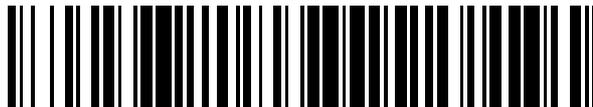


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 537 990**

51 Int. Cl.:

H04L 12/24 (2006.01)

H04L 29/08 (2006.01)

H04L 12/58 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.05.2012 E 12722321 (2)**

97 Fecha y número de publicación de la concesión europea: **04.03.2015 EP 2707996**

54 Título: **Un procedimiento de caracterización de una comunicación de red social usando motivos**

30 Prioridad:

10.05.2011 US 201113104904

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

16.06.2015

73 Titular/es:

**TELFÓNICA S.A. (100.0%)
C/ Gran Vía 28
28013 Madrid, ES**

72 Inventor/es:

**OLIVER RAMÍREZ, NURIA y
ZHAO, QIANKUN**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 537 990 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Un procedimiento de caracterización de una comunicación de red social usando motivos

Campo de la invención

5 La presente invención se refiere, en general, a redes de comunicación y, más específicamente, a la propagación de información a lo largo de las comunicaciones en tales redes.

Antecedentes de la invención

10 Las redes sociales representan los enlaces entre un conjunto de entidades conectadas entre sí con distintos tipos de relaciones. Por ejemplo, los artículos están enlazados por citas en una red de citas y los blogueros están enlazados por observaciones o listas de blogs en una red de blogs, mientras que los teléfonos celulares están conectados mediante llamadas telefónicas en una red de telefonía celular.

En la bibliografía, las redes sociales han sido extensamente estudiadas desde una perspectiva de la teoría de grafos (p. ej., las leyes de potencia, el fenómeno del mundo pequeño, la cobertura, etc.). Además, las propiedades de distintos tipos de redes complejas han sido comparadas.

15 Recientemente, los estudios de investigación sobre redes sociales, desde una perspectiva conductista, han recibido gran atención. Estas obras, que tratan de problemas tales como la identificación comunitaria y la detección del bombardeo publicitario, o que modelan flujos de información, tienen un montón de aplicaciones en sistemas de recomendaciones, búsqueda social, economía y publicidad.

20 Una cuestión fundamental en el análisis del flujo de información o de patrones de propagación dentro de redes sociales orientadas hacia la comunicación es cómo representar los datos de comunicación de modo que capturen cada trozo de información útil. En la bibliografía, unas pocas alternativas han sido propuestas y usadas para modelar las interacciones entre personas, donde habitualmente cada usuario es representado como un nodo en el gráfico y cada interacción como un borde en el gráfico, o acumulando interacciones entre usuarios añadiendo ponderaciones a la representación gráfica.

25 Estas representaciones son significativas y válidas en ciertas redes sociales, tales como las redes de amigos o de citas, donde la naturaleza de la relación está incrustada en, o puede obtenerse fácilmente de, los registros. Sin embargo, en el caso de redes sociales obtenidas de registros de comunicaciones (p. ej., llamadas telefónicas), es difícil deducir adecuadamente la naturaleza de las relaciones, debido a la multiplicidad de razones para hacer una llamada (p. ej., de negocios, personal, de servicio, etc.), y el papel que el contexto temporal desempeña en la comunicación. En otras palabras, una vez que un artículo haya citado a otro, la relación entre ambos artículos siempre es verdadera.

30 Sin embargo, las llamadas telefónicas se hacen por distintas razones y, por tanto, la naturaleza de una relación entre dos nodos en la red también puede depender del contexto temporal de las llamadas, es decir, una llamada hecha durante horas laborables es, probablemente, de naturaleza distinta a la de una llamada hecha a la noche. Lo mismo vale para otros atributos temporales, tales como la duración y la frecuencia de la interacción, o la distancia temporal entre dos llamadas (retardo temporal entre llamadas). Como resultado, las representaciones que se usan en los estudios existentes de propagación de la información no son válidas en el contexto de las comunicaciones telefónicas. Además, muchos estudios de propagación de la información suponen que las interacciones consecutivas transmiten el mismo trozo de información dentro de las redes deducidas, lo que no es necesariamente cierto en comunicaciones telefónicas.

35 Las soluciones como las propuestas por Zhang Yibin, y descritas en la solicitud de patente 2009 / 0287813, ofrecen procedimientos, aparatos y productos de programa de ordenador para analizar las relaciones de comunicación entre los dispositivos informáticos de un usuario y los usuarios de dispositivos informáticos remotos. Específicamente, Yibin divulga un aparato que puede incluir un procesador configurado para almacenar un registro de comunicaciones con un dispositivo remoto, y que puede ser adicionalmente configurado para calcular uno o más valores representativos de una relación de comunicación con el dispositivo remoto durante un periodo de tiempo basado, al menos en parte, en el registro almacenado de comunicaciones. El procesador puede ser adicionalmente configurado para determinar un patrón de relaciones de comunicación basado en dichos uno o más valores calculados.

45 La reciente disponibilidad de grandes cantidades de datos procedentes de una amplia variedad de redes (p. ej., redes sociales en línea, medios sociales y redes de contenidos generados por el usuario, proteínas, etc.) ha permitido el análisis de la propagación de la información en tales redes. Se ha hecho trabajo de investigación para analizar redes de telefonía celular, de mensajería instantánea, de blogs, de Flickr, de correo electrónico y de interacción de proteínas.

50 Más recientemente, las propiedades dinámicas de redes sociales a gran escala han sido estudiadas extensamente, donde las anotaciones temporales de cada comunicación son usadas para dividir el conjunto de datos total en una serie temporal de instantáneas, que es luego estudiada en términos de sus patrones temporales. Según estas líneas, se han llevado a cabo investigaciones sobre el estudio de cascadas de información, activadas por sucesos específicos. Otro

trabajo relacionado ha sido efectuado sobre la maximización de la influencia dentro de las redes sociales, a partir de patrones de comportamiento histórico, y con modelos probabilísticos, que resuelven el problema de mercadotecnia viral, dada la restricción de maximización de la influencia sobre la red.

5 Kleinberg et al., en el documento "La estructura de las trayectorias de la información en una red de comunicación social" (En KDD, 2008), proponen el concepto de distancia temporal para hallar una trayectoria de información en la red, llamada la estructura vertebral, donde la información tiene la más alta probabilidad de fluir en base a hábitos de comunicación temporal.

10 El concepto de *motivo* se originó en la biología, donde ha sido definido como patrones que reaparecen dentro de una red (de regulación de transcripción) mucho más a menudo que lo esperado aleatoriamente. Dentro de las redes de regulación de transcripción, se han efectuado investigaciones para mostrar experimentalmente que estos motivos de red son los cimientos de la red y que desempeñan papeles funcionales tales como la auto-regulación, los generadores de pulsos y los aceleradores de respuesta.

15 Los usuarios individuales pueden hacer y recibir un montón de llamadas telefónicas. Los enfoques tradicionales crean un enlace entre llamadas adyacente que comparten al menos un usuario, que pueden no ser suficientemente precisas como para reflejar cómo los usuarios se comunican colectivamente y cómo se propaga la información por la red. Por ejemplo, dos llamadas adyacentes que tratan de dos trozos distintos de información no deberían ser usadas para crear un trayecto de propagación de la información. Obsérvese que el problema de identificar el trozo de información que es propagado en cada llamada telefónica, o cualquier otra interacción social, es todavía un problema abierto, dado que, usualmente, el contenido asociado no está disponible o es demasiado sensible, en cuanto a la privacidad, para ser público.

20 Los estudios mencionados anteriormente se han centrado principalmente en el análisis de las propiedades estructurales o topológicas de distintos tipos de redes y modelos de construcción que explican los datos. Sin embargo, no ha habido ningún trabajo, hasta la fecha, que combine las propiedades topológicas globales de la red con patrones de comportamiento local, a fin de arrojar luz sobre los principios centrales de los patrones estructurales colectivos. Además, los enfoques anteriores han ignorado habitualmente los atributos temporales y la potencia de cada comunicación individual.

Sumario de la invención

La invención se refiere a un procedimiento de caracterización de una comunicación en red de acuerdo a la reivindicación 1. Las realizaciones preferidas del procedimiento están definidas en las reivindicaciones dependientes.

30 A fin de superar los problemas indicados en la sección anterior, la presente invención proporciona un procedimiento de caracterización de una red de comunicación, que está basado en la hipótesis de que la comunicación dentro de la red social está guiada por sucesos (p. ej., en una red de comunicación telefónica, cada llamada telefónica se hace con el fin de propagar u obtener un cierto elemento de información). Más específicamente, el procedimiento propuesto considera tanto la información temporal de las interacciones de comunicación dentro de la red como la duración de tales interacciones, como un parámetro indicativo de la información trasladada en tal interacción.

