



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

## *Predicción del alcance de propagación de información e influencia en redes sociales online*

**Raúl Ortiz-Gaona**

**ADVERTIMENT** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del repositori institucional UPCCommons (<http://upcommons.upc.edu/tesis>) i el repositori cooperatiu TDX (<http://www.tdx.cat/>) ha estat autoritzada pels titulars dels drets de propietat intel·lectual **únicament per a usos privats** emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei UPCCommons o TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a UPCCommons (*framing*). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del repositorio institucional UPCCommons (<http://upcommons.upc.edu/tesis>) y el repositorio cooperativo TDR (<http://www.tdx.cat/?locale-attribute=es>) ha sido autorizada por los titulares de los derechos de propiedad intelectual **únicamente para usos privados enmarcados** en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio UPCCommons. No se autoriza la presentación de su contenido en una ventana o marco ajeno a UPCCommons (*framing*). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the institutional repository UPCCommons (<http://upcommons.upc.edu/tesis>) and the cooperative repository TDX (<http://www.tdx.cat/?locale-attribute=en>) has been authorized by the titular of the intellectual property rights **only for private uses** placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading nor availability from a site foreign to the UPCCommons service. Introducing its content in a window or frame foreign to the UPCCommons service is not authorized (*framing*). These rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

Programa de Doctorado en Ingeniería Telemática

# **Predicción del alcance de propagación de información e influencia en redes sociales online**

Tesis doctoral

Autor: **Raúl Ortiz-Gaona**

Directores: Dr. José Luis Melús Moreno

Dr. Marcos Postigo-Boix

Departamento de Ingeniería Telemática, UPC.

Barcelona, España, 2020

## **Agradecimientos**

Agradezco al Departamento de Ingeniería Telemática de la Universidad Politécnica de Cataluña y a la Universidad de Cuenca-Ecuador por permitirme realizar mi sueño de estudiar el doctorado en Ingeniería Telemática. Agradezco a mis directores de tesis Dr. José Luis Melús Moreno y Dr. Marcos Postigo Boix, al Dr. Joan Serrat Fernández por su valiosa guía y consejos, al personal administrativo Mar Serre y Marta Herranz por siempre estar dispuestas a ayudarme, al personal técnico-informático Juan Paz y Armando Martín, y a Fina, señora encargada de la limpieza del Departamento de Ingeniería Telemática.

## Resumen

Las redes sociales online (*OSNs*) son cada vez más utilizadas por diferentes sectores de la sociedad, y tienen el propósito de informar e influir a la gente. Estos sectores son, entre otros: gobiernos, partidos políticos, organizaciones no gubernamentales, proveedores de bienes y servicios, etc. A ellos les sería muy útil poder pronosticar el alcance de la propagación de la información y de la influencia que transmiten a través de las redes sociales. Actualmente existen un sinnúmero de modelos matemáticos utilizados en trabajos científicos, que pronostican el alcance de esas propagaciones, muchos de los cuales se basan, a su vez, en uno de siguientes modelos base o referentes: *Linear Threshold Model (LTM)* y *Independent Cascade Model (ICM)*.

En este documento presentamos *Lucy Model (LM)*, un nuevo modelo matemático predictor, que incorpora varios elementos que no han sido tomados en cuenta en los modelos base ni en sus modelos derivados, pero que están presentes en la vida real y afectan el alcance de la propagación de mensajes en redes sociales online. Estos nuevos elementos son los siguientes: Nosotros distinguimos tres clases de mensajes: mensajes meramente informativos; mensajes que tienen la intención de influir apelando a las emociones y sentimientos de la gente; y mensajes que tienen la intención de influir apelando a las conveniencias e intereses personales. Nosotros planteamos que la intensidad de la relación entre individuos (*tie-strength*) depende de la relevancia y de la afinidad entre ellos. La probabilidad de que el mensaje sea recibido por un individuo es directamente proporcional a esta intensidad de relación. También, nosotros consideramos que un mensaje con mayor importancia tiene más probabilidad de propagarse entre la gente que un mensaje con menos importancia. Finalmente, nosotros sostenemos que los individuos son informados o influenciados (activados) luego de superar un umbral, el cual depende de la clase e importancia de mensaje y de la relevancia o afinidad, según el caso, entre individuos.

Por otro lado, determinamos una forma de cuantificar la afinidad entre nodos y la fortaleza de la relación entre nodos, y esta la ligamos a la probabilidad de que un nodo reciba un mensaje.

Para validar los resultados de *LM*, realizamos pruebas en ordenador de *LM*, *LTM*, *ICM* y *WCM* con una red *Facebook* rastreada de un usuario anónimo. Estos resultados los comparamos con pruebas empíricas de propagación de información e influencia en la plataforma *Facebook* en la red del mismo usuario anónimo. *LM* representa el proceso de propagación de mensajes de manera más fina y sus resultados son notablemente más cercanos a los resultados empíricos en comparación con los resultados que se obtuvieron con los modelos *LTM*, *ICM* y *WCM*.

Con el propósito de estudiar el comportamiento de *LM*, llevamos a cabo pruebas en ordenador adicionales, modificando sus parámetros y utilizando una red sintética (artificial) *Power-Law* y dos redes sociales reales: *YouTube* y la red social *Facebook* de un específico usuario.

*LM* es adecuado para ser utilizado en diferentes ámbitos de la actividad humana, por ejemplo:

- 1- Permitirá conocer de una forma más precisa a los diferentes sectores de la sociedad el impacto que producen los mensajes que ellos publican en las *OSNs*.
- 2- Puede ser incorporado en los modelos de maximización de propagación de influencia en redes sociales.
- 3- Puede ser utilizado en los modelos que pronostican la pérdida de clientes en los servicios de telefonía

móvil. 4. También se puede utilizar en los sistemas que simulan la propagación de virus informático en las redes de computadores, redes de telefonía celular y redes sociales online.

## Abstract

Online social networks (*OSNs*) are increasingly used by different sectors of society with the purpose of informing and influencing people. These sectors are: governments, political parties, non-governmental organizations, suppliers of goods and services, etc. They wish they could predict the extent of the spread of information and the influence they transmit through these networks. Currently there are many mathematical models that allow us to predict the scope of these propagations, many of which are based on *Linear Threshold Model (LTM)* or *Independent Cascade Model (ICM)*.

In this document we present Lucy Model, a new predictive mathematical model that incorporates several elements that have not been considered or taken into account in the base models or in their derived models, but which are present in real life and affect the extent of propagation of messages on online social networks. These new elements are the following: We distinguish three classes of messages: purely informative messages; messages that are intended to influence appealing to people's emotions and feelings; and messages that are intended to influence by appealing to personal interests and conveniences. We argue that the intensity of the relationship between individuals (tie-strength) depends on the relevance and affinity between them. The probability that the message is received by an individual is directly proportional to this relationship intensity. Also, we consider that a message with greater importance is more likely to spread among people than a message with less importance. Finally, we maintain that individuals are informed or influenced (activated) after exceeding a threshold, which depends on the type and importance of the message and the relevance or affinity, as the case may be, between individuals.

On the other hand, we determine a way to quantify the affinity between nodes and the strength of the relationship between nodes, and we link this to the probability that a node receives a message.

To validate the results of *LM*, we performed laboratory tests of *LM*, *LTM*, *ICM* and *WCM* with a Facebook network tracked of an anonymous user. These results have been compared with empirical tests of information propagation and influence on the *Facebook* platform in the network of the same anonymous user. *LM* represents the message propagation process in a finer way and its results are notably closer to the empirical results in comparison with the results that were obtained with the models *LTM*, *ICM* and *WCM*. In order to study the behavior of *LM*, we carry out additional laboratory tests, modifying its parameters and using a synthetic (artificial) *Power-Law* network and two real networks (*Facebook* and *YouTube*).

*LM* is suitable to be used in different spheres of human activity, for example: 1- It will allow to know more precisely the different sectors of society about the impact produced by the messages that they publish in the *OSNs*. 2- It can be incorporated into the maximization models of influence propagation in social networks. 3- It can be used in models that predict the loss of customers in mobile telephone services. 4. It can also be used in systems that simulate the spread of informatics viruses in computer networks, cell phone networks and online social networks.

# Índice General

## 1 Contenido

1	Introducción .....	1
1.1	Motivación de la tesis .....	1
1.2	Antecedentes .....	1
1.3	Objetivos .....	2
1.4	Descripción general del presente trabajo .....	3
1.5	Método de investigación .....	4
1.6	Utilidad de <i>Lucy Model</i> .....	4
1.7	Contenido de la Tesis.....	4
1.8	Artículos publicados .....	5
2	Trabajos relacionados .....	6
2.1	Linear Threshold Model.....	6
2.2	Independent Cascade Model.....	7
2.3	Revisión de los dos modelos base anteriores .....	8
2.4	Otros modelos de propagación de influencia .....	9
2.5	Cuantificación de <i>tie-strength</i> social .....	11
3	Descripción de <i>Lucy Model</i> .....	12
3.1	Introducción al modelo .....	12
3.2	<i>Tie strength</i> y probabilidad de recibir mensajes.....	14
3.2.1	Afinidad <b><i>A<sub>ij</sub></i></b> .....	16
3.2.2	Normalización de afinidad y relevancia .....	18
3.3	Umbrales de activación e importancia de mensajes .....	19
3.4	Ejemplo: cálculo de <i>tie strength</i> y umbrales de activación .....	21
3.5	Búsqueda de nodos activos y activación de nodos .....	23
3.6	<i>Lucy Model</i> en bloques .....	24
3.7	Resumen del capítulo .....	28
4	Método de validación del modelo <i>Lucy Model</i> .....	30
4.1	Objetivos de validación del diseño. ....	30
4.2	Diseño de la validación .....	30
4.3	Protocolo para la extracción de datos empíricos.....	32
4.4	Protocolo para las pruebas en ordenador .....	32
4.5	Resumen el capítulo .....	34

5	Resultados de validación y análisis .....	36
5.1	Experimentos de campo .....	36
5.2	Pruebas en ordenador .....	37
5.3	Hallazgos.....	46
5.4	Resumen del capítulo .....	47
6	<i>Lucy Model</i> en la Telemática .....	48
6.1	Simulación de distribución de apps para teléfonos móviles.....	48
6.2	Propagación de virus informático en las redes de telefonía móvil .....	48
6.3	Predicción de pérdida de clientes de los operadores de telefonía móvil .....	49
6.4	Resumen de capítulo .....	50
7	Conclusiones y trabajo futuro .....	51
7.1	Conclusiones.....	51
7.1.1	Beneficios derivados de <i>Lucy Model</i> .....	51
7.1.2	Limitación de la implementación de <i>Lucy Model</i> .....	52
7.1.3	Dificultad de validación .....	52
7.2	Trabajo futuro .....	52
8	Apéndices .....	55
9	Fuentes e información.....	57



# Índice de figuras

Figura 1. Linear Threshold Model. Activación del nodo $v$ .....	7
Figura 2. Independent Cascade Model: Activación del nodo $w$ .....	8
Figura 3. Proceso de activación de nodos .....	14
Figura 4. Grafo de una red eléctrica .....	17
Figura 5. Red eléctrica equivalente .....	17
<i>Figura 6. Gij equivalente</i> .....	17
<i>Figura 7. Afinidad y relevancia de los nodos</i> .....	18
<i>Figura 8. Grafo usado en el ejemplo</i> .....	22
<i>Figura 9. Proceso de propagación de activación de nodos</i> .....	23
<i>Figura 10. Propagación de mensaje no homofílico</i> .....	26
<i>Figura 11. Propagación de mensaje homofílico</i> .....	27
<i>Figura 12. Propagación de mensaje informativo</i> .....	28
<i>Figura 13. Red de Facebook rastreada</i> .....	36
Figura 14. Exactitud: Propagación de un mensaje no homofílico .....	38
Figura 15. Exactitud: Propagación de mensaje homofílico .....	39
Figura 16. Exactitud: Propagación de un mensaje informativo. ....	39
Figura 17. Propagación de un mensaje no homofílico .....	40
Figura 18. Lucy Model. Red Facebook rastreada. ....	41
Figura 19. Lucy Model. Red YouTube .....	42
Figura 20. Lucy Model. Red sintética Power-Law. ....	42

## Índice de tablas

Tabla 1. Relación entre $Is$ y el conjunto $\epsilon$ .....	21
Tabla 2. Cuantificación de la importancia del mensaje .....	21
Tabla 3. Cálculo de $wji$ .....	22
Tabla 4. Cálculo de los umbrales de activación.....	22
Tabla 5. Asignación preliminar de grado $k$ a cada nodo .....	31
Tabla 6. Matriz de adyacencia del grafo .....	31
Tabla 7. Características de los grafos utilizados.....	31
Tabla 8. Rendimiento de simuladores implementados con red de Facebook rastreada y diferentes estructuras de datos .....	45
Tabla 9. Tiempo de ejecución de los simuladores [mm:ss] con la red social Facebook rastreada y listas ligadas .....	45
Tabla 10. Exactitud: Raíz cuadrada del error cuadrático medio .....	45
Tabla 11. Precisión: Desviación estándar $\sigma$ e Intervalo de Confianza .....	46

## Notación, variables y parámetros

$A_{ij}$	Afinidad entre $i$ y $j$
$A_0$	Conjunto inicial de nodos activos o semillas
$Ar_{ji}$	Afinidad que $j$ siente por $i$ en términos relativos
$b_{v,w}$	Peso del enlace entre los nodos $v$ y $w$ ( <i>LTM</i> )
$d_v$	Grado del nodo a activar $v$ ( <i>WCM</i> )
$E$	Conjunto de enlaces del grafo
$G$	Grafo
$G_{ij}$	Conductancia eléctrica equivalente entre $i$ y $j$
$i$	Un nodo de $G$
$Ip_j$	Importancia del mensaje percibida por $j$
$Is$	Parámetro del sistema. Importancia del mensaje que tiene para el promedio de la gente
$j$	Un nodo de $G$
$k$	Número de enlaces de un nodo ( <i>Power-Law</i> )
$n$	Número de nodos del grafo
$n_{ij}$	Número de nodos conectados a $i$ y $j$
$p(k)$	Probabilidad de que un nodo tenga $k$ enlaces en un grafo <i>Power-Law</i>
$p.u.$	Valor normalizado en por unidad
$p_{v,w}$	Probabilidad de que el nodo $v$ active al nodo $w$ ( <i>ICM</i> )
$R$	Resistencia eléctrica
$R_i$	Relevancia de $i$
$R_{ij}$	Resistencia eléctrica entre $i$ y $j$
$Rr_{ij}$	Relevancia relativa de $i$ respecto de $j$
$Uh_{ji}$	Umbral de influencia homofílica que presenta $j$ a $i$
$Ui_j$	Umbral de información de $j$
$Unh_{ji}$	Umbral de influencia no homofílica que presenta $j$ a $i$

## Notación, variables, parámetros y acrónimos

$V$	Conjunto de vértices o nodos del grafo $G$
$w_{ji}$	Fortaleza del enlace, probabilidad de que $j$ reciba el mensaje enviado por $i$
$w_{Aji}$	Fortaleza del enlace, probabilidad de que $j$ reciba el mensaje enviado por $i$ debido a la afinidad que $j$ siente por $i$
$w_{Rji}$	Fortaleza del enlace, o probabilidad de que $j$ reciba el mensaje enviado por $i$ debido a la relevancia que $j$ siente que tiene $i$
$X_i$	Variable independiente aleatoria ( <i>Teorema de Bernoulli</i> )
$\bar{X}_n$	Valor promedio de $n$ variables aleatorias ( <i>Teorema de Bernoulli</i> )
$\alpha$	Parámetro usado en los umbrales de activación de $LM$
$\beta$	Parámetro usado en los umbrales de activación de $LM$
$\varepsilon$	Cualquier número positivo ( <i>Teorema de Bernoulli</i> )
$\epsilon$	Conjunto de valores que depende de $ls$ ( $LM$ )
$\gamma$	Exponente de un grafo sintético o artificial tipo <i>Power-Law</i>
$\mu$	Valor esperado ( <i>Teorema de Bernoulli</i> )
$\theta_v$	Umbral de influencia del nodo $v$ ( $LTM$ )
$\theta_w$	Número aleatorio ( $ICM$ )
$\zeta(\gamma)$	Función Zeta de Riemann (grafos sintéticos <i>Power-Law</i> )

# Capítulo 1

## 1 Introducción

### 1.1 Motivación de la tesis

Desde que surgieron tímidamente a finales del milenio anterior y más decididamente a inicios del actual, las redes sociales online (*OSNs*) se han convertido en parte de nuestras vidas (Abbas, 2013; Yun & Gloor, 2015). Estas redes son cada vez más utilizadas por diferentes sectores de la sociedad con el doble propósito de informar e influir en la gente. Algunos de estos sectores son: gobiernos, partidos políticos, movimientos sociales, proveedores de bienes y servicios, etc. Esta situación está generando grandes volúmenes de información en Internet (Jin et al., 2013). A estos sectores les gustaría poder pronosticar el alcance de la propagación de la información y de la influencia que ellos transmiten a través de estas redes.

Actualmente existen muchos modelos que representan la propagación de mensajes a través de las *OSNs*. Nosotros consideramos que estos modelos no toman en cuenta ciertos elementos que están presentes en la vida real y afectan de forma importante dichos procesos. Motivados por esta situación, hemos desarrollado un nuevo modelo que represente de manera más cercana a la realidad el proceso de propagación de mensajes a través de las *OSNs* y que pronostique de forma más exacta el alcance, amplitud o distancia de dicha propagación.

### 1.2 Antecedentes

Una manera de pronosticar el alcance de la propagación de mensajes en las *OSNs* es utilizando el *Análisis de Redes Sociales (SNA)*, el cual es actualmente muy usado por investigadores (Bandeli & Agarwal, 2018; Ito et al., 2018; Kosorukoff & Passmore, 2011; Yang et al., 2016), y que es lo que nosotros también hemos hecho en nuestra investigación.

El *Análisis de Redes Sociales* surge de la Sociología, Antropología y Psicología, y es una forma de entender y representar las redes sociales. El concepto de redes sociales fue introducido por el antropólogo John Arundel Barnes en 1954. *SNA* se enfoca en la asociación que existe entre los individuos (amigos, clientes, empleados, estudiantes, o cualquier otro grupo de personas) más que en los atributos de dichos individuos. *SNA* es un método de investigación que toma prestado del campo de la Sociología una serie de conceptos: *métricas de centralidad, influencia, umbrales, fortaleza de enlaces y homofilia*, conceptos describiremos más adelante. *SNA* también utiliza herramientas matemáticas como la Teoría de Grafos, probabilidad y estadística y ciencias de la computación.

## Capítulo 1. Introducción

Una red social se representa matemáticamente como un grafo, cuya notación es  $G = (V, E)$ , donde  $V$  es un conjunto finito de nodos, cada uno de ellos simbolizando a un individuo.  $E \subseteq V \times V$  es un conjunto de enlaces que representa la relación entre cada par de nodos. El análisis de redes se centra en la asociación que existe entre nodos más que en los atributos de cada nodo.

Dentro de esta perspectiva, *Linear Threshold Model (LTM)* e *Independent Cascade Model (ICM)* (Kempe et al., 2003) son dos modelos estocásticos de referencia, que pronostican indistintamente el alcance de la propagación de información y de influencia a través de redes sociales. Un modelo derivado de *ICM* es *Weighted Cascade Model (WCM)* (Kempe et al., 2003). *LTM* e *ICM* son estudiados y usados en muchos trabajos científicos importantes dirigidos a maximizar la propagación de información en redes sociales. Algunos ejemplos son los siguientes: *Arendt and Blaha 2015; Worrell et al. 2017; Bulumulla et al. 2018; Cui et al. 2018; da Silva et al. 2018; Fischetti et al. 2018; N. et al. 2018*. Por esta razón, nosotros hemos evaluado nuestra propuesta comparándola con los modelos base *LTM* e *ICM*, con el modelo *WCM*, y con resultados empíricos.

*The Cambridge Dictionary* define el término *inform* como sigue: “to tell someone about particular facts”; y el término *influence* lo define así: “to affect or change how someone or something develops, behaves, or thinks”. *Informar* e *influir* no solo son dos conceptos diferentes, sino que implican distintos procesos de propagación de mensajes entre la gente.

### 1.3 Objetivos

Los objetivos de la presente investigación son los siguientes:

- a. Desarrollar un modelo (*Lucy Model*) que represente y pronostique el alcance de la propagación de información e influencia en redes sociales online, considerando elementos que están presentes en estos procesos en la vida real, que afectan de manera importante los resultados de la propagación de mensajes, pero que los modelos de base, ni los que derivan de estos, no los consideran.
- b. Realizar pruebas empíricas en la plataforma de Facebook, en la red social de un usuario anónimo para determinar el alcance real de los mensajes que se propagan en dicha red.
- c. Llevar a cabo pruebas en ordenador con nuestro modelo *LM*, y con los modelos de base *LTM* e *ICM*, y con *WCM* en la red social *Facebook* de dicho usuario anónimo, previamente rastreada. Utilizamos *LTM* e *ICM* porque han sido y son en la actualidad la base de muchos trabajos científicos. Utilizamos *WCM* porque es uno de los modelos que se deriva de los dos anteriores.
- d. Comparar los resultados de laboratorio con las pruebas empíricas y determinar el modelo que presenta resultados más cercanos a las pruebas empíricas.

Nuestro trabajo no pretende establecer un método para encontrar el mejor conjunto de semillas a fin de maximizar la propagación de mensajes en una red social.

## 1.4 Descripción general del presente trabajo

Las características de *Lucy Model* son ocho: predictivo, numérico computacional (no analítico), iterativo, no basado en datos históricos, estocástico, orientado a grafos, independiente del tiempo y parametrizable.

En el presente trabajo nosotros contribuimos con cinco elementos que no son considerados en los modelos de línea base ni en sus modelos derivados. Estos elementos son:

1. Importancia del mensaje.
2. Existencia de diferentes clases de mensaje.
3. Manera de cuantificar el umbral de activación (influencia o información) de los nodos.
4. Modelo para cuantificar la intensidad o fortaleza de la relación entre nodos.
5. Manera de cuantificar la afinidad entre nodos.

Las diferentes clases de mensajes que *Lucy Model* considera son tres:

1. Mensajes únicamente informativos.
2. Mensajes para influenciar apelando a los sentimientos y emociones de la gente.
3. Mensajes para influenciar apelando a los intereses y conveniencias de las personas.

Los umbrales de activación de los nodos se cuantificarán en función de la importancia del mensaje y relevancia de los nodos o la afinidad entre ellos, según sea el caso. La relevancia de la gente se cuantifica utilizando métricas de centralidad (Newman, 2010). La afinidad entre la gente se cuantifica con un modelo nuevo que proponemos y que explicaremos más adelante.

