

Práctica 1

Análisis de Redes Sociales

Introducción

En la red que se facilita (hashtags_cleaned.graphml), los nodos son hashtags de un dataset recogido usando la técnica snowball, a partir del hashtag elections2020, durante las Elecciones de USA de 2020 en múltiples idiomas; una arista une dos nodos si ambos hashtags aparecieron en el mismo mensaje, y el peso (“weight”) indica el N de tweets en los que aparecieron a la vez.

Acceso al fichero: https://upm365-my.sharepoint.com/:u/g/personal/angel_panizo_upm_es/Ecce8Exo1VRBugoOMWx_lo4BuhzC3RbCZ1j9kqSf4AieA?e=H5b8LP

Ejercicios

Para realizar la práctica, se proponen una serie de ejercicios para obtener una buena representación de la red y la información que contiene.

1. Calcula las comunidades utilizando el algoritmo de Leiden. ¿Cuántas comunidades salen?
2. Utiliza Weisfeiler-lehman graph kernels para codificar cada una de las comunidades como si de grafos independientes se trataran. Basado en los resultados obtenidos, ¿qué comunidades son más similares?
3. El tamaño del grafo excede el tiempo que tenemos en la práctica, por ello, y para tener una representación más compacta, colapsa el grafo de forma que cada nodo sea una comunidad, manteniendo los atributos.
4. Aplica distintas métricas que puedan resultar útiles para tener una mejor idea de las comunidades, y crea visualizaciones diferentes en función de las métricas.
5. ¿Hay nodos centrales?, ¿qué métricas nos ayudan a saberlo? ¿Qué métricas macro nos servirían para explicar la estructura de la red, teniendo en cuenta que son comunidades y no nodos?
6. ¿Merece la pena hacer comunidades solapadas o no solapadas?
7. ¿Si visualizamos la/s comunidad/es más grandes y las más pequeñas; que se ve, que nos dice esa representación?
8. Describe los contenidos de cada comunidad. ¿Es posible, aunque sea intuitivamente, caracterizar una comunidad en función de sus hashtags más importantes?
9. Finalmente, utilizando técnicas de predicción de aristas, ¿qué comunidades tienen más probabilidades de estar unidas? Habiendo descrito dichas comunidades previamente, ¿tienen sentido estas predicciones?

Entrega

La práctica se realizará por parejas y se deberá entregar íntegramente en un cuadernillo Jupyter .ipynb donde se indicará el nombre de los participantes y, además del código correspondiente correctamente ejecutado (se deben ver los resultados de las celdas), se redactará en el propio cuadernillo las conclusiones y respuestas de cada apartado. Se pueden realizar subapartados si se considera necesario. La entrega se realizará a través de Moodle.