

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

Curso 2017-2018

Cuarto Curso del Grado en Ingeniería Informática

Práctica 1

Análisis Preliminar y Visualización Básica de una Red 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* ¹, disponible para su descarga en <https://gephi.org/users/download/>.

Para ello, se requerirá que el alumno genere una red de Facebook, 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 evalúa sobre un total de **1 punto**. La fecha límite de entrega será el **Domingo 15 de Octubre 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

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* ², nosotros usaremos la red de actividad de un grupo, que conecta a los usuarios a través de los *posts* que comparten, bien marcándolos con un *like* o bien comentándolos (opción *group data*). El resto de tipos de redes pueden ser empleadas para “experimentar” fuera del desarrollo de la práctica.

¹ 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.

² 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/>).



Netvizz v1.42

Netvizz is a tool that extracts data from different sections of the Facebook platform - in particular groups and pages - for research purposes. File outputs can be easily analyzed in standard software. Please reference [this paper](#) when using Netvizz for academic work.

For **questions**, please consult the [FAQ](#) and [privacy](#) sections. Non-commercial use only.

Netvizz is being updated regularly. If you encounter a **problem**, please check the [FAQ](#) for how to report it.


The following modules are currently available:

- group data** - creates networks and tabular files for user activity around posts on **groups**
- page data** - creates networks and tabular files for user activity around posts on **pages**
- page like network** - creates a network of **pages** connected through the likes between them
- page timeline images** - creates a list of all images from the "Timeline Photos" album on **pages**
- search** - interface to Facebook's **search function**
- link stats** - provides statistics for **links** shared on Facebook

Big pages or groups can take some time to process (minutes or hours). **Be patient and try not to reload!**

Developing and hosting netvizz costs time and money. If the tool is useful for you, please consider to [Donate](#)

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. El estudiante deberá **indicar el grupo escogido en la documentación** de la práctica, explicando la temática del mismo. Una vez seleccionada la opción *group data*, la app nos solicita el identificador del grupo en Facebook (que se puede obtener 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)³. Después de especificar los parámetros y pulsar en la opción *get group data*, se genera un *zip* que contiene otros cinco ficheros, tres de estadísticas y texto de los *posts*, y dos de redes en formato *GDF*. De esas dos redes usaremos la denominada *interactions*, que define una red **no dirigida y ponderada** en la que los nodos son los usuarios (**con los nombres anonimizados**, por las restricciones de seguridad) y los enlaces no dirigidos indican una relación entre dos usuarios vía un *like* o un comentario de algún *post* de uno de ellos por parte del otro indicando el peso del enlace el número de esas interacciones.



Netvizz v1.42

Group Data Module

This module gets posts (specify either last n or a date range) from a group and creates:

- A tabular file (tsv) that lists different metrics for each post.
- A tabular file (tsv) that lists basic stats per day for the period covered by the selected posts.
- A tabular file (tsv) that contains the text of user comments (**anonymized**).
- A bipartite graph file (gdf) that shows posts, users (**anonymized**), and connections between the two. A user is connected to a post if she commented or liked it.
- A monopartite graph file (gdf) that shows interactions between users (**anonymized**). Connections are made through liking or commenting on a post.

Attention: Processing time depends a lot on group size - may take up to an hour or more. The script may run out of memory or access credits for very large group (> 1M comments/likes). Consider grabbing stats only or working with smaller date blocks.

On the first run, *always* select "get only post statistics" to get an idea of the size of the group.

This module can only retrieve data for **open groups** at this time. If you are an admin, consider making the group open, run Netvizz, and then close it again. See the api reference documentation for the [group-id/feed endpoint](#) for documentation.

group id: (find group ids [here](#) or through Netvizz' [search module](#))

date scope: ☒ last posts (max. 999)
☐ posts between and

data to get: ☒ get only post statistics (no network and comment files, much faster and can deal with very large pages)

get [group data](#)

³ El tiempo requerido para la generación de la red depende del número de *posts* especificado y del número de interacciones que haya entre ellos. Cuando se indican muchos *posts* o hay muchos usuarios con muchos *likes* o muchos comentarios, 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 conexas, 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 estudiante, 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 *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 ⁴.

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.

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.

⁴ 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.

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 *MedidasRedesPractica1.xls*, renombrando el fichero de la misma forma. El uso de un formato común nos permitirá comparar posteriormente los resultados obtenidos por todos los alumnos para ver la similitud o diferencia entre las redes analizadas por cada uno.

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 interacción entre proteínas en las transparencias del Tema 2 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.

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 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 será entregado por internet a través del espacio de la asignatura en la plataforma PRADO2.