Nosotros consideramos que la relevancia de la gente y la afinidad entre ellas también afectan la intensidad de las relaciones interpersonales llamada *tie strength*. Basados en esta consideración, presentamos un nuevo modelo que cuantifica *tie strength* en redes sociales online.

El método que hemos utilizado para evaluar nuestra propuesta tiene los siguientes pasos:

1. Hemos creado un ambiente simulado implementado en computadora para los cuatro modelos predictivos *LM*, *LTM*, *ICM* y *WCM*.
2. Hemos rastreado la red social de un usuario anónimo de Facebook.
3. Hemos realizado simulaciones con los cuatro modelos predictivos sobre la red social rastreada antes indicada.
4. Para obtener los datos empíricos, hemos usado la plataforma de Facebook para enviar mensajes en la red social desde el usuario anónimo, y hemos constatado qué tan lejos se propagan dichos mensajes.

Los procesos de propagación en los cuatro modelos dependen de elementos aleatorios y por lo tanto sus resultados son probabilísticos. Estos resultados pueden ser determinados a través del método de *Montecarlo* (Bolthausen & Wüthrich, 2013), que a su vez se basa en el *Teorema de Jakob Bernoulli*, también llamado *Ley de los grandes números* (Dietz & Heesterbeek, 2002).

Las pruebas muestran concluyentemente que las predicciones de nuestro modelo (*LM*) son notablemente más exactas y precisas que las predicciones de los modelos base y *WCM*.

## Capítulo 1. Introducción

Con el propósito de estudiar el comportamiento de *LM*, realizamos pruebas en ordenador adicionales modificando sus parámetros en una red sintética *Power-Law* (Newman, 2010) y dos redes sociales reales: *YouTube* (<https://snap.stanford.edu/data/com-Youtube.html>) y la red social *Facebook* rastreada de un usuario particular.

### 1.5 Método de investigación

Los pasos de nuestro trabajo han sido guiados por el siguiente método de investigación.

- a. Estudio de los trabajos relacionados con la propagación de influencia y de información en redes sociales online, entre ellos: los modelos base *LTM* e *ICM*, el modelo *WCM*, y modelos de maximización de propagación de información e influencia actuales derivados de los modelos base.
- b. Revisión de ciertas áreas del conocimiento orientados al estudio de redes: Análisis de Redes Sociales, Teoría de Grafos, Teoría de Conjuntos, Teoría de Matrices, *Método de Montecarlo* para la resolución de problemas con variables aleatorias, Lenguaje de programación C/C++, Estructuras de datos.
- c. Diseño y construcción del simulador en computadora de nuestro modelo *Lucy Model*.
- d. Construcción de los simuladores en computadora de los modelos *LTM*, *ICM* y *WCM*.
- e. Pruebas de validación para verificar que el software desarrollado cumple con las especificaciones, es consistente con los datos de entrada, y logra su cometido.
- f. Simulación de propagación de mensajes en redes artificiales y reales con *LM*, *LTM*, *ICM* y *WCM*.
- g. Rastreo de una red de usuario anónimo de *Facebook*.
- h. Envío de mensajes en la plataforma de Facebook desde la cuenta de un usuario anónimo, y rastreo del alcance de la propagación de dichos mensajes.
- i. Análisis de resultados y conclusiones.

### 1.6 Utilidad de *Lucy Model*

*Lucy Model* es adecuado para ser usado en diferentes áreas de la actividad humana: para predecir el alcance de mensajes enviados a través de las *OSNs*; puede usarse como la base de nuevos modelos que maximicen la propagación de mensajes (Wang et al., 2018); para pronosticar el alcance de la propagación de virus informáticos en redes de computadoras (Piqueira & Araujo, 2009), en *OSNs* (Fan & Yeung, 2011; Luo et al., 2016), en redes telefónicas móviles, para predecir la pérdida de clientes de las compañías de telefonía móvil (Phadke et al., 2013), y en el amplio campo de la *ciber seguridad social* (comportamiento humano, político, gobernanza, seguridad nacional, bullying, segregación, etc.).

### 1.7 Contenido de la Tesis

Este documento tiene el siguiente contenido: El capítulo 2 explica con todo detalle los modelos predictores considerados referentes *Linear Threshold Model* e *Independent Cascade Model*, así como *Weighted Cascade Model*. Adicionalmente hacemos una breve descripción de varios otros modelos predictores, la mayoría de ellos se derivan de los modelos base o referentes. Finalmente hacemos mención de múltiples métodos de cuantificación de la



fortaleza de los enlaces *tie-strength* definidos por otros autores. En el capítulo 3 identificamos los elementos que afectan el alcance de la propagación de mensajes en una red social, los mismos que se recogen en *Lucy Model*. Luego de identificarlos, determinamos la manera de cuantificar dichos elementos. El capítulo 4 presenta el método de validación de *Lucy Model*, que consiste en describir las pruebas en ordenador y las pruebas de campo que se tienen que realizar para demostrar la validez de *Lucy Model*. El capítulo 5 muestra los resultados de las pruebas tanto de laboratorio como empíricas, con las cuales se demuestra en forma concluyente que nuestro modelo es más exacto y preciso que los modelos referentes. En el capítulo 6 describimos algunas aplicaciones en las cuales puede utilizarse *Lucy Model*, tales como: el análisis de la distribución de apps para teléfonos móviles; el estudio de la propagación de virus informático en las redes de telefonía móvil; y la predicción de la pérdida de clientes de los operadores de telefonía móvil. Finalmente, en el capítulo 7 resumimos las ventajas y desventajas de nuestro modelo *Lucy Model*, también relatamos las dificultades que tuvimos que superar en el proyecto de investigación. Finalmente, proponemos algunos nuevos trabajos que se podrían realizar y que se desprenden de la presente investigación.

### 1.8 Artículos publicados

La primera publicación relacionada con la presente tesis, fue el artículo que publicamos en diciembre del 2016 titulado “Centrality Metrics and Line-Graph to Measure the Importance of Links in Online Social Networks”. En este trabajo nosotros proponemos medir la fortaleza de las relaciones entre los nodos de una red o *tie-strength*, aplicando las métricas de centralidad de los nodos de un grafo. Para poder hacerlo nosotros convertimos los enlaces del grafo en nodos de un nuevo grafo, y entonces aplicamos las métricas de centralidad en esos nodos, que representan enlaces del grafo anterior. Los valores de relevancia de estos nodos son una medida de *tie-strength* del grafo original. Para convertir los enlaces de un grafo original en los nodos e un nuevo grafo, aplicamos el concepto matemático de Line-Graph. Este trabajo fue publicado en a la revista *International Journal of New Technology and Research*.

El segundo artículo publicado y motivado por la presente tesis fue “Extent prediction of the information and influence propagation in online social networks”, publicado en la revista *Computational & Mathematical Organization Theory*, con fecha 27 de Marzo del 2020. Este artículo constituye la base de la presente tesis de grado, por lo que aquí la describimos con todo detalle.

## Capítulo 2

### 2 Trabajos relacionados

El estudio de la propagación de mensajes a través de redes sociales online, ha despertado gran interés en la academia a lo largo de muchos años, es por esto que existen un sinnúmero de trabajos de investigación publicados relacionados con este tema, y aun así, el tema no se agota.

Actualmente, existen dos tipos de modelos predictivos de propagación de influencia: aquellos enfocados a grafos, y aquellos basados en datos de propagaciones anteriores. En estos últimos la propagación se representa a través de ecuaciones diferenciales, y se usan para pronosticar la propagación de enfermedades (Guille et al., 2013). El primer trabajo realizado para estudiar la propagación de enfermedades fue publicado por Daniel Bernoulli en 1766 (Dietz & Heesterbeek, 2002).

En esta sección, describimos doce modelos de propagación de influencia, que tienen una relación directa con *Lucy Model*. Primero describimos los dos modelos: *LTM* e *ICM*, que son pilares en la creación de otros modelos, algunos de los cuales también los describiremos. Hemos utilizado los modelos *Linear Threshold Model*, *Independent Cascade Model* y *Weighted Cascade Model* porque los dos primeros son considerados referentes, ya que ellos se están utilizando desde hace muchos años hasta la actualidad en innumerables trabajos de investigación, dirigidos a maximizar la propagación de información en redes sociales; y porque estos han servido de base para el desarrollo de otros modelos predictores como *Weighted Cascade Model*.

Todos estos modelos se centran en grafos. De aquí en adelante, a un nodo informado o influenciado, lo llamaremos indistintamente "*nodo activo*" o "*nodo activado*"

#### 2.1 Linear Threshold Model

Este modelo (Kempe et al., 2003) se basa en un umbral de influencia para cada nodo. Un nodo  $v$  es influenciado por cada vecino  $w$  acorde a un peso  $b_{v,w}$  tal que  $\sum_w \text{vecino de } v b_{v,w} \leq 1$ . Si todos los vecinos de  $v$  estuvieran activos la sumatoria sería igual a 1.

El proceso es el siguiente: cada nodo  $v$  elige un umbral  $\theta_v$  uniformemente en forma aleatoria dentro del intervalo cerrado  $[0, 1]$ . El umbral representa la fracción ponderada de vecinos de  $v$  que tienen que estar activos para que  $v$  se vuelva activo. Dada una elección aleatoria de umbrales, y un conjunto inicial de nodos activos  $A_0$  (con todos los otros nodos inactivos), el

proceso de difusión se desarrolla a través del tiempo en pasos discretos: todos los nodos que permanecieron activos en el paso  $t - 1$  continúan activos en el paso  $t$ , y cualquier nodo  $v$  se puede activar si el peso total de sus vecinos activos es al menos  $\theta_v$ :

$$\sum_{w \text{ neighbor of } v} b_{v,w} \geq \theta_v$$

De esta manera,  $\theta_v$  representa las diferentes tendencias latentes de los nodos para adoptar la innovación cuando sus vecinos lo hacen. El hecho de que esos son aleatoriamente seleccionados pretende modelar la falta de conocimiento de esos valores. Otro enfoque es fijar todos los umbrales a un valor conocido, por ejemplo  $\frac{1}{2}$ . Como se puede ver, *LTM* está orientado a los nodos inactivos porque se buscan nodos inactivos que cumplan la condición de activación.

Como ejemplo, en la Figura 1 mostramos el nodo inactivo  $v$  y sus vecinos ( $w_i$ ), de los cuales  $w_1$ ,  $w_4$  y  $w_5$  están inactivos. Los pesos  $b_{v,w_i}$  de los enlaces de  $v$  con sus vecinos activos se representan con líneas de diferente grosor. Si la suma de estos pesos es al menos igual a umbral  $\theta_v$ , entonces  $v$  será activado.

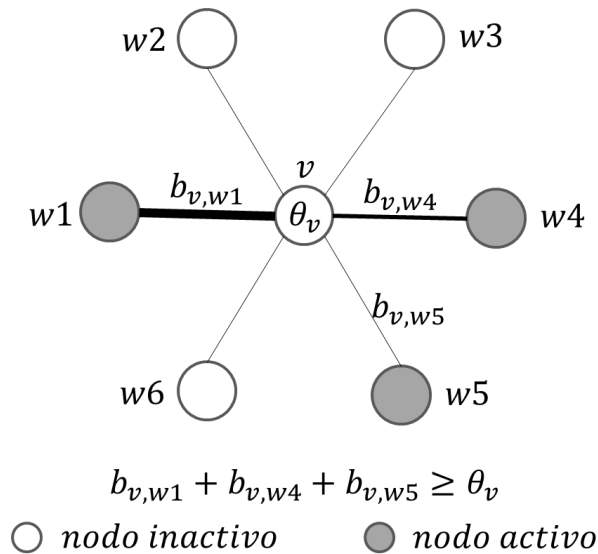


Figura 1. Linear Threshold Model. Activación del nodo  $v$

Sus autores manifiestan que este modelo es adecuado para un tipo de procesos llamado *contagio complejo*, en el que es más probable poder activar un nodo si se tiene una determinada cantidad de vecinos activos. Un ejemplo de esto es el proceso de propagación de influencia.

## 2.2 Independent Cascade Model

El proceso de propagación de influencia se inicia con un conjunto de nodos activos  $A_0$ , el proceso se realiza paso a paso de la siguiente manera. Cuando un nodo  $v$  se vuelve activo en el paso  $t$  este tiene sólo una oportunidad para activar a cada vecino  $w$  con una probabilidad  $p_{v,w}$  el cual es un parámetro del sistema. Sin explicaciones, sus autores establecen  $p_{v,w}$  en 1% y 10% en pruebas separadas, a mayor valor de  $p_{v,w}$ , es más probable que el nodo  $v$  influya en el nodo  $w$ .

Si  $w$  tiene nuevos vecinos recientemente activados, el intento de cada vecino de activar  $w$  es secuencial en orden arbitrario. Si  $v$  tiene éxito, entonces  $w$  se volverá activo en el paso  $t + 1$ ; de lo contrario  $v$  no hará ningún otro intento de activar  $w$  en las siguientes iteraciones. El proceso termina cuando no son posibles más activaciones. *ICM* está orientado a los nodos activos porque se busca estos nodos para que ellos traten de influir a sus vecinos inactivos.

En la Figura 2 mostramos el nodo inactivo  $w$ . Nosotros asumimos que  $p_{v,w} = 1\%$ .  $\theta_w$  es el umbral de influencia, un número aleatorio uniformemente distribuido en el intervalo  $[0, 1]$ . Si  $p_{v,w} \geq \theta_w$ , entonces  $w$  será activado.

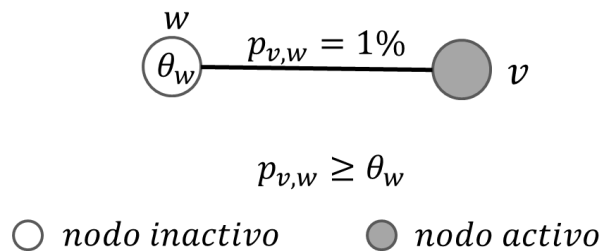


Figura 2. Independent Cascade Model: Activación del nodo  $w$

Sus autores manifiestan que este modelo es adecuado para procesos tipo *contagio simple*, donde una sola fuente es capaz de activar un nodo, como por ejemplo, la difusión información o de virus.

### 2.3 Revisión de los dos modelos base anteriores

Los autores de los modelos de referencia *LTM* e *ICM* (Kempe et al., 2003) consideran nodos de dos tipos: activos e inactivos. El nodo activo es aquel que ha adoptado una innovación. El nodo inactivo es el caso opuesto. Los autores dicen que influir en un nodo es activar ese nodo. La propagación de influencia o información comienza en un conjunto de nodos activos llamados semillas.

Con el tiempo, más y más vecinos de un nodo inactivo ( $v$ ) se activan. En algún momento, esto puede hacer que  $v$  se active, lo que a su vez puede hacer que los vecinos inactivos de  $v$  también se activen.

Los modelos de referencia tienen ocho restricciones:

1. Las fuentes de información e influencia están dentro de la red social, y los nodos no pueden ser influenciados por fuentes externas.
2. No se tiene en cuenta ningún contexto, como la región geográfica, la época del año, el entorno social, etc.
3. La estructura topológica del grafo permanece estática durante el proceso de propagación.
4. Los umbrales de activación de los nodos permanecen constantes a través de cada proceso de propagación de mensajes.
5. Un nodo que tiene un cierto estado no puede volver a un estado anterior, por lo que decimos que estos modelos son progresivos.
6. Ningún nodo puede negarse a ser semilla.

7. Ningún nodo puede ser semilla voluntariamente.
8. La influencia sólo puede ser propagada de nodo a nodo a través de los enlaces de la red.

Estos modelos tienen dos propiedades (Chen et al., 2013): *monotonidad*: agregar nodos a un conjunto de semillas no reduce el conjunto final de nodos activos; y *submodularidad*: se obtiene un incremento marginal decreciente cuando se agregan nodos a un conjunto de semillas. Los nodos se comportan de forma progresiva, es decir, un nodo puede pasar de inactivo a activo pero no de activo a inactivo.

## 2.4 Otros modelos de propagación de influencia

**Weighted Cascade Model.** Este modelo (Kempe et al., 2003) es considerado un caso especial de *Independent Cascade Model*, en el que en cada enlace  $\{u, v\}$  el nodo  $u$  tiene la probabilidad  $1/d_v$  para activar  $v$ , siendo  $d_v$  el grado (número de enlaces) del nodo  $v$ . *WCM* realiza iteraciones visitando los nodos del grafo en busca de nodos activos. Cada nodo activo solo tiene una oportunidad para activar sus vecinos. Si al final de la iteración actual hay al menos un nuevo nodo activado, se inicia una nueva iteración; de lo contrario, el proceso termina. Igual que con *ICM*, este modelo también es orientado a los nodos activos.

**Deterministic Linear Threshold Model (DLTM).** La diferencia de este modelo (Swaminathan, 2014) con *LTM* es que los umbrales de activación de *DLTM* se asignan determinísticamente.

**Deterministic Threshold Model (DTM).** En este modelo (Swaminathan, 2014), a cada enlace dirigido  $(u, v)$  se le asigna un peso fijo 1. A cada vértice  $v$  se le asigna un umbral fijo  $\theta_v \geq 1$  y es un número entero. Un vértice  $v$  cambia su estado  $S_v$ , de inactivo a activo ( $0 \rightarrow 1$ ), si la suma de los pesos de los enlaces que vienen de los nodos que están en estado activo es mayor o igual a  $\theta_v$ ; de lo contrario, el vértice  $v$  permanece inactivo.

**Triggering Model.** En este modelo (Kempe et al., 2003), cada nodo  $v$  elige un conjunto disparador aleatorio de vecinos  $T_v$ . Si  $v$  está inactivo en el tiempo  $t$ , y se ha activado un vecino en  $T_v$ , el nodo  $v$  también se activará en el tiempo  $t + 1$ .

**Only-Listen-Once Model.** Este (Kempe et al., 2003) es un caso especial del modelo de activación (Triggering Model). Aquí, cada nodo  $v$  tiene un parámetro  $p_v$ , de modo que el primer vecino de  $v$  que se activa, intenta activar a  $v$  con probabilidad  $p_v$ , y todos los intentos posteriores de activar a  $v$  fallan. En otras palabras,  $v$  solo escucha al primer vecino que intenta activarlo.

**Nonprogressive Processes.** En un proceso progresivo, los nodos solo pasan de inactividad a actividad, pero no viceversa. En el caso no progresivo (Kempe et al., 2003), los nodos pueden cambiar en ambas direcciones, de hecho puede reducirse al caso progresivo. Esto se logra con umbrales de influencia  $\theta_v(t)$  que varían con el tiempo, distribuidos uniformemente en el intervalo  $[0, 1]$ . El nodo  $v$  permanece activo si  $f_v(S) \geq \theta_v^t$  donde  $f$  es una función que determina la intensidad con la que el conjunto de vecinos  $S$  intentan activar  $v$ .

**Parallel Cascade model (PC)** (Samadi et al., 2016). Este es un modelo de propagación de influencia que utiliza la lógica de inferencia bayesiana (un tipo de inferencia estadística en la que las evidencias se emplean para actualizar o inferir la probabilidad de que una hipótesis

pueda ser cierta). Se consideran dos propuestas opuestas. Cada propuesta compite para prevalecer en un grupo de personas que pertenecen a una red social. Cada propuesta se difunde a través de la red comenzando en dos conjuntos de semillas  $S^+$  y  $S^-$  respectivamente. Cada nodo  $i$  recibe la influencia positiva  $L_{it}$  y la influencia negativa  $K_{it}$  en el tiempo  $t$ , que se acumulan con el tiempo. A cada nodo  $i$  se le asigna un umbral de influencia positiva  $\theta_i^+$  y un umbral de influencia negativa  $\theta_i^-$ . El nodo  $i$  está influenciado positivamente si  $L_{it} - K_{it} \geq \theta_i^+$ , o está influenciado negativamente si  $K_{it} - L_{it} \geq \theta_i^-$ ; de lo contrario, el nodo  $i$  permanece sin influencia. Cada nodo acumula influencia que proviene de sus vecinos, independientemente de si está influenciada positiva o negativamente. Solo un nodo influenciado puede propagar influencia positiva o negativa a sus vecinos.

**Partial Parallel Cascade model (PPC)** (Samadi et al., 2018). Este modelo es una variante del modelo *PC* que permite la activación parcial de nodos. En cada período de tiempo  $0 \leq t \leq T$ , cada nodo  $i$  entrega influencia positiva ( $X_{it}$ ) o negativa  $Y_{it}$  a sus vecinos salientes, donde  $0 \leq X_{it} \leq 1$  y  $0 \leq Y_{it} \leq 1$  es la activación parcial positiva (negativa) del nodo  $i$  en el tiempo  $t$  respectivamente. La cantidad de influencia positiva  $X_{it}$  ( $Y_{it}$  negativa) es una función de  $L_{it} - K_{it} \geq \theta_i^+$  ( $K_{it} - L_{it} \geq \theta_i^-$ ). Si  $L_{it} - K_{it} \geq \theta_i^+$  ( $K_{it} - L_{it} \geq \theta_i^-$ ), el nodo  $i$  se considera totalmente influenciado de manera positiva (negativa).  $\theta_i^+$  y  $\theta_i^-$  pueden o no ser iguales.

**Structural diversity model in social contagion** (Ugander et al., 2012). Sus autores encontraron a través del análisis empírico que la probabilidad de aceptar una recomendación depende más sutilmente de la estructura de la red, en lugar de la cantidad de amigos que están influenciados; es decir, esta aceptación depende de los diferentes grupos de vecinos en lugar del número de vecinos. Cada grupo o diversidad estructural representa un contexto social diferente al que pertenece el usuario: familia, compañeros de trabajo, compañeros de escuela, etc. En este modelo, solo los primeros amigos influenciados pueden influir en un amigo no influido.

**Influence propagation model** (Phadke et al., 2013). Este modelo se basa en la cuantificación de la variable llamada *tie strength*. Si el receptor es un amigo cercano del remitente, reflejado en una mayor fuerza de conexión con respecto a otros posibles remitentes, entonces es más probable que el receptor sea influenciado. La cantidad de influencia retenida por el receptor es relativa a la fuerza del vínculo social. Tan pronto como el receptor recibe cierta cantidad de influencia, envía la misma cantidad de influencia a todos sus vecinos. La influencia total acumulativa es la suma de las influencias recibidas por un nodo a través de sus vecinos activos. Esta influencia neta se utiliza como uno de los predictores para que el usuario tome una decisión.

Por otro lado, es interesante el estudio que (Vishwakarma et al., 2019) realizan para detectar noticias falsas en forma de imágenes que se difunden a través de las redes sociales. El sistema utiliza tecnologías para análisis de texto y *web scraping* (técnica utilizada mediante programas de software para extraer información de sitios web). Su estudio trabaja en la extracción de características de la imagen y el texto. El algoritmo aplicado utiliza varios recursos en línea para detectar la credibilidad de las noticias (canales de noticias como *Fox News*, *CNN* y sitios web de periódicos como *The Washington Post*). Los resultados del algoritmo dependen en

gran medida de las fuentes mencionadas anteriormente, porque los autores afirman que su confiabilidad es obligatoria por su buen nombre en el mercado.

