

TESIS DE MÁSTER

Sistema multiagente para el análisis de interacciones entre usuarios en medios sociales

Author: Carla Martínez Poveda

Directores:

Dr. Vicent Botti Navarro y

Dra. Elena del Val Noguera

Departamento de Sistemas Informáticos y Computación,

Universidad Politécnica de Valencia,

Camino de Vera, s/n

46022 Valencia, Spain

Diciembre de 2016



UNIVERSIDAD
POLITECNICA
DE VALENCIA

Índice

1	Introducción	1
1.1	Motivación	1
1.2	Objetivos	2
1.3	Organización del documento	3
2	Estado del arte	5
2.1	Introducción	5
2.2	Agentes Inteligentes y Sistemas Multiagente	5
2.3	Redes complejas	8
2.3.1	Tipos de redes	9
2.3.2	Modelos de red	11
2.4	Redes sociales online	15
2.5	Análisis de Redes Sociales Online	17
2.6	Conclusiones	19
3	Diseño del SMA	21
3.1	Introducción	21
3.2	Modelo de Organización Virtual	22
3.3	Dimensión estructural	22
3.4	Dimensión funcional	24
3.5	Dimensión dinámica	25
3.6	Conclusiones	26
4	Framework	27
4.1	Visión global	27
4.2	Plataforma y agentes	29
4.2.1	Rol Monitor: agentes <i>AgentTweet</i> y <i>Neo4jAgent</i>	29
4.2.2	Rol Analyzer: agentes <i>QueryNeo4jAgent</i> y <i>GraphAnalyzerAgent</i>	32
4.2.3	Rol User: agente <i>VisualAgent</i>	34
4.3	Protocolos de interacción	34
4.4	Conclusiones	38

5	Caso de Estudio	39
5.1	Introducción	39
5.2	Medidas a nivel de red	40
5.3	Medidas a nivel de nodo	42
5.4	Análisis de eventos	45
5.4.1	Evento #MuereBarberáARV	45
5.4.2	Evento #ChampionsLeague	48
5.4.3	Evento #pesadillazamora	51
5.5	Conclusiones	54
6	Conclusiones y trabajo futuro	57
6.1	Conclusiones	57
6.2	Trabajo Futuro	58
6.3	Publicaciones	59
	Bibliography	61

Lista de Figuras

2.1	Ejemplos de tipos de redes según el tipo de información considerada. Red social: citas entre alumnos de un instituto [36]. Red biológica: relaciones entre enfermedades a partir de los genes que tienen en común [12]. Red tecnológica: trayectos entre aeropuertos [21]. Red de información: citación de artículos [3].	11
2.2	Proceso aleatorio de redirección de enlaces $n=20$, $k=4$ siguiendo el modelo propuesto por Watts [50].	12
2.3	Estructura de Kleinberg. La estructura inicial es una cuadrícula regular. Las conexiones aleatorias entre dos nodos distantes se establecen con una probabilidad inversamente proporcional al cuadrado de la distancia entre ellos.	13
2.4	Árbol de distancias sociales jerárquicas presentado por Watts [49] y Kleinberg [28]. Los individuos se agrupan considerando características sociales: trabajo, ubicación geográfica o intereses. En el ejemplo, l indica el número de niveles de la jerarquía ($l = 4$), b es el factor de ramificación ($b = 2$) y x indica la separación entre dos individuos.	14
3.1	Dimensión Estructural del SMA propuesto. Esta dimensión proporciona una descripción de los elementos que componen el sistema, es decir, qué roles y qué agentes forman parte de la organización virtual.	23
3.2	Dimensión Funcional del SMA propuesto. Esta dimensión proporciona información sobre los servicios proporcionados por los agentes que forman parte del sistema.	24
3.3	Dimensión dinámica del SMA propuesto. Muestra las interacciones entre las entidades que forman parte del sistema para realizar la recogida y almacenamiento de información asociada a un evento en una red social online (<i>FollowEvent</i>).	25
3.4	Dimensión dinámica del SMA propuesto. Muestra las interacciones entre las entidades que forman parte del sistema para realizar el análisis de la red de interacciones generada a partir de la información recogida de un evento en una red social online (<i>RequestAnalysis</i>)	26
4.1	Estructura general del sistema. Indica la dirección de los flujos de interacción y/o conversación entre los diferentes agentes del sistema.	29

4.2	Ejemplo de cómo se almacenarían los distintos tipos de mensajes de interacción asociados a un evento utilizando el modelo propuesto para almacenar la información en Neo4j. Los nodos del grafo representan las entidades del modelo. Las entidades consideradas en el modelo son: eventos (rojo), interacciones (verde), usuarios (azul). Los enlaces del grafo representan los tipos de relaciones que hay en el modelo de datos (i.e., creates, mention, has_hashtag, y creates). Las relaciones que se muestran en el grafo son las generadas al insertar un tweet donde no se menciona a ningún usuario (rosa), un tweet que contiene una mención (morado), un tweet que es un reply-to (amarillo), y una mención a varios usuarios (mention). En la parte inferior se muestran los cuatro mensajes (tweets) insertados en la base de datos.	31
4.3	Red de interacciones generada a partir de la actividad de los usuarios en Twitter asociada a un evento.	33
4.4	Diagrama de secuencia para el escenario <i>FollowEvent</i> . Describe las interacciones entre los agentes <i>VisualAgent</i> , <i>ManagerAgent</i> , <i>AgentTweet</i> y <i>Neo4jAgent</i>	35
4.5	Diagrama de secuencia para el escenario <i>RequestAnalysis</i> . Describe las interacciones entre los agentes <i>VisualAgent</i> , <i>ManagerAgent</i> , <i>QueryNeo4jAgent</i> y <i>GraphAnalyzerAgent</i>	36
5.1	Evolución de las propiedades estructurales a nivel de red: longitud del camino medio, grado medio de conexión y diámetro de la red. Evento #MuereBarberáARV	45
5.2	Evolución de las propiedades estructurales a nivel de red: modularidad. Evento #MuereBarberáARV	46
5.3	Evolución de las propiedades estructurales a nivel de red: densidad. Evento #MuereBarberáARV	46
5.4	Evolución de las propiedades estructurales a nivel de red: número de nodos, número de ejes y componente gigante. Evento #MuereBarberáARV	47
5.5	Evolución de las propiedades estructurales a nivel de red: longitud del camino medio, grado medio de conexión y diámetro de la red. Evento #ChampionsLeague	49
5.6	Evolución de las propiedades estructurales a nivel de red: modularidad. Evento #ChampionsLeague	49
5.7	Evolución de las propiedades estructurales a nivel de red: densidad. Evento #ChampionsLeague	50
5.8	Evolución de las propiedades estructurales a nivel de red: número de nodos, número de ejes y componente gigante. Evento #ChampionsLeague	50
5.9	Evolución de las propiedades estructurales a nivel de red: longitud del camino medio, grado medio de conexión y diámetro de la red. Evento #pesadillazamora	52
5.10	Evolución de las propiedades estructurales a nivel de red: modularidad. Evento #pesadillazamora	53
5.11	Evolución de las propiedades estructurales a nivel de red: densidad. Evento #pesadillazamora	53

5.12 Evolución de las propiedades estructurales a nivel de red: número de nodos, número de ejes y componente gigante. Evento #pesadillazamora	54
--	----



Lista de Tablas

2.1	Comparativa de las características principales de trabajos relacionados con el análisis de redes sociales online. Se han considerado las siguientes características: qué fuentes de información se han utilizado para generar la red, qué tipo de información extraída de las redes analizan, qué tipo de relaciones han considerado para generar las redes, qué medidas de análisis se tienen en cuenta, cuándo se realiza el análisis, y qué tipo de framework ofrecen.	20
4.1	Resumen de la implementación de los agentes que forman parte del framework propuesto. Se describe para cada agente del sistema: los roles que juega, los objetivos en los que participa, los servicios que ofrece, los agentes que crea, y los agentes con los que interactúa.	37
5.1	Evolución de las propiedades estructurales a nivel de nodo durante el evento #MuereBarberáARV. Indica los 5 perfiles que tienen valores de centralidad más altos teniendo en cuenta distintos criterios de centralidad.	48
5.2	Evolución de las propiedades estructurales a nivel de nodo: indica un top 5 de los principales roles que tienen lugar durante el evento #ChampionsLeague. . .	51
5.3	Evolución de las propiedades estructurales a nivel de nodo: indica un top 5 de los principales roles que tienen lugar durante el evento #pesadillazamora. . . .	54

Introducción

1.1 Motivación	1
1.2 Objetivos	2
1.3 Organización del documento	3

1.1 Motivación

En los últimos años, la forma en la que las personas se comunican está cambiando. Las redes sociales online como Facebook, LinkedIn o Twitter han supuesto una evolución significativa en la forma de generar y percibir la información. Cada vez más personas hacen uso de ellas y cualquiera puede compartir opiniones, difundir ideas, entablar amistades y mantenerse actualizado del mundo que nos rodea. Según una encuesta del portal Eurostat [18], durante el pasado año en 28 países de la UE, cerca del 50% de los usuarios de entre 16 y 74 años usaban internet para conectarse a redes sociales, sugiriendo un aumento de uso en éstas en los próximos años. Teniendo en cuenta estos datos, resulta relevante el análisis de la gran cantidad de información generada por estas redes sociales online para poder entender como se comportan sus usuarios en unas condiciones y circunstancias determinadas.

Los mensajes y las interacciones entre usuarios en redes sociales online dejan una huella digital, siendo posible analizar la información que contienen e incluso generar nueva información. Ésta información puede ser utilizada para analizar el comportamiento de los usuarios durante un evento y de qué forma se propaga la información dentro de las redes sociales online. Para facilitar la tarea, es necesario hacer uso de herramientas que recuperen, procesen y analicen dicha

información relacionada con un evento determinado en un intervalo de tiempo específico. Una forma de representar los datos extraídos de las redes sociales es mediante una red (conjunto de nodos y enlaces que representan los usuarios y sus lazos sociales entre ellos). Los lazos entre usuarios pueden ser de diversos tipos (amistad, interacción, profesión, etc) y pueden cambiar, alterando así la estructura de la red que se crea. Un análisis de la estructura de la red y de cómo evoluciona hacen posible el estudio de las dinámicas entre usuarios y nos ayuda a entender el comportamiento social dentro de entornos virtuales.

Actualmente, existen diferentes formas analizar la actividad de una red social online. La mayoría se basan en la realización de un análisis basado en datos estadísticos donde se determina qué usuarios tienen más conexiones dentro de la red, qué usuarios son más activos, o qué cantidad de información se genera. Hay otros trabajos que plantean el análisis de la actividad de una red social online mediante una red, pero en la mayoría de los casos sólo se centran en una medida y no tienen en cuenta cómo evolucionan en el tiempo. En este trabajo se plantea un enfoque diferente en la extracción y el análisis de información procedente de redes sociales. Concretamente, se presenta un modelo distribuido basado en sistemas multiagente que recoge información de la actividad en Twitter asociada a un evento, la procesa y realiza un análisis de las estructuras de red que se generan a partir de las interacciones entre usuarios. El análisis proporciona información sobre el comportamiento de los usuarios como grupo así como de manera individual.

1.2 Objetivos

Este trabajo se presenta dentro del proyecto “*Analysis and prediction of social behaviour in virtual social networks*”¹ y pretende realizar un análisis de actividad en redes sociales mediante el uso de la tecnología de los sistemas multiagente y redes complejas. El objetivo del mismo es diseñar y desarrollar un sistema multiagente que permita analizar la actividad de los usuarios en eventos de Twitter, generar redes etiquetadas temporalmente y extraer propiedades estructurales que nos permitan analizar cómo están interactuando los usuarios. Este objetivo se puede dividir en los siguientes subjetivos:

- Proponer un diseño de un SMA para el análisis de eventos en redes sociales. Para esto,

¹*Analysis and prediction of social behaviour in virtual social networks* (Ref: SP20140800). Entidad financiadora: Ministerio de Educación, Cultura y Deporte. Programa Campus de Excelencia Internacional.

describiremos la arquitectura propuesta desde su dimensión estructural (entidades que componen el sistema), funcional (servicios ofrecidos por el sistema) y dinámica (los roles y sus interacciones).

- Desarrollar un framework utilizando tecnologías de agentes para el análisis de la actividad de los usuarios en un evento en redes sociales online. El framework está basado en la plataforma de Sistemas Multiagente Magentix2.
- Proponer un modelo basado en redes temporales para modelar la evolución de las interacciones de los usuarios en redes sociales online.
- Realizar un análisis de las propiedades estructurales de redes a nivel global e individual para extraer elementos significativos que nos permitan entender el comportamiento de los usuarios en determinados eventos.

1.3 Organización del documento

El proyecto se desglosará en cuatro apartados principales:

- El capítulo 2 realiza una introducción donde se describen los conceptos básicos en el área de los Sistemas Multiagente y las redes complejas que están directamente relacionados con el proyecto. Posteriormente, para profundizar más en el proyecto propuesto y qué queremos aportar a lo que actualmente existe, se realiza un análisis de otros trabajos realizados para el análisis de información de redes sociales online.
- El capítulo 3 describe todo lo relacionado con el diseño del Sistema Multiagente propuesto para llevar a cabo la tarea de recogida, procesamiento y análisis de la actividad en redes sociales online. El diseño se presenta siguiendo tres dimensiones: estructural, funcional y dinámica.
- En el capítulo 4 se expone la implementación de un framework basado en el diseño presentado en el capítulo anterior. El desarrollo del sistema se realiza sobre la plataforma de agentes Magentix2. En este capítulo se detalla cómo se ha implementado cada uno de los agentes del sistema.
- En el capítulo 5 expone casos de estudio para mostrar la funcionalidad y aplicación del framework desarrollado. Se han considerado varios eventos deportivos y de televisión y se ha analizado la evolución de las propiedades de las redes estructurales temporales asociadas a cada evento.

- Finalmente, en el capítulo 6 se presentan las conclusiones del trabajo realizado y posibles líneas de trabajo futuro.

Estado del arte

2.1	Introducción	5
2.2	Agentes Inteligentes y Sistemas Multiagente	5
2.3	Redes complejas	8
2.4	Redes sociales online	15
2.5	Análisis de Redes Sociales Online	17
2.6	Conclusiones	19

2.1 Introducción

La propuesta que se presenta en este trabajo está centrada en dos áreas principalmente: el área de sistemas multiagente y el área de redes complejas. En este capítulo se describen los conceptos relacionados con los agentes inteligentes y sistemas multiagentes así como una introducción al área de redes complejas (i.e., tipos de redes según la información que representan y modelos de redes). Finalmente, describimos con más detalle qué consideramos por redes sociales online. En una segunda parte, se realiza un resumen y un análisis de otras propuestas que se han realizado para hacer frente al análisis de la actividad de los usuarios en este tipo de redes.

2.2 Agentes Inteligentes y Sistemas Multiagente

El término **Agente** se ha utilizado para denotar objetos físicos que pueden controlar su destino, tales como robots. Sin embargo, también hace referencia a un elemento abstracto, normalmente *software*, con un objetivo determinado. En este caso, nos referimos a los Agentes Software y,

concretamente, a los Agentes Inteligentes.

A lo largo de la literatura, existen varias definiciones para el término de **Agentes Inteligentes**. Hípola y Vargas-Quesada lo demoninan “*una entidad software que, basándose en su propio conocimiento, realiza un conjunto de operaciones destinadas a satisfacer las necesidades de un usuario o de otro programa, bien por iniciativa propia o porque alguno de éstos se lo requiere*”[24]. También se puede encontrar la definición dada por Jimenez y Ramos para los que un agente inteligente es una “*pieza de software que ejecuta una tarea dada utilizando información recolectada del ambiente, para actuar de manera apropiada hasta completar la tarea de manera exitosa. El software debe ser capaz de auto-ajustarse basándose en los cambios que ocurren en su ambiente de forma tal que un cambio en las circunstancias producirá un resultado esperado*”[23].

Con esto, se puede llegar a la conclusión de que un Agente es una entidad capaz de entender su entorno e interactuar con él de forma inteligente para beneficio de los usuarios, reuniendo la siguientes características básicas:

- **Autonomía:** el agente inteligente puede actuar sin ningún tipo de intervención humana directa y tiene control de sus propios actos.
- **Capacidad de reacción:** los agentes pueden detectar cambios en su entorno y reaccionar ante ellos.
- **Adaptabilidad:** los agentes, a partir de la capacidad de reacción, son capaces de adaptarse a los cambios del entorno.
- **Comunicación:** el agente es capaz de comunicarse por medio de un lenguaje común con otros agentes o incluso con las personas.
- **Iniciativa:** el agente llega a sus objetivos mediante las herramientas de las que dispone.
- **Continuidad temporal:** los agentes no sólo realizan ejecuciones en un momento determinado, sino que pueden pasar a un estado de espera hasta que cualquier evento provocado por otro agente, usuario o por su entorno les haga reaccionar.

Un agente inteligente esta diseñado especialmente para prestar servicio al usuario con el que se comunica a través de consultas. Para ello, posee al menos uno de los siguientes elementos:

- **Capacidad de proceso:** es capaz de descomponer las consultas en subconsultas y asociar los términos resultantes de esta operación con otros afines.
- **Conocimiento del entorno donde se mueve:** un agente debe saber en todo momento a qué información acceder o a qué otro agente dirigirse para obtener esta información.

A parte de las características anteriormente mencionadas, los agentes pueden presentar otras peculiaridades tales como la **movilidad**, que consiste en la capacidad para moverse de un entorno a otro por decisión propia, y la **cooperación**, que permite interactuar con otros agentes en beneficio del objetivo. Aquí es donde introduciremos el concepto de Sistema Multiagente.

Un **Sistema Multiagente** (SMA o también MAS *Multi-agent System*) es un sistema compuesto por más de un Agente Inteligente que actúan en el mismo entorno con el objetivo de dividir las tareas en pequeñas subtarefas para poder resolver problemas complejos que con un solo Agente sería costoso o inviable. De esta forma, la metodología de resolución del problema se agiliza mediante la cooperación de diferentes agentes que pueden realizar funciones diferentes e independientes entre sí. Existen dos enfoques a la hora de modelar un SMA:

- **Enfoque formal:** Es la forma clásica que consiste en dotar a los agentes de todo el mayor conocimiento que dispondrán, de forma que pueden resolver un rango de problemas definido por el usuario en razón de las capacidades cognitivas de los mismos agentes.
- **Enfoque constructivista:** En este caso, a los agentes se les dota de la inteligencia necesaria y, a través de los mecanismos programados, éstos pueden “aprender” y generar ellos mismos un comportamiento inteligente a partir de los problemas que se les planteen (comportamiento emergente). Este tipo de enfoque generalmente es el más complejo de implementar, porque requiere unos procesos complicados para poder asignar ese aprendizaje a los agentes y que éstos infieran sobre el entorno de forma mucho más ágil.

