



Asociación Española para
la Inteligencia Artificial
(AEPIA)

UIMP
Universidad Internacional
Menéndez Pelayo



UNIVERSIDAD
DE GRANADA

Recuperación y extracción de información, grafos y redes sociales

Guión de Prácticas Tema 1: Análisis de Redes Sociales

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

**Máster Universitario en Investigación
en Inteligencia Artificial**

Práctica Tema 1: Análisis de Redes Sociales

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

1. Objetivos

El objetivo de esta 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.

Para ello, se requerirá que el alumno seleccione una red social, la cargue en la herramienta, la visualice y calcule los valores de una serie de medidas estándar de análisis de redes para estudiar las características principales de la misma así como la influencia de los distintos actores que la componen.

2. Trabajo a Realizar

En esta primera práctica, la red a analizar será una red escogida de entre las disponibles en la literatura ². Estas páginas web incluyen repositorios de redes que se pueden emplear:

- Mark Newman: <http://www-personal.umich.edu/~mejn/netdata/>
- SNAP: <http://snap.stanford.edu/data/>
- 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>

El estudiante puede escoger la red que desee de entre las disponibles en estos repositorios o en cualquier otro pero 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

¹ Aunque se recomienda el uso de *Gephi*, disponible en <https://gephi.org/users/download/>, y este guión de prácticas está personalizado para esa herramienta, el alumno puede optar por escoger cualquier otra de las herramientas de análisis y visualización de redes existentes.

² Hasta marzo de 2018 usábamos redes extraídas de Facebook utilizando la herramienta de scrapping *Netvizz* pero esta aplicación ya no puede descargar redes dados los cambios recientes de la política de privacidad de apps en Facebook.

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 la asignatura en la plataforma del máster en el que cada estudiante deberá incluir la red seleccionada.

2.1. Análisis Básico de la Red

Una vez descargada la red, se cargará en *Gephi* y se realizarán tareas básicas de análisis y visualización. Si la red presenta más de una componente conexa, se recomienda usar *Force Atlas 2* como algoritmo de *layout* (en la ventana *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 alumno, que puede probar las distintas variantes de algoritmos de *layout* implementados en *Gephi* y distintos valores 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 no 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 *Coefficiente 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 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.

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.

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 el de tres medidas de Centralidad (**intermediación**, **cercanía** y **excentricidad**), que emplearemos en la siguiente sección de la práctica.

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, distancias, etc. obtenidas. Será un análisis igual al que se realiza para las redes de proteínas de la levadura y de amistad de Facebook del profesor en las transparencias del Tema 1 del curso. No se trata de escribir mucho sino de hacer un análisis razonable considerando los conocimientos limitados que tenemos sobre el análisis de redes.

2.2. Estudio de la Centralidad de los Actores

El alumno realizará un pequeño análisis de redes sociales sobre la red basado en medidas de Centralidad. Determinará los 5 actores principales de la misma mediante las medidas de **grado**, **intermediación**, **cercanía** y **vector propio**.

El valor de tres de estas medidas ya está calculado con los pasos que hemos realizado en la sección anterior. La centralidad de grado (no normalizada) se generó al calcular el *Grado medio* en la ventana *Estadísticas*. Las de intermediación y cercanía se generaron con la opción *Diámetro de la red*. En este caso, sí que es posible especificar si se desean obtener normalizadas o no normalizadas con el *checkbox Normalizar centralidades en el rango [0,1]*. Finalmente, la *Centralidad de vector propio* se calcula en la opción del menú *Estadísticas* del mismo nombre.

Una vez ejecutadas las opciones de menú correspondientes, los valores de centralidad de cada nodo pueden visualizarse en la tabla *Nodos* de la pestaña *Tabla de datos*, junto con el resto de la información asociada a cada nodo. Se pueden ordenar los nodos por columnas simplemente pulsando sobre ellas.