## 2.5 Cuantificación de *tie-strength* social

Hasta ahora, hemos descrito los trabajos relacionados con la propagación de información e influencia. El número de nodos alcanzados por un mensaje que se propaga a través de una red social depende directamente de *tie strength* social entre los nodos. *Lucy Model* también presenta una forma de cuantificar *tie strength*. A continuación, describiremos primero algunos trabajos relacionados con la cuantificación de esta fortaleza.

(Phadke et al., 2013) presentan una forma de cuantificar *tie strength* entre dos usuarios de una red de telefonía móvil, utilizando la siguiente expresión matemática:  $w(x) = 1 - e^{-\frac{x}{\epsilon^2}}$ , donde  $x = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$ .  $\alpha_1, \alpha_2, \dots, \alpha_n$  y  $\epsilon$  se asignan empíricamente utilizando un conjunto de datos de entrenamiento  $x_1, x_2, \dots, x_n$  son atributos expresados en valores por unidad, por ejemplo: número de llamadas realizadas entre dos usuarios, la duración total de las llamadas realizadas entre dos usuarios, la proporción de vecinos que dos usuarios tienen en común entre sí.  $w(x)$  está restringido al intervalo  $[0, 1]$ .

*Edge betweenness centrality* de (Girvan & Newman, 2002) mide la cantidad de rutas más cortas que pasan a través de un enlace para conectar dos nodos de una red. Esta medida es una generalización de *point betweenness centrality* de Linton C. Freeman.

(Teixeira et al., 2013) definen la centralidad de expansión de un enlace  $e$  para un grafo ponderado y no dirigido como la relación entre el número de árboles de expansión mínima donde  $e$  participa y el número total de árboles de expansión mínima.

Donald S. Sade diseña la centralidad del nodo  $k$ -path que está generalizado por (De Meo et al., 2012) y se aplica a los enlaces, que se define como el número de caminos, a lo sumo, longitud  $k$  que se conecta con otros enlaces.

(Postigo-Boix & Melús-Moreno, 2018) proponen medir la distancia entre dos nodos  $p$  y  $q$ :  $distance_{pq} = \sqrt{w_1(p_1 - q_1)^2 + \dots + w_n(p_n - q_n)^2}$ ,  $1, 2, \dots, n$  son atributos.  $w_i$  es el peso asignado a cada dimensión, que se utiliza para normalizar o dar más o menos importancia a cada atributo. Los autores definen la afinidad como  $affinity = 1 - distance_{pq}$ . Si la afinidad es una medida de cercanía entre dos nodos, entonces la afinidad es una medida de la fuerza de unión entre los nodos.

## Capítulo 3

### 3 Descripción de *Lucy Model*

#### 3.1 Introducción al modelo

*Lucy Model* está orientado a las redes sociales modeladas matemáticamente como un grafo no dirigido  $G = (V, E)$ , donde  $V$  es un conjunto finito de vértices o nodos y  $E \subseteq V \times V$  es un conjunto de enlaces que conectan pares de nodos. Los nodos representan individuos y los enlaces representan relaciones entre individuos. Se aplican las mismas ocho consideraciones, restricciones y propiedades de los modelos de referencia (Sección 2.3). Como veremos más adelante, *LM* es configurable para que se adapte a las diferentes necesidades de cada situación, permitiendo flexibilidad de uso. *LM* tiene las características de *monotonidad* y *submodularidad*, (ver sección 2.3).

El propósito de una red social es para difundir mensajes de diferente clase. Nosotros hacemos una propia clasificación de un mensaje, en primer lugar, determinando cuál es su intención. Nosotros identificamos que, en general, existen dos clases de mensajes: mensajes que tienen la intención únicamente de informar, y mensajes que tienen la intención de influenciar en la gente para cambiar su comportamiento o su forma de pensar. Dentro de esta última clase, nosotros determinamos que la influencia puede conducirse a través de dos vías: apelando a los sentimientos y emociones de las personas; o apelando a los intereses y conveniencias individuales de las personas. Finalmente, nosotros encontramos que no aceptar la influencia de una clase de mensajes puede significar un riesgo para la gente; y, por otro lado, aceptar la influencia de otra clase de mensajes puede significar un riesgo para la gente.

*LM* considera la propagación de tres clases distintas de mensajes:

1. Mensajes para informar.
2. Mensajes para influir que apelan a los sentimientos.
3. Mensajes para influir que apelan a intereses y conveniencias personales.

Estos mensajes provocan en los individuos umbrales de activación distintos, que afectan el rango de propagación del mensaje y el número de nodos activados. Por otro lado, *LM* considera que la importancia de un mensaje y la intensidad de la relación entre las personas también afectan el alcance de su difusión.

Los modelos de referencia consideran que un nodo se activa sin distinguir si han sido informados o si han sido influenciados. *Lucy Model* considera que un nodo se activa de acuerdo con la clase de mensaje que se propaga en la red, como se indica en el párrafo



anterior. Por esta razón, *LM* está compuesto por tres módulos, uno para cada clase de mensaje.

Los mensajes se propagan gradualmente a través de la red social, así que los hemos modelado como procesos en tiempo discreto que son realizados paso a paso. Cada uno de esos pasos se representa con una iteración en los procesos de cada módulo. El proceso se inicia con un grupo de nodos llamados *semillas* que están previamente activados. Estas semillas tratan de activar a sus vecinos, estos a los suyos, y así sucesivamente, produciendo un proceso en cascada. (Zuo et al., 2016) sostiene que la influencia se propaga a un máximo de dos saltos de distancia de una semilla. Existe una relación directa entre el número de semillas y el número de nodos activados al final del proceso de propagación de mensajes en la red social. Solamente los nodos activos pueden tratar de activar a sus vecinos.

Los nodos están en uno de tres estados.

- Estado 1: nodo inactivo.
- Estado 2: nodo activo, habilitado para activar.
- Estado 3: nodo activo, inhabilitado para activar.

Si se activa un nodo en estado 1, entonces el nodo cambia a estado 2 y por lo tanto adquiere la capacidad de activar a sus vecinos. Un nodo en estado 2 tiene solamente una oportunidad de activar a sus vecinos. Después de que un nodo en estado 2 intenta activar a sus vecinos, pasa a estado 3. Al final del proceso solo habrá nodos en el estado 3 y eventualmente nodos en el estado 1. Dependiendo de la situación, un nodo se activa porque está influenciado o porque está informado. Para ambas situaciones en este documento usamos el término *activado*. El proceso de cambiar el estado de los nodos lo explicaremos en detalle más adelante en esta sección 3.5.

Debido a las consideraciones hechas en los párrafos anteriores de esta sección, *Lucy Model* necesita conocer cuatro elementos: la topología de la red, la importancia del mensaje, el número de semillas y su identidad.

Un mensaje que se propaga a través de la red sigue un proceso de tres pasos:

1. Buscamos en el grafo nodos activos habilitados para activar (nodos en estado 2).
2. Cada nodo encontrado en el estado 2 envía el mensaje a sus vecinos inactivos (nodos en el estado 1). Estos reciben el mensaje con probabilidad  $w_{ji}$  (Sección 3.2).
3. Si se recibió el mensaje, se verifica si el mensaje excede el umbral de activación del nodo; si este es el caso, el nodo pasa al estado 2.

Debido al paso (1), *Lucy Model* está orientado a los nodos activos.

En la sección 3.5, se describirán los algoritmos correspondientes a los pasos 1, 2 y 3. La Figura 3 explica los tres procesamientos de mensajes mencionados anteriormente.

A continuación describimos: *tie strength* y la probabilidad de recibir mensajes; umbrales de activación y la probabilidad de activar nodos; el proceso de propagación de mensajes; y el proceso de activación de nodos.

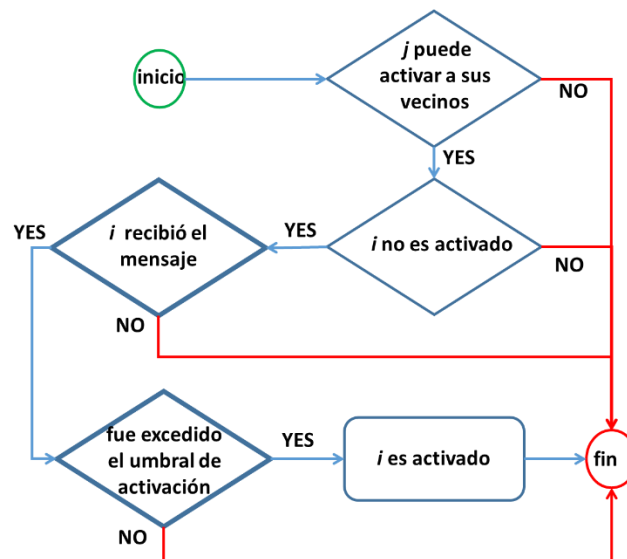


Figura 3. Proceso de activación de nodos

### 3.2 *Tie strength* y probabilidad de recibir mensajes

Es un hecho evidente que las relaciones que tenemos con las personas son de dos tipos: sentimentales, e interesadas. Relaciones sentimentales las mantenemos con personas cercanas: familiares, pareja sentimental y amigos; estas se conocen como *relaciones homofílicas*, con quienes nos unen sentimientos de afecto, afinidad y amistad, y tienden a perdurar en el tiempo. Las relaciones interesadas las mantenemos con personas que consideramos relevantes, quienes nos son convenientes, nos resuelven problemas, nos dan beneficios y satisfacen necesidades prácticas. Estas relaciones son circunstanciales, temporales y son las que mantenemos con compañeros de trabajo, colegas de profesión, vendedores y clientes, vecinos, simples conocidos, etc. Éstas se conocen como relaciones *no homofílicas*. Estas dos motivaciones están presentes y combinadas en mayor o menor medida en toda relación entre los individuos.

Las motivaciones sentimental y de interés que condicionan la fuerza de la relación entre las personas, también condicionan las relaciones entre los usuarios de las redes sociales online. *Lucy Model* define un modelo de fortaleza de enlaces (*tie strength*) para redes sociales online considerando esas dos motivaciones.

*Tie strength* en redes sociales es una propiedad que caracteriza el enlace entre dos nodos. *Tie strength* se puede cuantificar de tres maneras diferentes: (1) en términos de la topología de la red, (2) en términos funcionales, es decir, considerando la cantidad de mensajes que fluyen entre los nodos, y (3) en función de la Los atributos de los nodos: edad, género, raza, etc. (Granovetter, 1973) afirma que existe una relación entre *tie strength* y la estructura topológica de la red. Nosotros cuantificaremos *tie strength* en términos de la topología de la red, lo cual se describe más adelante.

Por definición, *tie strength* es una "combinación de la cantidad de tiempo, la intensidad emocional, la intimidad (confianza mutua) y los servicios recíprocos que caracterizan el enlace" (Granovetter, 1973). (Petróczi et al., 2006) afirma: "Los cuatro siguientes indicadores son componentes reales de *tie strength*: cercanía, duración y frecuencia, amplitud de temas,

y confianza mutua, mientras que las cinco contingencias contextuales siguientes: vecindario, afiliación, estatus socioeconómico, lugar de trabajo, y prestigio laboral son predictores. Las primeras cuatro contingencias contextuales anteriores son características de afinidad entre las personas”. Otras contingencias contextuales que caracterizan la afinidad entre las personas son, por ejemplo: gustos, pasatiempos, pasiones, religión, familia, edad, género, etnia, educación o personas que sufren la misma tragedia. La última contingencia contextual (prestigio ocupacional) se traduce en relevancia o importancia del individuo dentro de su red social. El fenómeno por el cual las personas tienden a establecer relaciones por razones de afinidad se llama *homofilia* (McPherson et al., 2001; Mondani, 2018). Por otro lado, si las personas están interesadas en interactuar con personas relevantes, estas son relaciones no homofílicas.

En resumen, los dos grandes predictores de *tie strength* son: la afinidad entre los individuos y la relevancia de estos individuos. Luego se deduce que las personas establecen relaciones por dos razones: por afinidad, por el interés de relacionarse con personas relevantes, o por ambas razones. *Lucy Model* propone una expresión matemática para cuantificar *tie strength* para las redes sociales online en función de esos dos grandes predictores.

La afinidad también es un sentimiento. La afinidad que siente el nodo  $i$  con respecto al nodo  $j$  no necesariamente es igualmente recíproco. Por otro lado, la relevancia de  $i$  en general es diferente de la relevancia de  $j$ . Por lo tanto, hay dos *tie strength* entre  $i$  y  $j$ , una de  $i$  a  $j$  y otra de  $j$  a  $i$ . Cuanto mayor es la afinidad que  $j$  siente por  $i$ , y mayor es la relevancia que  $j$  considera que tiene  $i$ , entonces *tie strength* de  $j$  hacia  $i$  (simbólicamente  $w_{ji}$ ) es mayor. Esto se traduce en una mayor predisposición de  $j$  para aceptar mensajes de  $i$ .

La afinidad absoluta entre los nodos  $i$  y  $j$  es  $A_{ij}$ .  $R_i$  es la cuantificación de la centralidad del nodo  $i$ .  $A_{ij}$  y  $R_i$  son magnitudes que tienen diferente naturaleza y escala. Para poder combinarlos en una expresión matemática para obtener  $w_{ji}$ , es necesario normalizar estas magnitudes, luego:

$$w_{Aji} = \text{normalized}(A_{ij})$$

$$w_{Rji} = \text{normalized}(R_i)$$

Y por tanto:

$$w_{ji} = f(w_{Aji}, w_{Rji})$$

$w_{Aji}$  es la afinidad que siente el nodo  $j$  con respecto al nodo  $i$ .  $w_{Aji}$ , al ser una magnitud normalizada, es la probabilidad de que el nodo  $j$  reciba mensajes del nodo  $i$  considerando la afinidad entre los dos nodos.  $w_{Rji}$  es la relevancia que  $j$  considera que tiene  $i$ .  $w_{Rji}$ , normalizado, es la probabilidad de que el nodo  $j$  reciba mensajes del nodo  $i$  considerando la relevancia del nodo  $i$ .  $w_{ji}$  es la probabilidad conjunta de dos eventos  $w_{Aji}$  y  $w_{Rji}$  probabilísticamente independientes entre sí. Luego:

$$w_{ji} = w_{Aji} * w_{Rji} \quad (1)$$

*Tie strength*  $w_{ji}$  está normalizado y representa la probabilidad de que el nodo  $j$  reciba mensajes del nodo  $i$ . Como se verá más adelante, en general  $w_{ji} \neq w_{ij}$ .

La ecuación (1) es consistente con lo expresado por Mark S. Granovetter en el sentido de que los vínculos dentro de las comunidades son fuertes y los vínculos entre estos grupos son

débiles. De hecho, los enlaces que unen a las comunidades tienen un valor pequeño  $w_{Aji}$  y, por lo tanto,  $w_{ji}$  es pequeño. Dentro de las comunidades, el valor de  $w_{Aji}$  es grande, obteniendo valores más grandes de  $w_{ji}$ . A continuación, encontramos expresiones para cuantificar  $A_{ij}$ ,  $w_{Aji}$  y  $w_{Rji}$ .

### 3.2.1 Afinidad $A_{ij}$

Para cuantificar *tie strength* debido a la afinidad  $w_{Aji}$ , necesitamos calcular la afinidad  $A_{ij}$  entre los nodos en función de la topología de la red. Jennifer Golbeck (Golbeck, 2013) afirma que "two friends are linked together by a strong tie if they have friends in common". Si esto es así, entonces podemos inferir que estos dos amigos se unen con sentimientos de afecto, afinidad y simpatía que es la amistad. Recordando lo que Mark S. Granovetter dijo que existe una relación entre *tie strength* y la estructura topológica de la red, nosotros encontraremos una expresión matemática para calcular  $w_{Aji}$  de acuerdo con el número de vecinos comunes que tiene cada par de nodos relacionados.

Observamos que existen las siguientes analogías entre una red eléctrica y una red social:

1. Las dos estructuras son redes que pueden representarse mediante un grafo.
2. Ambas redes pueden tener fuentes y admitir un flujo. En el primer caso, las fuentes son de energía y soportan el flujo de corriente eléctrica. En el segundo caso, las fuentes son de mensajes y soportan el flujo de dichos mensajes.
3. En un circuito eléctrico, la corriente fluye desde los puntos con mayor potencial eléctrico hacia puntos con menos potencial. En una red social, un mensaje fluye de las personas que tienen el mensaje a las personas que no lo tienen.
4. En un circuito eléctrico, los conductores tienen la propiedad de facilitar el flujo de corriente, y depende de las características físicas y eléctricas del material. Esta propiedad es la conductancia eléctrica  $G$ . En una red social, los enlaces tienen la propiedad de facilitar el flujo de mensajes, y depende de lo que luego veremos que es la afinidad  $A_{ij}$  que existe entre los individuos. A mayor afinidad, mayor *tie strength* entre un par de nodos, y por tanto, podemos ver una analogía entre  $G$  y  $A_{ij}$ .

Al identificar las analogías entre los dos tipos de redes, hacemos la abstracción de tratar una red social como si fuera una red eléctrica. Luego calculamos la conductancia eléctrica de los enlaces y asumimos que estas medidas son numéricamente iguales a las afinidades de los mismos enlaces en la red social. Anteriormente, para diferentes propósitos, ya se habían establecido similitudes entre dos sistemas pertenecientes a diferentes áreas de conocimiento; Por ejemplo, el concepto de ecosistema utilizado en Biología y Ciencias de la Vida se adaptó en un área de ciencias sociales para modelar el mundo de los negocios con el nombre de "Ecosistema de negocios" (Maturo et al., 2018; Rong et al., 2015). Otro ejemplo es la modelización de la propagación de virus informáticos a partir de modelos de propagación de enfermedades en poblaciones humanas (Piqueira & Araujo, 2009; Uddin et al., 2015). (Brandes & Fleischer, 2005) no determinaron la afinidad entre los individuos, sino que establecieron variantes de medidas de centralidad, considerando que la información se propaga eficientemente como una corriente eléctrica.

Nosotros procedemos a calcular la conductancia eléctrica equivalente  $G_{ji}$  entre los nodos  $j$  e  $i$  (Figura 6). Consideremos la porción de un grafo presentado en la Figura 4 como si fuera una red eléctrica. Los nodos 1, 2, ...,  $n$  están conectados a los nodos  $i$  y  $j$ , e interconectados entre sí. Suponemos que todos los enlaces tienen una resistencia eléctrica  $R = 1$ .

El cálculo de  $G_{ji}$  se puede hacer de dos maneras: aplicando el *Teorema de Kennelly* (Alexander & Sadiku, 2017), o de la siguiente manera. Si se aplica un voltaje entre los nodos  $i$  y  $j$ , los nodos 1, 2, ...,  $n$  estarán en el mismo potencial eléctrico, por lo tanto, no habrá corriente entre estos nodos (Alexander & Sadiku, 2017), entonces sus enlaces pueden ser eliminados, como se ve en la Figura 5.

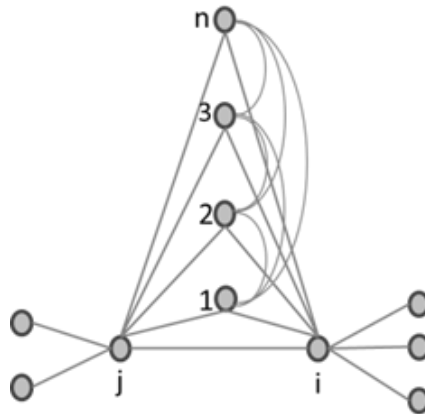


Figura 4. Grafo de una red eléctrica

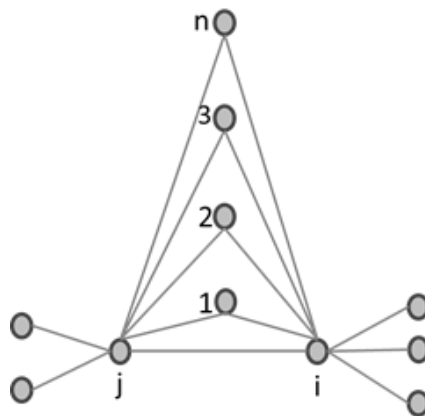


Figura 5. Red eléctrica equivalente

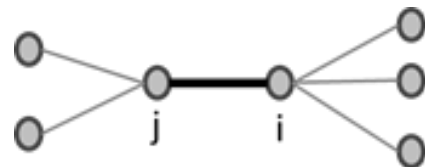


Figura 6.  $G_{ij}$  equivalente

$$R_{jni} = R_{jn} + R_{ni} = R + R = 2$$

$$R_{j1i} = R_{j2i} = \dots = R_{jni} = 2R$$

La conductancia eléctrica equivalente entre los nodos  $j$  e  $i$  (Figura 6) es:

$$G_{ji} = \frac{1}{R_{ij}} + \underbrace{\frac{1}{2R} + \frac{1}{2R} + \dots + \frac{1}{2R}}_n \quad (2)$$

$$G_{ji} = \frac{1}{R_{ij}} + \frac{n}{2R} \quad (3)$$

Si hemos asumido que cada enlace tiene una resistencia eléctrica igual a 1, entonces:

$$G_{ji} = 1 + \frac{n}{2} \quad (4)$$

done  $n$  es el número de nodos conectados a  $i$  y  $j$ .

Por la analogía existente ya indicada entre un circuito eléctrico y su red social correspondiente, concluimos que la conductancia equivalente  $G_{ji}$ , es numéricamente igual a la afinidad  $A_{ij}$ .

$$A_{ij} = 1 + \frac{n_{ij}}{2} \quad (5)$$

donde  $n_{ij}$  es el número de vecinos comunes a  $i$  y  $j$ . Observe que la ecuación (5) nos dice que la afinidad entre  $i$  y  $j$  es recíproca, es decir  $A_{ij} = A_{ji}$ .

### 3.2.2 Normalización de afinidad y relevancia

Los valores numéricos normalizados son valores relativos que poseen más información que sus respectivos valores absolutos. Recurrimos al proceso de normalización cuando los valores absolutos están expresados en escalas diferentes, o corresponden a magnitudes de distinta naturaleza. Los valores normalizados tienen unidades adimensionales homogéneas. Con estos valores podemos hacer comparaciones, operaciones aritméticas cuyos resultados pueden ser analizados para luego sacar conclusiones.

