

# Análisis de Redes Sociales

Guillermo Jiménez Díaz (gjimenez@ucm.es)  
Alberto Díaz (albertodiaz@fdi.ucm.es)

1 de octubre de 2014

## Prefacio

Estos son los apuntes de la asignatura Análisis de Redes Sociales, impartida en la Facultad de Informática de la Universidad Complutense de Madrid por los profesores Guillermo Jiménez Díaz y Alberto Díaz, del Departamento de Ingeniería del Software e Inteligencia Artificial.

Este material ha sido desarrollado a partir de distintas fuentes, destacando como referencia principal el libro *Network Science* de Laszlo Barabasi y el material de la asignatura *Social Network Analysis* impartido por Lada Adamic a través de Coursera.

## Tema 1: Teoría de Grafos

En este tema haremos un repaso de los principales conceptos de teoría de grafos necesarios para introducirnos en el análisis de las redes sociales.

### Introducción

TODO: Puentes de Königsberg

Euler construyó un grafo para representar y resolver el problema. Esta se puede considerar como la primera representación conocida de un grafo como medio para resolver un problema. De este hecho se pueden sacar dos conclusiones:

- Algunos problemas son más simples de resolver si se representan mediante un grafo.
- Los grafos presentan ciertas propiedades que limitan su comportamiento.

## Redes o grafos

Las redes o grafos se componen de:

- Un conjunto de nodos o vértices.
- Un conjunto de aristas o enlaces que los unen.

Los enlaces pueden ser de dos tipos:

- *Dirigidos*: Tienen una dirección. Si todos los enlaces de un grafo son dirigidos entonces este grafo se conoce como *grafo dirigido o digrafo*. Por ejemplo, un enlace de una página web se representaría como un grafo dirigido.
- *No dirigidos*: Representan relaciones bidireccionales. Si todos los enlaces de un grafo son no dirigidos entonces se le llama *grafo no dirigido*. Por ejemplo, la relación de estar casados entre dos personas es un enlace no dirigido.

Una arista o enlace puede tener atributos:

- Un peso, el cual puede representar información como la frecuencia de llamadas entre dos individuos. El peso puede tener un valor negativo, por ejemplo si lo usamos para representar la confianza-desconfianza entre dos individuos.
- Un tipo, el cual puede representar el tipo de relación entre dos individuos (familiares, amigos, ...)
- Otras propiedades que dependen del resto de la estructura del grafo, como el grado de intermediación (*betweeness*).

Un sistema se puede representar de diferentes formas dependiendo del conjunto de nodos usados y, sobre todo, de los enlaces que se definan. Un mismo conjunto de nodos puede definir distintos grafos.

## Propiedades básicas de los grafos

El número de nodos o vértices de un grafo ( $N$ ) representa el tamaño de la red mientras que el número de aristas o enlaces ( $L$ ) representa el número de interacciones entre los nodos.

El número máximo o total de enlaces ( $L_{max}$ ) es el número de enlaces supuesto que todos los nodos están conectados con todos y se calcula como:

$$L_{max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

En la mayoría de las redes reales se cumple que  $L \ll L_{max}$  por lo que se suele decir que estas redes son *dispersas* (*sparse*).

Otra propiedad básica pero bastante importante de los grafos (y que usaremos a menudo a lo largo de la asignatura) es el grado de un nodo  $i$  ( $k_i$ ), que es el número de enlaces que tiene  $i$  con otros nodos.

En un *grafo no dirigido*, si conocemos el grado de cada uno de sus nodos entonces podemos calcular el número de aristas ( $L$ ) de la siguiente forma:

$$L = \frac{1}{2} \sum_{i=1}^N k_i$$

El grado medio ( $\langle k \rangle$ ) de un grafo no dirigido se calcula como:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2L}{N}$$

En un *grafo dirigido* distinguimos entre grado de entrada ( $k_i^{in}$ ) y grado de salida ( $k_i^{out}$ ). El primero representa el número de enlaces que llegan al nodo mientras que el segundo representa el número de enlaces que salen del nodo. El grado total es la suma de ambos grados ( $k_i = k_i^{in} + k_i^{out}$ ). En este caso, el número de aristas del grafo se puede calcular como:

$$L = \sum_{i=1}^N k_i^{in} = \sum_{i=1}^N k_i^{out}$$

Y el grado medio ( $\langle k \rangle$ ) será el siguiente:

$$\langle k^{in} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{in} = \langle k^{out} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{out} = \frac{L}{N}$$

Para un grafo de tamaño conocido  $N$  podemos calcular la **distribución de los grados del grafo**, es decir, cuál es el porcentaje de nodos que tienen un determinado grado en dicho grafo. Esta distribución se puede representar mediante un histograma, donde en el eje X representamos los distintos grados presentes en el grafo ( $k$ ), mientras que en el eje Y representamos la proporción de nodos con grado  $k$  ( $p_k$ ):

$$p_k = \frac{N_k}{N}$$

donde  $N_k$  representa el número de nodos de grado  $k$  presentes en el grafo.  $p_k$  representa la probabilidad de que el grado de un nodo seleccionado aleatoriamente sea  $k$  y se cumple que:

$$\sum_{k=1}^{\infty} p_k = 1$$

Como veremos más adelante, la distribución de los grados será una propiedad importante que nos servirá para derivar otras propiedades de la red. Por ejemplo, a partir de  $p_k$  podemos calcular el grado medio del grafo:

$$\langle k \rangle = \sum_{k=1}^{\infty} k \cdot p_k$$

## Representación de un grafo

Existen distintas formas de representar un grafo. Las más comunes son las siguientes:

TODO: ejemplos de mismo grafo con distintas representaciones

- **Listas de enlaces.** Consiste en reopresentar un grafo como una lista de  $L$  pares  $(i, j)$  que representan los enlaces de la red:

$$G = \{(i_1, i_2), (i_1, i_n), \dots, (i_m, i_n)\}$$

- **Listas de adyacencia.** Es una forma reducida de la anterior. Para cada nodo escribimos una lista con cada uno de sus nodos adyacentes:

$$1 \{2, \dots n\}$$

$$2 \{1, \dots\}$$

$$\vdots$$

$$n \{1, \dots m\}$$

- **Matriz de adyacencia.** Es una matriz de tamaño  $N \times N$  de 0s y 1s. Si el elemento  $A_{ij} = 1$  entonces implica que hay un enlace que va de  $j$  hacia  $i$ . Si el elemento  $A_{ij} = 0$  entonces no existe tal enlace. En un grafo no dirigido se cumple que elemento  $A_{ij} = A_{ji}$  por lo que la matriz es simétrica. Si las aristas tienen un peso entonces  $A_{ij} = w_{ij}$ .

Las matrices de adyacencia nos ayudan a calcular fácilmente el grado de cada uno de los nodos:

- En un grafo no dirigido  $k_i = \sum_{j=1}^N A_{ij} = \sum_{i=1}^N A_{ij}$