



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Facultad de Ciencias

Grado en Física

Curso 2019/2020

Trabajo Fin de Grado

**REDES COMPLEJAS: DESCRIPCIÓN,
ANÁLISIS Y APLICACIONES**

Autor: **Pablo Ibáñez Porras**

Tutor: **José Miguel Mateos Roco**

Tutor: **Ángel Martín del Rey**



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Faculty of Science

Physics degree

Academic course 2019/2020

Final Degree Project

**COMPLEX NETWORKS: DESCRIPTION,
ANALYSIS AND APPLICATIONS**

Author: **Pablo Ibáñez Porras**

Tutor: **José Miguel Mateos Roco**

Tutor: **Ángel Martín del Rey**



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

REDES COMPLEJAS: DESCRIPCIÓN, ANÁLISIS Y APLICACIONES

D. José Miguel Mateos Roco y D. Ángel Martín del Rey autorizan la presentación de este Trabajo de Fin de Grado titulado “Redes complejas: descripción, análisis y aplicaciones” realizado bajo su tutela por el estudiante del Grado en Física D. Pablo Ibáñez Porras.

D. José Miguel Mateos Roco

D. Ángel Martín del Rey

En Salamanca, Julio de 2020

*“No conozco fórmulas que garanticen una victoria.
Sólo ocasiones que la facilitan o la impiden.”*

Rodrigo Díaz de Vivar

A. Pérez-Reverte, *SIDI*

Resumen

El estudio de las redes complejas supone una parte fundamental de la **modelización matemática**. Una red está formada por **nodos** (centros de información) y **aristas** (vías de transporte de información), así como una serie de conceptos de la teoría de grafos que se introducen en la primera parte del trabajo. De especial importancia es el concepto de **grado**, definido como el número de aristas conectadas a un nodo. Ligado a esta idea, a través de la **distribución de grado**, que determina la densidad de nodos por grado, se puede caracterizar la red en su totalidad en lugar de tratar los nodos como un conjunto discreto.

En el presente trabajo se definirán y estudiarán las principales propiedades de diferentes tipos de redes: homogéneas, aleatorias, de mundo pequeño y de escala libre. En las **redes homogéneas** todos los nodos tienen el mismo número de conexiones. Estas redes solamente son de utilidad para describir un conjunto reducido de fenómenos reales, por lo que en este trabajo se tratarán brevemente. Las **redes aleatorias**, generadas usualmente siguiendo el modelo de unión aleatoria de Erdős-Rényi, comienzan a acercarse a las redes reales pero fallarán a la hora de mostrar la existencia de nodos hiperconectados. Las **redes de mundo pequeño**, que comenzaron a ser estudiadas en el ámbito sociológico, mejoran en cuanto a su parecido con las redes reales en el **agrupamiento** (una característica que muestra la relación de unión entre nodos), pero fallan al explicar la existencia de los nodos hiperconectados. Por último, hablaremos de las **redes de escala libre**, basadas en el modelo de Barabási-Albert y caracterizadas por la unión preferencial, que permiten describir el fenómeno de crecimiento de muchas redes reales. De este modo, si un nodo se añade en una red, tenderá a unirse con aquellos que tengan un mayor grado, aumentando el número de conexiones de estos y dando lugar a la aparición de nodos hiperconectados que no permitían los demás modelos.

En la tercera parte se lleva a cabo un análisis de la **propagación de epidemias** que, gracias al desarrollo de la compartimentalización de la población, se puede estudiar haciendo uso de la modelización matemática. Se presentan y estudian tres modelos diferentes, **SI**, **SIS** y **SIR**, que muestran la variación del grupo de infectados, susceptibles y recuperados. Esta variación se determina en cada modelo mediante un sistema de ecuaciones diferenciales gobernado por el uso de distintas leyes de acción de masas, según el modelo a tratar.

En la última parte del trabajo, se trata la relación entre los modelos de propagación de epidemias y las redes complejas, lo que se introducirá a través de un factor que represente el número de individuos susceptibles que se encuentran en contacto con un nodo infectado. De esta forma, los modelos epidemiológicos se pueden conectar con la distribución de grado de la red considerada, con lo que cada topología de red podrá determinar una evolución diferente de la epidemia.

Palabras clave: *redes complejas, propagación de epidemias y modelización matemática.*

Abstract

The study of complex networks is a fundamental part of mathematics modelization. The networks are constructed by nodes (that are the point which receives information) and edges (the way that information follows), and other concepts of graph theory that it is going to be studied at the first part of this work. It is too important to study the importance of the degree, defined as the number of connections of a node. Using this idea, appears the degree distribution of the network, that is the density of nodes by degree. Then, the networks can be observed as a continuous, in place of a finite group of nodes.

In this project, the most important features of different complex networks will be defined and studied: homogeneous, random, small-world and scale-free. In **homogeneous networks** all nodes have the same number of connections, so they have not a good approximation to real networks and they are going to be introduced briefly. **Random networks** are usually generated by the Erdős-Rényi algorithm and they start to represent better real networks, but they fail when explaining the existence of **hubs** which are nodes with a great number of connections. The **small-world networks**, that started to be studied in the sociological ambit, are closer to real networks at the clustering (a feature that allows to study the relation between the nodes' connection) but also failed when explaining hubs. The last networks that are going to be studied are **scale-free networks**, which are based in Barabási-Albert algorithm and characterized by preferential attachment, and describe the growth of several kinds of networks. In this way, if a node is added to the network, it would tend to be connected to the nodes that have higher degree, increasing their number of connections and appearing the hubs that other models do not explain.

At the third part, the propagation of epidemics by the compartmentalization of the population will be studied, introducing three different models: **SI**, **SIS** and **SIR**. These models show the variation of susceptible, infected or recovered. This variation is determined in each model by a differential equation system controlled by different mass action laws, depending on the model to analyze.

At the last part of the project, it is studied the relation between the models of epidemics' propagation and complex networks, thanks to a factor that represents the number of susceptible individuals connecting with an infected one. Therefore, epidemiological models can be related with the degree distribution of networks, giving to each topology the key to develop the evolution of the epidemic.

Key words: *complex networks, epidemic propagation and mathematics modelization.*

Índice

| | |
|--|-----------|
| Resumen/Abstract | I |
| Glosario | VIII |
| Introducción | IX |
| I Teoría de grafos | 1 |
| II Redes complejas | 5 |
| 1. Redes homogéneas o regulares | 5 |
| 2. Redes aleatorias | 7 |
| 2.1. Modelo de Erdős-Rényi | 7 |
| 2.2. Propiedades de agrupamiento y distancia | 10 |
| 2.3. Dinámica temporal de la red aleatoria | 11 |
| 2.4. Otros modelos de generación | 15 |
| 3. Redes de mundo pequeño | 18 |
| 3.1. Modelo de Watts-Strogatz | 19 |
| 4. Redes de escala libre | 21 |
| 4.1. Hubs | 21 |
| 4.2. El papel del exponente de grado | 22 |
| 4.3. Modelo de Barabási-Albert | 24 |
| 4.4. Mejora de modelos | 31 |
| 4.4.1. Modelo de Bianconi-Barabási | 32 |
| 4.4.2. Atractivo inicial | 34 |
| 4.4.3. Aristas internas | 34 |
| 4.4.4. Eliminación de nodos | 35 |
| 4.4.5. Crecimiento acelerado | 36 |
| 4.4.6. Envejecimiento | 36 |
| 5. Correlación de grado | 37 |
| 5.1. Función de correlación de grado | 37 |

| | |
|---|-----------|
| 6. Comparación de redes | 39 |
| III Epidemias | 40 |
| 7. Modelización de epidemias | 40 |
| 8. Modelos | 43 |
| 8.1. Modelo SI | 43 |
| 8.2. Modelo SIS | 45 |
| 8.3. Modelo SIR | 47 |
| 9. El papel de las redes | 49 |
| 10. La dificultad de frenar una epidemia | 54 |
| 11. Redes temporales | 57 |
| 12. Desarrollos futuros | 60 |
| IV Conclusiones/Conclusions | 61 |
| Lista de figuras | 66 |
| Lista de tablas | 66 |

Glosario de términos

Teoría de grafos

| | |
|-----|---------|
| G | Grafo |
| a | Arista |
| v | Vértice |

Redes

| | |
|---------------------|--|
| L | Número de aristas |
| N | Número de nodos |
| k | Grado |
| $\langle k \rangle$ | Grado medio |
| ζ | Longitud de camino |
| d | Distancia |
| $\langle d \rangle$ | Distancia promedio |
| CD | Centralidad de grado |
| ρ_g | Densidad |
| C_i | Coficiente de agrupamiento |
| C | Coficiente de agrupamiento global |
| p_k | Función de distribución de grado |
| $\langle L \rangle$ | Número medio de aristas |
| p | Probabilidad de unión |
| m | Aristas iniciales para cada nuevo nodo |
| A | Atractivo inicial |
| β | Exponente dinámico |
| η | Coficiente de aptitud |
| r | Coficiente de eliminación de nodos |
| θ | Coficiente de crecimiento acelerado |
| ν | Coficiente de envejecimiento |
| k_{nn} | Función de correlación |

Epidemias

| | |
|----------------|---|
| S | Número de individuos susceptibles |
| I | Número de individuos infectados |
| R | Número de individuos recuperados |
| ρ_I | Densidad de infectados |
| ρ_S | Densidad de susceptibles |
| ρ_R | Densidad de recuperados |
| β | Tasa de infección |
| τ | Tiempo característico |
| μ | Tasa de recuperación |
| λ | Tasa efectiva |
| R_0 | Número reproductivo básico |
| θ | Probabilidad de que un nodo esté conectado a otro ya contagiado |
| $\omega_{i,j}$ | Peso |

Introducción

Usualmente, si uno consulta cualquier libro sobre redes complejas o teoría de grafos, encontrará en el primer capítulo el *problema de los puentes de Königsberg*. A primera vista, resulta un pasatiempo para el lector tratar de descifrar la solución a este tipo de juegos de intelecto, preguntándose la importancia que puede tener aplicarlo a situaciones más cotidianas. Lo cierto es que, según se avanza en su lectura, uno puede darse cuenta de la realidad que existe bajo ellos: todo lo que nos rodea se puede modelizar en términos de redes.

El objetivo de este trabajo se divide en tres partes con unos propósitos muy claros: explicar, estudiar y aplicar lo que es una red.

Cuando Euler trató el problema de Königsberg, simplificó la geografía de la ciudad hasta unirla con las matemáticas. Los mapas de metro, los árboles genealógicos, organigramas de empresas, relaciones entre distintas ecuaciones físicas, redes eléctricas, relaciones sociales, etimologías de palabras, ... todo sigue la misma línea: su posible representación como grafos. En el terreno de la física, se me ocurre la importancia que las representaciones muestran en: la electrónica con los diagramas de circuitos, la física de nuclear y de partículas con los diagramas de Feynman o la termodinámica con los diagramas termodinámicos. Pese a que no sean grafos que se puedan estudiar a partir de los conocimientos que se van a mostrar en este trabajo, la forma en la que buscan representar elementos reales y las relaciones entre ellos, les dotan de un carácter cercano a nuestro estudio.

Puesto que no se quiere llevar el lector a un trabajo extenso, no se comentarán aspectos que puedan tener un contenido más sociológico del estudio de redes, aunque tengan gran interés como *mis amigos tienen más amigos que yo* o *¿por qué tener amigos obesos hará que tenga más probabilidades de ser obeso?*, que se recomiendan como futuras lecturas. Tampoco se tratará el famoso problema de *los puentes de Königsberg*, ya mencionado; u otros como *el cartero chino* o *los pozos y las familias rivales*. Se recomienda la lectura de [1] donde se puede encontrar un manual divulgativo sobre redes.

