

Análisis de Redes Sociales

Guillermo Jiménez Díaz (gjimenez@ucm.es)

Alberto Díaz (albertodiaz@fdi.ucm.es)

27 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 Cordón García de la Universidad de Granada.

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

Tema 5: Centralidad

5.1 Concepto de centralidad

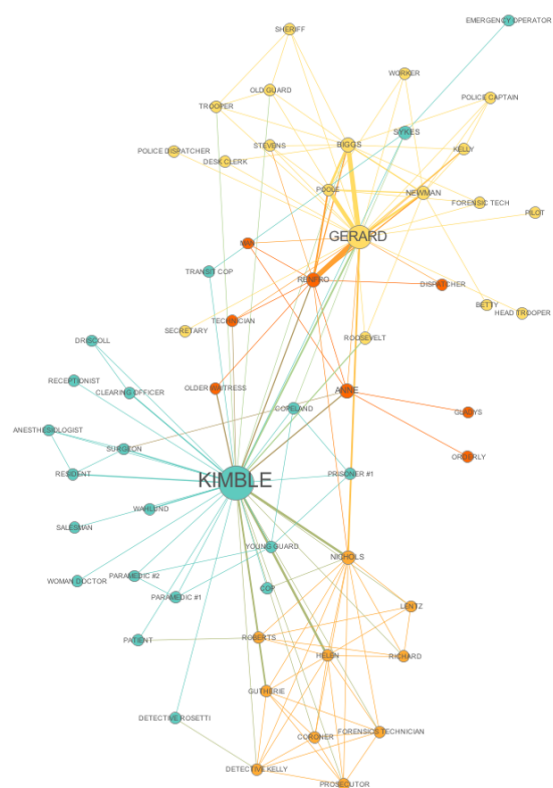
Los nodos con una “posición más central” (una mayor centralidad) tienen un acceso más fácil y rápido a los demás nodos de la red y una mayor capacidad para ejercer un control del flujo entre ellos.

5.1.1 Ejemplos de grafos de actores en películas

En los siguientes grafos se representan relaciones entre los actores que aparecen en distintas películas analizando los guiones. Se incluye una arista entre dos actores si un actor interacciona con otro actor en un punto concreto de la película.

Han sido extraídas de la web moviegalaxies.com

The Fugitive (1993)



moviegalaxies.com

Figure 1: The Fugitive (1993): Actores centrales: Dr. Kimble, Gerard

Memento (2000)

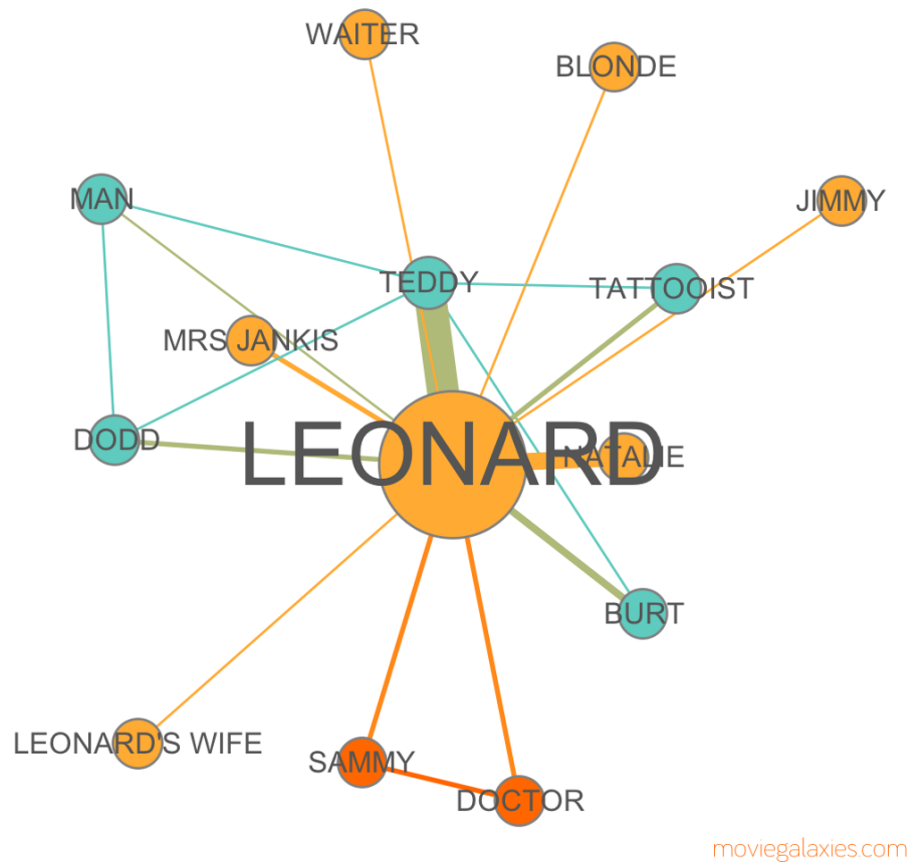


Figure 2: Memento: Actor central: Leonard

5.2 Medidas

Para analizar la estructura de las redes sociales existen 2 tipos de medidas:

