# Análisis de Redes Sociales

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

9 de enero 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 fuertes, 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.



## Tema 9: Robustez de las redes

El fallo de un componente sencillo en un sistema complejo puede provocar un error o un fallo general de todo el sistema. Sin embargo, muchos sistemas naturales y sociales tienen la capacidad de permanecer inmutables aunque varios de sus componentes fallen. En la mayoría de estos sistemas, la robustez se consigue gracias a densas interconexiones entre los componentes.

En este tema veremos el rol de las redes con el fin de asegurar la robustez de un sistema complejo. Veremos que la estructura de la red es una de las características básicas que hay que entender para conocer la robustez del sistema y para conocer sus debilidades en caso de ataques premeditados. Así mismo, estudiaremos los comportamientos de fallo en cascada y las leyes que lo gobiernan.

## 9.1 Robustez, errores y ataques

La robustez implica entender cómo se va a comportar una red en caso de que eliminemos nodos de la misma. En particular, vamos a fijarnos en la red desde dos puntos de vista distintos:

- **Fragmentación de la red**, es decir, cómo se rompe la red en comunidades aisladas a medida que eliminamos nodos.
- **Tolerancia a errores**, es decir, cómo de largas pasan a ser las distancias en la red a medida que eliminamos nodos.

Así mismo, estudiaremos la robustez con respecto a **errores** y **ataques**. Los *errores* hacen que se eliminen nodos de manera aleatoria. Sin embargo, los *ataques* consisten en la eliminación de nodos seleccionándolos de una manera deliberada.

## 9.2 Robustez en redes aleatorias

Primeramente vamos a estudiar cómo se comporta la robustez en una red aleatoria. Para ello vamos a estudiar un tipo muy particular de redes aleatorias relacionadas con la *Teoría de la Percolación* (Percolation theory).

### 9.2.1 Teoría de la percolación

Suponemos que tenemos una cuadrícula y que en cada intersección podemos poner un nodo. Dos nodos estarán conectados si están en intersecciones adyacentes. Podemos construir una red aleatoria a partir de esta cuadrícula decidiendo si ponemos o no un nodo en cada intersección con probabilidad $p$.

La teoría de la percolación explica, entre otras cosas, el tamaño medio de los clusters y el tamaño del cluster mayor suponiendo que vamos añadiendo nodos de manera aleatoria. En la siguiente figura podemos ver que, tal y como vimos en las redes aleatorias, existe un valor crítico $p_c$[1] a partir del cual emerge un componente gigante al que pertenecerán la mayoría de los nodos.

### 9.2.2 Fragmentación de la red

Podemos estudiar la robustez de esta red aleatoria usando, a la inversa, la teoría de la percolación. Si partimos de una cuadrícula en la que en todas las intersecciones hay nodos y eliminamos los nodos con una determinada probabilidad $f$ podemos observar que, al igual que durante la creación, existe una probabilidad umbral $f_c$ que nos permite distinguir tres fases:

[1] En las redes aleatorias se cumplía cuando $\langle k \rangle \approx 1 \rightarrow p_c = \frac{1}{N}$

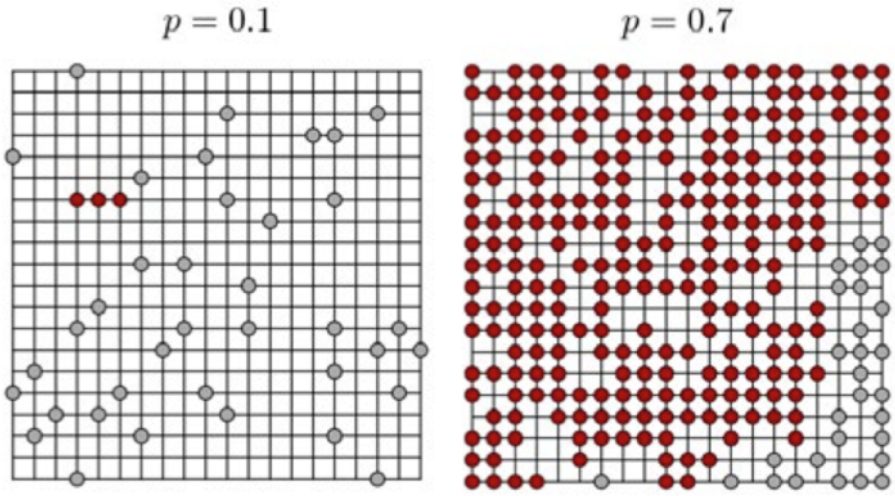


Figure 1: Redes creadas siguiendo la teoría de la percolación. En rojo están los nodos que pertenecen al cluster más grande