Sobre el trabajo, está dividido en tres partes. En la primera se tratarán definiciones básicas sobre teoría de grafos para que el lector conozca los conceptos que se irán tratando según se avance en el texto. La segunda, dedicada a introducir los distintos tipos de redes complejas: *homogéneas*, *aleatorias*, *de mundo pequeño* y *de escala libre*; además de mostrar distintos algoritmos de generación de cada una, y varias de las características que determinan cada tipo: *distancia*, *agrupamiento* y *distribución de grado*, entre otros. El comportamiento de esta última propiedad resultará clave a la hora de tratar las epidemias y será nuestro principal punto de análisis. Por último, en la tercera parte aparece una aplicación del estudio de redes complejas: las epidemias; donde se definirán, y tratarán los tres modelos básicos y más conocidos: SI, SIS y SIR. Se mostrará la importancia de las redes en la propagación y el contagio, así como en la lucha contra su avance.

La teoría de grafos es una disciplina propia, aplicable a cualquier otra que se nos pueda ocurrir. Mezcla de la formalidad matemática y los métodos de la física, con observaciones sociológicas, simulaciones informáticas y análisis de datos. Sin duda, las redes resultan totalmente apasionantes.

Parte I

Teoría de grafos

Sea un sistema de relaciones entre distintos elementos, se denomina grafo a la representación matemática de dicho sistema mediante un esquema gráfico. Formalmente,

Definición 1. Se denomina **grafo** G a la pareja $G = (V, A)$, donde V es el conjunto nodos o vértices y A es el conjunto de arcos o aristas. Cada arista $a \in A$ conecta dos vértices de V . Se escribe $a = (v_i, v_j)$, con $v_i, v_j \in V$ para indicar que la arista a conecta o une los vértices v_i y v_j , que diremos que son adyacentes.

A lo largo del trabajo se hablará de forma indistinta de redes y grafos pese a que exista una diferencia entre ambos términos. Una red representa las relaciones entre nodos de un sistema real y cada una de sus componentes tiene un sentido físico. Un grafo muestra la matemática detrás de la red, es una proyección de esta para poder estudiar sus propiedades.

Podemos definir los dos términos esenciales de una red de manera menos formal:

- **Nodo.** Componente que supone el punto de partida o llegada de la información a tratar en la red.
- **Arista.** Unión entre nodos que conlleva una relación entre ellos. Es la vía por la que circula la información.

En el caso de los grafos hablaremos de vértices en lugar de nodos.

Definición 2. Los grafos en los que el orden de los extremos de las aristas no importa se denominan **no dirigidos**, lo que significa que $a = (v_i, v_j)$ y $a' = (v_j, v_i)$ son iguales, con v_i y v_j dos vértices cualquiera del grafo.

Definición 3. Se considera **digrafos** a los grafos donde el orden de los extremos de las aristas sí importa, cumpliendo que $a \neq a'$.

Nótese que en un digrafo existe una dirección en la que se propaga la información, del vértice v_i al vértice v_j , o viceversa.

A lo largo del trabajo se tratarán grafos no dirigidos, por lo que todo lo que se introduce a continuación se relaciona únicamente con estos.

Definición 4. Se denomina **grado** del vértice al número de aristas que tienen un extremo conectado a éste y se representa con la letra k .

Definición 5. Se denomina **camino** a una secuencia de vértices y aristas consecutivas. La **longitud del camino** (ζ) se mide en el número de aristas que contiene.

Definición 6. Si el vértice inicial y el final del camino coincide, se dice que es un **ciclo**.

Definición 7. Se denomina **distancia** (d) entre dos vértices a la longitud del camino más corto que los une. La **distancia promedio**, $\langle d \rangle$, se define como el valor medio de las distancias entre dos vértices del grafo.

Definición 8. El **diámetro** de un grafo es la mayor distancia entre dos vértices cualquiera de la red o la distancia entre el par más alejado.

Definición 9. Se define un grafo **conexo** si existe, al menos, un camino que una cualquier par de vértices. En caso contrario se dirá que es **no conexo**.

Definición 10. Un grafo es **regular** si todos los vértices tienen el mismo grado.

Definición 11. Un grafo es **completo** si todos los vértices están unidos entre sí.

Dentro de la topología del grafo, podemos encontrar nodos o aristas que cobren mayor relevancia. Para cuantificar su envergadura, se hace uso de las medidas de centralidad, existiendo distintas formas de establecerla, pudiendo hablar de: grado, cercanía o intermediación; es decir, un vértice

puede ser central si está unido a un gran número de otros vértices (grado), si se encuentra en el centro de la distribución espacial de vértices (distancia) o si se encuentra en una posición determinante para permitir (o no) el paso de información entre otro par de vértices (intermediación). Por las características de este trabajo se presentará únicamente la centralidad de grado.

Definición 12. Se define la **centralidad de grado** (CD_i) como el grado normalizado del vértice v_i :

$$CD_i = \frac{k_i}{N - 1}, \quad (1)$$

con k_i el grado del vértice v_i y N el número total de vértices de la red.

Definición 13. Definimos **subgrafo** a un conjunto de vértices y aristas que se encuentran dentro del mismo grafo y cuyo número total de vértices depende de las condiciones que elijamos.

Definición 14. Se denomina **densidad**, ρ_g , a la relación entre las aristas del grafo y el número de estas que existirían si todos los vértices estuvieran unidos entre sí, es decir, si fuera un grafo completo:

$$\rho_g = \frac{2L}{N(N - 1)}. \quad (2)$$

Definición 15. Se define como **coeficiente de agrupamiento** o **clustering** (C_i) del vértice v_i a la densidad del subgrafo formado por los vecinos directos de v_i , esto es, la cantidad de aristas que existen entre los vértices unidos a v_i . Se determina como

$$C_i = \frac{2L_i}{k_i(k_i - 1)}, \quad (3)$$

donde L_i es el número de aristas entre vecinos del vértice v_i .

También se hará uso del **coeficiente de agrupamiento global**, C , que representa el promedio de los coeficientes de agrupamiento locales del grafo.

Definición 16. La **función de probabilidad de grado** (p_k) se define como la probabilidad de que un vértice seleccionado al azar tenga grado k

(con la misma probabilidad de elegir cualquier vértice). A partir de esta definición se fija la **distribución de grado**.

Definición 17. Se define el **grado medio** , $\langle k \rangle$, de un grafo no dirigido como el número medio de grados de los vértices del grafo. Se calcula como

$$\langle k \rangle = \frac{2L}{N}, \quad (4)$$

con N vértices y L el número de aristas.

Proposición 1. El grado medio de la red se obtiene a partir del momento de orden uno de la distribución de grado que siga el grafo

$$\langle k \rangle = \sum_{i=0}^{\infty} kp_k \quad (5)$$

donde además se cumple $\sum_{i=0}^{\infty} kp_k = \sum_{i=1}^{k_{max}} kp_k$, con k_{max} el mayor grado posible en un nodo del grafo.

Parte II

Redes complejas

1. Redes homogéneas o regulares

Definición 18. Se definen como **redes homogéneas o regulares** aquellas que tienen el mismo número de uniones en cada nodo, es decir, todos los nodos tienen el mismo grado.

Ejemplos especiales de este tipo de redes son (ver Figura 2):

- **Red completa.** Ver Definición 11.
- **Red anular.** Cada nodo está unido a alguno de sus primeros n – vecinos de forma periódica.

La principal característica a resaltar de estas redes es que la función de distribución de grado es una función tipo delta, pues al tener todos los nodos el mismo número de aristas, la probabilidad de que tengan grado k es $p = 1$, y en cualquier otro caso $p = 0$. En la Figura 1 se muestra la distribución que presenta esta clase de redes.

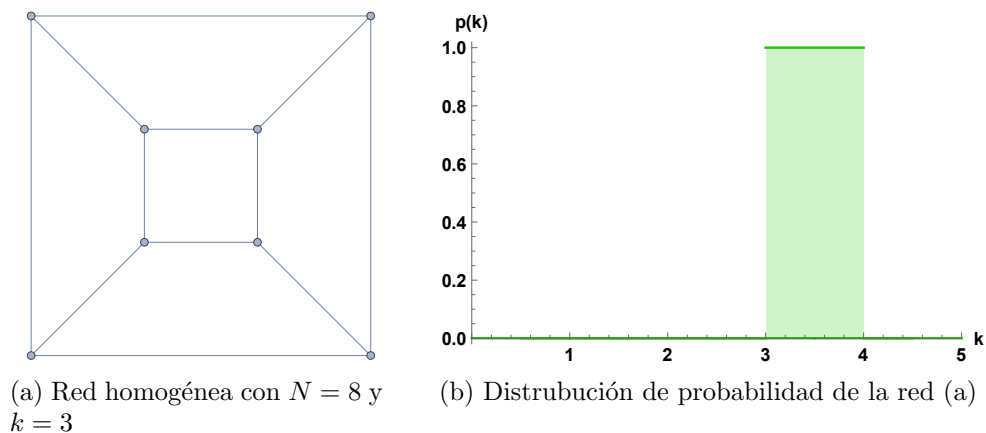
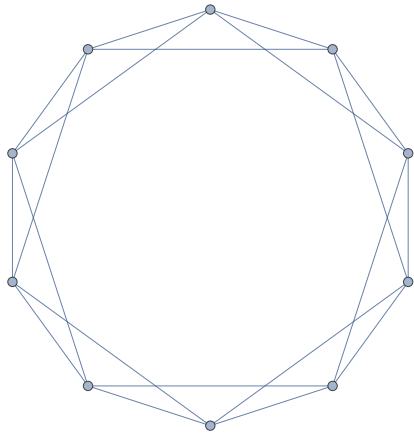
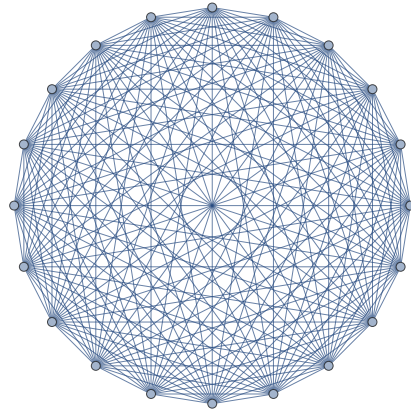


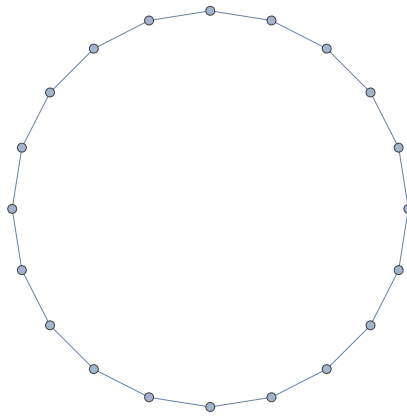
Figura 1: Ejemplo de red homogénea que se corresponde con una red cúbica, cuya proyección en tres dimensiones es un cubo.



(a) Red anular de 10 nodos, cada uno unido con sus dos primeros vecinos.



(b) Red completa con 20 nodos.



(c) Red anular con 20 nodos, cada uno unido a su primer vecino.

Figura 2: Algunos tipos de redes homogéneas.

2. Redes aleatorias

Definición 19. Las *redes aleatorias* son aquellas en las que las uniones entre sus nodos se establecen al azar.

2.1. Modelo de Erdős-Rényi

Con el uso de este modelo se pretende crear una red aleatoria partiendo de un conjunto de N nodos aislados y estableciendo conexiones aleatorias entre ellos con probabilidad $0 \leq p \leq 1$. Nótese que en el caso de $p = 0$, ningún nodo se unirá a otro. Para $p = 1$ no quedarán nodos aislados, encontrando una red completa con todos los nodos unidos entre sí. Estos dos casos extremos nos servirán más adelante para presentar la dinámica temporal de las redes aleatorias.

