

Análisis de Redes Sociales

Guillermo Jiménez Díaz (gjimenez@ucm.es)
Alberto Díaz (albertodiaz@fdi.ucm.es)

Curso 2016 - 2017

Prefacio

Estos son las prácticas de la asignatura Análisis de Redes Sociales, impartida en la Facultad de Informática de la Universidad Complutense de Madrid por los profesores Guillermo Jiménez Díaz y Alberto Díaz, del Departamento de Ingeniería del Software e Inteligencia Artificial.

Este material ha sido desarrollado a partir de distintas fuentes, destacando como referencia principal el libro *Network Science* de Laszlo Barabasi, el material de la asignatura *Social Network Analysis*, impartido por Lada Adamic a través de Coursera, y las transparencias de la asignatura Redes y Sistemas Complejos, creadas por Óscar Cerdón García de la Universidad de Granada.

Este obra está bajo una [licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional](#).

Práctica 1: Análisis de una red con Gephi

La práctica se dividirá en dos partes bien diferenciadas:

1. Procesado de datos: Se generarán unos datos que **no están en forma de grafo** y que habrá que procesar para que sean visualizables en Gephi.
2. Visualización y extracción de información: Los datos procesados serán cargados en Gephi para poder visualizar la red y extraer información adicional sobre los nodos y la estructura de la red en sí misma.

1.1 Procesado de los datos

Esta práctica tiene una parte de trabajo individual y otra de trabajo en grupo. En primer lugar cada alumno deberá analizar su propia red social de contactos en Google Mail y posteriormente se deberán discutir las diferencias y similitudes entre las redes de cada uno de los alumnos del mismo grupo.

En primer lugar hay que usar la hoja de cálculo de Google a la que podéis acceder a través de este enlace: <https://goo.gl/KXnINv>. Los pasos para poder extraer nuestra red son los siguientes:

- Accede a esta hoja de cálculo y abre sesión con tu cuenta UCM (o con una cuenta de Gmail que uses habitualmente para comunicarte con otros).
- Crea una copia de la página en tu drive: **Archivo > Crear una copia**.
- Ejecuta la opción que aparece en el menú **SNA > E-mail network**. Tendrás que dar varios permisos a la aplicación para que se ejecute. No os preocupéis ya que solo extrae información que quedará guardada en vuestra hoja de cálculo. Si quieres ver más de lo que hace el script ve a la opción **Herramientas > Editor de secuencia de comandos**.
- Indica el número de conversaciones (*threads*) que quieres analizar. Inicialmente prueba con unas pocas conversaciones para ver que el script funciona. Luego puedes aumentar el número de conversaciones para conseguir una red mayor (las que os mostramos en el ejemplo se han creado con 500 conversaciones, lo que ha llevado unos 5 minutos de ejecución).

Una vez ejecutado, la hoja de cálculo se rellenará con la siguiente información:

- **Mail-id:** Identificador único de un correo en Gmail.
- **Date:** Fecha de envío/recepción del mensaje.
- **Addresses:** Lista de las direcciones de correo que aparecen en el mensaje (*A*, *De*, *Cc*), separadas por `:`. Solo aparecen los mensajes en los que hay al menos 2 direcciones (tras haber filtrado).

Lo siguiente a realizar es procesar esta hoja de cálculo (se puede descargar en formato Excel o CSV) y crear la lista de nodos y de aristas en un archivo en [uno de los formatos que Gephi es capaz de cargar](#). Por ejemplo, si creamos un fichero de nodos y un fichero de aristas en formato CSV compatible con Gephi, el aspecto de ambos archivos será algo parecido a lo siguiente:

Lista de nodos:

```
Id;Label
0;fdi-despacho411-1@ucm.es
1;fdi-adscritos-1@ggrupos.ucm.es
2;support@phaser.io
3;pedro@sip.ucm.es
```

Lista de aristas:

```
Source;Target;Type
1;2;Undirected
1;2;Undirected
0;3;Undirected
1;3;Undirected
```

1.2 Análisis y visualización de las redes generadas

Para cada una de las redes generadas (una por miembro del grupo) hay que generar un informe que incluya la siguiente información:

1.2.1 Medidas globales de la red

Primeramente calcularemos las siguientes medidas para cada una de las redes generadas: número de nodos N , número de enlaces L y densidad de la red.