Una de las formas de modelar los sistemas multiagente donde colaboran un conjunto de agentes heterogéneos autónomos es a través del concepto de Organización Virtual (VO) [20]. Las VOs toman como base la Teoría de las Organizaciones Humanas [13] que permiten la descripción de los principales elementos de una organización: estructura (i.e., cuáles son las entidades del sistema), comportamiento (i.e., cómo interactúan entre ellas), y dinámica (i.e., cuál es la funcionalidad que ofrecen). El paradigma de las VOs ha sido empleado para el desarrollo de SMAs [19] y facilitan el desarrollo de sistemas con un alto nivel de abstracción que incorporan

tanto la perspectiva individual y organizacional así como su adaptación a los cambios que suceden en el entorno.

La propuesta de este trabajo intenta resolver el problema de la recogida, procesamiento y análisis de información mediante el diseño y desarrollo de un sistema multiagente donde cada uno de los agentes tiene un objetivo(s) individual(es) pero necesitan cooperar entre ellos para lograr el objetivo de la organización.

2.3 Redes complejas

El área de Redes Complejas está centrada en el estudio de las redes y que se basa en trabajos previos en el área de la teoría de grafos. Sin embargo, el área de Redes Complejas se diferencia de la visión más tradicional de la teoría de grafos en los siguientes aspectos [37]:

1. Ésta se centra en el estudio de las propiedades de las redes, no sólo desde un punto de vista teórico, sino también desde un punto de vista experimental. Las redes que se analizan emergen de una manera natural en el mundo real, evolucionando de una manera no supervisada y descentralizada (p. ej. World Wide Web, redes de transporte, redes de energía, redes económicas, colaboraciones científicas).
2. Las redes en el área de Redes Complejas se entienden, no sólo como la topología de los sistemas, sino también como un marco donde las entidades que forman parte de esas topologías interactúan [15].
3. En la teoría de grafos, las redes se asumían estáticas, sin embargo, en trabajos recientes en el área de Redes Complejas se ha tenido en cuenta la evolución de las redes respecto al tiempo de acuerdo a reglas dinámicas.

Una de las primeras aplicaciones conocidas al uso de las redes hace referencia a Euler y el problema de los puentes de Königsberg, cuyo planteamiento dio origen a la Teoría de Grafos tal y como la conocemos hoy en día. Sin embargo, el concepto de **Redes Complejas** no se estableció hasta finales de los años 50 cuando dos matemáticos, Erdős y Rényi, hicieron un gran avance en la teoría matemática clásica de grafos que revolucionaría la forma en que se pueden modelar estos problemas. Estos dos matemáticos describieron una red con una topología compleja por medio de un grafo aleatorio, estableciendo de esta manera los fundamentos de la teoría de redes aleatorias. Durante los últimos años, el estudio de las redes ha pasado del enfoque de análisis de pequeñas redes simples, sus propiedades y las propiedades de sus nodos y enlaces

a la consideración de propiedades estadísticas de redes a gran escala. Este nuevo enfoque ha sido impulsado por la disponibilidad de computadoras y redes de comunicación que permiten analizar datos en una escala mucho mayor de lo que era posible. Ahora no es raro ver estudios donde se analizan redes con millones o incluso miles de millones de nodos.

Una red es considerada una estructura compuesta por un conjunto de **nodos** y **enlaces**. Los nodos indican elementos que componen la red y los enlaces indican las conexiones que existen entre esos nodos. Los ejemplos de redes en contextos reales incluyen Internet, World Wide Web, redes sociales de conocidos u otras conexiones entre individuos, organizaciones, redes metabólicas, redes alimenticias, redes de citas entre artículos y muchas otras. La ubicuidad de las redes complejas (el hecho de encontrar características similares en muchas áreas de conocimiento) ha establecido un conjunto de problemas de investigación en diversos sectores. Es un hecho afirmar que las redes han sido utilizadas a lo largo de los años para denotar relaciones entre elementos y de esta forma estudiar su comportamiento y evolución. Ésto ha supuesto que sean elementos muy estudiados en ámbitos como las Matemáticas y las Ciencias Sociales.

2.3.1 Tipos de redes

Un conjunto de nodos unidos por enlaces es la versión más simple de red. Hay muchas maneras en que las redes pueden ser más complejas que esto. Por ejemplo, puede haber más de un tipo diferente de nodo en una red, o más de un tipo diferente de enlace. Los nodos y/o los enlaces pueden tener una variedad de propiedades, numéricas o de otro tipo, asociadas con ellos. Tomando como ejemplo de una red social de personas, los vértices pueden representar hombres o mujeres, personas de diferentes nacionalidades, lugares, edades, ingresos, o muchas otras cosas. Los enlaces pueden representar la amistad, pero también pueden representar similitud, o conocimiento o proximidad geográfica. También pueden llevar asociados pesos indicando el nivel de conocimiento que tiene una persona con la otra.

Teniendo en cuenta el tipo de sistema y los datos a representar, las redes se pueden clasificar en varios grupos: redes sociales, redes de información, redes tecnológicas, y redes biológicas [38] (ver Figura 2.1). A continuación describimos cada uno de estos grupos de redes.

Redes sociales. Una red social es un conjunto de personas o grupos de personas con algún patrón de conexión o interacción entre ellos. Los patrones de amistad entre individuos, las relaciones comerciales entre empresas, son todo ejemplos de redes que se han estudiado en el

pasado. Uno de los experimentos más conocidos es el de los “mundos pequeños” realizado por Milgram [47]. El experimento tenía como objetivo determinar la distribución de longitudes de trayecto en una red de conocidos. Para ello pidió a los individuos de la red pasar una carta a cada uno de sus conocidos hasta que alcanzara al destinatario. La mayoría de las cartas del experimento se perdieron, pero una cuarta parte de las cartas alcanzó el objetivo y pasó en promedio por las manos de aproximadamente seis personas. Este experimento fue el origen del concepto popular de “Seis grados de separación”.

Dentro de la categoría de redes sociales se encuentran las redes de afiliación en las que los grupos/individuos son los nodos y los enlaces son colaboraciones de un tipo u otro entre ellos. Un ejemplo clásico de este tipo de red es la red de colaboración de actores cinematográficos. Otra fuente de datos fiables para la generación de redes de colaboración son los registros de comunicación. Por ejemplo, se podría construir una red en la que cada enlace (dirigido) entre dos personas representaba una carta/email o paquete enviado por correo de uno a otro. Los autores Ebel, Mielsch y Bornholdt [16] reconstruyeron el patrón de comunicaciones por correo electrónico entre cinco mil estudiantes en la Universidad de Kiel a través de los registros almacenados por los servidores de correo electrónico. En esta red los nodos representan las direcciones de correo y los enlaces dirigidos representan un mensaje que pasa de una dirección a otra.

Redes de información. Las redes de información también son denominadas en algunos contextos como “redes de conocimiento”. El clásico ejemplo de una red de información es la red de citas entre artículos académicos [17]. La mayoría de los artículos citan trabajos previos de otros sobre temas relacionados. Estas citas forman una red en la que los nodos son artículos y un enlaces dirigidos van de un artículo A al artículo B indicando que A cita a B. La estructura de la red refleja la estructura de la información almacenada en sus nodos. Otro ejemplo muy importante de una red de información es el World Wide Web, que es una red de páginas Web que contienen información, unidas entre sí por hipervínculos de una página a otra [25].

Redes tecnológicas. Las redes tecnológicas son aquellas que han sido diseñadas por el hombre para la distribución de recursos como pueden ser la electricidad o la información. El caso de la red eléctrica es un buen ejemplo [40]. Otras redes de distribución que han sido estudiadas incluyen la red de rutas de líneas aéreas [51] y redes de carreteras, ferrocarriles [39] y tráfico peatonal [14]. La red telefónica y las redes de distribución tales como las utilizadas por la oficina de correos o empresas de entrega de paquetería también entrarían en esta categoría general. Otra red tecnológica muy estudiada es Internet, es decir, la red de conexiones físicas

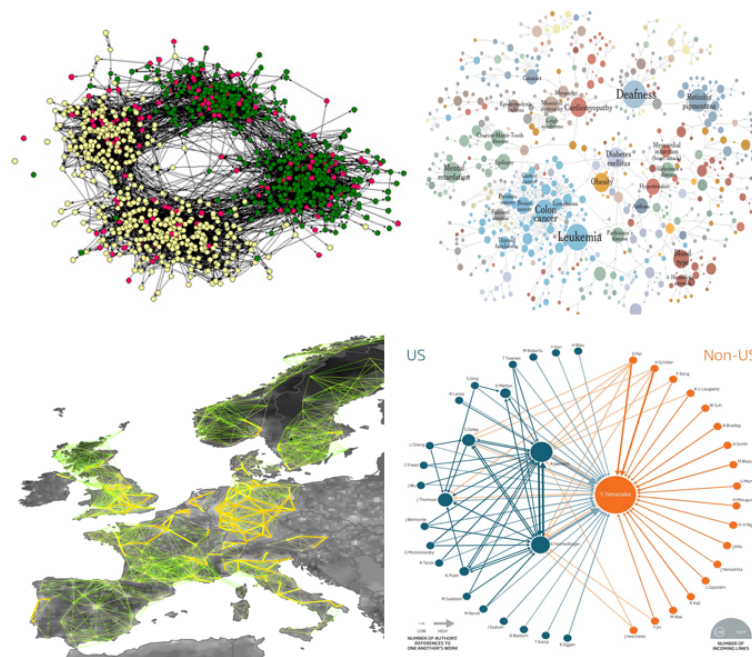


Figura 2.1: Ejemplos de tipos de redes según el tipo de información considerada. Red social: citas entre alumnos de un instituto [36]. Red biológica: relaciones entre enfermedades a partir de los genes que tienen en común [12]. Red tecnológica: trayectos entre aeropuertos [21]. Red de información: citación de artículos [3].

entre computadoras.

Redes biológicas. Los sistemas biológicos también pueden representarse en forma de redes. Las redes de reacciones metabólicas, las redes genéticas, los ecosistemas y cadenas tróficas, las redes neuronales o las vasculares son algunos de los ejemplos de redes biológicas analizadas desde la perspectiva de la teoría de redes. En el caso de las redes alimentarias, los nodos representan especies y los enlaces relacionan cada uno de esos nodos entre sí. El análisis del grado de salida y del grado de entrada de este tipo de redes posibilita la extracción de la complejidad e interconexión entre las distintas comunidades naturales [42].

2.3.2 Modelos de red

Se han propuesto varios modelos basados en funciones matemáticas para la generación de determinadas estructuras en redes reales. Estos modelos intentan reflejar cómo se establecen los vínculos entre los individuos de la red. En esta sección describimos dos de los modelos más importantes: *small-world* o *mundo pequeño* y *scale-free*.

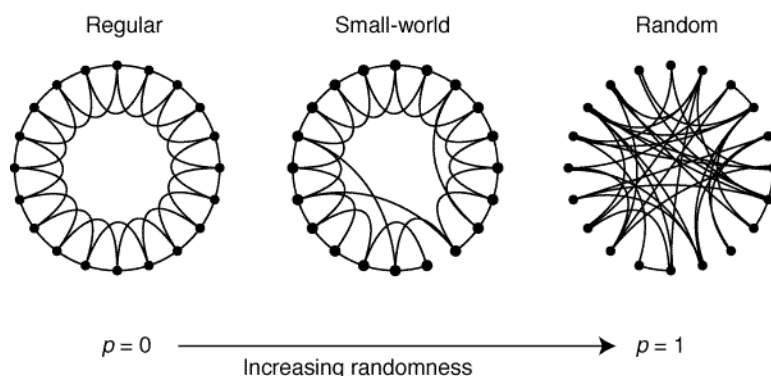


Figura 2.2: Proceso aleatorio de redirección de enlaces $n=20$, $k=4$ siguiendo el modelo propuesto por Watts [50].

Las redes de *small-world* se caracterizan por tener un alto número de enlaces entre nodos cercanos y unos pocos enlaces que conectan nodos que se encuentran distantes. Esta estructura tiene dos características principales: coeficiente de agrupamiento alto y trayectos cortos. El alto coeficiente de agrupamiento refleja que los vecinos de un nodo son también vecinos entre ellos (esto también se conoce comunmente con la descripción de los amigos de mis amigos son también mis amigos). La longitud corta de trayectorias se obtiene a través de las conexiones entre nodos que se encuentran distantes. Esta propiedad está relacionada con el diámetro de la red. El diámetro en las redes de *mundo pequeño* es exponencialmente menor que el tamaño de la red y puede ser limitado a $\log n$, donde n es el número de nodos. Este hecho hace que entre dos nodos en la red, por lo general, exista un camino corto.

Existen varios modelos que utilizan mecanismos aleatorios para la generación de propiedades de *mundo pequeño*. Podemos clasificar los modelos dependiendo de la estructura subyacente utilizada como base del modelo:

- Modelos rejilla-anillo: tienen una estructura regular (anillo, celosía o rejilla) donde se modifica alterando al azar algunos de los enlaces existentes o añadiendo nuevos para obtener características del mundo pequeño [29, 50].
- Modelos jerárquicos: son más realistas que los modelos regulares. En el proceso de generación, en lugar de utilizar una estructura regular (celosía o anillo), se utilizan estructuras jerárquicas que reflejan la estructura organizativa de un determinado dominio.
- Grid y modelos jerárquicos: toma características de los modelos anteriores para construir la red.

Watts y Strogatz [50] proponen un método para construir una red de *mundo pequeño* que

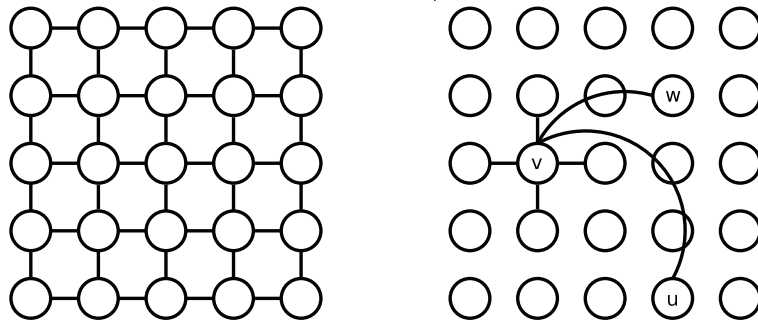


Figura 2.3: Estructura de Kleinberg. La estructura inicial es una cuadrícula regular. Las conexiones aleatorias entre dos nodos distantes se establecen con una probabilidad inversamente proporcional al cuadrado de la distancia entre ellos.

comienza con una red regular con n nodos situados en un anillo y k vecinos por cada nodo. Algunos enlaces (seleccionados al azar) se vuelven a reconectar con una probabilidad ρ , $0 \leq \rho \leq 1$ (véase Figura 2.2). Los autores concluyen que con valores intermedios de ρ se puede obtener una red con características de mundo pequeño (alto grado de agrupación y longitud de trayecto entre nodos pequeño).

Otro método para construir una red con propiedades de *mundo pequeño* fue propuesto por Kleinberg en [29]. La red se basa en un sistema de rejilla de $n \times n$ dimensiones donde todos los nodos de la red están conectados a los vecinos más cercanos (conexiones cortas). Entonces, las conexiones aleatorias largas son establecidas con una probabilidad p_{ij} que es inversamente proporcional al cuadrado de la distancia entre ellos:

$$p_{ij} \propto r_{ij}^{-\gamma}$$

donde r es la distancia de Manhattan (número de enlaces entre nodos) en la red entre los nodos i y j (véase la Figura 2.3). Kleinberg concluye que cuando γ es igual a la dimensión de la red, por medio de un algoritmo voraz se puede encontrar un nodo con caminos cuya distancia es inferior a $\log(n)^2$.

Kleinberg presenta un *modelo jerárquico* para la generación de redes con propiedades de *mundo pequeño* [28]. Primero, la jerarquía es modelada como un árbol que tiene n hojas. La distancia entre dos hojas u y v se define como la menor distancia al predecesor común $h(u, v)$ en el árbol. Basado en eso, Kleinberg define una red donde para un nodo v con grado k la probabilidad de establecer un enlace con un nodo w es proporcional a $b^{\beta h(v,w)}$ (véase Figura 2.4).

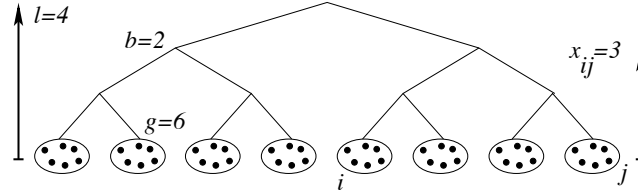


Figura 2.4: Árbol de distancias sociales jerárquicas presentado por Watts [49] y Kleinberg [28]. Los individuos se agrupan considerando características sociales: trabajo, ubicación geográfica o intereses. En el ejemplo, l indica el número de niveles de la jerarquía ($l = 4$), b es el factor de ramificación ($b = 2$) y x indica la separación entre dos individuos.

Otro modelo a considerar es la combinación de un modelo de grid basado en distancias y un modelo jerárquico. Esta idea es propuesta por Kleinberg en [27] donde define un nuevo modelo basado en una estructura de grupo. Kleinberg afirma que las personas que pertenecen al mismo grupo pequeño tienen más probabilidades de estar conectados. Por ejemplo, en un modelo de cuadrícula se pueden encontrar grupos (sub-cuadrículas) donde los nodos están más cerca. En las jerarquías, los grupos se pueden encontrar en subárboles donde dos nodos están dentro. El modelo se basa en la distancia entre dos nodos v y w que se define como el tamaño del grupo más pequeño que contiene ambos nodos. Un vínculo entre dos nodos v y w se establece con una probabilidad inversamente proporcional al tamaño del grupo más pequeño donde ambos nodos ($g(v, w)^\gamma$).

El otro grupo de modelos existentes que vamos a comentar es el de *Redes sin escala*. Este tipo de redes se caracterizan por el grado de los nodos que componen la red. El grado de la mayoría de los nodos es bajo mientras que unos pocos nodos presentan un alto grado de conexión. Esta característica se refleja en varias funciones matemáticas. La más conocida es la ley de potencias o *power-law*:

$$p_k \sim k^{-\alpha}$$

El parámetro α indica la tasa de decaimiento, un valor bajo de α significa un decaimiento más lento y un valor mayor de α significa un decaimiento más drástico del grado de conexión de los nodos.

