

Análisis de Redes Sociales

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

6 de noviembre de 2015

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 5: Modelos de redes: El modelo de red aleatoria

Los datos relativos a una red nos permiten calcular todas las propiedades vistas hasta ahora. Sin embargo, muchas de las redes reales son demasiado complejas como para tener todos los datos. En este caso es necesario aproximarlas mediante un modelo que las simplifique.

Un **modelo** es una representación simple de un sistema complejo del que podemos extraer y derivar propiedades matemáticamente. Por tanto, mediante un modelo que se aproxime a una red real podremos extraer y derivar propiedades de la misma. Además, los modelos nos ayudan a entender cómo y por qué se han formado las redes tal y como son ahora, prever cómo evolucionarán en un futuro y extraer resultados de las mismas.

En este y en sucesivos capítulos estudiaremos algunos de los modelos más conocidos de redes sociales. El primero de ellos es el **modelo de red aleatoria**.

5.1 Modelo de Red Aleatoria: algoritmo de construcción

El primero de los modelos que vamos a estudiar es el modelo de red aleatoria. Este modelo se basa en la siguiente suposición:

Los enlaces entre los nodos de una red se generan de manera aleatoria

Por ejemplo, podemos suponer que nuestras relaciones de amistad (enlaces) con otras personas (nodos) se ha generado de una manera completamente aleatoria.

Una red aleatoria es una red en la que cada uno de los enlaces entre dos nodos se ha creado siguiendo un proceso completamente aleatorio. De manera más matemática podemos decir que una red aleatoria es una red que tiene N nodos donde cada nodo puede estar conectado con otro con una probabilidad p . Se representa como $G(N, p)$ y se construye siguiendo el siguiente algoritmo:

1. Crear N nodos aislados

2. Seleccionar un par de nodos y generar un número aleatorio entre 0 y 1. Si es menor o igual que p entonces añadimos un enlace entre ellos. En otro caso, los dejamos desconectados.
3. Repetir el paso 2 para los $\frac{N(N-1)}{2}$ pares de nodos de la red.

El modelo de red aleatoria también se conoce como el modelo de Erdős-Renyi en honor de los dos matemáticos húngaros que las estudiaron inicialmente y que proporcionaron gran cantidad de información sobre las propiedades contenidas en este tipo de redes.

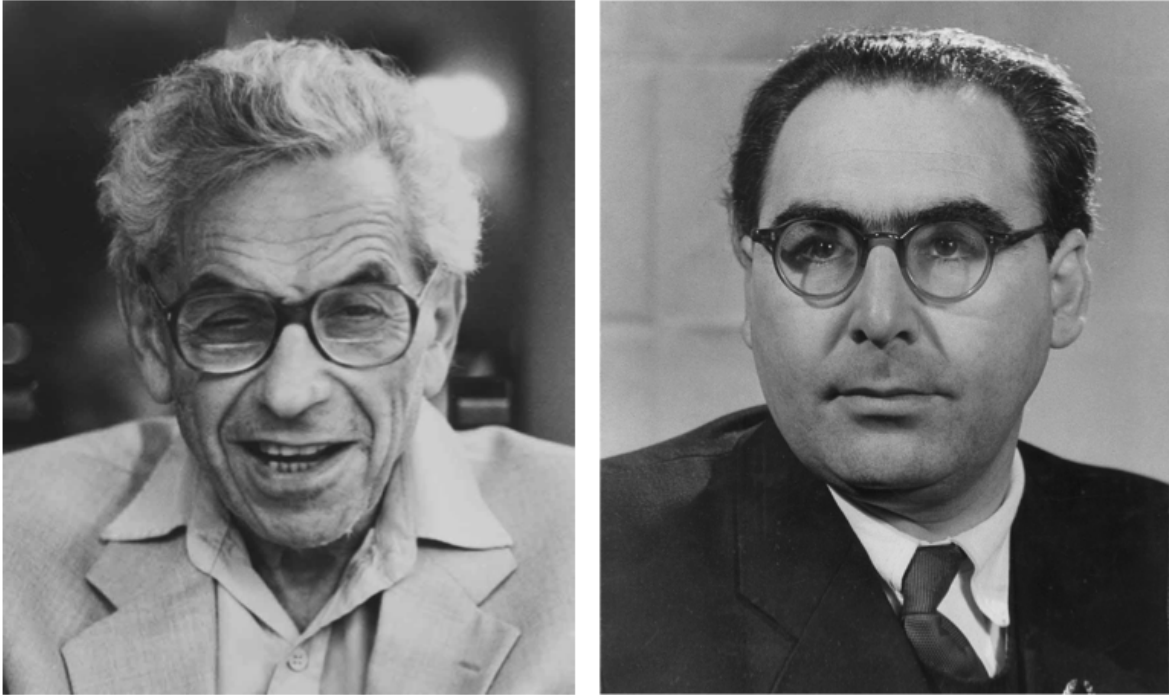


Figure 5.1: Pal Erdős y Alfred Renyi, matemáticos húngaros principales investigadores del modelo de red aleatoria.

