

# Análisis de Redes Sociales

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

24 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 Óscar Cerdón García de la Universidad de Granada.

Este obra está bajo una [licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional](#).

## Tema 4: Modelos de crecimiento

Tras haber visto los dos temas anteriores en los que hablábamos del modelo de red aleatoria y las redes libres de escala nos pueden surgir estas preguntas:

- ¿Por qué el modelo aleatorio no reproduce hubs ni sigue una distribución de ley potencial?
- ¿Por qué la mayoría de los sistemas reales complejos convergen a una topología de red libre de escala?

Para entender esta información estructural es necesario entender el mecanismo responsable de que aparezca la propiedad libre de escala. Por tanto, en este tema dejaremos un poco de lado los modelos de la topología de la red para centrarnos en algunos **modelos de crecimiento y evolución de los sistemas complejos**.

## 4.1 Los modelos de evolución influyen en la topología de la red

El modelo de evolución de una red aleatoria es muy sencillo:

- Partimos de un conjunto fijo de nodos  $N$ .
- Creamos aleatoriamente enlaces entre los nodos.

Este modelo de evolución crea una red con una determinada topología (ya estudiada en el tema 2).

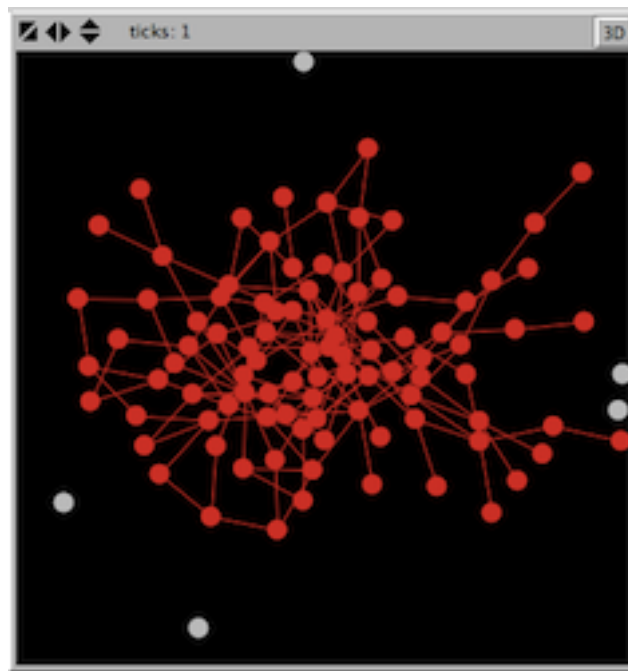


Figure 1: Topología de una red aleatoria

La intuición nos puede hacer pensar que si modificamos el modelo de evolución la red generada puede ser diferente. Para ello vamos a comenzar con unos modelos simples de evolución [^1] a partir del modelo aleatorio:

[^1] Estos modelos simples pueden verse en NetLogo usando el archivo **Modelos de Evolución Simples** que está en el Campus Virtual.

### 4.1.1 Modelo de presentación

Es un modelo similar al modelo aleatorio pero en el que aumentamos la probabilidad de los nodos que están enlazados al nodo al que se quiere añadir un nuevo

enlace. A modo de ejemplo, si suponemos que estamos en una red social, este modelo dice que la probabilidad de que me presenten (y me enlace) a un amigo de un amigo es mayor que la probabilidad de que me presenten a un desconocido.

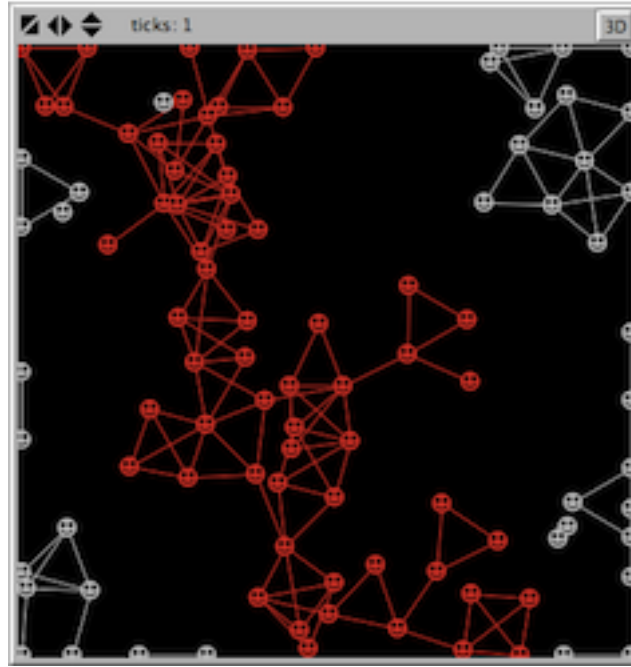


Figure 2: Topología de una red generada según el modelo de presentación

En este caso vemos que, en lugar de un único componente conexo, aparecen distintas componentes conexas, grupos de “amigos”.

