

Unidad 3. Detección de comunidades

Aglomeración jerárquica, espectral, markoviana, propagación de afinidad, basada en información. Fortalezas y debilidades de cada método. Comparación con el aglomeramiento métrico. Ejercitación: Comunidades en redes biológicas.

1. Construya un grafo G que consista de 3 subgrafos completos de 7 nodos cada uno, y agregue 3 aristas más, que unan los subgrafos individuales entre sí para formar un único grafo conexo. Grafíquelo. ¿Cuántas comunidades esperaría encontrar en él?
2. Sobre el grafo de 1, aplique los distintos métodos de detección de comunidades:
 - a. De agrupamiento aglomerativo
 - b. De agrupamiento divisivo
 - c. Espectral
 - d. Multinivel
 - e. Isomap¿Encuentra alguna diferencia entre los métodos?
3. Repita (2), pero esta vez sobre un grafo aleatorio de Erdos-Renyi con parámetros $n = 100$, $p = 0.05$. Antes de ejecutar los métodos, ¿qué estructura de comunidades esperaría encontrar? Mida la modularidad de los agrupamientos resultantes de cada método, compártelas entre sí y con los resultados del punto (2) ¿Puede decir algo al respecto?
4. Para el grafo del club de karate ya conocido, construya la matriz de similitud Ω propuesta en la Teórica

$$\omega(v, w) = \frac{|\Gamma(v) \cap \Gamma(w)|}{|\Gamma(v) \cup \Gamma(w)|},$$

Donde

- $\Omega_{v,w} = \omega(v, w)$ es la medida de similitud entre los nodos v y w ,
- $\Gamma(u)$ es el conjunto de nodos vecinos al nodo u y
- $|X|$ indica la cardinalidad (i.e., cantidad de elementos) del conjunto X .

¿En qué se diferencia esta medida de similitud de las dos alternativas que ofrece `igraph`, `similarity_jaccard` y `similarity_dice`?

5. Elija dos métodos de detección de comunidades (llamémoslos A y B), y un constructor de grafos aleatorio de su preferencia que acepte como parámetro el número de nodos, llamémoslo $C(n)$. Analice cómo varía el tiempo de ejecución de A y B sobre los grafos generados a partir de C , en función de n . Presente la información de manera clara en un único gráfico.
6. Para cada grafo producido por C en el punto (5), compare las comunidades encontradas por A y B, usando la métrica de “precisión” vista en clase. ¿Cómo varía esta métrica como función de n ? ¿Puede decir algo al respecto? Presente la información de manera gráfica.