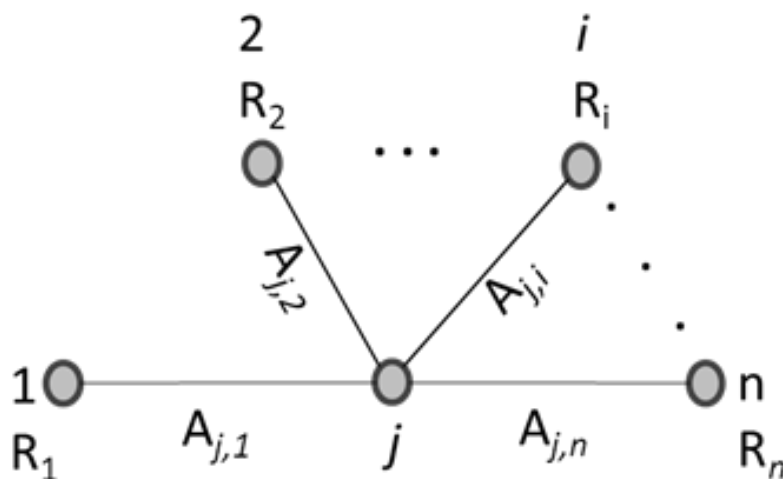


Figura 7. Afinidad y relevancia de los nodos

La probabilidad de que el nodo  $j$  reciba un mensaje del nodo  $i$  considerando la afinidad entre los dos nodos ( $w_{A_{ji}}$ ), es directamente proporcional a la afinidad entre  $i$  y  $j$  ( $A_{ij}$ ) en relación con la suma de las afinidades de  $j$  con todos sus vecinos (Figura 7), es decir

$$w_{Aji} = \frac{A_{ij}}{\sum_{\substack{k=1 \\ \{j,k\} \in E}}^n A_{jk}} \quad (6)$$

De manera similar, la probabilidad de que el nodo  $j$  reciba un mensaje del nodo  $i$ , considerando la relevancia del nodo  $i$  ( $w_{Rji}$ ), es directamente proporcional a la relevancia de  $i$  ( $R_i$ ) en relación con la suma de la relevancia de los vecinos de  $j$  (Figura 7).  $R_i$  se puede calcular utilizando algunas métricas de centralidad, es decir, *grado*, *PageRank*, *Eigenvector*, etc. (Bekiari & Hassanagas, 2015). Luego:

$$w_{Rji} = \frac{R_i}{\sum_{\substack{k=1 \\ \{j,k\} \in E}}^n R_k} \quad (7)$$

Las ecuaciones (6) y (7) son magnitudes normalizadas cuyos valores están en el intervalo [0, 1].

### 3.3 Umbrales de activación e importancia de mensajes

*LM* considera la propagación de tres clases distintas de mensajes, cada una de ellas con un propósito particular: mensajes de influencia que apelan a los intereses y conveniencias personales; mensajes para influir que apelan a los sentimientos; y mensajes para informar. Para cada clase de mensaje un individuo tiene un umbral de activación particular, teniendo en total tres umbrales de activación: umbral de influencia no homofílica, umbral de influencia homofílica, y umbral de información.

**Umbral de influencia no homofílica.** La primera clase de mensajes tiene la intención de influir en las personas apelando a intereses y conveniencias personales. Ejemplos de mensajes de esta clase son: consejos para la economía del hogar, recomendaciones de salud, de viaje, consejos para el manejo de una tarjeta de crédito, seguridad laboral, conductores de vehículos y peatones, etc. Las personas analizan estos mensajes de manera racional y objetiva, toman precauciones y buscan el consejo de otras personas. En esta clase de influencia, no aceptar la influencia podría significar riesgos. Esta clase de mensajes despierta en las personas un cierto umbral de influencia ( $Unh_{ji}$ ).

Con esta clase de mensajes, un individuo es más propenso a ser influenciado por personas que considera que tienen mayor relevancia que él. También, en esta clase de influencia, los mensajes que hacen referencia a situaciones más importantes tienen mayor probabilidad de influenciar a la gente. Estas consideraciones se cumplen en la siguiente ecuación matemática (8).

$$Unh_{ji} = 1 - (\alpha Rr_{ij} + \beta Ip_j) \quad (8)$$

Donde  $Rr_{ij}$  es la relevancia relativa de  $i$  con respecto a  $j$ ;  $Ip_j$  es la importancia del mensaje percibido por  $j$ ;  $\alpha$  y  $\beta$  son parámetros que dan pesos a las variables.

**Umbral de influencia homofílica.** Esta clase de mensajes trata de influir en las personas apelando a los sentimientos y las emociones, donde cuenta la afinidad, el afecto y la simpatía entre ellas. Ejemplos de mensajes de esta clase son un favor solicitado por alguien amado, el llamado a la solidaridad para alguien querido, la llamada a celebrar a alguien querido, etc. La

gente analiza estos mensajes con su corazón. En esta clase de influencia, la decisión tomada no significa riesgos. Esta clase de mensajes despierta en las personas un determinado umbral de influencia ( $U_{h_{ji}}$ ).

Cuanto mayor es la afinidad que siente el receptor  $j$  por el remitente  $i$  que envía un mensaje, o mayor es la importancia de dicho mensaje ( $I_{p_j}$ ), menor es el umbral de influencia del individuo  $j$  que recibe el mensaje, siendo más fácil influir en él. Estas consideraciones se cumplen en la siguiente ecuación matemática (9).

$$U_{H_{ji}} = 1 - (\alpha Ar_{ji} + \beta I_{p_j}) \quad (9)$$

Donde  $Ar_{ji}$  es la afinidad relativa que  $j$  siente por  $i$ .

Como se ha podido observar en las expresiones (8) y (9), un individuo no presenta el mismo valor de umbral de influencia a todos sus vecinos por igual, sino que éste depende de la relevancia o de la afinidad con cada vecino.

**Umbral de información.** Los mensajes no intentan influir, solo intentan informar. Ejemplos de mensajes de este tipo son escándalos, tragedias, eventos políticos o deportivos. Esta clase de mensajes despierta en las personas un cierto umbral de interés  $U_{i_j}$ . El nodo receptor  $j$  se activa si la noticia supera su umbral de interés. El umbral de interés  $U_{i_j}$  depende de la importancia del mensaje  $I_{p_j}$ . Un mayor interés en las noticias provoca un umbral más bajo  $U_{i_j}$  en el nodo  $j$ . Estas consideraciones se reflejan en la siguiente ecuación matemática (10).

$$U_{i_j} = 1 - I_{p_j} \quad (10)$$

**Relevancia relativa y afinidad relativa.** Cada individuo tiene una relevancia  $R_j$  dentro de su entorno social. La relevancia relativa  $R_{ij}$  usada en la ecuación (8) es la cuantificación de cuán relevante es el nodo  $i$  en comparación con el nodo  $j$ . Tanto  $R_{ij}$  como  $I_{p_j}$  deben ser valores dentro del rango  $[0, 1]$  para que  $U_{nh_{ji}}$  esté en este mismo rango. Por esta razón, la siguiente expresión que calcula la relevancia relativa de  $i$   $Rr_{ij} = \frac{R_i}{R_j}$  no nos ayuda. Una expresión que cumple los requisitos anteriores es la siguiente:

$$Rr_{ij} = \frac{R_i}{R_i + R_j} \quad (11)$$

No existe la posibilidad de que  $R_j = R_i = 0$ .

Los valores que pueden tomar  $Rr_{ij}$  están en el intervalo  $[0, 1]$ . Para cuantificar la relevancia de los nodos utilizamos la métrica de centralidad de *PageRank*, ya que es ampliamente utilizada en el análisis de redes sociales (Bekiari & Hassanagas, 2015; Yun & Gloor, 2015).

En la ecuación (9) usamos  $Ar_{ji}$ , que es el valor relativo de  $A_{ji}$ . De acuerdo con la Figura 7, definimos  $Ar_{ji}$  de la siguiente manera:

$$Ar_{ji} = \frac{A_{ji}}{\text{Max}_{\{j,k\} \in E} (A_{jk})} \quad (12)$$

**Importancia de un mensaje.** En general, las personas no perciben la importancia de un mensaje de la misma manera; Algunas personas dan más importancia y otras dan menos



importancia al mismo mensaje. Por ejemplo, si las personas, en general, sienten que un mensaje es importante, un individuo en particular podría percibirlo como moderadamente importante, mientras que otro lo encontrará muy importante.

En las ecuaciones (8)-(10) el término importancia del mensaje ( $I_{p_j}$ ) está presente. La importancia es percibida por un individuo particular  $j$ . Para tratar la variabilidad en la percepción de la importancia de un tema dado, relacionamos  $I_{p_j}$  con  $I_s$  en la siguiente ecuación (13), donde  $I_s$  es un parámetro del sistema que representa la importancia de un mensaje para la gente promedio. Definimos una escala de cinco valores cualitativos con sus valores cuantitativos correspondientes para  $I_s$ : *muy poco importante* (0.2), *poco importante* (0.4), *moderadamente importante* (0.6), *importante* (0.8), y *muy importante* (1.0).

$$I_{p_j} = I_s + \text{random}(\epsilon) \quad (13)$$

$\epsilon$  es un conjunto de valores que depende del parámetro  $I_s$ , como se ve en la Tabla 1. Para calcular  $I_{p_j}$ , se toma al azar un valor del conjunto  $\epsilon$  correspondiente a  $I_s$  y se suma a  $I_s$ .

Tabla 1. Relación entre  $I_s$  y el conjunto  $\epsilon$

Parameter $I_s$	0.2	0.4, 0.6, 0.8	1.0
Set $\epsilon$	{0.0, 0.2}	{-0.2, 0.0, 0.2}	{-0.2, 0.0}

Para aclarar cómo se cuantifica la importancia percibida por un individuo en particular  $j$   $I_{p_j}$ , damos el siguiente ejemplo: Supongamos que la importancia de un mensaje para el promedio de la gente  $I_s$  es “importante”, es decir su valor cuantitativo es 0.8, entonces  $I_{p_j}$  podría tener uno los valores que se muestra en Tabla 2.

Tabla 2. Cuantificación de la importancia del mensaje

$I_s$	Valor aleatorio $\epsilon$	$I_{p_j} = I_s + \text{random}(\epsilon)$	Importancia cualitativa del mensaje
0.8	-0.2	0.6	Moderadamente importante
	0	0.8	Importante
	0.2	1.0	Muy importante

En resumen, el alcance de la propagación de un mensaje en una red social depende de *tie strength* entre nodos y de sus umbrales de activación. *Tie strength* y los umbrales de activación dependen, a su vez, de la afinidad y relevancia de los nodos, y de la clase e importancia que tiene el mensaje.

### 3.4 Ejemplo: cálculo de *tie strength* y umbrales de activación

En la Tabla 3 mostramos un ejemplo de cálculo de  $w_{ji}$  para algunos enlaces del grafo de la Figura 8. Para hacer el ejemplo más explícito, la métrica de centralidad  $R_i$  es el grado de los nodos.

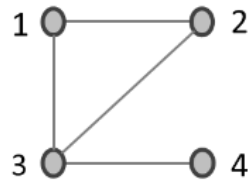


Figura 8. Grafo usado en el ejemplo

En la Tabla 4 mostramos un ejemplo de cálculo de los tres tipos de umbrales de activación para algunos nodos de la Figura 8. Grafo usado en el ejemplo con respecto a cada vecino. Los parámetros  $\alpha = \beta = 0.5$  y la importancia percibida del mensaje  $I p_j = 0.2$ .

Tabla 3. Cálculo de  $w_{ji}$

j,i	$n_{ji}$	$A_{ji} = 1 + \frac{n_{ji}}{2}$	$A_{jk}$	$\sum A_{jk}$	$w_{Aji} = \frac{A_{ji}}{\sum A_{jk}}$	$R_i$	$R_k$	$\sum R_k$	$w_{Rji} = \frac{R_i}{\sum R_k}$	$w_{ji} = w_{Aji} * w_{Rji}$
1,2	1	1.5	$A_{12}=1.5$ $A_{13}=1.5$	3	0.5	2	$R_2=2$ $R_3=2$	4	0.5	0.25
1,3	1	1.5	$A_{12}=1.5$ $A_{13}=1.5$	3	0.5	3	$R_2=2$ $R_3=2$	4	0.75	0.375
3,1	1	1.5	$A_{31}=1.5$ $A_{32}=1.5$ $A_{34}=1$	4	3.75	2	$R_1=2$ $R_2=2$ $R_4=1$	5	0.4	0.15
3,4	0	1	$A_{31}=1.5$ $A_{32}=1.5$ $A_{34}=1$	4	0.25	1	$R_1=2$ $R_2=2$ $R_4=1$	5	0.2	0.05
4,3	0	1	$A_{43}=1$	1	1	3	$R_3=3$	3	1	1

En la Tabla 3:  $i, j$  enlace del nodo  $i$  al nodo  $j$ ;  $\sum A_{jk}$  suma de las afinidades de  $j$  con sus vecinos;  $w_{Aji}$  fortaleza del enlace desde  $j$  hacia  $i$  debido a la afinidad entre nodos;  $R_i$  y  $R_k$  relevancia de nodos  $i$  y  $k$ ;  $\sum R_k$  suma de las relevancias de los vecinos de  $j$ ;  $w_{Rji}$  fortaleza del enlace desde  $j$  hacia  $i$  debido a la relevancia de nodos.

Tabla 4. Cálculo de los umbrales de activación

j,i	$R_j$	$R_i$	$Rr_{ij}$	$A_{ji}$	$A_{jk}$	$\max(A_{jk})$	$Ar_{ji}$	$Unh_{ji} = 1 - (\alpha Rr_{ij} + \beta I p_j)$	$Uh_{ji} = 1 - (\alpha Ar_{ji} + \beta I p_j)$	$Ui_j = 1 - I p_j$
1,2	2	2	0.5	1.5	$A_{12}=1.5$ $A_{13}=1.5$	1.5	1	0.65	0.4	0.8
1,3	2	3	0.6	1.5	$A_{12}=1.5$ $A_{13}=1.5$	1.5	1	0.6	0.4	0.8
3,1	3	2	0.4	1.5	$A_{31}=1.5$ $A_{32}=1.5$ $A_{34}=1$	1.5	1	0.7	0.4	0.8
3,4	3	1	0.25	1	$A_{31}=1.5$ $A_{32}=1.5$ $A_{34}=1$	1.5	0.67	0.775	0.565	0.8
4,3	1	3	0.75	1	$A_{43}=1$	1	1	0.525	0.4	0.8

En la Tabla 4:  $i, j$  enlace del nodo  $i$  al nodo  $j$ ;  $R_i$  y  $R_j$  relevancias de  $i$  y  $j$  respectivamente;  $Rr_{ij}$  relevancia relativa de  $i$  respecto de  $j$ ;  $A_{ij}$  afinidad entre  $i$  y  $j$ ;  $\max(A_{jk})$  máxima afinidad entre  $j$  y  $k$ ;  $Ar_{ji}$  afinidad relativa de  $j$  hacia  $i$ ;  $Unh_{ji}$ ,  $Uh_{ji}$ ,  $Ui_j$  umbrales de activación de nodos no homofilico, homofilico y de información respectivamente.

### 3.5 Búsqueda de nodos activos y activación de nodos

Una vez determinados la fortaleza de los enlaces del grafo (*tie strength*) y los umbrales de activación de cada nodo, *Lucy Model* procede a realizar dos procesos: recorrer el grafo en busca de nodos activos habilitados para activar, para luego, con cada uno de estos nodos, intentar activar a sus vecinos. Estos dos procesos se realizan para las tres clases de propagación de mensajes.

*Lucy Model* realiza sucesivas iteraciones, visitando en cada una de ellas a todos los nodos del grafo de forma secuencial, en busca de nodos en estado "activo, habilitado para activar" (estado 2). Estos nodos solo tienen una oportunidad para intentar activar a sus vecinos inactivos (estado 1), después de lo cual los nodos en estado 2 pasan a ser nodos "activos, inhabilitados para activar" (estado 3). Los nodos inactivos (estado 1) que logran ser activados, pasan a ser nodos activos, habilitados para activar (estado 2). Este proceso es el mismo para las tres clases de propagación de mensajes (no homofílica, homofílica, e informativa).

En la Figura 9 mostramos un ejemplo del proceso de propagación de activación de nodos en un pequeño grafo. Inicialmente existen dos semillas: los nodos 1 y 4 (Figura 9(a)), los otros nodos están en estado inactivo. El recorrido del grafo inicia en el nodo 1. En la Figura 9(b) el nodo 1, luego de intentar activar a sus vecinos, puede activar solo al nodo 2. El nodo 1 pasa a estado activo, deshabilitado para activar, y el nodo 2 pasa a estado activo, habilitado para activar. En la Figura 9(c) el nodo 2 intenta activar a sus vecinos inactivos, logrando activar solo al nodo 6. El nodo 2 pasa a estado activo, deshabilitado para activar y el nodo 6 pasa a estado activo, habilitado para activar. El siguiente nodo activo, habilitado para activar es el nodo 4. En la Figura 9(d) el nodo 4 pasa a estado activo, deshabilitado para activar, luego de no poder activar a ninguno de sus vecinos. En la Figura 9(e) se inicia el segundo recorrido del grafo, comenzando con el primer nodo activo, habilitado para activar que es el nodo 3, pero este no logra activar a su vecino el nodo 5. En la Figura 9(f) el nodo 3 pasa a estado activo, deshabilitado para activar. Como ya no hay más nodos en estado activo, habilitados para activar, el proceso termina.

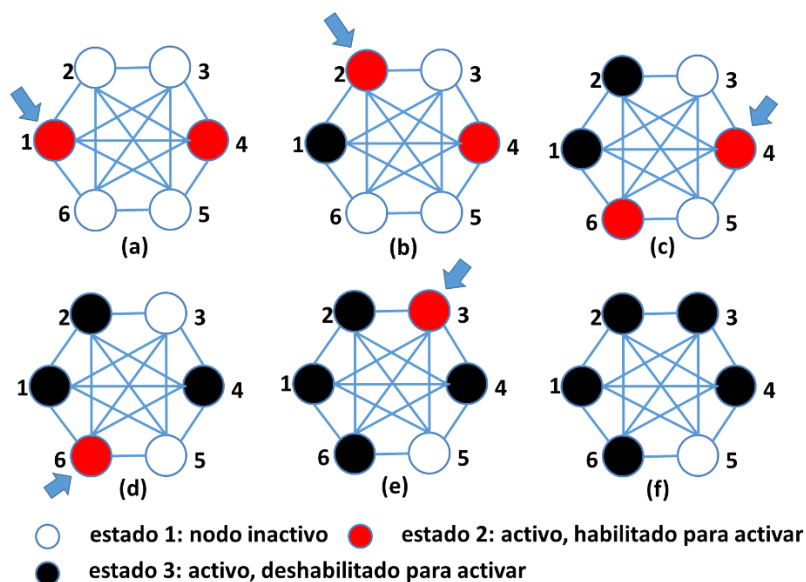


Figura 9. Proceso de propagación de activación de nodos

---

#### Algoritmo 1. Búsqueda de nodos activos habilitados para activar

---

**Input:** grafo  
**Output:** semillas (nodos activados)  
**Begin**  
  **Repeat**  
    nuevoCiclo  $\leftarrow$  0  
    **for**  $i = 1$  **to**  $i = n$  **do**  
      **if**  $i$  está habilitado para activar: ( $estado[i] == 2$ ) **then**  
         $i$  es deshabilitado para activar:  $estado[i] \leftarrow 3$   
        **for**  $j = 1$  **to**  $j = n$  **do**  
          **if**  $j$  está inactivo ( $estado[j] == 1$ ) **then**  
            Algoritmo 2: Envía mensajes y active nodos  
          **end if**  
        **end for j**  
      **end if**  
    **end for i**  
  **Until** nuevoCiclo == 0  
**End.**

---

El proceso de activación de nodos de *Lucy Model* se resume a manera de pseudocódigo en el Algoritmo 2.

---

#### Algoritmo 2. Envío de mensajes y activación de nodos

---

**Begin**  
  Calcula la afinidad en cada enlace:  $A_{ij}$   
  Normaliza la afinidad:  $w_{Aji}$   
  Calcula la relevancia de los nodos:  $R_i \leftarrow$  centralidad[ $i$ ]  
  Normaliza la relevancia:  $w_{Rji}$   
  Calcula tie strength:  $w_{ji} \leftarrow w_{Aji} * w_{Rji}$   
  **if** el nodo  $j$  no recibe el mensaje del nodo  $i$  **then end**  
  Lee la importancia del mensaje  $I_s$   
  Calcula la relevancia percibida para cada nodo:  $I_{pj}$   
  Calcula la relevancia relative del nodo remitente  $i$ :  $R_{Rij}$   
  Calcula la afinidad y relevancia relativas  $j$ :  $A_{Rji}$   
  **if** el tipo de propagación es influencia no homofílica **then**  
    Calcula el umbral de influencia:  $U_{NHji}$   
  **end if**  
  **if** el tipo propagación es información **then**  
    Calcula el umbral de información:  $U_{Ij}$   
  **end if**  
  **if** el tipo de propagación es influencia homofílica **then**  
    Calcula el umbral de influencia:  $U_{Hji}$   
  **end if**  
  **if** el valor aleatorio[0, 1]  $\geq$  umbral **then**  
    el nodo  $j$  es activado y habilitado para activar:  $estado[j] \leftarrow 2$   
    Al final de a presente iteración, una nueva iniciará: nuevoCiclo  $\leftarrow 1$   
  **end if**  
**End.**

### 3.6 *Lucy Model* en bloques

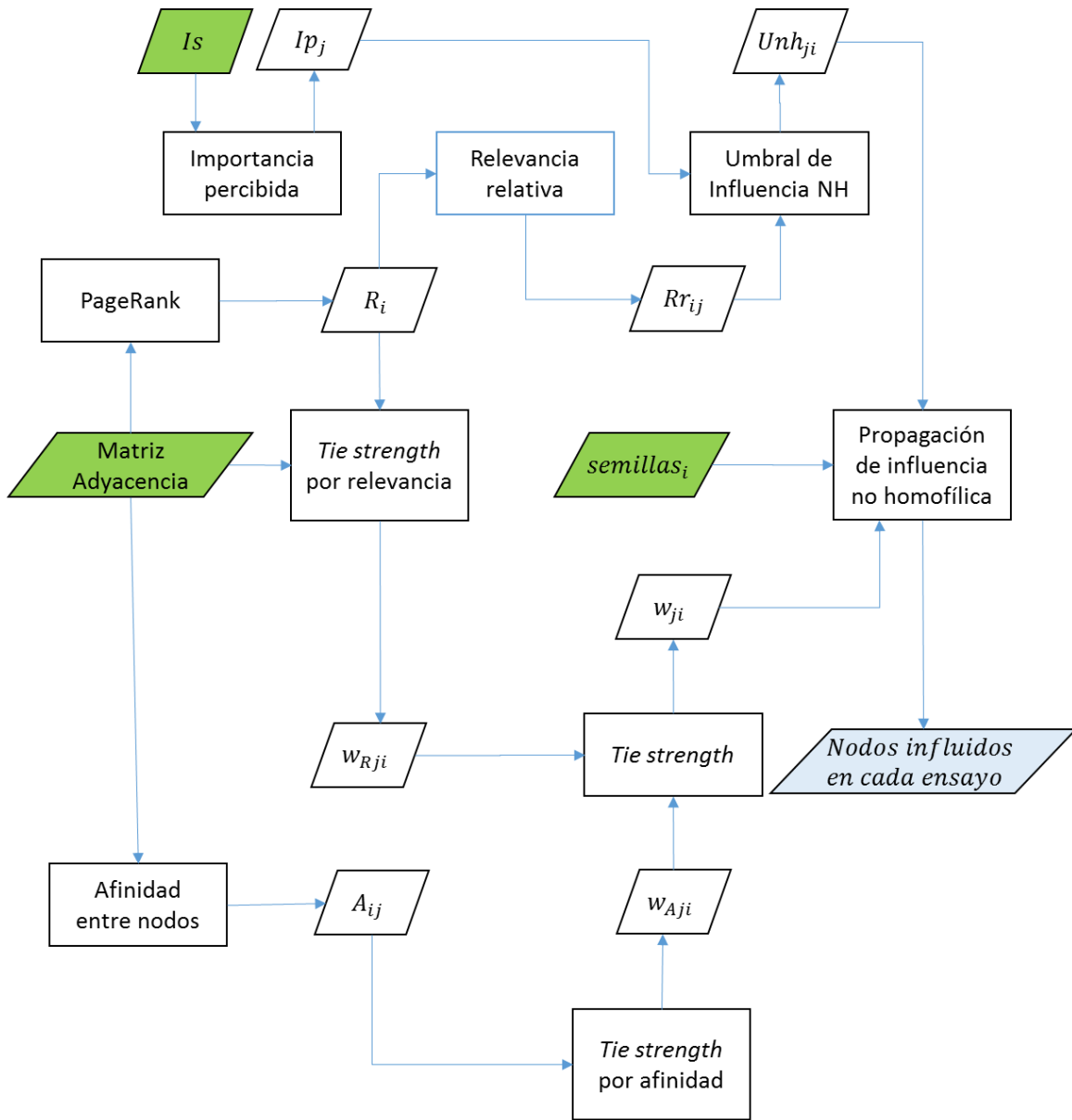
En los siguientes tres diagramas de flujo explicamos, a un mayor nivel de abstracción, la secuencia de los procesos que *Lucy Model* realiza para activar nodos con cada clase de mensaje que se propaga por una red social. En estas figuras se observa que algunos procesos se pueden realizar en forma paralela. Estas figuras indican con claridad, utilizando diferentes formas y colores, los datos de entrada, los procesos parciales, los resultados intermedios y el resultado final.

