un alto grado de estabilidad y rápido desarrollo. MySQL está disponible para múltiples plataformas. La interfaz de programación de aplicaciones (API) se encuentra disponible en C, C++, Eiffel, Java, Perl, PHP, Python, Ruby y TCL. Está optimizado para equipos de múltiples procesadores y es muy destacable su velocidad de respuesta. Se puede utilizar como cliente-servidor o incrustado en aplicaciones. Soporta múltiples métodos de almacenamiento de las tablas, con prestaciones y rendimientos diferentes para poder optimizar el gestor de base de datos a cada caso concreto. Su administración se basa en usuarios y privilegios. Sus opciones de conectividad abarcan TCP/IP, *sockets* UNIX y *sockets* NT, además de soportar completamente ODBC. Es altamente confiable en cuanto a estabilidad se refiere (Gibert Ginesta, et al., 2004).

### **XES: eXtensible Event Stream**

XES es un estándar basado en XML para registros de eventos. Su propósito es proporcionar un formato reconocido, generalmente para el intercambio de datos de registro de eventos entre las herramientas y los dominios de aplicación. Su objetivo principal es para la Minería de Procesos, es decir, el análisis de procesos operativos en base a sus registros de eventos. El estándar ha sido adoptado por la IEEE Task Force on Process Mining como el formato de intercambio de registros de eventos por defecto. Además, XES ha sido diseñado para ser también adecuado para la minería en general de datos, minería de texto, y análisis estadístico. Cuando se diseñó el estándar XES se fijaron los siguientes principios (Buijs, et al., 2011).

- **Sencillez**

Utilizar la forma más sencilla posible para representar la información. Los registros XES debe ser fácil de analizar y generar, además de ser legibles (XES-Standard, 2012).

- **Flexibilidad**

El estándar XES debe ser capaz de captar los registros de eventos de cualquier origen, sin importar de qué dominio de aplicación o de soporte de tecnológico provenga. Por lo tanto, XES pretende ir más allá de la Minería de Procesos y procesos de negocio, y se esfuerza por ser una norma general para los datos de registros de eventos (XES-Standard, 2012).

- **Extensibilidad**

Debe ser fácil de añadir a la norma en el futuro. La extensión del estándar debe ser lo más transparente posible, manteniendo al mismo tiempo compatibilidad con las nuevas herramientas y las antiguas (XES-Standard, 2012).

- **Expresividad**

Los registros de eventos serializados en XES deben tener la menor pérdida de información posible. Por lo tanto, todos los elementos de información deben ser fuertemente tipados y deben poseer un método genérico para adjuntar la interpretación humana con la semántica (XES-Standard, 2012).

### **1.9 Conclusiones del capítulo**

- El estudio de enfoques y tendencias para el análisis de redes sociales permitió identificar diversos campos de aplicación emergentes, denotando el amplio interés de la comunidad científica internacional en el objeto de estudio de la investigación.
- Las métricas definidas de Minería de Procesos para la inferencia de redes sociales, permiten descubrir relaciones basándose en causalidad, casos y actividades en común, así como por tipos de eventos especiales. Sin embargo, limitan su análisis al empleo de un registro de eventos y descartan otros contextos de interacción.
- El estudio de influencia en la red permite determinar la capacidad que tiene un nodo de cambiar el comportamiento de otro.
- El análisis de las principales fuentes bibliográficas permitió identificar los aspectos que caracterizan la influencia en la red, entre los que se señalan la autoridad, reciprocidad, coherencia, simpatía, aprobación social y escasez.

### **Capítulo 2: Propuesta de solución**

#### **2.1. Introducción**

En este trabajo se presenta un nuevo algoritmo para la obtención de diferentes enfoques de la inferencia de redes sociales a partir de las actividades realizadas por los usuarios y sus datos personales e institucionales en múltiples contextos.

#### **2.2 Presentación de solución**

Para el cumplimiento del objetivo general del presente trabajo se realizó el complemento **Mine for an Authority Social Network**. Este complemento estudia el comportamiento de los nodos en la red a través de la autoridad que unos ejercen sobre otros. Dicho complemento está implementado en Java bajo el marco de trabajo ProM.

**Mine for an Authority Social Network** a diferencia de los actuales complementos desarrollados para ProM para el análisis de redes sociales mediante la Minería de Procesos, permite añadir más de un registro de eventos. Esta funcionalidad da la posibilidad de analizar el comportamiento de los usuarios en más de un contexto de interacción. Para determinar la relevancia que posee cada contexto cargado, en la pantalla de configuración del complemento se permite dar un índice de importancia a cada contexto. Una vez asignada la puntuación, también permite otorgar un índice de importancia a las actividades contenidas en los mismos. Estas puntuaciones otorgadas modificarán la fortaleza de las interrelaciones de los nodos.

Como novedad en las herramientas para la visualización de las redes sociales el complemento brinda un sistema de leyenda el cual permite distinguir a los nodos según sus características. Esta leyenda se basa en colores y formas para lograr discernir nodos. Además, este complemento brinda la opción de filtrar nodos según determinadas características y datos personales seleccionados.

Además de la métrica para medir la autoridad de un nodo sobre otros, se puede calcular también el índice de autoridad el cual mide o determina un coeficiente de autoridad para cada nodo de la red.

## **2.3 Propuesta de solución**

### **2.3.1 Análisis de varios registros de eventos**

Un registro de eventos contiene información sobre cierto contexto de interacción. Cuando se quiere analizar un grupo determinado de personas las cuales interactúan en más de un contexto, resulta interesante visualizar el comportamiento de estos usuarios en cada uno de esos escenarios. Para ello el complemento implementa la opción de cargar más de un registro de eventos, a los cuales según el interés y la importancia que posean se les asigna una puntuación. Dicha puntuación será un valor entre cero y cinco donde cero representa la menor importancia y cinco la mayor. Esta puntuación le será dada en el panel de configuración del complemento al iniciar el mismo. Del mismo modo, a las actividades contenidas en cada uno de estos contextos se le atribuirá un valor de relevancia. En caso de no dárselo, las actividades obtendrán una misma puntuación, lo cual significará que tienen la misma importancia.

En el complemento, cada registro o combinación de estos se representará como una red social mediante un grafo. Se podrá seleccionar la combinación de registros a visualizar mediante un panel de cajas de verificación.

### **2.3.2 Métrica para detección de autoridad**

En el estudio, para el desarrollo del algoritmo planteado se define una métrica para la detección de autoridad en varios registros de eventos, para la cual se entiende como premisa que si un individuo X, iniciador de un proceso, logra que otros individuos interactúen con dicho proceso, entonces se infiere que el individuo X tiene autoridad sobre ellos.

#### **2.3.2.1 Definición de métrica para detección de autoridad**

Sea  $R_{zq}$  el valor correspondiente a la evaluación de la autoridad del individuo z sobre el individuo q en el conjunto de contextos a analizar, se tiene que:

$$R_{zq} = \frac{\sum_{x=1}^e (\mu_x * \beta_{xzq})}{e * M}$$

**Fórmula 2.** Autoridad del individuo z sobre el individuo q.

Siendo  $e$  la cantidad de contextos a analizar,  $\beta_{xzq}$  el valor correspondiente al componente de autoridad asociado a los individuos  $z$  y  $q$  en el contexto  $x$ .  $\mu_x$  la

importancia del contexto con referencia a los demás y **M** el mayor valor de importancia asignado a los contextos.

$$\beta_{xzq} = \frac{\sum_{a=1}^y (\Theta_{zqa})}{\sum_{w=1}^p (\sum_{g=1}^h \Theta_{wqg})} * \Psi_q$$

**Fórmula 3.** Componente de autoridad asociado a los individuos **z** y **q**.

Siendo **y** la cantidad de acciones de **q** que aportan a la autoridad de **z**, **p** la cantidad de usuarios que se relacionan con **q** y **h** la cantidad de acciones de **z** que aportan a **w**.  $\Theta_{zqa}$  es el valor correspondiente a la interacción **a** entre **z** y **q**.  $\Psi_q$  el valor neto del usuario **q**, el cual representa que tan influenciable es **q**.

$$\Psi_q = \frac{\sum_{w=1}^p (\sum_{g=1}^h \Theta_{wqg})}{T * \Lambda}$$

**Fórmula 4.** Valor neto del usuario **q**.

Sea **T** la mayor cantidad de acciones realizadas por algún usuario y **Λ** el mayor valor asignado a las interacciones.

### 2.3.2.2 Pseudocódigo del algoritmo para el cálculo de $\beta_{xzq}$

**Entrada del algoritmo:** Registros de Evento.

**Salida:** redes sociales.

1: Crear una estructura **D** capaz de representar un grafo dirigido y ponderado. Crear un arreglo de tipo **D** capaz de almacenar y representar diferentes grafos dirigidos y ponderados.

**Para** cada registro de eventos **Hacer**

2: Crear tres mapas hash **aValue**, **origValues** y **sumTi**. **aValue** tiene como objetivo almacenar el valor asociado a cada actividad. **origValues** almacena la sumatoria del valor de las actividades ejecutadas por cada recurso así como la cantidad de estas con el objetivo de calcular posteriormente el valor neto del recurso. **sumTi** guarda de cada recurso el valor de las actividades que tributan a su índice de autoridad, esto relacionado con el usuario que ejecuta la actividad.