Barabasi y Albert [5] presentan un modelo matemático para crear una red con características de *power-law*. El método para crear la red realiza lo siguiente: cuando un nuevo nodo llega a la red, tiene más probabilidad de establecer una conexión con nodos que están altamente conectados que con nodos que tienen pocas conexiones. La probabilidad de que un nodo que ya está en la red reciba un nuevo enlace es $p(k_i) = c k_i$, donde c es una constante de normalización

y k_i es el grado de conexión del nodo i .

Una vez revisado los tipos de redes existentes y los principales modelos estructurales podemos afirmar que en este proyecto el uso de redes nos ayuda a modelar escenarios complejos donde el número de actores que intervienen es elevado. Concretamente, en este trabajo las redes que se utilizarán para modelar las interacciones entre los usuarios durante un evento determinado serán redes de información donde los nodos serán las cuentas de los usuarios y los enlaces las comunicaciones entre ellos. A partir de estas redes, se analizarán propiedades como las que se han comentado en los modelos estructurales (longitud de caminos, grados de conexión, etc.) para comprender qué comportamientos comunicativos emergen.

2.4 Redes sociales online

En términos esenciales, una red social online es considerada una estructura compuesta por un conjunto de usuarios (individuos u organizaciones) relacionados en base a un criterio, tales como parentesco, amistad, relación profesional, etc. y que estas relaciones se han establecido utilizando un medio online. En los últimos años, en Internet han aflorado muchas comunidades locales en forma de redes sociales online y su número de usuarios activos se ha ido incrementando con el auge de las nuevas tecnologías y el acceso a los nuevos canales de información. Existen diferentes tipos de redes sociales online: de tipo profesional, relacionadas con la amistad, con aficiones, etc.

Las redes sociales online como Facebook, Twitter, LinkedIn o Flickr contienen millones de usuarios que comparten pensamientos, preferencias, creencias políticas así como establecen enlaces entre los usuarios. Cada una de sus contribuciones deja una huella digital y para poder realizar cualquier tipo de análisis es necesario procesar y gestionar ese volumen de datos. Actualmente, gracias a la tecnología disponible dentro del área de Big Data, estos datos pueden ser almacenados y procesados de una manera eficiente [32]. Este hecho permite que se pueda analizar y extraer información sobre el comportamiento de las personas y las interacciones entre usuarios.

El análisis de las redes sociales online tiene una aplicación directa en diversas áreas. Concretamente, el análisis de redes sociales virtuales ha sido utilizado para campañas de marketing on-line y estrategias de persuasión [31], diseño de interfaces de usuario, sistemas de recomendación, para determinar consumidores potenciales [1] o para determinar la personalidad de los usuarios a través de sus interacciones [22]. El análisis de las interacciones en las redes sociales facilita la comprensión de los flujos de información y la localización de usuarios tienen una

posición influyente en la red [44, 45]. Otros ámbitos donde se ha aplicado el análisis de las redes sociales son la personalización de los resultados de búsquedas basándose en los intereses de nodos vecinos en la red [10], así como para el terrorismo [43] o el cyberbullying [6].

En el trabajo presentado, nos centraremos principalmente en una red social online: Twitter. La razón por la que nos decantamos por esta red social y no por otra es porque Twitter proporciona un API que facilita el acceso a la información y los usuarios que utilizan esta red tienen un perfil público en su mayor parte.

Twitter [48] nació en el año 2006 y es una red social catalogada como un servicio de *microblogging*. Su principal característica es que permite al usuario publicar mensajes de texto plano de una longitud determinada, establecida en 140 caracteres. Dichos mensajes son denominados *tweets*. Éstos se muestran en la página principal del usuario y pueden estar sometidos a ciertas acciones por parte del resto de los usuarios. Dichas acciones son las siguientes:

- Cualquier usuario puede realizar **comentarios** en los *tweets* (que no sobrepase el máximo establecido). A través de estos se puede **citar** directa o indirectamente a otro usuario.
- De la misma forma, un tweet puede ser **retweeteado** (difundir sin cambiar el contenido por otro usuario) o **citado** (hacer referencia a él, pero con un texto o usuario que lo acompañe).
- Un usuario puede indicar afinidad por el contenido del tweet con la opción de “**Me gusta**”. Sin embargo, con esta acción no se establece ningún tipo de lazo en la red social, simplemente indica una preferencia.

Los **usuarios** son los encargados de subir contenido a la red, abarcando innumerables temas y discusiones. Los usuarios que comienzan a seguir a un usuario, pasan a ser denominados *followers* de ese usuario. Un usuario puede conectarse con otro a través de mensajes públicos (los tweets) o por **mensajes directos**, que es un tipo de mensajería privada disponible dentro de Twitter.

Hay muchas maneras de asociar un tema en Twitter, pero la manera más fácil es mediante los **hashtags**. Los hashtags son etiquetas de texto precedidas por un símbolo hash (#) que contiene el nombre, frase o descripción del tema. Además, gracias a los hashtags es posible asociar **tendencias** o establecer *trending topics*, que son hashtags muy utilizados en Twitter por los usuarios en un momento determinado. Por ejemplo, para las Elecciones Generales del 20N, los

hashtags utilizados para seguir el evento fueron *#EleccionesGenerales2015* y *#Elecciones20D*, ambos trending topic del momento durante los meses previos al evento y después de éste.

Twitter ha sido usado para gran variedad de propósitos, en diferentes situaciones y lugares. Sin embargo, ha tenido gran relevancia a la hora de seguir determinados eventos considerados “de actualidad” o “candentes”. Un ejemplo de ello es que fue usado para organizar y reportar diferentes protestas como la de Egipto en 2011 [11] o la conocida Primavera Valenciana de 2012 [2].

2.5 Análisis de Redes Sociales Online

Cómo se comportan los usuarios en las redes sociales online, cuan complejas son y cómo se establecen a través de la comunicación y la interacción digital han sido y sigue siendo temas de interés en distintas áreas de conocimiento. Hablar de complejidad es un hecho, ya que cuanto mayor sea el número de personas en una red social online, mayor es el número de conexiones y posibles relaciones entre individuos.

A través de las redes sociales online, y concretamente en Twitter, se le brinda a los usuarios la posibilidad compartir experiencias, divulgar opiniones y establecer lazos de “amistad”. Sin embargo, desde el punto de vista del análisis estas redes, Huberman et. al [26] consideraron que muchas de las relaciones de “amistad” son irrelevantes , ya que hay usuarios que pueden seguir o ser amigos de otros usuarios sin que hayan interactuado nunca y que por tanto no haya un vínculo real entre ellos. Teniendo en cuenta este hecho, lo que realmente interesa es conocer la red social que emerge considerando las interacciones entre usuarios.

Varios trabajos se han centrado en estudiar y medir las características de diferentes redes sociales online. Mislove et. al [35] realizaron un estudio de cuatro redes populares: *Flickr*, *Youtube*, *LiveJournal* y *Orkut*, con el objetivo de analizar la estructura de los grafos generados por las relaciones entre individuos. Se demostró que cada red tenía una estructura social diferente, y sin embargo, a pesar de su crecimiento incremental, dicha estructura permanecía invariable. Así mismo, también se llegó a la conclusión de que la interacción entre usuarios tenía un alto grado de reciprocidad: los usuarios participaban y respondían mensajes entre ellos.

Otros trabajos están relacionados con el análisis de la red social online Twitter debido a su facilidad para recuperar información. Estos trabajos estudian y analizan la estructura de su red. Smith *et al.* [46] estudiaron las redes creadas en conversaciones políticas en Twitter y las

clasificaron en seis estructuras o arquetipos diferentes que difieren entre ellas en función de dos factores: el tema a tratar y las personas que intervienen en su conversación. Por otra parte, Lothan *et al.* [34] se centraron en el análisis de las interacciones de usuarios y cómo se generan sus flujos de datos en eventos políticos tales como la revolución egipcia y tunecina del año 2011.

La mayoría de las redes sociales online del mundo real evoluciona con el tiempo y genera una enorme cantidad de datos temporalmente anotados. Actualmente, muchos de los trabajos como los anteriormente descritos se centran en analizar las propiedades en un determinado momento. Un número limitado de trabajos se han centrado en estudiar en profundidad la evolución de las estructuras que emergen en redes sociales online. Hay trabajos previos que analizan la evolución en el tiempo de algunas propiedades estructurales o métricas. Por ejemplo, Meeyoung *et al.* [35] considera las relaciones de amistad y *likes* en *Flickr*. Los autores realizan un análisis que incluye la topología de la red (es decir, la longitud de los caminos, el coeficiente de agrupamiento, el diámetro y la distribución del grado). Este análisis es realizado al final de un período de tiempo. Hay una métrica que si que es analizada en el tiempo: el número de seguidores de la foto. Benevenuto *et al.* [7] consideran las redes sociales de *Orkut*, *MySpace*, *Hi5* y *LinkedIn*. Su análisis se centra en datos detallados de los clicks que realizan los usuarios de redes sociales. Analizan métricas relacionadas con las sesiones de los usuarios en las redes sociales. La métrica que se analiza en el tiempo es el número de usuarios únicos en línea. Kumar *et al.* [30] presentaron un análisis evolutivo más completo de las propiedades estructurales en *Flickr* y *Yahoo! 360*. Sin embargo, el análisis se realiza a partir de la evolución de la red de amistades la cual no proporciona una visión realista de las relaciones “reales” que hay entre los usuarios.

Borge-Hoelthofer *et al.* [8] estudió los patrones estructurales y dinámicos de una red formada por usuarios de Twitter durante el fenómeno social 15M en España. Analizan la evolución de la componente gigante, los tweets recibidos y enviados, y la popularidad. Hay otras propiedades tales como distribuciones de grado y la estructura de las comunidades que sólo se analizan en un punto específico del tiempo. Hay otros trabajos que se centran en la comprensión de las interacciones de los usuarios relacionados con los acontecimientos políticos y analizan la evolución de algunas propiedades en puntos específicos del tiempo [9, 34, 41]. Borondo *et al.* [9] analizan las propiedades topológicas de las redes de menciones y retweets, la distribución acumulativa de la actividad de los usuarios y el porcentaje de retweets y menciones dirigidas a los políticos y a las cuentas oficiales de los medios de comunicación al final del evento. Sólo realizan un análisis de la evolución de los tweets acumulados en el tiempo. Lotan *et al.* [34] analizan las revoluciones tunecina y egipcia en Twitter. Analizan la distribución de los tipos de actores sobre los usuarios que participan en los eventos, la participación de las cuentas oficiales

frente a las cuentas individuales y la información sobre el número de mensajes generados en ambos eventos. El análisis de la evolución de las propiedades de los eventos a medida que pasa el tiempo se centra sólo en el número de tweets. Peña et. al [41] toma como referencia el mismo evento del 15M y lo compara con el 25S. Los autores analizan la participación de los usuarios en los dos eventos, la distribución de los tipos de mensajes, la caracterización demográfica y la detección de comunidades una vez terminado el evento. Además, los autores aportan varias cifras que muestran la evolución de las relaciones establecidas en cada etapa del 15M.

Resumiendo, las obras existentes analizan la mayoría de las métricas asociadas a la actividad de los usuarios al final del evento (ver Tabla 2.1). En algunos de ellos, hay una propiedad específica que se analiza durante un período de tiempo, pero en general, las propiedades se analizan sin tener en cuenta su evolución. En los enfoques presentados por Benevenuto et. Al, Kumar *et al.* y Meeyoung et. al, los autores se centran en las relaciones de amistad en lugar de las interacciones, que proporcionan una visión menos realista de dichas relaciones. Nuestra propuesta trata de superar estas limitaciones proporcionando un marco multiagente que automatiza la recuperación de la información y un análisis en profundidad de la evolución de las métricas asociadas a un evento en las redes sociales online. El análisis se puede realizar en tiempo real y después del evento. El análisis considera la interacción entre los usuarios como los enlaces de las redes generadas. El análisis que ofrece nuestra propuesta puede ser configurable y proporcionar diferentes niveles de detalle (es decir, análisis a nivel de red e individual). El uso de la tecnología sistemas multiagente facilita la escalabilidad y la distribución de la recopilación de datos, el procesamiento y el análisis de la actividad de los usuarios.

2.6 Conclusiones

Las redes sociales pueden considerarse procesos dinámicos donde, con el paso del tiempo, los individuos se unen, abandonan, crean o desactivan los lazos sociales, alterando así la estructura de la red. El análisis de la evolución de las redes de interacción en tiempo real se considera una valiosa fuente de información sobre el comportamiento social humano. El monitoreo de la actividad de los usuarios en tiempo real y la realización de un análisis es, sin embargo, una tarea compleja para los seres humanos debido al conocido problema de sobrecarga de información. El principal problema es cómo automatizar el proceso de monitoreo de las interacciones de los usuarios con la intención de detectar patrones de comportamiento en usuarios y/o grupos de una manera distribuida y eficiente para manejar la demanda de los usuarios.

Con el fin de hacer frente a este problema, en este trabajo se propone el uso de sistemas multi-

Aproximación	Qué analiza?	Enlaces de la red	Dónde?	Cuándo?	Qué?	Framework de análisis configurable
Meeyoung et. al	propiedad única: fans asociadas a fotos (por día)	amistad y favoritos marcados	Flickr	después del evento	longitud del camino, coeficiente de clustering, diámetro, distribución de los grados, fans por día.	✗
Benevenuto et. al	propiedad única: usuarios on-line (por hora)	datos click-stream	Orkut, MySpace, Hi5, and LinkedIn	después del evento	métricas relacionadas con sesiones de usuario	✗
Kumar et. al	✓ (por semana)	amistad	Flickr Yahoo! 360	después del evento	singletons, componente gigante, región intermedia	✗
Borge-Hoelthoefer et. al	✓ componente gigante tweets recibidos /enviados popularidad	interacción de eventos: 15M	Twitter	después del evento	Crecimiento de la red, comunidad estructura, fuentes de información y sumideros	✗
Borondo et. al	✓ tweets acumulados	interacciones en las elecciones presidenciales en España	Twitter	después del evento	actividad de usuarios, propiedades topología, mensajes orientados a políticos.	✗
Lotan et. al	solo número de tweets durante ese tiempo	interacciones	Twitter	después del evento	tipos de actores, participación organizativa vs individuos, núm. de mensajes	✗
Peña-Lopez et. al	solo relaciones durante ese tiempo	interacciones	Twitter	después del evento	tipo de mensajes, demografía, detección de comunidades	✗
Propuesta	✓ (configurable)	interacciones	Twitter	durante y después	a nivel de red, a nivel de nodo	✓

Tabla 2.1: Comparativa de las características principales de trabajos relacionados con el análisis de redes sociales online. Se han considerado las siguientes características: qué fuentes de información se han utilizado para generar la red, qué tipo de información extraída de las redes analizan, qué tipo de relaciones han considerado para generar las redes, qué medidas de análisis se tienen en cuenta, cuándo se realiza el análisis, y qué tipo de framework ofrecen.

agente, ya que éstos pueden ayudar a comprender las redes sociales. Ésto es debido a que los SMAs son una tecnología apropiada para hacer frente al análisis distribuido de la evolución de la actividad de los usuarios en las redes sociales online. Además, son adecuados para implementar sistemas de software complejos y distribuidos. Los agentes como entidades de computación social pueden colaborar en el proceso de recopilación, procesamiento y análisis de información de forma distribuida. Ésto hace que los SMAs puedan adaptarse a los cambios en el entorno mientras siguen logrando los objetivos generales del sistema frente a la alta dinámica de Internet. De esta forma, proporcionan una base para crear una arquitectura que se ocupe de la reducción de la complejidad, la flexibilidad y la escalabilidad necesarias en el análisis automático de la actividad de los usuarios en eventos en redes sociales en línea.

Diseño del SMA

3.1	Introducción	21
3.2	Modelo de Organización Virtual	22
3.3	Dimensión estructural	22
3.4	Dimensión funcional	24
3.5	Dimensión dinámica	25
3.6	Conclusiones	26

3.1 Introducción

La tecnología de agentes se considera adecuada para dar soporte a los paradigmas de computación donde ésta es considerada como una actividad inherentemente social, en lugar de solitaria. Si se extrapola al análisis de la actividad de los usuarios en las redes sociales online, esta tecnología permite a las entidades computacionales interactuar para lograr un objetivo en un entorno totalmente dinámico en el tiempo. En este capítulo, se presenta una descripción detallada del diseño del sistema propuesto. Dicho diseño está definido en base al paradigma de organizaciones virtuales, donde mediante un conjunto de dimensiones (estructural, funcional y dinámica) se describen el conjunto de roles del sistema (es decir, monitor, analizador, administrador y usuario), qué servicios son ofrecidos por cada rol, y cómo interactúan las entidades que forman parte del sistema.

3.2 Modelo de Organización Virtual

La arquitectura propuesta se basa en una **organización virtual (VO)** que permite la descripción de los principales aspectos de una organización: estructura, comportamiento y dinámica. Un modelo basado en VOs ofrece la posibilidad de definir cuáles son las entidades (roles principales) del sistema, cómo se relacionan entre ellas, cuál es su funcionalidad, qué servicios necesitan y/o ofrecen, etc. La principal ventaja de un modelo basado en VOs es que está dirigido hacia entornos abiertos donde agentes heterogéneos y autónomos trabajan colaborando entre ellos.

El diseño de la arquitectura del SMA propuesto se ha realizado bajo la metodología GORMAS [4]. Basada en la Teoría de las Organizaciones, y dado que el concepto de organización ha adquirido gran relevancia en el área de los sistemas multiagente, esta metodología facilita el análisis y diseño de mecanismos de coordinación y colaboración complejos en sistemas abiertos. El modelo de organización GORMAS permite describir los principales aspectos de las organizaciones: estructura, funcionalidad, normalización, dinamicidad y entorno. La propuesta que nosotros planteamos está basada en una VO (SocialNetAnalyzer) que considera tres elementos fundamentales:

1. **Dimensión estructural:** Describe los elementos del sistema.
2. **Dimensión funcional:** Detalla la funcionalidad del sistema.
3. **Dimensión dinámica:** Describe los elementos de interacción del sistema.

A continuación se describe con detalle cada una de las tres dimensiones definidas en la VO.

