

Práctica 1. Análisis de una red

Fecha de entrega: 2 de noviembre de 2022

El objetivo de esta práctica es analizar una red con las herramientas que se presentan en la asignatura (Python/Pandas/Gephi).

Elección de los datos

Hay que buscar **una red previamente construida**. La red seleccionada deberá contener datos que el grupo sea capaz de interpretar y que, en la medida de lo posible, resulte de vuestro interés. **La red debe tener más de 500 nodos**.

Algunos enlaces interesantes son:

- [ASU Social Computing data repository](#)
- [AMiner Datasets for Social Network Analysis](#)
- [Network Repository](#)
- [Stanford Large Network Dataset Collection](#)
- [UCI Network Data Repository](#)
- [Mark Newman's Network Datasets](#)
- [Gephi datasets](#)
- [Tore Opshal's Datasets](#)

La idea es que cada grupo analice una red distinta y que no se analicen las mismas redes que trabajaron vuestros compañeros el curso pasado, por tanto, la red que se elija no debe aparecer en la lista publicada en el campus virtual y debe de ser validada previamente por el profesor.

Análisis y visualización de la red

Hay que utilizar las dos herramientas presentadas en la asignatura, es decir, **Python+NetworkX** (posiblemente con matplotlib y Pandas, y otras librerías que podáis considerar interesantes) y Gephi.

Las herramientas se usarán para analizar la red y extraer toda la información relevante que se pueda encontrar en la misma. Esta información debe ser apoyada con imágenes de la red, tablas de métricas de la red, y gráficas que resuman métricas de la red.

¿Qué hay que entregar?

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

- **Informe en pdf** que deberá incluir, al menos, el siguiente contenido:
 - ✗ • Portada con el número de grupo y nombre y apellidos de sus integrantes.

- ✗ • Enlace a la página concreta desde que se descargó la red (la página donde se encuentra la red) y el enlace de descarga de la propia red.
- ✗ • Descripción de qué muestra la red (previa al análisis). Esto consiste básicamente en reproducir en castellano la descripción que se encuentra en la web donde la habéis descargado.
- Análisis realizado en **Gephi y NetworkX**, apoyado por **gráficos, capturas, tablas y datos** indicando de dónde se han extraído.
- Una valoración de qué para qué aspecto se ha considerado **más útil y menos útil** cada una de las herramientas utilizadas (NetworkX y Gephi).
- **Referencias bibliográficas** u otro tipo de material distinto del proporcionado en la asignatura que se haya consultado para realizar la práctica (incluyendo enlaces si se trata de recursos web). Incluye tanto el **manejo de referencias de ciencia de redes como del dominio del problema**, en caso de haber sido necesario, para validar u orientar el análisis realizado.

- **Notebook de Jupyter con lo siguiente**

- ✗ • Primera celda con el número de grupo y nombre y apellidos de sus integrantes.
- **Trabajo realizado con los datos descargados para cargarlos tanto en NetworkX como en Gephi** (si hubiera que haber realizado dicho trabajo).
 - Dicho trabajo debe ser **reproducibile en el notebook a partir del fichero de la red** para el que proporcionáis el enlace de descarga (ver más arriba).
- Análisis que incluya **los valores y las visualizaciones, y celdas de texto** donde se explique qué se está haciendo.
- **Enlace a un vídeo de máximo 6 minutos** y menos de 100MB en el que participen todos los integrantes del grupo, identificándose (“Hola soy...”) y donde cada integrante explique las principales conclusiones de la parte del trabajo que él o ella ha llevado a cabo o liderado. El archivo del vídeo puede estar en un servicio de almacenamiento en la nube o en un servidor accesible.

Falta la parte
relacionada con
NetworkX

¿Qué elementos debe incluir el análisis?

- Descripción global de la red mediante las métricas vistas en clase y gráficos de apoyo.
 - ¿Qué nos dicen dichas medidas sobre la red en el contexto del problema o dominio que representa? ¿Hasta qué punto las medidas de resumen (normalmente se trata de medias) son representativas de los individuos de la red? ¿Qué nos dice eso del problema analizado?

- Identificación de nodos relevantes en la red según las distintas medidas a nivel de nodo.
 - ¿Qué nodos tienen un papel destacado en la red? ¿Qué papel es ese en términos de ciencia de redes y cómo se interpreta dicho papel en el contexto del problema analizado?
- Identificación de distintas regiones de la red y/o comunidades dentro de la misma
 - ¿Qué comunidades existen en la red? ¿Existen regiones de la red con topologías o características notablemente distintas (p.ej. núcleo y periferia, o regiones más densas que otras)? ¿Con qué se corresponden dichas comunidades y/o regiones encontradas?
- Para responder las preguntas anteriores debes intentar superponer información del dominio que tienen los nodos y/o enlaces de la red (si la tuvieran) con la información extraída del análisis de la red (nodos relevantes, comunidades, etc).
 - Es posible que los ficheros de nodos y enlaces no tengan información adicional pero nosotros podemos ser capaces de añadirla al fichero o de hacer una interpretación “a ojo” aunque no figure. Por ejemplo, si estamos tratando una red de países es posible que no tenga información de continentes o si se trata de países ricos o pobres, pero puede ser posible para nosotros interpretar la red en esos términos.
 - Si el problema a analizar no tiene información adicional o no podemos inferirla, por ejemplo, porque analizamos una red de relaciones entre individuos anónimos, debes hacer una composición de qué rol juegan los elementos en la red, etc de forma supuesta.
- Para extraer la información se debe trabajar tanto con Gephi como con NetworkX y para cada medida calculada o gráfico se debe indicar con qué herramienta se ha obtenido.

¿Qué se valora en el trabajo?

- La puesta en juego de todos los conceptos vistos en clase y relativos a los cuatro primeros temas del curso.
- La calidad del análisis entendiendo esto como una puesta en contexto de los resultados observados tanto dentro de la red que se está trabajando, como del dominio o problema del mundo real que esa red representa.
- La calidad de la memoria como documento en el que se transmite la información encontrada, que describe las características más relevantes de la red, las contextualiza dentro del dominio del mundo real al que pertenece, y justifica cómo se han obtenido esas conclusiones adecuadamente a través de tablas, gráficos o capturas amigables al lector y efectivas (es decir, que destaquen los aspectos relevantes).

Entrega de la práctica

La práctica se entregará a través del Campus Virtual a través de una herramienta de entrega de tareas habilitada al efecto.

Solamente ha de entregar la práctica uno de los miembros del grupo.

La fecha límite para entregar la Práctica 1 será el miércoles 2 de noviembre a las 23:55. Se recomienda no dejar la entrega para el último momento para evitar problemas de última hora.