## CAPÍTULO 2. Propuesta de solución

**Para** cada traza **Hacer**

- 3:           **Extraer** el recurso originador de la traza y actualizar su valor neto  
              //El recurso es el usuario que realiza una actividad.  
              //Una actividad es una acción realizada por un usuario.
- 4:           **Añadir** el recurso originador a *sumTi* si no existía antes
- 5:           **Para** cada actividad de la traza empezando desde la número dos **Hacer**
- 6:                   **Si** el recurso de la actividad es igual al originador **Entonces**
- 7:                           **Saltar** a la siguiente actividad
- 8:                   **FinSi**
- 9:           **Extraer** el recurso de la actividad y actualizar su valor neto
- 10:           **Si** el recurso de la actividad no había tributado anteriormente al originador **Entonces**
- 11:                   **Añadir** el recurso de la actividad al listado de *sumTi* del originador y actualizar sus valores
- 12:                   **FinSi**
- 13:           **Si no Hacer**
- 14:                   **Actualizar** los valores del recurso que ejecuta la actividad para el originador en *sumTi*
- 15:                   **FinSi no**
- 16:           **FinPara**
- 17:           **Para** cada recurso en las trazas **Hacer**
- 18:                   **Si** el recurso existe como un originador en *sumTi* **Entonces**
- 19:                           **Para** cada recurso que se relacionó con el originador actual **Hacer**

## CAPÍTULO 2. Propuesta de solución

20:  $\psi = Q / T * \Lambda$   
// Q: suma del valor de todas las actividades realizadas por recurso que se relacionó con el originador.

21:  $\beta = ((Su/SuCant) * \psi)$   
// Su: suma del valor de las actividades realizadas por el recurso actual que tributan a la autoridad del originador.  
// SuCant: suma del valor de las actividades realizadas por el recurso actual.

22: **Añadir** a *D* la relación de autoridad entre el originador y el recurso actual con un valor de  $\beta$

23: **FinPara**

24: **FinSi**

25: **FinPara**

26: **FinPara**

27: **FinPara**

28: **Retornar** *D*

### 2.3.2.3 Seudocódigo del algoritmo para el cálculo de $R_{zq}$

El algoritmo para el cálculo de  $R_{zq}$  es el encargado de la unificación de varios registros de eventos.

**Entrada del algoritmo:** Redes Sociales.

**Salida:** Red Social.

1: Crear una nueva red social *sn* y un mapa hash *nodesIdentifiers sn* guardará la red social producto de la unión y *nodesIdentifiers* tiene como objetivo almacenar todos nodos presentes en cada una de las redes a unir.

**Para** cada red social **Hacer**



## CAPÍTULO 2. Propuesta de solución

- 2: **Para** cada nodo de la red actual **Hacer**
- 3:       **Si** el nodo actual no existe en *nodesIdentifiers* **Entonces**
- 4:               **Añadir** el nodo actual a *sn* y a *nodesIdentifiers*
- 5:       **FinSi**
- 6: **FinPara**
- 7: **FinPara**
- 8: Crear un mapa hash *edgesValue* y una lista de arreglos *edgesIdentifiers*  
**Para** cada red social **Hacer**
- 9:       **Para** cada arista de la red actual **Hacer**
- 10:                $edgeValue = \beta_{xzq} + \mu_x$
- 11:       **Si** la arista actual existe en *edgesIdentifiers* **Entonces**  
          //se define que una arista es igual a otra si su nodo origen y nodo destino son iguales.
- 12:       **Actualizar** *edgesValue* con la suma de *edgeValue* y el valor anterior de la arista en *edgesValue*
- 13:       **FinSi**
- 14:       **Si no Hacer**
- 15:               **Añadir** la arista a *edgesIdentifiers*
- 16:               **Insertar** el valor de la arista a *edgesValue*
- 17:       **FinSi no**
- 18: **FinPara**
- 19: **FinPara**
- 20: **Para** cada arista en *edgesIdentifiers* **Hacer**
- 21:        $R_{zq} = (\text{al valor de la arista actual en } edgeValue) / (e * M)$

22: **Añadir** la arista actual a  $sn$  con el valor  $R_{zq}$

23: **FinPara**

24: **Retornar**  $sn$

### 2.3.3 Índice de Autoridad

Sea  $M_i$  el valor correspondiente a la evaluación del índice de autoridad del individuo  $i$  en el conjunto de contextos a analizar, se tiene que:

$$M_i = Y_i + \sum_{j=1}^n \Phi_{ij}$$

**Fórmula 5.** Evaluación del índice de autoridad del individuo  $i$ .

Siendo  $Y_i$  el valor neto de autoridad del usuario  $i$  y  $\Phi_{ij}$  el índice de ganancia o pérdida de autoridad entre  $j$  e  $i$ , siendo  $n$  la cantidad de nodos de la red social.

$$Y_t = \left( \sum_{j=1}^n R_{tj} \right) \div \left( \sum_{k=1}^n \left( \sum_{l=1}^n R_{kl} \right) \right)$$

**Fórmula 6.** Valor neto de autoridad del usuario  $i$ .

El valor neto de autoridad  $Y_i$  da el porcentaje de relaciones que posee el usuario  $i$  con respecto a toda la red.

$$\Phi_{xy} = \begin{cases} Y_y * (R_{xy} - R_{yx}) \times 0.5, & R_{xy} - R_{yx} > 0 \\ Y_x * (R_{xy} - R_{yx}) \times 0.5, & R_{xy} - R_{yx} < 0 \\ 0, & R_{xy} - R_{yx} = 0 \end{cases}$$

**Fórmula 7.** Índice de ganancia o pérdida de autoridad entre  $j$  e  $i$ .

El índice  $\Phi_{ij}$  representa tanto la cuantía de autoridad que hereda el nodo  $i$  de  $j$  así como la pérdida de la misma dependiendo de qué nodo posee la mayor autoridad y la

intensidad de la misma.

### **2.3.3.1 Pseudocódigo del algoritmo para el cálculo de $M_i$**

**Entrada del algoritmo:** Red Social.

**Salida:** Mapa de nodos con índices de autoridad.

1: Crear un mapa hash *hm* el cual almacenará la relación de nodos con los índices de autoridad, una variable *s* de tipo doble para guardar la suma total y un mapa *rankings* que guardará la relación de cada nodo con su índice de autoridad.

**Para** cada arista **Hacer**

2:     **Sumar** el peso de la arista actual a *s*

3:     **Si** el nodo origen de la arista actual existe en *hm* **Entonces**

4:         **Actualizar** *hm* en la posición del nodo de origen de la arista actual con la suma del peso de la arista actual y el valor anterior de la arista en *hm*

5:     **Si no**

6:         **Añadir** el nodo origen de la arista actual a *hm* y asignarle como valor el peso de la arista

7:     **FinSi**

8: **FinPara**

9: **Para** cada nodo **Hacer**

10:     **Si** *hm* contiene al nodo actual **Entonces**

11:          $Y_i = (\text{valor del nodo actual en } hm) / s$

12:     **Si** ( $Y_i > 0$ ) **Entonces**

13:         Crear un mapa hash *fi* que guardará de cada nodo relacionado con el nodo actual, la diferencia entre la fortaleza de sus relaciones. Definir una lista de arreglos *vertexRelation* que almacenará los id de cada nodo relacionado con el actual.

**Para** cada arista saliente del nodo actual **Hacer**

## CAPÍTULO 2. Propuesta de solución

- 14:                   **Añadir** el nodo destino de la arista actual a *fi* con el valor del peso de la arista
- 15:                   **Añadir** el nodo destino de la arista actual a *vertexRelatio*
- 16:                   **FinPara**
- 17:                   **Para** cada arista entrante del nodo actual **Hacer**
- 18:                   **Si** el nodo origen de la arista actual existe en *fi* **Entonces**
- 19:                   **Actualizar** *fi* en la posición del nodo de origen de la arista actual con la **diferencia** entre el valor anterior en *fi* y el peso de la arista actual
- 20:                   **FinSi**
- 21:                   **Si no Hacer**
- 22:                   **Añadir** el nodo origen de la arista actual a *fi* con el valor del peso de la arista **multiplicado** por -1
- 23:                   **Añadir** el nodo destino de la arista actual a *vertexRelation*
- 24:                   **FinSi no**
- 25:                   **FinPara**
- 26:                   Crear  $\phi_{ij}$  de tipo doble e igual a 0  
**Para** cada nodo en *vertexRelation* **Hacer**
- 27:                   **Si** ((el valor en *fi* del nodo actual > 0) Y (*hm* contiene al nodo actual)) **Entonces**
- 28:                    $\phi_{ij} += ((\text{valor del nodo actual en } hm) / s) * ((\text{valor del nodo actual en } fi)) * 0.5$
- 29:                   **FinSi**
- 30:                   **Si no Y** (el valor en *fi* del nodo actual < 0) **Entonces**

## CAPÍTULO 2. Propuesta de solución

- 31:  $\phi_{ij} += ((\text{valor del nodo en } hm) / s) * ((\text{valor del nodo actual en } fi)) * 0.5$
- 32: **FinSi no**
- 33: **FinPara**
- 34: **FinSi**
- 35:  $M_i = Y_i + \phi_{ij}$
- 35: **Añadir** a *rankings* el nodo actual con el valor  $M_i$
- 36: **FinSi**
- 37: **Si no Hacer**
- 38: **Añadir** a *rankings* el nodo actual con el valor 0
- 39: **FinSi no**
- 40: **FinPara**
- 41: **Retornar** *rankings*

### 2.4 La graficación en el complemento

Para la graficación en el complemento desarrollado se utilizó la librería JUNG (Universal Network Java / Graph). Esta librería es de código abierto e implementada en Java, diseñada para soportar el modelado, análisis y visualización de datos que se pueden representar en forma de grafos. Se centra en las aplicaciones de gráficos matemáticos y algorítmicos relacionados con los ámbitos de análisis de redes sociales, la visualización de información, el descubrimiento de conocimiento y la Minería de Datos. Sin embargo, no tiene que utilizarse específicamente para estos campos ya que se puede utilizar para muchas otras aplicaciones que utilicen gráficos y redes (Java Universal Network/Graph framework, 2010).