35 El procedimiento propuesto para la caracterización de una comunicación en red, para cada interacción de comunicación entre dos usuarios, u_i^o , u_i^d , de dicha red, toma los siguientes datos como entrada:

- una anotación temporal t_i de cada interacción de comunicación, y
- una duración δ_i de cada interacción de comunicación;

40 y una primera interacción de comunicación y una segunda interacción de comunicación dentro de la red son consideradas para la caracterización de tal red y forman parte de un grafo de comunicación C si:

- a) dichas interacciones de comunicación primera y segunda tienen al menos un usuario en común; y
- b) el tiempo transcurrido entre el final de la primera interacción de comunicación y el comienzo de la segunda interacción de comunicación está dentro de una ventana temporal W pre-establecida;

45 estando cada grafo de comunicación C_i^j formado por i usuarios únicos y j interacciones de comunicación que cumplen las condiciones a) y b) fijadas anteriormente.

La restricción temporal dada por la ventana temporal W es el filtro central para eliminar algo del ruido en los datos. De tal modo, el procedimiento propuesto usa los atributos temporales y la duración de las distintas interacciones (p. ej., llamadas) dentro de una red a fin de caracterizarla.

Por tanto, considerando que las anotaciones temporales de interacciones sociales están disponibles, en el procedimiento

propuesto:

- cada interacción entre dos usuarios y su correspondiente sello horario se toma como evidencia de la propagación de información entre los usuarios;

5 - es probable que dos interacciones adyacentes (es decir, interacciones que comparten al menos un usuario común) sean acerca del mismo elemento de información, si están cercanas temporalmente;

- cada elemento de información solamente es válido para una ventana temporal W (p. ej., 1 hora, 2 días, etc., según el tipo de red); y

- o bien los usuarios propagan esta información mediante interacciones inmediatas dentro de W , o bien no la propagarán ya más.

10 El procedimiento, preferiblemente, comprende además la etapa de separar por filtración cada grafo de comunicación con una restricción de frecuencia: un grafo de comunicación es considerado para su caracterización si el mismo tipo de grafo de comunicación aparece más de un cierto número de veces en la red de interacción social.

De acuerdo al procedimiento propuesto, dos grafos de comunicación son preferiblemente considerados del mismo tipo cuando comparten la misma estructura de topología.

15 Preferiblemente, para un umbral de soporte Θ dado, cualquier clase de equivalencia del grafo de comunicación $S = \{C_{ij}(1), C_{ij}(2), C_{ij}(3), \dots, C_{ij}(n)\}$, basada en la equivalencia de grafos con no menos que Θ grafos de comunicación, es mencionada como un motivo de comunicación.

Cada motivo de comunicación puede ser extraído usando un proceso basado en una primera búsqueda temprana, EFS.

Un motivo de comunicación es un motivo de máximo flujo cuando maximiza el flujo de información dentro de la red.

20 De esta manera, el procedimiento propuesto caracteriza los patrones de comunicación funcional de redes de interacción social que no solamente ocurren con frecuencia, sino que también son indicativos del proceso de propagación de la información en la red.

25 El procedimiento propuesto aquí incluye un enfoque genérico – mediante los llamados motivos de comunicación – que caracteriza distintos tipos de sucesos y los motivos de comunicación propuestos caracterizan cómo la información se propaga desde una perspectiva tanto funcional como cuantitativa.

El tamaño de la ventana temporal W depende, preferiblemente, del tipo de red.

Un grafo de comunicación tiene que aparecer más de un cierto número de veces en la red a fin de ser considerado.

Dicha anotación temporal t_i puede ser un sello horario del comienzo de la interacción de comunicación, o un sello horario del final de la interacción de comunicación.

30 **Breve descripción de los dibujos**

Para completar la descripción y a fin de facilitar la comprensión de la invención, se proporciona un conjunto de dibujos. Dichos dibujos forman parte integral de la descripción e ilustran realizaciones preferidas de la invención, que no deberían ser interpretadas como restrictivas del ámbito de la invención, sino solamente como ejemplos de cómo la invención puede ser realizada. Los dibujos comprenden las siguientes figuras:

35 la Figura 1 ilustra varias posibles representaciones de grafos de datos de Registros de Detalles de Llamadas (CDR).

La Figura 2 muestra varias posibles representaciones de grafos de datos de CDR, teniendo en consideración restricciones temporales de acuerdo al procedimiento propuesto.

La Figura 3 muestra una representación de grafo de un grupo de interacciones a partir de otro CDR.

La Figura 4 muestra el soporte para tres motivos de comunicación hallados en los grafos mostrados en la Figura 2.

40 La Figura 5 ilustra distintos tipos de motivos de comunicación.

La Figura 6 muestra los motivos más frecuentes, extraídos de los conjuntos de datos usados en el experimento (obtenidos con una ventana temporal $W = 4$ horas).

La Figura 7 muestra la correlación entre el soporte y la clasificación de los motivos.

La Figura 8 muestra gráficamente el impacto de la longitud de la ventana temporal W.

La Figura 9 muestra la cobertura de usuarios para los motivos de comunicación más frecuentes en el experimento.

Descripción detallada de las realizaciones preferidas

Se hará ahora referencia en detalle a una realización preferida del procedimiento para caracterizar redes sociales.

- 5 Una cuestión fundamental en el análisis del flujo de información o de los patrones de propagación dentro de redes sociales orientadas a la comunicación es cómo representar los datos de comunicación de modo que capturen cada elemento de información útil. Consideremos el ejemplo en la Tabla 1, que muestra varias entradas de un conjunto de datos de registros de detalles de llamadas (CDR) telefónicas, donde se registran las llamadas entre usuarios y sus duraciones y sellos horarios asociados, con una entrada por llamada. Se muestran en la Figura 1 un cierto número de representaciones alternativas de los datos de CDR en la Tabla 1. El primer grafo G1 está construido tomando a cada usuario (llamador / llamado) como un nodo y a cada llamada entre dos usuarios cualesquiera como un borde entre los dos nodos correspondientes. El grafo G2 extiende G1 añadiendo ponderaciones a los bordes, proporcionales a la frecuencia de llamadas entre dos nodos cualesquiera.
- 10

Tabla 1: CDR (Registros de Detalles de Llamadas)

<i>llamador</i>	<i>llamado</i>	<i>sello horario</i>	<i>duración</i>
1	3	9:00, 11 de agosto de 2008	2 minutos
3	8	9:10, 11 de agosto de 2008	5 minutos
8	7	9:21, 11 de agosto de 2008	3 minutos
7	6	9:41, 11 de agosto de 2008	2 minutos
1	2	9:59, 11 de agosto de 2008	9 minutos
2	4	10:08, 11 de agosto de 2008	6 minutos
4	1	10:43, 11 de agosto de 2008	6 minutos
1	4	10:36, 11 de agosto de 2008	3 minutos
4	1	11:08, 11 de agosto de 2008	2 minutos
1	4	11:36, 11 de agosto de 2008	5 minutos
2	5	11:36, 11 de agosto de 2008	1 minuto
2	3	11:46, 11 de agosto de 2008	2 minutos
2	6	11:56, 11 de agosto de 2008	6 minutos
2	8	13:36, 11 de agosto de 2008	5 minutos
7	4	14:30, 11 de agosto de 2008	2 minutos
5	2	17:36, 11 de agosto de 2008	6 minutos
5	6	17:56, 11 de agosto de 2008	1 minuto
5	7	18:24, 11 de agosto de 2008	3 minutos
5	8	18:38, 11 de agosto de 2008	2 minutos

- 15 Estas representaciones mostradas en la Figura 1 son significativas y válidas en ciertas redes sociales tales como las redes de amigos o de citas, donde la naturaleza de la relación está incrustada en – o puede ser fácilmente obtenida de – los registros. Sin embargo, para las redes sociales obtenidas de los registros de comunicación, es difícil deducir adecuadamente la naturaleza de las relaciones, debido a la multiplicidad de razones para hacer una llamada (p. ej., de negocios, personal, de servicio, etc.) y la implicación del contexto temporal. En otras palabras, una vez que un artículo
- 20

haya citado a otro, la relación citada entre ambos artículos siempre es verdadera. Sin embargo, las llamadas telefónicas se hacen por distintas razones y, por tanto, la naturaleza de una relación entre dos nodos en la red también puede depender del contexto temporal de las llamadas, es decir, una llamada hecha durante horas laborables es, probablemente, de naturaleza distinta a la de una llamada hecha a la noche. Lo mismo vale para otros atributos temporales tales como la duración y la frecuencia de la interacción, o la distancia temporal entre dos llamadas (retardo temporal entre llamadas). Como resultado, las representaciones que se usan en los estudios existentes de propagación de la información no son válidas en el contexto de las redes de comunicación telefónica. Además, muchos estudios de propagación de la información suponen que las interacciones consecutivas transmiten el mismo elemento de información dentro de las redes deducidas, que tampoco es una hipótesis válida.