3.3 Dimensión estructural

La dimensión estructural describe el conjunto de componentes del sistema (roles y agentes) y las relaciones entre ellos. El modelo define una VO (SocialNetAnalyzer) como una entidad social que está formada por el conjunto mínimo de agentes que juegan unos roles y, en función a éstos, llevan a cabo una tarea determinada siguiendo un protocolo de comunicación predefinido. La VO definida juega un papel de punto de encuentro virtual donde los agentes pueden entrar y salir de ella de forma dinámica mediante la adopción de los roles definidos.

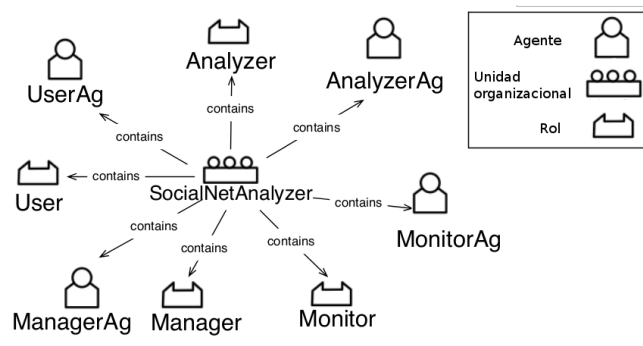


Figura 3.1: Dimensión Estructural del SMA propuesto. Esta dimensión proporciona una descripción de los elementos que componen el sistema, es decir, qué roles y qué agentes forman parte de la organización virtual.

La VO propuesta presenta una topología basada en una jerarquía simple, en la cual existe un agente supervisor (*Manager*) que controla a otros miembros. Una VO define una serie de roles que pueden ser adoptados por agentes miembros. En *SocialNetAnalyzer* se han definido cuatro roles diferentes: *User*, *Manager*, *Monitor*, *Analyzer* y un agente por cada rol (ver figura 3.1).

- El agente *User* es el intermediario entre el sistema y el usuario final. El usuario interactúa con este agente a través de la interfaz gráfica del programa y le indica a éste el evento que quiere seguir o analizar. El agente *User* recibe la información y se la comunica al agente *Manager*.
- El agente *Manager* es el responsable de habilitar y gestionar todos los agentes del sistema. Indica a cada uno de ellos en qué momento debe realizar su función y les aporta los datos necesarios para que realicen la tarea (como, por ejemplo, el nombre del evento a analizar).
- El agente *Monitor* se encarga de recuperar la información, en este caso los tweets relacionados con ese evento que el usuario ha indicado inicialmente. Cuando la información se recupera, ésta se almacena en una base de datos a la espera de su reconstrucción como grafo y a su análisis estadístico por parte del agente *Analyzer*.
- El agente *Analyzer* tiene la tarea de crear la red, extraer la información necesaria, sacar las estadísticas y mostrarlas a través de la interfaz gráfica. Esta acción se realiza bajo demanda del usuario.

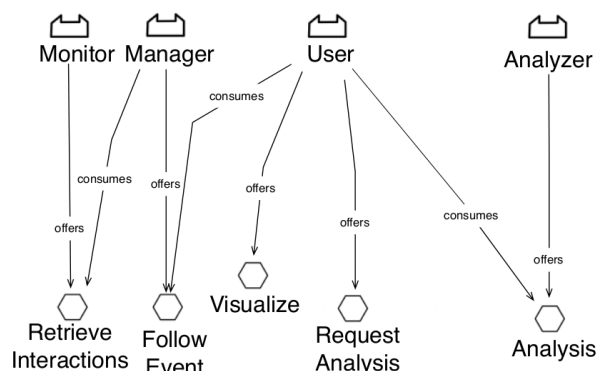


Figura 3.2: Dimensión Funcional del SMA propuesto. Esta dimensión proporciona información sobre los servicios proporcionados por los agentes que forman parte del sistema.

3.4 Dimensión funcional

Cada rol definido en la unidad organizacional *SocialNetworkAnalyzer* ofrece y/o consume un conjunto de servicios que define la funcionalidad del sistema. Ésta se describe de la siguiente manera (ver figura 3.2):

- *RequestAnalysis*: Es un servicio ofrecido por los agentes *User* con el objetivo de obtener información relacionada con el análisis de la evolución de un evento. Esta información es procesada y mostrada al usuario a través de una interfaz web.
- *FollowEvent*: El agente *User* ofrece este servicio que consiste en comenzar a seguir el evento y recopilar tweets asociados con el mismo. Este mismo agente se encarga de enviar un mensaje al agente *Manager* para añadir los agentes *Monitor* necesarios.
- *RetrieveInteractions*: Este servicio está ofrecido por el agente *Monitor* y consiste en la recolección de los mensajes asociados con el evento en Twitter.
- *Analysis*: Este servicio es el responsable del análisis de las interacciones del usuario en un evento. Lo proporciona el agente *Analyzer* y dicho análisis, una vez efectuado, se dirige al agente *User*.
- *Visualize*: Lo proporciona el agente *User* y el servicio consiste en la visualización de los resultados del análisis proporcionado por el agente *Analyzer* al usuario final.

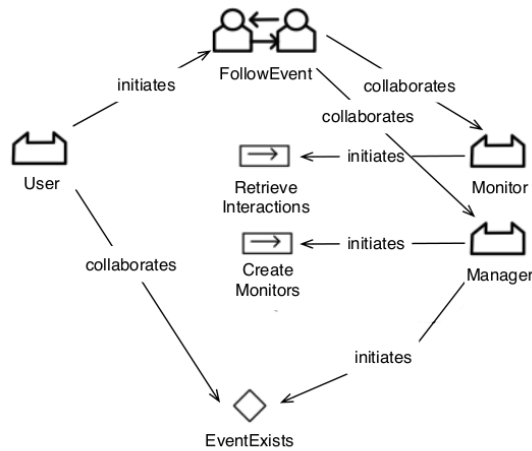


Figura 3.3: Dimensión dinámica del SMA propuesto. Muestra las interacciones entre las entidades que forman parte del sistema para realizar la recogida y almacenamiento de información asociada a un evento en una red social online (*FollowEvent*).

3.5 Dimensión dinámica

En esta dimensión se definen las interacciones y sus roles participantes. Se consideran dos entidades de interacción:

- *FollowEvent*: Aquí participan los roles *Monitor*, *User* y *Manager* (ver figura 3.3). El agente *User* inicializa la interacción y los agentes *Monitor* y *Manager* colaboran iniciando las actividades *CreateMonitors* y *RetrieveInteractions*. Se considera que el evento existe en el sistema. El *Manager* crea un nuevo *Monitor* para monitorizar los mensajes que se disponen en el evento.
- *RequestAnalysis*: Aquí participan los roles *Monitor*, *User* y *Analyzer* (ver figura 3.4). En este caso, tanto el agente *Monitor* como el *User* pueden iniciar la unidad de interacción *RequestAnalysis*, pero solo éste último agente puede iniciar la actividad *ShowResults*. El agente *Analyzer* puede iniciar *NetworkAnalysis*, *NodeAnalysis* y *StatisticsAnalysis*.

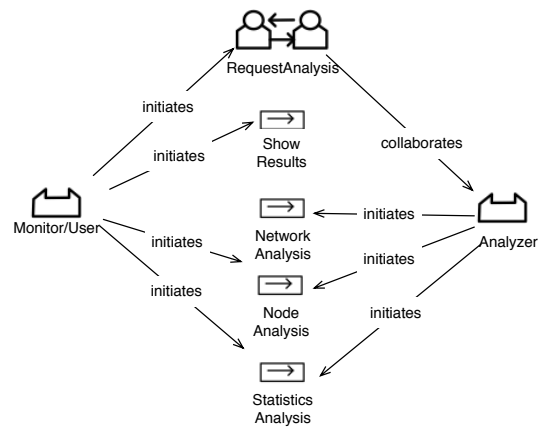


Figura 3.4: Dimensión dinámica del SMA propuesto. Muestra las interacciones entre las entidades que forman parte del sistema para realizar el análisis de la red de interacciones generada a partir de la información recogida de un evento en una red social online (*RequestAnalysis*)

3.6 Conclusiones

El diseño del SMA propuesto para llevar a cabo la recogida y análisis de la información de Twitter se ha realizado utilizando la metodología GORMAS. Esta metodología nos ha permitido definir el sistema por medio de tres dimensiones: estructural (nos especifica qué roles intervienen en el sistema), funcional (describe qué funcionalidad/servicios ofrece cada uno de los roles) y finalmente la dinámica (especifica qué interacciones y escenarios de interacción se van a dar entre las entidades del sistema en función del rol que desempeñen). El uso de la metodología nos ha permitido definir con detalle el diseño del sistema que se utilizará como base para el desarrollo del framework multiagente.

Framework

4.1	Visión global	27
4.2	Plataforma y agentes	29
4.3	Protocolos de interacción	34
4.4	Conclusiones	38

En este trabajo se propone un sistema multiagente para la automatización y distribución del proceso de recolección de datos de la actividad de los usuarios en redes sociales online. Posteriormente, se analiza su evolución en diferentes niveles de granularidad basados en métricas estructurales de redes complejas. El sistema multiagente propuesto está organizado en organizaciones virtuales donde hay un conjunto de roles (i.e., *Manager*, *Monitor*, *Analyzer* y *User*) que definen los servicios ofrecidos. El papel más relevante es el analizador que se caracteriza por ofrecer servicios relacionados con un análisis configurable de la información recopilada que está asociada a un evento. Estos servicios ofrecen un análisis de lo que ocurre antes, durante y después de un evento a nivel global e individual. Toda la información generada durante éste se le proporciona al usuario final a través de una interfaz. Los agentes que forman parte del MAS funcionan asincróna y simultáneamente, lo que da como resultado una eficiencia computacional. Los nuevos requerimientos o funcionalidades pueden ser abordados agregando nuevos agentes o reconfigurando los existentes.

4.1 Visión global

En esta sección presentamos la implementación de un framework para el análisis de la actividad de los usuarios asociada a un evento en redes sociales online, concretamente Twitter. El

sistema está basado en el diseño propuesto en el capítulo anterior y cuenta con seis agentes encargados de realizar la extracción, almacenamiento, procesamiento y análisis de la información procedente de Twitter. Concretamente el sistema lo componen los siguientes roles: *Manager*, *Monitor*, *Analyzer* y *User*. Algunos agentes pueden tener varios roles y realizar sus funciones en base a ellos.

En total los agentes que componen el sistema son los siguientes: el agente *VisualAgent* es el intermediario entre el MAS y el usuario final. El usuario interactúa con este agente para indicar el evento sobre el que está interesado en recoger información y/o analizar. El agente *VisualAgent* recibe esta información y envía un mensaje al agente *ManagerAgent* para iniciar el seguimiento del evento. El *ManagerAgent* es responsable de gestionar los eventos seleccionados por los usuarios. El agente *AgentTweet* es responsable de recopilar los mensajes de generados durante un evento. Estos mensajes son enviados al agente *Neo4jAgent* para ser almacenados. El agente *QueryNeo4jAgent* se encarga de extraer la información de la base de datos y crear la red que se enviará al agente *GraphAnalyzerAgent*. Éste agente realiza un análisis bajo demanda del usuario y todos los datos se envían al agente *VisualAgent* que muestra los resultados del análisis de actividad asociada al evento seleccionado por el usuario. En los siguientes párrafos, describimos con detalle los principales agentes del framework.

La idea principal y su proceso general se puede dividir en dos partes:

1. **Búsqueda de la actividad asociada a un evento.** A partir de una interfaz gráfica, el usuario introduce el nombre del hashtag en Twitter y presiona el botón de búsqueda. Los agentes que participan son *ManagerAgent*, *VisualAgent*, *AgentTweet* y *Neo4jAgent*, que se encargan recoger los datos del evento por parte del usuario, recopilar todos los tweets que pueden encontrar relacionado con ese evento y almacenarlos en una base de datos.
2. **Análisis de los datos.** A través de la misma interfaz, el usuario introduce el hashtag que quiere analizar y un intervalo de fechas inicial y final. Hay que tener en cuenta que deben existir datos del evento en la base de datos para efectuar esta acción. Después, cuando el usuario presiona el botón de analizar, el sistema se encarga de extraer los mensajes correspondientes a ese intervalo de tiempo, crear la red y analizar sus características. Los agentes que participan son *ManagerAgent*, *VisualAgent*, *QueryNeo4jAgent* y *GraphAnalyzerAgent*.

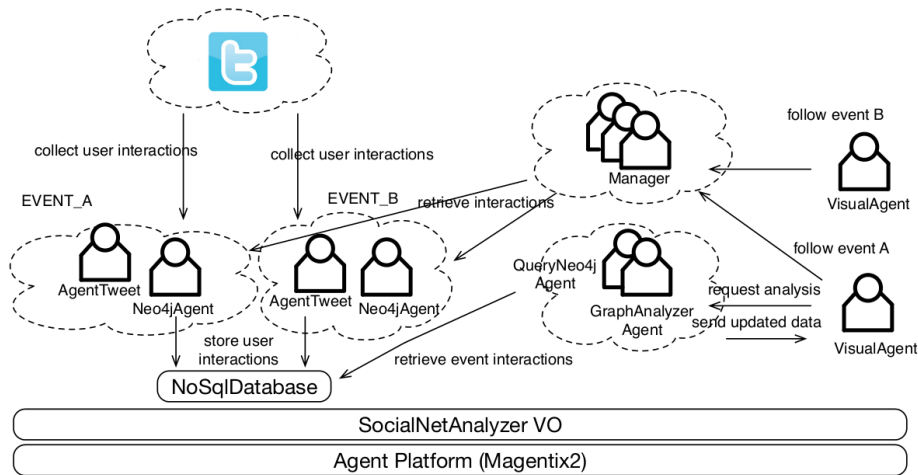


Figura 4.1: Estructura general del sistema. Indica la dirección de los flujos de interacción y/o conversación entre los diferentes agentes del sistema.

4.2 Plataforma y agentes

El sistema se basa en un conjunto de agentes desarrollados utilizando la plataforma Magentix2 [33]. Magentix2 es una plataforma para SMAs abiertos que soporta protocolos y conversaciones de interacción flexibles, comunicación indirecta e interacciones entre las organizaciones de agentes. Además, la plataforma permite la definición de un contexto normativo, que se puede adaptar fácilmente a los cambios en los requisitos del sistema. Hoy en día, Magentix2 ofrece soporte a tres niveles: *Nivel organizacional* que consiste en tecnologías y técnicas relacionadas con las sociedades de agentes; *Nivel de interacción* que soporta protocolos y conversaciones de interacción flexibles; y a nivel de *Agente* que ofrece diferentes clases de agentes como CAgent (i.e., un agente que permite la creación automática de conversaciones simultáneas) o JasonAgent (i.e., agentes BDI que pueden participar en conversaciones simultáneas).

4.2.1 Rol Monitor: agentes *AgentTweet* y *Neo4jAgent*

El agente *AgentTweet* utiliza el API de Twitter4J¹ para la recuperación de los datos públicos generados por los usuarios durante un evento. Dichos datos, una vez recopilados, son almacenados en Neo4J² por el agente *Neo4jAgent*. Neo4j es una base de datos de código abierto NoSQL implementada en Java y Scala, es altamente escalable y utiliza las relaciones como

¹<http://twitter4j.org/en/index.html>

²<https://neo4j.com/>

aspecto central de su modelo de datos, por lo que es capaz de almacenar, procesar y consultar conexiones de manera eficiente. Los modelos de datos en Neo4J consideran tres componentes: *nodos* para poder definir las entidades, *etiquetas* que se utilizan para agrupar los nodos en conjuntos y *relaciones* que conectan nodos y los organizan en estructuras arbitrarias.

En el modelo propuesto, hemos considerado los siguientes **nodos**:

- *hashtag*: representa el evento en el que un usuario está interesado. Identificamos un solo atributo asociado al hashtag: el nombre del evento.
- *tweet*: representa el mensaje (i.e., mensaje de texto, foto, video, etc.) que un usuario crea en Twitter. Identificamos un conjunto de atributos asociados al mensaje como coordenadas geográficas, anotaciones temporales, referencias a otros usuarios, número de likes/favoritos, etc.
- *usuario*: representa a un usuario de Twitter que ha participado a través de la publicación de un mensaje o que ha sido mencionado en el contenido del mismo. Identificamos un conjunto de atributos asociados al usuario que han sido extraídos del propio mensaje como: anotaciones temporales, amigos/seguidores, likes, etc.

Las **relaciones** que conectan los nodos en el modelo de datos propuesto son (ver Figura 4.2):

- *has_hashtag*, que conecta un tweet con un evento.
- *mentions*, que conecta un tweet (i.e., mensaje de texto, foto, video, etc.) con uno o varios usuarios a los que se menciona o hace referencia en el tweet.
- *creates*, que conecta a un usuario con un tweet que él mismo ha creado.

Utilizando estas entidades y relaciones, hemos podido almacenar eficientemente toda la información generada por los usuarios sobre un determinado evento. Como ejemplo, se muestra el uso de este modelo para el caso de almacenar la actividad de los usuarios durante un evento en Twitter, pero puede extrapolarse a otras redes sociales donde las interacciones podrían ser modeladas como mensajes entre los usuarios sobre un tema o evento. En Twitter, hemos considerado los siguientes tipos de mensajes:

- *Mensajes globales* que se utilizan cuando la actualización está pensada para cualquier persona que se preocupe por leerla. Este es el tipo más habitual de mensaje en Twitter. Esta información y sus relaciones se representan utilizando el modelo propuesto para Neo4j como sigue:

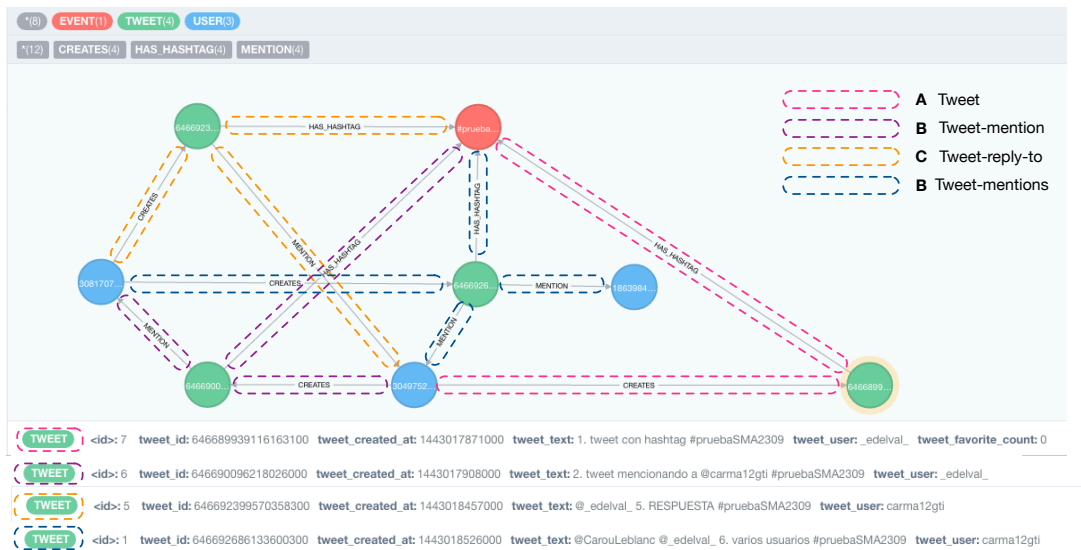


Figura 4.2: Ejemplo de cómo se almacenarían los distintos tipos de mensajes de interacción asociados a un evento utilizando el modelo propuesto para almacenar la información en Neo4j. Los nodos del grafo representan las entidades del modelo. Las entidades consideradas en el modelo son: eventos (rojo), interacciones (verde), usuarios (azul). Los enlaces del grafo representan los tipos de relaciones que hay en el modelo de datos (i.e., creates, mention, has_hashtag, y creates). Las relaciones que se muestran en el grafo son las generadas al insertar un tweet donde no se menciona a ningún usuario (rosa), un tweet que contiene una mención (morado), un tweet que es un reply-to (amarillo), y una mención a varios usuarios (mencion). En la parte inferior se muestran los cuatro mensajes (tweets) insertados en la base de datos.