5.2.1 Medidas locales (a nivel de nodos).

- Todas estas medidas están basadas en el concepto general de centralidad (redes no dirigidas) o prestigio (redes dirigidas).
- La centralidad es una medida general de la posición de un nodo en la estructura global de la red social.
- Estas medidas se usan para identificar los nodos claves de una red.
- Muestran como las relaciones se concentran en unos pocos nodos (individuos), dando una idea de su poder social.

5.2.2 Medidas globales (a nivel de red)

- Proporcionan información más compacta que permite evaluar la estructura global de la red.
- Proporcionan información sobre propiedades importantes de los fenómenos sociales subyacentes.
- Algunas de estas medidas las hemos utilizado en temas anteriores: diámetro, distancia media, grado medio, densidad, coeficiente de clustering.

5.3 Medidas locales de centralidad

Existen varias medidas distintas de centralidad:

- Grado
- Intermediación (betweenness)
- Cercanía (closeness)
- Excentricidad
- Centralidad de vector propio
- Coeficiente de clustering

5.4 Centralidad de grado

Mide el número de enlaces con otros nodos. Hay que distinguir entre grafos dirigidos y no dirigidos.

5.4.1 Grafos no dirigidos.

Centralidad de grado de un nodo: número de enlaces que lo conectan con otros.

$$C_D(N) = k_N$$

$C_D(N)$ se define en el intervalo $\{0, g-1\}$, siendo g el número de nodos de la componente conexa.

Interpretación: *Los nodos con más enlaces son más centrales.* En una red social de personas con relaciones de amistad, los individuos con más amigos son más centrales. *Sólo mide la importancia con respecto a los vecinos más cercanos.* Se asume que las conexiones de los vecinos no importan, sólo importa lo que se pueda hacer directamente con los vecinos.

5.4.2 Grafos dirigidos.

En grafos dirigidos, se define el prestigio de entrada (in-degree), denominado Soporte, y el prestigio de salida (out-degree), denominado Influencia:

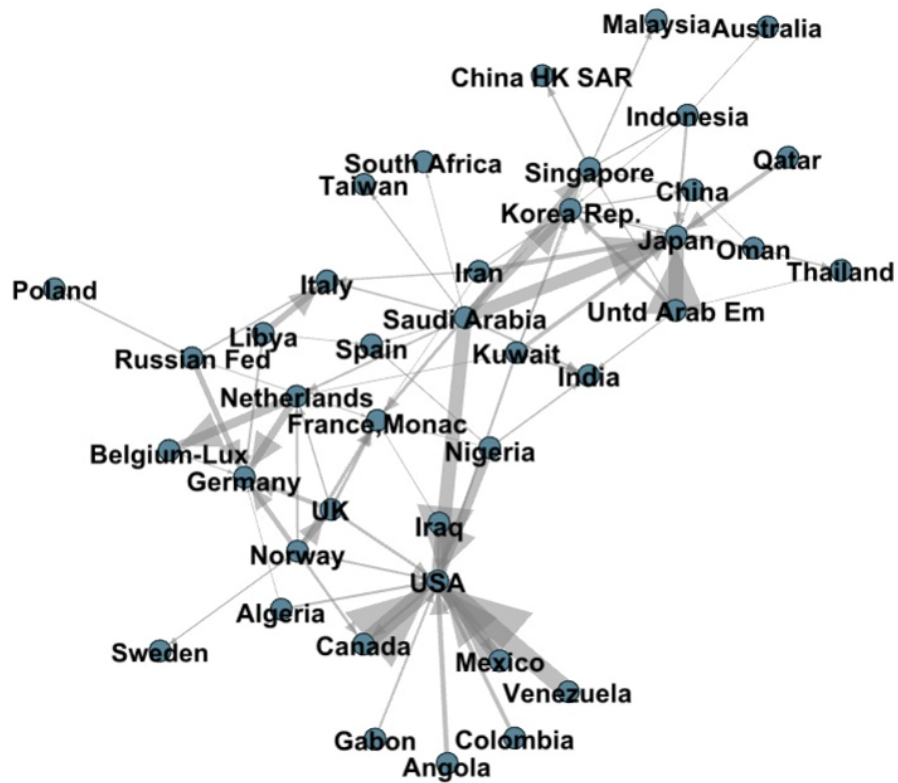
$$\text{Soporte} = P_D^{in} = k_N^{in}$$

