

Análisis de Redes Sociales

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

5 de diciembre 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.

Para este capítulo se ha utilizado, adicionalmente, material de los libros *Analyzing the Social Web* de Jennifer Goldbeck (capítulo 10) y *Networks, Crowds, and Markets* de Easley y Kleinberg (capítulos 19 y 21).

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

Tema 8: Propagación y Difusión en redes

Las conexiones presentes en la redes permiten modelar la propagación de todo tipo de elementos entre sus nodos: enfermedades, vídeos virales, rumores, virus informáticos, productos, anuncios, información... En general, la mayoría de estos modelos de propagación son similares independientemente de lo que se pretenda propagar.

Existe desde hace muchos años un estudio intenso en la propagación de enfermedades. El conocimiento de cómo se propagan a través de una red de individuos nos puede servir para entender cómo se puede propagar cualquier otro tipo de información en dicha red. Por este motivo, en este tema vamos a hablar los

modelos fundamentales de propagación de enfermedades y vamos a estudiar como aplicarlos a las redes, analizando cómo la estructura de la misma afecta enormemente a estos fenómenos de propagación.

8.1 Modelos de contagio simple

La epidemiología es la ciencia que estudia la salud y control de enfermedades en una población, así como la predicción de expansión de dichas enfermedades. El modelo general epidemiológico se basa en dos hipótesis:

- **Modelo compartimental:** cada individuo puede estar en un determinado estado dependiendo de en qué fase se la enfermedad se encuentra. El modelo más simple, que será el que usemos, supone 3 estados:
 - **Susceptible (S):** El individuo está sano y puede ser infectado.
 - **Infectado (I):** El individuo está infectado y puede contagiar a otros individuos.
 - **Recuperado (R):** El individuo estuvo contagiado pero se ha recuperado y no puede volver a ser contagiado. También se utiliza para modelar los individuos que no han superado la enfermedad y que han muerto a causa de ella.
- **Mezcla homogénea:** Cualquier individuo tiene la misma probabilidad de estar en contacto con un individuo infectado. Esta hipótesis puede interpretarse como que la red de contactos está modelada mediante una red aleatoria aunque, realmente, elimina la necesidad de conocer los contactos de los individuos y se puede asumir que cualquiera puede infectar a cualquiera.

Los modelos generados a partir de estas hipótesis observan el comportamiento (los cambios de estado) de los individuos a lo largo del tiempo para predecir el alcance y la velocidad de propagación de la enfermedad, entre otras. A continuación vamos a estudiar la dinámica de los modelos de propagación clásicos, que combinan las letras del modelo compartimental.

8.1.1 Modelo SI

Es el modelo más sencillo, en el que un individuo susceptible puede quedar infectado pero, una vez infectado, no se puede recuperar. Un ejemplo de este tipo es el virus del VIH (o los zombies).

En este modelo suponemos que cada individuo tiene $\langle k \rangle$ contactos (enlaces) y que en cada instante de tiempo la enfermedad se propaga con una tasa de contagio δ , que representa la probabilidad de que un individuo infectado transmita la enfermedad a uno susceptible.

Para entender la dinámica del modelo vamos a definir los siguientes parámetros.

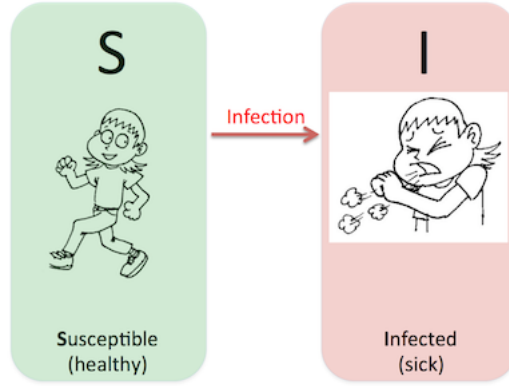


Figure 1: Esquema del Modelo SI

- N es el tamaño de la población y es $N = S(t) + I(t)$, donde $S(t)$ (número de individuos que están en el estado susceptible en el tiempo t) e $I(t)$ (número de individuos que están en el estado infectado en el tiempo t).
- En lugar de manejar valores absolutos vamos a manejar las proporciones o ratios de susceptibles e infectados. De este modo $s(t) = s = \frac{S(t)}{N}$ representa la proporción de individuos susceptibles de la población mientras que $i(t) = i = \frac{I(t)}{N}$.
- β es la tasa de transmisión que incluye el grado medio de cada individuo, esto es, $\beta = \delta \cdot \langle k \rangle$.