$(u : USER) - [: CREATES] -> (t : MESSAGE)$

$(t : MESSAGE) - [: HAS_HASHTAG] -> (e : EVENT)$

- *Mensajes individuales* que son los que involucran a otro usuario. Los mensajes individuales en Twitter pueden ser: *retweets*, *menciones* y *respuestas* a los usuarios.
 - Los *retweets* son mensajes que previamente fueron publicados por otro usuario.
 - Las *menciones* son mensajes que se usan cuando un usuario pretende informar a través de un tweet a una persona específica. A menudo, dos o más usuarios tendrán conversaciones mediante la publicación de menciones entre sí.
 - Las *respuestas* a los usuarios son mensajes en los que un usuario menciona a otro usuario en un tweet anterior.

La información sobre los mensajes individuales y sus relaciones se representan usando el modelo propuesto en Neo4j como sigue:

$(t : MESSAGE) - [: MENTIONS] -> (u : USER)$

4.2.2 Rol Analyzer: agentes *QueryNeo4jAgent* y *GraphAnalyzerAgent*

El agente *QueryNeo4jAgent* es responsable de extraer la información asociada a un evento durante un intervalo de tiempo. Para ello hace uso del lenguaje Cypher³. Cypher es un lenguaje de consulta de grafos declarativo que permite la ejecución eficiente de consultas y la actualización de la base de datos. Cypher toma prestada su estructura de SQLqueries. Las cláusulas se encadenan juntas y alimentan conjuntos de resultados intermedios entre sí. El agente *QueryNeo4jAgent* utiliza este lenguaje para obtener información sobre las interacciones de los usuarios asociadas a un evento durante un período de tiempo. Como ejemplo, mostramos una instancia de una consulta en Cypher para la recuperación de información en Twitter que se ha generado sobre un evento “*myEvent*” en un intervalo de tiempo determinado [time_start, time_end]:

```
MATCH (u:USER) -[:CREATES] -> (t:MESSAGE) -[:HAS_HASHTAG]
      -> (e:EVENT{Hashtag:"myEvent"}),
      OPTIONAL MATCH (u2:USER) <-[:MENTION] - (t:MESSAGE),
      WHERE t.tweet_created_at >="time_start"
      AND t.tweet_created_at <="time_end",
      RETURN u.User_screen_name, t.tweet_id, u2.User_screen_name)
```

Basándose en la información obtenida de la base de datos Neo4j, el agente *QueryNeo4jAgent* crea una red de interacciones del evento anotada temporalmente. Un usuario A se convierte en un nodo de la red cuando participa escribiendo un mensaje global o individual con el *hashtag* asociado al evento o cuando otro usuario B le hace referencia en un mensaje individual dentro del intervalo temporal seleccionado. Cada nodo de la red (usuario) tiene una etiqueta asociada que representa el instante cuando se unió a la red (i.e., en instante cuando participó por primera vez en el evento). Los enlaces de la red se establecen cuando un usuario escribe un mensaje individual donde se menciona a un usuario que ya estaba presente en la red (i.e., un usuario que ya había participado anteriormente en el evento con un mensaje) o a un usuario que todavía no forma parte de la red. Por lo tanto, el tipo de red es dirigido (ver Figura 4.3). Es importante observar que trabajamos con datos acumulados (i.e., la red en el tiempo t también incluye los nodos y enlaces de un momento anterior $t' < t$). El agente *QueryNeo4jAgent* crea la red utilizando GRAPHML, un lenguaje para describir estructuras de redes complejas y sus datos asociados.

Una vez que el agente *QueryNeo4jAgent* ha construido la red de interacciones de usuarios asociada a un evento durante un intervalo temporal, el agente *GraphAnalyzerAgent* inicia el

³<https://neo4j.com/developer/cypher-query-language/>

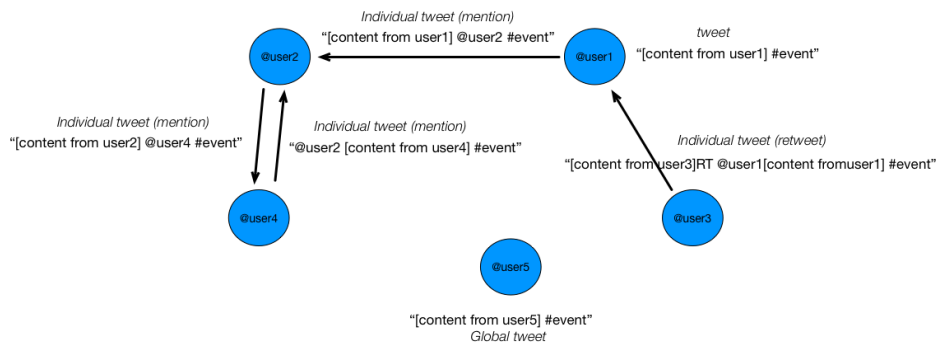


Figura 4.3: Red de interacciones generada a partir de la actividad de los usuarios en Twitter asociada a un evento.

análisis de las propiedades estructurales a nivel global y/o individual. Para llevar a cabo esta tarea, el agente utiliza JUNG⁴ y Gephi Toolkit⁵. Jung, así como Gephi Toolkit, son una serie de librerías software que facilitan el modelado, análisis y visualización de datos que pueden ser representados como una red. El agente *GraphAnalyzerAgent* ofrece dos tipos de análisis de las interacciones de los usuarios:

- **Análisis de nivel de red.** A nivel de red, el agente *GraphAnalyzerAgent* estudia la evolución de las siguientes propiedades a lo largo del tiempo:
 1. Nodos, es decir, la evolución de la participación de los usuarios durante un evento o eventos.
 2. Enlaces, es decir, cómo fluye la información en la red e influye en la formación del componente gigante.
 3. Longitud y diámetro medio de la red para entender en qué momento hay un cambio en la estructura de la red y, por lo tanto, en la comunicación.
 4. Modularidad y grado medio, cómo de conectadas están las comunidades.
 5. Densidad, un indicador de la proporción de enlaces que hay con respecto a todos los posibles.
 6. Número de nodos en la componente gigante, nivel de participación de usuarios dentro de la componente principal.
- **Análisis a nivel individual.** El agente *GraphAnalyzerAgent* estudia la evolución de las propiedades de centralidad de los usuarios (i.e., nodos de la red) en un evento. Las propiedades de centralidad determinan qué usuarios son los más importantes basados en

⁴<http://jung.sourceforge.net/>

⁵<https://gephi.org/toolkit/>

su ubicación en la estructura de la red siguiendo distintos criterios. El agente *GraphAnalyzerAgent* considera las siguientes métricas de centralidad:

1. *Intermediación*, indica si el usuario actúa como puente/intermediario a través de las rutas más cortas entre otros dos usuarios.
2. *In-degree*, indica si éste usuario es una referencia para otros usuarios.
3. *Out-degree*, indica si éste usuario es una persona activa en el evento.
4. *Eigenvector*, indica si el usuario está conectado a otros usuarios que a su vez también tienen un alto grado de conectividad.
5. *Page Rank*, indica cómo de relevante es ese usuario en la red.
6. *Closeness, eccentricity*, indican si a través de ese usuario se transmite la información más rápido.
7. *Betweenness*, indica si el usuario está más conectado a otros usuarios a la hora de transmitir la información (i.e., si actúa normalmente de puente a lo largo del recorrido de dicha información).

Una vez que el análisis se ha realizado, se almacena durante un período de tiempo en la base de datos.

4.2.3 Rol User: agente *VisualAgent*

El agente *VisualAgent* actúa como un agente mediador entre los usuarios y el sistema a través de una interfaz. La interfaz ofrece a los usuarios dos posibles acciones: recoger información asociada a un evento o analizar un evento. Para seguir un evento, el usuario introduce un identificador de evento (i.e., *hashtag*, palabra clave, nombre de usuario, etc). Para analizar un evento, el usuario debe proporcionar la palabra clave del evento y el intervalo de tiempo durante el cual quiere analizar un evento del que anteriormente se haya recogido información.

4.3 Protocolos de interacción

Los agentes del sistema interactúan entre sí para recopilar, almacenar, procesar y analizar lo que está sucediendo en un evento en una red social. La dinámica del sistema se muestra en los siguientes protocolos de interacción: *FollowEvent* y *RequestAnalysis*.

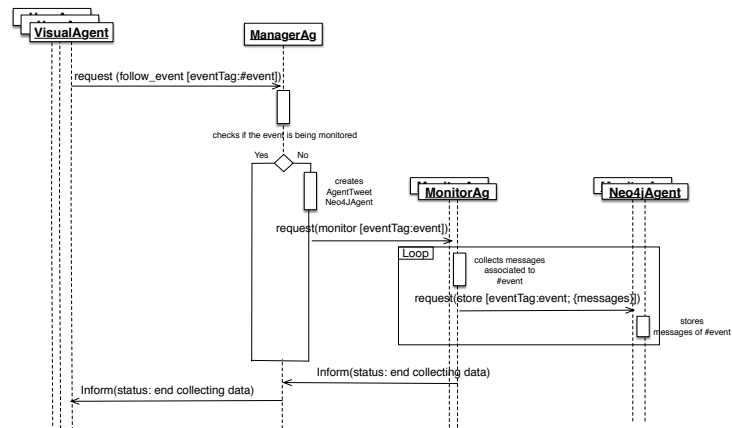


Figura 4.4: Diagrama de secuencia para el escenario *FollowEvent*. Describe las interacciones entre los agentes *VisualAgent*, *ManagerAgent*, *AgentTweet* y *Neo4jAgent*.

El escenario *FollowEvent* comienza cuando el agente que representa al usuario (*VisualAgent*) envía un mensaje *request* al agente *ManagerAgent* para comenzar a seguir un evento (ver Figura 4.4). El mensaje contiene la palabra clave que identifica el evento o *hashtag*. El agente *ManagerAgent* comprueba si el nombre del evento es válido (no se considera el símbolo de *hashtag* # en dicho nombre). Si todo es correcto, el *ManagerAgent* crea los agentes *AgentTweet* y *Neo4jAgent*, les transmite el nombre del evento solicitado. Llegados a este punto, cada uno de los dos agentes realiza sus tareas establecidas.

Por una parte, el agente *AgentTweet* establece los parámetros de Twitter4J y comienza la búsqueda y recopilación de tweets que contengan ese nombre de evento en su contenido del mensaje. Debido al formato obtenido del mensaje, cada vez que un tweet es recuperado, éste se codifica en formato JSON y se le envía al agente *Neo4jAgent*. Este agente tiene en cuenta también la restricción que existe en el API de Twitter y, cuando ese umbral de tweets obtenidos supera el umbral permitido, éste se queda a la espera para poder seguir recopilando tweets. El proceso termina cuando ya no hay tweets por recuperar o API de Twitter4J considera que su recopilación ha llegado a su fin y no es posible obtener más mensajes. Cuando esto sucede, el agente *AgentTweet* le envía al agente *Neo4jAgent* el número total de tweets obtenidos y notifica al agente *ManagerAgent*, que a su vez hace lo mismo para que el agente *VisualAgent* muestre los datos obtenidos al usuario.

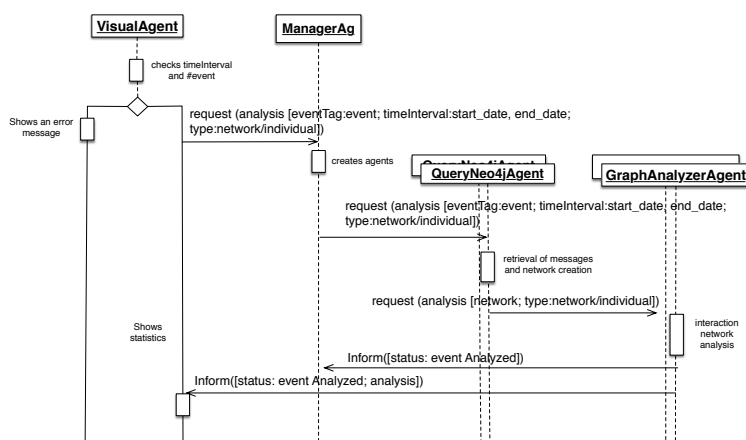


Figura 4.5: Diagrama de secuencia para el escenario *RequestAnalysis*. Describe las interacciones entre los agentes *VisualAgent*, *ManagerAgent*, *QueryNeo4jAgent* y *GraphAnalyzerAgent*.

Por otra parte, el agente *Neo4Agent* permanece a la espera de recibir los tweets en formato JSON. Éstos son decodificados y se comprueba que cada uno de los tres elementos (hashtag, tweet y usuario) no se encuentren ya en la BD de Neo4j. Esto impide insertar elementos repetidos y la misma comprobación es realizada también para las relaciones entre esos nodos (*has_hashtag*, *mention* y *creates*). Si, por ejemplo, el tweet que se quiere insertar no existe en la DB, pero menciona a un usuario que ya ha sido insertado anteriormente, el agente recupera el nodo que hace referencia a ese usuario para poder establecer la relación *mentions* con el nuevo tweet. Cuando el agente recibe el número final de tweets obtenidos, realiza las operaciones de cierre para la base de datos.

Un usuario que estuviera interesado en seguir un evento también puede solicitar un análisis de la actividad de los usuarios en el evento cada vez que lo desee. El escenario *RequestAnalysis* comienza cuando un usuario solicita un análisis del evento en el que está interesado durante un intervalo de tiempo (ver Figura 4.5). En primer lugar, el agente *VisualAgent* comprueba si el nombre del evento es correcto y si las fechas establecidas en el intervalo de tiempo para el análisis están en el formato adecuado. Si no es así, el agente envía un mensaje de error al usuario. En caso contrario, se notifica al agente *ManagerAgent* la operación a realizar. Éste se encarga de crear los agentes *QueryNeo4jAgent* y *GraphAnalyzerAgent* y transmitirles los datos necesarios.

Una vez que el agente *QueryNeo4jAgent* recibe una solicitud para el análisis de un evento, realiza una consulta a la base de datos de Neo4j obteniendo así todos los nodos hashtag, tweet y usuario relacionados con ese evento y en el intervalo de tiempo en el que se generó el tweet. Después, se extraen los datos y para poder crear el fichero de texto en formato GRAPHML que definirá los nodos y los enlaces de la red de interacciones. Una vez se dispone de la red, el agente *QueryNeo4jAgent* lo envía al agente *GraphAnalyzerAgent* para su análisis. El agente *GraphAnalyzerAgent* se encarga de procesar, analizar y mostrar el análisis de la red. Las métricas que analiza se pueden clasificar en dos tipos: a nivel individual (considerando el nodo) y a nivel global (considerando el grafo entero). Los resultados se envían al agente *VisualAgent* para mostrarlos al usuario. En la tabla 4.1 se expone un resumen de las operaciones y agentes implicados mencionados anteriormente.

Rol	Agente	Servicios	Objetivos	Agentes creados	Agentes con los que interactúa
X	MainSMA	Inicio	Inicializar el SMA	ManagerAgent	X
Manager	ManagerAgent	X	Crear y gestionar los agentes	AgentTweet, VisualAgent, Neo4jAgent, QueryNeo4jAgent, GraphAnalyzerAgent	X
User	VisualAgent	Recopilar tweets y análisis	Ofrecer entorno visual y recoger los eventos de usuario	X	ManagerAgent
Monitor	AgentTweet	Recopilar tweets	Obtener los tweets en base al nombre del evento y enviarlos en formato JSON al Neo4jAgent	X	ManagerAgent, Neo4jAgent
Monitor	Neo4jAgent	Recopilar tweets	Recoger los tweets de AgentTweet y almacenarlos en la DB	X	ManagerAgent, AgentTweet
Analyzer	QueryNeo4jAgent	Análisis	Recuperar tweets de la DB, realizar una query y crear el grafo	X	ManagerAgent
Analyzer	GraphAnalyzerAgent	Análisis	Procesar y analizar el grafo	X	ManagerAgent

Tabla 4.1: Resumen de la implementación de los agentes que forman parte del framework propuesto. Se describe para cada agente del sistema: los roles que juega, los objetivos en los que participa, los servicios que ofrece, los agentes que crea, y los agentes con los que interactúa.

4.4 Conclusiones

En este capítulo hemos presentado un framework multiagente para el análisis de la evolución del comportamiento de la red y de los usuarios durante eventos en redes sociales online. El sistema es responsable de recopilar, procesar, analizar y presentar la información recuperada de las actividades de los usuarios en Twitter. El uso de un sistema multiagente para el marco proporciona características intrínsecas como flexibilidad, adaptabilidad y escalabilidad necesarias en un entorno donde hay una demanda variable de los usuarios y una actividad variable de los mismos dependiendo del evento. En el marco propuesto, consideramos un período de tiempo en el análisis del comportamiento de los usuarios durante un evento en lugar de un solo punto. Además, el marco ofrece una visión completa de lo que sucede durante un evento teniendo en cuenta las propiedades estructurales de la red temporalmente anotada generada a partir de las interacciones de los usuarios. El sistema ofrece un análisis a nivel de red (es decir, ofrece una vista del comportamiento global de los usuarios durante un evento) y a nivel individual (es decir, qué usuarios juegan un papel clave durante un evento considerando diferentes criterios de centralidad).