- Si $0 < f < f_c$ entonces continuamos teniendo un componente gigante en la red.
- Si $f = f_c$ entonces el componente gigante comienza a desvanecerse y la red se empieza a fragmentar.
- Si $f > f_c$ entonces la red queda completamente rota en muchos clusters pequeños.

A modo de conclusión podemos afirmar que la fragmentación de la red debido a fallos no es un proceso gradual. Inicialmente, la eliminación de una pequeña fracción de nodos no afecta a la integridad de la red. Sin embargo, existe un punto crítico a partir del cual la red se rompe abruptamente en pequeños grupos de nodos desconectados.

Se han hecho estudios que han hecho un cálculo aproximado de este umbral crítico $f_c$[2] dando como resultado:

$$f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$$

Para una red aleatoria, en la que el segundo momento es conocido, podemos decir que este umbral es:

[2]Para conocer más sobre cómo se calcula este umbral, revisar el Tema 8 (pp 10-11 y Temas avanzados B y C) del libro *Network Science* de Barabasi.

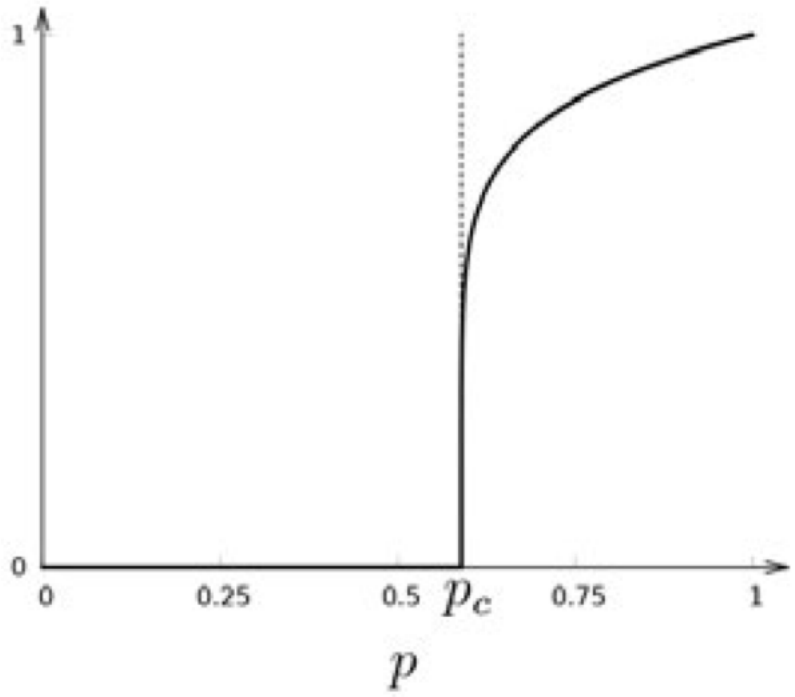


Figure 2: Probabilidad de que un nodo pertenezca a un componente gigante a partir de la teoría de la percolación

