

Proyecto Final

Análisis de Redes Sociales 2024/25 (1C)

Fecha	Descripción de la versión
2024-11-22	Enunciado del proyecto
2024-11-24	Cambiado el título del apartado “Optativos” para reflejar mejor su ponderación en la nota

Objetivo

El objetivo del proyecto final de la asignatura es desarrollar las habilidades y competencias que hemos adquirido, aplicándolas a una o varias redes sociales de nuestra elección.

Las redes a estudiar deberán ser las analizadas preliminarmente al principio del cuatrimestre y que fueron asignadas a nuestro grupo. Para hacer cualquier cambio en esto, escribir al profesor cuanto antes.

Requisitos mínimos

R1) Utilizar python con `uv`, `click` y `black`.

R2) **Operativas.** El programa deberá ser único, con un `main.py` como punto de entrada. Este programa deberá ser capaz de realizar distintas operativas según los argumentos y opciones de terminal. Cada una de las operativas se debe de poder ejecutar por separado.

- R2A) El programa deberá ser capaz de descargarse los datos de las redes sociales elegidas, guardando el conjunto de ficheros `html` resultantes. Se puede utilizar API, scraping, o cualquier combinación.
- R2B) El programa deberá ser capaz de analizar los datos de los `html` guardados, extrayendo de ellos la estructura de grafo (nodos, aristas, y atributos crudos de ambos), y guardando el resultado en el dataset.
- R2C) El programa deberá ser capaz de analizar los datos crudos obtenidos sobre los nodos, y extraer de ellos datos derivados (procesamiento de texto, estadísticas numéricas, etc.), guardando el resultado en el dataset.
- R2D) El programa deberá ser capaz de analizar la estructura de red, calculando diversas propiedades. Como mínimo, se deberá calcular la distribución de grados, el coeficiente de clústering promedio, e identificar los hubs.

- R2E) El programa deberá ser capaz de convertir los datos crudos y derivados obtenidos para R2C y R2D en gráficas que los permitan visualizar y entender.

R3) **Dataset.** El resultado de las operativas se guardará en un conjunto de datos o “dataset”.

- R3A) El programa deberá ser capaz de crear el dataset si hace falta, ampliar los datos contenidos si ya existe, y utilizar estos datos como punto de partida para cada una de las operativas.
- R3B) El dataset deberá usar formatos estándar e interoperables, y estar razonablemente ordenado y su estructura documentada. Algunas opciones son: directorio con subdirectorios y ficheros crudos, csvs, json, base de datos sqlite...

R4) **Informe.** Se deberá crear un informe con los resultados obtenidos. Este informe deberá al menos:

- R4A) Dar una breve introducción a las redes sociales estudiadas.
- R4B) Explicar y justificar la elección de representación en forma de grafo (cuáles son los nodos y las aristas, cuáles son sus atributos).
- R4C) Dar una breve explicación de la estructura del código, cómo se han obtenido los datos, y cómo se han analizado.
- R4D) Analizar los resultados, obteniendo conclusiones sobre las redes sociales estudiadas, comparándolas con el modelo teórico de red libre de escala y entre sí.
- R5E) El informe deberá redactarse en markdown, y se almacenará su código fuente en el GitHub de la asignatura. Para su entrega, se deberá convertir a PDF.

R5) **Defensa.** El trabajo realizado y las conclusiones obtenidas deberán presentarse en exposición oral.

- R5A) Todos los integrantes del equipo deberán participar en la defensa. El orden de presentación será determinado al azar el mismo día de la defensa, por lo que todos deberán conocer los contenidos íntegros de la presentación.
- R5B) Se recomienda utilizar materiales de apoyo visual, como una presentación de diapositivas. Se recomienda preparar la presentación en PDF o Google Slides, y compartirla con el profesor con antelación para que se puedan proyectar sin problemas. Si se quieren usar otros materiales se deberá comprobar previamente que funcionan en el aula.
- R5C) El tiempo de presentación será de 5 ± 1 minutos por cada integrante del equipo, más 5–15 minutos de preguntas.