Caso de Estudio

5.1	Introducción	39
5.2	Medidas a nivel de red	40
5.3	Medidas a nivel de nodo	42
5.4	Análisis de eventos	45
5.5	Conclusiones	54

5.1 Introducción

El objetivo de este capítulo es validar el sistema multiagente propuesto utilizando un contexto específico que nos permita comprobar la funcionalidad y los resultados proporcionados por el SMA. Se han considerado varios escenarios en los que un usuario decide comenzar a seguir un conjunto de eventos en Twitter para analizar la evolución de la actividad en dichos eventos. El usuario interactúa con el marco a través de una interfaz que elige las palabras clave y el periodo de tiempo. En el proyecto nos hemos centrado en la recuperación de información de Twitter. Una vez especificada toda esta información, el SMA comienza a seguir la actividad de los usuarios de manera transparente para el usuario. El análisis que ofrece el SMA se centra en dos niveles: nivel de red (es decir, métricas que proporcionan información sobre la actividad global), y nivel individual (es decir, métricas que proporcionan información sobre usuarios relevantes que consideran criterios diferentes).

En este capítulo, primero presentaremos las medidas estructurales a nivel de red y a nivel de nodo que se han tenido en cuenta y cómo se interpretan en el contexto de las redes sociales online y más concretamente en el tipo de redes de interacciones que genera el SMA desarrollado. En una segunda parte, presentaremos varios casos prácticos donde se analizan las medidas en

distintos tipos de eventos.

5.2 Medidas a nivel de red

El primer conjunto de medidas estudia la red asociada al evento como un todo. Las medidas a nivel de red determinan principalmente las propiedades estructurales o topológicas, es decir, cómo se conectan unos nodos con otros. Una vez determinadas las propiedades, éstas nos ayudan a determinar si una red sigue un determinado patrón ya definido como puede ser una red de tipo small world, scale-free, o random y por lo tanto inferir propiedades propias del tipo de red.

Antes de entrar en la definición de algunas de las propiedades más importantes que determinan la estructura de la red, vamos a presentar la siguiente notación. \mathcal{G} representa una red que consiste en un conjunto de nodos N , donde cada nodo $i \in \{1, \dots, n\}$ representa a un usuario. Los nodos están conectados a través de enlaces $E \in N \times N$. Utilizamos la matriz de adyacencia A para representar la red. Un enlace entre un nodo i y un nodo j se representa como $A_{ij} = 1$. Si no existe un enlace entre i y j se representa como $A_{ij} = 0$. Considerando esta definición, presentamos las siguientes propiedades que se han tenido en cuenta en el análisis.

Nodos. Es un indicador del número de nodos (i.e., en el contexto del proyecto los nodos hacen referencia a cuentas de usuario) que han participado en un evento. La evolución de los nodos permite ver el ritmo al cual nuevos usuarios se incorporan al mismo con su primer mensaje que usa el hashtag asociado al evento. Es interesante tener en cuenta la evolución del número de nodos porque nos permite detectar en qué momento se unen más usuarios a la red (al evento).

Enlaces. Los enlaces representan las interacciones entre usuarios (i.e., menciones, retweets y mensajes) y muestran cómo fluye la información en el evento. La evolución del número de enlaces ayuda a ver en qué momento los usuarios interactúan más y como consecuencia se forma la componente gigante. La componente gigante es la componente conectada de la red que contiene una mayor parte de los nodos.

Longitud media de los caminos y diámetro de la red. Se define como la media de las distancias entre todos los pares de nodos, es decir, la separación típica entre pares de usuarios.

Si consideramos la red \mathcal{G} con un conjunto de vértices N , siendo $d(i, j)$ la distancia más corta entre i y j y asumiendo que $d = (i, j) = 0$ si i no puede ser alcanzado por j . La longitud media del camino más corto l_g será:

$$l_g = \frac{1}{|N| \cdot (|N| - 1)} \cdot \sum_{i \neq j} d(i, j)$$

donde $|N|$ es el número de nodos de la red \mathcal{G} .

La evolución del camino medio ayuda a ver en qué momento hay un cambio estructural en la red. El diámetro es la distancia máxima de entre todas las distancias cortas entre dos usuarios de la red. Estas medidas nos indican cuántos retweets, menciones, o reply-to serían necesarios para que un mensaje se propagara por toda la red (diámetro) o llegara a un usuario particular (camino medio).

Densidad. La densidad determina la conectividad en la red entre pares, es decir, si un usuario está conectado a otro usuario y éste a su vez al primero. Esto se calcula dividiendo el número de relaciones existentes entre las relaciones totales posibles. Esta medida resulta útil para conocer si entre usuarios existe reciprocidad entre los mensajes que se envían (si el usuario A escribe al usuario B y éste último responde de vuelta al usuario A). Si en un evento hay muchos usuarios que escriben pero no son respondidos, la densidad de la red disminuye.

Modularidad. La modularidad es la fracción de los enlaces que caen dentro de los grupos establecidos menos la fracción esperada si los enlaces se distribuyeran al azar. El valor de la modularidad está en el rango $[-1/2, 1]$. Es positivo si, por casualidad, el número de enlaces dentro de los grupos excede el número esperado sobre la base. Para una división determinada de los vértices de la red, la modularidad refleja la concentración de los enlaces dentro de los módulos en comparación con la distribución aleatoria de enlaces entre todos los nodos independientemente de los módulos.

Usuarios que son parte de la componente gigante de la red. El análisis de esta propiedad nos facilita estimar en qué momento durante un evento la componente gigante se forma, es decir, en qué momento se forma un núcleo de interacciones que contiene a la mayor parte de los usuarios que participan en el evento. Esta medida es importante porque a partir del momento en el que se forma la componente gigante se determina el mejor momento para difundir información.

En las medidas no se tiene en cuenta el número de contactos de los usuarios. Es una medida sencilla de obtener y frecuentemente se emplea para dar una idea de los impactos que tiene un determinado tweet. Sin embargo, nosotros consideramos que las medidas a partir de las menciones directas a usuarios tienen una mayor visibilidad y, por lo tanto, un mayor impacto.

5.3 Medidas a nivel de nodo

El segundo conjunto de medidas se centra en la relevancia de los usuarios dentro de la red, con el fin de identificar aquellos usuarios más importantes para la formación de la red o para la transmisión de información. Habitualmente estas medidas se realizan sobre la red completa final, pero en nuestro caso, la evolución de la importancia de los usuarios a lo largo del evento es muy relevante y, como se muestra más adelante, varía mucho, especialmente en cierto tipo de eventos.

Existen muchas definiciones de importancia y por lo tanto muchas medidas de centralidad para redes. En el análisis realizado en el proyecto hemos considerado las siguientes:

Centralidad basada en grado (degree centrality). Una de las medidas más simples es tener en cuenta el grado de un nodo de la red, el número de enlaces conectados al nodo. Aunque es una medida sencilla, puede ser relevante. Por ejemplo, en una red social parece razonable suponer que aquellos nodos que están muy conectados puedan tener más acceso a más información o más prestigio que aquellos nodos que tienen menos. Un ejemplo claro donde se utiliza esta medida es en redes de citaciones de artículos, donde la importancia de un artículo (nodo) depende del grado de entrada de ese artículo (in-degree).

La expresión que calcula el valor de centralidad basada en grado es la siguiente:

$$C_{D_i} = \sum_j A_{ij}$$

En las redes dirigidas, cada nodo tiene dos grados. El grado de entrada es el número de enlaces entrantes y el grado de salida es el número de enlaces salientes. En el contexto de las redes sociales, para un nodo i , el grado de entrada representa el número de mensajes generados por otros usuarios donde se hace referencia a i . Esta métrica indica si el usuario es significativo para otra gente, es decir, que ese usuario es una persona conocida dentro de la comunidad que podríamos definir como un referente dentro de la red. El grado de salida para un nodo i de la

red representa el número de mensajes que un usuario crea haciendo referencia a otros usuarios. El grado de salida indica la actividad del usuario, es decir, que el grado de interacción del usuario A con el resto de participantes en el evento.

$$C_{D_i}^{out} = \sum_{j=1}^N A_{ij}$$

$$C_{D_i}^{in} = \sum_{j=1}^N A_{ji}$$

Centralidad basada en vector de valores propios (eigenvector centrality). Una extensión natural de la centralidad basada en el grado es la centralidad basada en el vector de valores propios. En esta medida de centralidad los nodos vecinos no se consideran todos iguales. En muchas circunstancias, la importancia de un nodo de la red aumenta teniendo en cuenta a qué nodos está conectado y si estos son también importantes.

La centralidad basada en vectores propios del nodo v puede ser definida como:

$$C_{V_i} = \frac{1}{\lambda} \sum_{j \in k_i} C_{v_j} = \frac{1}{\lambda} \sum_{j \in \mathcal{G}} A_{ij} C_{v_j}$$

donde k_i es el conjunto de vecinos de i y λ es una constante.

En el contexto de las redes sociales, un usuario con un alto valor de eigenvector es una autoridad que se le sigue por los contenidos que emite. Dentro de esta misma categoría, también se tiene en cuenta el PageRank, un tipo de medida que determina la relevancia de los nodos en la red en base al mayor valor del eigenvector en una matriz de adyacencia normalizada A' .

Centralidad basada en cercanía (Closeness centrality). Una medida de centralidad diferente a las anteriores es la proporcionada por la centralidad basada en cercanía, que mide la distancia media de un nodo al resto de nodos. Formalmente, la cercanía de un nodo i se define como:

$$C_{CLO_i} = \frac{|N|}{\sum_j S_{ij}}$$

donde S es la matriz de distancias de la red, es decir, aquella matriz cuyos elementos (i,j)

corresponden a la distancia más corta desde el nodo i hasta el nodo j .

Esta medida se puede interpretar como la rapidez que tomará la propagación de la información desde un nodo a todos los demás o como la accesibilidad de un nodo en la red.

Uno de los problemas de esta medida es que el rango de valores de centralidad entre el máximo y el mínimo es muy estrecho. La consecuencia de esto es que es difícil distinguir entre valores más o menos centrales utilizando esta medida. Los valores de centralidad tienen a acumularse en un pequeño intervalo y es necesario mirar los dígitos del final para determinar cuál es más relevante.

Centralidad basada en intermediación (Betweenness centrality). Un concepto de centralidad que es diferente a los anteriores es la centralidad basada en la intermediación. Esta medida de centralidad para un nodo a mide cuántos caminos entre otros dos nodos de la red (i, j) pasan a través de él. Los nodos con un alto valor de betweenness pueden ser considerados como influyentes en una red ya que controlan la información que otros nodos se están pasando. Los nodos que tienen el valor más alto de betweenness en escenarios donde hay transmisión de información son aquellos nodos por los que más mensajes pasan y en el caso de que obtuvieran un beneficio por hacer de “pasarela” de mensajes, podrían adquirir una gran relevancia en la red. Los vértices que tienen un alto valor de betweenness son también aquellos que al ser eliminados interrumpirán más comunicaciones entre otros nodos de la red ya que a través de ellos pasan la mayor parte de los caminos que recorren los mensajes. La idea de centralidad basada en intermediación puede ser considerada como una aproximación a la influencia que tienen los nodos sobre el proceso de comunicación.

Formalmente, la intermediación C_{BET_i} de un nodo i en una red se define como:

$$C_{BET_i} = \sum_{j,k} \frac{b_{jik}}{b_{jk}}$$

donde b_{jk} es el número de caminos más cortos desde el nodo j hasta el nodo k , y b_{jik} el número de caminos más cortos desde j hasta k que pasan a través del nodo i .

Un usuario con un valor alto de betweenness indica que es un punto de conexión entre distintos grupos de usuarios y suele tener cierto control sobre los flujos de información en la red. Un nodo con un betweenness alto no es necesario que tenga un grado de entrada o de salida alto. Si quisiéramos difundir un mensaje en una red social, sería interesante que los usuarios con alto valor de intermediación reenviaran el mensaje.

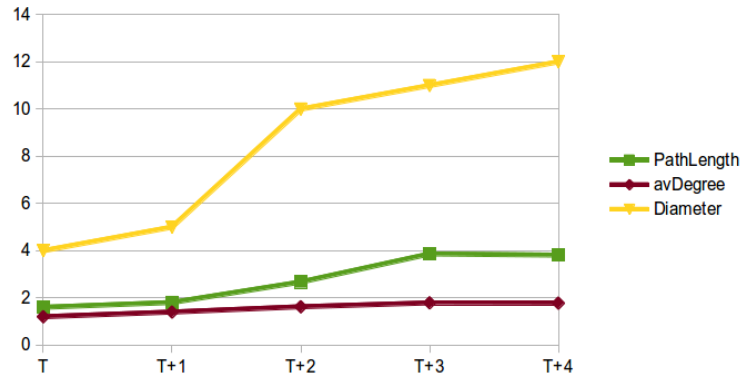


Figura 5.1: Evolución de las propiedades estructurales a nivel de red: longitud del camino medio, grado medio de conexión y diámetro de la red. Evento #MuereBarberáARV

5.4 Análisis de eventos

En esta sección vamos a realizar el análisis de las estructuras comunicativas que emergen a partir de las interacciones entre usuarios durante un evento en Twitter. Veremos las propiedades estructurales a nivel de red y a nivel de nodo que se comentaron en la sección anterior y analizaremos su evolución. A partir de estas medidas podremos detectar patrones de comunicación que se siguen en función del tipo de evento. Los eventos que hemos considerado son los siguientes: #MuereBarberáARV (un evento de televisión que trata temas de actualidad política), #ChampionsLeague (un evento de tipo deportivo) y #pesadillazamora (un evento de televisión de entretenimiento).

5.4.1 Evento #MuereBarberáARV

Teniendo en cuenta la evolución de las propiedades estructurales, podemos observar cómo se comportan los usuarios en el evento. Su evolución se puede dividir en tres etapas: inicio, durante, y final del programa. Esta división nos muestra que el comportamiento de los usuarios no es uniforme durante el programa “Al Rojo Vivo”, si no que los usuarios modifican su forma de interactuar a lo largo del programa.

La figura 5.1 muestra la evolución del diámetro, el grado medio de conexión y la longitud del camino medio. El *diámetro de la red* sigue una evolución parecida al camino medio de la red. Al principio del evento el diámetro tiene un valor entorno a cinco, esto significa que la distancia mayor entre dos usuarios de la red es cinco. Esto es un indicador de que al principio del evento

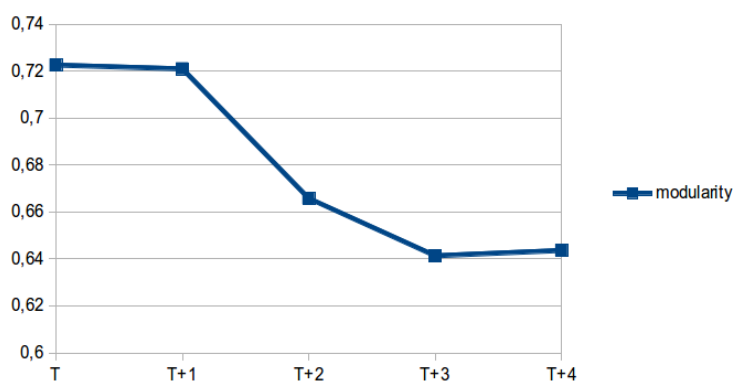


Figura 5.2: Evolución de las propiedades estructurales a nivel de red: modularidad. Evento #MuereBarberáARV

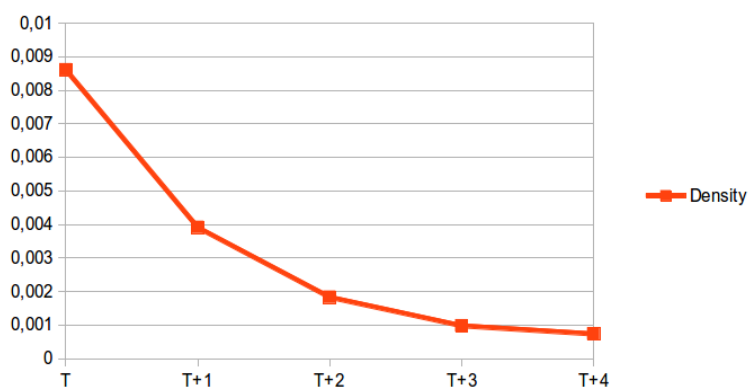


Figura 5.3: Evolución de las propiedades estructurales a nivel de red: densidad. Evento #MuereBarberáARV

los usuarios que han participado mediante mensajes son pocos y entre ellos ha habido una comunicación recíproca. Respecto al *camino medio*, éste va aumentando conforme crece la red, lo que significa que las comunicaciones entre los usuarios no se mantiene constante, sino que cambia. El momento más relevante de este cambio es durante el transcurso del programa (entre los instantes de tiempo T+1 y T+3). Este cambio en la estructura se debe a la participación de nuevos usuarios (nuevos nodos que se incorporan a la red) y la generación de mensajes de interacción (ver figura 5.4). El *grado de conexión* de los nodos no evoluciona de una manera menos significativa que el camino medio. El valor del grado de conexión está alrededor de 1.5. Esto significa que de media un usuario participa en la red con uno o dos mensajes en la red.

La figura 5.2 nos muestra la evolución de la *modularidad*. La evolución de la *modularidad* nos proporciona información sobre las comunidades que se forman en la comunicación durante este

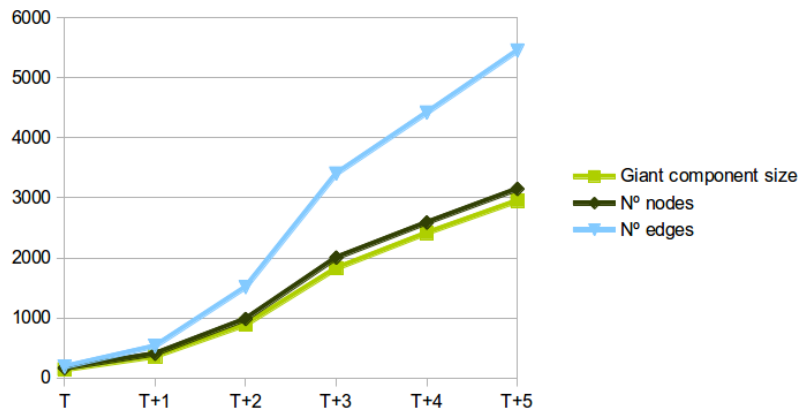


Figura 5.4: Evolución de las propiedades estructurales a nivel de red: número de nodos, número de ejes y componente gigante. Evento #MuereBarberáARV

evento. En general, durante todo el evento, se mantiene el número constante de comunidades, esto es un indicador de que los usuarios siguen interactuando con el mismo grupo de cuentas de Twitter durante la evolución del programa. La figura 5.3 nos muestra la *densidad* de la red. Ésta tiene un valor bajo que va decreciendo a lo largo del evento. Esto es debido a que hay una muy baja reciprocidad en la comunicación entre los usuarios (i.e., los mensajes generados por un usuario A hacen referencia a otro(s) usuario(s) pero éste (o éstos) no interactúan con A).

