

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 3:
Análisis Preliminar y Visualización Básica
de una Red Social con *Gephi*

Curso 2020-2021

Máster en Ingeniería Informática

Práctica 3

Análisis Preliminar y Visualización Básica de una Red Social con *Gephi*

1. Objetivos

El objetivo de esta primera práctica es doble. Por un lado, familiarizarse con los procedimientos de análisis de redes y con las medidas habitualmente consideradas para esta tarea. Por otro, aprender el manejo de una herramienta estándar de análisis y visualización de redes como *Gephi* ¹, disponible para su descarga en <https://gephi.org/users/download/>.

Para ello, se requerirá que el estudiante realice un análisis de una red social cargándola en la herramienta, visualizándola y calculando los valores de una serie de medidas estándar de análisis de redes para estudiar las características principales de la misma.

La práctica se realizará principalmente en las propias sesiones de laboratorio de la asignatura. No es necesario entregar ningún informe resultante del trabajo realizado, aunque el estudiante podrá entregarlo **voluntariamente** para la evaluación continua del curso. La entrega se realizaría en el espacio de la asignatura en la plataforma PRADO, con fecha límite el **domingo 6 de junio de 2021** a las 23:55 horas.

2. Trabajo a Realizar

En esta primera práctica, la red a analizar será una red escogida de entre las disponibles en la literatura. Estas páginas web incluyen repositorios de redes que se pueden emplear, así como cualquier otra propuesta por el estudiante:

- Mark Newman: <http://www-personal.umich.edu/~mejn/netdata/>
- SNAP: <http://snap.stanford.edu/data/>
- Network Repository: <http://networkrepository.com/index.php>
- UciNet: <http://vlado.fmf.uni-lj.si/pub/networks/data/UciNet/UciData.htm>
- Pajek: <http://vlado.fmf.uni-lj.si/pub/networks/data/default.htm>
- Gephi: <https://github.com/gephi/gephi/wiki/Datasets>
- Alex Arenas: <http://deim.urv.cat/~alexandre.arenas/data/welcome.htm>

¹ Aunque se recomienda el uso de *Gephi* y este guión de prácticas está personalizado para esa herramienta, el estudiante puede optar por realizarla con cualquier otra de las herramientas de análisis y visualización de redes existentes.

El estudiante puede escoger la red que desee de entre las disponibles en estos repositorios o en cualquier otro pero en ese último caso deberá comunicarlo al profesor para que le confirme la asignación antes de comenzar a realizar la práctica. Esto se hace con objeto de que no se escojan las mismas redes en las prácticas de varios estudiantes. Para establecer una comunicación fluida en las dos direcciones (comunicación con el profesor y conocimiento de las redes que han escogido los compañeros), se ha habilitado un wiki en el espacio de Prado de la asignatura. **Se valorará positivamente el uso de redes con un tamaño razonable, al menos de unos pocos cientos de nodos.**

Una vez escogida la red, se cargará en *Gephi* y se realizarán tareas básicas de análisis y visualización. Si la red está en formato *Gephi*, la carga será inmediata. En caso contrario, para formatos como *GraphML* o *Net*, será necesario importarla siguiendo los pasos del wizard de *Gephi* ².

Si la red presenta más de una componente conexa, se recomienda usar *Force Atlas 2* como algoritmo de *layout* (*Distribución*). Para evitar que las componentes conexas queden fuera de la vista principal que muestra la componente gigante, fijar el valor del parámetro *Gravedad* en *Puesta a punto* a un valor entre 10 y 20. Si todo queda demasiado amontonado, se puede probar a marcar la opción *Disuadir Hubs* y/o *Evitar el solapamiento*. Los aspectos estéticos de la visualización se dejan al parecer del propio estudiante, que puede probar las distintas variantes de algoritmos de *layout* implementados en *Gephi* y de parámetros para determinar cuál le proporciona la distribución que más le guste.

Para los primeros pasos del análisis, comenzaremos por anotar los valores de las **medidas globales** básicas: número de nodos N y número de enlaces L , que aparecen directamente en la ventana *Contexto*, además de calcular manualmente el número máximo de enlaces L_{max} . Posteriormente, calcularemos otra medida global, el grado medio $\langle k \rangle$, ejecutando la opción correspondiente en la ventana *Estadísticas*. En el caso en que se nos preguntara, deberíamos especificar que la red es dirigida. Al realizar el cálculo del grado medio, obtendremos también la distribución de grados de la red completa, que debemos grabar (*Gephi* lo guarda en una carpeta con una imagen *png* y un fichero *html*).