5
10
15

Proponemos un procedimiento novedoso y significativo para modelar los comportamientos de los usuarios y los patrones de flujos de información. Por ejemplo, en base a los datos de CDR en la Tabla 1, se construyen cuatro secuencias distintas de llamada g1-g2-g3-g4, teniendo en cuenta los sellos horarios de llamadas, según se muestra en la Figura 2. En estos grafos, dos bordes están conectados si y solo si sus correspondientes sellos horarios no difieren en más de 30 minutos entre sí. En comparación con los grafos de la Figura 1, las cuatro secuencias en la Figura 2 tienen en cuenta la información temporal de las llamadas, permitiendo una caracterización más precisa del comportamiento y los flujos de información en las redes sociales.

De hecho, según lo demostrado por nuestros resultados, los atributos temporales asociados a los datos de comunicación son críticos para modelar la propagación de la información y el propósito detrás de cada comunicación. La duración de cada comunicación es un buen indicador para cuantificar la cantidad de información que se está propagando.

20

Por tanto, a fin de modelar cómo se propaga la información en redes sociales, se propone el concepto de motivo de comunicación y de motivo de flujo máximo. Se proponen dos algoritmos para descubrir tales motivos.

A continuación, usamos redes de llamadas telefónicas creadas a partir de Registros de Detalles de Llamadas (CDR) como un ejemplo de una red de comunicación.

25

Indiquemos con G una base de datos de los CDR. Cada registro en G está representado como un cuarteto $e_i = \langle u_i^o, u_i^d, t_i, \delta_i \rangle$, donde u_i^o representa al llamador, u_i^d al llamado, t_i al sello horario cuando la llamada fue establecida y δ_i a la duración de la llamada. Para representar el flujo de información en este tipo de red, definimos un grafo de comunicación como:

30

Definición de **Grafo de Comunicación**: Para una base de datos de CDR dada, representada como G, y una ventana temporal W que indica el lapso vital de cada elemento de información, un grafo de comunicación se define como un subconjunto de registros de G, indicado como $C = \{e_1, e_2, \dots, e_k\}$, de modo que $\forall e_i \in C, 1 \leq i \leq k$, existe al menos un usuario $e_j \in C$ e $i \neq j$ tal que:

- 1) $|\{u_i^o, u_i^d\} \cap \{u_j^o, u_j^d\}| > 0$, y
- 2) $0 < t_i - t_j - \delta_j < W$ o $0 < t_j - t_i - \delta_i < W$.

35

Obsérvese que, incluso aunque hay k entradas en el grafo de comunicación, el número de usuarios únicos implicados puede ser distinto, según las características de las llamadas que forman este grafo de comunicación. Por lo tanto, cada grafo de comunicación está anotado con dos números C_i^j , donde i indica el número de usuarios únicos (nodos en C) y j indica el número de interacciones (bordes) que forman este grafo de comunicación.

Por ejemplo, y sin pérdida de generalidad, dada la lista de llamadas en la base de datos de CDR mostrada en la Tabla 1 (p. ej., 1→3, 3→8, ...), y fijando la ventana temporal W en 30 minutos, entonces el primer grafo g1 en la Figura 2 se indica como C_5^4 , es decir, un grafo de comunicación con 5 usuarios y 4 llamadas (bordes).

40
45

La restricción *temporal* dada por la ventana temporal W es el filtro central para eliminar algo del ruido en los datos. Por ejemplo, en los tradicionales enfoques de análisis de flujo de información de redes sociales, si hay un enlace fuerte entre el usuario A y B, y hay un enlace fuerte entre B y C, entonces se supone que la información propagada desde A a B tiene una alta probabilidad de ser propagada a C. Sin embargo, obviamente, esto no es necesariamente el caso, según lo ilustrado en los siguientes escenarios: (1) Todas las interacciones entre A y B ocurrieron hace años y no han interactuado desde entonces, mientras que B y C solamente comenzaron a comunicarse recientemente; y (2) las interacciones entre A, B y C ocurrieron en el mismo año. La mayor parte del tiempo, después de que A y B interactúan, B interactúa con D y E, en lugar de con C. Sin embargo, B y C todavía tienen muchas interacciones, pero no después de las interacciones entre A y B. Los dos casos presentados anteriormente no son gestionados de manera efectiva, o son ignorados, por los más recientes enfoques del flujo de información.

50

Obsérvese también que el grafo de comunicación individual es una simplificación de la realidad. En los datos reales, es común observar que una persona está implicada en múltiples cadenas de comunicación, incluso durante un breve periodo de tiempo, dado que él (ella) puede recibir o iniciar múltiples llamadas, desde o a, más de una persona dentro de la

ventana temporal W predefinida. Por ejemplo, consideremos los CDR en la Tabla 2 a continuación, que incluye seis registros:

Tabla 2: CDR (Registros de Detalles de Llamadas)

<i>llamador</i>	<i>llamado</i>	<i>sello horario</i>	<i>duración</i>
1	2	10:09, 11 de junio	20 segundos
6	2	10:12, 11 de junio	30 segundos
7	2	10:15, 11 de junio	50 segundos
2	3	10:20, 11 de junio	60 segundos
3	4	10:40, 11 de junio	120 segundos
3	5	10:46, 11 de junio	20 segundos

5 La Figura 3 muestra el correspondiente grafo de comunicación para $W = 30$ minutos. No es claro, a partir del grafo, cuál elemento de información se propaga desde 2 a 3, dado que 2 ha recibido información desde 1, 6 y 7. Como resultado, algunos de los grafos de comunicación individuales pueden ser engañosos y no significativos semánticamente. Para tratar esta dificultad, proponemos filtrar los grafos de comunicación con una restricción de *frecuencia*; a fin de ser considerado, el mismo tipo de grafo de comunicación debe aparecer más de un cierto número de veces en la red de interacción social. La idea subyacente de usar la frecuencia está basada en el hecho de que los patrones de comunicación auténticos permanecerán a pesar del ruido.

En consecuencia, introducimos los conceptos de equivalencia de grafos y motivos de comunicación.

Definición de **Equivalencia de Grafos**: Para dos grafos de comunicación cualesquiera, $C_1 = \langle e_1, e_2, \dots, e_k \rangle$ y $C_2 = \langle e_1', e_2', \dots, e_k' \rangle$, sean S_1 y S_2 , respectivamente, el conjunto de usuarios distintos de los grafos C_1 y C_2 . Si hay una función biunívoca f entre los usuarios en S_1 y S_2 , tal que $\forall e_i \in C_1, \exists (f(U_i^a) \rightarrow f(U_i^d) \in C_2)$, y viceversa, entonces estos grafos de comunicación son mencionados como equivalentes.

Básicamente, dos grafos son equivalentes si comparten la misma estructura de topología. En base a la definición anterior de equivalencia de grafos, definimos el motivo de comunicación.

Definición de **Motivo de Comunicación**: Dado un umbral de soporte Θ , entonces cualquier clase de equivalencia de los grafos de comunicación $S = \{C_i^l(1), C_i^l(2), C_i^l(3), \dots, C_i^l(n)\}$ basada en la equivalencia de grafos con no menos de Θ grafos de comunicación se denomina un motivo de comunicación. Además, la cardinalidad de la clase de equivalencia S se denomina el soporte del motivo de comunicación.

Ejemplo: Tomando la base de datos de CDR de muestra en la Tabla 1 y los cuatro grafos de comunicación g_1 a g_4 mostrados en la Figura 2, la Figura 4 ilustra el soporte para tres motivos de comunicación que se hallan en los grafos. El primer motivo de comunicación tiene un soporte de 5 (3 veces en el primer grafo y 2 veces en el segundo grafo), el segundo tiene un soporte de 3, mientras que el último tiene un soporte de 2.

Finalmente, obsérvese que, a fin de tratar sesgos potenciales y el ruido local, nos gustaría que los gráficos de comunicación de los motivos estuvieran uniformemente distribuidos en la red social, en particular, dado que se supone que los motivos capturan propiedades intrínsecas de la red. Por tanto, nos gustaría evitar una situación donde la mayoría de los grafos de comunicación en un motivo comienzan desde el mismo nodo (p. ej., el número telefónico). Por tanto, introducimos el concepto de soporte-L para fijar una cota superior sobre el número de grafos de comunicación por origen que son tenidos en cuenta en el soporte global de los motivos de comunicación.

En resumen, el procedimiento propone caracterizar los patrones de comunicación funcional de redes de interacción social que no solamente ocurren con frecuencia, sino que también son indicativos del proceso de propagación de información en la red.

Los motivos han sido definidos como una representación de las funcionalidades incrustadas dentro de la red social. Se pretende que capturen patrones del comportamiento colectivo de los usuarios en la red. Sin embargo, si queremos aplicar los motivos a escenarios del mundo real tales como la mercadotecnia viral, necesitamos definir una medida para cuantificar la probabilidad de propagación de información dentro de cada motivo.