### Capítulo 3. Descripción de *Lucy Model*

En las tres figuras se describe el cálculo de la fortaleza de enlaces  $w_{ji}$  en función de la fortaleza de enlaces debido a la afinidad  $w_{Aji}$  y de la fortaleza de enlaces debido a la relevancia  $w_{Rji}$ .

Los procesos que se describen en las figuras *Figura 10* y *Figura 11* son casi idénticos. La única diferencia en estos dos diagramas es que en la *Figura 10* se calcula el umbral de activación de nodos para mensajes no homofílicos  $Unh_{ji}$  en función de la relevancia relativa  $Rr_{ij}$ , mientras que en la *Figura 11* el umbral de activación de nodos para mensajes homofílicos  $Uh_{ji}$  se calcula en función de la afinidad relativa  $Ar_{ji}$ .

La *Figura 12* describe los procesos que se deben realizar para activar los nodos de un grafo cuando se propaga un mensaje con fines informativos. Para activar nodos con esta clase de mensajes no se necesita cuantificar la relevancia relativa  $Rr_{ij}$  ni la afinidad relativa  $Ar_{ji}$ . El umbral de información se calcula solamente en función de la importancia percibida del mensaje  $Ip_j$ .



Simbología:

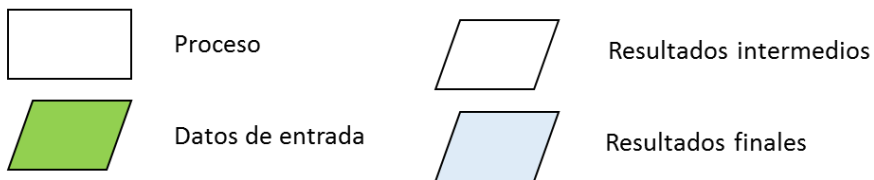
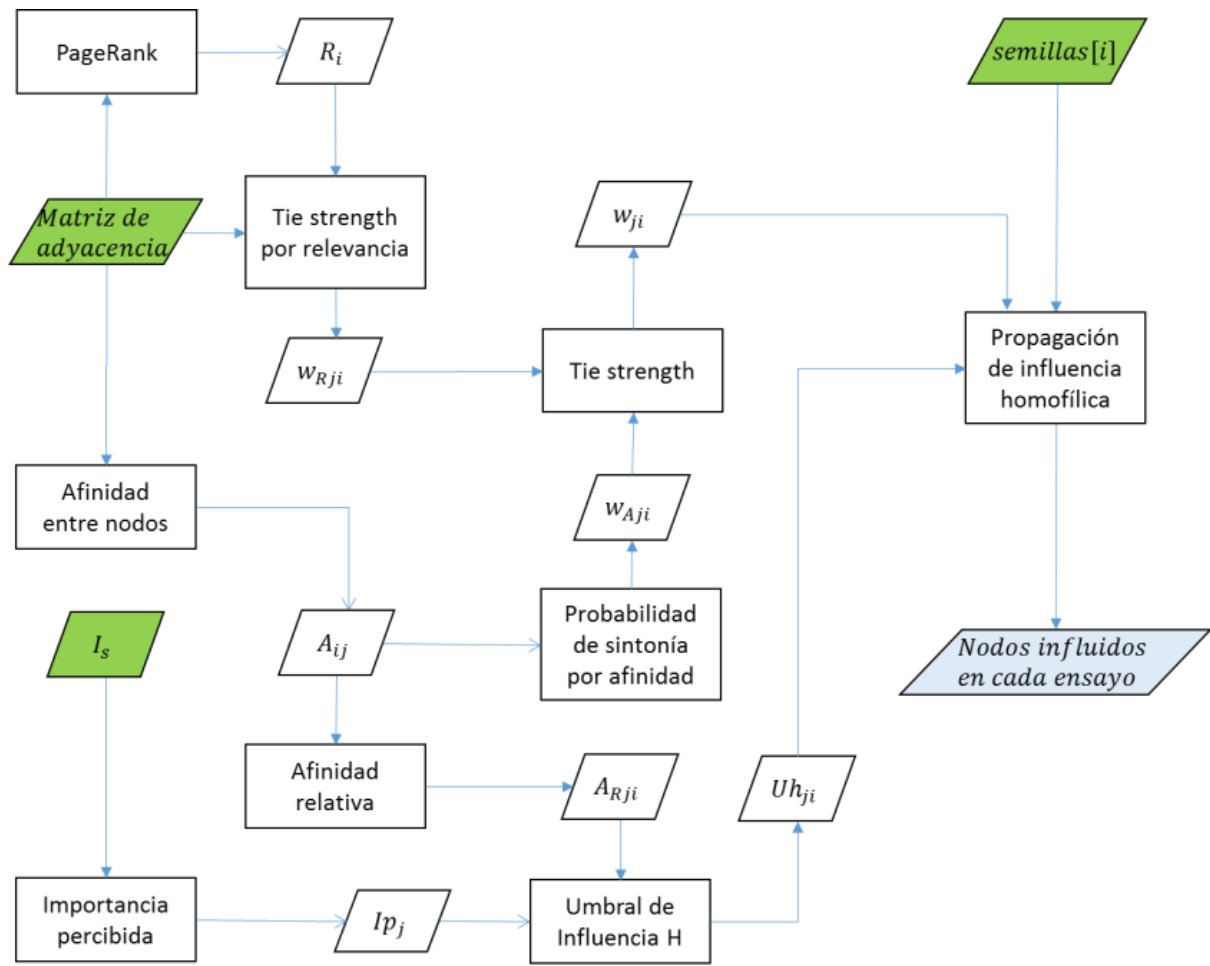


Figura 10. Propagación de mensaje no homofílico



Simbología:

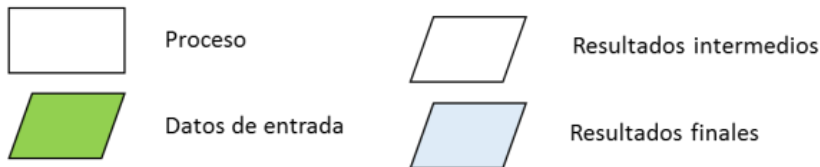


Figura 11. Propagación de mensaje homofílico

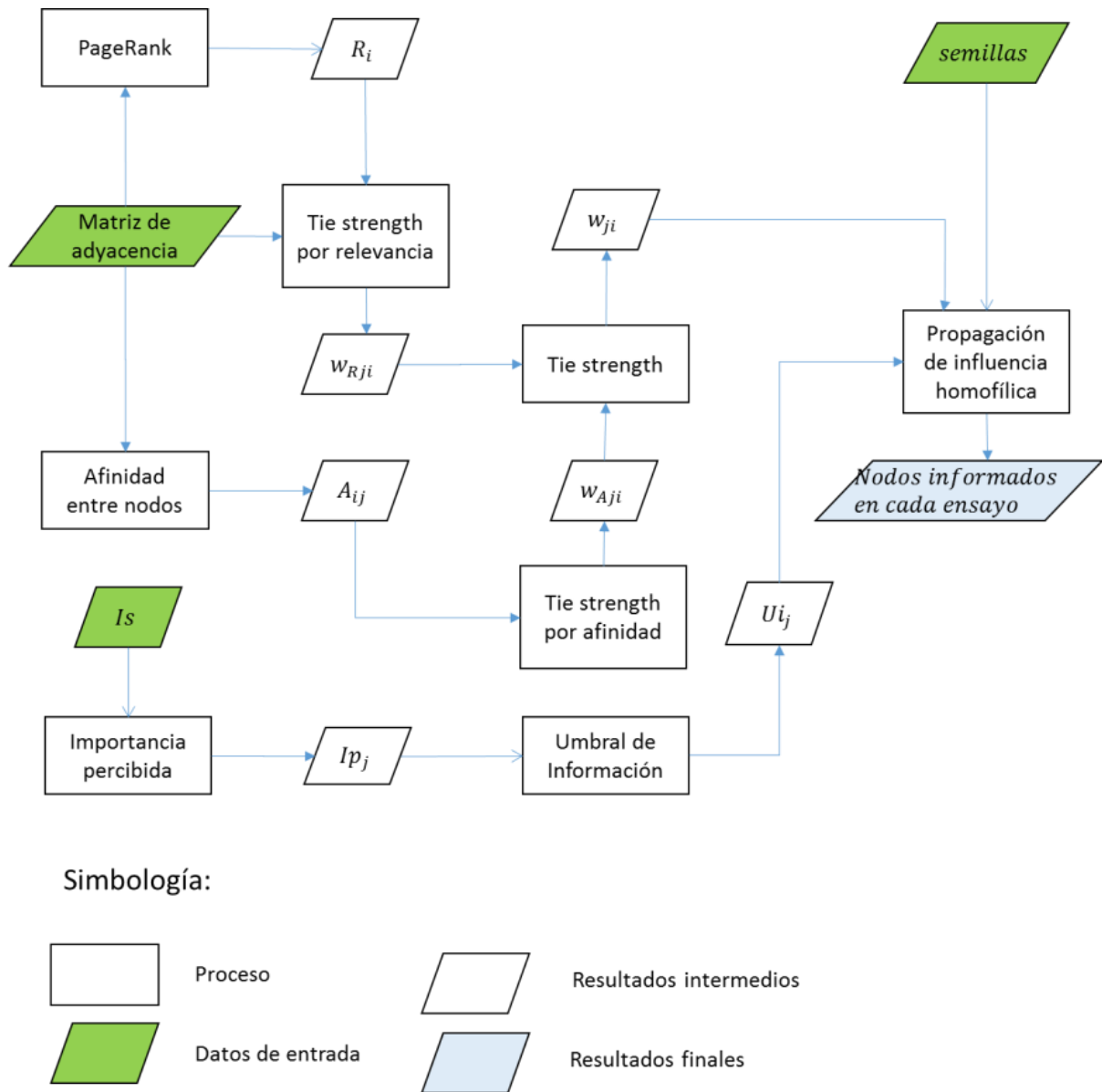


Figura 12. Propagación de mensaje informativo

### 3.7 Resumen del capítulo

En este capítulo identificamos los elementos que afectan el alcance de la propagación de mensajes en una red social, los mismos que se recogen en el modelo de *Lucy Model*. Luego de identificarlos, determinamos la manera de cuantificar dichos elementos, los mismos que son los siguientes:

1. Clase de mensaje y los umbrales de influencia y de información de los individuos.
2. Importancia del mensaje.
3. Intensidad de las relaciones entre las personas a través de las cuales se propaga un mensaje.



*Lucy Model* considera tres clases de mensajes. Cada clase de mensaje trata de influir a los individuos de una manera diferente. Estas clases son:

1. Clase de mensaje que trata de influir apelando a los sentimientos de afinidad y afecto.
2. Clase de mensaje que trata de influir apelando a los intereses y conveniencias de los individuos.
3. Clase de mensaje que simplemente trata de informar.

Con cada clase de mensaje recibido, los individuos reaccionan de modo diferente, esta reacción se manifiesta en el umbral de influencia de cada individuo, es decir, hay un umbral de influencia para cada clase de mensaje. Nosotros hemos determinado una manera de cuantificar los umbrales de influencia por cada clase de mensaje y para cada individuo.

Un mismo mensaje, independientemente de la clase a la que pertenece, no tienen el mismo grado de importancia para todos los individuos. Nosotros establecemos su valor con base en el grado de importancia que tienen el mensaje para la gente en general, el cual se determina estadísticamente, y de una variable aleatoria que se aplica a cada individuo.

Las relaciones entre las personas o *tie-strength* dentro de un círculo social se basan en la combinación de dos motivaciones:

1. Sentimientos de afecto, afinidad y amistad.
2. Razones de interés y conveniencia personal.

La intensidad de las relaciones entre las personas depende la intensidad de estas dos motivaciones. *Lucy Model* determina una manera de cuantificar la intensidad de las relaciones entre las personas cuantificando precisamente las dos motivaciones indicadas. Nuestro modelo determina la intensidad de cada *tie-strength* con base en la topología de la red social. Para cuantificar el sentimiento de afecto, afinidad y amistad nosotros hacemos una analogía entre una red social y una red eléctrica. Para cuantificar la relación por razones de interés y conveniencia personal, nos basamos en la relevancia de los individuos dentro de la red social, utilizando medidas de centralidad.

## Capítulo 4

### 4 Método de validación del modelo *Lucy Model*

El criterio utilizado por los modelos de referencia (*LTM*, *ICM*), *WCM* y nuestro modelo (*LM*) para predecir el rango de propagación de mensajes es el número de nodos activados. De esta forma, hemos podido determinar la precisión de los resultados de nuestro modelo comparándolos con los de los otros autores. Para demostrar la validez de *Lucy Model*, realizamos los siguientes pasos.

#### 4.1 Objetivos de validación del diseño.

Las pruebas que hemos llevado a cabo tienen tres objetivos: Primero, demostrar de manera concluyente que las predicciones de *Lucy Model* son notablemente más precisas y exactas que las predicciones de los modelos de referencia y *WCM*. En segundo lugar, mostrar la flexibilidad de *Lucy Model* al modificar sus parámetros para adaptarse a diferentes situaciones. En tercer lugar, mostrar el comportamiento de *Lucy Model* en tres tipos diferentes de redes: una red sintética *Power-Law* (Newman, 2010), una red social de *YouTube* y una red de *Facebook* rastreada.

#### 4.2 Diseño de la validación

La herramienta que utilizamos para generar la red sintética fue *GenRndPowerLaw* del *Stanford Network Analysis Project SNAP* (Leskovec & Sosič, 2016). La red social de *YouTube* también se obtuvo de *SNAP*.

*Power-Law* (Newman, 2010) es una relación matemática utilizada por algoritmos que generan redes sintéticas. Las redes de *Power-Law* son una clase de redes aleatorias porque los enlaces se agregan aleatoriamente a un conjunto estático de nodos. La distribución de grados obedece la ley de potencia  $p(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$ , donde  $p(k)$  es la probabilidad de que un nodo tenga grado  $k$ ,  $\zeta(\gamma)$  es la *función Zeta de Riemann* definida como  $\zeta(\gamma) = 1 + \frac{1}{2^\gamma} + \frac{1}{3^\gamma} + \dots$ , y  $\gamma > 1$  es el exponente que toma valores entre 2 y 3 para las redes sociales. Cuando ocurre  $2 \leq \gamma \leq 3$ , estas redes se llaman redes sin escala porque hay una alta probabilidad de que los nodos con alto grado estén conectados a otros nodos con un alto grado (Li et al., 2005), característica que es típica de las redes sociales.

Elegimos usar una red *Power-Law* porque muchas redes sociales existentes son de este tipo, por ejemplo: redes de amigos (Hein et al., 2006), redes de llamadas telefónicas ( $\gamma = 2.1$ ), redes de correo electrónico ( $\gamma = 2.0$ ), *World Wide Web* ( $\gamma = 2.1$ ) (Newman, 2010) y redes sociales online (Duong-Ba, Thuan Hong, 2014).

Para explicar cómo se genera un grafo *Power-Law* presentamos el siguiente ejemplo. Suponemos que tenemos un grafo de 5 nodos y  $\gamma = 2.0$ , por lo tanto  $\zeta(\gamma = 2.0) = 1.644$ . Utilizamos la ley de potencia para el grado  $k = 1, 2, 3, y 4$ :  $p(1) = 0.608$ ;  $p(2) = 0.152$ ;  $p(3) = 0.067$ ;  $p(4) = 0.038$ . Generamos un número aleatorio  $r$  en el intervalo cerrado  $[0, 1]$ . El grado de un nodo se determina, en principio, según la Tabla 5.

Tabla 5. Asignación preliminar de grado  $k$  a cada nodo

$r$	$0 \leq r \leq 0.038$	$0.038 < r \leq 0.067$	$0.067 < r \leq 0.152$	$0.152 < r \leq 0.608$	$0.608 < r \leq 1.0$
$k'$	4	3	2	1	0

Tabla 6. Matriz de adyacencia del grafo

nodo	1	2	3	4	5	$r$	$k'$	$k$
1	X	1	1	X	X	0.10	2	2
2	1	X	1	1	X	0.42	3	3
3	1	1	X	X	X	0.47	1	2
4	X	1	X	X	X	0.72	0	1
5	X	X	X	X	X		0	0

Al comienzo la Tabla 6 está vacía. Iniciamos con la fila del nodo 1 generando un número aleatorio  $r = 0.10$ , que corresponde a un grado  $k' = 2$ . En esta fila, marcamos con 1 en las celdas los dos nodos 2 y 3, y los siguientes nodos marcamos con X. Copiamos la fila 1 en la columna 1. Continuamos con la fila del nodo 2 generando  $r = 0.42$ , que corresponde a  $k' = 3$ . Como el nodo 2 ya tiene un enlace, en la fila 2 solo marcamos con 1 las celdas de los nodos 3 y 4, y la celda del nodo 5 marcamos con X. Copiamos la fila 2 en la columna 2. Continuamos con el nodo 3, generamos  $r = 0.47$  que corresponde a  $k' = 1$ , pero el nodo 3 ya tiene dos enlaces, entonces solo marcamos con X las celdas restantes de la fila 3. Copiamos la fila 3 en la columna 3. Continuamos con la fila del nodo 4 generando  $r = 0.72$ , que corresponde a  $k' = 0$ , pero el nodo 4 ya tiene un enlace, entonces solo marcamos con X las celdas restantes de la fila 4. Copiamos la fila 4 en la columna 4. En la última fila ya no es necesario hacer nada, en este caso, el nodo 5 no tiene enlaces. El grado de cada nodo se indica en la última columna de la Tabla 6.

En la Tabla 7, mostramos las características de los grafos que estamos utilizando.

Tabla 7. Características de los grafos utilizados

Grafo	Número de nodos	Número de enlaces	Grado máximo	Grado medio
<i>Facebook</i>	968,810	2,742,523	4,804	5.66
<i>YouTube</i>	1,134,890	2,987,624	28,754	5.34
<i>Power-Law</i> exponente $\gamma=2.1$	1,000,000	3,382,218	143,309	6.76

A continuación, elaboramos los protocolos de validación para validar y presentar los resultados.

### 4.3 Protocolo para la extracción de datos empíricos

Para validar los resultados de *Lucy Model*, extrajimos datos empíricos de propagación de mensajes en la red social de *Facebook* de un usuario anónimo, que tiene un perfil personal. Para este propósito, difundimos mensajes a través de diferentes usuarios semilla. Cada mensaje publicado incluía un hipervínculo a una página de un servidor web. De esta manera, el usuario de *Facebook* que recibió el mensaje, al hacer *click* en el hipervínculo, podría acceder a la información completa relacionada con ese mensaje. El número de personas que accedieron al servidor web lo contamos como personas activadas.

Los usuarios *semilla* fueron veintitrés estudiantes de la asignatura de *Matemática Discreta* del período marzo-julio de 2018 de la Facultad de Ingeniería de la Universidad de Cuenca-Ecuador.

Se diseñaron y publicaron tres clases de mensajes: un mensaje no homofílico, un mensaje homofílico y un mensaje informativo. Específicamente, los mensajes fueron: espionaje en *WhatsApp*, abandono de mascotas, y racismo, respectivamente. Cada semilla publicó cada clase de mensaje por una sola vez, programada de la siguiente manera: Día 1: semilla 1. Día 2: semilla 2. Día 3: semilla 3. Día 4: semilla 4. Día 5: semilla 5. Día 6: semillas 6, 7, 8, 9 y 10. Día 7: semillas 11, 12, 13, 14 y 15. Día 8: semillas 16, 17, 18, 19, 20, 21, 22 y 23. La participación de las semillas fue en orden descendente según su número de amigos en *Facebook*. Programamos la publicación de mensajes de esta manera para conocer la contribución de cada semilla o grupo de semillas en el número total de nodos activados. Tuvimos el mismo propósito cuando realizamos las pruebas en ordenador con los modelos *LM*, *LTM*, *ICM* y *WCM*. Por otro lado, tratamos de asegurarnos de que la publicación de cada mensaje se realice en el menor tiempo posible para que los mensajes no pierdan relevancia e interés en los nodos que se activarán. La publicación de cada clase de mensaje se hizo en un lapso de ocho días. La publicación de las tres clases de mensajes duró diez días, y tuvo lugar entre el 25 de junio y el 4 de julio del año 2018. El 25 de junio se comenzó a publicar el mensaje no homofílico, el 26 de junio se comenzó a publicar el mensaje homofílico y el 27 de junio se comenzó a publicar el mensaje informativo.