La ecuación diferencial que modela la tasa a la que varía el número de infectados es:

$$\frac{di}{dt} = \beta \cdot i \cdot s = \beta \cdot i \cdot (1 - i)$$

Si resolvemos esta ecuación nos queda que:

$$i = \frac{i_0 \exp(\beta t)}{1 - i_0 + i_0 \exp(\beta t)}$$

donde i_0 representa a la tasa de infectados en el instante $t = 0$. De la representación gráfica de esta función extraemos las siguientes conclusiones:

- Inicialmente el número de infectados crece exponencialmente.
- A medida que el número de infectados se hace mayor, hay menos individuos susceptibles por lo que el crecimiento de infectados se ralentiza y la infección termina cuando todos están infectados ($i(t \rightarrow \infty) = 1$).

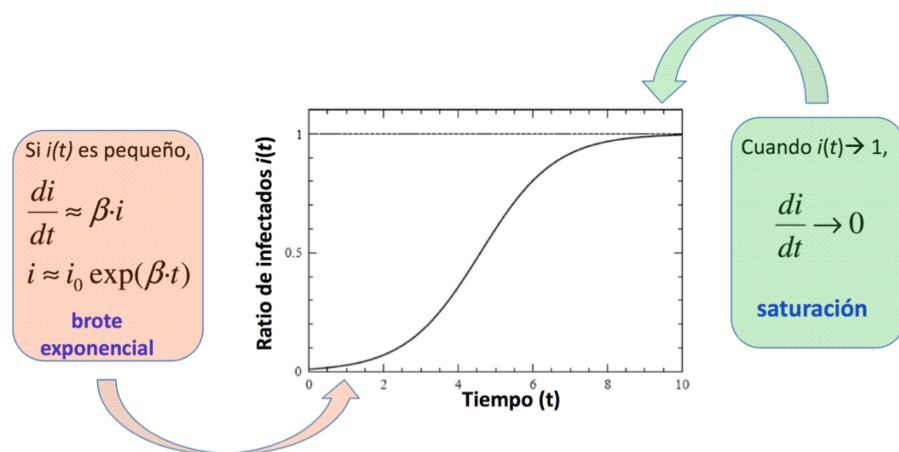


Figure 2: Representación de la proporción de infectados en el Modelo SI

8.1.2 Modelo SIS

Es similar al anterior salvo en que, en este modelo, los individuos infectados se pueden recuperar, volviendo al estado susceptible. Un ejemplo de este modelo es el resfriado común.

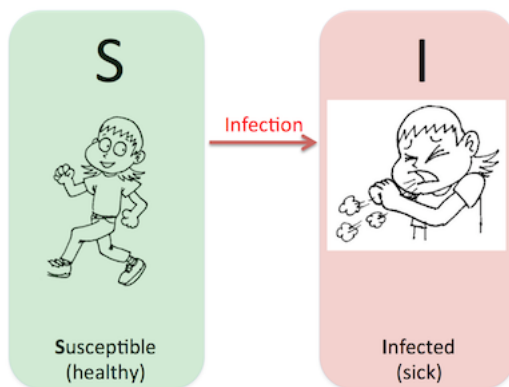


Figure 3: Esquema del Modelo SI

Para este modelo necesitamos, además de los parámetros del anterior, la tasa de recuperación μ , que representa la proporción de infectados que se recuperan y pasan al estado susceptible en cada instante de tiempo.