Extendemos a continuación la definición de motivo añadiendo medidas cuantitativas basadas en las propiedades temporales de cada interacción dentro del motivo. Definimos el motivo de flujo máximo como una estructura (motivo)

donde es probable que la información se propague con la mínima cantidad posible de pérdida de información. Los motivos de flujo máximo son motivos estándar con la restricción añadida de maximizar el flujo de información. Es decir, se espera que cada nodo (llamada telefónica) dentro del motivo de flujo máximo maximice el flujo de información.

5 Definición de **Maximización del flujo de información**: Dadas dos llamadas conectadas $e_i = \langle u_i^o, u_i^d, t_i, \delta_i \rangle$ y $e_j = \langle u_j^o, u_j^d, t_j, \delta_j \rangle$, e_i y e_j maximizan el flujo de información si, para todas las llamadas conectadas G' , donde $e_i, e_j \in G'$, no hay ningún otro e_j tal que $\delta_j / (\delta_i \times (t_j - t_i)) \leq \delta_i / (\delta_i \times (t_j - t_i))$.

Por ejemplo, en la red en la Figura 3, dada una llamada $7 \rightarrow 2$, el motivo de máximo flujo de información, originado en esta interacción, sería $7 \rightarrow 2$, $2 \rightarrow 3$ y $3 \rightarrow 4$, pero $3 \rightarrow 5$ no estará incluido en el motivo de flujo máximo, porque su flujo de información (duración de la llamada) es más pequeño que la llamada $3 \rightarrow 4$.

10 En base a las definiciones anteriores, los motivos de comunicación pueden revelar aspectos de los principios y la dinámica en las comunicaciones que tienen lugar dentro de la red. La Figura 5 ilustra cuatro tipos básicos de motivos de comunicación con distintos patrones de flujos de información: (a) Cadena larga: que es un grafo de comunicación con una lista extremadamente larga de participantes únicos, tales como 3; 5; 6; ..., 20 en la Figura; (b) Ping-pong: que representa un patrón con muy pocos participantes que se comunican repetidamente una y otra vez entre sí, tal como 3; 5 en la Figura; (c) Bucle: donde hay un bucle en la comunicación entre algunos de los participantes – pero no un bucle de ping-pong -, tal como desde 3 a 4, 8, 9 y nuevamente a 3 en la Figura; y (d) Estrella: donde todas las interacciones comienzan / terminan en un número limitado de miembros del grafo.

En general, un motivo de comunicación implica un cierto número de participantes distintos y una combinación de los motivos básicos definidos anteriormente. La distribución de los motivos básicos dentro del motivo más grande es una indicación de cómo fluye la información dentro del motivo, de cuán rápidamente puede esparcirse y de la topología de la red social.

Un objetivo importante del estudio de los motivos es entender mejor los patrones de propagación de información en las redes sociales. Por lo tanto, el interés principal está en los motivos que implican a más de dos usuarios, ya que una interacción única entre solamente dos usuarios y / o las estructuras de ping-ping de 2 personas no son funcionalmente adecuadas para propagar información.

Los motivos de comunicación y los motivos de flujo máximo se extraen de los datos de acuerdo a procesos de generación de motivos, que se explican en detalle más adelante. En el caso de los motivos de comunicación, debido a que las entradas en el conjunto de datos están almacenadas en orden temporal, se usa una estrategia de primera búsqueda temprana (EFS) para extraer los motivos. Por el contrario, se usa una estrategia de primera búsqueda de flujo máximo (MFS) para el descubrimiento de motivos de flujo máximo, ya que se desean las entradas del conjunto de datos que propagan grandes volúmenes de información.

El algoritmo 1 anterior esboza un proceso de extracción de motivos basado en la primera búsqueda temprana (EFS). Dado un usuario en una red social y una comunicación iniciada por este usuario, el algoritmo de EFS recorre los bordes adyacentes siguiendo esta comunicación inicial (entrada en el conjunto de datos) en orden no descendente de los sellos horarios de los bordes. Toda vez que se halla que un borde está dentro de la ventana temporal W dada, se añade al sub-grafo. Luego, el algoritmo continúa visitando más bordes que estén conectados con el sub-grafo, hasta que no pueda hallarse ningún borde nuevo que satisfaga la restricción de la ventana temporal. El sub-grafo hallado se añade al repositorio de candidatos, que es filtrado más tarde por el umbral de soporte Θ .

40 **Algoritmo 1:** Extracción de motivos de comunicación

Entrada: Grafo de comunicación G , ventana temporal W , umbral Θ

Salida: Una lista de motivos de comunicación M con su correspondiente soporte

Generar un conjunto de entradas seminales S en G ;

45 $M \leftarrow \Theta$;

para cada entrada $e \in S$ **hacer**

$C \leftarrow \Theta$; /* C es el sub-grafo a extraer */

$P \leftarrow \{e\}$; /* P almacena las entradas pendientes de ser procesadas */

mientras P no esté vacío **hacer**

$e_i \leftarrow$ la entrada con el más pequeño sello horario en P;

$C \leftarrow C \cup \{e_i\}$

para cada e_j adyacente a e_i **hacer**

5 $\text{si } 0 < t_j - t_i - \delta_i \leq W$ entonces $P \leftarrow P \cup \{e_j\}$

Hacer anónimos los Identificadores de usuario en C;

$M \leftarrow M \cup \{C\}$;

Aumentar en 1 el soporte de C en M;

para cada $C \in M$ **hacer**

10 $\text{si } C.\text{soporte} < \theta$ entonces eliminar C de M;

devolver M

15 Los motivos de máximo flujo de información son identificados con el algoritmo de primera búsqueda de flujo máximo enunciado en el Algoritmo 2. A partir de una entrada dada en el conjunto de datos, el algoritmo visita iterativamente el borde adyacente que lleva el máximo volumen de información, y lo añade al sub-grafo de motivo local. De tal modo, el algoritmo recorre la red siguiendo el trayecto que propaga el más grande volumen de información.

Algoritmo 2: Extracción del motivo de máximo flujo de información

Entrada: Grafo de comunicación G, ventana temporal W, umbral θ

20 **Salida:** Una lista de motivos de máximo flujo de información, M_i

Generar un conjunto de entradas seminales S en G;

$M \leftarrow \emptyset$;

para cada entrada $e \in S$ **hacer**

$C \leftarrow \emptyset$; /* C es el sub-grafo a extraer */

25 $P \leftarrow \{g\}$; /* P almacena las entradas pendientes de ser procesadas */

mientras P no esté vacío **hacer**

$e_i \leftarrow$ la entrada con el más pequeño sello horario en P;

$C \leftarrow C \cup \{e_i\}$;

$e_j \leftarrow$ la entrada adyacente de e_i dentro de W que maximice $\delta_j / (\delta_i (t_j - t_i))$;

30 $P \leftarrow P \cup \{e_j\}$; /* e_j podría ser \emptyset */;

Hacer anónimos los Identificadores de usuario en C;

$M_i \leftarrow M_i \cup \{C\}$;

Aumentar en 1 el soporte de C en M_i ;

para cada $C \in M_i$ **hacer**

35 $\text{Si } C.\text{soporte} < \theta$ entonces eliminar C de M_i ;

devolver M_i

Estos nuevos conceptos de motivos de comunicación y motivos de flujo máximo han sido introducidos para modelar cómo se propaga la información en las redes sociales. Los conceptos de motivos propuestos son una extensión del motivo de red en redes biológicas, que se refiere a patrones que se repiten dentro de una red mucho más a menudo que aleatoriamente.

5 El comportamiento de la propagación de información en estas redes sociales ha sido caracterizado usando estos motivos como una medición, y ha llevado a los siguientes hallazgos:

* Los patrones de comportamiento funcionales de la propagación de información dentro de una red de comunicación son estables a lo largo del tiempo.

10 * Los patrones de propagación de información en redes sociales síncronas y asíncronas son distintos y sensibles al coste de la comunicación.

* La cantidad de información que se propaga y su velocidad están correlacionadas y dependen de los perfiles de usuario.

15 La información temporal se usa para filtrar algo del ruido que está presente en los enfoques actuales del análisis de las redes sociales, en particular, cuando el contenido de la comunicación es privado o está ausente. El algoritmo propuesto de descubrimiento de motivos no usa ninguna información de contenido, pero la información temporal no es la única restricción; hay también restricciones topológicas y de frecuencia para asegurar que los motivos descubiertos sean estadísticamente significativos cuando se comparen con los sub-grafos en una red aleatoria equivalente.