El modelo de red aleatoria también puede caracterizarse como $G(N, L)$, donde L es el número de enlaces de la red. En este caso, la red se forma seleccionando aleatoriamente 2 nodos y, si no existe un enlace entre ellos, se añade. Este proceso se repite hasta conseguir un total de L enlaces. Esta forma de caracterizar la red se usa menos.

5.2 Número de enlaces

Las redes aleatorias $G(N, p)$ tienen un número variable de enlaces (L). Sin embargo, usando probabilidades básicas podemos calcular la probabilidad de que la red tenga exactamente L enlaces.

Para aproximar este cálculo se utiliza la *distribución binomial* que describe el número de éxitos (x) que se pueden conseguir en la realización de n experimentos independientes donde la probabilidad de acierto es p y la de fracaso, $1 - p$.

$$B(n, x, p) \rightarrow p_x = \binom{n}{x} p^x (1-p)^{n-x}$$

De esta función de distribución conocemos las siguientes métricas:

- Media de la distribución: $\langle x \rangle = \sum_{x=1}^N x \cdot p_x = n \cdot p$
- Varianza de la distribución: $\sigma_x^2 = p \cdot (1-p) \cdot n$
- Desviación estándar de la distribución: $\sigma_x = [p \cdot (1-p) \cdot n]^{\frac{1}{2}}$

De acuerdo a esto podemos representar la probabilidad de una red $G(N, p)$ de tener exactamente L enlaces como:

$$p_L = B(L_{max}, L, p) \rightarrow p_L = \binom{L_{max}}{L} p^L (1-p)^{L_{max}-L}$$

Recordemos que el número máximo de enlaces L_{max} es el número posible de pares distintos de nodos que podemos formar. Esto se calcula como:

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

De acuerdo a esto podemos calcular el número medio de enlaces esperados en $G(N, p)$ como:

$$\langle L \rangle = p \cdot L_{max} = p \cdot \binom{N}{2} = p \cdot \frac{N(N-1)}{2}$$

Por tanto, el grado medio de esta red¹ es:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p \cdot (N-1)$$

De acuerdo a esto, una de las primeras características que podemos extraer de las redes aleatorias es que cuanto mayor sea p mayor es el grado medio y, por tanto, más densa se vuelve la red.

¹Si la red se representa como $G(L, p)$ entonces ya sabemos el número exacto de enlaces por lo que $\langle k \rangle = \frac{2L}{N}$.

5.3 Distribución del grado de los nodos

Al igual que antes, podemos representar la distribución del grado de los nodos como una binomial.

$$p_k = B(N-1, k, p) \rightarrow p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

- El grado medio será: $\langle k \rangle = (N-1) \cdot p$
- La varianza de k será: $\sigma_k^2 = p \cdot (1-p) \cdot (N-1)$
- La desviación estándar de k será: $\sigma_k = [p \cdot (1-p) \cdot (N-1)]^{\frac{1}{2}}$

Sin embargo, sabemos que la mayoría de las redes reales son *dispersas*, lo que implica que $\langle k \rangle \ll N$. En este caso, la distribución de grados de los nodos se aproxima mejor usando una *distribución de Poisson*:

$$P(x, \lambda) \rightarrow p_x = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

En esta distribución sabemos que:

- Media de la distribución: $\langle x \rangle = \lambda$
- Varianza de la distribución: $\sigma_x^2 = \lambda$
- Desviación estándar: $\sigma_x = \lambda^{\frac{1}{2}}$

Podemos ver visualmente las diferencias entre la distribución binomial y la Poisson:

