

**UNIVERSIDAD DE GRANADA**  
**E.T.S.I. INFORMÁTICA Y TELECOMUNICACIÓN**



**UNIVERSIDAD  
DE GRANADA**



Departamento de Ciencias de la  
Computación e Inteligencia Artificial

**Gestión de Información en la Web**

**Guión de Prácticas**

**Práctica 4:**  
**Caso Práctico de Análisis y**  
**Evaluación de Redes en Twitter**

Curso 2020-2021

Máster en Ingeniería Informática

# Práctica 4

## Caso Práctico de Análisis y Evaluación de Redes en Twitter

### 1. Objetivos

El objetivo de esta segunda práctica es formalizar todos los conocimientos adquiridos en el curso aplicándolos a un caso real de análisis de una red social on-line generada a partir de un medio social. Para ello, el alumno seleccionará un medio social concreto, planteará una pregunta de investigación, obtendrá un conjunto de datos del medio en cuestión, construirá una red social on-line adecuada y la analizará para responder a la pregunta planteada. En principio, nos plantearemos el uso de redes de Twitter pero el alumno puede considerar cualquier red obtenida de cualquier otro medio social (*Facebook, Instagram, Pinterest, Youtube, Flickr, Wikipedia, etc.*).

La práctica se realizará en modalidad individual. Se evalúa sobre un total de **4.5 puntos**. La fecha límite de entrega será el **domingo 30 de mayo de 2021** antes de las 23:55 horas. **Adicionalmente, se defenderá en una clase programada para la semana siguiente**. La entrega de la práctica se realizará por Internet a través del espacio de la asignatura en PRADO.

### 2. Trabajo a Realizar

El proyecto implica la realización de los siguientes pasos:

1. Selección de un medio social, definición de una pregunta de investigación y obtención de un conjunto de datos asociado.
2. Construcción de la red social on-line a analizar y visualizar.
3. Cálculo de los valores de las medidas de análisis.
4. Determinación de las propiedades de la red.
5. Cálculo de los valores de las medidas de análisis de redes sociales.
6. Descubrimiento de comunidades en la red.
7. Visualización de la red social.
8. Discusión de los resultados obtenidos.

Las subsecciones siguientes describen las actividades a realizar en cada uno de los pasos.

## 2.1. Selección del medio social, definición de la pregunta y obtención del conjunto de datos asociado

Seleccionar un medio social concreto, identificar una pregunta de investigación que sea de interés para el alumno y obtener un conjunto de datos de medios sociales que puede ayudar a responder a la pregunta planteada.

Ejemplos de preguntas podrían ser:

- ¿Los usuarios  $x_1$ ,  $x_2$  y  $x_3$  son los más relevantes en la discusión de Twitter sobre la temática  $z$  y del día  $z$ ?
- ¿Existen unos políticos concretos de cada partido que se encargan de comunicarse en Twitter o en blogs políticos con los políticos de otros partidos?
- ¿Cuáles son los actores fundamentales en la red de amistad de un personaje famoso en Twitter o Facebook?
- ¿Cuáles son los usuarios que iniciaron una discusión en Twitter sobre un hashtag concreto en un momento específico?
- ¿La localización de los usuarios tiene influencia en una discusión concreta de Twitter, creando comunidades más cohesivas?
- ¿Existe una co-ocurrencia de dos hashtags determinados en las publicaciones de los usuarios en un medio social superior a la de esos dos hashtags por separado con cualquier otro hashtag?
- ¿Las wikipedias de distintos idiomas tienen una cobertura de contenidos distinta?
- etc.

El conjunto de datos seleccionado debe permitir responder a la pregunta planteada. Se valorará la originalidad de la misma y el tamaño del conjunto de datos generado. La información podrá ser obtenida usando cualquier *scraper*, por ejemplo, usando el plugin *Twitter Streaming Importer* de *Gephi* o el *NodeXL* de *Excel* estudiados en clase. En el caso de *NodeXL* habrá que realizar la exportación de los datos en formato *csv* para poder importarlos en *Gephi*. Se valorará positivamente el empleo de cualquier otra herramienta de acceso a la API del medio social considerado.