El procedimiento propuesto para caracterizar las redes sociales puede ser usado para un cierto número de importantes aplicaciones comerciales, que incluyen, pero no se limitan a:

20 a. *Mercadotecnia Viral y Recomendación de Interacción*: La estabilidad temporal de los motivos sugiere que pueden ser una caracterización general y útil de la red, y útil para la mercadotecnia viral. Se podrían aplicar estrategias incentivadoras o recomendaciones para proveedores de servicios móviles o sedes de redes sociales, para aumentar el nivel de las interacciones de usuarios en base a los motivos descubiertos. Por ejemplo, los proveedores de servicios móviles tienen habitualmente un plan familiar o grupal para un grupo de usuarios. La frecuencia, la cobertura y el tamaño medio de los motivos podrían ser usados para optimizar las estrategias del plan. Los operadores móviles podrían diseñar un plan familiar o grupal optimizado con un número específico de usuarios y una cierta duración de tiempo de charla libre, midiendo el tamaño de los motivos que tienen un porcentaje mínimo de cobertura de usuarios y una duración mínima de las comunicaciones dentro de los motivos.

30 Las sedes en línea de redes sociales están comenzando a recomendar comunicaciones dentro de la red – tales como las recientemente introducidas por Facebook -, las cuales, a su vez, aumentarán los niveles de actividad en la red. El proceso de extracción de motivos permitirá seleccionar las interacciones recomendadas a fin de optimizar la cobertura. Además, la influencia de los usuarios individuales puede ser cuantificada estudiando las propiedades de los motivos en que están implicados, lo que puede ser críticamente importante a fin de mejorar las estrategias de mercadotecnia viral desde una perspectiva distinta a los enfoques existentes.

35 b. *Identificación Automática de Roles de Comunicación*: Las características de los motivos, tales como los tipos de motivos con los que está implicado un usuario, la posición (o rol) del usuario en el motivo, la proporción del usuario en grado incluyente o en grado excluyente, la cardinalidad del motivo y la frecuencia y distribución del motivo, pueden ser usadas para deducir relaciones entre pares de usuarios. Por ejemplo, un usuario A que está en el centro de un motivo en forma de estrella juega un rol distinto en la relación que un usuario B que está en la cola de un largo motivo en forma de cadena. Incluso dentro del mismo tipo de motivo, los usuarios con distintos grados incluyentes y / o grados excluyentes, implicados en interacciones más tempranas o más tardías dentro del motivo, desempeñan roles distintos en la comunicación. Explorando las propiedades de los usuarios individuales en el contexto de los motivos de comunicación, podrían mejorarse significativamente la precisión del rol y la identificación de la relación en la comunicación y las redes sociales.

45 c. *Optimización de costes de comunicación*: La correlación entre costes de comunicación y los motivos que se hallan en una red de comunicación puede ser usada para entender el rol que desempeñan los costes en la determinación de los patrones de comunicación entre los miembros de la red. Según aumenta el coste de la comunicación, los usuarios tienden a comunicarse solamente con un número reducido de sus contactos, creando por tanto motivos distintos (con un diámetro y grado incluyente más pequeños) que cuando el coste de la comunicación es bajo. Dadas estas correlaciones entre coste y motivos, los proveedores de redes pueden hacer predicciones eficaces del éxito de sus estrategias de mercadotecnia viral y maximizar por tanto la dispersión de información con restricciones de recursos. Además, en base al coste de la comunicación y sus patrones, podemos modelar adicionalmente los incentivos y las estrategias asociadas, ya sea para aumentar o para reducir la actividad y la cobertura dentro de la red. Gracias al descubrimiento automático y a la caracterización de los motivos, los proveedores de redes pueden tanto maximizar los ingresos como la influencia con un presupuesto limitado. Por ejemplo, un proveedor de servicios móviles puede ajustar la tarificación, dado el objetivo de

maximizar el nivel de actividad y, por tanto, los ingresos, mientras que un anunciante puede maximizar su influencia en la red, dado un cierto presupuesto, seleccionando los clientes que sean parte de motivos de flujo máximo de información.

Experimentos

5 Un análisis exhaustivo de dos conjuntos de datos reales de comunicación de redes sociales ha sido llevado a cabo a fin de estudiar las características de los motivos dentro de estas redes sociales.

Los dos conjuntos de datos reales de comunicación de redes sociales usados en los experimentos son:

1) Un conjunto de datos de Registros de Detalles de Llamadas (CDR) que ha sido recogido de una gran ciudad occidental durante tres meses, y para un proveedor específico de servicios móviles. Contiene 17.800.000 entradas (llamadas telefónicas) entre 245.600 usuarios únicos.

10 2) La historia de cartel mural de Facebook, recogida para la red de Facebook de Nueva Orleans durante más de dos años, con 800.000 interacciones entre 65.000 usuarios únicos.

Los motivos de comunicación están extraídos tanto de los CDR como de Facebook, y los motivos de máximo flujo de información de los CDR, usando la duración de cada llamada.

Tabla 3: Distribución global de motivos (W = 4 horas)

<i>Conjunto de datos</i>	<i>CDR</i>	<i>Facebook</i>	<i>Al azar</i>
Entradas aisladas	74,3%	69,7%	99,99%
Ping-pong de 2 personas	11,4%	14,5%	0,01%
Motivos de comunicación	14,3%	15,8%	0

15 La Tabla 3 presenta la distribución global de tres tipos de cimientos básicos de comunicación en tres redes (CDR, Facebook y al azar): entradas aisladas (es decir, entradas que no tienen otras entradas relacionadas dentro de la ventana temporal dada); motivos de ping-pong de 2 personas (entradas repetidas que implican solamente a 2 personas) y otros motivos de comunicación. La red al azar es una red de comunicación sintética que: (1) tiene exactamente la misma topología que el conjunto de datos de los CDR (es decir, el número de llamadas entre dos usuarios cualesquiera es idéntico), pero donde (2) los sellos horarios para todas las llamadas que se hacen dentro del periodo temporal dado están aleatorizados. Los algoritmos de extracción de motivos también fueron ejecutados sobre este conjunto de datos sintéticos. En todos los conjuntos de datos, podemos ver que la mayoría de las entradas están aisladas. Sin embargo, si bien tanto los motivos de ping-pong de 2 personas como los motivos de comunicación comprenden alrededor del 15% de las entradas totales en los conjuntos de datos de los CDR y de Facebook, tales motivos apenas pueden hallarse en las redes al azar. Este resultado sugiere marcadamente que los motivos restringidos temporalmente en las redes de comunicación social están guiados por las características especiales de las actividades humanas (tales como la propagación de información).

30 Como se ha mencionado anteriormente, el análisis de los datos está solamente basado en los motivos que implican a tres o más usuarios (representan alrededor del 15% de las interacciones en los conjuntos de datos).

Distribución de frecuencia de motivos

35 En primer lugar, se extraen los motivos de ambas redes sociales y se calcula cuán frecuentes son dentro de cada red. La Figura 6 muestra los seis motivos más frecuentes procedentes de los conjuntos de datos (obtenidos con una ventana temporal W = 4 horas). Obsérvese que en lo sucesivo el soporte se convierte en un porcentaje, basado en la razón entre la frecuencia y el número total de entradas en el conjunto de datos, de modo que los resultados entre los conjuntos de datos sean comparables. Es claro que hay motivos frecuentes que son compartidos por ambos conjuntos de datos, tales como las topologías de 2 cadenas y de 2 estrellas. Esto indica que los comportamientos de comunicación de la gente son similares incluso en entornos distintos. Por otra parte, también podemos observar algunas diferencias en las listas de motivos frecuentes: el conjunto de datos de Facebook contiene considerablemente más estructuras de estrella que el conjunto de datos de los CDR, mientras que el conjunto de datos de los CDR tiene más cadenas.

40 La Figura 7 muestra la correlación entre el soporte y la clasificación de los motivos. Aquí, obsérvese que cuanto más pequeño es el valor en eje de las X, más frecuente es el motivo. En ambos conjuntos de datos, y con distintos tamaños de ventana temporal (W = 4 horas en los CDR y W = 8 horas en los conjuntos de datos de Facebook), el soporte y la clasificación de los motivos siguen una distribución de ley de energía. Por lo tanto, según aumenta la categoría, podemos

esperar que el soporte de los motivos caiga abruptamente.

Impacto de la ventana temporal

5 A continuación, se examina el impacto de la ventana temporal W en el soporte de los motivos, variando su valor desde los 30 minutos a las 8 horas. Obsérvese que W impone una restricción sobre la proximidad temporal con la cual tienen lugar las interacciones. Como puede verse en la Figura 8, si bien crece el número absoluto de soportes según aumenta la ventana temporal, la distribución de soportes relativos de los motivos no es muy sensible al tamaño de la ventana temporal. Sin embargo, hay dos excepciones interesantes: el motivo de 2 estrellas en el conjunto de datos de los CDR y el motivo de 2 cadenas en el conjunto de datos de Facebook tienen razones decrecientes en la distribución del soporte según crece el tamaño de la ventana temporal. Este resultado, en particular, cuando se combina con la comparación de motivos con la red al azar en la Tabla 2, sugiere que los motivos restringidos temporalmente en las redes de comunicación social, aunque no son insensibles en general a la ventana temporal, están guiados por las características especiales de las actividades humanas (tales como la propagación de información).