JUNG fue creado por tres estudiantes de Ciencias de la Información y Computación de la Universidad de California, Irvine: Joshua O'Madadhain, Danyel Fisher y Scott White. La arquitectura de JUNG está diseñada para soportar una gran variedad de representaciones de entidades y sus relaciones, tanto grafos dirigidos como no dirigidos,

## **CAPÍTULO 2. Propuesta de solución**

grafos multimodales, grafos con aristas paralelas e hipergrafos. JUNG provee un mecanismo para la anotación de grafos, entidades y relaciones con metadatos. Esto facilita la creación de herramientas de análisis de conjuntos de datos complejos que pueden examinar las relaciones entre las entidades, así como los metadatos asociados a cada entidad y relación (Java Universal Network/Graph framework, 2010).

La distribución utilizada de JUNG incluye implementaciones de una serie de algoritmos de teoría de grafos, Minería de Datos y análisis de redes sociales, tales como las rutinas de la agrupación, la descomposición, la optimización, generación de gráficos al azar, análisis estadístico, y el cálculo de las distancias de redes, flujos y medidas de importancia (centralidad, PageRank, hits, etcétera) (Java Universal Network/Graph framework, 2010).

JUNG también proporciona un marco de visualización que facilita la construcción de herramientas para la exploración interactiva de datos de la red. Los usuarios pueden utilizar uno de los algoritmos de diseño previstos, o utilizar el marco para crear sus propios diseños personalizados. Además, se proporcionan mecanismos de filtrado que permiten a los usuarios centrar su atención o sus algoritmos sobre partes específicas de la gráfica (Java Universal Network/Graph framework, 2010).

### **2.5 Opciones y paneles de visualización**

El complemento presentará las siguientes opciones y paneles de visualización:

**Panel de distribución:** permite modificar la distribución de los nodos en la red, empleando los siguientes algoritmos para representación gráfica: KKLAYOUT, Circle Layout, FRLAYOUT, SpringLayout, ISOMLAYOUT, Ranking View.

**Panel de ránking:** otorga una puntuación a los nodos según determinadas clasificaciones y ofrece las siguientes opciones Degree, In degree, Out Degree, Closeness, Bary Center, Betweenness, Authority.

**Authority:** otorga una puntuación a los nodos según el índice de autoridad definido en este estudio.

**Panel de modo del ratón:** permite modificar el modo en que se usa el ratón presentando dos opciones: Transforming, Picking.

**Otras opciones:**

**Size by ranking:** asigna el tamaño del nodo según el valor del Panel de ranking seleccionado.

**Show vertex name:** muestra en el grafo el nombre de los vértices.

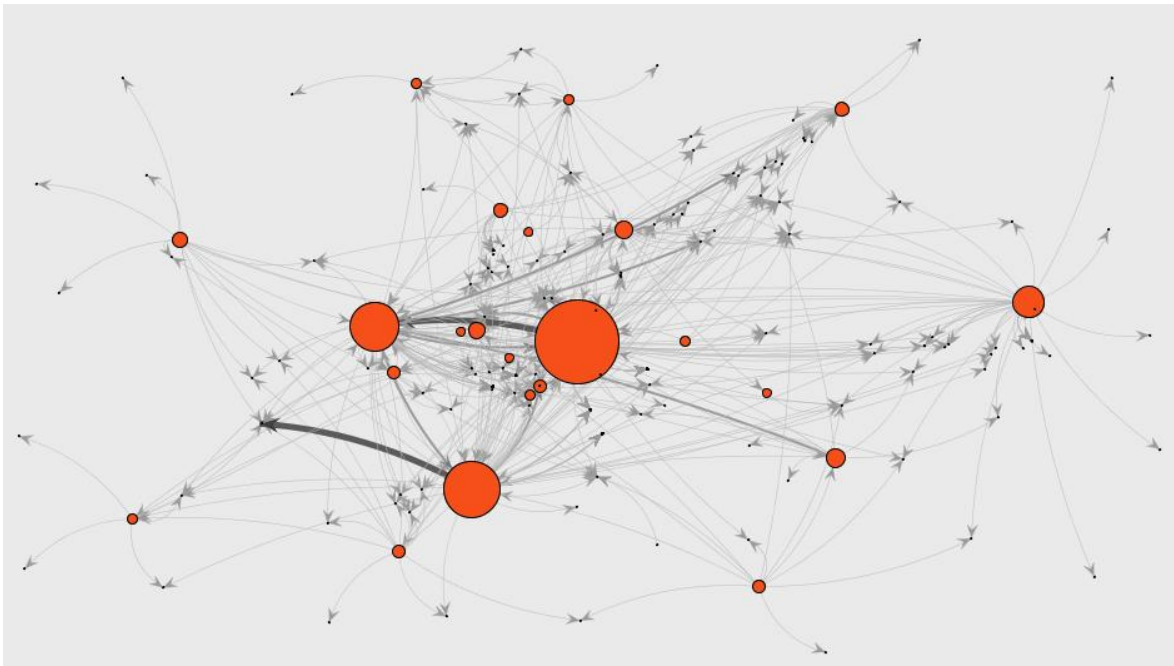
**Show edge weight value:** muestra el peso de las aristas.

**Show edge:** muestra las aristas.

**Stroke highlight on selection:** dado un nodo seleccionado esta opción resalta el borde de los nodos que están conectados a él.

### 2.5.1 Filtrado de nodos

El complemento permite filtrar los nodos de la red en dependencia de las variables de composición<sup>1</sup> seleccionadas. Una vez seleccionados los atributos<sup>2</sup> que se desean analizar, los nodos que no posean dichos atributos se reducirán en tamaño y se atenuará su color. El filtrado de nodos permite la personalización y segmentación de las búsquedas atendiendo al interés del usuario. Además, eleva los niveles de recuperación de los datos.



---

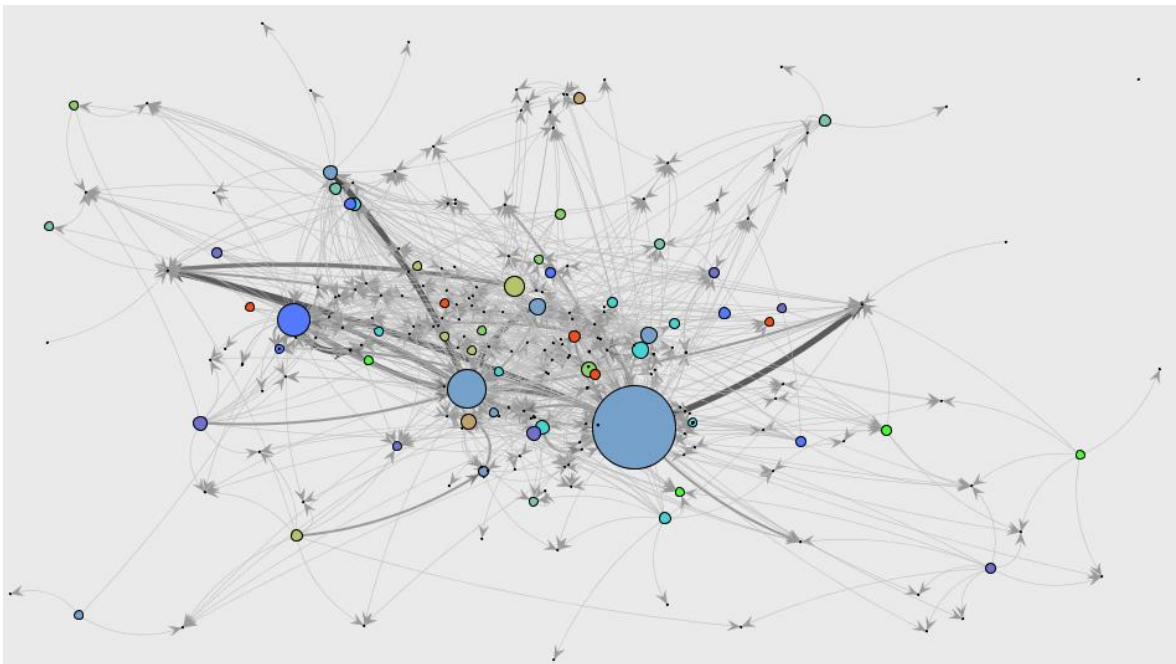
<sup>1</sup> **Las variables de composición:** son las categorías de las características de los nodos. Por ejemplo provincia.

<sup>2</sup> **Los atributos:** son los distintos valores que pueden tomar estas variables de composición. Por ejemplo Matanzas.

**Figura 14.** Filtrado de nodos por autoridad.

### **2.5.2 Leyenda**

El complemento posee como novedad un sistema de leyenda que asigna determinado color o forma a los nodos en dependencia de la característica seleccionada. Este recurso favorece la capacidad de realizar análisis a simple vista y elevar la comprensión y legibilidad de las representaciones.