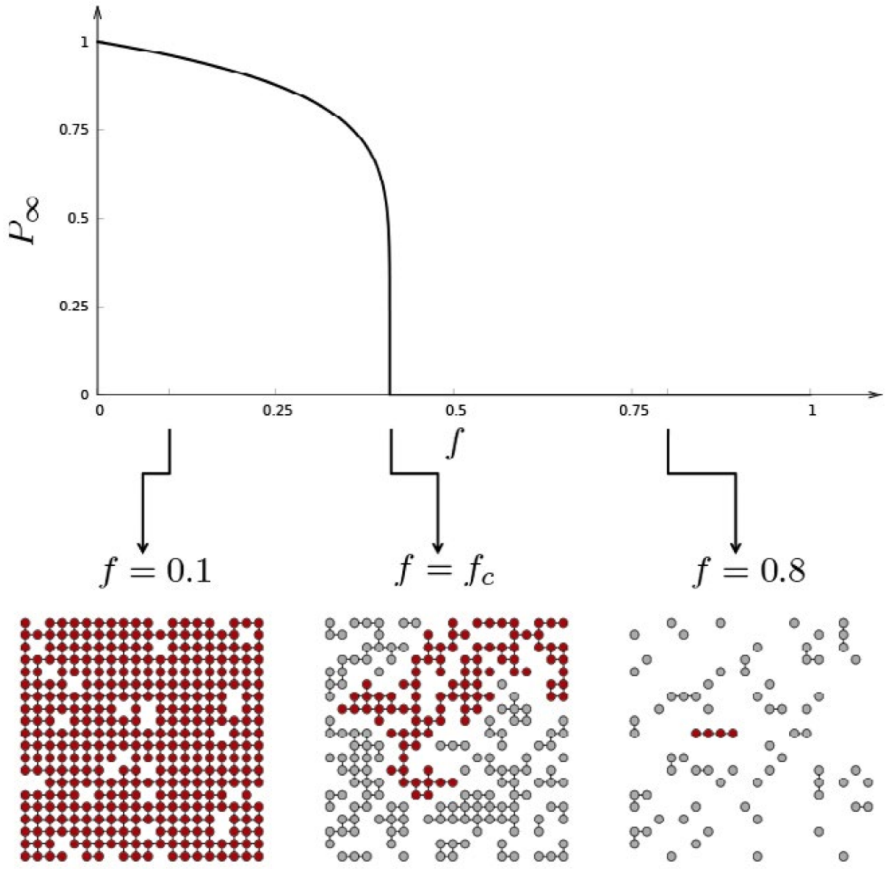


Figure 3: Probabilidad de que un nodo pertenezca al componente gigante a medida que aumentamos la probabilidad $f$ con la que eliminamos nodos de la red

$$f_c^{ER} = 1 - \frac{1}{\langle k \rangle}$$

Esto implica que cuanto más densa sea la red, mayor es el umbral $f_c$, por lo que más robusta es la red.

### 9.2.3 Tolerancia a errores

Para comprender la tolerancia a errores de una red aleatoria vamos a estudiar cómo evoluciona el diámetro de la red a medida que eliminamos nodos de manera aleatoria. Para ello calculamos el diámetro de la red a medida que aumentamos $f$.

En la figura podemos observar que el diámetro de la red crece de manera monótona para valores muy pequeños de $f$. Esto se debe a que, como la mayoría de los nodos tienen aproximadamente el mismo grado, todos contribuyen aproximadamente de la misma forma al diámetro de la red, por lo que la desaparición de cualquiera de ellos hace que las distancias vayan creciendo. Además, en el momento en el que alcancemos $f_c$ el diámetro divergirá ya que romperemos el componente gigante.

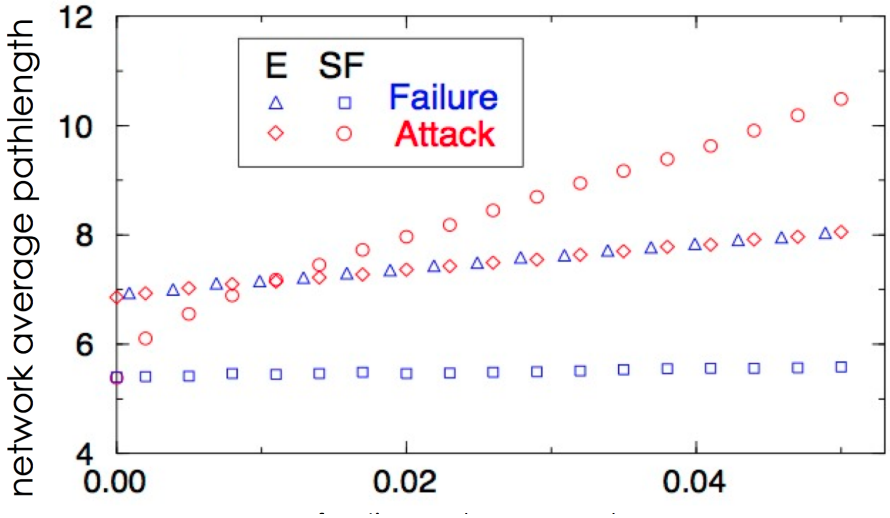


Figure 4: Diámetro de la red a medida que aumentamos el número de nodos que eliminamos de la red. Aquí se muestra solo para una pequeña fracción de nodos eliminados (E=Red aleatoria; SF=Red libre de escala)