Destacar que, debido al papel del azar al elegir los nodos, cada vez que se reinicia el algoritmo (manteniendo la misma probabilidad) se obtiene una red distinta de la anterior, como se puede observar en la Figura 3 donde aparecen tres redes generadas con el método de Erdős-Rényi para mismos valores de p y N , pero con distinto resultado.

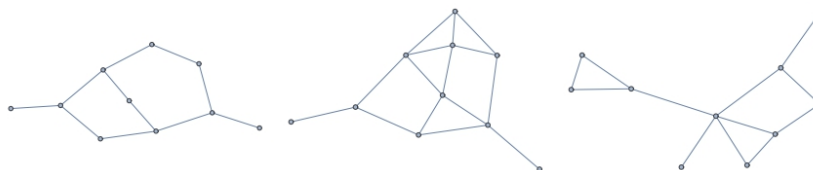


Figura 3: Tres redes aleatorias generadas con el algoritmo de ER con 10 vértices y probabilidad de unión $p = 0,3$.

Sobre el número de aristas

Sea p_L la probabilidad de que una red aleatoria contenga L aristas. Esta seguirá una distribución binomial¹, de forma que

$$p_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2} - L}. \quad (6)$$

La ecuación se compone de tres términos

¹La distribución binomial ocurre en sucesos donde la respuesta es binaria. En nuestro caso si se produce la unión o no.

- El número combinatorio $\binom{N(N-1)}{L}$ representa de cuántas formas podemos conectar los nodos según el número de aristas que estemos estudiando.
- p^L muestra la probabilidad de que se formen las L aristas.
- $(1 - p)^{\frac{N(N-1)}{2} - L}$ es la probabilidad de que el resto de uniones no se acaben formando.

El porqué de $\frac{N(N-1)}{2}$ viene al determinar que cada nodo de la red se puede unir a cualquier otro restante ($N - 1$). Se divide entre dos porque estamos tomando parejas de nodos. Así, el factor resulta en el número total de aristas posibles de la red.

El uso de la distribución binomial nos permite calcular el número medio de aristas presentes en la red como

$$\langle L \rangle = \sum_{l=0}^{\frac{N(N-1)}{2}} L p_L = p \frac{N(N-1)}{2}, \quad (7)$$

de donde se obtiene directamente el grado medio de la red al sustituir:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1) \approx pN, \quad (8)$$

que no es más que aplicar la probabilidad de unión al resto de nodos de la red y suponer $N \gg 1$.

Distribución de Poisson

Podemos estudiar la distribución de grado, p_k , a partir de la distribución normal para encontrar la probabilidad de que un nodo (escogido aleatoriamente de entre todos los que componen la red) tenga grado k , de forma que ²

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \quad (9)$$

Considerando que el tamaño de la red es muy grande, podemos llegar a la siguiente aproximación

²En este caso, el suceso binomial es que el nodo escogido tenga o no grado k .

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (10)$$

que no es más que la distribución de Poisson. Una demostración de la convergencia de la distribución normal en la de Poisson se puede encontrar en [2]. En la Figura 4 se muestra un ejemplo de red aleatoria y su distribución de grado, donde se observa que dicha distribución sigue la forma de una distribución de Poisson.

Proposición 2. La posición del máximo de la distribución depende de la densidad de la red.

Demostración. Al variar el valor de la probabilidad de unión p , también lo hace la densidad de la red (aumenta o disminuye la probabilidad de que se formen aristas entre nodos). Siguiendo la ecuación (8), también se modifica el valor de grado medio y, al coincidir con el máximo, se cumple la proposición anterior. \square

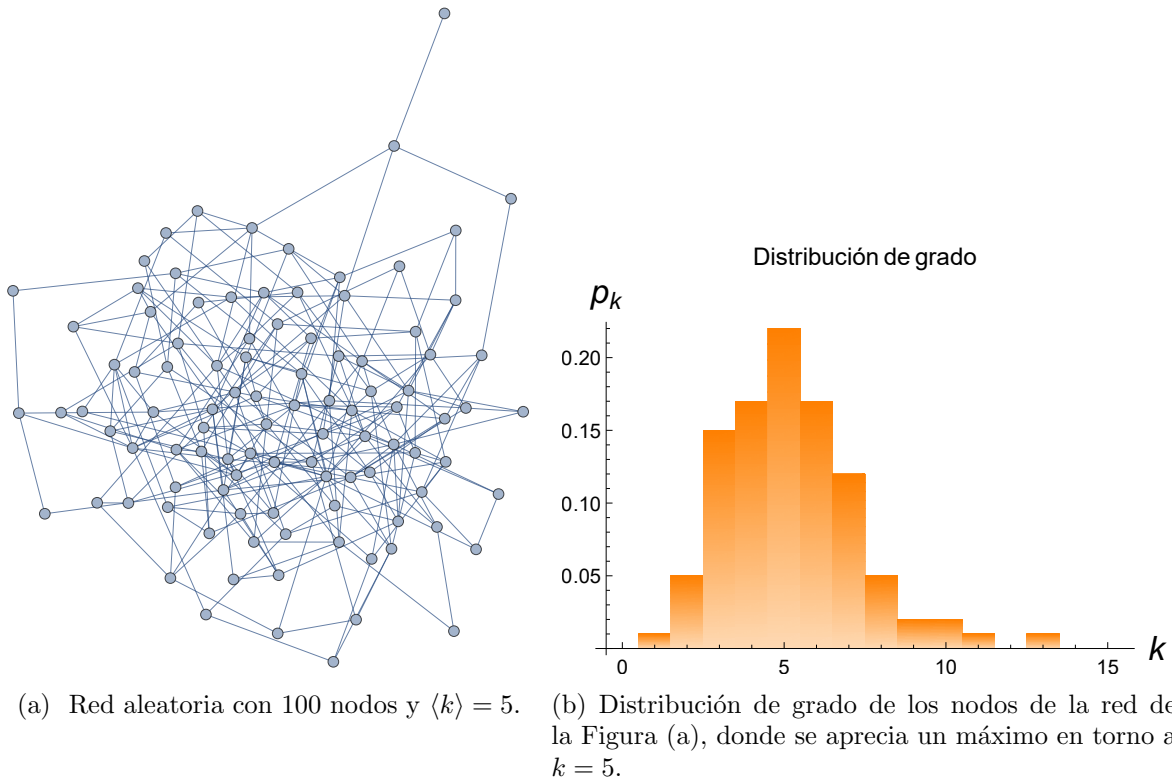


Figura 4: Red aleatoria generada por el método de Erdős-Rényi (a) y su correspondiente distribución de grado (b).

Una de las características de este tipo de distribución, es la aparición de un máximo en torno al grado medio de la red y una caída muy rápida a partir del pico. La explicación de este descenso se debe al denominador de la ecuación (10). Si se aplica la aproximación de Stirling

$$k! \approx [\sqrt{2\pi k}] \left(\frac{k}{e}\right)^k \rightarrow p_k = \frac{e^{-\langle k \rangle}}{\sqrt{2\pi k}} \left(\frac{e\langle k \rangle}{k}\right)^k,$$

encontrando que, una vez k supera el valor de $e\langle k \rangle$, los términos dependientes de k ($1/\sqrt{k}$ y $e\langle k \rangle/k$), son menores que la unidad y la distribución decrece rápidamente. Esto impide que existan valores de p_k para grados altos, haciendo que no puedan aparecer nodos hiperconectados o *hubs* (que serán estudiados más adelante), propios de las redes reales.

2.2. Propiedades de agrupamiento y distancia

Proposición 3. El coeficiente de agrupamiento local de una red aleatoria construida por el algoritmo Erdős-Rényi viene determinada como

$$C_i = \frac{\langle k \rangle}{N}. \quad (11)$$

Demostración. Si se considera el nodo i de la red y estudiando la relación entre sus k_i vecinos, se podrá encontrar que el número medio de aristas entre ellos es

$$\langle L_i \rangle = p \frac{k_i(k_i - 1)}{2},$$

donde p es la probabilidad de unión entre dos nodos de la red y que, por tanto, siguen también los nodos vecinos de i y $\frac{k_i(k_i-1)}{2}$ es el número máximo posible de aristas que se pueden formar entre los vecinos de i . Sustituyendo esta ecuación en la Definición 15, se obtiene directamente como resultado la proposición anterior, donde se hace uso de $\langle k \rangle = pN$, obtenida en la ecuación (8). \square

Proposición 4. La longitud de una red aleatoria construida por el algoritmo Erdős-Rényi (ER) es

$$\zeta \sim \frac{\ln N}{\ln \langle k \rangle}.$$

Demostración. Sea N el número de nodos de una red ER y $\langle k \rangle$ su grado medio. Si un nodo tiene grado $\langle k \rangle$, estará unido con un $N_1 = \langle k \rangle$ nodos de la red, que serán sus primeros vecinos. El número de segundos vecinos del nodo será $N_2 = N_1 \langle k \rangle$, es decir, el número de nodos que conectan los primeros vecinos del nodo inicial, resultando en $N_2 = \langle k \rangle^2$. La cantidad de n -vecinos, será entonces $N_n \approx \langle k \rangle N_{n-1} \approx \langle k \rangle^n$. La distancia entre el nodo inicial y el n -vecino viene dada directamente por n (la distancia entre el nodo inicial y el primer vecino será $\zeta = 1$ que coincide con n , entre el inicial y el segundo vecino será $\zeta = 2, \dots$), por lo que se puede considerar $n \sim \zeta$, resultando $N_n = \langle k \rangle^\zeta$. Tomando logaritmos en ambos lados se obtiene la proposición anterior. \square

2.3. Dinámica temporal de la red aleatoria

Se puede considerar que la probabilidad de unión entre nodos varía en función del tiempo, haciendo que se modifique el grado de la red. El estudio de este dinamismo conlleva a la existencia de la **componente gigante**, un agrupamiento de muchos nodos conectados.

Ya se ha comentado que existen dos casos límites según p ($p = 0$ y $p = 1$), ahora se va a relacionar dicha probabilidad con $\langle k \rangle$, pues si $p = 0$ ningún nodo estará conectado con otro y el grado medio de la red será $\langle k \rangle = 0$. Para $p = 1$, $\langle k \rangle = N - 1$.

Entre estos dos extremos, podemos encontrar estados intermedios que componen el dinamismo de la red.

Definición 20. Se define **componente gigante (CG)** como la parte de la red donde existe una mayor conexión entre nodos.

Proposición 5. Una componente gigante aparece sí y solo sí cada nodo está conectado con, al menos, otro nodo. Es decir, si $\langle k \rangle \geq 1$.

Demostración. Sea s la relación entre el número de nodos pertenecientes a la componente gigante frente al total de nodos de la red y u la fracción de nodos que no pertenecen a la CG, de forma que

$$s = \frac{N_G}{N} \quad \text{y} \quad u = 1 - \frac{N_G}{N} = 1 - s,$$

con N y N_G el número de nodos de la red y de la componente gigante respectivamente.

Supóngase un nodo i que no pertenece a la CG y un nodo j cualquiera de la red. Las posibilidades para una potencial unión son:

- Que i y j no estén unidos. Asumiremos que entonces j pertenece a la CG. La probabilidad de que dos nodos de la red no se unan es $1 - p$, con p la probabilidad de unión.
- Que estén unidos, lo que significa que j no puede pertenecer a la CG. La probabilidad de unir dos nodos que no pertenezcan a la CG es pu .