Cobertura de usuarios

15 También fue llevado a cabo un análisis de cobertura de usuarios, a fin de medir la parte de usuarios implicados en algunos motivos. Si muchos usuarios están cubiertos, eso significa que los motivos son comúnmente vistos, o están uniformemente distribuidos, en las actividades de todos. La Figura 9 muestra que mientras los 10 motivos principales representan solamente alrededor del 8% del conjunto de datos totales, juntos cubren más del 30% de los usuarios en la red. Por lo tanto, creemos que los motivos son sub-estructuras significativas para representar redes sociales. La cobertura de usuarios de los motivos en la red de Facebook es levemente mayor que la de la red de los CDR, lo que puede indicar que más usuarios tienden a adoptar los motivos cuando necesitan diseminar información, en lugar de tan solo una comunicación casual.

20 Según se ha indicado anteriormente, la presente invención se refiere a un procedimiento para caracterizar las comunicaciones de redes sociales. Ha de entenderse que la divulgación precedente es una ejemplificación de los principios de la invención y que no limita la invención a las realizaciones descritas.

25

REIVINDICACIONES

1. Un procedimiento de caracterización de una comunicación de red, para cada interacción de comunicación entre dos usuarios, u_i^o , u_i^d , de dicha red, procedimiento que toma los siguientes datos como entrada:
- una anotación temporal t_i de cada interacción de comunicación, y
- 5 - una duración δ_i de cada interacción de comunicación,
- el procedimiento está **caracterizado por** ser consideradas una primera interacción de comunicación y una segunda interacción de comunicación dentro de la red para la caracterización de tal red y por formar parte de un grafo de comunicación C si:
- a) dichas interacciones de comunicación primera y segunda tienen al menos un usuario en común; y
- 10 b) el tiempo transcurrido entre el final de la primera interacción de comunicación y el comienzo de la segunda interacción de comunicación está dentro de una ventana temporal W pre-establecida;
- en el que cada grafo de comunicación C_i^j está formado por i usuarios únicos y j interacciones de comunicación que cumplen las condiciones a) y b) fijadas anteriormente.
- 15 2. Procedimiento de acuerdo a la reivindicación 1, que comprende además la etapa de separar por filtración cada grafo de comunicación con una restricción de frecuencia.
3. Procedimiento de acuerdo a la reivindicación 2, en el cual un grafo de comunicación es considerado para la caracterización si aparece el mismo tipo de grafo de comunicación más de un cierto número de veces en la red de interacciones.
- 20 4. Procedimiento de acuerdo a la reivindicación 3, en el que dos grafos de comunicación son considerados del mismo tipo, o equivalentes, cuando comparten la misma estructura de topología.
5. Procedimiento de acuerdo a la reivindicación 1, en el que, para un umbral de soporte Θ dado, cualquier clase de equivalencia de los grafos de comunicación $S = \{C_i^j(a), C_i^j(2), C_i^j(3), \dots, C_i^j(n)\}$ basada en la equivalencia de grafos con no menos de Θ grafos de comunicación es mencionada como un motivo de comunicación.
- 25 6. Procedimiento de acuerdo a la reivindicación 5, en el que cada motivo de comunicación es extraído usando un proceso basado en una primera búsqueda temprana EFS.
7. Procedimiento de acuerdo a cualquiera de las reivindicaciones 5 a 6, en el que un motivo de comunicación es un motivo de flujo máximo cuando maximiza el flujo de información dentro de la red.
8. Procedimiento de acuerdo a cualquiera de las reivindicaciones 1 a 7, en el que el tamaño de la ventana temporal depende del tipo de red.
- 30 9. Procedimiento de acuerdo a cualquiera de las reivindicaciones 1 a 8, en el que un grafo de comunicación tiene que aparecer más de un cierto número de veces en la red a fin de ser considerado.
10. Procedimiento de acuerdo a la reivindicación 1, en el que dicha anotación temporal t_i es un sello horario del comienzo de la interacción de comunicación.
- 35 11. Procedimiento de acuerdo a la reivindicación 1, en el que dicha anotación temporal t_i es un sello horario del fin de la interacción de comunicación.

