

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



Departamento de Ciencias de la  
Computación e Inteligencia Artificial

# **Gestión de Información en la Web**

## **Guión de Prácticas**

**Práctica 1:**  
**Análisis Preliminar y Visualización Básica**  
**de una Red Social de Facebook con *Gephi***

Curso 2016-2017

Máster en Ingeniería Informática

# Práctica 1

## Análisis Preliminar y Visualización Básica de una Red Social de Facebook 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* <sup>1</sup>.

Para ello, se requerirá que el alumno genere una red de Facebook (en este caso, una red de *likes* de páginas), 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.

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 alumno 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 **Lunes 25 de Mayo de 2017** a las 23:55 horas.

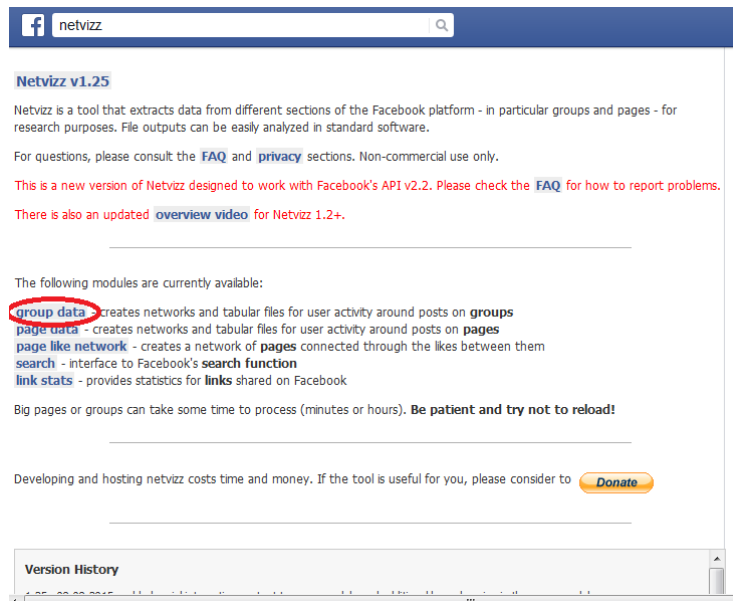
### 2. Trabajo a Realizar

En esta primera práctica, la red a analizar será una de las redes sociales de Facebook. Para obtenerla, se empleará *Netvizz*, una *app* disponible en <https://apps.facebook.com/netvizz/>, que permite generar la red en formato *GDF*, un formato estándar empleado por muchas de las herramientas de análisis de redes. Existe la posibilidad de crear distintos tipos de redes de Facebook con *Netvizz* <sup>2</sup>, nosotros usaremos la red de páginas conectadas a través los *likes* que comparten (opción *group data*). El resto de tipos de redes pueden ser empleadas para “experimentar” fuera del desarrollo de la práctica.

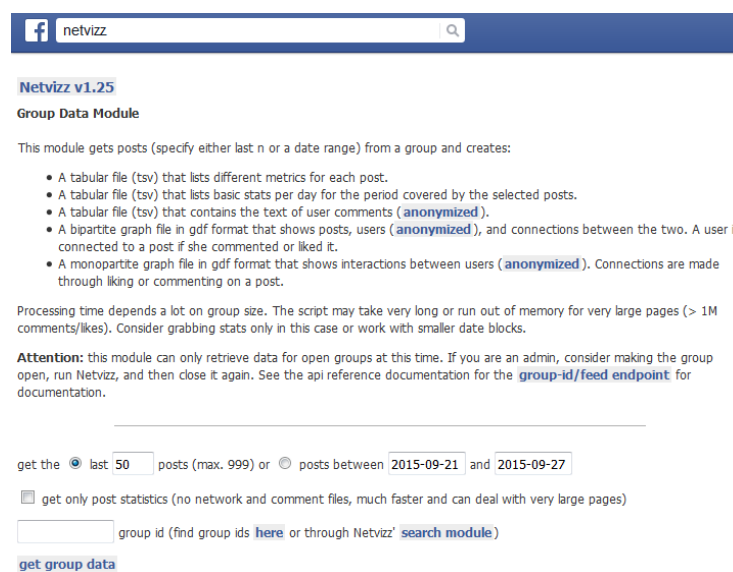
---

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

<sup>2</sup> Hasta el pasado 29 de Enero de 2015, *Netvizz* y otras aplicaciones como el *Social Network Importer* de *NodeXL* permitían obtener directamente la propia red social de amigos de un usuario o un grupo en Facebook (*personal or group friend networks*). Sin embargo, el cambio de la política de privacidad de apps en Facebook ha dado lugar a la eliminación de esta funcionalidad (<http://thepoliticsofsystems.net/2015/01/the-end-of-netvizz/>).



Debido a las últimas restricciones de seguridad de Facebook, sólo se puede extraer información de páginas de grupos **abiertos** de los que se sea miembro. Una vez seleccionada la opción *group data*, la app nos solicita el identificador del grupo en Facebook (que puede ser obtenido proporcionando la dirección de su página) y el alcance de la red (definido en forma de los  $x$  posts más recientes o de los realizados en un periodo de tiempo concreto)<sup>3</sup>. Después de especificar los parámetros y pulsar en la opción *get group data*, se genera un zip con 5 ficheros, tres de datos y dos de redes en formato GDF. La primera es una red bipartita en la que los nodos son usuarios (anonimizados) y posts mientras que los enlaces son conexiones entre ellos (usuarios que comentaron un post o lo marcaron con un “me gusta”). La segunda es una red en la que los nodos son usuarios y los enlaces son conexiones directas entre usuarios a los que les gustaba o comentaron un mismo post. Trabajaremos con esta última, una red no dirigida y ponderada ya que los pesos de los enlaces indican el número de interacciones entre dos usuarios.



<sup>3</sup> El tiempo requerido para la generación de la red depende de su tamaño. En grupos grandes, puede consumir bastante y generar ficheros muy grandes.

Una vez generada la red, se cargará en *Gephi* y se realizarán tareas básicas de análisis y visualización. **Al leer el fichero de la red en *Gephi*, se debe especificar que la red es no dirigida.** 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* 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 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.

Ahora pasaremos a analizar la **conectividad de la red**. En primer lugar, obtendremos el número de componentes conexas ejecutando la opción *Componentes conexos* 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.

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, y la distribución de distancias, que guardaremos, así como otras muchas medidas, varias de las cuales estudiaremos en temas de teoría posteriores como por ejemplo la Centralidad.

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 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 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 **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 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 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 alumno (Ej. Pérez Pérez Manuel.zip). Este fichero se entregará en el espacio de la asignatura en la plataforma PRADO.