Uniendo ambas opciones, el resultado de que el nodo i no pertenezca a la CG (que equivale a que no esté unido con ningún nodo de la CG) resulta como $(1 - p + pu)^{N-1}$, es decir, $u = (1 - p + pu)^{N-1}$.

Tomando $p = \frac{\langle k \rangle}{N-1}$, sustituyendo en u , aplicando logaritmos y utilizando el desarrollo en serie $\ln(1+x) \approx x + \dots$, pues $\langle k \rangle \ll N$, resulta

$$u = e^{-\langle k \rangle(1-u)} \quad \text{y} \quad s = 1 - e^{-\langle k \rangle s}.$$

Para encontrar el valor de $\langle k \rangle$ en el cuál aparece la CG (transición de fase), derivamos respecto de s para $s = 0$ y obtenemos

$$\left. \frac{d}{ds} \left(1 - e^{-\langle k \rangle s} \right) \right|_{s=0} = 1 \quad \Rightarrow \quad \langle k \rangle = 1.$$

□

Definición 21. Se define **probabilidad de grado crítica** (p_c) como aquella probabilidad a partir de la cual se considera que existe una componente gigante.

Proposición 6. La probabilidad de grado crítica es inversamente proporcional al tamaño de la red:

$$p_c = \frac{1}{N-1} \approx \frac{1}{N}. \quad (12)$$

Demostración. Basta sustituir la condición de la Proposición 5 en la ecuación (8). □

Una característica de las redes reales es que se cumple la Proposición 5, lo que conlleva la existencia de una componente gigante apreciable (ver Tabla 1).

Dependiendo del valor de p y de su relación con p_c , podemos diferenciar cuatro etapas en el crecimiento de la componente gigante.³

- **Régimen subcrítico** ($p < p_c$). No existe una componente gigante en la red.
- **Punto crítico** ($p = p_c$). En este caso sigue sin existir una componente gigante, pero sí existen múltiples conjuntos pequeños de nodos. Marca la separación entre el régimen subcrítico (no aparición de una componente gigante) y el régimen supercrítico (componente gigante).
- **Régimen supercrítico** ($p_c < p < p_c \ln N$). Existe una componente gigante, pero no todos los nodos de la red están conectados, de forma que existen islas de nodos aisladas.
- **Régimen completamente conectado** ($p > p_c \ln N$). Todos los nodos de la red pertenecen a la componente gigante.

Proposición 7. Para que exista régimen supercrítico se debe cumplir que $1 \leq \langle k \rangle \leq \ln N$.

Proposición 8. Existirá régimen completamente conectado si $\langle k \rangle > \ln N$.

Demostración. Punto de transición entre regímenes ($\langle k \rangle = \ln N$).

Sea N_G el tamaño de la componente gigante de la red (número de nodos que pertenecen a esta componente) y sea N el número de nodos de la red, de forma que, si $N_G/N = 1$, todos los nodos de la red pertenecen a la componente gigante.

Sea p^{N_G} la probabilidad de que un nodo pertenezca a la CG, por tanto, $(1 - p)^{N_G}$, será la probabilidad de que un nodo no pertenezca.

³Nótese la semejanza en la existencia o no de la componente gigante con la ordenación de los spines según la temperatura o las distintos estados de un material en un diagrama termodinámico presión-temperatura [2].

Considérese, además, $N_G \sim N$. Entonces, el número de nodos que no se encuentran incorporados a la CG, I_N , será

$$I_N = N(1 - p)^{N_G} \approx N(1 - p)^N = N \left(1 - \frac{N \cdot p}{N}\right)^N \approx N e^{-Np}$$

donde se ha hecho uso de la caracterización del número e , ($e = \lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$) y se considera N lo suficientemente grande.

Haciendo la red considerablemente densa (aumentando p) tanto como para que el número de nodos sin conectar a la CG sea $I_N = 1$, despejando p y utilizando la ecuación (8), resulta

$$p = \frac{\ln N}{N} \Rightarrow \langle k \rangle = \ln N \quad (13)$$

que es el punto de transición entre el régimen supercrítico y el régimen completamente conectado [2].

□

Al analizar diferentes redes reales, se puede ver que no se cumple la Proposición 8 y, por tanto, no todos los nodos estarán unidos y dentro de la componente gigante, sino que quedarán desconectadas algunas islas de nodos. Se considera que las redes reales pertenecen al **régimen supercrítico** (ver Tabla (1)).⁴

| Red | N | L | $\langle k \rangle$ | $\ln N$ |
|----------------------------|--------|--------|---------------------|---------|
| Internet | 192,24 | 609,06 | 6,34 | 12,17 |
| Redes eléctricas | 4,94 | 6,59 | 2,67 | 8,51 |
| Citaciones científicas | 23,13 | 94,44 | 8,08 | 10,05 |
| Redes de actores | 702,39 | 29,40 | 83,71 | 13,46 |
| Interacciones de proteínas | 2,02 | 2,93 | 2,90 | 7,61 |

Tabla 1: Valores del número de nodos (N), aristas (L), grado medio ($\langle k \rangle$) y $\ln N$ para distintas redes reales. Se observa que se cumple $\langle k \rangle > 1$ para todas ellas, pero no $\langle k \rangle > \ln N$, por lo que el régimen es supercrítico. Datos obtenidos de [2].

⁴Existen redes sociales en las que es obligatorio recibir una invitación de otro miembro de la red para que un nuevo usuario se pueda registrar. De esta forma se unen los perfiles en el mismo momento en el que se crea la nueva cuenta, obligando a que la red social tenga siempre $\langle k \rangle \geq 1$ y que aparezca la componente gigante.

2.4. Otros modelos de generación

Pese a que el algoritmo de Erdős-Rényi es el más utilizado para generar redes aleatorias, existen otros como los que se presentan a continuación.

Modelo configuracional

El objetivo de utilizar este algoritmo es crear una red en la que cada nodo tenga un grado ya determinado pero donde las uniones sean aleatorias [3], lo que nos puede devolver distintas redes cada vez que repitamos el proceso. Este algoritmo se basa en:

1. Asignar un grado generado por una función de distribución p_k a cada nodo. Aún no se une nada pero vamos a representar el grado con un *stub*⁵.
2. Unir los nodos eligiendo al azar un par de *stubs*.
3. Repetir el paso 2 hasta que cada *stub* esté conectado a otro.

Se obtiene así una red aleatoria pero con una distribución p_k determinada.

Una de las características de esta red es que se pueden producir nodos con multiaristas y lazos (nodos unidos consigo mismos), elementos que no suelen estar presentes en las redes reales.

Sean dos nodos al azar i y j , con grados k_i y k_j , respectivamente. Si se parte de un *stub* del nodo i , habrá k_i posibles a escoger. Elegido uno, quedarán $(2L - 1)$ *stubs* libres. La probabilidad de que los nodos i y j se unan, queda como

$$p_{ij} = \frac{k_i k_j}{2L - 1}. \quad (14)$$

Red aleatoria con conservación de grado

Este algoritmo servirá para estudiar redes distintas donde el grado de cada nodo es constante, manteniendo fijo el número de nodos y aristas, pero con uniones formadas aleatoriamente. El algoritmo para formar este tipo de redes se puede ver en la Figura 5.b y sigue:

⁵Marca de inicio o fin de una arista.

1. Se selecciona una arista al azar de la red inicial y se separa uno de sus extremos.
2. Se escoge un nuevo nodo aleatoriamente y se une al extremo desconectado.
3. De este último nodo se selecciona una arista conectado a él y se separa la conexión, tomando un nuevo nodo y uniéndolo al extremos desconectado.
4. Se repite hasta que todos los nodos se reconectan y mantienen el mismo grado que en la red original, recordando que el primer nodo que se ha modificado tiene un grado menos y también se debe unir.

Aunque se ha tomado una arista para la presentación del algoritmo, el método se puede llevar a cabo reconectado varias a la vez. De esta forma, se obtiene una red donde las uniones se han producido de forma aleatoria pero donde se mantiene la misma distribución de grado que la red original [4].

Modelo de aleatoriedad total

Parecido al modelo anterior, en esta ocasión no se busca mantener el grado de cada nodo constante, formándose una red totalmente aleatoria en la distribución de grado, partiendo de una red inicial, pero con el mismo número de aristas y nodos. El algoritmo de generación se puede observar en la Figura 5.c y sigue:

- Tómnese dos nodos conectados de la red y sea desconectado uno de ellos manteniendo unido el otro.
- Se coloca el extremo suelto de la arista en otro nodo.
- Repetir el proceso para todas las aristas.

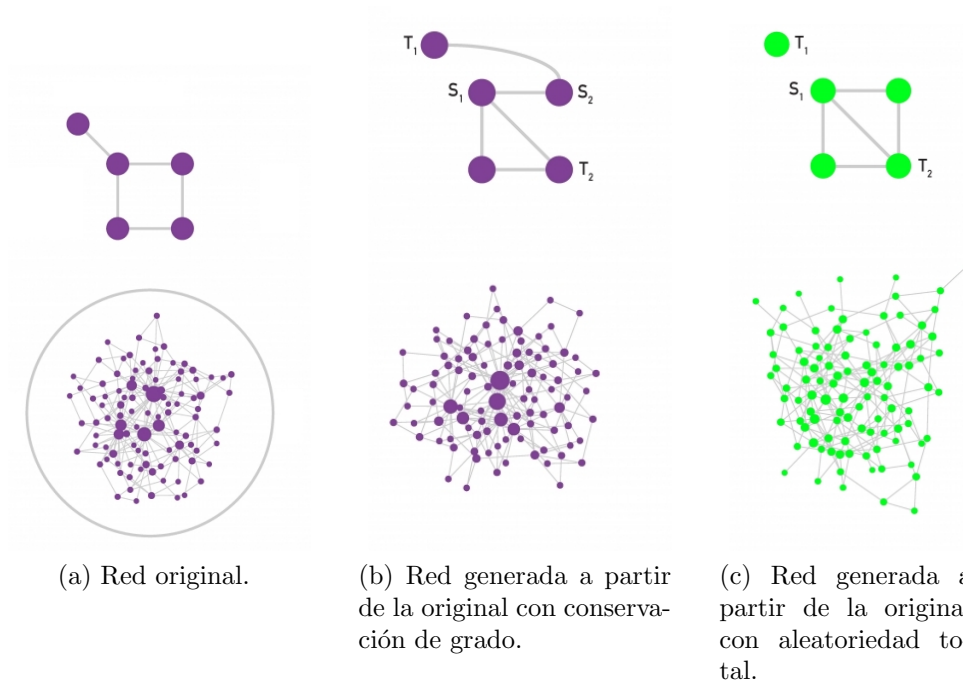


Figura 5: Diferencia entre una red aleatoria con conservación de grado (b) y una con aleatoriedad total (c) partiendo de una red inicial (a).

3. Redes de mundo pequeño

En general, se consideran como redes de mundo pequeño aquellas donde, dos nodos elegidos al azar que no estén unidos directamente entre sí, están conectados a través de un número pequeño de aristas. En este tipo de redes se presenta un alto nivel de agrupamiento.

Es interesante resaltar en el estudio de este tipo de redes la figura del psicólogo S. Milgram⁶ quien observó que para conectar dos personas (desconocidas entre sí) se necesitaban de media tan solo de unos cinco individuos intermedios. Se introdujo así la idea de los **seis grados de separación** dando nombre a este tipo de redes, que constituyen una muy buena aproximación a las redes sociales reales.

Se propone la lectura de [5] para comprender la expectación que siguen produciendo estas redes. Utilizando una técnica similar a la de Milgram se encontró que entre dos miembros de la red social Facebook (sin conexión directa entre sí) existe una distancia media $\langle d \rangle = 4,74$ [6].