Finalmente, la figura 5.4 muestra la evolución del número de nodos y enlaces así como el tamaño de la componente gigante. Como hemos comentado anteriormente, es durante el evento cuando se produce un aumento significativo de la participación. Es en ese periodo cuando el número de usuarios, mensajes y el tamaño de la componente gigante aumentan. Cabe destacar que el número de mensajes es mayor que el número de nodos por lo que durante el evento los usuarios participan de una manera más activa comentando lo que está sucediendo con más de un mensaje. Esto hace que el grado medio de conexión aumente (ver figura 5.1). Es importante fijarse en el momento donde la componente gigante crece. Esto es interesante a la hora de decidir cuándo es relevante generar un mensaje cuyo objetivo sea llegar a la mayor parte de usuarios que están participando en el evento.

El análisis que ofrece el framework permite detectar qué usuarios juegan un papel relevante en la comunicación teniendo en cuenta distintos criterios de centralidad. En la tabla 5.1, mostramos el perfil de los usuarios que son relevantes si consideramos las medidas de centralidad comentadas en la sección anterior en distintos momentos del evento. Hemos considerado los siguientes tipos de perfiles: **celebrities** (i.e., actores, presentadores, participantes de un

T	IN DEGREE	OUT DEGREE	PAGE RANK	EIGENVECTOR	CLOSENESS	ECCENTRICITY	BETWEENNESS
INICIO	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CELEBRITIES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CELEBRITIES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
MEDIO	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CELEBRITIES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	CELEBRITIES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CELEBRITIES	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CELEBRITIES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES
	CELEBRITIES	M. COMUNICACION	CELEBRITIES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
FINAL	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS
	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	M. COMUNICACIÓN	M. COMUNICACIÓN	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS
	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS

Tabla 5.1: Evolución de las propiedades estructurales a nivel de nodo durante el evento #MuereBarberáARV. Indica los 5 perfiles que tienen valores de centralidad más altos teniendo en cuenta distintos criterios de centralidad.

concurso, personajes políticos, etc.), **cuentas oficiales** (i.e., cuentas asociadas a programas de TV, cuentas oficiales de un evento, etc.), **cuentas de grupos** (i.e., club de fans, asociaciones, partidos políticos, etc.), **medios de comunicación** (i.e., blogs, bloggers, radios, televisiones, periodistas, etc), y **usuarios anónimos**.

En el evento #MuereBarberáARV se observa que las cuentas oficiales asociadas al programa son las más relevantes en cuanto a grado de entrada, page rank, eigenvector, excentricity. Esto refleja que las cuentas oficiales no son cuentas que generen conversaciones con otros usuarios, sino que son cuentas que reciben muchos mensajes de usuarios anónimos y sólo hacen referencia a otras cuentas de personajes o grupos relevantes. Las cuentas de personajes relevantes juegan un papel importante durante el evento si tenemos en cuenta los mensajes recibidos de otras cuentas (in-degree) porque están relacionados con otras cuentas importantes (i.e., alto valor de page-rank e eigenvector). Los usuarios anónimos en este evento también juegan un papel principal en la generación de mensajes (i.e., un alto valor en el grado de salida) y en la conexión de comunidades diferentes (i.e., betweenness).

5.4.2 Evento #ChampionsLeague

Este evento de la jornada de Liga de Campeones tuvo lugar el 23 de noviembre, donde se hicieron seis partidos de forma simultánea en el que intervinieron varios equipos de fútbol. Basándonos en la evolución de las propiedades estructurales, podemos observar que el comportamiento de los usuarios durante el evento #ChampionsLeague fue variable en el tiempo.

La figura 5.5 muestra la evolución del diámetro, el grado medio de conexión y la longitud del camino medio. Al principio del evento, el *diámetro* de la red es dos. Que la mayor distancia entre dos usuarios de la red sea tan baja indica que hay nodos que están conectados a muchos otros nodos de la red, esto hace que disminuyan las distancias entre éstos. El *camino medio* de

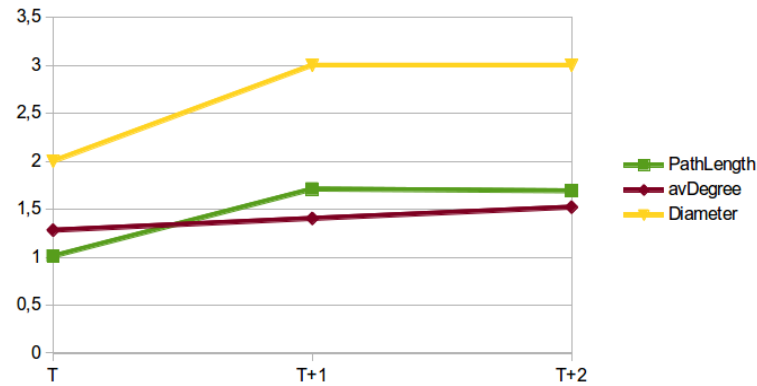


Figura 5.5: Evolución de las propiedades estructurales a nivel de red: longitud del camino medio, grado medio de conexión y diámetro de la red. Evento #ChampionsLeague

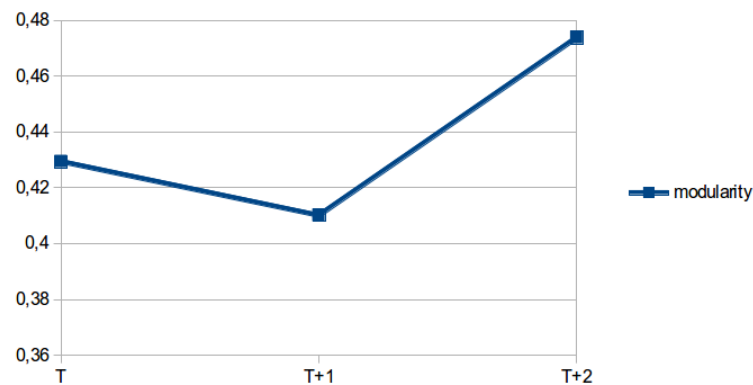


Figura 5.6: Evolución de las propiedades estructurales a nivel de red: modularidad. Evento #ChampionsLeague

la red crece en la primera mitad del evento (a medida que crece la red) y se mantiene constante hasta finalizar. Esta evolución en la estructura indica que muchos usuarios se unieron al evento desde T a T+1. En el intervalo entre T+1 a T+2 el aumento más significativo del número de mensajes (i.e., mensajes) que de usuarios hizo que el camino medio se mantuviera constante. Un valor tan bajo tiene relación con que la mayoría de los usuarios están conectados a cuentas oficiales y medios de comunicación que hace de hub y acortan los caminos entre los usuarios. El *grado de conexión* no evoluciona de forma significativa. Mantiene un valor medio alrededor de 1.4, lo que indica que un usuario participa en la red con uno o dos mensajes como mucho.

La figura 5.6 nos muestra la evolución de la *modularidad*. Podemos ver que ésta es casi constante y al tener un valor alto indica que el nivel de cohesión en las comunidades detectadas

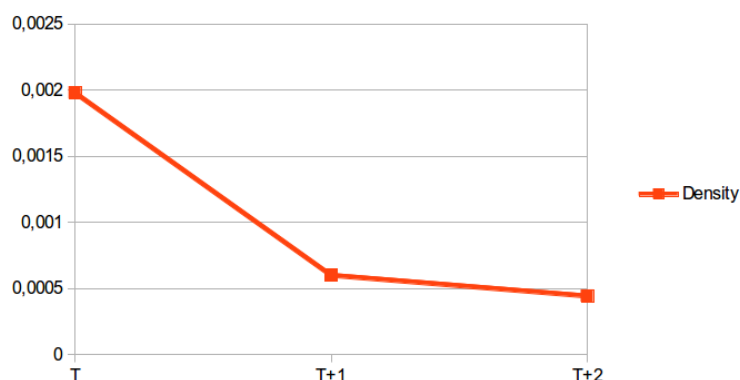


Figura 5.7: Evolución de las propiedades estructurales a nivel de red: densidad. Evento #ChampionsLeague

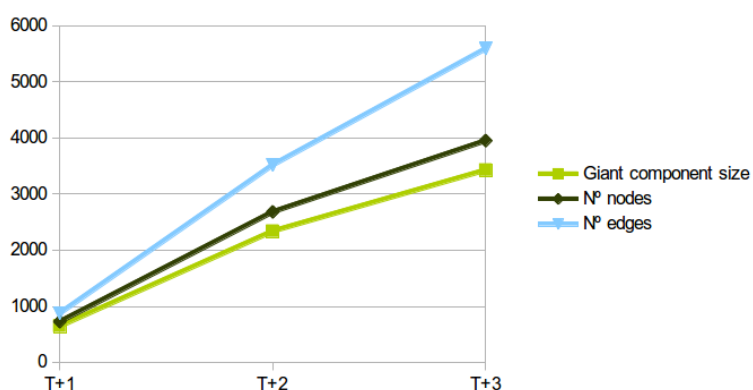


Figura 5.8: Evolución de las propiedades estructurales a nivel de red: número de nodos, número de ejes y componente gigante. Evento #ChampionsLeague

es alto con respecto a lo esperado. La figura 5.7 muestra la *densidad*, la cual tiene unos valores muy bajos si la comparamos con el evento anterior. Estos valores reflejan que no hay un intercambio de mensajes entre usuarios que se encuentran cercanos en la red y que no hay conversaciones entre dos usuarios que interactúan.

Por último, en la figura 5.8 muestra el número de nodos, el número de enlaces y el número de nodos en la componente gigante. La evolución del número de nodos en la red indica que el número de usuarios aumenta a medida que el evento avanza, así como su componente gigante. Se puede deducir que en la primera parte del evento es cuando más usuarios participan, pero en la segunda mitad del mismo el número de nodos crece más despacio y el número de enlaces continúa creciendo. Esto representa que durante el evento el número de usuarios nuevos es

T	IN DEGREE	OUT DEGREE	PAGE RANK	EIGENVECTOR	CLOSENESS	ECCENTRICITY	BETWEENESS
INICIO	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	M. COMUNICACION
	CUENTAS OFICIALES	M. COMUNICACION	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES
MEDIO	CUENTAS OFICIALES	M. COMUNICACION	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES
	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	M. COMUNICACION	M. COMUNICACION	M. COMUNICACION
	CUENTAS OFICIALES	M. COMUNICACION	CUENTAS OFICIALES	CUENTAS OFICIALES	M. COMUNICACION	M. COMUNICACION	M. COMUNICACION
	USUARIOS ANÓNIMOS	M. COMUNICACION	M. COMUNICACION	USUARIOS ANÓNIMOS	M. COMUNICACION	M. COMUNICACION	USUARIOS ANÓNIMOS
	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	M. COMUNICACION	CUENTAS OFICIALES	M. COMUNICACION	M. COMUNICACION	USUARIOS ANÓNIMOS
	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES	CUENTAS OFICIALES
FINAL	CUENTAS OFICIALES	M. COMUNICACION	CUENTAS OFICIALES	CUENTAS OFICIALES	M. COMUNICACION	M. COMUNICACION	M. COMUNICACION
	CUENTAS OFICIALES	CUENTAS DE GRUPOS	CUENTAS OFICIALES	CUENTAS OFICIALES	M. COMUNICACION	M. COMUNICACION	M. COMUNICACION
	CUENTAS OFICIALES	M. COMUNICACION	M. COMUNICACION	CUENTAS OFICIALES	M. COMUNICACION	M. COMUNICACION	M. COMUNICACION
	USUARIOS ANÓNIMOS	M. COMUNICACION	M. COMUNICACION	USUARIOS ANÓNIMOS	M. COMUNICACION	M. COMUNICACION	M. COMUNICACION

Tabla 5.2: Evolución de las propiedades estructurales a nivel de nodo: indica un top 5 de los principales roles que tienen lugar durante el evento #ChampionsLeague.

menor pero la participación en general de todos los nodos aumenta.

En la tabla 5.2 se muestran la evolución de los usuarios relevantes del evento en base a las medidas de centralidad anteriormente comentadas. Debido a la naturaleza del evento #ChampionsLeague, se puede observar como los principales roles que tienen mayor grado de entrada, page rank y eigenvector son, en su mayor parte, cuentas oficiales relativas al partido. Esto indica que dichas cuentas reciben muchas interacciones, pero no responden a los mensajes. En los valores de closeness, eccentricity y betweeness los roles más participativos hacen referencia a medios de comunicación, que son los más propensos a propagar y difundir la información, así como establecer comunicaciones en forma de noticias que son comentadas, citadas o retweeteadas por el resto de usuarios. Se aprecia también como, al principio del evento, los roles con mayor grado de salida son usuarios anónimos. Pero poco a poco los medios de comunicación se hacen con los primeros puestos a la hora de propagar los mensajes referentes a la #ChampionsLeague.

5.4.3 Evento #pesadillazamora

En este evento, se han analizado los tweets obtenidos del programa de televisión “Pesadilla en la Cocina”, en concreto el correspondiente a #pesadillazamora. De la misma forma que los anteriores, se han estudiado diferentes propiedades a nivel estructural y a nivel de nodo. La evolución en ambos entornos se ha tenido en cuenta en función del tiempo. En este caso, al tratarse de un programa de TV de una temática radicalmente diferente a los eventos que hemos visto anteriormente, se busca analizar a través de las métricas obtenidas diferentes aspectos en la estructura de la red generada.

La imagen 5.9 nos muestra la evolución del diámetro, el grado medio de conexión y la longitud del camino medio. En el *diámetro* de la red se aprecia el grado de interacción entre usuarios: al

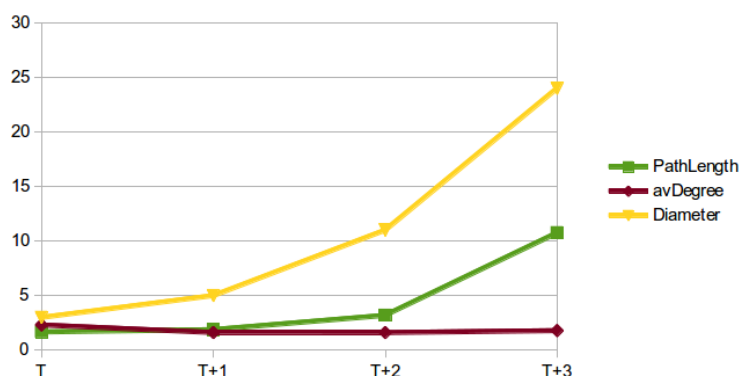


Figura 5.9: Evolución de las propiedades estructurales a nivel de red: longitud del camino medio, grado medio de conexión y diámetro de la red. Evento #pesadillazamora

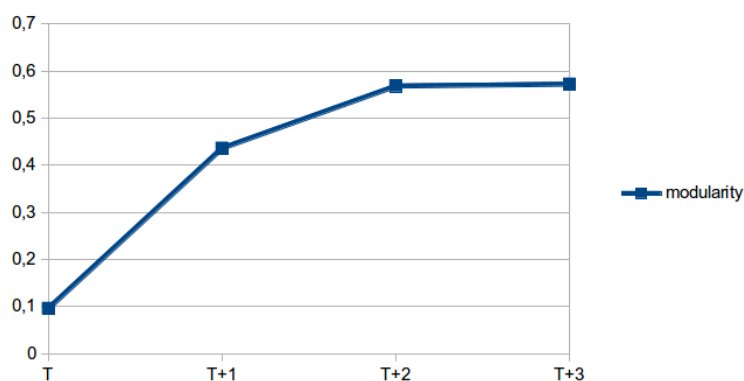


Figura 5.10: Evolución de las propiedades estructurales a nivel de red: modularidad. Evento #pesadillazamora

principio del evento el número de usuarios interactuando era menor y por lo tanto el diámetro de la red también. Después, durante el evento, el número de usuarios que se unen a la red es mucho mayor y el tipo de interacciones más frecuente es unidireccional, con lo que aumenta el diámetro de la red. En lo referente al *camino medio*, este presenta un crecimiento significativo a partir del momento T+2. Con esto se deduce que cada vez más usuarios se unen a la red con un nivel alto de interacción, incluso momentos después de que el evento termine. El *grado de conexión* inicialmente se encuentra alrededor de 2.5 y conforme avanza el programa decrece. Esto refleja que durante el programa nuevos usuarios se unen al evento participando a través de una única interacción. En la parte final del programa, el valor del grado medio de conexión vuelve a incrementarse un poco debido a una mayor participación por parte de los usuarios que ya habían participado con anterioridad.

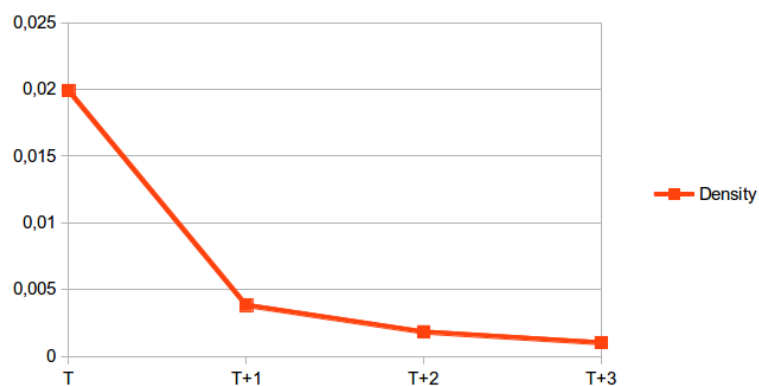


Figura 5.11: Evolución de las propiedades estructurales a nivel de red: densidad. Evento #pesadillazamora

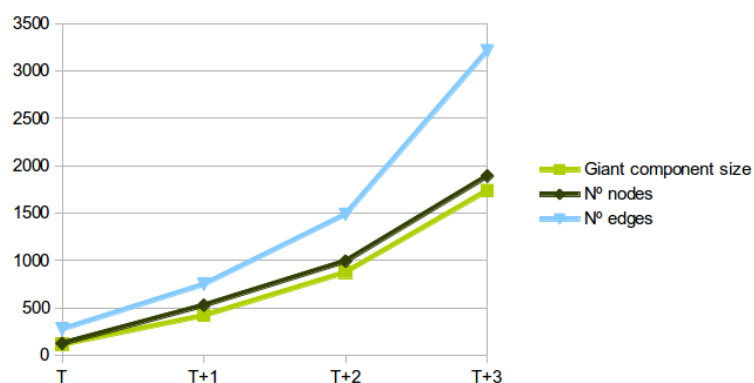


Figura 5.12: Evolución de las propiedades estructurales a nivel de red: número de nodos, número de ejes y componente gigante. Evento #pesadillazamora

En la imagen 5.10 se muestra la evolución de la modularidad. Si nos fijamos en la *modularidad*, podemos ver que es casi constante. Al tener un valor de modularidad alto, esto significa que el nivel de cohesión en las comunidades detectadas es alto con respecto a lo esperado. La evolución de la densidad de la red se puede ver en la figura 5.11 y mantiene un valor bajo que va decreciendo durante el evento. Esta medida refleja que hay pocas interacciones con respecto a todas las posibles que podría haber en el grafo.