Funcionalidad Adicional

O1) Se podrá utilizar cualquier librería a elección de los estudiantes para sustituir o mejorar el trabajo implementado manualmente en las prácticas, o implementar funcionalidad adicional.

O2) Se podrán utilizar algoritmos avanzados de análisis de redes sociales, como los que ofrece `networkx`, para por ejemplo identificar comunidades en el grafo, estimar el diámetro y la longitud de camino medio, etc.

O3) Se podrán identificar los hubs, centros u otros nodos importantes de la red, y realizar un análisis más profundo sobre ellos, por ejemplo con análisis textual de su contenido.

O4) Se podrá utilizar funcionalidad adicional de `pandoc`, como plantillas `Latex`, opciones tipográficas, etc., para mejorar el aspecto del informe resultante.

O5) Se podrá redactar el informe y realizar la presentación íntegramente en **inglés**.

Ideas y recomendaciones

Un posible análisis sería combinar algunos datos derivados, calculados en el análisis de contenido (práctica 2), con propiedades estructurales (práctica 3). Por ejemplo, polaridad del sentimiento comparada con grado del nodo, palabras más usadas en los hubs frente al resto de la red, etc. Si se dispone de datos de carácter temporal, ver la evolución en el tiempo de alguna característica, o la difusión de algún término por la red.

Si tenemos más de una red para analizar, un buen análisis podría ser hacer el mismo cálculo en todas ellas y compararlo, extrayendo conclusiones sobre las redes y sus diferencias o similitudes en base a los resultados.

En cuanto al código, aunque la mayor parte del trabajo de base ya está hecho, no conviene subestimar la complejidad de combinarlo y organizarlo para el proyecto. Es recomendable (y se valorará en la nota) hacer un análisis preliminar de las tareas a realizar, convertirlas en “issues” de GitHub, y utilizar las herramientas de GitHub para coordinar el trabajo de manera ágil y efectiva. Una primera aproximación sencilla sería directamente convertir los requisitos del proyecto en “issues”, y extraer a partir de ellas tareas más concretas.

Entrega y evaluación

1. El informe final del proyecto se entregará en moodle, en formato PDF, en la tarea habilitada para tal efecto.
2. El código se entregará a través del mismo repositorio git utilizado para las prácticas. Se deberá crear una etiqueta `entrega_proyecto` en el repositorio, que será la versión del código que el profesor corregirá. El commit deberá estar subido al repositorio remoto a tiempo para la fecha límite de entrega.
3. El repositorio debe incluir todas las instrucciones necesarias para ejecutar el programa en el archivo `README`, así como la documentación adicional que se considere oportuna. El repositorio debe incluir también cualquier archivo de configuración o

plantilla que sea necesaria para ejecutar el código.

4. La evaluación de los entregables (informe y código) se realizará en el campus virtual. Se ha incluido una rúbrica donde se puede ver los criterios de corrección que se usarán, que se recomienda consultar antes de empezar a trabajar.
5. La nota del proyecto será individual, según el siguiente desglose:
 - 6 puntos de la nota de los entregables.
 - 3 puntos de la presentación oral.
 - 1 punto de evaluación del trabajo en equipo, otorgado anónimamente por los compañeros.
 - +1 punto si se hace el informe y la presentación en inglés.
 - Hasta +1 punto extra, por la presentación preliminar que se hizo al principio del cuatrimestre.
 - En caso de obtenerse más de 10 puntos, el remanente se usará para subir la nota de prácticas.

Fechas importantes

La fecha límite para los entregables (informe y código): **2025-01-12 23:55**.

La fecha para la presentación oral del proyecto será: **2025-01-09 13:00** (fecha asignada para el examen final de la asignatura).

Se podrán incorporar mejoras al informe y el código en base a los comentarios recibidos en la defensa del proyecto, pero lógicamente la mayor parte del trabajo deberá estar hecha a tiempo para la presentación oral.