Distancia

Considérese una red en el que el grado de cada nodo sea el grado medio de la red. En ese caso, un nodo i elegido al azar tendrá como primeros vecinos a otros $\langle k \rangle$ nodos, que se encuentran a distancia uno. Sus segundos vecinos serán $\langle k \rangle^2$ nodos, que se encuentran a distancia dos, y así hasta los d -ésimos vecinos, a distancia d . Así, el número total de nodos que se encuentren unidos al nodo i , serán

$$N(d) = \sum_{n=0}^d \langle k \rangle^n = \frac{1 - \langle k \rangle^{d+1}}{1 - \langle k \rangle} \approx \langle k \rangle^d,$$

donde se ha considerado $\langle k \rangle \gg 1$. Nótese la semejanza entre este desarrollo y la Proposición 4. El número de nodos a los que se encuentra conectado el nodo i tiene que ser finito, pues el tamaño máximo de la red se define a partir del número total de nodos que la componen, N . De esta forma $N(d_{max}) = N$, resultando

⁶Reconocido, entre otros estudios, por su experimento sobre la obediencia a la autoridad.

$$\langle k \rangle^{d_{max}} \approx N \quad \Rightarrow \quad d_{max} \approx \frac{\ln N}{\ln \langle k \rangle}.$$

Esta ecuación queda incompleta, pues necesita de varias correcciones que exceden del objetivo del trabajo y que se puede encontrar en [7].

Sin embargo, el resultado de la ecuación sí se corresponde con valores obtenidos para el valor medio de la distancia, $\langle d \rangle$, por lo que se puede describir como

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}. \quad (15)$$

La distancia entre nodos dependerá del tamaño de la red ($N \gg \ln N$) y del grado medio. Cuánto más densa y más pequeña sea la red, menor será la distancia entre nodos, pues $\langle k \rangle \sim pN$.

3.1. Modelo de Watts-Strogatz

Sea una red anular con N nodos, cada uno de ellos conectado con sus primeros n -vecinos, de forma que estas uniones sean simétricas (izquierda y derecha). El algoritmo es como sigue:

1. Se elige una arista cualquiera de la red, de forma que uno de los nodos que conecta se mantiene unido y en el otro se separa la arista con una probabilidad q .
2. La parte de la arista que queda suelta se conecta a otro nodo de la red escogido al azar.
3. Se repite hasta que todas las aristas han sido escogidas.

Dos casos límites de este modelo son:

- Para $q = 0$, la red seguirá siendo homogénea.
- Para $q = 1$, todas las aristas se habrán separado de uno de los nodos que unían en la red homogénea y se habrán vuelto a unir con otro nodo al azar, creando una red aleatoria.

Propiedades de agrupamiento y distancia

Proposición 9. El coeficiente de agrupamiento de una red construida utilizando el algoritmo de Watts-Strogatz es

$$C(q) = \frac{3(k-2)}{2(2k-1)}(1-q)^3.$$

Demostración. Véase [3]. □

Proposición 10. La distancia de una red construida utilizando el algoritmo de Watts-Strogatz es

$$L(q) = \frac{2n}{k} f(2nq/k),$$

con

$$f(x) = \begin{cases} c, & \text{si } x \ll 1 \\ \ln x/x, & \text{si } x \gg 1 \end{cases} \quad (\text{normalmente } c = 1/4). \quad (16)$$

Demostración. Ver [3]. □

Modelo de Newman-Watts

En este algoritmo no se produce la reorganización de las aristas ya existentes, sino que cada par de nodos que se encuentran desconectados en la red anular original se unirán con una probabilidad $0 < p \ll 1$. Los casos límite pasan a ser:

- Para $p = 0$, la red seguirá siendo homogénea.
- Para $p = 1$, todos los nodos estarán conectados entre sí, obteniendo una red completamente conectada.

Proposición 11. El coeficiente de agrupamiento para una red generada por el algoritmo de Newman-Watts, sigue

$$C(p) = \frac{3(K-2)}{4(K-1) + 4Kp(p+2)}.$$

Demostración. Ver [3]. □

4. Redes de escala libre

En los estudios del mapeo de Internet se observa la existencia de nodos hiperconectados que, como ya se ha visto, la distribución de Poisson no contempla (Sección 2). Se llegó a la conclusión de que la distribución de grado de esta red seguía una ley de potencias tipo

$$p_k \approx k^{-\gamma},$$

donde γ se conoce como **exponente de grado**.

Definición 22. *Se denominan **redes de escala libre** a aquellas redes cuya distribución de probabilidad de grado sigue una ley de potencias.*

4.1. Hubs

Definición 23. *Se denomina **hub** al nodo con un número de conexiones mucho mayor que la media. Considerando el nodo i :*

$$k_i \gg \langle k \rangle.$$

La existencia de supernodos se predice directamente por la ley de potencias que siguen las redes de escala libre, pues la caída exponencial obliga a que existan nodos que tengan un grado muy alto, aunque su cantidad sea reducida.

Proposición 12. Sea k_{max} el máximo valor de grado (alejado del grado medio para cumplir la condición de *hub*) y k_{min} el grado del nodo con menor número de enlaces de la red. Se puede encontrar una relación entre el tamaño del *hub* y el de la red como

$$k_{max} = k_{min} N^{\frac{1}{\gamma-1}}. \quad (17)$$

Cuanto mayor sea la red, mayor será el tamaño del *hub*, pues tendrá como grado k_{max} al ser el nodo de mayor tamaño.

Demostración. Sea una red con una distribución de grado normalizada. Consideraremos que existe un único nodo con $k > k_{max}$. Esto es:

$$\int_{k_{min}}^{\infty} p_k dk = 1 \quad \text{y} \quad \int_{k_{max}}^{\infty} p_k dk = \frac{1}{N}.$$

Descomponiendo los límites

$$\int_{k_{min}}^{\infty} p_k dk = \int_{k_{min}}^{k_{max}} p_k dk + \int_{k_{max}}^{\infty} p_k dk = 1 \Rightarrow \int_{k_{min}}^{k_{max}} p_k dk = 1 - \frac{1}{N}.$$

Se puede resolver la parte izquierda de la ecuación anterior haciendo uso de $p_k = (\gamma - 1)k_{min}^{(\gamma-1)}k^{-\gamma}$, que resulta de normalizar la distribución de grado de las redes de escala libre y cuya demostración se puede encontrar en [2], resultando

$$\begin{aligned} \int_{k_{min}}^{k_{max}} p_k dk &= \int_{k_{min}}^{k_{max}} (\gamma - 1)k_{min}^{(\gamma-1)}k^{-\gamma} dk = (\gamma - 1)k_{min}^{\gamma-1} \left[\frac{k^{(1-\gamma)}}{(1-\gamma)} \right]_{k_{min}}^{k_{max}} = \\ &= k_{min}^{(\gamma-1)} \left(k_{min}^{(1-\gamma)} - k_{max}^{(1-\gamma)} \right) = 1 - k_{max}^{(1-\gamma)} k_{min}^{(\gamma-1)}. \end{aligned}$$

Juntando todo queda

$$1 - k_{max}^{(1-\gamma)} k_{min}^{(\gamma-1)} = 1 - \frac{1}{N} \Rightarrow k_{max}^{(1-\gamma)} = k_{min}^{(1-\gamma)} \frac{1}{N},$$

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

□

4.2. El papel del exponente de grado

Ante una red que siga la distribución de escala libre existirán distintos valores del exponente de grado. Para analizar los regímenes que se pueden encontrar en función de γ , es necesario tratar sobre el valor esperado y el momento de orden dos.

El valor esperado se define como

$$\langle k \rangle = \int_{k_{min}}^{k_{max}} k p_k dk,$$

mientras que el momento de orden dos,

$$\langle k^2 \rangle = \int_{k_{min}}^{k_{max}} k^2 p_k dk.$$

Se puede encontrar una solución genérica a estas ecuaciones tomando el momento n -ésimo, $\langle k^n \rangle$; con n el orden de magnitud. El resultado de la integral queda

$$\langle k^n \rangle = \int_{k_{min}}^{k_{max}} k^n p_k dk = B \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n - \gamma + 1},$$

siendo B una constante de integración.

Una consideración que se obtiene directamente de la ecuación es que si el momento diverge para un orden n , los órdenes superiores ($n + 1, n + 2, \dots$), también lo harán. Para los momentos a tratar:

$$\langle k \rangle = \frac{k_{max}^{2-\gamma} - k_{min}^{2-\gamma}}{2 - \gamma} \quad \text{y} \quad \langle k^2 \rangle = \frac{k_{max}^{3-\gamma} - k_{min}^{3-\gamma}}{3 - \gamma}. \quad (18)$$

El valor de k_{max} viene determinado por la ecuación (17), con dependencia del número de nodos. Si se toma una red con un número de nodos muy elevado ($N \rightarrow \infty$), tendremos tres posibles regímenes:

- **Régimen anómalo** ($\gamma < 2$). Tanto $\langle k \rangle$ como $\langle k^2 \rangle$ divergen, ya que $\langle k \rangle \sim k_{max}^{2-\gamma} \sim N^{\frac{2-\gamma}{\gamma-1}}$, siendo $\frac{2-\gamma}{\gamma-1} > 0$ y se produce la divergencia.
- **Régimen de escala libre** ($2 < \gamma < 3$). $\langle k \rangle$ tiene un valor finito mientras $\langle k^2 \rangle$ diverge. Se caracteriza como de escala libre porque no se puede calcular la varianza de la distribución, σ_k , lo que impide conocer con precisión si un nodo escogido aleatoriamente de la red tendrá un grado k determinado, pues se correspondería con $k = \langle k \rangle \pm \sigma_k$, con σ_k función de $\langle k^2 \rangle$, que diverge. Así, la información que aporta el valor medio pierde relevancia.
- **Régimen aleatorio** ($\gamma > 3$). Tanto $\langle k \rangle$ como $\langle k^2 \rangle$ convergen. Se puede calcular la varianza de la distribución y la red pierde el carácter de escala libre.

4.3. Modelo de Barabási-Albert

Para explicar por qué se ajusta con tanta exactitud las redes de escala libre a las redes reales, AL. Barabási y R. Albert proponen el estudio de dos características que se dan en ellas: el crecimiento de la red y la unión preferencial.

- **Crecimiento.** Es un hecho que las redes reales se encuentran en constante aumento al incrementar el número de nodos.

Ejemplo 4.1

Si se considera YouTube como una red donde los nodos son los vídeos subidos y las aristas la relación entre ellos, el resultado es una red en la que se produce crecimiento^a.

^aSe estima que cada hora se suben unas 300 horas de vídeo a la red. Fuente *Statistic Brain Research Institute*.

- **Unión preferencial.** Cuando un nuevo nodo se incorpora a la red, las uniones al resto de nodos no se produce de forma aleatoria, sino que tienden a generar aristas con aquellos nodos que están más conectados (mayor grado).

Algoritmo de Barabási-Albert

Sea una red con m_0 nodos con las aristas entre ellos creadas al azar, de forma que en cada nodo se cumpla $k \geq 1$.

- **Crecimiento.** Se incorpora un nuevo nodo a la red por cada paso de tiempo, siendo el grado de este nuevo nodo $k = m$, con $m \leq m_0$.
- **Unión preferencial.** La probabilidad de que el nuevo nodo se una a otros nodos de la red no está definida de forma aleatoria, sino que depende del grado del nodo al que se vaya a unir. De esta forma, la probabilidad de que el último nodo incorporado se una al nodo i es

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (19)$$

Este modelo representa el llamado *Efecto Mateo*. Se basa en el dicho popular de *dinero llama a dinero*: cuanto mayor sea el grado del nodo, mayor

será la probabilidad de que un nuevo nodo establezca una unión y aumente su grado. Queda claro entonces que es posible la aparición de *hubs* con este modelo.