## 2.2. Construcción de la red social a analizar y visualizar

El alumno construirá una red social apropiada para resolver la pregunta de investigación planteada a partir del conjunto de datos obtenido en el paso anterior. Se valorará positivamente el manejo de redes con un tamaño considerable para todos los apartados de la práctica, aunque pueda ser necesario reducir su dimensión para proceder a su visualización o para cualquier otra operación a realizar. **Si la red tiene una dimensión demasiado grande** habrá que reducirla para hacerla manejable para nuestro análisis. El alumno deberá aplicar un procedimiento de reducción básico pero adecuado. Por ejemplo, no es correcto seleccionar simplemente las  $x$  primeras líneas del fichero de datos obtenidos. En lugar de eso, habrá que filtrar los nodos y/o los enlaces de la red, escogiendo los nodos de mayor grado y/o los enlaces de mayor peso, en caso de ser una red ponderada; aplicando una poda Pathfinder o por umbrales de pesos. En caso de que la red

no sea ponderada se puede considerar un filtrado, por ejemplo un *k-core* (que permite seleccionar la subred más conectada al exigir un grado mínimo de  $k$  a todos los nodos que la componen), o la selección de los nodos de mayor grado y añadir todos aquellos nodos que estén conectados a los primeros en una distancia menor o igual que  $q$ .

### 2.3. Cálculo de los valores de las medidas de análisis

Se obtendrán los valores de las **medidas globales** básicas de la red social: número de nodos  $N$  y número de enlaces  $L$ , densidad  $D$ , grado medio  $\langle k \rangle$ , diámetro  $d_{max}$ , distancia media  $\langle d \rangle$ , distancia media para la red aleatoria equivalente  $\langle d_{aleatoria} \rangle = \log n(N) / \log n(\langle k \rangle)$ , coeficiente de clustering medio  $\langle C \rangle$  y coeficiente de clustering medio para la red aleatoria equivalente  $\langle C_{aleatoria} \rangle = \langle k \rangle / N$ .

El alumno analizará la **conectividad de la red**, obteniendo el número de componentes conexas así como el número de nodos y enlaces de la componente gigante y sus porcentajes con respecto a la red total.

Se almacenarán todos los valores de las medidas de análisis en una tabla, que se incluirá en la documentación de la práctica.

### 2.4. Determinación de las propiedades de la red

Se obtendrán y representarán las distribuciones de grados, distancias y coeficientes de clustering de la red completa. Si la red es dirigida, se determinarán las dos distribuciones de grados, la de entrada y la de salida.

El alumno determinará si la red social (o las redes de entrada y salida) es libre de escala analizando la distribución de grados. Para ello, debe seguir la ley de la potencia.

Se determinará si la red es un mundo pequeño de acuerdo a los valores de distancias obtenidos en la sección anterior. Para ello, la distancia media de la red debe tener una escala logarítmica con respecto al tamaño de la red  $\log n(N) / \log n(\langle k \rangle)$  o incluso menor  $\log n(N) / \log n(\log n(N))$  (*mundo ultra-pequeño*).

El alumno analizará el coeficiente de clustering medio de la red e indicará cómo de transitiva es.

### 2.5. Cálculo de los valores de las medidas de análisis de redes sociales

En el caso en que la pregunta de investigación planteada lo requiera, el alumno realizará un análisis de redes sociales básico sobre la red social diseñada usando medidas de Centralidad. Determinará los actores principales de la red mediante las medidas de *grado*, *intermediación*, *cercanía* y *vector propio*, y determinará qué medida es la más relevante para su estudio. Igualmente, si fuera necesario, realizaría un estudio de las triadas de la red.