**Figura 15.** Leyenda por color y filtrado de nodo.



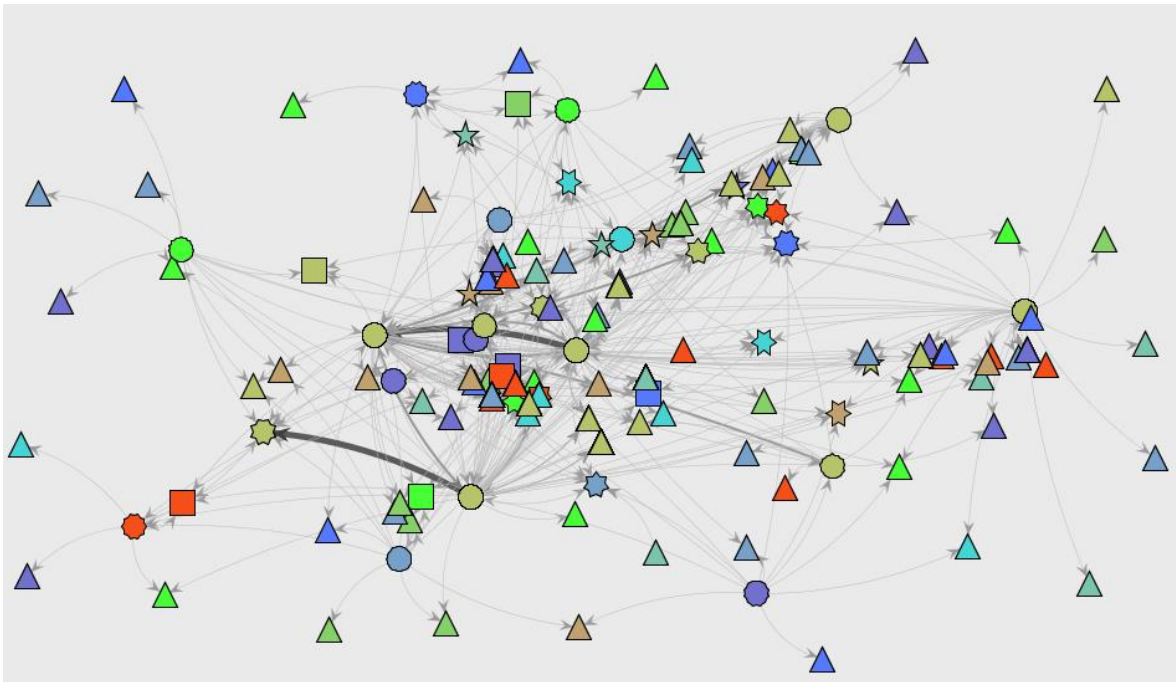


Figura 16. Leyenda por forma.

### 2.5.3 Resumen estadístico

Se definen un conjunto de datos que son mostrados en forma de resumen estadístico para apoyar las opciones de visualización. Estos datos relacionan las variables de composición de interés para el análisis y permiten identificar modos de comportamientos en la red. Las variables serán seleccionadas por los usuarios. La vista detallada de los datos permite asociar los atributos de las variables, los niveles de presencia y las frecuencias relativas según el índice de autoridad aportado por la métrica.

Var	Attribute	Fr	% Aparition	Average R with attribute
Sexo	Masculino	2,5	95	0,5
Provincia	Habana	2,2	47	0,55
Provincia	Matanzas	1,99	27	0,67
Provincia	Pinar del Río	1	30	0,6
Sexo	Femenino	0,7	5	0,7

Figura 17. Resumen estadístico.

**Var:** representa las variables de composición.

**Attribute:** representa los atributos de las variables de composición.

**Fr:** representa la sumatoria del peso de las relaciones de los nodos que cumplan determinado atributo.

**%Aparition:** representa el porcentaje de aparición de un determinado atributo en la red.

**Average R with attribute:** representa el promedio de las relaciones salientes de los nodos que compartan un atributo determinado.

### **2.6 Conclusiones del capítulo**

- La obtención del diseño de un algoritmo para la inferencia de redes sociales a partir de las actividades realizadas por los usuarios y sus interacciones en diferentes contextos, posibilitó incorporar el análisis del comportamiento de los usuarios en más de un contexto de interacción.
- La implementación del complemento **Mine for an Authority Social Network** a partir del algoritmo diseñado para la inferencia de redes sociales, facilitó el estudio del comportamiento de los nodos en la red a través de la autoridad que unos ejercen sobre otros. Dicho complemento está implementado en Java, bajo el marco de trabajo ProM.
- Las nuevas opciones de visualización implementadas contribuyeron a elevar la comprensión de los resultados representados en la red social.
- Se destacan como novedades del complemento un sistema de leyenda, opciones de filtrado de nodos y resúmenes estadísticos. Estas opciones permiten favorecer la capacidad de realizar análisis a simple vista, elevar la comprensión y legibilidad de las representaciones, la personalización y segmentación de las búsquedas atendiendo al interés del usuario.

## **CAPÍTULO 3. Validación de la solución**

### **3.1 Introducción**

En el presente capítulo se realizará la validación del algoritmo propuesto. Con el fin de mostrar la utilidad del algoritmo se utilizó la Comunidad UCI de PHP (<http://php.uci.cu>) como un entorno real. En esta comunidad se realizaron las pruebas y validaciones que certificaron la calidad de la solución propuesta.

### **3.2 Aplicación en un entorno real**

La Comunidad UCI de PHP es un espacio creado por estudiantes, profesores y trabajadores de la Universidad de las Ciencias Informáticas. Fue creada el 1ro de marzo del 2011 y es una comunidad temática que aborda temas de interés sobre lenguajes de programación, principalmente lenguajes web, enfocándose en PHP. Dicha comunidad brinda un espacio para que interesados en estas temáticas compartan noticias, artículos orientados al desarrollo web, descargas de recursos y herramientas para desarrolladores y un foro de discusión. Además, cuenta con varios espacios de intercambio directo entre sus miembros (peñas, ferias expositivas, conferencias, talleres).

### **3.3 Extracción de los registros de eventos**

A partir de la base de datos y el estudio de las secciones existentes dentro de la comunidad, se realizó un análisis de los diferentes contextos existentes, seleccionándose Artículos, Noticias y Descargas por acumular la mayor cantidad de interacciones dentro de la comunidad. Estos contextos se componen por una serie de actividades que se relacionan a continuación.

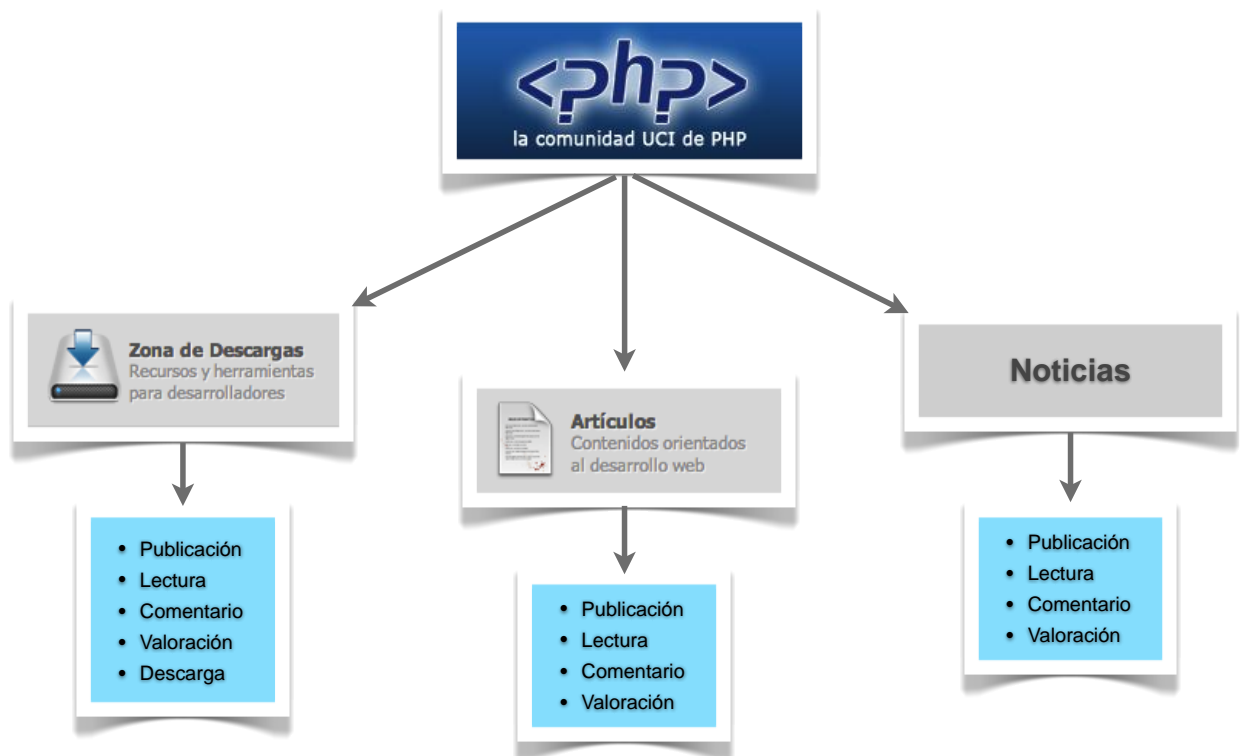


Figura 18. Contextos identificados en la Comunidad UCI de PHP.

Al definir cada uno de los contextos y las actividades que los componen, se analizó la base de datos de la comunidad, con el fin de determinar si contaba con el nivel de madurez necesario para realizar la extracción de los registros de eventos.

### 3.3.1 Calidad de los registros de eventos

Hay varios criterios para juzgar la calidad de los datos de los registros de eventos. Los eventos deben ser confiables (debería ser seguro asumir que los eventos registrados realmente ocurrieron y que los atributos de los eventos son correctos). Los registros de eventos deberían ser completos (dado un determinado contexto, no puede faltar ningún evento). Cualquier evento registrado debe tener una semántica bien definida. Además, los datos de eventos deben ser seguros en el sentido que se tengan en cuentas consideraciones de privacidad y seguridad al registrar los eventos. Por ejemplo, los actores deben ser conscientes del tipo de eventos que se registra y la forma en que se utilizan (van der Aalst, 2012).