Dinámica temporal de grado

Las nuevas uniones que se producen como consecuencia del crecimiento de la red hacen que el grado de los nodos que se encontraban ya presentes varíe.

Proposición 13. El grado del nodo i varía con el tiempo, debido a la unión con los nuevos nodos que aparecen en la red según

$$k_i(t) = m \left(\frac{t}{t_i} \right)^\beta, \quad (20)$$

donde β se denomina **exponente dinámico**.

Demostración. Considérese el nodo i , que se añade a la red en el instante t_i . Supóngase k como una variable continua, de forma que $\Pi(k_i)$ es la tasa de cambio que sufre la variable, además de una cierta constante que, por condición inicial, es el número de aristas iniciales con las que el nuevo nodo se une a la red. Así, la variación de grado del nodo i será

$$\frac{dk_i}{dt} = m\Pi(k_i) = m \frac{k_i}{\sum_{j=1}^{N-1} k_j}. \quad (21)$$

El sumatorio del denominador representa el grado de todos los nodos de la red menos el del nuevo nodo, esto es $\sum_{j=1}^{N-1} k_j = 2mt - m$, lo que lleva a

$$\frac{dk_i}{dt} = \frac{k_i}{2t - 1} \approx \frac{k_i}{2t}.$$

La integral de la ecuación anterior nos lleva directamente a (20), con la condición inicial $k_i(t_i) = m$. Esta condición de integración representa que en el momento de incorporar el nodo i a la red (t_i), el número de conexiones posibles es m , pues es la condición de crecimiento del algoritmo de Barabási-Albert (BA). \square

Algunas consideraciones a tratar a partir de la ecuación (20) para el modelo BA:

- El exponente dinámico (β) tiene un valor $\beta = 1/2$ para todos los nodos. El que se encuentre en el régimen sublineal (i.e., que k tenga un crecimiento menor que si fuera una relación lineal con t), viene de la condición de crecimiento. Según avanza t , cada vez habrá más nodos en la red, de forma que el último en aparecer tiene más posibles elecciones que el anterior. Visto al revés, habrá cada vez una mayor cantidad de nodos que competirán por unirse con el nuevo.
- Cuanto antes se incorpore el nodo a la red, mayor será su grado. Esto se conoce como *ventaja del primer movimiento*.
- Se puede obtener la velocidad de crecimiento del grado del nodo i derivando la ecuación (20), resultando

$$\frac{dk_i(t)}{dt} = \frac{m}{2} \frac{1}{\sqrt{t_i t}}. \quad (22)$$

Para conocer el valor de la preferencia de unión, se compara una misma red en dos tiempos distintos, conocido el grado del nodo i en ambos instantes. Entonces, el incremento del grado de dicho nodo será $\Delta k_i = k_i(t + \Delta t) - k_i(t)$, con Δt el incremento de tiempo entre ambos instantes. De acuerdo con la ecuación (21), tendremos

$$\frac{\Delta k_i}{\Delta t} \sim \Pi(k_i) \quad (23)$$

Distribución de grado

Proposición 14. La distribución de grado de una red de escala libre formada a partir del modelo de Barabasi-Albert sigue

$$p_k \sim 2m^2 k^{-3}. \quad (24)$$

Demostración. Partiendo de la probabilidad de que el nodo i tenga grado menor de k , $P(k_i < k)$, se aplica la ecuación (20), de forma que

$$P(k_i < k) = P\left(t_i > t \frac{m^2}{k^2}\right) = 1 - P\left(t_i \leq t \frac{m^2}{k^2}\right),$$

$P\left(t_i \leq t \frac{m^2}{k^2}\right)$ es la probabilidad de que el nodo i se incorpore a la red en un instante de tiempo t_i anterior a $t \frac{m^2}{k^2}$. Esta se corresponde con la probabilidad de que un nuevo nodo aparezca en cada instante de tiempo

$$P\left(t_i \leq t \frac{m^2}{k^2}\right) = \sum_{j=1}^{t \frac{m^2}{k^2}} P(t_j) = t \frac{m^2}{k^2} P(t),$$

donde $P(t_j)$ es la probabilidad de que se una un nuevo nodo j a la red en el instante t_j , con $t_j \leq t_i$ y $P(t)$ la probabilidad de que se incorpore un nodo a la red a tiempo t . Como a cada paso de tiempo aparece un nuevo nodo, la distribución toma una función uniforme y resulta $P(t) = \frac{1}{m_0+t}$, resultando

$$P(k_i < k) = 1 - \left(\frac{m}{k}\right)^2.$$

Para obtener la densidad de probabilidad, se deriva con respecto a la variable que considerada

$$p(k) = p(k_i(t) = k) = \frac{d}{dk} P(k_i < k) = \frac{d}{dk} \left[1 - \left(\frac{m}{k}\right)^2\right],$$

obteniendo

$$p_k \sim 2m^2 k^{-3}.$$

□

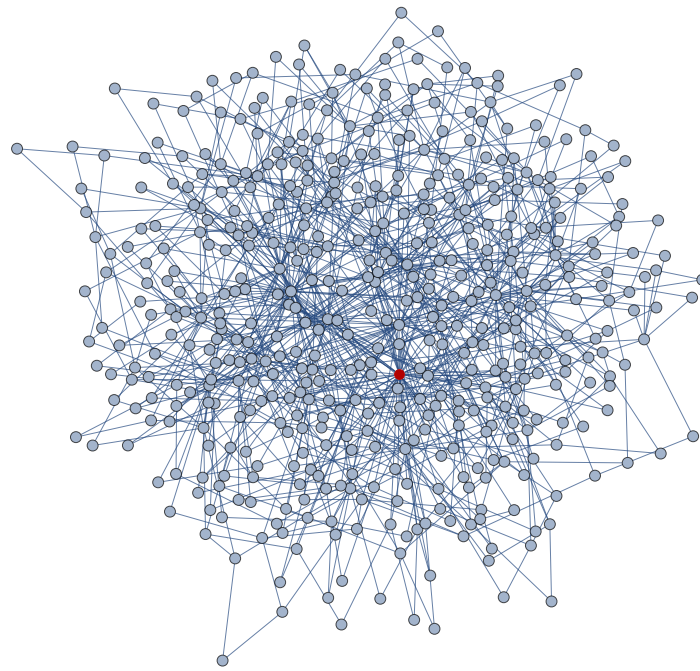
Proposición 15. El resultado exacto de la distribución para el modelo BA sigue

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}$$

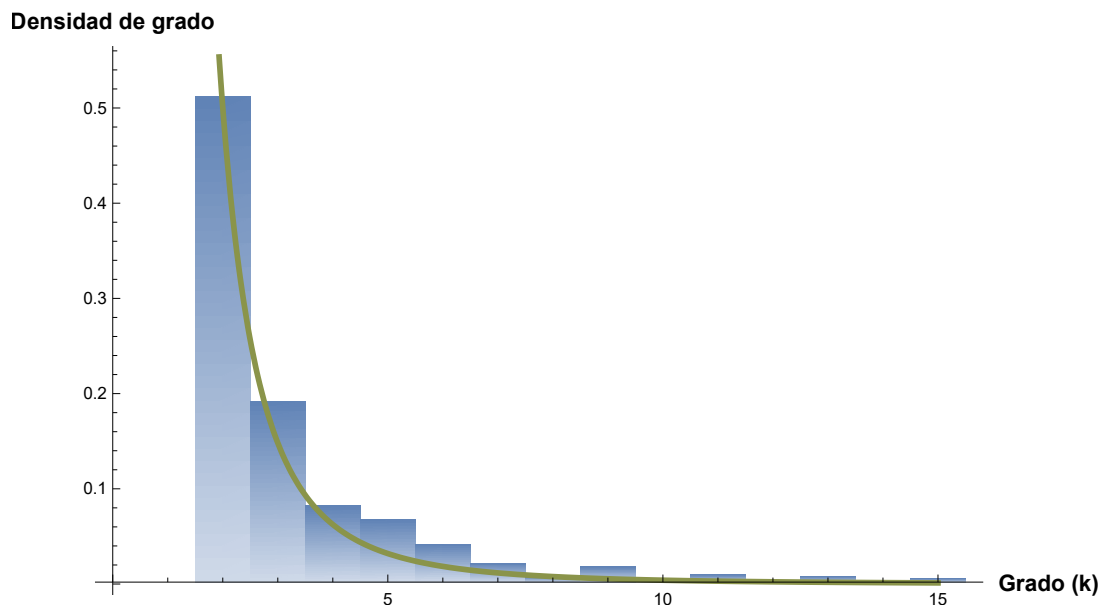
Demostración. Ver [2].

□

En la Figura 6 se muestra una red de escala libre creada a través del algoritmo BA (a) y su correspondiente distribución de grado (b), donde se observa que se cumple la ley de potencias de la ecuación (24).



(a) Red de escala libre con el *hub* señalado en rojo.



(b) Distribución de grado (barras) de la red y función de distribución de grado teórica (ecuación (24)) en verde.

Figura 6: Red de escala libre formada por el algoritmo de Barabási-Albert con 500 nodos y $m = 2$ (a) y su distribución de grado (b), donde se observa que sigue la ecuación (20) (pintada en verde).

Propiedades de distancia y agrupamiento

Proposición 16. El coeficiente de agrupamiento global de una red generada por el algoritmo de Barabasi-Albert sigue

$$C \sim \frac{(\ln N)^2}{N}.$$

Demostración. Ver [2]. □

Proposición 17. La distancia promedio en una red generada por el algoritmo de Barabasi-Albert es

$$\langle d \rangle \sim \frac{\ln N}{\ln \ln N}.$$

Demostración. Ver [8]. □

Otros modelos con preferencia de unión

Los modelos que se muestran a continuación están relacionados con el de Barabási-Albert, pues también presentan sus dos pasos principales: crecimiento y preferencia de unión. El crecimiento en todos los modelos es igual: cada vez que se reinicia el algoritmo se introduce un nuevo nodo a la red. Lo que varía es la manera en la que el nodo introducido se une con los demás.

Modelo de selección de arista

Una vez que se ha producido el crecimiento de la red, se escoge una arista al azar y se une el nuevo nodo con uno de los que se encuentran conectados a través de esta arista.

Si llamamos q_k a la probabilidad de que el nodo escogido para unirse al nuevo tenga grado k , entonces

$$q_k = \frac{k p_k}{\langle k \rangle}, \tag{25}$$

donde $\langle k \rangle$ en el denominador se obtiene de la normalización [2].

En este modelo, q_k representa la probabilidad de que el nuevo nodo se una a otro, es decir, la preferencia de unión. Al igual que en el modelo de Barabási-Albert, cuanto mayor sea el grado del nodo escogido al azar, mayor será la probabilidad de unión. Por otro lado, ambos algoritmos muestran una relación lineal con k , base de la preferencia de unión.

Ejemplo 4.2

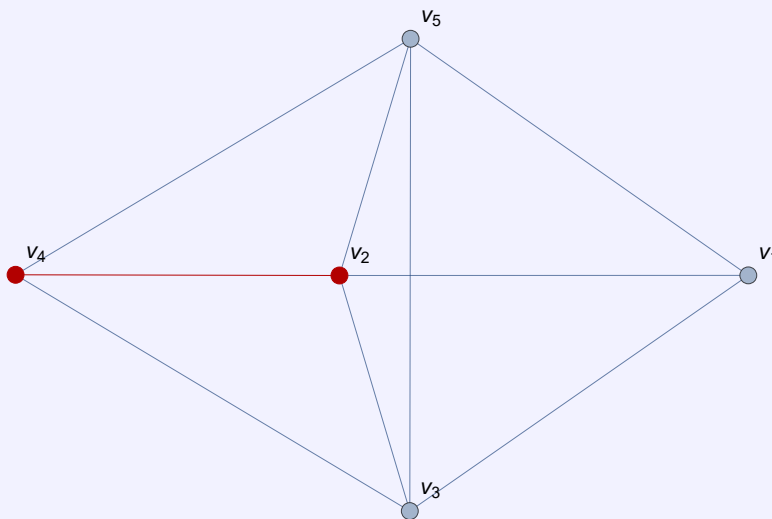


Figura 7: El nuevo vértice se añadirá a la red mediante la unión a uno de los vértices conectados a través de la arista remarcada en rojo.