### 9.2.4 Comportamiento frente a ataques

Ahora queremos observar el comportamiento de la red en caso de que no se produzcan errores sino ataques, es decir, se decida deliberadamente qué nodos de la red queremos eliminar en cada momento. Una forma sencilla de simular un ataque es eliminar los nodos en orden decreciente de su grado $k$, es decir, primero eliminamos el de mayor grado, luego el siguiente de mayor grado, etc.

Desde el punto de vista de la fragmentación podemos observar que la red evoluciona de la misma forma ya sea por errores o por ataques. Esto se debe a que todos los nodos tienen un grado similar por lo que un ataque deliberado no es más efectivo que los errores aleatorios en el caso de una red aleatoria.

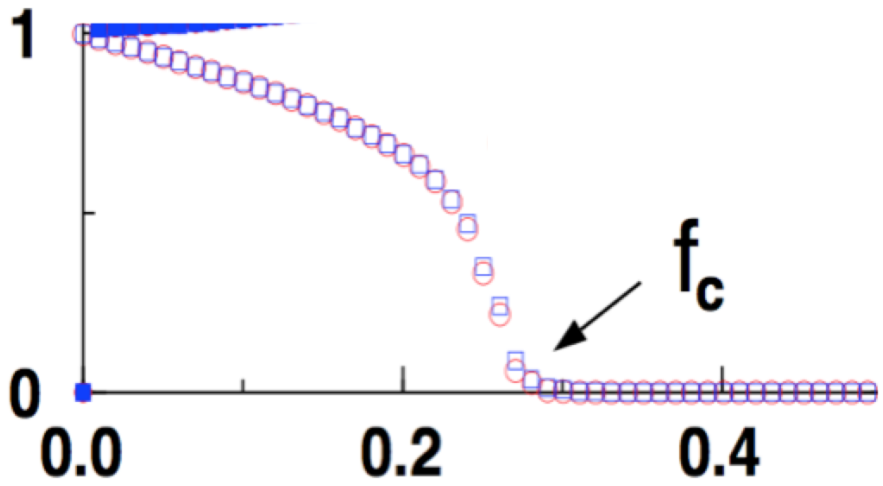


Figure 5: Probabilidad de pertenecer al componente gigante a medida que variamos la probabilidad de que los nodos sean eliminados. Los cuadrados azules muestran la evolución debida a fallos mientras que las circunferencias rojas muestran la evolución debida a ataques.

En cuanto al diámetro de la red observamos el mismo comportamiento viendo la figura de la sección anterior: el ataque a los nodos de mayor grado sigue haciendo que la distancia crezca de manera monótona y no se ve diferencia apreciable con respecto a la tolerancia a errores.

A modo de resumen podemos decir que las redes aleatorias son resistentes a ataques y que su resistencia es similar para errores aleatorios que para ataques.

## 9.3 Robustez en redes libres de escala

Lo visto hasta ahora para redes aleatorias no es válido para las redes libres de escala, principalmente debido a la distribución de grados tan diferente que hay

entre una y otra red. A continuación veremos cómo se comportan las redes libres de escala en cuanto a robustez.

### 9.3.1 Fragmentación de la red

Las simulaciones con datos reales (Internet) y con redes creadas a partir del modelo de Barabasi-Albert (estudiado en el tema 4) demuestran que la teoría de la percolación no es aplicable a este tipo de redes. En realidad se ha observado que el umbral $f_c$ a partir del cual se rompe el componente gigante y la red queda rota en pequeños clusters es cercano a 1. Esto implica que las redes libres de escala son extremadamente robustas a errores aleatorios. Desde el punto de vista de Internet, esta conclusión nos dice que sería necesario que prácticamente todos los routers existentes fallasen para que esta gran red quedase fragmentada.

