

Big data y Ciencias Sociales Computacionales

Triangulación: no quedarse con los datos de un dataset y con los supuestos del mismo.

Propiedades de los grafos para las redes sociales:

- Los nodos de un grafo son las personas generalmente.
- Los nodos se contagian entre sí. Se habla de **viralidad**.

Idea de centralidad e influencia: un nodo por el que pasan muchos caminos mínimos entre pares de nodos es un nodo de alta centralidad.

Homofilia: tendencia de “amigos” a parecerse entre sí. Uno puede calcular la distancia entre dos personas en un grafo (grafo de personas), calcular la distancia entre gustos y así encontrar si hay homofilia en tal gusto a partir de un grafo. Es clave explotar este concepto en grafos. Por ejemplo se puede usar para deducir datos faltantes.

Difusión compuesta: de todos estos atributos nos interesan, para aprender algo en particular se debe aprender de un subgrupo en particular.

Mundo pequeño: si dada una red que crece, la distancia entre nodos crece a una velocidad logarítmica (no aumenta a la misma velocidad que el crecimiento de la red).

Los nodos de alta centralidad achican los grafos, funcionan como **puentes**. Además generan más relaciones de homofilia.

Análisis de texto

- contar palabras
- sentiment analysis: Tratar de entender la emoción expresada en una palabra o si una crítica es positiva o negativa. Herramienta: LIWC, es una lista de palabras “positivas”, “negativas”, y otras categorías.
- detectar odio, fake news, “clickbait”
- entender semántica
- “culturomics” y humanidades digitales

A veces se pueden chequear noticias falsas en base a los comentarios de la gente. También en base al embedding del texto de la noticia se puede relacionar con otras noticias.