#### 4.1.2 Modelo geográfico estático

Este modelo tiene en cuenta la distancia a la que se encuentran los nodos. Cada nodo se conectará directamente a los  $k$  nodos más cercanos.

En este caso, la topología de la red depende de la configuración espacial inicial de los nodos. Dependiendo de ésta, podemos conseguir redes muy distintas.

#### 4.1.3 Modelo de encuentros aleatorios

Es un modelo similar al anterior pero, en este caso, los nodos se mueven aleatoriamente y dos nodos se unen si chocan. Cada nodo tiene un número máximo de enlaces.

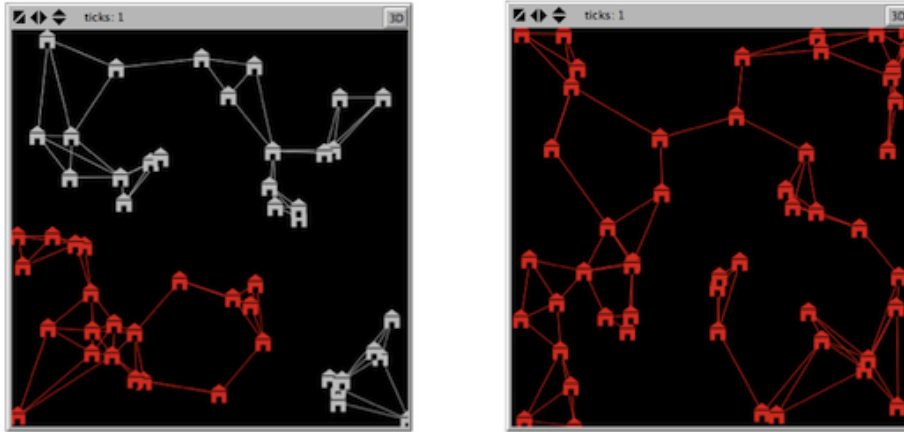


Figure 3: Dos redes distintas siguiendo el modelo estático geográfico

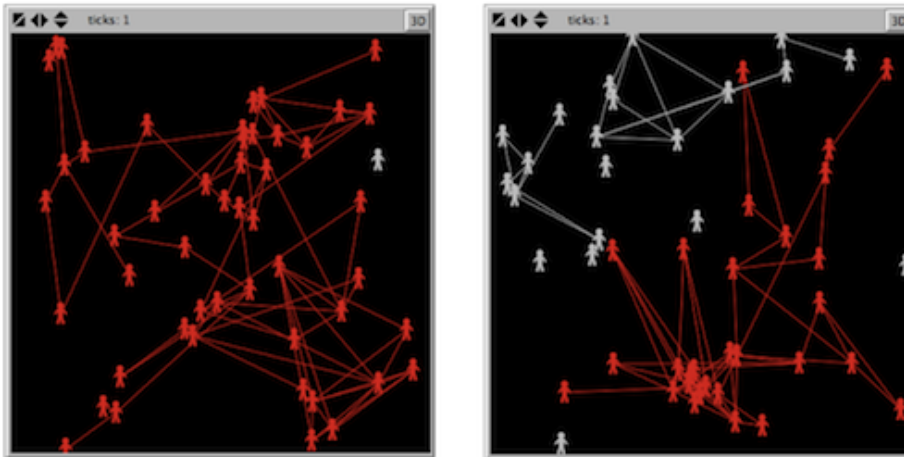


Figure 4: Dos redes distintas siguiendo el modelo de encuentros aleatorios

Al igual que antes, la topología de la red depende de la configuración espacial inicial de los nodos. Sin embargo, este modelo permite que nodos alejados entre sí puedan llegar a unirse si se mueven lo suficiente. Al igual que antes, la topología de la red es muy variable.

En resumen, vemos que la forma en la que se crea una red tiene cierta influencia en la topología final de la misma.

## 4.2 Ingredientes para la creación de una red libre de escala

En 1999, los investigadores Barabasi y Albert destacaron dos supuestos que no eran tenidos en cuenta en el modelo de red aleatoria pero que existían en las redes reales:

- **La red crece a lo largo del tiempo.**

Las redes reales se expanden mediante la adición de nuevos nodos y nuevos enlaces. Recordemos que el modelo de red aleatoria supone un número fijo de nodos  $N$  inicial y que este modelo crea enlaces nuevos pero no modifica el número de nodos existentes. En las redes reales, el número de nodos no es fijo sino que crece con el tiempo. Por ejemplo, en 2001, la WWW tenía un nodo (la página inicialmente creada por Tim Berners-Lee) mientras que en la actualidad está en torno a los 1000 millones.