Se definen cinco niveles de madurez de un registro de eventos que van desde excelente calidad (nivel 5) a mala calidad (nivel 1). Las técnicas de Minería de Procesos se recomiendan que sean aplicadas a registros de eventos en los niveles cinco, cuatro o tres, aunque en principio también es posible aplicar Minería de Procesos en registros

### CAPÍTULO 3. Validación de la solución

de eventos en niveles dos o uno. Sin embargo, el análisis de dichos registros de eventos es generalmente problemático y los resultados no son confiables (van der Aalst, 2012).

**Tabla 3.** Niveles de madurez de los registros de eventos (van der Aalst, 2012).

Nivel	Caracterización
<b>Nivel 5</b>	Nivel más alto: el registro de eventos es de excelente calidad (por ejemplo confiable y completo) y los eventos están bien definidos. Los eventos se registran de manera automática, sistemática, confiable, y segura. Se toman en cuenta adecuadamente consideraciones acerca de la privacidad y la seguridad. Además, los eventos registrados (y todos sus atributos) tienen una semántica clara. Esto implica la existencia de una o más ontologías <sup>3</sup> .
<b>Nivel 4</b>	Los eventos se registran automáticamente y de manera sistemática y confiable (por ejemplo, los registros de eventos son confiables y completos). A diferencia de los sistemas operando a nivel tres, se da soporte de manera explícita a nociones tales como instancia de proceso (caso) y actividad.
<b>Nivel 3</b>	Los eventos se registran automáticamente, pero no se sigue un enfoque sistemático para registrar los eventos. Sin embargo, a diferencia de los registros de eventos en el nivel dos, hay algún nivel de garantía que los eventos registrados calzan con la realidad (por ejemplo, el registro de eventos es confiable pero no necesariamente completo).
<b>Nivel 2</b>	Los eventos se registran automáticamente, por ejemplo, como un subproducto de algún sistema de información. La cobertura varía, por ejemplo, no se sigue un enfoque sistemático para decidir qué eventos se registran. Además, es posible pasar por alto el sistema de información. Por lo tanto, podrían faltar eventos o estos podrían no registrarse correctamente.

<sup>3</sup> En informática el término “ontologías” hace referencia a la formulación de un exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados.

<b>Nivel 1</b>	Nivel más bajo: los registros de eventos son de mala calidad. Los eventos registrados podrían no corresponder a la realidad y podrían faltar eventos. Los registros de eventos en los cuales los eventos se registran manualmente suelen tener dichas características.
----------------	--

### 3.3.2 Estado de la base de datos de la comunidad UCI de PHP y extracción de registro de eventos

Durante el análisis de la base de datos del sitio PHP se descubrieron una serie de dificultades que obstaculizaron el proceso de extracción de los registros de eventos, aun cuando los mismos poseían un nivel 4 de madurez. Dentro de las dificultades encontradas resaltan que los datos no estaban normalizados, no todas las actividades eran gestionadas por el sistema de trazas y no existía un estándar para el almacenamiento de ellas.

Las actividades se encontraban dispersas en diferentes tablas de la base de datos (ver tabla 4 y 5), almacenadas en algunos casos varias actividades en una misma tabla pero con diferentes formatos (ver tabla 6).

**Tabla 4.** Relación id-contexto.

<b>Id</b>	<b>Contexto</b>
1	Artículo
2	Descarga
3	Noticias

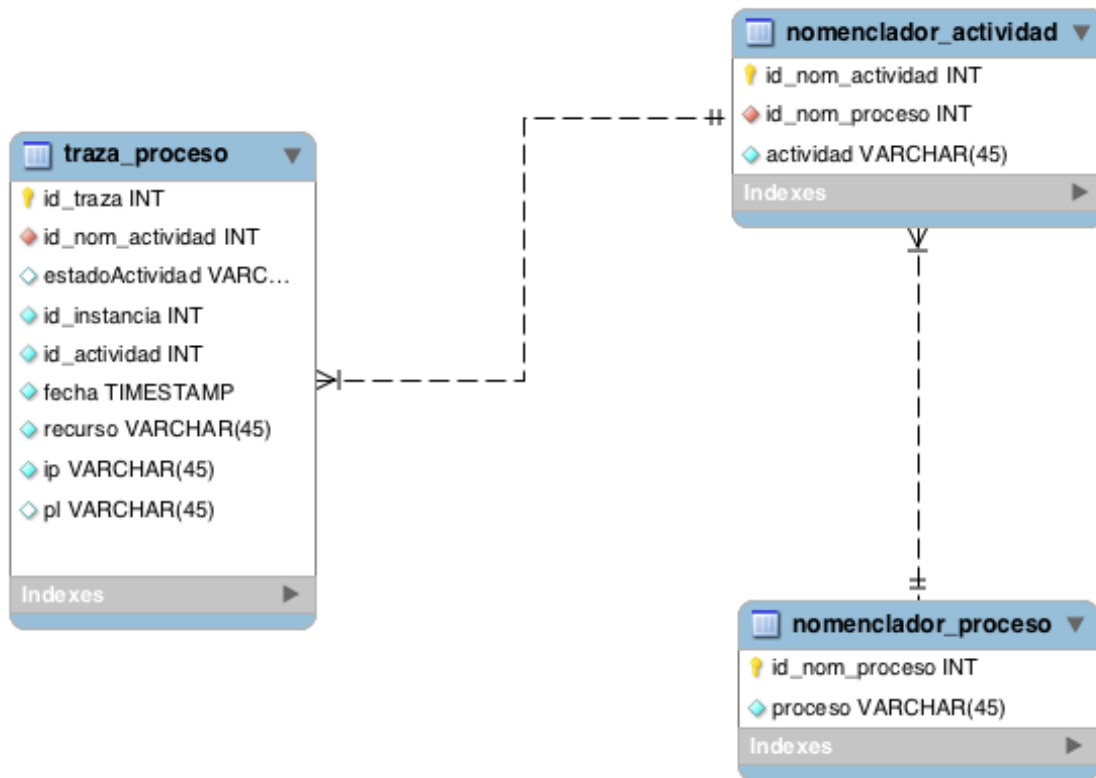
**Tabla 5.** Ubicación de registros de actividades en la base de datos.

<b>Id de contexto</b>	<b>Actividad</b>	<b>Donde se encuentra</b>
1, 2, 3	Lectura	phpv2_gtp_guests
2	Descarga	phpv2_gtp_guests
1, 2, 3	Valorar	phpv2_ratings
1, 2, 3	Comentar	phpv2_comments
1	Publicar	phpv2_articles
2	Publicar	phpv2_downloads
3	Publicar	phpv2_news

**Tabla 6.** Forma de almacenamiento de las trazas en la tabla phpv2\_gtp\_guests.

Id	Actividad	Log
1	Lectura	articles.php?article_id=87
2	Descarga	downloads.php?cat_id=1&file_id=368
3	Lectura	news.php?readmore=729
2	Lectura	/downloads.php?cat_id=5& download_id=134

Por todas las dificultades antes mencionados se confeccionó una base de datos de transición (ver figura 19), la cual se estructuró para soportar explícitamente nociones de instancia de procesos y actividad. Para realizar el paso de las trazas de la base de datos de la comunidad PHP a la nueva base de datos, se realizó una aplicación en java. Dicha aplicación tomó los datos originales de la comunidad PHP, los normalizó, agrupó e insertó en la nueva base de datos.



**Figura 19.** Diagrama entidad relación de la base de datos de transición.

### 3.4 Utilización de un sociograma para la validación del componente desarrollado

Para la validación de la métrica implementada se aplicó una técnica tradicional de la sociología: el sociograma. Se conoce como sociograma a una técnica que, a través de la observación y la evaluación de un contexto, refleja en un gráfico los diferentes

### **CAPÍTULO 3. Validación de la solución**

vínculos entre los miembros de un grupo. De esta manera, logra dejar en evidencia los grados de influencia y los lazos de preferencia que se presentan en él. Los sociogramas son recursos útiles para graficar la estructura de una organización, ya sea grande o pequeña (Ander-Egg, 1982).

El sociograma está considerado como una técnica de carácter sociométrico, es decir, un método que permite medir las relaciones sociales entre los integrantes de un grupo humano, donde sus elementos se conocen, poseen objetivos en común y se influyen mutuamente. A nivel gráfico, un sociograma representa las relaciones interpersonales por medio de nodos (los individuos) que aparecen unidos por una o más aristas (las relaciones interindividuales) (Ander-Egg, 1982).

El gráfico obtenido permite apreciar afinidades, detectar subgrupos dentro del conjunto principal y ubicar a los líderes sociométricos (las personas más influyentes). De esta forma, es posible actuar sobre el grupo para activar vínculos potenciales o desactivar aquellos existentes (Ander-Egg, 1982).

Todos estos conocimientos se consiguen a través de la elaboración de una serie de preguntas que conformarán un cuestionario muy completo y preparado. En concreto, esas cuestiones serán de dos tipos. Por un lado, estarán las llamadas intelectuales que son las que se refieren a la actividad productiva de la clase y por otro, estarán las afectivas que se refieren a las elecciones que realizarían las personas para desarrollar actividades de tipo lúdico (Ander-Egg, 1982).

Así, por ejemplo, entre las preguntas intelectuales se encontrarían algunas preguntas como “¿A quién elegirías para realizar el trabajo de Plástica?”. Mientras, una muestra de lo que son las cuestiones afectivas sería la siguiente: “¿A qué compañero elegirías para ir de excursión?”.