En este caso, la ecuación diferencial que modela la tasa a la que varía el número de infectados es:

$$\frac{di}{dt} = \beta \cdot i \cdot s - \mu \cdot i = \beta \cdot i \cdot (1 - i) - \mu \cdot i$$

La resolución de esta ecuación nos da el siguiente resultado:

$$i = \left(1 - \frac{\mu}{\beta}\right) \frac{C \cdot e^{(\beta-\mu)t}}{1 + C \cdot e^{(\beta-\mu)t}}$$

$$C = \frac{\beta \cdot i_0}{\beta - \mu - \beta \cdot i_0}$$

En este caso, las conclusiones que podemos extraer de la representación gráfica de la función son las siguientes:

- Como la recuperación es posible, el sistema alcanza un *estado endémico* en el que la tasa de infectados es constante:

$$i(\infty) = 1 - \frac{\beta}{\mu}$$

En este estado endémico, la proporción de infectados no varía con el tiempo y sólo se produce cuando la tasa de recuperación es inferior a la tasa de transmisión ($\mu < \beta$)

- En caso de que la tasa de recuperación sea mayor que la tasa de transmisión ($\mu > \beta$) entonces llegado a un determinado punto la proporción de infectados comienza a decrecer exponencialmente, alcanzado un estado libre de enfermedad, en la que todos los individuos se han recuperado y no hay infectados.

El ritmo reproductivo básico (R_0) representa el número promedio de individuos susceptibles que serán infectados por un individuo infectado:

$$R_0 = \frac{\beta}{\mu}$$

Tal y como hemos visto antes, si $R_0 < 1$ entonces la enfermedad termina desapareciendo de la población. Si $R_0 > 1$ entonces la enfermedad se propagará. Cuanto mayor sea R_0 , más rápido es el proceso de propagación de la enfermedad. Por ejemplo, el sarampión (que se contagia por el aire) tiene un $R_0 = 12 - 18$ mientras que la gripe tiene un $R_0 = 2 - 3$.

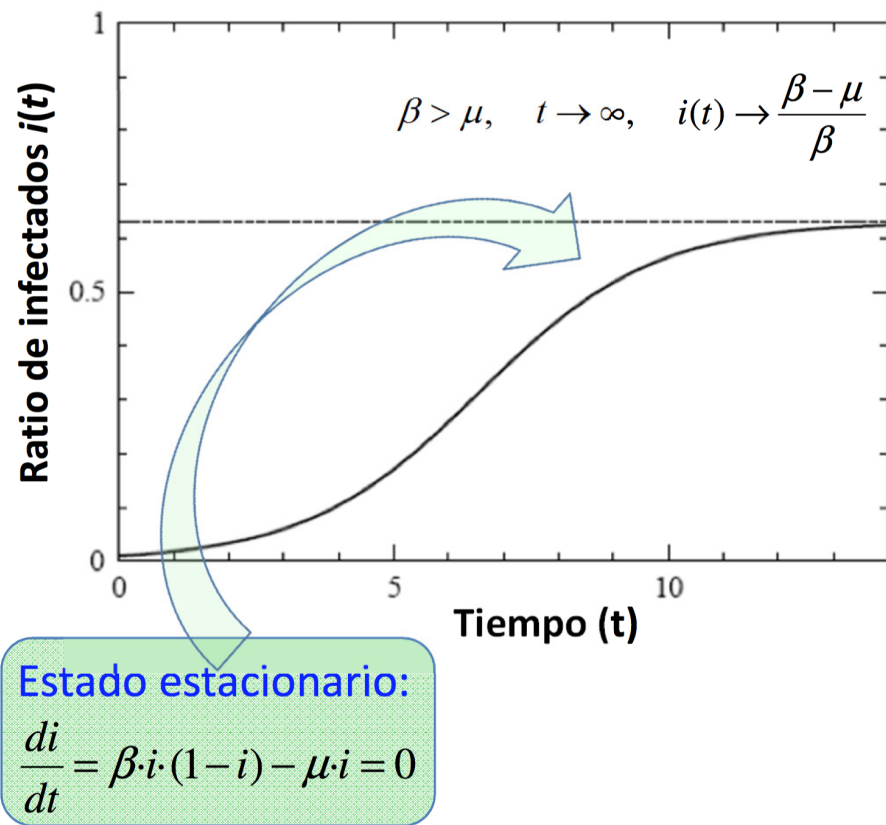


Figure 4: Representación de la proporción de infectados en el Modelo SIS

8.1.3 Modelo SIR

En este modelo, los individuos infectados no vuelven a ser susceptibles sino que desarrollan una inmunidad a la enfermedad (o mueren) y pasan a un estado recuperado¹ en el que no afectan al modelo de propagación: no pueden ser infectados ni pueden infectar a otros.