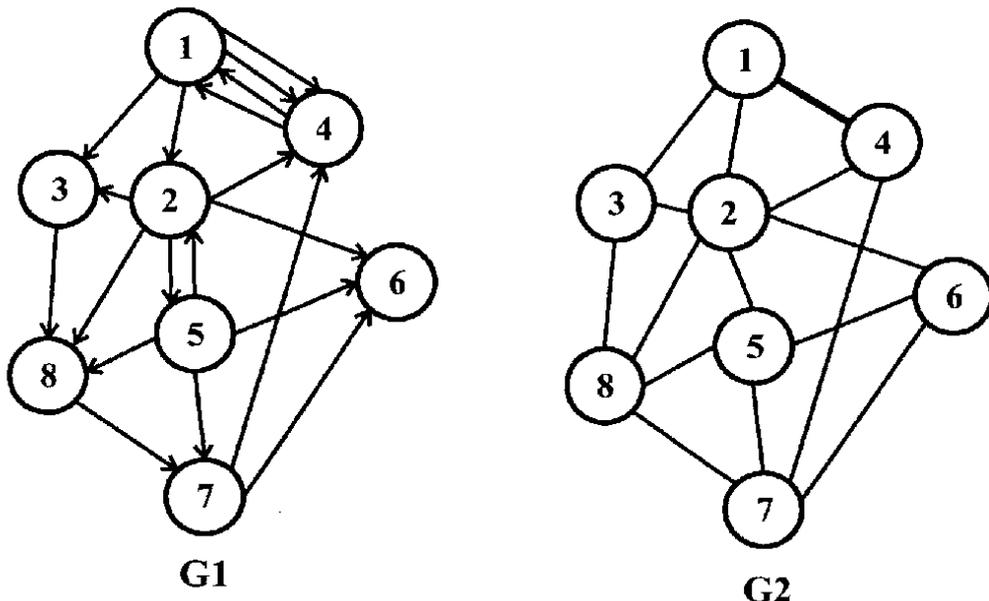


FIG. 1

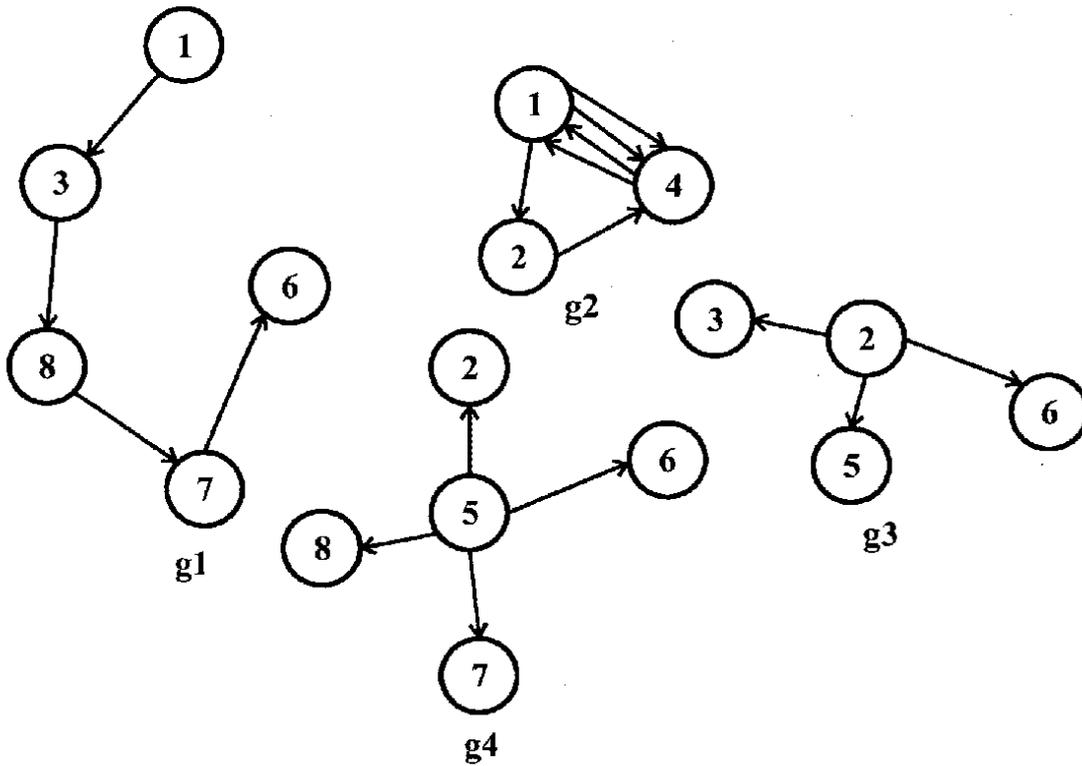


FIG. 2

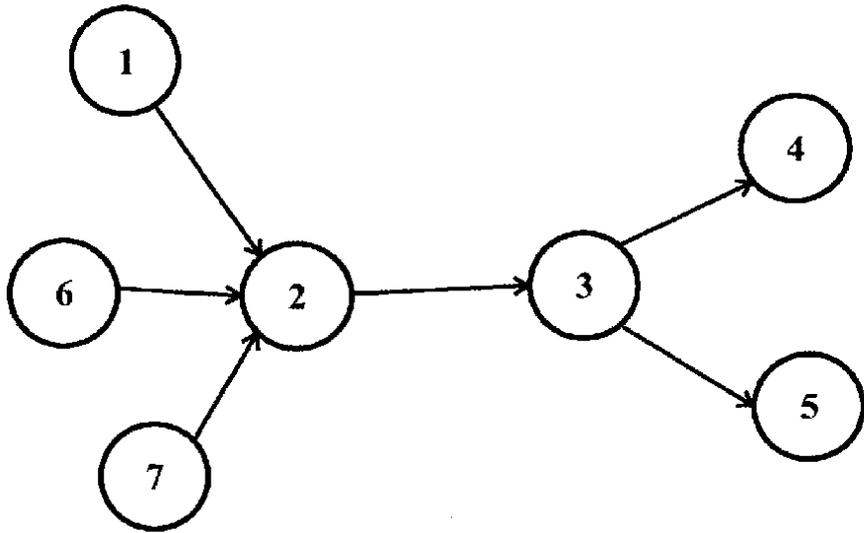


FIG. 3

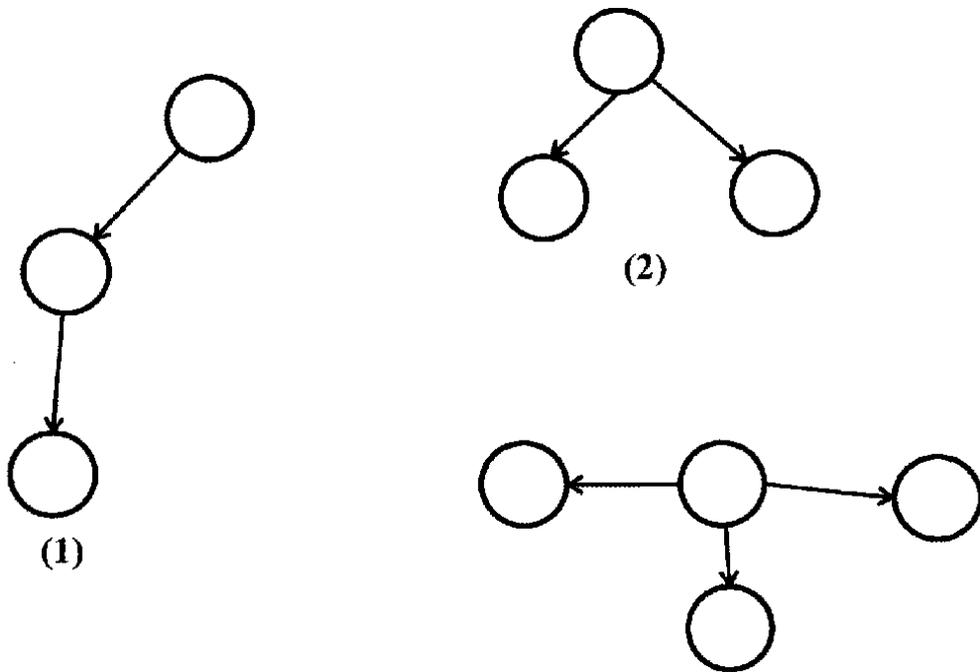


FIG. 4

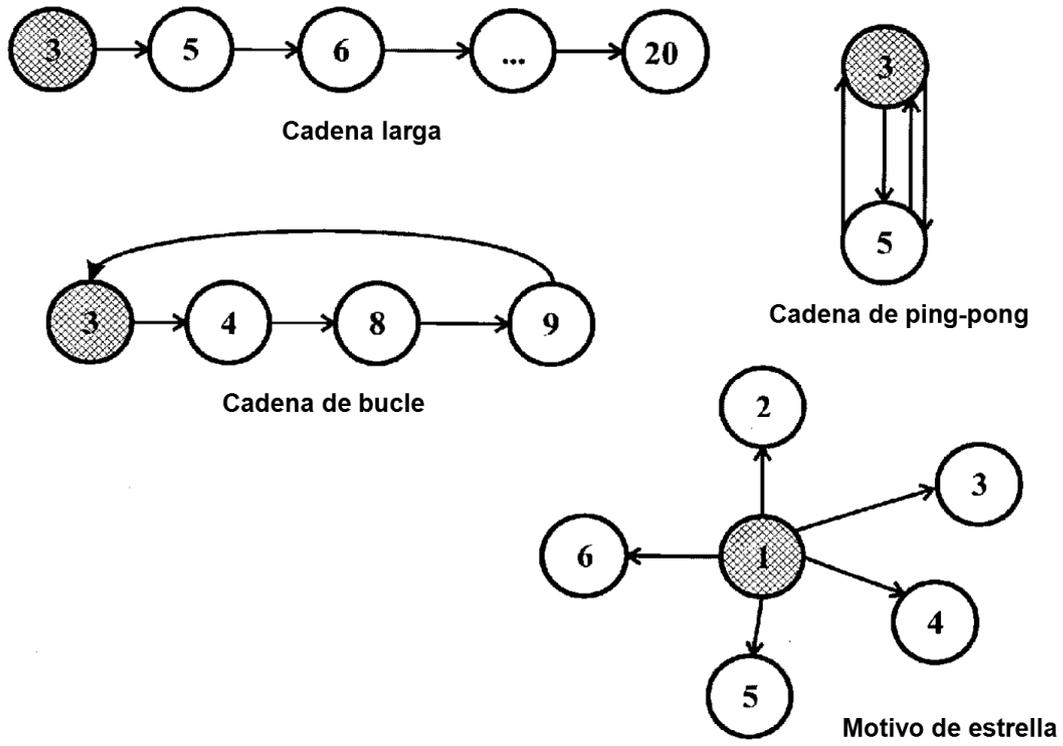


FIG. 5

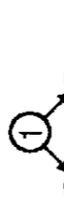
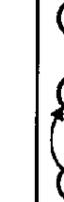
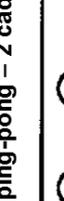
CDR	CDR (Máximo flujo)	Facebook
<p>43,4%</p>  <p>2 cadenas</p>	<p>42,4%</p>  <p>2 cadenas</p>	<p>38,7%</p>  <p>2 estrellas</p>
<p>29,6%</p>  <p>2 estrellas</p>	<p>30,9%</p>  <p>2 estrellas</p>	<p>30,1%</p>  <p>2 cadenas</p>
<p>5,2%</p>  <p>1 cadena - 2 ping-pong</p>	<p>5,8%</p>  <p>1 cadena - 2 ping-pong</p>	<p>7,7%</p>  <p>3 estrellas</p>
<p>4,6%</p>  <p>1 ping-pong - 2 cadenas</p>	<p>3,7%</p>  <p>1 ping-pong - 2 cadenas</p>	<p>6,5%</p>  <p>1 cadena - 2 ping-pong</p>
<p>3,2%</p>  <p>1 cadena - 2 estrellas</p>	<p>3,6%</p>  <p>3 cadenas</p>	<p>3,5%</p>  <p>4 estrellas</p>
<p>2,2%</p>  <p>3 cadenas</p>	<p>3,2%</p>  <p>1 cadena - 2 estrellas</p>	<p>2,4%</p>  <p>1 cadena - 2 estrellas</p>