## 2.6. Descubrimiento de comunidades en la red

Se aplicarán al menos dos métodos de detección de comunidades sobre la red estudiada para determinar la estructura modular de la red. El primero de ellos será el método de Lovaina, disponible en *Gephi*. Para el segundo método, se podrán emplear los disponibles como plugins de *Gephi*, como el de Girvan-Newman o el de Leiden; los existentes en la herramienta *DetCom* proporcionada; o cualquier otra implementación disponible. Se mostrarán los valores de la medida de modularidad asociados a cada particionamiento obtenido y se analizará la composición de las comunidades generadas para determinar si tienen algún tipo de influencia en la estructura de la red, comparando el comportamiento de los métodos ejecutados, y para tratar de identificarlas/etiquetarlas desde un punto de vista experto. Se valorará el considerar distintos valores de parámetros de los algoritmos (si existieran) para obtener detectar distintas estructuras de comunidades.

Además, se valorará positivamente realizar un análisis de centralidad particularizado a alguna de las comunidades más relevantes.

## 2.7. Visualización de la red social

Se visualizará la red social generada empleando cualquiera de los paquetes software estudiados en la asignatura (*Gephi*, *NodeXL*, etc.). El alumno empleará el algoritmo de *layout* (*distribución*) más adecuado para obtener una visualización estética. Se obtendrán distintas visualizaciones resaltando las propiedades más importantes de la red. Se usarán códigos de color y etiquetado para resaltar los nodos más importantes. Se visualizará la red completa y las partes más relevantes de la misma. Se podarán nodos y/o enlaces en caso de que sea necesario para obtener una mejor visualización. El alumno discutirá brevemente las propiedades estructurales más significativas.

## 2.8. Discusión de los resultados obtenidos

Se discutirán los resultados obtenidos. El alumno indicará si dichos resultados han permitido responder a la pregunta de investigación inicial. Se dará más importancia a la calidad de la discusión que en la cantidad.

Se compararán y discutirán las propiedades de la red obtenida con las de redes relacionadas reportadas a lo largo de la asignatura.

# 3. Documentación y Ficheros a Entregar

La **documentación** de la práctica será un fichero *pdf* que deberá incluir, al menos, el siguiente contenido:

- a) Portada con el número y título de la práctica, el curso académico y el nombre, DNI y dirección de e-mail del alumno.

- b) Una sección que describa todos los pasos realizados, organizados en subsecciones. Como mínimo, habrá de incluir:
- La pregunta de investigación planteada y una descripción del conjunto de datos y de cómo se ha obtenido.
  - Una explicación de la estructura de red obtenida a partir de los datos de medios sociales indicando claramente que representan los nodos, los enlaces, los pesos de éstos (en caso de existir), si la red es dirigida, etc.
  - Una tabla con los valores de las medidas estudiadas y un análisis de las mismas.
  - Gráficas con las distribuciones de grado y coeficientes de clustering. Una explicación de las propiedades de la red.
  - Una tabla con los valores de las medidas de centralidad para los nodos principales y un análisis de dichos valores.
  - Los resultados de los métodos de descubrimiento de comunidades y su análisis. Una explicación razonada de si tienen una influencia real o no en la red.
  - Las visualizaciones obtenidas para la red, incluyendo la explicación del procedimiento seguido para obtenerlas.
  - El análisis de resultados final en el que se comparan los valores de las medidas y las propiedades obtenidas para la red analizada con las de otras redes existentes y se responde a la pregunta planteada.
- c) Referencias bibliográficas u otro tipo de material distinto del proporcionado en la asignatura que se haya consultado para realizar la práctica (en caso de haberlo hecho).

Aunque lo esencial es el contenido, también debe cuidarse la presentación y la redacción.

El fichero *pdf* de la documentación, el fichero original del conjunto de datos estructurado obtenido, los ficheros de la red y todos los demás ficheros relacionados se comprimirán conjuntamente en un fichero *.zip* etiquetado con los apellidos y nombre del alumno (Ej. *Pérez Pérez Manuel.zip*). Este fichero será entregado por internet a través del espacio de la asignatura en PRADO.