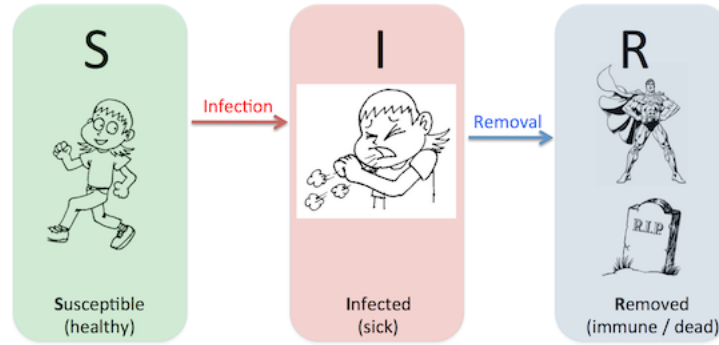


Figure 5: Esquema del Modelo SIR

En este modelo μ representa la tasa de recuperación que, a diferencia del anterior, es la tasa de individuos infectados que pasan al estado recuperado. Para este modelo, la población es la suma de los infectados, susceptibles y recuperados ($R(t)$), por lo que la proporción de infectados es $i = 1 - s - r$.

Las ecuaciones diferenciales de este modelo son las siguientes:

$$\frac{di}{dt} = \beta \cdot i \cdot s - \mu \cdot i; \quad \frac{ds}{dt} = -\beta \cdot i \cdot s; \quad \frac{dr}{dt} = \mu \cdot i$$

En este caso el cálculo es más complejo pero podemos llegar a la siguiente representación gráfica de las tres funciones:

- Cuando $\beta > \mu$ la proporción de infectados crece hasta un pico máximo y luego decrece hasta valer 0.
- La proporción de susceptibles decrece de forma monótona. Aunque satura, no llega nunca a 0 ya que cuando $i \rightarrow 0$ ya no hay individuos que puedan infectar. Esto implica que los individuos que se mantienen susceptibles hasta fases avanzadas pueden no llegar a infectarse nunca.
- La proporción de recuperados crece de manera monótona. De manera similar a los susceptibles, la proporción de recuperados nunca llega a valer 1. Su valor asintótico representa el número de individuos afectados y se calcula como:

¹En inglés, el estado es *removed*, que es más adecuado para describir el proceso.

