

Proyecto final de Análisis de Redes Sociales

El objetivo del proyecto es poner en práctica los conocimientos adquiridos en el curso, bien pudiéndolos aplicar a un caso real de análisis de redes complejas, o bien desarrollando simulaciones o algoritmos similares a los que serán descritos durante el curso y haciendo un análisis de los mismos.

La realización del proyecto consistirá en el desarrollo de una memoria explicativa del trabajo realizado y las conclusiones obtenidas, así como una presentación pública del mismo.

Los proyectos deberán ser acordados con el profesor de la asignatura y se deberán cumplir los plazos descritos en la Sección Planificación.

Tal y como aparece en la ficha docente de la asignatura, el proyecto supone un 60% de la nota final de la asignatura.

Análisis de un problema real usando la ciencia de redes

El objetivo general consiste en modelar por medio de una red un fenómeno del mundo real de forma que al hacerlo ganemos una mayor comprensión sobre dicho fenómeno y que usando los conocimientos vistos en clase seamos capaces de extraer conocimiento que, en el mejor de los casos, no se pueda obtener de otra forma.

Para la realización de este tipo de proyecto es necesario realizar (y describir en la memoria) las siguientes fases:

1. Obtención de datos: Descripción detallada de cómo se han obtenido los datos incluyendo al menos

- Cuál ha sido la fuente o fuentes desde la que se han obtenido, detallando el proceso necesario para extraer los datos de la fuente (p.ej. consultas realizadas a una API).
- Manipulación y transformación de los datos obtenidos para llegar a la representación que da lugar a la forma de red, detallando el proceso necesario (p.ej. filtrado o selección, agregación de elementos, etc).
- Se deben incluir el código fuente necesario para recuperar los datos (si lo hubiera) y los notebooks necesarios para procesar y transformar los datos (si lo hubiera).

Se valorará la originalidad de los datos, del enfoque para analizarlos, y el trabajo realizado para obtenerlos y manipularlos.

2. Modelado de la red. Análisis de las posibles alternativas barajadas para la representación en forma de red de los datos extraídos respondiendo preguntas como:

- ¿Qué representan los nodos y las aristas en cada alternativa considerada? Si es un grafo dirigido, ¿qué representa la dirección de las aristas? Si es un grafo con pesos, ¿qué representan los pesos? ¿Qué fenómeno del problema real representaría la red resultante?
- ¿Qué tamaño (al menos aproximado) de nodos y enlaces tendría la red?
- ¿Qué fortalezas y/o debilidades tiene usar esa alternativa para representar el fenómeno real?
- ¿Es factible su representación o análisis con los medios disponibles (fuente de datos, software, etc.)? ¿Qué decisiones se han tomado sobre los datos para llegar a esa propuesta? P.ej. sería deseable contar con los pesos, pero no es factible, o solamente se han

muestreado un conjunto de nodos pero no todos porque no es posible, etc.

Se valorará el proceso seguido en la definición de las distintas propuestas, el análisis de las mismas y las razones para elegir la representación o representaciones finales de la red.

3. Análisis de la red

- Descripción de la red mediante métricas globales: valores globales de la red (nodos, aristas, coeficiente de clustering global...), medios (grado medio, camino mínimo medio, coeficiente de clustering medio...), distribuciones de las principales magnitudes (grado, camino mínimo,...).
- Descripción de la red a nivel de nodo usando para ello las métricas vistas en clase. Para cada métrica se debe:
 - o Indicar si tiene sentido usarla en nuestra red y cuál sería su interpretación a la luz del problema real (esto es distinto a entender qué mide la métrica en un grafo abstracto).
 - o Identificar aquel nodo (o nodos) con valores más relevantes en estas métricas e interpretar a la luz de esa información qué rol desempeñan en el problema real en el que estamos trabajando.
- Descripción de la red a nivel de comunidades, regiones o patrones recurrentes.
 - o Identificar las principales comunidades de la red utilizando alguno de los algoritmos de detección de comunidades vistos en clase y asegurando que es una buena partición en comunidades tanto a matemático (modularidad más alta o similar a otras particiones en comunidades posibles que da el mismo algoritmo con otras parametrizaciones u otros

algoritmos), como a nivel de sentido dentro del problema que estamos analizando.