FIG. 6

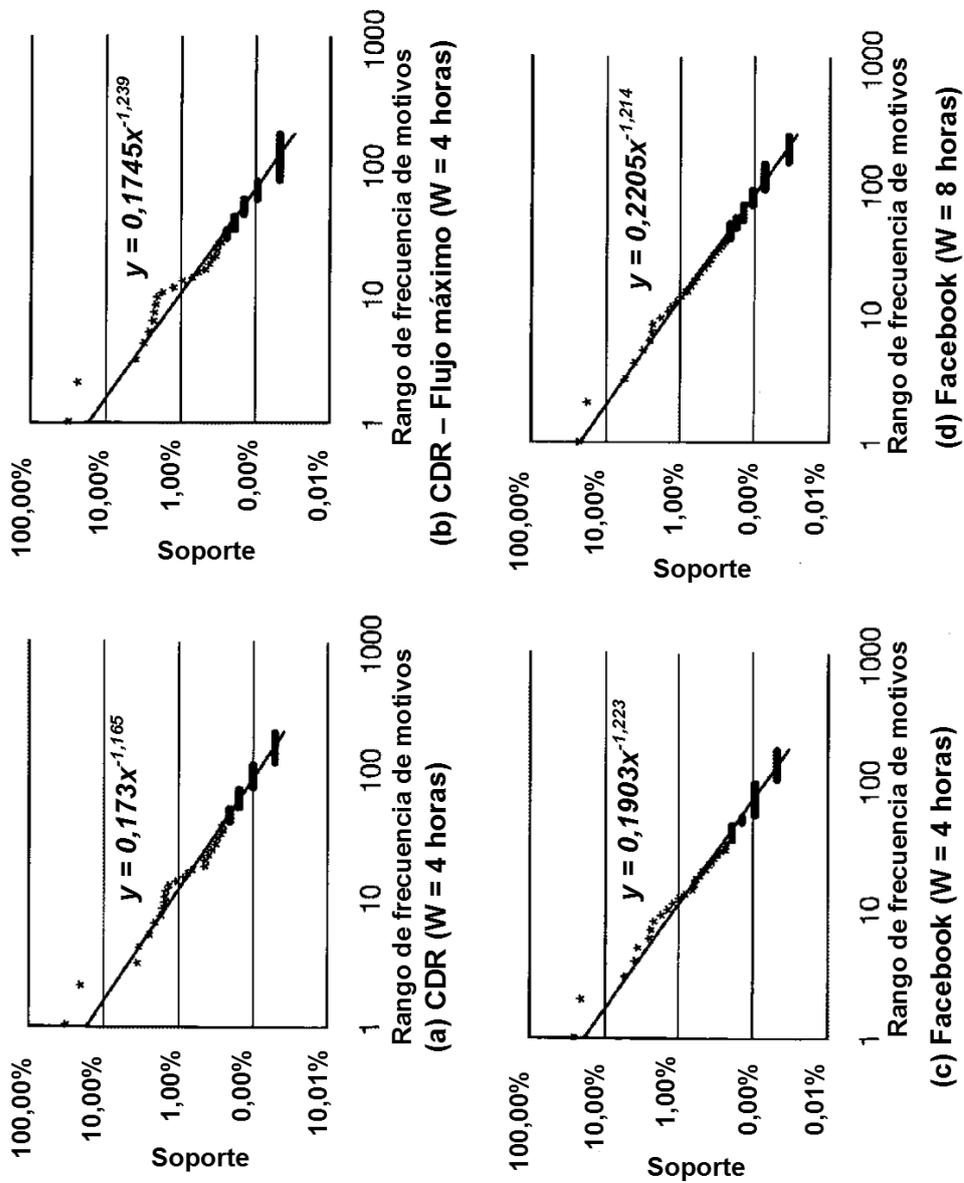
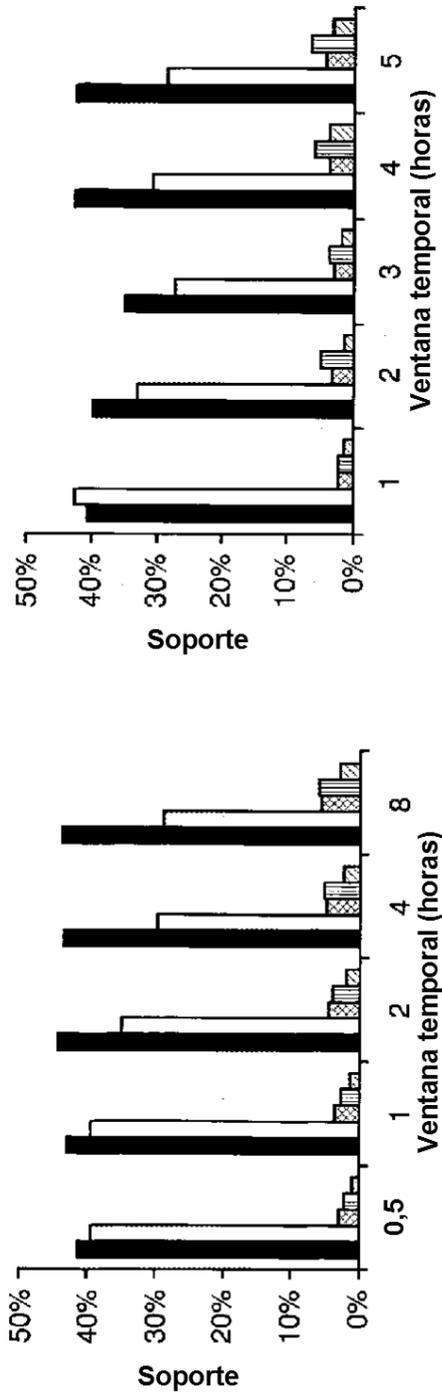
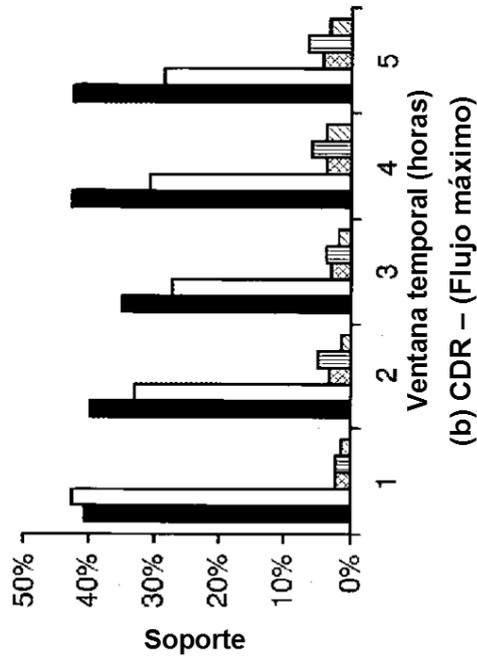


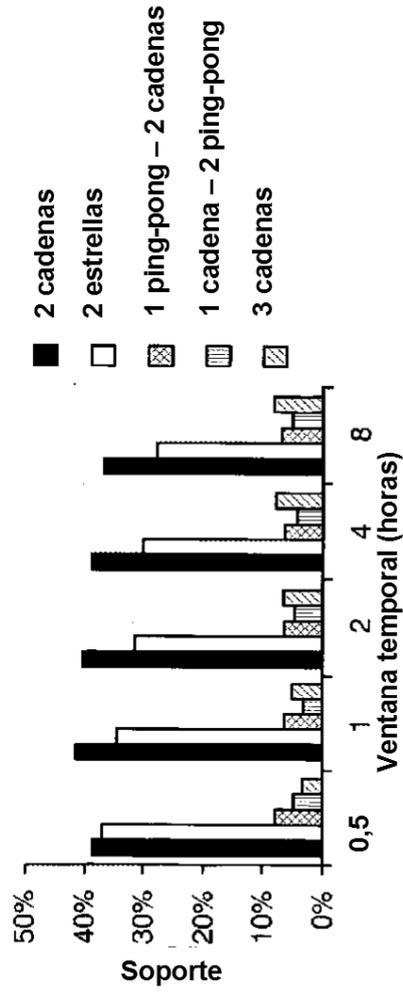
FIG. 7



(a) Conjunto de datos de los CDR



(b) CDR – (Flujo máximo)



(b) Conjunto de datos de Facebook

FIG. 8

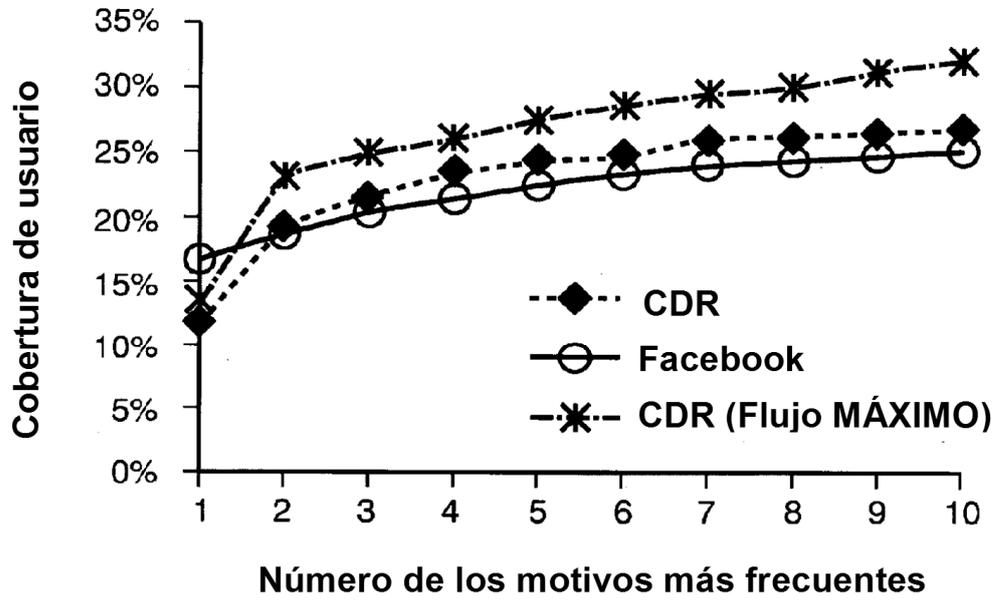


FIG. 9