$$\text{Influencia} = (P_D^{out} = k_N^{out})$$

Ambos se definen en el intervalo $\{0, g-1\}$

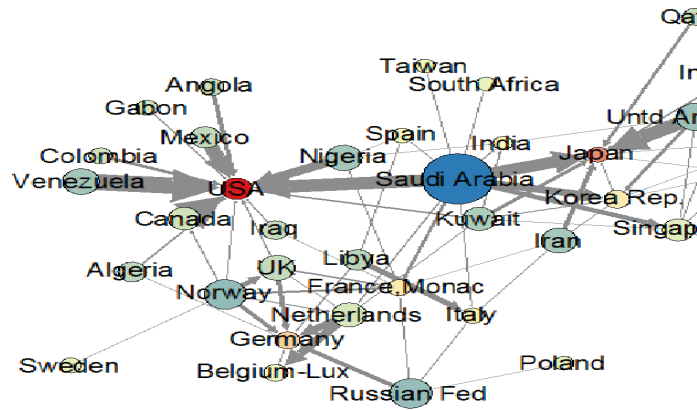
Interpretación Soporte: *Los nodos con muchos enlaces de entrada son prominentes.* La idea básica es que muchos nodos procuran tener enlaces directos a ellos, por lo que se puede considerar como una medida de su importancia.

Interpretación Influencia: *Los nodos que tienen muchas conexiones directas de salida con otros son influyentes.* Pueden transferir información rápidamente a muchos otros nodos.



Comercio de petróleo y derivados, 1998. Fuente: NBER-United Nations Trade Data. Cada nodo es un país y hay un enlace entre el exportador y el importador.

¿Qué países importan de muchos otros? ¿Arabía Saudí, Japón, Iraq, USA, Venezuela?



Comercio de petróleo y derivados, 1998. Fuente: NBER-United Nations Trade Data. Cada nodo es un país y hay un enlace entre el exportador y el importador. El grosor de la arista indica el volumen exportado.

¿Qué país exporta a pocos países pero lo hace en gran cantidad? ¿Arabía Saudí, Japón, Iraq, USA, Venezuela?

5.4.3 Normalización del grado

Es habitual normalizar los valores de las centralidades de grado ($C_D(N)$, P_D^{in} , P_D^{out}) dividiéndolas por su valor máximo $g-1$ (siendo g el número de nodos de la componente conexa donde está el nodo).

5.4.4 Problemas del grado como medida de centralidad

El grado es una medida adecuada para evaluar la centralidad de un nodo en una red social pero desde una perspectiva muy local: * mide la importancia y la influencia del nodo con respecto a sus vecinos más cercanos, * pero tiene una limitación importante: no tiene en cuenta la estructura global de la red.

5.5 Intermediación

La intermediación es una medida pensada para capturar como de central es un nodo desde el punto de vista de cuantos caminos que conectan nodos pasan por él.

$$C_B(i) = \sum_{j < k} \frac{g_{jk}(i)}{g_{jk}}$$

donde g_{jk} es el número de caminos mínimos que conectan los nodos j y k , y $g_{jk}(i)$ es el número de esos caminos que incluyen al nodo i .

$C_B(i)$ se define en el intervalo $\{0, (g-1)(g-2)\}$ en redes dirigidas y en $\{0, (g-1)(g-2)/2\}$ en no dirigidas.

Interpretación: * Un nodo tiene una posición más favorable (mayor intermediación) en la medida en que dicho nodo esté situado en los caminos geodésicos (caminos más cortos) de todos los demás. * En otras palabras, cuanto más caminos geodésicos pasen por un nodo más central será.

Es habitual considerar la medida normalizada:

$$C'_B(i) = \frac{C_B(i)}{(g-1)(g-2)/2}$$

donde el denominador representa el número de pares de nodos excluyendo el propio nodo i .

Los nodos con una intermediación alta ocupan roles críticos en la estructura de una red, puesto que suelen ocupar una posición que les permite trabajar como interfaces entre subgrupos de nodos fuertemente unidos.

Son elementos vitales en la conexión entre distintas regiones de una red.