El sociograma es una herramienta muy usual en el campo de la Sociología, la ciencia que hace énfasis en la examinación y análisis de los grupos sociales. Con este tipo de métodos, esta ciencia investiga las relaciones que los sujetos mantienen entre sí y con el sistema, y el grado de cohesión que existe en la estructura social. Cabe destacar, que los métodos de la sociología pueden ser cualitativos (con descripciones detalladas de situaciones, comportamientos y personas) o cuantitativos (que se encarga de las características y variables que pueden ser representadas por valores numéricos). El



## **CAPÍTULO 3. Validación de la solución**

objetivo de la aplicación de esta técnica en el presente trabajo es verificar la utilidad y el nivel de exactitud que presenta la métrica implementada en el algoritmo.

### **3.5 Aplicación del sociograma**

Para la aplicación del sociograma se realizó un sistema de encuesta el cual se les aplicó a los diez usuarios detectados con los mayores niveles de autoridad en la comunidad PHP. La encuesta contó con 3 preguntas, cada pregunta tenía como objetivo que cada uno de los diez usuarios escogiera a 3 de sus compañeros para la realización de determinadas actividades (ver anexo 1). El objetivo final sería detectar por quien se sentían más influenciados en cada uno de los escenarios encuestados. El sistema de preguntas fue desarrollado por un sociólogo de la Universidad de las Ciencias Informáticas.

### **3.6 Correlación entre los resultados del sociograma y los resultados del complemento desarrollado**

Se entiende como correlación al grado de relación existente entre dos variables. Para verificar la correspondencia de los resultados obtenidos con el sociograma y con el complemento desarrollado se aplicaron dos pruebas: el coeficiente de correlación de Spearman y el coeficiente de correlación de Pearson. Se aplicó también un diagrama de dispersión para poder observar de forma gráfica la dispersión de los datos.

#### **3.6.1 Coeficiente de correlación de Spearman**

El coeficiente de correlación de Spearman permite medir la relación entre atributos ordenados. Se aplica para obtener correlaciones entre ordenaciones. Sea A y B dos atributos en escala ordinal y  $a_1, \dots, a_n$  y  $b_1, \dots, b_n$  sus diferentes categorías. Se suponen ordenadas estas categorías de menor a mayor. Se denomina rango a la posición de cada categoría en la muestra ordenada. Los rangos toman entonces los valores  $\{1, 2, \dots, n\}$ . Se denota por  $x_i$  e  $y_i$ ,  $i=1, 2, \dots, n$  los rangos correspondientes a los dos atributos (Sarabia Alegría, et al., 2005).

Se denomina coeficiente de correlación ordinal de Spearman o coeficiente de correlación por rango de Spearman al coeficiente de correlación lineal entre los dos rangos (Sarabia Alegría, et al., 2005).

### CAPÍTULO 3. Validación de la solución

Si se suponen dos atributos A y B en escala ordinal, cuyos valores han sido transformados en los rangos  $(x_i, y_i)$ ,  $i = \{1, 2, \dots, n\}$ , el coeficiente de correlación de rangos de Spearman viene dado por:

$$r = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)}$$

**Fórmula 8.** Coeficiente de correlación de rangos de Spearman (Sarabia Alegría, 2005).

Donde  $d_i = x_i - y_i$ ,  $i=1,2,\dots, n$

Al tratarse de un coeficiente de correlación su valor varía entre -1 y +1. Si la concordancia entre los rangos es perfecta entonces  $d_i = x_i - y_i = 0$ ,  $i=1,2,\dots, n$  y por lo tanto  $r = 1$ . Cuando la discordancia entre los rangos es perfecta entonces  $r = -1$ . Si los dos criterios de clasificación son independientes entonces  $r = 0$  (Sarabia Alegría, et al., 2005).

#### 3.6.2 Coeficiente de correlación lineal o Coeficiente de Pearson

El coeficiente de correlación lineal o Coeficiente de Pearson es una medida de dependencia lineal adimensional entre dos variables cuantitativas X e Y (Sarabia Alegría, et al., 2005). La cual se define como:

$$r = \frac{S_{XY}}{S_X * S_Y}$$

**Fórmula 9.** Coeficiente de correlación de Pearson (Sarabia Alegría, 2005).

Donde  $S_{XY}$  representa la covarianza y  $S_X$  y  $S_Y$  las desviaciones típicas de X e Y, respectivamente.

Propiedades del Coeficiente de correlación de Pearson:

1. El coeficiente de correlación lineal es adimensional, es decir, carece de unidades de medida.
2. Es invariante frente a transformaciones lineales en los datos. En concreto, si a partir de una variable bidimensional (X, Y) se construye una nueva variable  $(a_1 + b_1X, a_2 + b_2 Y)$  se verifica que  $(b_1, b_2 \neq 0)$ :  
 $r(a_1 + b_1X, a_2 + b_2 Y) = \text{signo}(b_1 * b_2) * r(X, Y)$

3. Es acotado:  $-1 \leq r \leq 1$ .
4. Cuando  $|r|$  está cercano a 1, existe una relación aproximadamente lineal entre las variables X e Y. Si el valor de r está cercano a 0, no existe relación de tipo lineal entre las variables (Sarabia Alegría, et al., 2005).

### **3.6.3 Diagrama de dispersión**

Un diagrama de dispersión es una representación gráfica de la relación entre dos variables, muy utilizada en las fases de comprobación de teorías e identificación de causas raíz y en el diseño de soluciones y mantenimiento de los resultados obtenidos. Estudios revelan que resulta más fácil identificar las relaciones en un diagrama de dispersión que en una tabla de números (Excelencia, 2006).

El análisis de un diagrama de dispersión consta de un proceso de cuatro pasos: se elabora una teoría razonable, se obtienen los pares de valores y se dibuja el diagrama, se identifica la pauta de correlación y se estudian las posibles explicaciones. Las pautas de correlación más comunes son: correlación fuerte positiva (Y aumenta claramente con X), correlación fuerte negativa (Y disminuye claramente con X), correlación débil positiva (Y aumenta algo con X), correlación débil negativa (Y disminuye algo con X), correlación compleja (Y parece relacionarse con X pero no de un modo lineal) y correlación nula (no hay relación entre X e Y) (Excelencia, 2006).

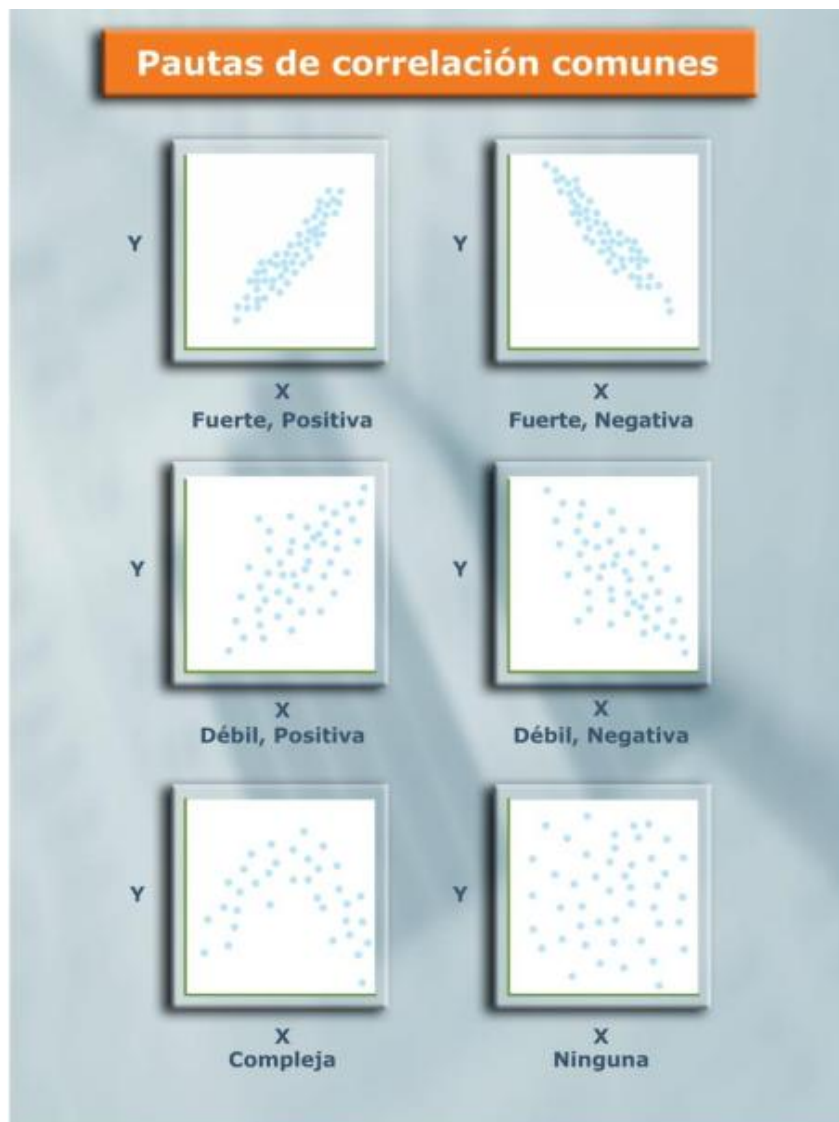


Figura 20. Diagrama de dispersión (Excelencia, 2006).

### 3.7 Aplicación del coeficiente de correlación de Spearman y coeficiente de Pearson

Para la aplicación del coeficiente se definieron las siguientes premisas: si se obtiene un valor de correspondencia mayor o igual que 0,6 y menor que 0,8, entonces el valor de correspondencia es adecuado; si se obtiene un valor de correspondencia mayor o igual que 0,8, entonces el valor de correspondencia es muy confiable.