Sea el grafo de la Figura 7. Se selecciona al azar una arista, por ejemplo, la que une los nodos v_2 y v_4 . Haciendo uso de la ecuación (25), sabiendo $k_2 = 4$ y $k_4 = 3$

$$q_4 = \frac{4p_4}{18/5} \quad \text{y} \quad q_3 = \frac{3p_3}{18/5}.$$

Como hay un mayor número de nodos con $k = 4$ que $k = 3$, podemos considerar $p_4 > p_3$, resultando

$$q_4 > q_3.$$

Luego, es más probable que el nuevo nodo se una al nodo $v_2 = 2$.

Modelo de copiado

Una vez que aparece el nuevo nodo en la red, se siguen dos pasos:

1. *Conexión aleatoria.* El nuevo nodo tiene una probabilidad p de unirse a un nodo de la red, que llamaremos u .
2. *Copiado.* Tomando probabilidad $(1 - p)$, elegimos uno de los nodos conectados a u al azar y se une al nuevo nodo.

Se denomina modelo de copiado porque el nuevo nodo que se introduce en la red tiende a copiar a u para unirse a alguno de los nodos con los que está conectado. Si se consideran las relaciones humanas, cuantos más amigos tiene una persona, mayor es la probabilidad de que se conozca a nueva gente a través de su red de amigos (en [2] se explica este modelo llegando a decir “sin amigos, es difícil hacer amigos”).

La probabilidad de que el nuevo nodo conecte con uno de grado k , queda

$$\Pi(k) = \frac{p}{N} + \frac{1-p}{2L}k,$$

que resulta de combinar la probabilidad de la unión en los dos pasos del modelo [2]. Volviendo a comparar con la ecuación (19) se observa de nuevo la dependencia en k , ofreciendo la preferencia de unión.

4.4. Mejora de modelos

El modelo de Barabási-Albert recrea aspectos básicos del comportamiento de las redes reales, pero al comprobar la dinámica de grado de los nodos, se mencionó que la evolución de k es la misma para todos los nodos y dependía del tiempo que llevaran en la red. Si el nodo a estudiar estaba en la red desde el inicio, su grado sería mayor. De esta forma, los más antiguos deberían de ser los *hubs*. Pero en las redes reales se observa que no siempre ocurre.

Ejemplo 4.3

Si se estudia la red de las redes sociales, Instagram nació en 2010, 4 años después de Twitter. En abril de 2019, la primera tenía aproximadamente mil millones de miembros, mientras que Twitter rondaba los 330 millones^a. La diferencia entre ambas es muy grande y no sigue las expectativas del modelo *BA*.

^aFuente: Statista

En esta sección se van a plantear ajustes para poder afinar el parecido con las redes reales.

4.4.1. Modelo de Bianconi-Barabási

Aparece entonces una característica para determinar la preferencia por un nodo u otro. Se trata de una propiedad interna de cada nodo: la aptitud o *fitness*.

La existencia de distintas capacidades en cada nodo para poder despuntar por encima del resto, aún cuando ha surgido más tarde, hace que haya que introducir un nuevo parámetro en la ecuación (19) para tenerlas en cuenta. G. Bianconi y AL. Barabási lo denominan coeficiente de aptitud (η) [9]. El modelo se basa en el propuesto por AL. Barabási y R. Albert, pero incluyendo esta nueva variable que tiene un fuerte impacto en la preferencia de unión.

- **Crecimiento.** Se introduce un nuevo nodo en la red con m aristas y un coeficiente de aptitud η .
- **Preferencia de unión.** El cambio de la probabilidad traduce la ecuación (19) a

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}, \quad (26)$$

donde ahora aparece el coeficiente de aptitud del nodo i y del resto de los de la red. La dependencia en las dos variables indica que, aunque un nodo de la red se encuentre poco conectado en las etapas iniciales, si su parámetro de aptitud es alto puede convertirse en un *hub*.

Si ahora tomamos la ecuación (20), de forma que aplicamos una dependencia a la tasa de aptitud de cada nodo

$$k(t, t_i, \eta_i) = m \left(\frac{t}{t_i} \right)^{\beta(\eta_i)}, \quad (27)$$

se puede llegar a encontrar la relación entre los valores de los coeficientes dinámico y de aptitud

$$\beta(\eta) = \frac{\eta}{\xi} \quad \text{con} \quad \xi = \int \rho(\eta) \frac{\eta}{1 - \beta(\eta)} d\eta. \quad (28)$$

La demostración de esta ecuación se puede encontrar en [2].

De acuerdo a esta ecuación, cada nodo tendrá una evolución distinta de su grado en función de su coeficiente de aptitud.

También se puede obtener la distribución de probabilidad de grado como

$$p_k \approx \xi \int d\eta \frac{\rho(\eta)}{\eta} \left(\frac{m}{k} \right)^{\frac{\xi}{\eta} + 1}. \quad (29)$$

De nuevo, se dirige al lector a [2] para la demostración de la ecuación.

Redes reales

El valor del coeficiente de aptitud que se aplica a cada nodo se establece como una forma de comparar el despunte de este sobre el resto, dando una visión global de la red. Al principio de la sección se trataba sobre redes sociales, donde Instagram generaba mayor despunte frente a otras. Pero esto no quiere decir que sea la red favorita para todos, pues habrá quienes asignen un valor para este coeficiente menor que a otras. En definitiva, se puede tomar la aptitud como una medida de la preferencia de un nodo sobre otro, atendiendo a los datos.

Este coeficiente cobrará gran importancia en las primeras etapas de un nodo. Además, si un nodo aparece en la red con un cierto coeficiente de aptitud η , conocido el valor de éste podremos augurar su crecimiento.

Para conocer el valor de η se hace un seguimiento del crecimiento que tenga el nodo durante los primeros instantes en la red. Tomando logaritmos en la ecuación (27),

$$\ln k(t, t_i, \eta_i) = \beta(\eta_i) \ln t + B_i, \quad \text{con} \quad B_i = \ln \frac{m}{t_i^{\beta(\eta_i)}}. \quad (30)$$

Como además, $\beta(\eta) = \frac{\eta}{\xi}$, encontramos la relación que buscábamos. Tomando un considerable número de nodos, podemos encontrar la función de distribución $\rho(\eta)$ y estudiar su comportamiento.

4.4.2. Atractivo inicial

Partiendo de las ecuaciones (19) o (26), un nodo de la red que se encuentre aislado no podrá ser conectado por ningún otro nodo ($k_i = 0 \rightarrow \Pi(k_i) = 0$). Para remediar este problema que no tiene por qué darse en redes reales, se incluye un nuevo coeficiente A , que representa la probabilidad de que el nodo se una con otro por primera vez, de forma que la preferencia de unión cambia a

$$\Pi(k) \approx A + k. \quad (31)$$

Nótese que, debido a la dependencia $k \sim t^\beta$ (ecuación (27)), para t suficientemente grande, el coeficiente de atractivo inicial es despreciable.

Se puede relacionar el coeficiente de atractivo inicial con el exponente de grado [10] como

$$\gamma \sim \frac{A}{m}.$$

Esta ecuación muestra cómo el atractivo inicial aumenta significativamente p_k , haciendo que la red se vuelva más homogénea. Esto se debe a que A prevalece en la ecuación (31) si el grado del nodo es pequeño, favoreciendo la unión del nuevo nodo a este y resultando en un reparto más equilibrado de nuevas uniones.

4.4.3. Aristas internas

En el modelo BA, solo se considera que la generación de aristas se produce cuando entra un nuevo nodo a la red. En la mayoría de redes reales

esto no ocurre así, sino que se suelen establecer aristas entre nodos que se encuentran ya en la red. Las medidas experimentales proponen que la probabilidad de formar nuevas aristas en el interior de la red, es

$$\Pi(k, k') \approx (A + Bk)(A + Bk'), \quad (32)$$

donde k y k' representan el grado de los nodos que se conectan [11].

4.4.4. Eliminación de nodos

La evolución de una red también implica que el número de nodos varíe no solo al añadir nuevos, sino también al eliminar los que ya la forman.

Se considera un coeficiente que indique el número de nodos que desaparecen de la red según avance el tiempo, r . Una de las conclusiones [12] a las que se llega utilizando este modelo (sumando las mejoras aportadas anteriormente), es la existencia de una transición de fase en función del parámetro del atractivo inicial A y del número de nodos que se eliminan de la red

$$A = \frac{m(1 - r)}{r(1 + r)}, \quad (33)$$

donde se considera un solo nodo añadido en cada ciclo del proceso.

Tomando $r^*(A)$ como el valor para el cuál, no se producen cambios en el tamaño de la red, es decir, si $r = r^*(A)$, se compensa la aparición de nuevos nodos con la eliminación de los ya existentes. Tendremos entonces tres posibles estados en función de r :

- **Escala libre** ($r < r^*(A)$). Se eliminan menos nodos de los que se crean y el tamaño de la red aumenta, estableciendo un régimen de escala libre.
- **Exponencial estirada** ($r = r^*(A)$).⁷ El tamaño de la red no varía.
- **Exponencial** ($r > r^*(A)$). El tamaño de la red disminuye.

⁷https://en.wikipedia.org/wiki/Stretched_exponential_function

4.4.5. Crecimiento acelerado

El modelo principal implicaba que los nuevos nodos que llegaban a la red tenían el mismo grado inicial, es decir, el mismo número de aristas iniciales. Un estudio de la evolución de la WWW determinó que el número de aristas de cada nuevo nodo que se añade, m , sigue una dependencia temporal

$$m(t) = m_0 t^\theta, \quad (34)$$

con $m(t = 0) = m_0$ [13] y θ el **exponente de crecimiento acelerado**.

De esta forma, el número de aristas que presenta cada nuevo nodo en la red aumenta con el tiempo a través de θ .

4.4.6. Envejecimiento

Al contrario que ocurre con el crecimiento acelerado, también puede ser que los nodos que ya se encuentran en la red dejen de admitir nuevas uniones. Es un límite que se introduce en los nodos más antiguos de la red, basándose en redes reales (por ejemplo, un actor de mayor edad tiende a disminuir el número de obras en las que aparece debido a sus limitaciones físicas, de ahí que esta corrección se denomine envejecimiento).