### 4.4 Protocolo para las pruebas en ordenador

La simulación para predecir el alcance de la propagación de mensajes con los respectivos modelos, se realizó en la red de *Facebook* rastreada del mismo usuario anónimo utilizado en los experimentos de campo. Tuvimos que rastrearlo hasta tres saltos del usuario anónimo porque los mensajes propagan hasta dos saltos de una semilla (Afrasiabi Rad & Benyoucef, 2012; Zuo et al., 2016). El seguimiento en varios niveles de profundidad no está permitido por las *API* oficiales de *Facebook*, debido a sus políticas de seguridad. Por este motivo utilizamos el *scraping script* para *Facebook* *FBS* (González Toral, Santiago 2018). Este *script* utiliza el *Front-end* de *Facebook*, está desarrollado en el lenguaje *Python*, permite el seguimiento automático y tiene en cuenta que *Facebook* podría bloquear la cuenta del usuario.

Web scraping o “escarbar una web”, es una técnica utilizada para extraer información de sitios web mediante programas de software. Usualmente, estos programas simulan la navegación de un humano en la World Wide Web, ya sea utilizando el protocolo HTTP manualmente, o incrustando un navegador en una aplicación. El script que nosotros hemos utilizado simula a un usuario de Facebook, el cual es una aplicación web, para que recursivamente ingrese a su perfil de usuario y extraiga su lista de amigos, luego ingresa a cada perfil de sus amigos para, a su vez, extraer la lista de amigos de sus amigos, y así sucesivamente. De esta manera se construye la red social del usuario inicial.

El proceso de rastreo de la red fue muy lento ya que se realizó en veintisiete días desde el 28 de mayo hasta el 23 de junio del año 2018.

Las predicciones hechas por los cuatro modelos *LM*, *LTM*, *ICM* y *WCM* dependen de elementos probabilísticos. En el caso de *LM*, estos elementos son tres:

1. Probabilidad de que los nodos reciban el mensaje [Ec. 1].
2. Probabilidad de exceder el umbral de activación de cada nodo receptor (8)-(10)
3. Estimar la importancia del mensaje percibido por cada nodo receptor [Ec. 13].

Hemos usado las mismas 23 semillas y las hemos organizado de la misma manera que lo hicimos en los experimentos de campo, es decir, de manera descendente por el número de vecinos o grado. Formamos los siguientes conjuntos de nodos:  $A = \{1\}$ ,  $B = \{1,2\}$ ,  $C = \{1,2,3\}$ ,  $D = \{1,2, \dots, 4\}$ ,  $E = \{1, 2, \dots, 5\}$ ,  $F = \{1,2, \dots, 10\}$ ,  $G = \{1,2, \dots, 15\}$  y  $H = \{1,2, \dots, 23\}$ . Lo hicimos de esta manera para determinar cuánto contribuye cada semilla o grupo de semillas en la activación de los nodos.

Aplicamos el *Método Montecarlo* (Bolthausen & Wüthrich, 2013) para predecir el número de nodos activados, ejecutando mil veces cada uno de los cuatro modelos (*LM*, *LTM*, *ICM*, *WCM*), y para cada grupo de semillas, comenzando con el grupo A. De esta manera, los resultados convergen a un valor promedio del número de nodos activados. En cada ejecución generamos uniformemente números aleatorios en el intervalo  $[0, 1]$ , con el propósito de determinar si el mensaje es recibido por cada nodo ( $w_{ji}$ ), la importancia del mensaje percibida por cada nodo ( $Ip$ ), y si el mensaje supera el respectivo umbral de activación ( $Unh_{ji}$ ,  $Uh_{ji}$ ,  $Ui_j$ ).

El *Método Montecarlo* se basa en el *Teorema de Jakob Bernoulli*, también llamado *Ley de grandes números*. El teorema establece que si  $X_1, X_2, X_3 \dots$  es una secuencia infinita de variables aleatorias independientes que tienen el mismo valor esperado  $\mu$  y desviación estándar  $\sigma$ , entonces  $\bar{X}_n = (X_1 + \dots + X_n)/n$  converge en probabilidad a  $\mu$ . Es decir, para cualquier número positivo  $\epsilon$  tenemos:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

Este es un método de investigación estocástica numérica que permite obtener soluciones aproximadas a problemas complejos, deterministas y probabilísticos, como es nuestro caso (Masayuki Yano et al., 2013).

En el caso de *Lucy Model*, este proceso se realizó con cada una de las tres clases de propagación de mensajes: informativo, no homofílico y homofílico. Hemos asignado a los tres mensajes el valor de 0.2 a  $Is$ , ya que estos mensajes tienen muy poca importancia para el

promedio de la gente, debido que los temas que tratan no son de actualidad y porque esos mensajes no fueron diseñados con criterios para campañas publicitarias. Por otro lado, hemos asignado pesos  $\alpha$  y  $\beta$  a las variables de los umbrales de activación, para así obtener valores máximos y mínimos de propagación de mensajes a través de las redes.

Los autores de *LTM* (D. Kempe et al. 2003), no especifican si un nodo que no ha logrado ser influenciado tiene una nueva oportunidad o no de ser influenciado, si se incrementa el número de sus vecinos activos. Nosotros hemos implementado *LTM* considerando las dos posibilidades. Con la primera posibilidad el número de nodos activados nos pareció exageradamente alto, razón por la cual decidimos elegir la segunda implementación.

(D. Kempe et al. 2003) aplican su modelo *LTM* a una red de coautores en donde la fortaleza de los enlaces ( $b_{v,w}$ ) se calcula en función del número de *papers* publicados como coautores. Ellos no explican cómo calcular la fortaleza de los enlaces en otros tipos de redes sociales. Para que los resultados de *LTM* y *LM* sean comparables en condiciones similares, hemos decidido que  $b_{v,w}$  de *LTM* es igual a  $w_{ji}$  [Eq. 1] de *LM*. En *ICM* fijamos  $p_{v,w} = 1\%$ .

Los modelos de propagación de mensajes se implementaron mediante simulación. Las herramientas utilizadas fueron: lenguaje de programación C / C ++, biblioteca de minería de grafos *SNAP*, analizador de red *NodeXL*, hoja de cálculo *Excel* y procesador PC i7 core, sistema operativo *Ubuntu* de 8 bits y 64 bits.

#### 4.5 Resumen el capítulo

Par demostrar la validez de nuestro modelo *LM* hemos realizado pruebas de dos tipos: pruebas en ordenador y pruebas empíricas, para luego poder comparar los resultados de laboratorio con los resultados empíricos. Estas pruebas tuvieron tres objetivos:

1. Demostrar de manera concluyente que las predicciones de *Lucy Model* son notablemente más precisas y exactas que las predicciones de modelos existentes importantes.
2. Mostrar la flexibilidad de *Lucy Model* que permite modificar sus parámetros para adaptarse a diferentes situaciones, como son: distintas clases de mensaje, múltiples niveles de importancia del mensaje, y poder asignar diferentes pesos a la afinidad entre nodos y la relevancia de ellos para la cuantificación de los umbrales de influencia y de información.
3. Mostrar el comportamiento de *Lucy Model* en distintos tipos de redes: una red sintética *Power-Law*, una red social de YouTube y una red Facebook rastreada de un usuario particular anónimo.

Las pruebas en ordenador consistieron en predecir el rango o alcance de propagación de mensajes, reflejado en el número de nodos activados (influenciados o informados según sea el caso). Para ello se utilizaron los modelos predictores *LM*, *LTM*, *ICM* y *WCM* en una red Facebook rastreada de un usuario particular anónimo. Hemos utilizado los modelos *Linear Threshold Model*, *Independent Cascade Model* y *Weighted Cascade Model* porque los dos primeros son considerados referentes, ya que ellos se están utilizando desde hace muchos años hasta la actualidad en innumerables trabajos de investigación, dirigidos a maximizar la propagación de información en redes sociales; y porque estos han servido de base para el desarrollo de otros modelos predictores como *Weighted Cascade Model*.

En algunas de las pruebas en ordenador usamos una red *Power-Law* porque muchas redes sociales existentes son de este tipo.

Los cuatro modelos predictores son estocásticos, por lo que los resultados de laboratorio pueden ser determinados aplicando el Método de Montecarlo, que a su vez se basa en el Teorema de Jakob Bernoulli, también llamado Ley de los grandes números. El método consiste en realizar muchas pruebas para que los resultados converjan a un valor promedio.

Las pruebas empíricas se realizaron utilizando la plataforma Facebook con la misma red rastreada de un usuario anónimo utilizada en las pruebas en ordenador. Esta red se rastreó utilizando un scraping script desarrollado en el lenguaje Python. Este script usa el Front-end de Facebook. El rastreo de esta red fue un proceso realmente lento ya que se lo pudo realizar en el lapso de veintisiete días, trabajando las 24 horas al día.

Los datos empíricos se obtuvieron difundiendo mensajes en la red social Facebook de un determinado usuario anónimo. La difusión de mensajes se inició con un determinado número de usuarios semilla. Cada mensaje publicado incluía un hipervínculo a una página de un servidor web. El usuario de Facebook que recibía el mensaje, al hacer click en el hipervínculo, podía acceder a la información completa relacionada con ese mensaje. El número de personas que accedieron al servidor web lo contamos como personas activadas.

## Capítulo 5

### 5 Resultados de validación y análisis

#### 5.1 Experimentos de campo

Fue indispensable realizar pruebas de campo para poder comparar las predicciones de *Lucy Model* y de los otros tres modelos existentes *LTM*, *ICM* y *WCM* con datos empíricos. De esta manera se pudo determinar qué modelo predictor se acerca más a los datos reales.

Basado en el protocolo de extracción de datos empíricos descrito en la sección anterior, mostramos en la Figura 13 los resultados de los experimentos de campo.

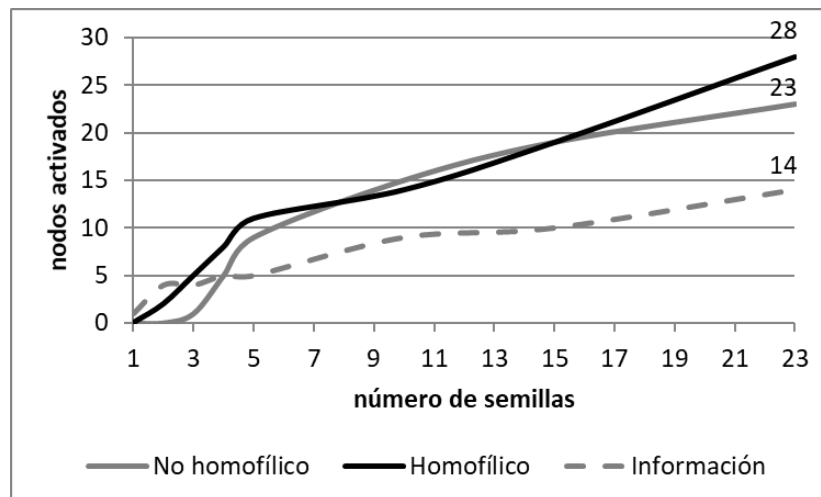


Figura 13. Red de Facebook rastreada

En las pruebas de campo realizadas en la plataforma *Facebook*, se determinó el número de personas que fueron influenciadas o informadas con un mensaje no homofílico, un mensaje homofílico y un mensaje informativo respectivamente. Estos resultados se presentan gráficamente en la *Figura 13* utilizando un sistema de ejes de coordenadas cartesianas. En el eje horizontal se representan las veintitrés semillas dispuestas en forma descendente según el número de amigos que tienen, y en el mismo orden que publicaron cada mensaje a lo largo de los ocho días que duró la publicación. Cada semilla publicó cada mensaje por una sola vez. En el eje vertical de cada gráfico se representa el número de nodos que fueron activados.

En los resultados de las pruebas empíricas notamos las siguientes dos características:



1. Existe un comportamiento similar en la propagación del mensaje homofílico y el mensaje no homofílico.
2. El número de nodos activados con el mensaje informativo es menor que el alcanzado por las otras dos clases de mensajes.

## 5.2 Pruebas en ordenador

Las figuras 11, 12 y 13 muestran el grado de exactitud alcanzado por las predicciones de laboratorio realizadas por *Lucy Model*, *LTM*, *ICM* y *WCM*, en comparación con la evidencia empírica. En el caso de *LM*, estas cifras muestran el número de nodos activados como resultado de la propagación de un mensaje no homofílico, homofílico e informativo, respectivamente. En estas figuras hay una banda gris dentro de la cual se encuentran todas las curvas de *Lucy Model*, etiquetadas como B. Cada curva corresponde a diferentes valores asignados a los parámetros  $(\alpha, \beta, I_s)$  de los umbrales de activación (8)-(10). Las curvas que delimitan las bandas grises se obtuvieron con  $\alpha = 0$  y  $\beta = 1$ . De esta forma,  $U_{nh_{ji}} = U_{h_{ji}} = U_{i_j} = 1 - I_{p_j}$ . Las curvas que marcan el límite superior e inferior de estas bandas se obtuvieron con  $I_s = 0.2$  e  $I_s = 1$ , respectivamente.

En estas figuras, observamos claramente que la banda de *LM* está mucho más cerca del resultado de los experimentos empíricos, en comparación con las predicciones del modelo de línea de base y *WCM*.

Los bordes inferior y superior de las bandas grises de las tres figuras siguientes marcan el número mínimo y máximo de nodos activados predichos para cada clase de mensaje. La curva de color negro que está dentro de cada banda gris indica el número de nodos predichos para  $I_s = 0.2$  y  $\alpha = 0.5$ .

Las pruebas en ordenador que realizamos con *LM* y los experimentos de campo dieron como resultado un número muy pequeño de nodos activados en comparación con los casi un millón de nodos de la red de *Facebook* rastreada y los millones de nodos de la plataforma de *Facebook*. Esto se debió al hecho de que, en ambas pruebas, teníamos disponible un conjunto de semillas con muy poca relevancia con respecto a otros nodos en la red.

En las figuras Figura 14, Figura 15 y

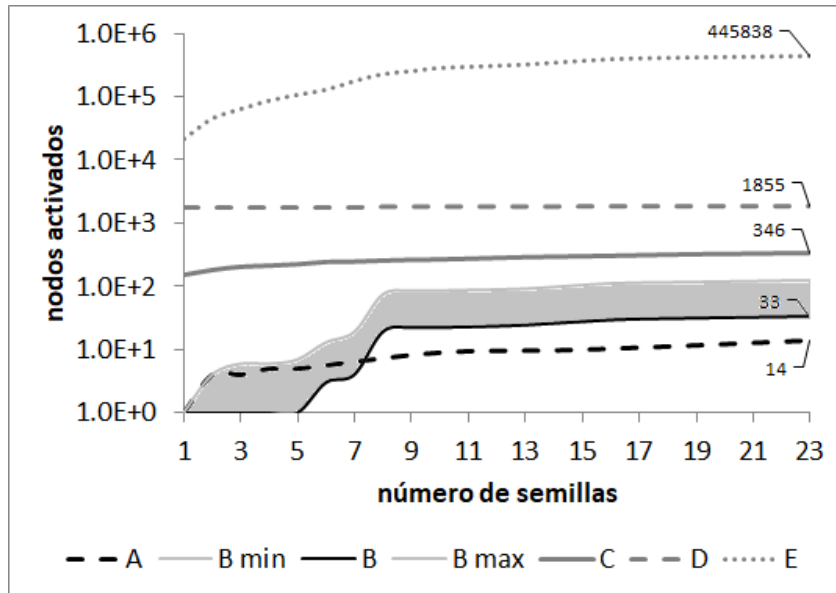


Figura 16, *LM* tiene con la semilla número 8 el mayor crecimiento en el número de nodos activados. Esto se debe a que los vecinos tienen una mayor fuerza de unión con la semilla 8 en comparación con las semillas anteriores. Por lo tanto, la probabilidad de que los vecinos reciban el mensaje y se activen es mayor.

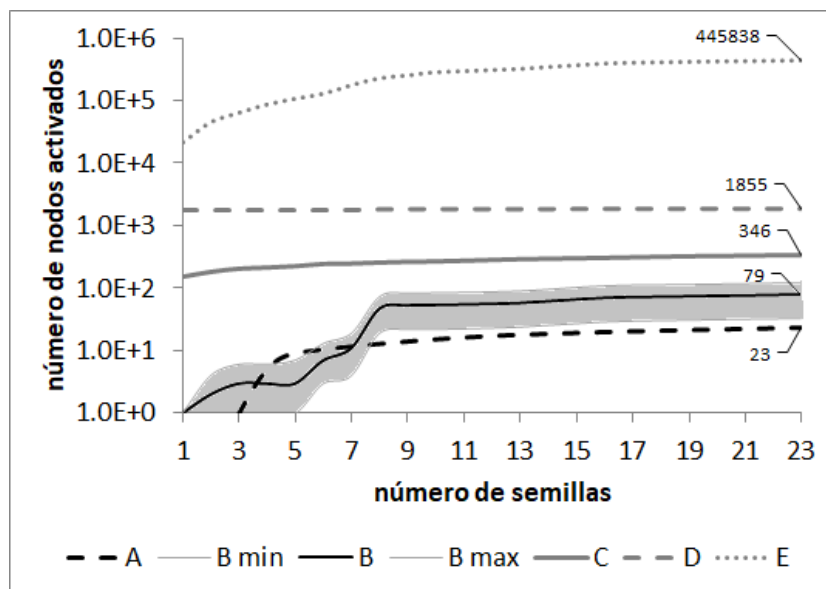


Figura 14. Exactitud: Propagación de un mensaje no homofílico.

En la figura anterior las curvas que se presentan tienen la siguiente descripción. A: datos empíricos. B mín: *LM* mín. B: *LM* ( $I_s = 0.2, \alpha = 0.5$ ). B máx: *LM* máx. C: *ICM* ( $p_{v,w} = 1\%$ ). D: *LTM*. E: *WCM*.

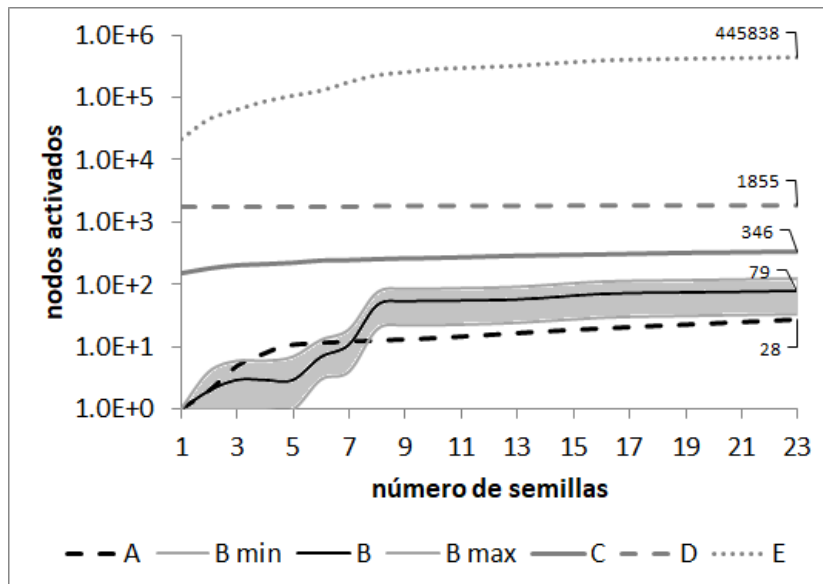


Figura 15. Exactitud: Propagación de mensaje homofílico.

En la figura anterior las curvas que se presentan tienen la siguiente descripción. A: datos empíricos. B mín: *LM* mín. B: *LM* ( $l_s=0.2, \alpha=0.5$ ). B máx: *LM* máx. C: *ICM* ( $p_{v,w} = 1\%$ ). D: *LTM*. E: *WCM*

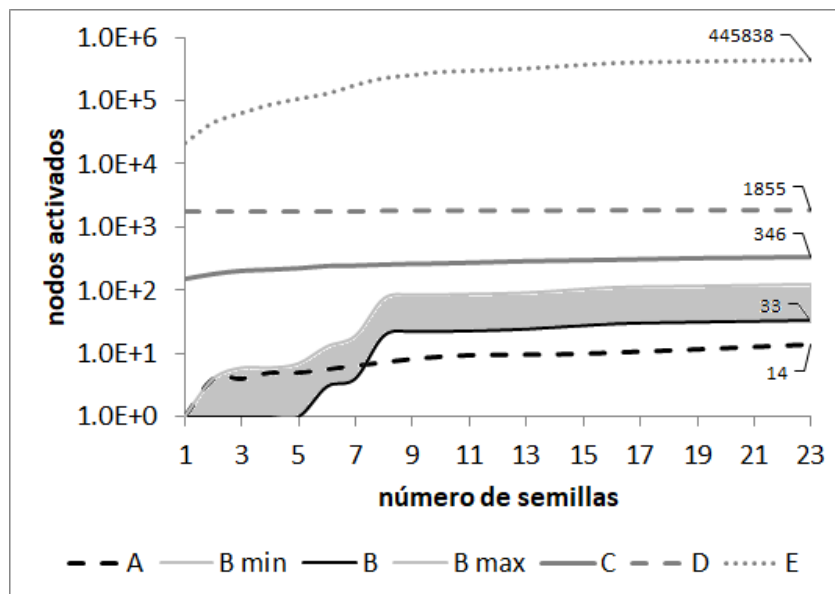


Figura 16. Exactitud: Propagación de un mensaje informativo.

En la figura anterior las curvas que se presentan tienen la siguiente descripción. A: datos empíricos. B mín: *LM* mín. B: *LM* ( $l_s=0.2, \alpha=0.5$ ). B máx: *LM* máx. C: *ICM* ( $p_{v,w} = 1\%$ ). D: *LTM*. E: *WCM*.

Las figuras 14 a 16 también muestran la flexibilidad de los módulos de *Lucy Model* para adaptarse a los diferentes niveles de importancia del mensaje, así como los diferentes pesos que el usuario del modelo desea dar a los parámetros de los umbrales de activación de los nodos.

*LM* activa un número mucho menor de nodos en relación con los otros modelos porque consideramos la probabilidad de que los nodos reciban el mensaje ( $w_{ji}$ ), mientras que los

otros modelos no. Si en *LM* solo consideramos la probabilidad de influencia  $Unh_{ji}$ , sin considerar  $w_{ji}$ , los resultados son los siguientes (Figura 17).

