

Seminario de Ciencia de Datos

2^{do} Cuatrimestre de 2017

Módulo 4 - Redes

El archivo `data.csv` contiene una lista de publicaciones de física de altas energías del repositorio ArXiv. En esta práctica vamos a estudiar el grafo de colaboraciones entre autores, en el que los nodos representan autores y las aristas coautorías. Vamos a usar un grafo simple no dirigido, en el que basta una coautoría para que haya una conexión entre dos autores.

1. Creación del Grafo

A partir del archivo de datos, genere el grafo de colaboraciones entre autores. Cuente el número de nodos y aristas. Haga una representación gráfica del grafo.

2. Distribución de Grado

Grafique la distribución del *grado* de los autores, en escalas lineales, semilogarítmica, y log-log. Argumente, a partir de los gráficos, de qué tipo de distribución se trata.

3. Componentes Conexas

Calcule el número de componentes conexas del grafo. Muestre el tamaño de la componente mayor, o *componente gigante*.

4. Tamaños de Vecindades

Trabajando con la componente gigante del grafo, estudie, parándose en un nodo al azar, cómo aumenta el número de autores alcanzados a medida que se aleja del nodo semilla. Grafique el número de autores alcanzados en función de la distancia al nodo semilla. Grafique también el número de nuevos autores que se agregan en cada paso, y estime el máximo de esta función. Pruebe con varios nodos semillas y analice la robustez de este resultado. Discuta el significado de los gráficos y su relación con el fenómeno de *seis grados de separación*.

5. Mundos Pequeños

Compute el coeficiente de clustering C y el camino mínimo medio l para la componente gigante. Genere un grafo aleatorio con la misma distribución de grado y compute las mismas medidas para este grafo. Compare e interprete los resultados. ¿Se trata de un grafo con estructura de mundos pequeños?

6. Estrellas

Discuta cómo haría para individuar a los autores “estrella” del campo. Evalúe quiénes son estos autores según por lo menos dos métricas diferentes. Analice qué sucedería con la comunidad si estos autores desapareciesen. En particular, determine cuántos autores deberían desaparecer para que desaparezca la componente gigante del grafo.