

Ejercicio creación de grafos en Networkx



Introducción

En este ejercicio vamos a crear un grafo en networkx a partir de datos en bruto obtenidos de Kaggle. En concreto usaremos el dataset de reviews de cervezas¹. Crearemos dos grafos, el primero será un grafo bi-partido donde los nodos serán usuarios y cervezas y los enlaces la review de un usuario sobre una cerveza. El segundo, será la proyección sobre las cervezas del grafo bi-partido anterior.

Grafo bi-parido de review

Crea un grafo bi-partito con las reviews de los usuarios y los siguientes datos:

- De los usuarios, me interesa conocer su nombre.
- De las cervezas, me interesa el nombre y su estilo.
- De las reviews, me interesa su puntuación “review_overall”

Limpieza

1. Elimina los usuarios que hayan hecho menos de 3 reviews.
2. Elimina las cervezas que hayan recibido menos de 10 review.
3. Quédate con la componente conexas más grande.

¹ <https://www.kaggle.com/datasets/thedevastator/1-5-million-beer-reviews-from-beer-advocate/>

Grafo proyecciones de cervezas

Para crear el grafo sigue estos pasos:

1. Selecciona solo aquellas reviews con una puntuación de 5. Usar el método `edge_subgraph2`
2. Quédate con la componente conexas más grande.
3. Haz la proyección en las cervezas.
4. Elimina aquellas aristas con un peso menor que 4 (que no hay al menos 4 usuarios que les gusten ambas cervezas).
5. Quédate con la componente conexas más grande.

Overview

Vamos a hacer análisis de redes para mostrar las cervezas más relevantes del dataset, para ello:

1. Calcula las comunidades del grafo proyección ¿Qué representa cada comunidad?
2. Muestra las 3 cervezas con pagerank más alto de cada comunidad. Estas cervezas son las más relevantes del dataset ¿Por qué?

²

https://networkx.org/documentation/stable/reference/classes/generated/networkx.Graph.edge_subgraph.html