Posteriormente, calcularemos las restantes medidas globales: grado medio ($\langle k \rangle$), diámetro d_{max} y distancia media $\langle d \rangle$. Para ello ejecutaremos las opciones correspondientes en la ventana **Estadísticas**.

Al realizar el cálculo del grado medio obtendremos también el histograma con la distribución de grados de la red completa.

Finalmente, ejecutaremos la opción Coeficiente medio de clustering para obtener la medida del mismo nombre, $\langle C \rangle$. Dicha opción nos proporcionará también la distribución del coeficiente de clustering en la red.

1.2.2 Medidas de centralidad

El cálculo del diámetro nos proporciona el valor de la distancia media y la distribución de distancias, así como el valor de varias medidas de centralidad (intermediación, cercanía, excentricidad). Las opciones de PageRank y Centralidad de vector propio nos permiten calcular el valor de estas otras medidas de centralidad.

Indica en el informe qué nodos tienen mayor valor en cada una de las medidas de centralidad. Explica por qué esos individuos tienen esos valores de acuerdo al significado de la red concreta que se está analizando.

1.2.3 Conectividad

Ahora pasaremos a analizar la conectividad de la red. En primer lugar, obtendremos el número de componentes conexas. 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**, en 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.

Por último, hay que obtener los valores asociados al cálculo de modularidad y el número de comunidades presentes en cada red. Juega con los parámetros del cálculo de la modularidad para intentar encontrar comunidades que tengan sentido en el contexto de la red que estamos analizando (correos de mi grupo de fútbol, de una práctica de una asignatura de la carrera...)

1.2.4 Visualización

En el informe hay que incluir al menos una imagen de cada una de las redes estudiadas, generada en Gephi, que apoye las principales conclusiones obtenidas de las redes. Por ejemplo, se podría utilizar distintos colores para los nodos que están en distintos módulos (comunidades) y distintos tamaños para reflejar una determinada medida de centralidad. Busca la visualización que más información pueda dar de la red analizada.

Para las tareas de visualización, si la red presenta más de una componente conexa, se puede 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, fijad el valor del parámetro **Puesta a punto >> Gravedad** en torno a 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 de parámetros para determinar cuál le proporciona la distribución que resulte más informativa.

1.2.5 Comentarios sobre el informe

Con todos los datos calculados de cada red hay que realizar un análisis de los resultados obtenidos con estas medidas. ¿Qué conclusiones se pueden sacar de esos datos? ¿Qué diferencias/similitudes vemos entre las distintas redes analizadas? El análisis se realizará a partir de los valores de las medidas y de las gráficas de distribución obtenidas para las redes generadas.

No se trata de escribir mucho sino de hacer un análisis razonado de la red que estamos viendo, uniendo los conceptos vistos en clase con los conocimientos que tenemos de nuestra propia red.

1.3 Entrega

La práctica se entregará en el Campus Virtual, **antes de las 23:55 del día 20 de noviembre de 2016**.

La entrega de la práctica será un archivo **.zip** (etiquetado con el número de grupo *GrupoXX*) con los siguientes contenidos:

- *Documentación.pdf*: Un archivo pdf que deberá incluir, al menos, el siguiente contenido:
 - Portada con el número y título de la práctica.
 - Número de grupo.
 - Nombre y apellidos de los integrantes del grupo.
 - Para cada una de las redes estudiadas:
 - * Tablas y gráficas con los valores de las medidas analizadas.
 - * Visualizaciones significativas de la red.
 - * Un análisis de la red en función de los datos anteriores.
 - Referencias bibliográficas u otro tipo de material distinto del proporcionado en la asignatura que se haya consultado para realizar la práctica.

- Archivos de las redes generadas y Ficheros de los proyectos Gephi de los análisis realizados **anonimizados** (eliminar las etiquetas `label`).

El archivo puede ser subido por cualquiera de los integrantes del grupo (sólo una entrega).