Finalmente, la figura 5.12 muestra la evolución del número de nodos, el número de enlaces y número de nodos en la componente gigante. Anteriormente hemos hablado del aumento de la participación de los usuarios, dato que aumenta desde el principio del evento de manera constante. Es en ese punto donde crecen el número de nodos, de enlaces y la componente

T	IN DEGREE	OUT DEGREE	PAGE RANK	EIGENVECTOR	CLOSENESS	ECCENTRICITY	BETWEENNESS
INICIO	CELEBRITIES	USUARIOS ANÓNIMOS	CELEBRITIES	CELEBRITIES	CELEBRITIES	CELEBRITIES	CELEBRITIES
	CUENTAS OFICIALES	M. COMUNICACIÓN	CUENTAS OFICIALES	CUENTAS OFICIALES	M. COMUNICACIÓN	M. COMUNICACIÓN	M. COMUNICACIÓN
	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	M. COMUNICACIÓN	M. COMUNICACIÓN	CELEBRITIES	CELEBRITIES	CUENTAS OFICIALES
	CELEBRITIES	M. COMUNICACIÓN	CELEBRITIES	CELEBRITIES	M. COMUNICACIÓN	M. COMUNICACIÓN	M. COMUNICACIÓN
MEDIO	M. COMUNICACIÓN	CELEBRITIES	M. COMUNICACIÓN	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	M. COMUNICACIÓN
	CELEBRITIES	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	CELEBRITIES	CELEBRITIES	CELEBRITIES	CELEBRITIES
	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	CELEBRITIES	CUENTAS OFICIALES	M. COMUNICACIÓN	M. COMUNICACIÓN	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	CUENTAS OFICIALES	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
FINAL	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	M. COMUNICACIÓN	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	CELEBRITIES	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	M. COMUNICACIÓN	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS	USUARIOS ANÓNIMOS
	M. COMUNICACIÓN	USUARIOS ANÓNIMOS	CELEBRITIES	M. COMUNICACIÓN	M. COMUNICACIÓN	M. COMUNICACIÓN	M. COMUNICACIÓN

Tabla 5.3: Evolución de las propiedades estructurales a nivel de nodo: indica un top 5 de los principales roles que tienen lugar durante el evento #pesadillazamora.

gigante. Si nos fijamos en el número de enlaces, podemos ver como el número de mensajes aumenta, no solo durante, sino hacia el final del evento (cuando el desarrollo y la conclusión del programa tienen lugar). Con respecto a los roles que interactúan en mayor medida, en la figura 5.3 se puede ver como los usuarios que reciben más mensajes son aquellas que hacen referencia al programa, como la cuenta oficial del programa y de alguno de sus integrantes. Basándonos en los resultados de los valores de page rank y eigen vector, estos mismos usuarios también son considerados los que más relevancia tienen en la red y, por lo tanto, es más fácil para estos propagar la información. De la misma forma, se generan muchos mensajes de parte medios de comunicación y usuarios anónimos.

5.5 Conclusiones

En este capítulo hemos expuesto una serie de medidas estructurales a nivel de red y a nivel de nodo para tener una perspectiva general de las pruebas del grafo. De la misma forma, se ha validado el sistema multiagente propuesto a partir de una serie de experimentos. Hemos demostrado que dicho sistema es capaz de recopilar los mensajes generados a partir de un evento, de diseñar una red que abarca entre dos instantes de tiempo y de analizar sus características a nivel global y a nivel individual (de nodo) en la red. Dependiendo del evento y su naturaleza (tipo de programa, el día y hora en el que es emitido, si tiene difusión nacional o internacional, etc) hemos podido extraer características que nos permiten entender los usuarios que participan en dicho evento, cómo se propaga la información y el perfil general del público del evento (los roles de los usuarios que intervienen) entre otras. Como comentábamos anteriormente, Twitter es una red social que, debido a su gran número de perfiles abiertos (esto es, no privados) y a su popularidad en los últimos años, se presta a ser una herramienta interesante para estudiar la interacción entre los usuarios a partir de los mensajes generados.

Conclusiones y trabajo futuro

6.1 Conclusiones	57
6.2 Trabajo Futuro	58
6.3 Publicaciones	59

6.1 Conclusiones

El número de personas que utilizan las redes sociales online como una nueva forma de comunicación es creciente en la última década. Los mensajes que un usuario escribe en estas redes y sus interacciones con otros dejan un rastro digital que se puede almacenar y posteriormente procesar y analizar. Gracias a este hecho y a la aplicación de conocimientos del área de teoría de redes complejas, se pueden realizar análisis de los mensajes, las interacciones del usuario y de las estructuras comunicativas y sociales que emergen. Además, la información generada en las redes sociales online está etiquetada temporalmente, lo que hace posible ir un paso más en el análisis y poder estudiar la dinámica de los patrones de interacción entre usuarios.

En este trabajo, hemos diseñado y desarrollado un sistema multiagente que permite de manera distribuida y automatizada realizar el proceso de recogida, procesado y análisis de la actividad en redes sociales online. Para su diseño, hemos utilizado conceptos de organizaciones virtuales en entornos multi-agente abiertos que nos han facilitado la definición de los principales componentes del sistema. Para ello, se ha realizado un diseño del sistema a tres niveles: funcional, estructural, y dinámico. Basándonos en el diseño, hemos desarrollado una implementación utilizando la plataforma de agentes Magentix2. Los servicios ofrecidos por cada uno de los agentes han incorporado tecnologías adecuadas para la manipulación y análisis de estructuras

de red etiquetadas temporalmente. Finalmente, basándonos en medidas proporcionadas en el área de redes complejas hemos analizado varios eventos en la red social online Twitter. Este análisis lo hemos llevado a cabo teniendo en cuenta la evolución de las propiedades a nivel de red y a nivel de nodo. Como resultado de este análisis, hemos detectado patrones de evolución de la red y las características estructurales comunes, así como las diferencias entre los distintos eventos.

Durante el desarrollo de este trabajo se han alcanzado los siguientes objetivos:

- La propuesta de un diseño de SMA para el análisis de eventos en redes sociales. Se ha descrito la arquitectura propuesta desde su dimensión funcional, estructural, y dinámica.
- Se ha desarrollado un framework basado en tecnologías de agentes para el análisis de la actividad de los usuarios en las redes sociales online durante un evento. El framework está basado en una plataforma de SMA Magentix2.
- Se ha establecido un tipo de red de información etiquetada temporalmente que representa los flujos de información durante un evento.
- Se ha realizado un análisis de varios casos de estudio de eventos donde se han extraído elementos significativos que nos han permitido entender mejor lo que estaba sucediendo en la comunicación entre usuarios durante el evento.

6.2 Trabajo Futuro

El trabajo presentado compone las bases para continuar en esta línea de investigación. Entre algunos de los trabajos futuros que se contemplan están: (i) integrar más redes sociales online como pueden ser Instagram o Facebook que permita hacer un análisis de redes multicapa; (ii) mejorar la visualización de las redes de interacción que contienen un alto número de nodos y enlaces y que permita aplicar filtros para centrar la atención en determinadas partes de la red; (iii) incorporar nuevas medidas para el análisis de las interacciones teniendo en cuenta el dominio; (iv) realizar una interfaz web que permita una mejor visualización de los análisis desde cualquier tipo de dispositivo.

6.3 Publicaciones

Como resultado del trabajo presentado se han realizado dos publicaciones científicas. La primera se ha publicado en un track de Agentes Inteligentes dentro de la Conferencia Internacional de Modelos de Softcomputing en Entornos Industriales. El segundo trabajo ha sido publicado en una revista con un JCR de 1.630 (Q2).

- E. del Val, **C. Martínez**, and V. Botti. *Analyzing users' activity in online social networks over time through a multi-agent framework*. *Soft Computing* 20.11 (2016): 4331-4345.
- E. del Val, **C. Martínez** and V. Botti. *A Multi-agent Framework for the Analysis of Users Behavior over Time in On-Line Social Networks* 10th International Conference on Soft Computing Models in Industrial and Environmental Applications pp. 191-201. (2015)

Bibliography

- [1] AHN, Y.-Y., HAN, S., KWAK, H., MOON, S., AND JEONG, H. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web* (2007), ACM, pp. 835–844.
- [2] ALBERO-GABRIEL, J. Twitter,# primaveravalenciana y generación de noticias. *Cuadernos de Información y Comunicación* 19 (2014), 253.
- [3] ALDHOUS, P. The stem cell wars: data, methods and results. <https://www.newscientist.com/article/dn18996-the-stem-cell-wars-data-methods-and-results/>, 2016. [Online; accessed 22-Nov-2016].
- [4] ARGENTE, E., BOTTI, V., AND JULIAN, V. Gomas: An organizational-oriented methodological guideline for open mas. In *International Workshop on Agent-Oriented Software Engineering* (2009), Springer, pp. 32–47.
- [5] BARABASI, A. L., AND ALBERT, R. Emergence of scaling in random networks. *Science (New York, N.Y.)* 286 (1999), 509–512.
- [6] BASTIAENSENS, S., VANDEBOSCH, H., POELS, K., VAN CLEEMPUT, K., DESMET, A., AND DE BOURDEAUDHUIJ, I. Cyberbullying on social network sites. an experimental study into bystanders’ behavioural intentions to help the victim or reinforce the bully. *Computers in Human Behavior* 31 (2014), 259–271.
- [7] BENEVENUTO, F., RODRIGUES, T., CHA, M., AND ALMEIDA, V. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference* (2009), ACM, pp. 49–62.

- [8] BORGE-HOLTHOEFER, J., RIVERO, A., GARCÍA, I., CAUHÉ, E., FERRER, A., FERRER, D., FRANCOS, D., IÑIGUEZ, D., PÉREZ, M. P., RUIZ, G., ET AL. Structural and dynamical patterns on online social networks: the spanish may 15th movement as a case study. *PLoS One* 6, 8 (2011).
- [9] BORONDO, J., MORALES, A., LOSADA, J. C., AND BENITO, R. M. Characterizing and modeling an electoral campaign in the context of twitter: 2011 spanish presidential election as a case study. *Chaos: an interdisciplinary journal of nonlinear science* 22, 2 (2012), 023138.
- [10] CARMEL, D., ZWERDLING, N., GUY, I., OFEK-KOIFMAN, S., HAR'EL, N., RONEN, I., UZIEL, E., YOGEV, S., AND CHERNOV, S. Personalized social search based on the user's social network. In *Proceedings of the 18th ACM conference on Information and knowledge management* (2009), ACM, pp. 1227–1236.
- [11] CASTELLS, M., AND HERNÁNDEZ, M. *Redes de indignación y esperanza: los movimientos sociales en la era de Internet*. Alianza Editorial, 2012.
- [12] CUSICK, M. V. A.-L. B. M. Mapping the Human 'Diseasome'. http://www.nytimes.com/interactive/2008/05/05/science/20080506_DISEASE.html, 2008. [Online; accessed 22-Nov-2016].
- [13] DAFT, R. *Organization theory and design*. Cengage learning, 2006.
- [14] DEL VAL, E., PALANCA, J., AND REBOLLO, M. U-Tool: A Urban-Toolkit for enhancing city maps through citizens' activity. In *14th International Conference on Practical Applications of Agents and Multi-Agent Systems* (2016), pp. 243–246.
- [15] DEL VAL, E., REBOLLO, M., AND BOTTI, V. Does the type of event influence how user interactions evolve on Twitter? *PLOS ONE* 10(5) (2015), 1–32.
- [16] EBEL, H., MIELSCH, L.-I., AND BORNHOLDT, S. Scale-free topology of e-mail networks. *Physical review E* 66, 3 (2002), 035103.
- [17] EGGHE, L., AND ROUSSEAU, R. *Introduction to informetrics: Quantitative methods in library, documentation and information science*.
- [18] EUROSTAT. Information society statistics - households and individuals. http://ec.europa.eu/eurostat/statistics-explained/index.php/Information_society_statistics_-_households_and_individuals, 2014. [Online; accessed 19-May-2016].

- [19] FERBER, J., GUTKNECHT, O., AND MICHEL, F. From agents to organizations: an organizational view of multi-agent systems. In *International Workshop on Agent-Oriented Software Engineering* (2003), Springer, pp. 214–230.
- [20] FOSTER, I., KESSELMAN, C., AND TUECKE, S. The anatomy of the grid: Enabling scalable virtual organizations. *International journal of high performance computing applications* 15, 3 (2001), 200–222.
- [21] GLEAMVIZ. gleamviz: The Global Epidemic and Mobility Model. <http://www.gleamviz.org/gleam-in-detail/>, 2010. [Online; accessed 22-Nov-2016].
- [22] GOLBECK, J., ROBLES, C., AND TURNER, K. Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems* (2011), ACM, pp. 253–262.
- [23] HÍPOLA, P., AND VARGAS-QUESADA, B. Agentes inteligentes: definición y tipología: los agentes de la información. *El profesional de la información* 8, 4 (1999), 13–21.
- [24] HÍPOLA, P., VARGAS-QUESADA, B., AND MONTES, A. Descripción y evaluación de agentes multibuscadores. *El profesional de la información* 8, 11 (1999).
- [25] HUBERMAN, B. A. *The laws of the Web: Patterns in the ecology of information*. mit Press, 2001.
- [26] HUBERMAN, B. A., ROMERO, D. M., AND WU, F. Social networks that matter: Twitter under the microscope. *Available at SSRN 1313405* (2008).
- [27] KLEINBERG, J. Small-world phenomena and the dynamics of information. In *In Advances in Neural Information Processing Systems (NIPS) 14* (2001), MIT Press, p. 2001.
- [28] KLEINBERG, J. Complex networks and decentralized search algorithms. In *In Proceedings of the International Congress of Mathematicians (ICM)* (2006).
- [29] KLEINBERG, J. M. Navigation in a small world. *Nature* 406 (2000).
- [30] KUMAR, R., NOVAK, J., AND TOMKINS, A. Structure and evolution of online social networks. In *Link mining: models, algorithms, and applications*. Springer, 2010, pp. 337–357.
- [31] LESKOVEC, J., ADAMIC, L. A., AND HUBERMAN, B. A. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)* 1, 1 (2007), 5.

- [32] LOHR, S. The age of big data. *New York Times* 11 (2012).
- [33] LÓPEZ FOGUÉS, R. Magentix2: Una nueva plataforma para sistemas multiagente abiertos.
- [34] LOTAN, G., GRAEFF, E., ANANNY, M., GAFFNEY, D., PEARCE, I., ET AL. The arab spring— the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International journal of communication* 5 (2011), 31.
- [35] MISLOVE, A., MARCON, M., GUMMADI, K. P., DRUSCHEL, P., AND BHATTACHARJEE, B. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (2007), ACM, pp. 29–42.
- [36] NEWMAN, M. High school dating. <http://www-personal.umich.edu/~mejn/networks/>, 2004. [Online; accessed 22-Nov-2016].
- [37] NEWMAN, M., BARABASI, A.-L., AND WATTS, D. J. *The structure and dynamics of networks*. Princeton University Press, 2011.
- [38] NEWMAN, M. E. The structure and function of complex networks. *SIAM review* 45, 2 (2003), 167–256.
- [39] OUYANG, M., ZHAO, L., HONG, L., AND PAN, Z. Comparisons of complex network based models and real train flow model to analyze chinese railway vulnerability. *Reliability Engineering & System Safety* 123 (2014), 38–46.
- [40] PAGANI, G. A., AND AIELLO, M. The power grid as a complex network: a survey. *Physica A: Statistical Mechanics and its Applications* 392, 11 (2013), 2688–2700.
- [41] PEÑA-LÓPEZ, I., CONGOSTO, M., AND ARAGÓN, P. Spanish indignados and the evolution of 15m: towards networked para-institutions. *Big Data: Challenges and Opportunities* (2013), 25–26.
- [42] PERIANES-RODRÍGUEZ, A., OLMEDA-GÓMEZ, C., AND DE-MOYA-ANEGÓN, F. Introducción al análisis de redes. *El profesional de la información* (2008).
- [43] PERLIGER, A., AND PEDAHZUR, A. Social network analysis in the study of terrorism and political violence. *PS: Political Science & Politics* 44, 01 (2011), 45–50.
- [44] REBOLLO PEDRUELO, M., DEL VAL NOGUERA, E., CARRASCOSA CASAMAYOR, C., PALOMARES CHUST, A., AND PEDROCHE SÁNCHEZ, F. Consensus over multiplex network to calculate user influence in social networks. In *International Journal of Complex Systems in Science* (2013), vol. 3, Interlude, pp. 71–75.

-
- [45] ROMERO, D. M., MEEDER, B., AND KLEINBERG, J. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web* (2011), ACM, pp. 695–704.
- [46] SMITH, M. A., RAINIE, L., SHNEIDERMAN, B., AND HIMELBOIM, I. Mapping twitter topic networks: From polarized crowds to community clusters. *Pew Research Center 20* (2014).
- [47] TRAVERS, J., AND MILGRAM, S. An experimental study of the small world problem. *Sociometry* (1969), 425–443.
- [48] TWITTER. Twitter developers. <https://dev.twitter.com/>, 2016. [Online; accessed 19-May-2016].
- [49] WATTS, D. J., DODDS, P. S., AND NEWMAN, M. E. J. Identity and search in social networks.
- [50] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature* 393 (1998), 440–442.
- [51] ZANIN, M., AND LILLO, F. Modelling the air transport with complex networks: A short review. *The European Physical Journal Special Topics* 215, 1 (2013), 5–21.