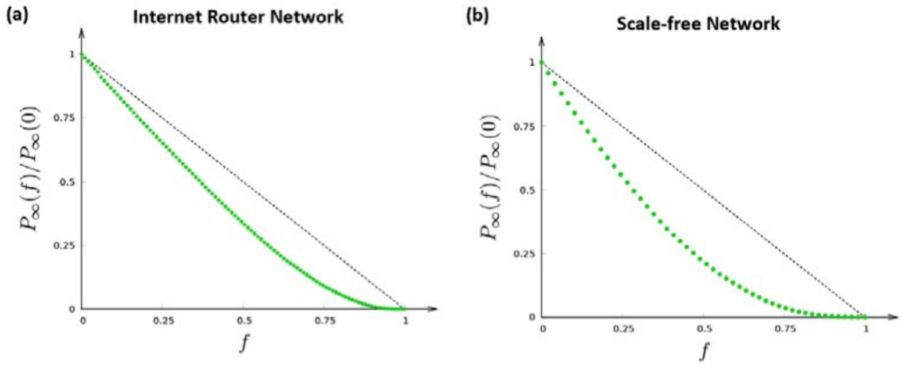


Figure 6: Probabilidad de que un nodo pertenezca al componente gigante a medida que aumentamos la probabilidad $f$ con la que eliminamos nodos de la red. Resultados de la simulación con (a) datos de Internet y (b) una red creada siguiendo el modelo de Barabasi-Albert

Matemáticamente podemos verificar este resultado con la aproximación del umbral $f_c$ mostrada anteriormente:

$$f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$$

Tal y como vimos en anteriores temas, para redes libres de escala con $\gamma < 3$ y $N \to \infty$ el segundo momento diverge por lo que $f_c$ converge a 1. Esto confirma que es necesario eliminar casi todos los nodos de este tipo de red para que se produzca una ruptura en la red.

Aunque este umbral se calcula para redes en las que $N \to \infty$, en las redes reales de gran tamaño el cálculo se aproxima bastante bien a la realidad. Por ejemplo, usando los datos de Internet en los que $\langle k \rangle = 6,34$ y $\sigma = 14,14$:

$$\langle k^2 \rangle = \sigma^2 + \langle k \rangle^2 = 240,1296$$

$$\frac{\langle k^2 \rangle}{\langle k \rangle} = 37,8753$$

$$f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1} = 0,9728$$

Esto implica que sería necesario eliminar aleatoriamente el 97% de los routers para provocar una ruptura de la red. Esto implica que deberían fallar unos 220000 routers (que son aproximadamente el 97% de esta red de $N = 228263$). Esta es una de las razones por lo que esta red es tan segura a fallos.

Esta propiedad se debe a que los hubs son mucho menos abundantes que los nodos con menor grado, por lo que los errores aleatorios eliminarán, con mayor probabilidad, uno de estos nodos con grado pequeño antes que un hub. Estos pequeños nodos contribuyen poco en mantener la integridad de la red.

### 9.3.2 Tolerancia a errores

Desde el punto de vista del diámetro de la red, las conclusiones son bastante similares: el diámetro permanece prácticamente inalterable ante errores aleatorios en la red. Al igual que antes, esto se debe a que los hubs son los grandes responsables de que las distancias dentro de la red sean cortas y un error aleatorio tiene una baja probabilidad de afecta a estos hubs, lo que mantiene las distancias prácticamente inalterables.

### 9.3.3 Comportamiento frente a ataques

Los resultados vistos hasta ahora son bien distintos en el caso de que se produzcan ataques. Tal y como vimos en una sección anterior, podemos simular los ataques eliminando los nodos en orden decreciente de grado. Esto hace que los primeros nodos eliminados en un ataque a una red libre de escala son los hubs, haciendo que la red se fragmente rápidamente con unos pocos nodos eliminados.

Se puede ver que el umbral $f_c$ cae drásticamente en el caso de producirse ataques, siendo extremadamente bajo. Esto implica que es suficiente atacar solo una pequeña fracción de nodos para romper la red en pequeños grupos.