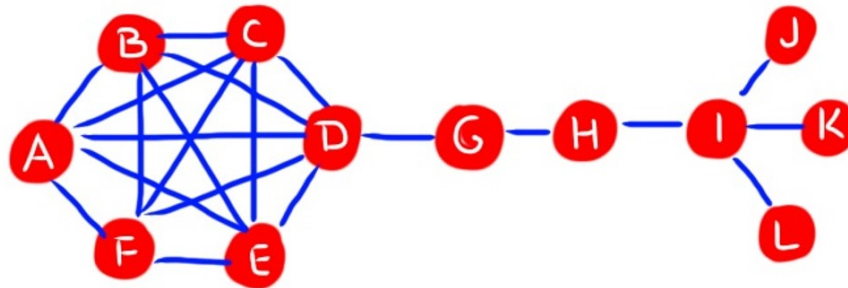


Figure 3: Ejercicio

- Encontrar un nodo con un valor alto de intermediación pero con un grado bajo
- Encontrar un nodo con un valor bajo de intermediación pero con un grado alto

5.6 Cercanía

Esta medida da importancia a la facilidad de acceso al resto de la red, a que se pueda llegar al resto de los nodos de la red con pocos saltos, o dicho de otra

forma que la distancia al resto de los nodos de la red sea pequeña.

En esta medida no se tiene en cuenta ni tener muchos vecinos directos ni estar situado entre otros nodos. Se le da importancia a estar “en medio” de la red.

La suma de las distancias geodésicas (distancias de los caminos mínimos) de un nodo de la red a todos los demás es la lejanía de un nodo al resto. La inversa de dicha suma es la medida de cercanía.

Centralidad de cercanía: $C_C(i) = 1 / \sum_j d(i, j)$

Centralidad de cercanía normalizada: $C'_C(i) = C_C(i) / g - 1$

Si hay varias componentes conexas puede haber distancias infinitas, se puede calcular la centralidad de cercanía como la suma de las inversas de las distancias.

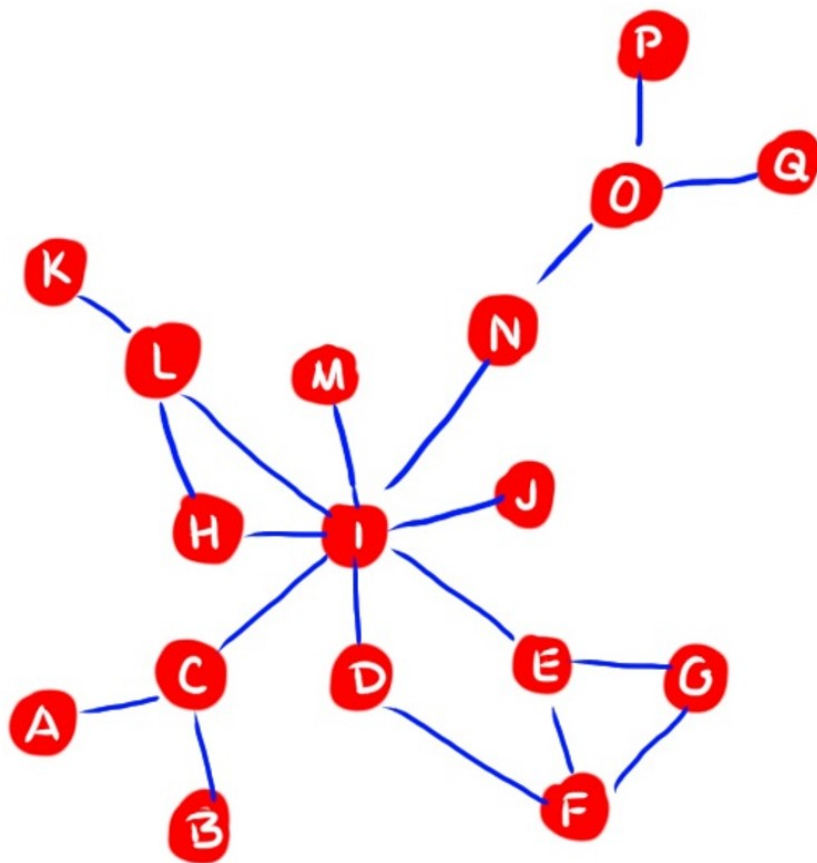


Figure 4: Ejercicio

- Encontrar un nodo con grado relativamente alto pero con un valor de

cercanía bajo

5.6.1 Cercanía en redes dirigidas

Se pueden definir dos medidas de cercanía distintas considerando sólo los enlaces de entrada o de salida:

- Cercanía de entrada (p.ej.: prestigio en redes de citación)
- Cercanía de salida (alcance de la influencia de un nodo)