**$0,6 \leq x < 0,8$  Adecuado**

**$x \geq 0,8$  Muy confiable**

Para validar el supuesto de que al analizar varios contextos de interacción se eleva la

### CAPÍTULO 3. Validación de la solución

precisión de los análisis del comportamiento de los individuos, se aplicó el coeficiente de correlación de Spearman y de Pearson a cada uno de los contextos por separado y luego a la unión de todos ellos.

Las tablas 7, 8 y 9 fueron desarrolladas para aplicar el coeficiente de Spearman y Pearson en tres contextos diferentes. La primera columna representa los usuarios seleccionados como muestra para realizar la prueba. La segunda columna, los valores asociados a cada usuario obtenidos al aplicar la técnica del sociograma. La tercera, los valores obtenidos del complemento desarrollado. X e Y representan los datos ordenados por rango, X el orden de los valores del complemento e Y el orden de los valores del sociograma.

Contexto Artículos.

**Tabla 7.** Coeficiente de correlación del registro de eventos Artículos.

ID	Sociograma	Complemento	X	Y	D= x-y	D <sup>2</sup>
A	49	0,272	1	1	0	0
C	31	0,153	3	2	1	1
B	30	0,03	5	3	2	4
G	24	0,198	2	4	-2	4
L	8	0,042	4	5	-1	1
O	5	0,011	7	6	1	1
D	1	0,019	6	9,5	-3,5	12,25
J	3	0,006	8	7,5	0,5	0,25
I	1	0,005	9	9,5	-0,5	0,25
K	3	0	10	7,5	2,5	6,25
<b>Total</b>						<b>30</b>

Para los datos obtenidos en el contexto Artículos se obtuvo un valor de correspondencia de Spearman de 0,818 y un valor de correspondencia Pearson de 0,817.

### CAPÍTULO 3. Validación de la solución

Contexto Descargas.

**Tabla 8.** Coeficiente de correlación del registro de eventos Descargas.

ID	Sociograma	Complemento	X	Y	D= x-y	D <sup>2</sup>
A	49	0,281	1	1	0	0
C	31	0,115	2	2	0	0
B	30	0,022	5	3	2	4
G	24	0,092	3	4	-1	1
L	8	0	9,5	5	4,5	20,25
O	5	0,023	4	6	-2	4
D	1	0,0015	8	9,5	-1,5	2,25
J	3	0,019	6	7,5	-1,5	2,25
I	1	0	9,5	9,5	0	0
K	3	0,008	7	7,5	-0,5	0,25
<b>Total</b>						<b>34</b>

Para los datos obtenidos en el contexto Descargas se obtuvo un valor de correspondencia de Spearman de 0,793 y un valor de correspondencia Pearson de 0,792.

Contexto Noticias.

**Tabla 9.** Coeficiente de correlación del registro de eventos Noticias.

ID	Sociograma	Complemento	X	Y	D= x-y	D <sup>2</sup>
A	49	0,5149	1	1	0	0
C	31	0,191	2	2	0	0
B	30	0,04	3	3	0	0
G	24	0,025	5	4	1	1
L	8	0,036	4	5	-1	1
O	5	0,0015	7	6	1	1
D	1	0,001	8,5	9,5	-1	1
J	3	0	10	7,5	2,5	6,25
I	1	0,001	8,5	9,5	-1	1
K	3	0,017	6	7,5	-1,5	2,25
<b>Total</b>						<b>13,5</b>

Para los datos obtenidos en el contexto Noticias se obtuvo un valor de correspondencia

### CAPÍTULO 3. Validación de la solución

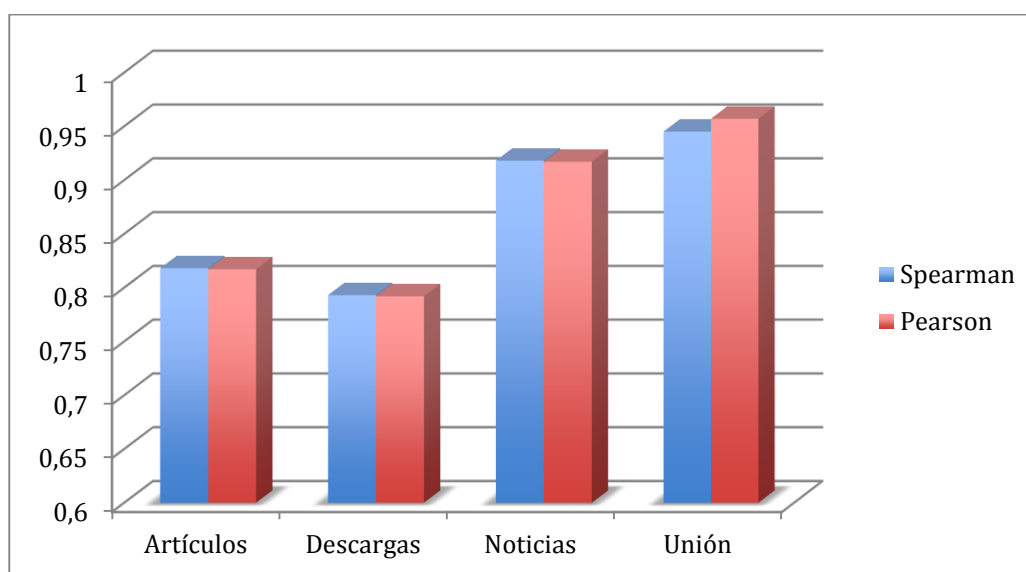
de Spearman de 0,918 y un valor de correspondencia Pearson de 0,917.

**Tabla 10.** Coeficiente de correlación de la unión de todos los registros de eventos

ID	Sociograma	Complemento	X	Y	D= x-y	D <sup>2</sup>
A	49	0,33	1	1	0	0
C	31	0,15	2	2	0	0
B	30	0,028	4	3	1	1
G	24	0,12	3	4	-1	1
L	8	0,025	5	5	0	0
O	5	0,013	6	6	0	0
D	1	0,008	8	9,5	-1,5	2,25
J	3	0,009	7	7,5	-0,5	0,25
I	3	0,002	10	9,5	0,5	2,25
K	1	0,0072	9	7,5	1,5	2,25
<b>TOTAL</b>						<b>9</b>

Al aplicar el método de Spearman se obtuvo un valor de correspondencia de 0,945 y al aplicar el método de Pearson se obtuvo un valor de correspondencia de 0,957.

Se puede concluir que según el valor de correspondencia de Spearman y el de Pearson la correlación del sociograma con respecto a la del complemento desarrollado en cada contexto por separado, es menor que la correlación de todos los registros de eventos unidos.

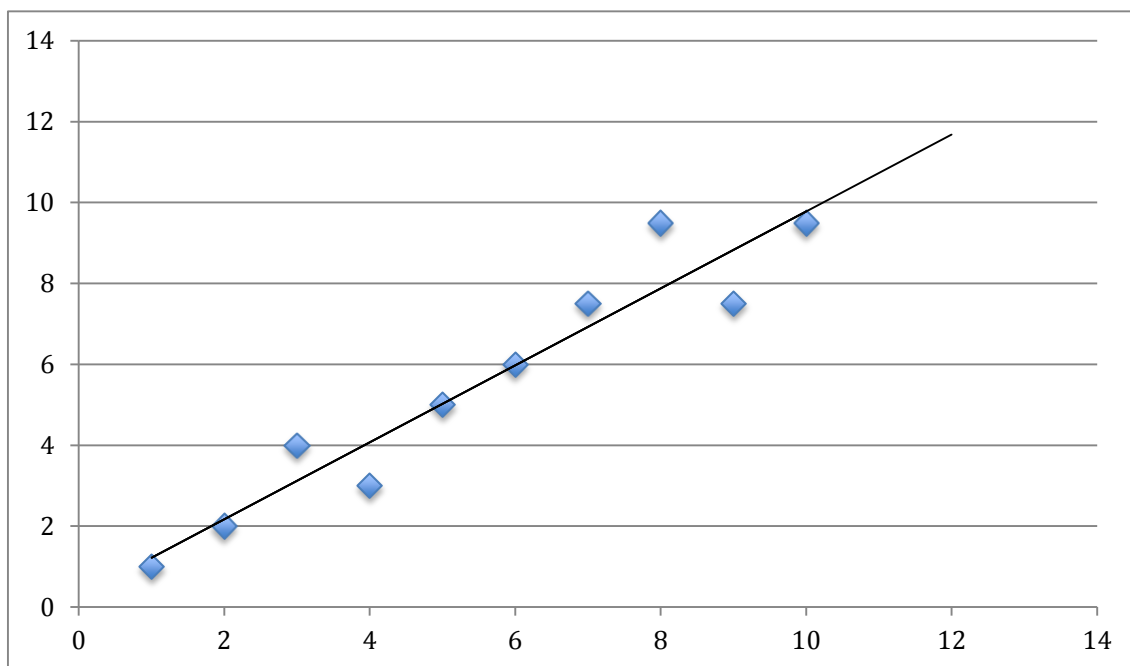


**Figura 21.** Comparación de los resultados obtenidos de correlación.

### CAPÍTULO 3. Validación de la solución

Avalándose que: si se desarrolla un algoritmo de Minería de Procesos para la inferencia de redes de interacción social a partir de las actividades realizadas por los usuarios, sus datos personales e institucionales, entonces se eleva la precisión de los análisis del comportamiento de los individuos en sus contextos de interacción.