- **Los nodos nuevos prefieren conectarse a los nodos más conectados.**

Cuando un nuevo nodo llega a la red en ella ya existen enlaces por lo que cada nodo puede decidir a quién conectarse en función del número de enlaces que otro nodo ya tiene. Sin embargo, el modelo de red aleatoria asume que la probabilidad de conectarse a otro nodo es la misma para todos los nodos (es completamente aleatoria). Esta preferencia existente en las redes reales de conectarse con los nodos con más conexiones se conoce como **conexión preferencial** o *preferential attachment*. Por ejemplo, en la WWW solemos tener más información de las páginas con más enlaces (Facebook, Google...) por lo si creamos nuestra propia web es muy probable que nos conectemos a una de ellas en lugar de a otra menos conocida.

La conexión preferencial es un concepto antiguo y conocido por muy distintos nombres:

- Un matemático húngaro lo referencia por primera vez en 1923 y lo llama el *proceso Polya*.
- En estadística es conocido como el *proceso Yule*.
- En 1923, Gibrart lo denomina *crecimiento proporcional*.

- En 1955, Simon lo usa para demostrar la naturaleza de la cola ancha de algunas distribuciones como el tamaño de las ciudades o el número de citas.
- En 1965, Price, basándose en el anterior trabajo, lo denomina *ventaja acumulativa* (lo veremos más adelante).
- En sociología se conoce como el *efecto Matthew*.
- El término de *conexión preferencial* es acuñado por Barabasi y Albert en 1999.

### 4.3 Modelo de Price

El primer modelo que tiene en cuenta estos dos ingredientes para modelar su evolución es el modelo de Price (1965). Este modelo fue usado por este investigador para explicar la cola ancha que presenta la distribución de citas de los artículos de investigación. Un artículo de investigación (también conocido como *paper*) es un documento publicado que siempre ha de tener un mínimo número de citas bibliográficas o referencias a otros artículos que ya han sido publicados anteriormente y que sirven como soporte a ciertas ideas que se presentan en dicho artículo.

Este modelo es el siguiente:

- Cada artículo nuevo tiene  $m$  citas. En términos de una red esto implica que cada vez que añadamos un nodo al modelo, este tendrá  $m$  enlaces a otros nodos ya existentes en la red.
- Un artículo nuevo cita a otro ya publicado con una probabilidad proporcional al número de citas que este último tiene. En términos de una red, la probabilidad de enlazarse a un nodo depende del grado de este nodo.
- El modelo necesita hacer una suposición y es que todos los artículos tienen al menos una cita. Por tanto, la probabilidad de enlazarse a un nodo no va a ser proporcional a  $k$  sino a  $k + 1$ .

El estudio y la simulación de este modelo demostró que las redes generadas siguiendo este modelo tienen una distribución de grados que sigue una ley potencial con exponente  $\gamma = 2 + \frac{1}{m}$ .

### 4.4 Modelo de Barabasi-Albert

Este modelo fue usado para modelar la distribución de los nodos de la WWW. En este caso, el modelo sigue las siguientes etapas:

- Partimos de una distribución inicial aleatoria de  $m_0$  nodos y todos los nodos tienen al menos un enlace.
- La red evoluciona realizando los dos siguientes pasos.

1. En cada momento de tiempo  $t$  se añade un nuevo nodo a la red con  $m \leq m_0$  enlaces que se conectarán a  $m$  nodos ya existentes en la red.
2. La probabilidad de que uno de los enlaces del nuevo nodo se conecte a un nodo ya existente  $i$  depende del grado de dicho nodo  $k_i$ . Esta probabilidad se puede representar como:

$$\pi(k) = \frac{k_i}{\sum_{j \neq i} k_j}$$

## 4.5 Resumen de los modelos de crecimiento

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**  
**Scale-free networks**

**DEGREE DISTRIBUTION**

Discrete form:

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

Continuous form:

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

**SIZE OF THE LARGEST HUB**

$$k_{\max} \sim k_{\min} N^{\frac{1}{\gamma-1}}.$$

**MOMENTS OF  $p_k$**

$2 < \gamma < 3$ :  $\langle k \rangle$  finite,  $\langle k^2 \rangle$  diverges when  $N \rightarrow \infty$ .

$\gamma > 3$ :  $\langle k \rangle$  and  $\langle k^2 \rangle$  finite.

**DISTANCES IN A SCALE-FREE NETWORK**

$$d \sim \begin{cases} \text{const.} & \text{if } \gamma=2, \\ \frac{\ln \ln N}{\ln(\gamma-1)} & \text{if } 2 < \gamma < 3, \\ \frac{\ln N}{\ln \ln N} & \text{if } \gamma=3, \\ \ln N & \text{if } \gamma > 3. \end{cases}$$

Figure 5: Cuadro resumen (extraído de *Network Science*, cap 4, pp. 37)