- o Identificar regiones de la red relevantes que pudieran ser relevantes por su topología (diferente a otras regiones de la red) o por los distintos valores que adoptan las métricas de centralidad (sin necesidad de repetir las comunidades ya identificadas salvo que vuelvan a aparecer usando un criterio distinto)
 - o Definir si algún patrón (motif) de tamaño pequeño (3, 4) podría tener especial significación en el problema del mundo real que la red representa y buscar si dicho patrón es más frecuente de lo esperado.
- Visualizaciones que faciliten la comprensión de los datos a los distintos niveles (global, comunidad/región e individual). Al menos una visualización por nivel. En cada visualización debe quedar claro qué representa el tamaño y el color del nodo y los enlaces (si es que desempeñan algún papel relevante en la visualización), si se ha realizado algún tipo de filtrado, etc.
- Descripción de las tareas realizadas durante el análisis de datos.
 - o Se debe identificar qué métricas se ha calculado con cada una de las herramientas vistas en clase.
 - o Se deben incluir los notebook con análisis en NetworkX que se han llevado a cabo.
 - o Para el uso de Gephi, se debe indicar donde sea preciso qué opciones de Gephi se han usado (parametrizaciones) para obtener los resultados (p.ej. algoritmo de comunidades usado y su parametrización) o las visualizaciones.

Se valorará el rigor en el análisis, la completitud a la hora de poner en juego todos los conceptos vistos en clase y la capacidad para extraer

información de la red y transformarlo en conocimiento útil para el problema, así como que el análisis se describa de tal forma que sea fácilmente reproducible (instrucciones paso a paso).

4. Estudio de los modelos de red sintéticos

- Ajuste de cuatro modelos de redes teóricos (Erdos-Renyi, Barabasi-Albert, configuration model y otro a elegir libremente) para intentar obtener una red que globalmente se parezca la analizada (si se han analizado varias, coger al menos una).
 - Si el modelo lo permite, el proceso de ajuste debe probar distintos parámetros en el modelo sintético hasta determinar aquellos que obtienen una red de características más similares a la real.
- Determinar si cada una de las redes simuladas se parecen o no a la real y la razón por la que esto sucede. Para ello se debe usar la distribución de grado, el grado medio, el número de enlaces, el camino mínimo medio (en la componente gigante, si hubiera más de una componente), etc

5. Conclusiones del análisis

- Detallar las limitaciones encontradas durante el proceso ya sea de los datos utilizados, del modelo utilizado o de alcance de las conclusiones extraídas.
- Describir las conclusiones más relevantes encontradas durante el análisis a nivel del problema del mundo estáis representando.
- Proponer acciones que se pueden llevar a cabo gracias a las interpretaciones que se hacen de la red. Es decir, a quién le puede servir el análisis que habéis realizado y qué acciones concretas podrían realizar teniendo en cuenta la información que de él habéis extraído. Por ejemplo, si se trata de una red de telefonía,

proponer acciones de captación o retención sobre clientes o grupos de clientes particulares.

Planificación

- El proyecto tendrá que realizarse teniendo en cuenta la siguiente planificación:
- Para el V18 de noviembre cada grupo tendrá que esbozar el proyecto en una presentación pública de 4 minutos con unas pocas diapositivas. Hay que mandar por correo el material a utilizar antes del mismo día de la presentación antes de las 13 horas.
- En todo caso, el V25 de noviembre se deberá haber cerrado la temática del proyecto con el profesor y tener definido de forma tentativa el tamaño de la red (o redes) a analizar. En el caso de que para esa fecha el proyecto no se haya cerrado, el proyecto estará suspenso.
- El día V2 de diciembre se realizará una presentación preliminar del estado del proyecto. Consistirá en una presentación pública de 6 minutos explicando el tema elegido, los pasos que se han dado, conclusiones preliminares y trabajo por hacer.
- El día X11 de enero a las 23.59h se deberá realizar la entrega de toda la documentación del proyecto a través del Campus Virtual.
- El día V13 de enero a las 13.00h (día del examen de SOC) se realizará la presentación final pública del proyecto. En este caso se dispondrá de 12 minutos para explicar de forma completa el tema elegido, las tareas realizadas y las conclusiones finales e interpretación de los resultados obtenidos.

Entrega del proyecto

El proyecto se entregará en el Campus Virtual, antes de las 23:59 del día 11 de enero de 2022. 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.
 - Memoria del trabajo
 - Referencias bibliográficas u otro tipo de material distinto del proporcionado en la asignatura que se haya consultado para realizar la práctica.
- Código fuente de las herramientas desarrolladas durante la realización del trabajo, así como sus versiones ejecutables e instrucciones de uso.
- Todos los archivos adicionales que el grupo considere relevantes (por ejemplo, archivos de Gephi, hojas excel con los datos utilizados durante el análisis y comparación, etc.).

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