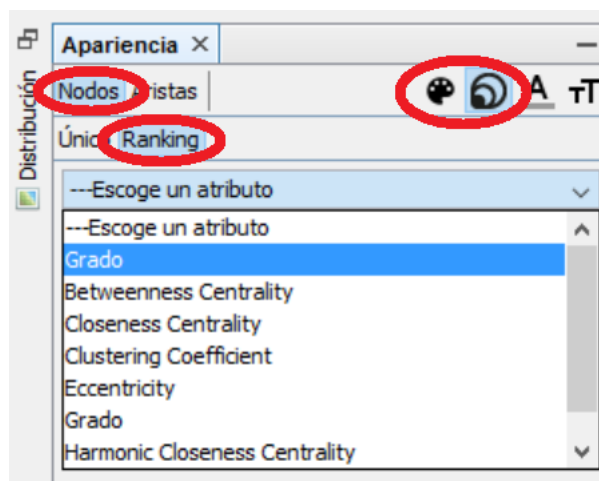
El alumno anotará los nombres de los 5 actores con mejor valor **para cada una** de las cuatro medidas anteriores, así como el valor de la medida **en cada caso** y los almacenará en una tabla como la siguiente:

Centralidad de Grado	Centralidad de Intermediación	Centralidad de Cercanía	Centralidad de Vector propio
<i>Nombre 1er actor: valor 1er actor</i>	<i>Nombre 1er actor: valor 1er actor</i>	<i>Nombre 1er actor: valor 1er actor</i>	<i>Nombre 1er actor: valor 1er actor</i>
<i>Nombre 2o actor: valor 2o actor</i>	<i>Nombre 2o actor: valor 2o actor</i>	<i>Nombre 2o actor: valor 2o actor</i>	<i>Nombre 2o actor: valor 2o actor</i>
<i>Nombre 3er actor: valor 3er actor</i>	<i>Nombre 3er actor: valor 3er actor</i>	<i>Nombre 3er actor: valor 3er actor</i>	<i>Nombre 3er actor: valor 3er actor</i>
<i>Nombre 4o actor: valor 4o actor</i>	<i>Nombre 4o actor: valor 4o actor</i>	<i>Nombre 4o actor: valor 4o actor</i>	<i>Nombre 4o actor: valor 4o actor</i>
<i>Nombre 5o actor: valor 5o actor</i>	<i>Nombre 5o actor: valor 5o actor</i>	<i>Nombre 5o actor: valor 5o actor</i>	<i>Nombre 5o actor: valor 5o actor</i>

Finalmente, realizará un pequeño análisis de los actores más importantes de la red desde una perspectiva global en función de los valores de estas medidas y el conocimiento adquirido en el Tema 1 del Bloque I-Redes Sociales del curso.

Se valorará adicionalmente la realización de gráficas adicionales tales como:

- Representaciones de la red en las que se visualicen dos de las medidas anteriores (por ejemplo, la intermediación en el tamaño de los nodos y la centralidad de vector propio en el color de los mismos) como las mostradas en las transparencias de la Sesión II.1 del curso. Estas visualizaciones pueden realizarse directamente en *Gephi*, usando las opciones *Nodos* y *Ranking* en la ventana *Apariencia*. Los dos iconos con la paleta y las bolas de distinto tamaño de la parte superior derecha de la pantalla permiten escoger qué valor de medida se desea emplear para definir el color y el tamaño de los nodos en la visualización, respectivamente:



- Gráficos que representen los valores de dos de las medidas para todos los actores de la red en ejes de coordenadas como los estudiados en la Sesión II.1. Para realizarlos, puede exportar los valores de la Tabla de datos de la red en *csv* con la opción *Exportar tabla* y generarlos fácilmente usando Excel.

3. Documentación y Ficheros a Entregar

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 *MedidasRedesPracticaParteI-1.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 título de la práctica, el curso académico y el nombre, DNI y dirección e-mail del alumno.
- b) Una sección que incluya:
 - 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, distancia, etc.
- c) Una sección que incluya el análisis de la red en función de los datos mostrados en la Sección 2.1.
- d) Una sección que describa el análisis de la centralidad de los actores de la red desarrollado en la Sección 2.2.
- e) Una sección con las visualizaciones y gráficos adicionales (**en caso de haberlos realizado**).
- f) 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 alumno (Ej. Pérez Pérez Manuel.zip). Este fichero será entregado por internet a través del espacio de la asignatura en la plataforma antes de la fecha indicada en la misma.