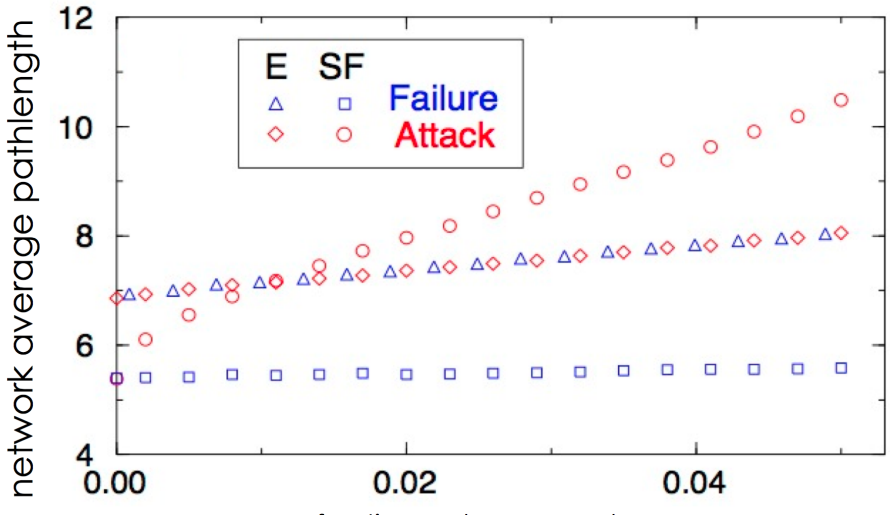


Figure 7: Diámetro de la red a medida que aumentamos el número de nodos que eliminamos de la red. Aquí se muestra solo para una pequeña fracción de nodos eliminados (E=Red aleatoria; SF=Red libre de escala). Se observa que, ante fallos, la SF permanece prácticamente inalterable

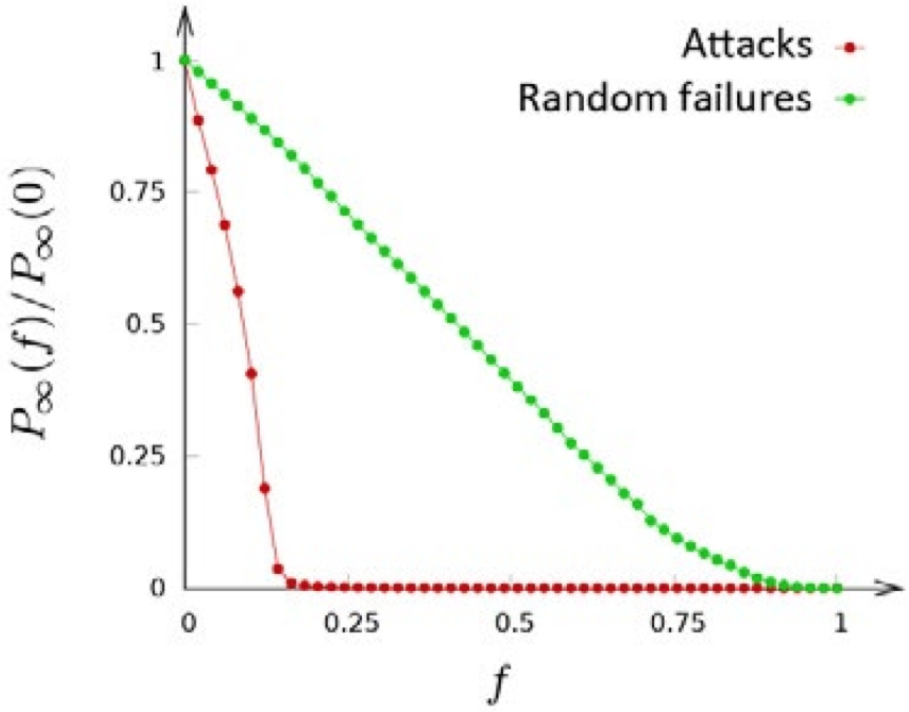


Figure 8: Probabilidad de pertenecer al componente gigante a medida que variamos la probabilidad de que los nodos sean eliminados. La línea verde representa la evolución debida a fallos aleatorios mientras que la roja muestra como el umbral cae debido a ataques