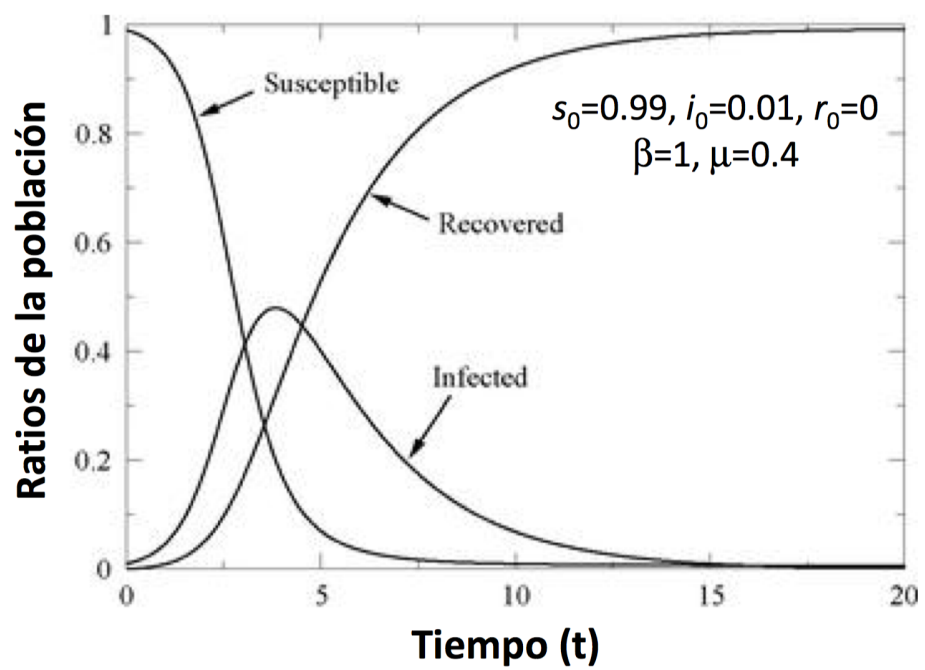


Figure 6: Representación de la proporción de infectados, susceptibles y recuperados en el Modelo SIR

$$r = 1 - s_0 \cdot e^{-\beta \frac{r}{\mu}}$$

Las condiciones iniciales más habituales son:

$$i_0 = \frac{c}{N}; s_0 = 1 - \frac{c}{N}; r_0 = 0$$

8.1.4 Comportamientos importantes de los modelos epidemiológicos

Existen principalmente dos comportamientos destacables en estos modelos:

Comportamiento temprano. Es el patrón de comportamiento en las fases iniciales. Es importante para saber cuánto tiempo tenemos para el desarrollo de vacunas e intervenciones médicas. La mejor forma de detener o contener la epidemia en esta fase es mediante vacunación temprana o la cuarentena.

En todos los modelos el número de infectados en la fase temprana es bajo pero crece exponencialmente. Generalmente, el modelo SI es el más relevante para describir este comportamiento.

Comportamiento tardío. Es el patrón de comportamiento en las fases más avanzadas de la epidemia (cuando $t \rightarrow \infty$). Permite predecir el alcance, número de infectados, etc.

En este caso, cada modelo realiza una predicción distinta:

- En el modelo SI todos terminan infectados.
- En el modelo SIS se alcanza un estado endémico en el que una proporción de la población queda infectada ($R_0 > 1$) o en el que la enfermedad desaparece ($R_0 < 1$)
- En el modelo SIR todos terminan recuperados (en el estado susceptible o recuperado, pero no infectados)

En resumen, las características básicas de los modelos epidemiológicos son los siguientes:

Tal y como hemos indicado, estos modelos no tienen en cuenta la red de contactos ya que suponen que hay una mezcla homogénea. Tal y como veremos a continuación, para predecir con mayor precisión las dinámicas de la propagación tendremos que tener en cuenta la estructura de la red ya que ésta modificará el comportamiento de estos modelos simples.

	SI	SIS	SIR
Comportamiento temprano:	$i(t) = \frac{i_0 \exp(\beta \cdot t)}{1 - i_0 + i_0 \exp(\beta \cdot t)}$	$i(t) = \left(1 - \frac{\mu}{\beta}\right) \frac{C \cdot e^{(\beta - \mu) \cdot t}}{1 + C \cdot e^{(\beta - \mu) \cdot t}}$	Crecimiento exponencial tipo SIS
	Crecimiento exponencial de los individuos infectados		
Comportamiento tardío:	$i(t) \rightarrow 1$	$i(t) \rightarrow 1 - \frac{\mu}{\beta}$	$i(t) \rightarrow 0$
	Saturación cuando $t \rightarrow \infty$		

Figure 7: Características básicas de los modelos epidemiológicos

8.2 Modelos de contagio basados en redes

Antes de entrar en los detalles más técnicos vamos a utilizar los modelos que conocemos hasta ahora y las simulaciones en NetLogo para observar el comportamiento de las epidemias teniendo en cuenta la estructura de la red.

8.2.1 Redes aleatorias

Vamos a simular un modelo SI en una red aleatoria. Si utilizamos el simulador de [Difusión en una red aleatoria](#)² podemos ver la influencia de la densidad de la red en los procesos de contagio.

Como se puede ver en la simulación, la densidad de la red afecta a la velocidad de infección y al número de individuos infectados: a mayor densidad, mayor es el número de individuos infectados y mayor es la velocidad de propagación.