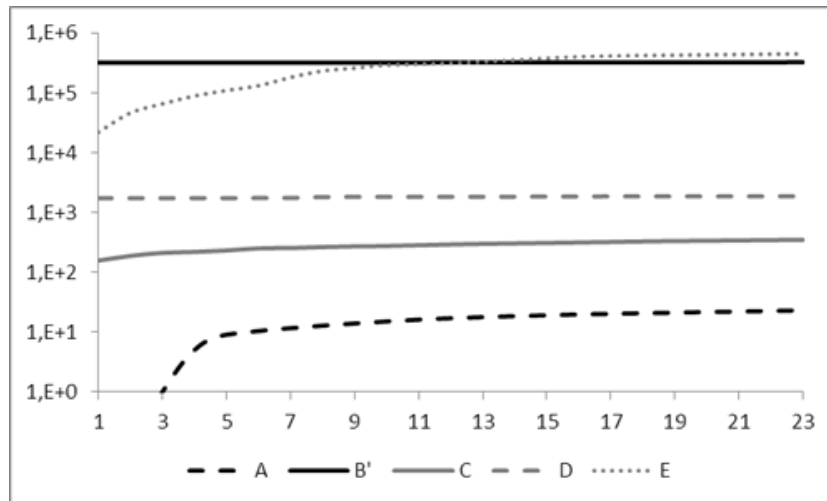


Figura 17. Propagación de un mensaje no homofílico.

En la figura anterior las curvas que se presentan tienen la siguiente descripción. A: datos empíricos. B: *LM* ( $Is=0.2, \alpha=0.5$ ). C: *ICM* ( $p_{v,w} = 1\%$ ). D: *LTM*. E: *WCM*.

En las figuras de la Figura 14 a la

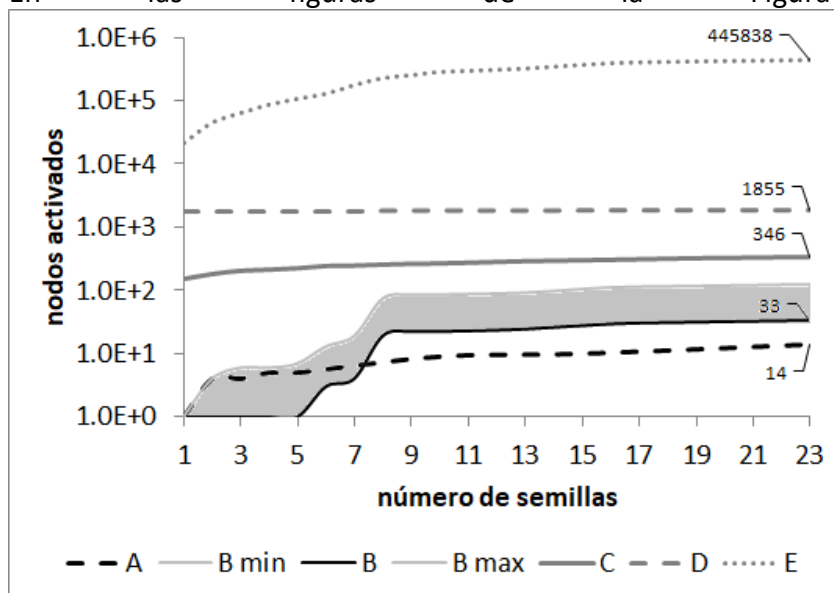
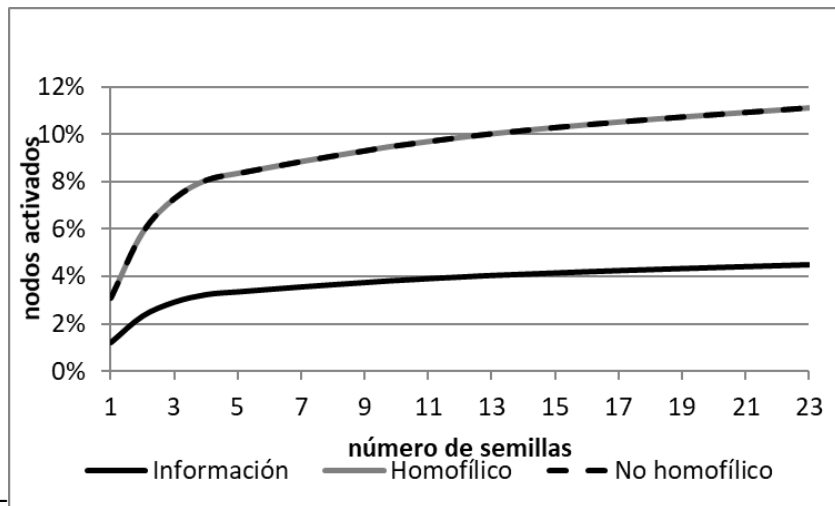


Figura 16 observamos que *Lucy Model* alcanza el menor número de nodos activados con relación a *LTM*, *ICM* y *WCM*. Esto se debe a que en nuestro modelo deben cumplirse dos condiciones: primero debe lograrse que el nodo a activar reciba el mensaje, y en segundo lugar se debe intentar activar dicho nodo. Los modelos de los otros autores contemplan solo una condición: intentar activar a los nodos.



En las figuras Figura 18-

Figura 20 mostramos el comportamiento de *Lucy Model* en diferentes tipos de redes: la red de Facebook rastreada, una red de *YouTube* y una red sintética de *Power-Law*. Las pruebas en ordenador nos permitieron hacer algunos hallazgos. En tres módulos de *Lucy Model* usamos los mismos parámetros:  $I_s = 0.2, \alpha = 0.5$ .

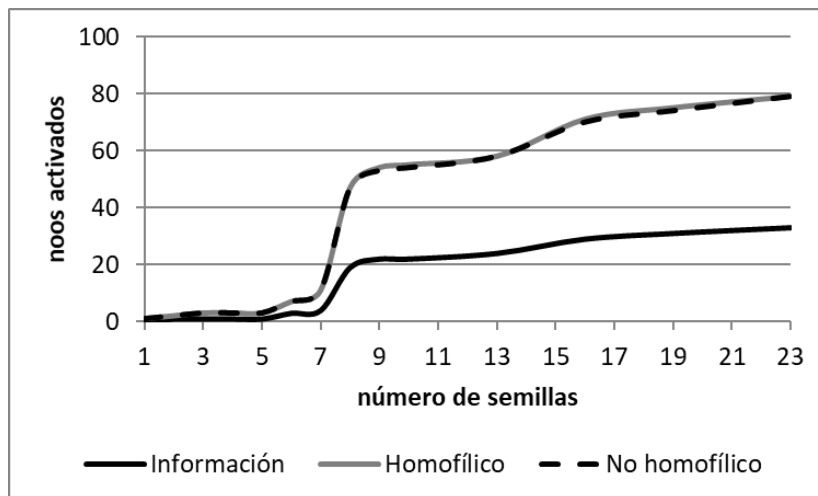


Figura 18. *Lucy Model*. Red Facebook rastreada.

En las figuras Figura 19 y Figura 20 puesto que usamos otras redes, tuvimos que usar otras semillas. Elegimos los 23 nodos con el valor más alto de centralidad de PageRank como semillas y los ordenamos de forma descendente. Con estas semillas, la activación de los nodos es mucho mayor como vemos en las figuras respectivas.

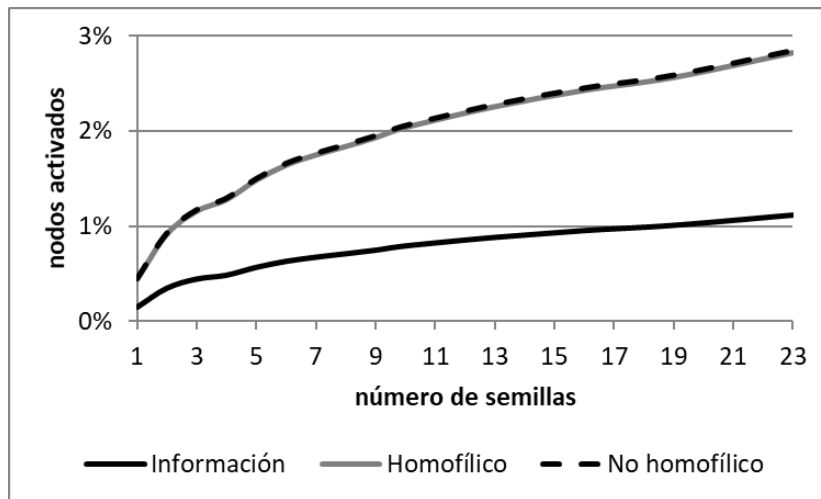


Figura 19. Lucy Model. Red YouTube.

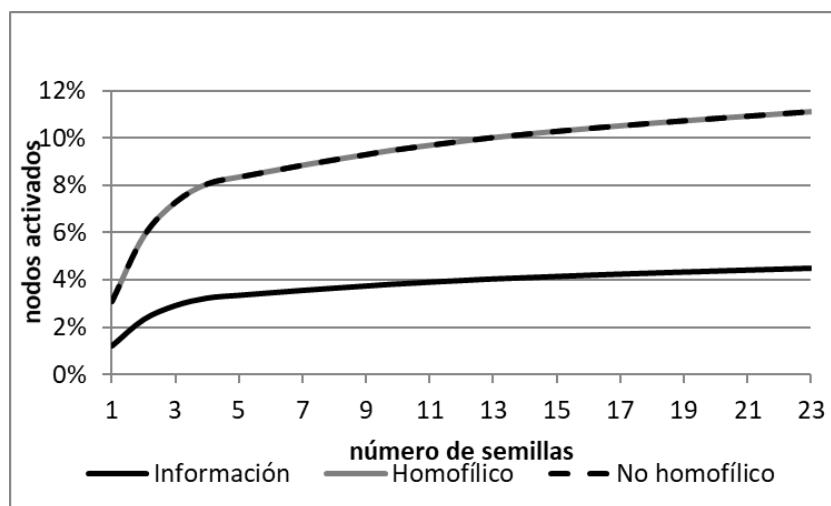


Figura 20. Lucy Model. Red sintética Power-Law.

En primer lugar, se observan dos interesantes resultados:

1. El mensaje no homofílico y el mensaje homofílico tienen casi el mismo número de nodos activados. De hecho, la diferencia relativa promedio en el número de nodos activados entre estos dos tipos de mensajes en las redes de *Facebook*, *YouTube* y *Power-Law* es 0.46%, 0.85% y 0.19% respectivamente. Obtuvimos resultados similares en los experimentos empíricos.
2. El mensaje informativo alcanza la propagación más pequeña con respecto a los otros tipos de mensajes. Este resultado es el mismo obtenido en los experimentos de campo (Figura 13).

Para explicar estos resultados, la activación de los nodos depende de *tie strength* y de los umbrales de activación. *Tie strength* entre nodos es el mismo puesto que la propagación de las tres clases de mensajes se realiza en el mismo grafo. Por lo tanto, la activación de los nodos depende solamente de los umbrales de activación. Si el número de nodos que se activaron fue casi el mismo con las dos clases de mensajes no homofílica y homofílica, y fueron mayores al número de nodos activados con el mensaje informativo, entonces los umbrales de activación se relacionan de la siguiente manera:

$$U_{ij} > Unh_{ji} \approx Uh_{ji} \quad (14)$$

por lo tanto:  $1 - (\alpha Rr_{ij} + \beta Ip_j) \approx 1 - (\alpha Ar_{ji} + \beta Ip_j)$ . Si a las dos clases de mensajes se les asignó la misma importancia ( $Ip_j$ ), diferenciándose en que (Ec. 8) depende de la relevancia de los nodos, mientras que (Ec. 9) depende de la afinidad entonces:

$$\begin{aligned} Ar_{ji} &\approx Rr_{ij} \\ Ar_{ij} &\approx Rr_{ji} \end{aligned} \quad (15)$$

Por otro lado, en la ecuación (11) se tiene:  $Rr_{ij} = \frac{R_i}{R_i + R_j}$ , análogamente:  $Rr_{ji} = \frac{R_j}{R_i + R_j}$ , Entonces:

$$Rr_{ij} + Rr_{ji} = 1 \quad (16)$$

Reemplazando la ecuación (15) en la ecuación (16) tenemos:

$$Ar_{ji} + Ar_{ij} = 1 \quad (17)$$

Finalmente, reemplazando la ecuación (15) en la ecuación (17):

$$\begin{aligned} Ar_{ji} + Rr_{ji} &= 1 \\ Ar_{ij} + Rr_{ij} &= 1 \end{aligned} \quad (18)$$

Las ecuaciones de la (15) a la (18) son interesantes relaciones entre la afinidad y la relevancia de los nodos.

Los grafos pueden ser representados a través de diferentes estructuras de datos, dos de ellas son: la matriz de adyacencia y las listas ligadas de adyacencia. En la gran mayoría de los casos, las listas ligadas de adyacencia son estructuras mucha más pequeñas que una matriz de adyacencia para representar un mismo grafo, requiriendo por lo tanto menos espacio de memoria a la hora de implementarlo en computadora. Por otro lado, siendo mucho más pequeña la estructura de datos, el recorrido secuencial exhaustivo, que realizamos en los modelos *LM*, *LTM*, *ICM* y *WCM*, es mucho más rápido con listas ligadas de adyacencia que con una matriz de adyacencia. Utilizar una estructura de datos mucho más pequeña para representar un grafo y poderlo recorrer en mucho menor tiempo, no altera el espacio de búsqueda, es decir, el conjunto de todas las posibles soluciones (en nuestro caso, el alcance de la propagación de un mensaje). Finalmente, utilizar listas ligadas de adyacencia no compromete la calidad de los resultados ni su validación. Las dos ventajas (menores tiempos y espacio) que ofrece las listas ligadas de adyacencia frente a las matrices de adyacencia las pudimos constatar cuando implementamos los cuatro modelos con los dos tipos de estructuras de datos. Solo para fines didácticos, nosotros hemos explicado el funcionamiento de nuestro modelo *LM*, representando los grafos con matrices de adyacencia (algoritmos 1 y 2).

Para ver la diferencia entre las dos representaciones de un grafo, hacemos el siguiente análisis. La red de Facebook rastreada tiene 968.810 nodos y 2.742.523 enlaces, con un grado promedio de nodos de solo 5,66. Para representar esta red con una matriz de adyacencia, necesitamos una matriz de  $(968.810)^2$  celdas. Para representar la misma red con listas de adyacencia ligadas, solo necesitamos  $(968.810) * (5.66)$  ubicaciones de memoria. La cantidad

## Capítulo 5. Resultados de validación y análisis

de memoria requerida con las matrices de adyacencia es 171.167,84 veces mayor que la cantidad de memoria requerida con las listas ligadas de adyacencia. El espacio de memoria utilizado está en el orden de  $O(n * \langle k \rangle)$  donde  $n$  es el número de nodos y  $\langle k \rangle$  es el grado promedio del grafo. Claramente  $n \gg \langle k \rangle$ .

En la Tabla 8 comparamos el rendimiento de los modelos con las dos estructuras de datos. En la Tabla 9 presentamos los tiempos de ejecución de los cuatro modelos implementados con listas ligadas. El rendimiento de los modelos con los grafos de *YouTube* y *Power-Law* fue muy similar a los resultados presentados en la Tabla 8 y la Tabla 9.



Tabla 8. Rendimiento de simuladores implementados con red de Facebook rastreada y diferentes estructuras de datos

<i>LM, LTM, ICM, WCM</i>			
Estructura de datos	Posiciones de memoria requerida	Tamaño relativo de memoria	Tiempo de ejecución
Matriz de adyacencia	$n^2$	$\frac{n}{\langle k \rangle} = 171.167,84$	No fue posible ejecutar el simulador porque el tamaño de la matriz desbordaba la memoria del computador
Listas ligadas	$n * \langle k \rangle$	1	Menos de 20 minutos

Tabla 9. Tiempo de ejecución de los simuladores [mm:ss] con la red social Facebook rastreada y listas ligadas

	<i>LM</i>		<i>LTM</i>	<i>ICM</i>	<i>WCM</i>
<i>No homofílica</i>	<i>Homofílica</i>	<i>Informativa</i>			
13:14	18:39	15:20	6:22	6:32	19:27

Nosotros justificamos la validez estadística de los resultados de *LM* a través de las medidas de exactitud y precisión de los mismos. La exactitud; es decir, qué tan cerca están los resultados obtenidos de los valores reales (sesgo), la expresamos en términos de la raíz cuadrada del error cuadrático medio (Tabla 10). La precisión; es decir, la dispersión de los resultados la expresamos en términos de la desviación estándar y del intervalo de confianza (Tabla 11).

$$RMSE = \sqrt{\frac{\sum_1^n (\text{Pronosticado}_i - \text{Real}_i)^2}{n}}$$

El error de *LM* con respecto a los resultados empíricos se calculó con la curva que marca el límite superior de las bandas grises, que es la que produce el error máximo.

Tabla 10. Exactitud: Raíz cuadrada del error cuadrático medio

Clase de mensaje	<i>LM</i>	<i>LTM</i>	<i>ICM</i>	<i>WCM</i>
No homofílico	53	238	1,759	241,352
Homofílico	52	236	1,757	241,350
Informativo	58	241	1,761	241,357

La Tabla 11 muestra la dispersión de los datos o precisión en términos de desviación estándar  $\sigma$  y del intervalo de confianza de las predicciones de *LM*, *LTM*, *ICM* y *WCM* que se mostraron

en las figuras Figura 14, Figura 15 y

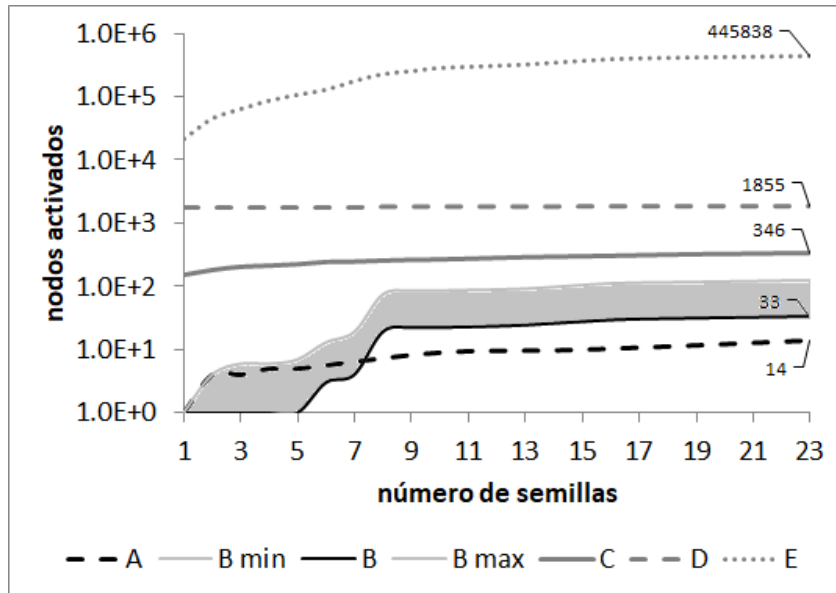


Figura 16. Nosotros encontramos que *Lucy Model* obtiene resultados significativamente más precisos en comparación con los otros modelos.

$$\text{Desviación estándar } \sigma = \sqrt{\frac{(x_i - \bar{x})^2}{N}}$$

$$\text{Intervalo de Confianza} = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

Tabla 11. Precisión: Desviación estándar  $\sigma$  e Intervalo de Confianza

	<i>Lucy Model</i>			<i>LTM</i>	<i>ICM</i>	<i>WCM</i>
	No homofílico	Homofílico	Informativo			
Desviación estándar $\sigma$	12.47	15.69	6.99	861	81	356,667.49
Intervalo de confianza	0.43	0.97	0.43	53.40	5.03	22,106.07

### 5.3 Hallazgos

Un mensaje de influencia homofílica y un mensaje de influencia no homofílica son de naturaleza diferente y, por lo tanto, sus umbrales de activación respectivos son función de distintas variables. Pero descubrimos que sus curvas de activación de nodos son prácticamente las mismas en pruebas en ordenador y similares en los experimentos de campo. Es decir, las dos clases de mensajes producen los mismos resultados. Estos resultados nos llevaron a la conclusión de que no es necesario definir dos umbrales de activación de nodos sino solo uno, unificando las dos variables (afinidad y relevancia) en una sola variable.

#### 5.4 Resumen del capítulo

Fue indispensable realizar pruebas de campo para poder comparar las predicciones de *Lucy Model* y de los otros tres modelos existentes *LTM*, *ICM* y *WCM* con los datos empíricos. De esta manera se pudo determinar qué modelo predictor se acerca más a los datos reales.

En las pruebas de campo realizadas en la plataforma *Facebook*, se determinó el número de personas que fueron influenciadas o informadas con un mensaje no homofílico, un mensaje homofílico y un mensaje informativo respectivamente. Estos resultados se presentan gráficamente utilizando un sistema de ejes de coordenadas cartesianas. En el eje horizontal se representan las veintitrés semillas dispuestas en forma descendente según el número de amigos que tienen, y en el mismo orden que publicaron cada mensaje a lo largo de los ocho días que duró la publicación. Cada semilla publicó cada mensaje por una sola vez. En el eje vertical de cada gráfico se representa el número de nodos que fueron activados.

Los primeros gráficos muestran en forma concluyente que las predicciones hechas en todos los casos por *Lucy Model*, son mucho más cercanas a los datos empíricos obtenidos, en comparación con las de los otros modelos.

Los últimos gráficos muestran los resultados de las pruebas en ordenador realizadas solamente con *Lucy Model*, con las tres clases de mensajes, y utilizando tres redes diferentes: red Facebook rastreada, red YouTube y red sintética *Power-Law*. Con estas pruebas pudimos encontrar que, en los tres tipos de redes, la propagación de los mensajes homofílico y no homofílico tienen prácticamente el mismo alcance, y estos son siempre mayores al alcance que tiene la propagación del mensaje informativo. Estos resultados de laboratorio se confirman con los resultados obtenidos experimentalmente.

Con los resultados de laboratorio que se confirman con los experimentos de campo, pudimos deducir cuatro expresiones matemáticas que resultan interesantes:

1. Las relevancias relativas mutuas entre dos nodos suman 1.
2. Las afinidades relativas mutuas entre dos nodos suman 1.
3. La afinidad relativa más la relevancia relativa entre dos nodos suman 1.

En cuanto a los programas informáticos que implementan los cuatro modelos predictores, estos se desarrollaron, no utilizando matrices de adyacencia, sino listas indexadas. De esta manera se pudo lograr dos cosas: en primer lugar, que se puedan ejecutar los programas y, en segundo lugar, que se puedan ejecutar en tiempos muchísimo menores.

Finalmente, mostramos que *LM* es mucho más exacto en términos de la raíz cuadrada del error cuadrático medio *RMSE* y mucho más preciso en términos de la desviación estándar  $\sigma$  con respecto de los otros modelos.

## Capítulo 6

### 6 *Lucy Model* en la Telemática

*Lucy Model* puede ser ajustado para representar la propagación de mensajes en aplicaciones telemáticas orientadas a distintos ámbitos.