La opción *Densidad de grafo* nos mide la relación entre número de enlaces L y el número máximo de enlaces L_{max} . La ejecutaremos y anotaremos el valor.

Posteriormente, ejecutaremos la opción *Coeficiente medio de clustering* para obtener la medida del mismo nombre, $\langle C \rangle$. Dicha opción nos proporcionará también la distribución de coeficientes de clustering de la red, que guardaremos ³.

Ahora pasaremos a analizar la **conectividad de la red**. En primer lugar, obtendremos el número de componentes conexas ejecutando la opción *Componentes conexas* y lo anotaremos. Luego nos centraremos en la componente gigante y

² Existen varios formatos estándar de redes que *Gephi* puede importar directamente. Lo normal es que las redes estén disponibles en alguno de esos formatos.

³ Hay veces que *Gephi* falla y devuelve una gráfica de coeficiente de clustering vacía. En ese caso, habrá que generarla a mano usando *Excel*. Para ello, basta con entrar en la pestaña *Laboratorio de datos* de *Gephi*, exportar los datos correspondientes en formato *csv* e importarlos en *Excel* para generar la gráfica correspondiente.

calcularemos su número de nodos. Para ello, iremos a *Filtros*, seleccionaremos *Topología→Componente gigante* y arrastramos el filtro a la ventana de abajo llamada *Consultas* donde pone *Arrastrar filtro aquí*. Entonces pulsaremos en el botón *Filtrar* con la flecha verde en la esquina inferior izquierda de la pantalla. La visualización cambiará y sólo mostrará la componente gigante. La ventana *Contexto* en la esquina superior izquierda nos mostrará el número de nodos y enlaces de dicha componente y sus porcentajes con respecto a la red total, los cuales anotaremos.

Finalmente, calcularemos las restantes **medidas globales** (diámetro d_{max} y distancia media d) sobre la componente gigante de la red ejecutando la opción correspondiente al *Diámetro de la red* en la ventana *Estadísticas*. El cálculo del diámetro nos proporciona también el valor de la distancia media, que anotaremos, así como otras muchas medidas, varias de las cuales estudiaremos en temas de teoría posteriores como por ejemplo la Centralidad.

Después de efectuar todos los pasos anteriores, el alumno guardará el proyecto desde *Gephi* nombrándolo con sus apellidos y su nombre propio. Luego almacenará todos los valores obtenidos en la tabla incluida en el fichero Excel disponible en el espacio de la asignatura en la plataforma, llamado *MedidasRedesPractical.xls*, renombrando el fichero de la misma forma.

La última tarea a realizar será escribir un pequeño análisis de la red estudiada a partir de los valores de medidas y de las gráficas de distribución de grados, etc. obtenidas. Será un análisis similar al que se realiza para la red de amistad de Facebook del profesor en las transparencias del Tema 1 de teoría. No se trata de escribir mucho sino de hacer un análisis razonable considerando los conocimientos limitados que todavía tenemos sobre el análisis de redes sociales.

3. Documentación y Ficheros a Entregar

En caso de querer entregar un informe con el trabajo realizado en la práctica, el estudiante guardará el proyecto desde *Gephi* nombrándolo con sus apellidos y su nombre propio. Luego almacenará todos los valores obtenidos en la tabla incluida en el fichero Excel disponible en el espacio de la asignatura en la plataforma, llamado *MedidasRedesPractical.xls*, renombrando el fichero de la misma forma.

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 estudiante.
- b) Una sección que incluya:
 - El nombre, descripción y fuente (página web, repositorio, etc. del que se ha obtenido) de la red escogida.
 - Una imagen de la red completa y otra de la componente gigante con una visualización lo más estética posible.
 - La tabla Excel con los valores de las medidas estudiadas incrustada.
 - Los gráficos de las distribuciones de grado, etc.

- c) Una sección que incluya el análisis de la red en función de los datos mostrados en la sección anterior.
- d) 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 *GDF* de la red, el fichero del proyecto *Gephi* y el fichero Excel con los valores de las medidas se comprimirán conjuntamente en un fichero .zip etiquetado con los apellidos y nombre del estudiante (Ej. Pérez Pérez Manuel.zip). Este fichero se entregará en el espacio de la asignatura en la plataforma PRADO.