8.2.2 Redes libres de escala

A continuación simularemos un modelo SI en una red libre de escala creada siguiendo el modelo de Barabasi-Albert. Si recordamos, para que se presente la propiedad de ser libre de escala es necesario que exista enlace preferencial. Por este motivo vamos a ver el efecto de la existencia de enlace preferencial en estas redes. Para ello usaremos el simulador de [Difusión en una red libre de escala](#).

En este caso podemos observar que el enlace preferencial favorece el contagio. Esto se debe a que el enlace preferencial posibilita la existencia de Hubs, que son los responsables de ayudar a difundir más rápidamente la infección.

²Todos los modelos de NetLogo que se ven en este tema están disponibles en el Campus Virtual.

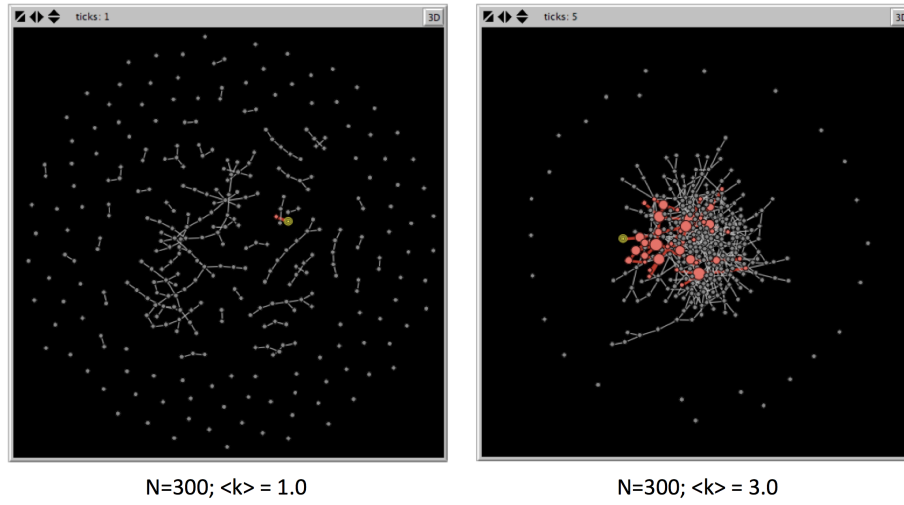


Figure 8: Influencia de la densidad de la red aleatoria en los procesos de contagio

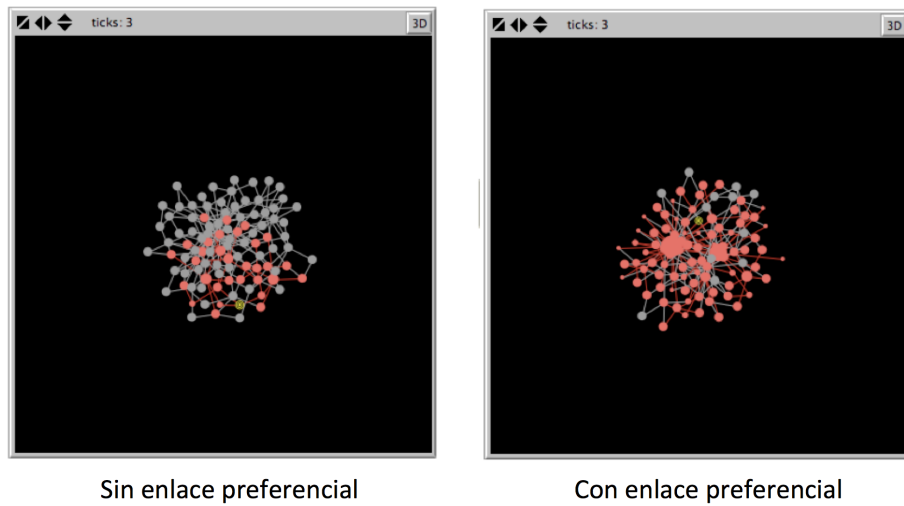


Figure 9: Influencia del enlace preferencial (redes libres de escala) en los procesos de contagio