A continuación, describimos algunas de las aplicaciones telemáticas en las cuales se puede adaptar *LM* para representar la propagación de información e influencia.

#### 6.1 Simulación de distribución de apps para teléfonos móviles

*Lucy Model* puede usarse para pronosticar el alcance de la distribución y actualización de aplicaciones informáticas para smart phones (*apps*). Por ejemplo, en el afán de dar un mejor servicio al usuario, una compañía proveedora de un cierto servicio puede desear distribuir a sus clientes una *app* que les permita acceder a información detallada de las características del servicio que ella provee. El proveedor envía la aplicación a través de un mensaje.

La compañía envía el mensaje sólo a los usuarios más importantes, quienes, si lo desean, lo pueden instalar en su móvil. El proveedor del servicio sugiere al usuario reenviar la aplicación a sus contactos. Los usuarios que reciben la aplicación de otro usuario lo toman como una recomendación de un amigo o de alguien importante, por lo que estarían más dispuestos a leer el mensaje. La compañía proveedora ahorra ancho de banda porque ella no tiene que repartir la *app* directamente a todos sus clientes, sino que utilizan el ancho de banda de sus clientes. En el contexto de propagación de mensajes, *activar un nodo* significa que el cliente instala la *app* en su teléfono.

La probabilidad de que el usuario destino lea el mensaje depende de qué tan amigo y relevante es para él el remitente (*tie strength*). Una vez que el usuario recibe la *app*, decide instalarlo si le es útil (Umbral de información).

#### 6.2 Propagación de virus informático en las redes de telefonía móvil

Un virus informático es un software que altera el normal funcionamiento de un dispositivo, tales como un Smartphone. Una manera de propagar virus es a través del intercambio de archivos de diferente tipo. Las redes sociales son un medio propicio para la propagación de virus de telefonía móvil.

La seguridad de los teléfonos móviles frente a las amenazas de los virus informáticos es un tema que preocupa a usuarios y a operadores del servicio. Una manera de abordar este tema

es estudiando los factores que influyen en su propagación, y para encontrar una forma de pronosticar la propagación que alcanzaría a través de las redes sociales de telefonía móvil.

La *activación de nodos* en el presente contexto significa que los Smartphone se contaminen de virus.

La probabilidad de que un usuario reciba el archivo infectado depende de qué tan amigo o importante es el remitente (*tie strength*). La probabilidad de que el usuario abra el archivo depende de qué tan cercano es el remitente.

### 6.3 Predicción de pérdida de clientes de los operadores de telefonía móvil

A medida que el mercado móvil se satura es más difícil para los proveedores conseguir nuevos clientes. Debido a la competencia entre operadores y la capacidad de los clientes de moverse desde un proveedor a otro, todos los proveedores de telefonía móvil sufren la pérdida de clientes (*churn*). La pérdida de clientes depende de aspectos intrínsecos como precio y calidad de servicio, pero también depende de aspectos extrínsecos como la opinión de la gente respecto del servicio y de la imagen de la empresa. Es más caro conseguir un nuevo cliente que mantener un cliente existente que se quiere ir. Por esta razón, muchas operadoras están enfocándose en conservar clientes existentes más que en conseguir nuevos.

Muchas veces la satisfacción o insatisfacción de la gente con el servicio y las recomendaciones que realiza, se expresa a través de las redes sociales online. Las empresas están utilizando las *OSNs* para conseguir recomendaciones de clientes claves. Luego de detectar a los usuarios que dan señales de querer abandonar el servicio (usuarios que llaman con más frecuencia a usuarios de otras telefónicas), el proveedor se dirige a los más relevantes para ofrecerles incentivos adicionales para mantener su interés en el servicio, así poder conservarlos, y por otro lado lograr que estos clientes recomienden a sus contactos las bondades del servicio. La recomendación desinteresada de un amigo puede producir una reacción en cadena de recomendaciones, con un impacto mayor sobre el mercado que cualquier campaña publicitaria interesada.

*Lucy Model* se puede adaptar fácilmente para pronosticar el número de clientes que piensan abandonar el servicio de telefonía móvil (*churn*). La forma de hacerlo es publicando un mensaje recomendando abandonar el servicio a través de los clientes más relevantes de la red social móvil. Este mensaje se propaga solamente en un simulador, así que no existe ningún peligro.

La probabilidad de que el usuario destino lea el mensaje depende de qué tan amigo y relevante es para él el remitente (*tie strength*).

En este contexto, la importancia del asunto percibido por el cliente  $I p_j$  debe reemplazarse por el grado de satisfacción del servicio  $S p_j$  que percibe el cliente, que a su vez está en función de la calidad de servicio  $QoS$  que brinda el proveedor.

#### **6.4 Resumen de capítulo**

Lucy Model puede ser ajustado para representar la propagación de mensajes en aplicaciones telemáticas orientadas a distintos ámbitos como son las siguientes:

1. Simulación de distribución de apps para teléfonos móviles
2. Propagación de virus informático en las redes de telefonía móvil
3. Predicción de pérdida de clientes de los operadores de telefonía móvil

Lucy Model puede usarse para pronosticar el alcance de la distribución y actualización de aplicaciones informáticas para smart phones (apps). Por ejemplo, en el afán de dar un mejor servicio al usuario, una compañía proveedora de un cierto servicio puede desear distribuir a sus clientes una app que les permita acceder a información detallada de las características del servicio que ella provee. El proveedor envía la aplicación a través de un mensaje.

Una manera de propagar virus en una red celular es intercambiando archivos a través de las redes sociales online, tales como WhatsApp y Messenger. El alcance de esta contaminación puede ser simulado y pronosticado con el sistema Lucy Model.

Lucy Model se puede adaptar fácilmente para pronosticar el número de clientes que piensan abandonar el servicio de telefonía móvil (churn). La forma de hacerlo es publicando un mensaje recomendando abandonar el servicio a través de los clientes más relevantes de la red social móvil. Este mensaje se propaga solamente en un simulador, así que no existe ningún peligro.

# Capítulo 7

## 7 Conclusiones y trabajo futuro

### 7.1 Conclusiones

#### 7.1.1 Beneficios derivados de Lucy Model

Luego de realizar pruebas en ordenador con *Lucy Model* y con los modelos base, y después de confrontarlos con los resultados de los experimentos empíricos, determinamos de manera concluyente e irrefutable que las predicciones realizadas por *Lucy Model* son notablemente más exactas y precisas que las realizadas por los modelos de línea base, *Linear Threshold Model* y *Independent Cascade Model*, y *Weighted Cascade Model*.

La contribución de nuestro trabajo es identificar nuevos elementos que están presentes en los procesos de propagación de información e influencia en la vida real, y que no se han tenido en cuenta antes. Estos elementos se incorporaron a *LM*, permitiendo que el modelo represente estos procesos de una manera más fina y cercana a la realidad. Los elementos son los siguientes:

- Tres clases de propagación de mensajes en las redes sociales (influencia homofílica, influencia no homofílica e informativa).
- El concepto de afinidad entre nodos y la métrica para calcularlo.
- El concepto de *tie strength* definido como la probabilidad de que un nodo reciba un mensaje y la expresión matemática para cuantificarlo.
- El concepto de umbral de activación definido como la probabilidad de activación de un nodo. Cada nodo a influir ( $j$ ) tiene un umbral de activación diferente para cada clase de propagación de mensajes y para cada nodo que influye ( $i$ ).
- La importancia de un mensaje para las personas en general y la importancia percibida para cada individuo en particular.
- *LM* es configurable para permitir flexibilidad de uso para adaptarse a diferentes niveles de importancia de un mensaje, así como para asignar distintos pesos a algunas variables que intervienen en el modelo. Es importante indicar que diseñamos *LM* antes de conocer los datos experimentales.

Por otro lado, los resultados de la activación de nodos del *LM*, *LTM*, *ICM* y *WCM*, tienen las características de *monotonidad* y *submodularidad*, conceptos que definimos en la Sección 2.3.

Con respecto al rendimiento de nuestro modelo, el uso de listas ligadas de adyacencia en lugar de matrices de adyacencia nos ha permitido usar un tiempo de procesamiento y espacio de memoria muy pequeños, y realizar experimentos con grafos mucho más grandes del orden de millones de nodos a velocidades mucho más altas, demostrando la efectividad de *LM*.

En relación con la utilidad de *Lucy Model*, algunos sectores de la sociedad (gobiernos, movimientos sociales, vendedores, etc.) que usan *OSNs* les gustaría conocer el alcance de la difusión de información y la influencia que transmiten a través de estas redes. *LM* facilitará a los ingenieros de telecomunicaciones la tarea de realizar diferentes tipos de análisis sobre las características de la propagación de información e influencia a través de las redes sociales en línea, satisfaciendo las necesidades de estos sectores.

Nuestra propuesta tiene muchos nuevos elementos que pueden usarse para ayudar a desarrollar nuevos modelos que maximicen la propagación de mensajes; para predecir la propagación de virus tanto en redes de computadoras como en redes de teléfonos móviles que podrían propagarse incluso a través de las redes sociales en línea. Nuestro modelo también puede ayudar a predecir la pérdida de clientes en telefonía móvil.

### **7.1.2 Limitación de la implementación de Lucy Model**

*WCM* fue el modelo más simple y más fácil de implementar en comparación con los otros tres modelos. *LM* es el modelo más complejo de implementar ya que incorpora varios elementos que los otros modelos no lo consideran.

La implementación de *LM* permite trabajar con gráficos de hasta cinco millones de nodos, en una computadora PC core i7, memoria de 8 GB y sistema operativo *Ubuntu* (64 bit). Para gráficos más grandes, necesitamos hardware con más capacidad.

### **7.1.3 Dificultad de validación**

Para validar los resultados de este tipo de modelos, es necesario rastrear la red social real correspondiente, una tarea que es muy engorrosa y muy lenta, ya que podría durar varias semanas e incluso meses. Esto se debe a las medidas de seguridad impuestas por las compañías propietarias de las redes sociales en línea como *Facebook*. Para superar esta dificultad, están a disposición las redes artificiales o sintéticas.

## **7.2 Trabajo futuro**

Otras clases de mensajes adicionales a los estudiados en esta tesis, y que podrían considerarse en el futuro o en otros modelos son las siguientes:

- Mensajes que tratan de influenciar apelando al interés y conveniencia de las personas, donde aceptar la influencia signifique riesgos; por ejemplo: adquirir una nueva tecnología,



emprender un nuevo negocio, enrolar a un nuevo personal, realizar una inversión en un campo no conocido.

- Mensajes que tratan de influenciar a la gente apelando al mismo tiempo a los sentimientos y a los intereses y conveniencias personales; por ejemplo: propaganda electoral, anuncios comerciales tipo “protege a tus pequeños hijos con un seguro de salud”.

Noticias falsas que podrían provocar pánico social que, a su vez, conduzca a la quiebra de una institución financiera, la inestabilidad política y social de un país, o afectar la seguridad nacional.

El efecto de estas cuatro últimas clases de mensajes se podrá determinar luego de modelar cada uno de ellas y probarlas en redes sociales tanto sintéticas como reales, para luego contrastar sus resultados con los respectivos resultados empíricos.

Es posible que existan trabajos realizados, relacionados con algunos de los temas que proponemos como trabajo futuro, pero que van más allá del alcance de nuestra investigación, razón por la cual no los hemos incluido dentro de nuestra revisión bibliográfica. Nosotros estamos recomendando realizar estos trabajos, utilizando como modelo base a *Lucy Model*. El investigador que decida recoger alguna de nuestras recomendaciones sabrá realizar la revisión bibliográfica correspondiente.

*Lucy Model* considera en forma separada solo procesos homofílicos o no homofílicos de propagación de influencia. Pero en la vida real, los mensajes que se publican y que tienen la intención de influenciar a la gente, apelan tanto a la conveniencia y el interés personal, así como a los sentimientos. Por esta razón, sería importante diseñar un modelo que considere ambos tipos de influencia al mismo tiempo.

Por otro lado, los mensajes que se publican en la vida real en las redes sociales, unos pretenden influir en un sentido y otros quieren influir en sentido opuesto o simplemente diferente. Nuestro modelo únicamente considera la propagación de influencia en un sentido particular; sería muy interesante modelar la propagación de la influencia en al menos dos sentidos o con influencias opuestas.

Otro trabajo futuro podría ser predecir no solo cuántos nodos se activan sino también cuáles son los nodos activados. De esta manera se pueden hacer estudios de comportamiento y reacciones de los diferentes sectores de sociedad frente a una determinada situación en particular.

Los mensajes que se publican y se difunden a través de una red no es en forma instantánea, sino que podría tomar inclusive algunos días. En el transcurso de esos días la red social podría cambiar de tipología. Por este motivo, recomendamos estudiar los procesos de propagación con redes que evolucionan su topología en el tiempo, utilizando el modelo *LM*.

Muchas veces se necesita conocer qué tan rápido se propaga un mensaje a través de una red social, para a su vez poder predecir la rapidez de reacción de la gente frente a determinada situación, por ejemplo un tsunami, una inundación, el corte de un servicio público básico, etc. Por esta razón, otro trabajo futuro podría ser determinar el tiempo que lleva la activación de nodos en una red.

## Capítulo 7 Conclusiones y trabajo futuro

Por otro lado, puede ser interesante estudiar las características de propagación de sentimientos negativos (homofobia, xenofobia, racismo, etc.), malas noticias y noticias falsas, para que las autoridades de gobierno puedan tomar las medidas necesarias.

Finalmente, es muy útil poder determinar de qué manera se puede llegar con un mensaje al mayor número de gente posible. Es por esto que recomendamos estudiar la maximización de la activación utilizando *LM*.

## 8 Apéndices

## Apéndice A

### Abreviaturas y acrónimos

<i>apps</i>	Aplicaciones para teléfonos móviles inteligentes
<i>ICM</i>	Independent Cascade Model
<i>LM</i>	Lucy Model
<i>LTM</i>	Linear Threshold Model
<i>OSNs</i>	Online Social Networks
<i>QoS</i>	Quality of Service
<i>RMSE</i>	Root main square error
<i>SNA</i>	Social Network Analysis
<i>SNAP</i>	Stanford Network Analysis Project
<i>SMS</i>	Short Message Service
<i>WCM</i>	Weighted Cascade Model

## 9 Fuentes e información

- Abbas, S. M. A. (2013). An agent-based model of the development of friendship links within Facebook. *Computational and Mathematical Organization Theory*, 19(2), 232-252.  
<https://doi.org/10.1007/s10588-013-9156-z>
- Afrasiabi Rad, A., & Benyoucef, M. (2012). Measuring propagation in online social networks: The case of youtube. *Journal of Information Systems Applied Research*, 5(1), 26.
- Alexander, C. K., & Sadiku, M. N. O. (2017). *Fundamentals of electric circuits* (Sixth edition). McGraw-hill Education.
- Bandeli, K. K., & Agarwal, N. (2018). Analyzing the role of media orchestration in conducting disinformation campaigns on blogs. *Computational and Mathematical Organization Theory*.  
<https://doi.org/10.1007/s10588-018-09288-9>
- Bekiari, A., & Hassanagas, N. (2015). Verbal Aggressiveness Exploration through Complete Social Network Analysis: Using Physical Education Students' Class as an Illustration. *Int'l J. Soc. Sci. Stud.*, 3, 30.
- Bolthausen, E., & Wüthrich, M. V. (2013). Bernoulli's Law Of Large Numbers. *ASTIN Bulletin: The Journal of the IAA*, 43(2), 73–79.
- Brandes, U., & Fleischer, D. (2005). Centrality measures based on current flow. *Annual symposium on theoretical aspects of computer science*, 533–544.
- Chen, W., Lakshmanan, L. V. S., & Castillo, C. (2013). Information and Influence Propagation in Social Networks. *Synthesis Lectures on Data Management*, 5(4), 1-177.  
<https://doi.org/10.2200/S00527ED1V01Y201308DTM037>

## Fuentes de información

- De Meo, P., Ferrara, E., Fiumara, G., & Ricciardello, A. (2012). A novel measure of edge centrality in social networks. *Knowledge-based systems, 30*, 136–150.
- Dietz, K., & Heesterbeek, J. A. P. (2002). Daniel Bernoulli's epidemiological model revisited. *Mathematical biosciences, 180*(1-2), 1–21.
- Duong-Ba, Thuan Hong. (2014). *Resource allocation optimization in large scale distributed systems* [Dissertation, Oregon State University].  
[https://ir.library.oregonstate.edu/concern/graduate\\_thesis\\_or\\_dissertations/ff365874s](https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/ff365874s)
- Fan, W., & Yeung, K. H. (2011). Online social networks—Paradise of computer viruses. *Physica A: Statistical Mechanics and Its Applications, 390*(2), 189-197.  
<https://doi.org/10.1016/j.physa.2010.09.034>
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*(12), 7821-7826.  
<https://doi.org/10.1073/pnas.122653799>
- Golbeck, J. (2013). *Analyzing the Social Web*. Newnes.
- Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360-1380.
- Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *ACM Sigmod Record, 42*(2), 17–28.
- Hein, O., Schwind, M., & König, W. (2006). Scale-free networks. *Wirtschaftsinformatik, 48*(4), 267–275.
- Ito, S., Vymětal, D., Šperka, R., & Halaška, M. (2018). Process mining of a multi-agent business simulator. *Computational and Mathematical Organization Theory, 24*(4), 500-531.  
<https://doi.org/10.1007/s10588-018-9268-6>
- Jin, L., Chen, Y., Wang, T., Hui, P., & Vasilakos, A. V. (2013). Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine, 51*(9), 144–150.

## Fuentes de información

- Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146.
- Kosorukoff, A., & Passmore, D. L. (2011). *Social Network Analysis: Theory and Applications*. Passmore, D. L. <https://books.google.com.ec/books?id=LrAnswEACAAJ>
- Leskovec, J., & Sosič, R. (2016). SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM Transactions on Intelligent Systems and Technology*, 8(1), 1-20.  
<https://doi.org/10.1145/2898361>
- Li, L., Alderson, D., Doyle, J. C., & Willinger, W. (2005). Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, 2(4), 431–523.
- Luo, W., Tay, W. P., & Leng, M. (2016). Infection Spreading and Source Identification: A Hide and Seek Game. *IEEE Transactions on Signal Processing*, 64(16), 4228-4243.  
<https://doi.org/10.1109/TSP.2016.2558168>
- Masayuki Yano, James Douglass Penn, George Konidaris, & Anthony T Patera. (2013, agosto). *Math, Numerics, & Programming (for Mechanical Engineers)—Buscar con Google*.  
[https://ocw.mit.edu/ans7870/2/2.086/S13/MIT2\\_086S13\\_Textbook.pdf](https://ocw.mit.edu/ans7870/2/2.086/S13/MIT2_086S13_Textbook.pdf)
- Maturo, F., Migliori, S., & Paolone, F. (2018). Measuring and monitoring diversity in organizations through functional instruments with an application to ethnic workforce diversity of the U.S. Federal Agencies. *Computational and Mathematical Organization Theory*.  
<https://doi.org/10.1007/s10588-018-9267-7>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1), 415-444.  
<https://doi.org/10.1146/annurev.soc.27.1.415>
- Mondani, H. (2018). The underlying geometry of organizational dynamics: Similarity-based social space and labor flow network communities. *Computational and Mathematical Organization Theory*, 24(3), 378-400. <https://doi.org/10.1007/s10588-017-9260-6>

## Fuentes de información

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199206650.001.0001/acprof-9780199206650>

Petróczi, A., Bazsó, F., & Nepusz, T. (2006). *Measuring tie-strength in virtual social networks*.

Phadke, C., Uzunalioglu, H., Mendiratta, V. B., Kushnir, D., & Doran, D. (2013). Prediction of subscriber churn using social network analysis. *Bell Labs Technical Journal*, 17(4), 63–76.

Piqueira, J. R. C., & Araujo, V. O. (2009). A modified epidemiological model for computer viruses. *Applied Mathematics and Computation*, 213(2), 355-360.

<https://doi.org/10.1016/j.amc.2009.03.023>

Postigo-Boix, M., & Melús-Moreno, J. L. (2018). A social model based on customers' profiles for analyzing the churning process in the mobile market of data plans. *Physica A: Statistical Mechanics and its Applications*, 496, 571–592.

Rong, K., Hu, G., Lin, Y., Shi, Y., & Guo, L. (2015). Understanding business ecosystem using a 6C framework in Internet-of-Things-based sectors. *International Journal of Production Economics*, 159, 41–55.

Samadi, M., Nagi, R., Semenov, A., & Nikolaev, A. (2018). Seed activation scheduling for influence maximization in social networks. *Omega*, 77, 96-114.

<https://doi.org/10.1016/j.omega.2017.06.002>

Samadi, M., Nikolaev, A., & Nagi, R. (2016). A subjective evidence model for influence maximization in social networks. *Omega*, 59, 263-278. <https://doi.org/10.1016/j.omega.2015.06.014>

Swaminathan, A. (2014). *An Algorithm for Influence Maximization and Target Set Selection for the Deterministic Linear Threshold Model*. Virginia Polytechnic Institute and State University.

Teixeira, A. S., Monteiro, P. T., Carriço, J. A., Ramirez, M., & Francisco, A. P. (2013). Spanning edge betweenness. *Workshop on mining and learning with graphs*, 24, 27–31.

Uddin, S., Khan, A., Hossain, L., Piraveenan, M., & Carlsson, S. (2015). A topological framework to explore longitudinal social networks. *Computational and Mathematical Organization Theory*, 21(1), 48-68. <https://doi.org/10.1007/s10588-014-9176-3>



## Fuentes de información

Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. (2012). Structural diversity in social contagion.

*Proceedings of the National Academy of Sciences*, 109(16), 5962-5966.

<https://doi.org/10.1073/pnas.1116502109>

Vishwakarma, D. K., Varshney, D., & Yadav, A. (2019). Detection and veracity analysis of fake news via

scrapping and authenticating the web search. *Cognitive Systems Research*, 58, 217-229.

<https://doi.org/10.1016/j.cogsys.2019.07.004>

Wang, F., Jiang, W., Li, X., & Wang, G. (2018). Maximizing positive influence spread in online social

networks via fluid dynamics. *Future Generation Computer Systems*, 86, 1491-1502.

<https://doi.org/10.1016/j.future.2017.05.050>

Yang, S., Keller, F. B., & Zheng, L. (2016). *Social Network Analysis: Methods and Examples*. SAGE

Publications.

Yun, Q., & Gloor, P. A. (2015). The web mirrors value in the real world: Comparing a firm's valuation

with its web network position. *Computational and Mathematical Organization Theory*, 21(4),

356-379. <https://doi.org/10.1007/s10588-015-9189-6>

Zuo, X., Blackburn, J., Kourtellis, N., Skvoretz, J., & Iamnitchi, A. (2016). The power of indirect ties.

*Computer Communications*, 73, 188–199.