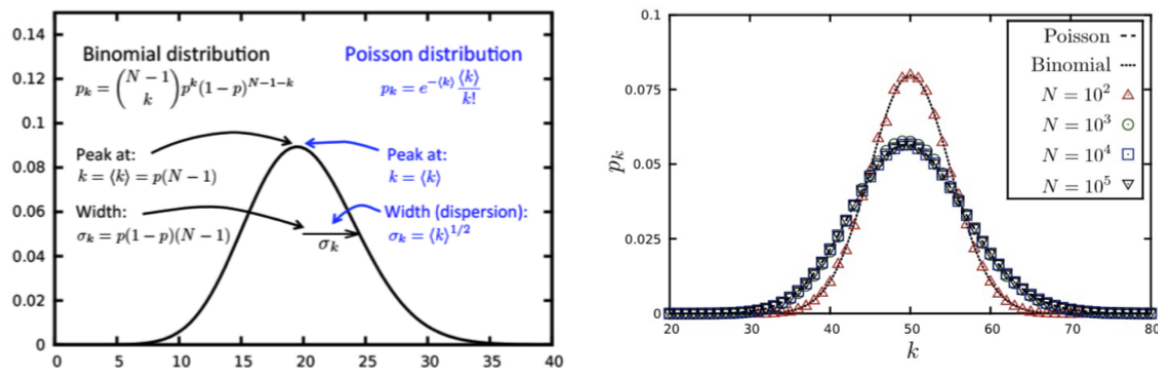


Figure 5.2: Representación gráfica de la distribución binomial y de la distribución de Poisson

Vemos que ambas distribuciones tienen propiedades comunes:

- Tienen un pico en $\langle x \rangle$ de modo que si modificamos p entonces el pico se desplaza hacia la derecha.
- Cuanto más densa es la distribución entonces más ancha es la distribución.

Las principales ventajas de la Poisson frente a la binomial son que sus principales propiedades (media, varianza...) tienen una forma más simple y que la Poisson no depende de N .

Si estamos seguros de que tenemos una red en la que $N \gg k$ y aproximamos su distribución de grados mediante una Poisson entonces se cumple que:

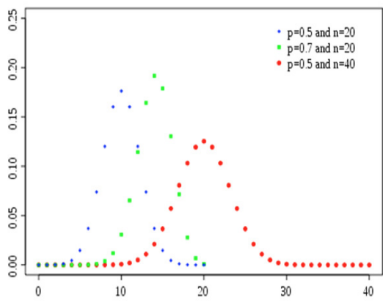
$$p_k = P(k, \langle k \rangle) \rightarrow p_k = e^{-\langle k \rangle} \cdot \frac{\langle k \rangle^k}{k!}$$

De modo que:

- Varianza de k : $\sigma_x^2 = \langle k \rangle$
- Desviación estándar de k : $\sigma_x = \langle k \rangle^{\frac{1}{2}}$

La siguiente figura resume las diferencias entre una y otra función de distribución:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

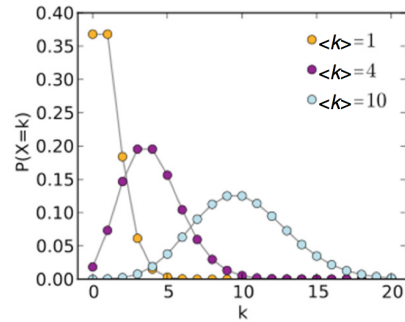


$$\langle k \rangle = (N-1)p$$

$$\langle k^2 \rangle = p(1-p)(N-1) + p^2(N-1)^2$$

$$\sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = [p(1-p)(N-1)]^{1/2}$$

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



$$\langle k \rangle = \langle k \rangle$$

$$\langle k^2 \rangle = \langle k \rangle (1 + \langle k \rangle)$$

$$\sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = \langle k \rangle^{1/2}$$

Figure 5.3: Comparativa de distribuciones de grados (izq: Binomial; der: Poisson)

De acuerdo a la distribución de Poisson podemos sacar las siguientes propiedades de las redes aleatorias dispersas:

- La distribución de grados no depende de N por lo que dos redes con igual $\langle k \rangle$ y distinto tamaño N tienen funciones de distribución de grado indistinguibles.

- La mayoría de los nodos tienen un grado entorno a la media ($\langle k \rangle$) y los nodos de mayor grado tienen solo unos pocos más que la media.
- Las redes aleatorias no tienen **concentradores o hubs**: nodos con conectividad o grado muy alto ya que la probabilidad de tener nodos con grado muy alto es extremadamente baja.

Por ejemplo, en una red con $\langle k \rangle = 1000$ se puede calcular que $\sigma_x = 31,62$ por lo que los nodos tienen entre 970 y 1030 enlaces. También podemos realizar aproximaciones² para calcular el grado máximo y mínimo, que quedaría en $k_{max} = 1185$ y $k_{min} = 816$.

5.4 Evolución de una red aleatoria

²Los cálculos para aproximar el máximo y el mínimo se pueden consultar en el libro “Network Science”, cap 3, pp 72.