

# Análisis de Redes Sociales

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

22 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, el material de la asignatura *Social Network Analysis*, impartido por Lada Adamic a través de Coursera, y las transparencias de la asignatura Redes y Sistemas Complejos, creadas por Oscar Cordón García de la Universidad de Granada.

## Tema 3: Redes libres de escala

Tal y como hemos visto en el capítulo anterior el modelo de Erdős-Renyi o de red aleatoria presenta algunas propiedades que no se cumplen en las redes reales. De entre ellas destacamos que el modelo de redes aleatoria no contempla la existencia de hubs o concentradores, nodos con un grado mucho mayor que el resto de nodos de la red.

En este tema vamos a estudiar un nuevo modelo de red que se ajusta mucho mejor a las redes reales, conocido como **redes libres de escala**.

### 3.1 La WWW: el primer mapa de red no aleatoria

En 1999 un grupo de investigadores realizó el primer mapa de una porción de la WWW con el objetivo de verificar las propiedades de los modelos de red aleatoria en este sistema complejo de páginas y enlaces. Hasta aquel momento había razones suficientes para suponer que esta red seguía este modelo de red

aleatoria. Compusieron el mapa del dominio `nd.edu`, con un total de 300000 documentos (nodos) y 1,5 millones de enlaces.

Aunque a primera vista parecía que este mapa cumplía esa aleatoriedad predicha por el modelo. Sin embargo, se pudo observar que, conviviendo con un gran número de nodos de grado pequeño, existen nodos con un número excepcional de enlaces: los hubs o concentradores. Recordemos que estos nodos son extremadamente improbables en las redes aleatorias por lo que este estudio dio lugar a replantearse que tal vez este modelo no era el que servía para modelar esta red. Así mismo, como veremos más adelante, se observó que estos hubs aparecen en otras muchas redes reales.

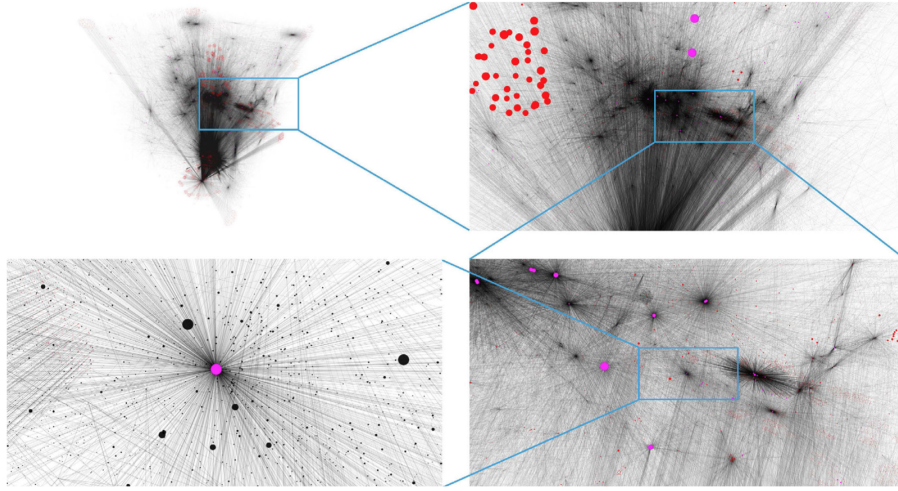


Figure 1: Mapa de la WWW: se puede ver la existencia de hubs.

El análisis de los datos de la WWW muestra que la distribución de grados no se asemeja a la distribución de Poisson que predice el modelo de redes aleatorias sino que dicha distribución se asemeja a una **ley potencial** o *power law*. Esta distribución se caracteriza por tener una cola mucho más ancha que una exponencial. La diferencia entre una y otra es mucho más clara si las representamos en una escala logarítmica.

De acuerdo a esto, la distribución de grados en la WWW se puede representar como:

$$p_k \sim k^{-\gamma} \rightarrow \log p_k \sim -\gamma \cdot \log k$$

Como la WWW es una red dirigida podemos distinguir entre la distribución de grados de entrada y la distribución de grados de salida:

$$p_{k_{in}} \sim k^{-\gamma_{in}} \quad p_{k_{out}} \sim k^{-\gamma_{out}}$$

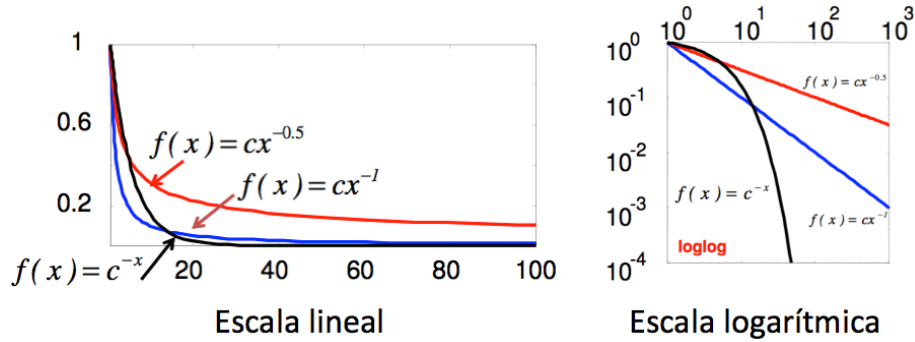


Figure 2: Comparativa de una función exponencial (negro) frente a funciones de ley potencial (azul y rojo). A la izquierda se presenta en escala lineal mientras que a la derecha se ve en escala logarítmica

De acuerdo a los datos usados por Barabasi y Albert se pudo estimar que  $\gamma_{in} \approx 2,1$  mientras que  $\gamma_{in} \approx 2,45$ .

