

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



**UNIVERSIDAD  
DE GRANADA**



Departamento de Ciencias de la  
Computación e Inteligencia Artificial

## **Redes y Sistemas Complejos**

### **Guión de Prácticas**

#### **Práctica 4:** **Caso Práctico de Análisis y Evaluación de Redes**

Curso 2017-2018

Cuarto Curso del Grado en Ingeniería Informática

# Práctica 4

## Caso Práctico de Análisis y Evaluación de Redes

### 1. Objetivos

El objetivo de esta cuarta práctica es formalizar todos los conocimientos adquiridos en el curso aplicándolos a un caso real de análisis de redes complejas. Para ello, el alumno seleccionará un dominio real concreto, planteará una pregunta de investigación, obtendrá un conjunto de datos del dominio en cuestión, construirá una red adecuada y la analizará para responder a la pregunta planteada.

La práctica puede ser realizada tanto individualmente como por parejas. Se evalúa sobre un total de **2 puntos**. La fecha límite de entrega será el **viernes 15 de diciembre de 2017** antes de las 23:59 horas. La entrega de la práctica se realizará en el espacio de la asignatura en la plataforma PRADO.

### 2. Trabajo a Realizar

El proyecto implica la realización de los siguientes pasos:

1. Selección de un dominio, definición de una pregunta de investigación y obtención de un conjunto de datos estructural inicial asociado.
2. Construcción de la red compleja 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 (*sólo en caso de que la red manejada sea una red social*)
6. Descubrimiento de comunidades en la red.
7. Visualización de la red compleja.
8. Discusión de los resultados obtenidos y comparación con los existentes.

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

## 2.1. Selección del dominio, definición de la pregunta y obtención del conjunto estructural inicial

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

Ejemplos de preguntas podrían ser:

- Todas las redes complejas de tipo  $xx$  tienen la misma distribución  $yy$ .
- Existe una diferencia significativa entre la selección de los términos de búsqueda usados en distintos momentos del día.
- Aún existe un sesgo geográfico en la co-autoría de artículos científicos en la era de Internet (es decir, se siguen escribiendo artículos con co-autores localizados cerca).
- Las wikipedias de distintos idiomas tienen una cobertura de contenidos distinta.
- Los actores  $x_1$ ,  $x_2$  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.
- La influencia del juego de Messi es distinta en el Barcelona y la selección argentina.
- etc.

El conjunto de datos seleccionado debe permitir responder a la pregunta planteada. Se puede considerar un conjunto de datos asociado a algún tipo de red ya documentada como las estudiadas en el tema de redes libres de escala, en el artículo científico siguiente:

- R. Albert, A.-L. Barabasi. Statistical mechanics of complex networks. Reviews of Modern Physics 74:1 (2002) 47-97 ([http://www.stanford.edu/class/msande337/notes/200201-30\\_RevModernPhys-StatisticalMech.pdf](http://www.stanford.edu/class/msande337/notes/200201-30_RevModernPhys-StatisticalMech.pdf)).

o los disponibles en algún repositorio de Internet. Aun así, se valorará la originalidad del conjunto de datos escogido, puntuándose mejor en el caso en que éste haya sido generado por el propio alumno.

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

El alumno construirá una red compleja apropiada para resolver la pregunta de investigación planteada a partir del conjunto de datos obtenido en el paso anterior. ***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 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, seleccionando los nodos de mayor grado y/o los enlaces de mayor peso, en caso de ser una red ponderada, aplicando un filtrado o una poda Pathfinder. En el caso que la red no

sea ponderada se pueden seleccionar los nodos de mayor grado y añadir todos aquellos nodos que estén conectados a los primeros en una distancia menor o igual que  $q$ .

*Se valorará positivamente el manejar redes con un tamaño considerable para todos los apartados de la práctica*, aunque pueda ser necesario reducir su dimensión para proceder a su visualización.

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

Se obtendrán los valores de las **medidas globales** básicas: 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  $\langle d_{aleatoria} \rangle$ , coeficiente de clustering medio  $\langle C \rangle$  y coeficiente de clustering medio para la red aleatoria equivalente  $\langle C_{aleatoria} \rangle$ .

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 distribuciones de grados de entrada y salida.

El alumno determinará si la red (o las redes de entrada y salida) es libre de escala analizando la distribución de grados. Si la red cumple la propiedad de ser libre de escala, se valorará positivamente la obtención del exponente de grado, *aunque no es obligatorio para la práctica*.

Se determinará si la red es un mundo pequeño de acuerdo a los valores de distancias obtenidos en la sección anterior. El alumno analizará el coeficiente de clustering medio de la red e indicará si se comporta como una red regular, aleatoria o libre de grado.

### 2.5. Cálculo de los valores de las medidas de análisis de redes sociales (sólo en caso de que la red manejada sea una red social)

En el caso en que la red compleja diseñada sea una red social, el alumno realizará un análisis de redes sociales básico sobre ella usando medidas de Centralidad. Determinará los actores principales de la red mediante las medidas de grado, intermediación, cercanía y vector propio.

## 2.6. Descubrimiento de comunidades en la red

Se aplicará al menos un método de detección de comunidades sobre la red estudiada para determinar la estructura modular de la red. En caso de aplicar dos, se recomienda escogerlos entre el método de Lovaina (disponible en *Gephi*), el método de Girvan-Newman y un clustering jerárquico aglomerativo. Se mostrarán los valores de la medida de modularidad asociados a cada particionamiento realizado y se analizará la composición de las comunidades generadas para determinar si tienen algún tipo de influencia en la estructura de la red. Se tendrá en cuenta que la estructura de comunidades obtenida no tiene por qué ser significativa ya que no todas las redes presentan una estructura modular. Se valorará el considerar distintos valores de parámetros de los algoritmos (si existieran) para obtener detectar distintas estructuras de comunidades.

## 2.7. Visualización de la red compleja

Se visualizará la red compleja obtenida empleando cualquiera de los paquetes software estudiados en la asignatura. 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 y comparación con los existentes

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 y en la literatura especializada. En caso de haberlo calculado, comparar el coeficiente de grafo de la red libre de escala con el de otras redes de ese tipo. Se caracterizará la red en función del valor de dicho exponente. Para la comparativa con otras redes libres de escala se pueden considerar el artículo científico mostrado en la sección 2.1 y cualquier otro trabajo que se considere relevante.

# 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 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 diseñada para la red 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 grados, distancias y coeficientes de clustering. Una explicación de las propiedades de la red.
  - **Si se trabaja con una red social**, 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 reportadas en la literatura.
- 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 la plataforma PRADO2.