#### Diagrama de dispersión



**Figura 22.** Diagrama de dispersión sociograma / complemento.

Al observar la gráfica se puede concluir que la dispersión obtenida presenta una correlación fuerte positiva (los valores de Y aumentan con los de X).

#### 3.8 Conclusiones del capítulo

- Al extraer los registros de eventos de la comunidad UCI de PHP se pudo determinar que estos poseen un nivel 4 de madurez (en una escala de 5), garantizando la confiabilidad y completitud de los datos.
- Se aplicó una técnica tradicional de carácter sociométrico para evaluar el algoritmo implementado, evidenciando los grados de influencia y lazos de preferencia que se representan en la red.
- El sociograma aplicado a los usuarios detectados con mayores niveles de autoridad en la comunidad permitió recolectar las elecciones y distinguir los tipos de relaciones existentes.



### ***CAPÍTULO 3. Validación de la solución***

- La aplicación del coeficiente de correlación de Spearman y el coeficiente de correlación de Pearson verificó la correspondencia entre los datos obtenidos con el sociograma y aquellos mediante el complemento desarrollado.
- Se demostró que los niveles de correlación aumentan en la medida que se incorporan mayor cantidad de contextos de interacción, elevando la precisión de los análisis y con ello la fiabilidad de los resultados obtenidos.

### Conclusiones generales

En el presente trabajo se realizó un análisis de las principales métricas existentes en la Minería de Procesos que son empleadas para el descubrimiento de redes sociales. Se creó un algoritmo para la identificación de redes sociales basándose en una métrica de autoridad, el cual permitió incluir nuevas funcionalidades que elevó la precisión de los análisis del comportamiento de los individuos en sus contextos de interacción. En función de los objetivos de la investigación los resultados permitieron arribar a las siguientes conclusiones:

- El análisis de las principales fuentes bibliográficas permitió identificar los aspectos que caracterizan la influencia en la red, entre los que se señalan la autoridad, reciprocidad, coherencia, simpatía, aprobación social y escasez.
- Se implementó el complemento **Mine for an Authority Social Network** a partir del algoritmo diseñado para la inferencia de redes sociales, teniendo en cuenta la métrica definida. Este complemento estudia el comportamiento de los nodos en la red a través de la autoridad que unos ejercen sobre otros.
- Se destacan como novedades del complemento un sistema de leyenda, opciones de filtrado de nodos y resúmenes estadísticos. Estas opciones favorecen la capacidad de realizar análisis a simple vista, elevan la comprensión y legibilidad de las representaciones, la personalización y segmentación de las búsquedas atendiendo al interés del usuario.
- Se aplicó una técnica tradicional de carácter sociométrico para evaluar el algoritmo implementado, evidenciando los grados de influencia y lazos de preferencia que se representan en la red. La aplicación del coeficiente de correlación de Spearman y el coeficiente de correlación de Pearson, permitió verificar la correspondencia entre los datos obtenidos con el sociograma y aquellos mediante el complemento desarrollado.
- Se demostró que los niveles de correlación aumentan en la medida que se incorporan mayor cantidad de contextos de interacción, permitiendo elevar la precisión de los análisis y con ello la fiabilidad de los resultados obtenidos.

**Bibliografía**

**Ailenei, Irina María. 2011.** *Process Mining Tools: A Comparative.* s.l. : EINDHOVEN University of Technology, 2011.

**Ander-Egg, Ezequiel. 1982.** *Técnicas de investigación social.* Buenos Aires : Universidad Nacional Abierta. Dirección de investigaciones y Postgrado, 1982.

**Battista, G, et al. 1999.** *Graph Drawing: Algorithms for Visualization of Graphs.* 1999.

**Buijs, J.C.A.M, et al. 2011.** *XES, XESame and ProM 6.* [ed.] Department of Mathematics and Computer Science. Eindhoven : Technische Universiteit Eindhoven, 2011.

**Butts, Carter T. 2014.** *Tools for Social Network Analysis.* 2014.

**Carrasco Jiménez, José C, et al. 2012.** *Social Interaction Discovery: a Simulated Multiagent Approach.* Monterrey : s.n., 2012.

**Chambers , John , Bates, Douglas and Bivand , Roger .** What is R? *r-project.* [Online] r-project. [Cited: Marzo 05, 2014.] <http://www.r-project.org/>.

**Chiavenato, Idalberto. 2004.** *Introducción a la Teoría General de la Administración.* s.l. : McGraw-Hill, 2004.

**Cialdini, Robert B. 2009.** *Influence. The Psychology of Persuasion.* s.l. : HarperCollins, 2009.

**Cook, Diane J and Holder, Lawrence B. 2007.** *Mining Graph Data.* EU : Wiley Interscience a John Wiley & Sons INC, 2007.

**Cruz, Isabel F and Tamassia, Roberto. 2009.** *Graph Drawing Tutorial.* 2009.

**Eades, P. 1984.** *A Heuristic for Graph Drawing.* 1984. Vol. 42.

**Excelencia, Centro de. 2006.** Diagrama de dispersión. *Centro de Excelencia.* [Online] 2006. [Cited: 05 05, 2014.] <http://www.centrosdeexcelencia.com/entidades/herram/dispersion>.

**Fruchterman, Thomas M and Reingold, Edward M. 1991.** *Graph drawing by force-directed placement.* 1991.

**Gibert Ginesta, Marc, Casillas Santillán, Luis A and Pérez Mora, Oscar. 2004.** *Bases de datos en MySQL*. s.l. : Universidad Abierta de Cataluña, 2004.

**Gosling, James, et al. 2005.** *The Java Language Specification*. third. California : s.n., 2005.

*Informe Estado de la Ciencia 2010*. **2010**. s.l. : UNESCO, 2010.

**Java Universal Network/Graph framework. 2010.** *Java Universal Network/Graph Framework*. [Online] 2010. [Cited: 05 05, 2014.] <http://jung.sourceforge.net/faq.html>.

**Kamada, Tomihisa and Kawai, Satoru. 1989.** *An algorithm for drawing general undirected graphs*. s.l. : Faculty of Science, University of Tokyo, 1989.

**Leonidas Aguirre, Julio. 2011.** *Introducción al Análisis de Redes Sociales*. Buenos Aires : Centro Interdisciplinario para el estudio de políticas públicas, 2011.

**Manifiesto sobre Minería de Proceso. 2011.** Manifiesto sobre Minería de Proceso. [Online] 2011. <http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:pmm-spanish-v1.pdf>.

**Martig, Sergio and Señas, Perla. 2000.** *Grafo Integrador de un Mapa Conceptual Hipermedial*. [ed.] Grupo InE Departamento de Ciencias de la Computación. Bahía Blanca : Universidad Nacional del Sur, 2000.

**Mayer, Bernd. 1998.** Self-organizing graphs. A neutral network perspective of graphs layout. [ed.] Sue H. Withesides. *Graph Drawing*. 1998.

**McFarland, Dan and Bender-deMoll, Skye. 2014.** SoNIA. Social Network Image Animator. *SoNIA-Social Network Image Animator*. [Online] University of Stanford, 2014. <http://web.stanford.edu/group/sonia/>.

**Mejias Olivares, Cristian Paolo. 2010.** *Análisis de Redes Sociales a Gran Escala*. Mexico, D.F : Centro de investigación y de estudios avanzados del instituto politécnico nacional. Departamento de computación., 2010.

**Mena, Néstor. 2010.** *El análisis de redes sociales, una herramienta para la gestión del conocimiento en las organizaciones*. La habana : CITMA, 2010.

**Monsalve, Mauricio. 2008.** Análisis de redes sociales: un tutorial. 2008.

## **BIBLIOGRAFÍA**

**Sarabia Alegría, José María and Pascual Sáez, Marta. 2005.** *Curso básico de estadística para economía y administración de empresas.* s.l.: Universidad de Cantabria, 2005.

**van der Aalst, Wil M.P, Reijers, Hajo A and Song, Minseok. 2005.** *Discovering Social Networks from Event Logs.* South Korea : s.n., 2005.

**van der Aalst, Will M. P. 2012.** *Process Mining Manifesto.* 2012.

**van der Aalst, Will M. P and Song, Minseok. 2004.** *Mining Social Networks: Uncovering Interaction.* 2004.




**van der Aalst, Will M.P . 2011.** *Process Mining: Discovery, Conformance and Enhancement of Business Processes.* 2011.

**WebMining Consultores. 2012.** Análisis de Redes Sociales con R. *Web Mining. Powering web intelligence.* [Online] WebMining Consultores Limitada, Abril 07, 2012. [Cited: Marzo 05, 2014.] <http://www.webmining.cl>.

**XES-Standard. 2012.** XES. *XES-Standard.* [Online] 2012. [Cited: Mayo 05, 2014.] <http://www.xes-standard.org>.

## Anexos

Anexo 1. Encuesta realizada para la aplicación del sociograma.

<p>El grupo de investigación de Minería De Proceso está realizando un estudio que permitirá incrementar los servicios personalizados de la comunidad <b>PHP</b>, teniendo en cuenta las relaciones que establecen sus miembros. Pedimos su colaboración y la sinceridad en las respuestas. Garantizamos su anonimato y la confidencialidad de los datos. Muchas gracias por su contribución.</p>
<p>0 </p>
<p>¿Con quién te gustaría publicar un post en la comunidad?</p> <p>_____</p> <p>1                      2                      3</p>
<p>1 </p>
<p>¿A quién le pedirías sugerencias antes de compartir código con la comunidad?</p> <p>_____</p> <p>1                      2                      3</p>
<p>2 </p>

Me parecen muy útiles los post que publica...

\_\_\_\_\_

1                      2                      3

Fin