Lo realmente interesante de este estudio no fueron los datos resultantes obtenidos sino la gran diferencia entre la distribución esperada (una Poisson) y la distribución final obtenida (una ley potencial). En la siguiente figura se observa claramente estas diferencias obtenidas:

Esto dio a entender que esta red real no se comportaba de acuerdo al modelo de red aleatoria sino que era “otro tipo de red”

## 3.2 Redes libres de escala

Una **red libre de escala** es una red cuya distribución de grados sigue una ley potencial. Esta distribución se caracteriza por tener una larga cola. Para esta distribución se pueden emplear dos formulaciones distintas aunque las propiedades de la red son independientes del formalismo empleado.

### 3.2.1 Formalismo discreto

En este caso consideramos que el grado de un nodo es siempre un entero positivo ( $K = 0, 1, 2 \dots$ ). De acuerdo a esto, la probabilidad de que un nodo aleatorio tenga grado  $k$  se puede expresar como:

$$p_k = C \cdot k^{-\gamma}$$

La constante  $C$  se puede calcular teniendo en cuenta que  $\sum_{k=1}^{\infty} p_k = 1$ :

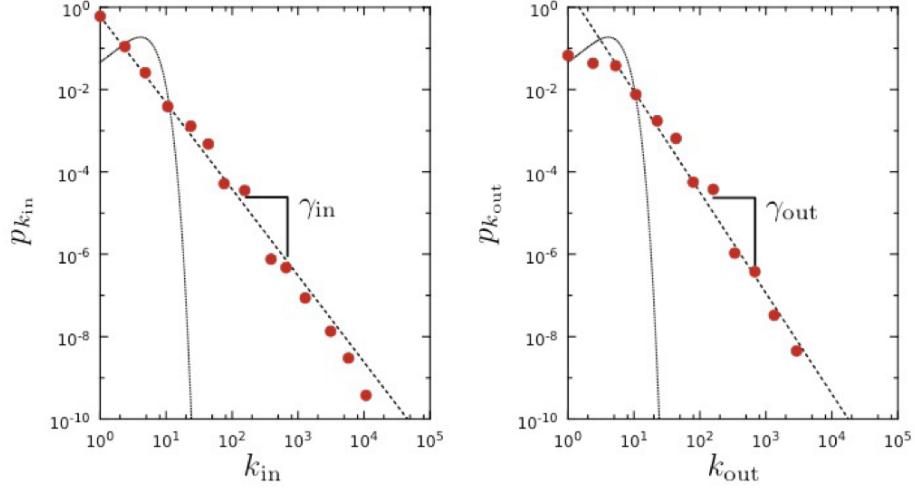


Figure 3: Distribución de grados de entrada y de salida de la WWW frente a la Poisson (en escala logarítmica)

$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)}$$

donde  $\zeta(\gamma)$  es la función Zeta de Riemann.

Para valores de  $k > 0$  se cumple que la función de distribución es:

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Los nodos aislado ( $k = 0$ ) tienen que tratarse de manera independiente.

### 3.2.2 Formalismo continuo

En este formalismo consideramos que  $k$  puede tomar cualquier valor real positivo. En este caso:

$$p(k) = C \cdot k^{-\gamma}$$

La constante  $C$  se puede calcular teniendo en cuenta que  $\int_{k_{min}}^{\infty} p(k) dk = 1$ :

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1) \cdot k_{min}^{\gamma-1}$$

De acuerdo a esto, la función de distribución según el formalismo continuo es:

$$p(k) = (\gamma - 1) \cdot k_{min}^{\gamma-1} \cdot k^{-\gamma}$$

Donde  $k_{min}$  es el menor grado de la red (o función de distribución). En este caso, la función de distribución se ha de interpretar como la probabilidad de que un nodo tenga un grado entre dos valores  $k_1$  y  $k_2$ :

$$\int_{k_1}^{k_2} p(k) dk$$

Finalmente, podemos ver que las redes reales cumplen otra nueva ley que es la **propiedad libre de escala**:

Muchas redes reales presentan distribuciones de cola ancha. Esto implica que nodos de grado bajo conviven con nodos con un grado excepcionalmente grande: los hubs.

### At a glance: Random networks

- *Definition:*  $N$  nodes, where each node pair is connected with probability  $p$ .
- *Average degree:*  $\langle k \rangle = p(N-1)$
- *Average number of links:*  $\langle L \rangle = \frac{p(N-1)}{2}$
- *Degree distribution:*  $p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$ .

For sparse networks ( $k \ll N$ ),  $p_k$  has the Poisson form

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}.$$

- *Giant component ( $N_g$ ):*

$\langle k \rangle < 1$ : no giant component ( $N_g \sim \ln N$ )

$1 < \langle k \rangle < \ln N$ : one giant component and disconnected clusters

$$\left( N_g \sim N^{\frac{2}{3}} \right)$$

$\langle k \rangle > \ln N$ : all nodes join the giant component  $N_g \sim (p - p_i)N$

- *Average distance:*  $\langle d \rangle \propto \frac{\log N}{\log \langle k \rangle},$
- *Clustering coefficient:*  $C = \frac{\langle k \rangle}{N}.$

**Box 3.8**

Figure 4: Cuadro resumen (extraído de *Network Science*, pp. 70)