Para describir matemáticamente este modelo, se define el *factor de envejecimiento*, φ , de forma que la probabilidad de unión del nuevo nodo con otro dentro de la red queda como

$$\Pi(k_i, t - t_i) \sim k(t - t_i)^{-\varphi} \quad (35)$$

Se pueden destacar dos casos a partir del valor de φ :

- $\varphi < 0$. Se sigue produciendo la unión preferencial, siguiendo la forma de una red de escala libre y, en el límite $\varphi \rightarrow -\infty$, todos los nodos se unen al más antiguo.
- $\varphi > 0$. Predomina la unión con los nodos más recientes y para el caso $\varphi \rightarrow 1$, cada nodo se enlaza con su nodo directamente anterior, perdiendo la propiedad de escala libre. Además, se observa que la red se vuelve homogénea cuando se supera $\varphi > 1$. [14]

5. Correlación de grado

La correlación marca la relación entre nodos según su grado. De esta forma, se estudia si en una red, los nodos que la forman tienden a unirse con aquellos cuyo grado sea similar al suyo o, en cambio, prefieren vínculos con nodos de grado distinto. Se trata entonces si es posible la unión entre nodos grandes o *hubs* (con grado muy alto) con nodos pequeños (grado muy bajo). En función de la relación entre estos tipos de nodos se definen tres tipos de redes:

- **Asortativas.** Son aquellas en las que los nodos de grado similar se unen entre sí. Por ejemplo, los *hubs* tenderían a unirse a otros *hubs*.
- **Disortativas.** Son las redes en las que los nodos de grados similares tienden a rechazarse, existiendo preferencia en la unión entre *hubs* y nodos de grado pequeño.
- **Neutras.** Redes en las que la probabilidad de enlace viene dada por

$$p_{k,k'} = \frac{k k'}{2 L} \quad (36)$$

con k y k' los grados de los nodos escogidos aleatoriamente.

Son redes donde la preferencia se basa en que cualquier nodo, tenga el tamaño que tenga, querrá unirse con un *hub*. Puede parecer una red asortativa, pero no se cumple la definición para nodos de pequeño tamaño, pues en una red neutra no tenderían a unirse entre sí, sino a aquellos con mayor grado.

- **Perfectamente asortativa.** Es el caso límite de las redes asortativas, cada nodo de grado k se une exclusivamente con otro nodo de grado k , lo que hará que evolucionen en redes homogéneas, pues todos los nodos tendrán el mismo grado (ver Sección 1).

5.1. Función de correlación de grado

Una de las formas de estudiar la correlación es utilizar una *función de correlación de grado*, cuya expresión se deduce haciendo uso de la *probabilidad condicionada*, $P(k'|k)$, que es la probabilidad de, partiendo de un nodo de grado k y siguiendo una de sus aristas, encontrar un nodo de grado

k' . Entonces, la función de correlación representaría el grado medio de los vecinos de todos los nodos de grado k :

$$k_{nn}(k) = \sum_{k'} k' P(k'|k). \quad (37)$$

Mediante un mapeo de Internet [15], se obtuvo una aproximación empírica para la función de correlación

$$k_{nn} \sim ak^\mu, \quad (38)$$

con a una constante.

Esta ecuación se puede estudiar en función de los tipos de redes definidas a partir de la correlación de grado:

- **Redes neutras.** Cumplen $\mu = 0$, reduciendo la función de correlación a $k_{nn} = a$, pudiendo definir el valor de a , siendo independiente de k . Al no existir correlación entre k y k' , se tendrá que $k_{nn} \sim k' P(k')$, es decir, que depende de la distribución de grado de k' .
- **Redes asortativas.** Se cumple $\mu > 0$. Si la función de correlación muestra el grado de los vecinos del nodo de grado k y en las redes asortativas los supernodos se tienden a unir con otros supernodos, a mayor valor de k , mayor será el grado de los vecinos de nodos con grandes grados. Igual pasa para nodos más pequeños.
- **Redes disortativas.** Su cumple $\mu < 0$. La función de correlación toma valores más pequeños para k grandes debido a que los supernodos tienden a unirse con nodos de grado menor.

6. Comparación de redes

A lo largo de las secciones anteriores se han estudiado dos conceptos clave de las redes: la distribución de grado y el coeficiente de agrupamiento. La primera de ellas, será de gran importancia al tratar las epidemias, una de las muchas aplicaciones del estudio de la teoría de grafos.

Para recoger toda la información, se muestra en la Tabla (2) el resultado de dichos parámetros para los modelos más conocidos de las redes estudiadas.

En cuanto a su aceptación o no como representaciones fieles de la realidad, ya se ha comentado la importancia de las redes de mundo pequeño en cuanto al coeficiente de agrupamiento. El problema que presentan aparece en la distribución de grado, que descarta la aparición de *hubs*. En cambio, los modelos de escala libre permiten la aparición de estos nodos hiperconectados, ajustando la distribución a la realidad. Por el contrario, pierden información al ajustarse al coeficiente de agrupamiento.

Se puede ejemplificar como si fuera un balancín: cuando las redes de escala libre superan a las de mundo pequeño al encajar en la distribución de grado, estas últimas se elevan al hablar del coeficiente de agrupamiento.

| Tipo de red (algoritmo) | Distribución de grado p_k | Coficiente de agrupamiento |
|--------------------------------|---|--|
| Homogénea | $\delta(k)$ | |
| Aleatoria (Erdős-Rény) | $e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$ | $C_i = \frac{\langle k \rangle}{N}$ |
| Mundo pequeño (Watts-Strogatz) | $e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$ | $C(p) = \frac{3(k-2)}{2(2k-1)}(1-p)^3$ |
| Escala libre (Barabási-Albert) | $\sim 2m^2 k^{-3}$ | $C \sim \frac{(\ln N)^2}{N}$ |

Tabla 2: Recopilación de la distribución de grado y el coeficiente de agrupamiento para distintos tipos de redes.

Parte III

Epidemias

7. Modelización de epidemias

La Real Academia Española define una epidemia como una enfermedad que se propaga durante algún tiempo por un país, acometiendo simultáneamente a gran número de personas. Se preferirá tratar las **epidemias** como la propagación de un agente infeccioso a través de una zona geográfica extensa que se presenta en forma de enfermedad sobre los individuos infectados, teniendo estos, o no, la suficiente capacidad para frenar la propagación y, por tanto, la enfermedad.

Si una epidemia deja de presentarse en una zona localizada concreta y afecta a un área de gran tamaño (comparándolo con la dimensión del planeta), se considera que ha evolucionado a *pandemia*, definida por la Organización Mundial de la Salud como la propagación mundial de una nueva enfermedad.

La enfermedad y su desarrollo tiene una evolución temporal, existiendo varias etapas:

- **Infección.** El agente biológico responsable de la enfermedad se introduce en el organismo anfitrión. En esta etapa también se considera el crecimiento.
- **Incubación.** Tiempo entre que el individuo se infecta y la aparición de síntomas. A lo largo del trabajo se considerarán modelos en los que este periodo de latencia es despreciable.
- **Periodo agudo.** Punto álgido del desarrollo de la enfermedad, donde se producen los síntomas más descriptivos.
- **Periodo de declive.** Comienzo de mejora del individuo afectado, donde los síntomas dejan de agravarse.
- **Periodo de convalecencia.** El infectado comienza a recuperar fuerzas.

La etapa de la enfermedad en la que se encuentre el individuo será clave para poder estudiar el contagio. No en todas ellas el propagador puede contagiar nuevos individuos sanos, característica que depende de cada agente.

A lo largo de este trabajo se dividirá y simplificará la población muestra en tres grupos:

- **Susceptibles (S).** Individuos que se encuentran sanos, sin haberse infectado pero que pueden ser contagiados.
- **Infectados (I).** Población que se encuentra en alguna de las etapas de la enfermedad y que puede infectar a los susceptibles.
- **Recuperados (R).** Individuos que han finalizado el periodo infeccioso y se pueden considerar sanos. Estos individuos presentan inmunidad frente a la enfermedad, es decir, no pueden volver a infectar.

La segmentación de la población se denomina **compartimentalización** y cumple $S + I + R = N$, siendo N total de individuos. Podría estudiarse también el sistema si N fuera variable, tratando la dinámica vital (muertes no relacionadas con la enfermedad y nacimientos). En este trabajo se fijará N como constante, suponiendo que el número de estas muertes y nacimientos se compensan. Para el fallecimiento que se puedan producir fruto de la enfermedad, se introducen en este trabajo dentro de los recuperados.

El papel de las redes en las epidemias se encuentra en el contacto entre individuos. Consideraremos la red real de las relaciones humanas, donde los nodos sean personas y las aristas sean relaciones personales que puedan dar lugar a contagio ⁸. Si dos nodos están unidos y uno de ellos desarrolla la enfermedad, existe la probabilidad de que el otro nodo también acabe infectado. A esto, hay que sumarle una muy compleja trama de relaciones y probabilidades. Considerar que todos los individuos tienen la misma probabilidad de infectarse es lo mismo que pensar cualquiera de los infectados puede estar en contacto con un susceptible, pasando a ser infectado. Esta hipótesis se conoce como **mezcla homogénea** de población.

Para simplificar la notación, se hablará de densidad de susceptibles, infectados o recuperados (ρ_s , ρ_i , ρ_r , respectivamente) representando el número de individuos sobre el total de población N . De esta forma

⁸Estas relaciones pueden variar desde una conversación, contacto físico directo o sexual, contacto indirecto a través de objetos, ...

$$\rho_S(t) = \frac{S(t)}{N}, \quad \rho_I(t) = \frac{I(t)}{N} \quad \text{y} \quad \rho_R(t) = \frac{R(t)}{N}.$$

A lo largo del capítulo se hará referencia a individuos y densidades indistintamente.

Se asumirá que la transmisión del agente infeccioso sigue una ley de acción de masas, estableciendo que la variación de cada compartimento es proporcional a susceptibles, infectados o recuperados según el modelo y tipo a estudiar. En cada modelo se explicará más detalladamente la ley de acción de masas empleada.

Un concepto importante es diferenciar entre infección y contagio, acciones relacionadas que pueden llevar a confusión.

- **Infección.** Un patógeno se introduce en el individuo. Es una acción propia del agente biológico.
- **Contagio.** Si dos nodos (uno infectado y otro susceptible) han mantenido contacto entre sí, el nodo infectado ha podido transmitir el patógeno al individuo susceptible, convirtiéndose en infectado. Es una característica que depende de la topología de la red.

8. Modelos

Para llevar a cabo el estudio de la propagación de epidemias se tratarán tres modelos distintos.

8.1. Modelo SI

La dinámica de este modelo se basa en el paso de los individuos susceptibles a infectados cuando se produce el contagio, pero sin que sea posible el paso de infectados a susceptibles o recuperados. Así, $R(t) = 0 \quad \forall t$.

Se introduce la **tasa de infección**, β , como la probabilidad de que la enfermedad se transmita de un infectado a un individuo susceptible por unidad de tiempo y que consideraremos propia de la enfermedad.

Se supondrá por ahora que cada persona puede estar en contacto con otros $\langle k \rangle$ individuos, recordando que $\langle k \rangle \approx pN$ (ecuación (8)).

Al considerar la hipótesis de mezcla homogénea, el número de individuos sanos conectados a uno infectado es $\langle k \rangle S(t)$, pues el número de contactos que tiene cada nodo es $\langle k \rangle$ y $S(t)$ es el número de susceptibles en el instante t .

Se pueden encontrar entonces una ecuación diferencial para obtener la densidad de infectados utilizando una ley de acción de masas. Como cada infectado tendrá $\langle k \rangle S(t)$ contactos con susceptibles, en total serán $I(t) \langle k \rangle S(t)$ posibles contagios, todo ello multiplicado por el factor β :

$$\frac{d\rho_I(t)}{dt} = \beta \langle k \rangle \rho_I(t) (1 - \rho_I(t)), \quad (39)$$

donde se ha considerado $\rho_S(t) = 1 - \rho_I(t)$, esto es, que la población sana es la que no está infectada. Con la condición inicial $\rho_I(t = 0) = \rho_{I,0}$ y absorbiendo el coeficiente $\beta \langle k \rangle$ como β , considerándolo como la tasa de contagio por unidad de tiempo, resulta

$$\rho_I(t) = \frac{1}{1 + \left(\frac{1}{\rho_{I,0}} - 1 \right) e^{-\beta t}} \quad (40)$$

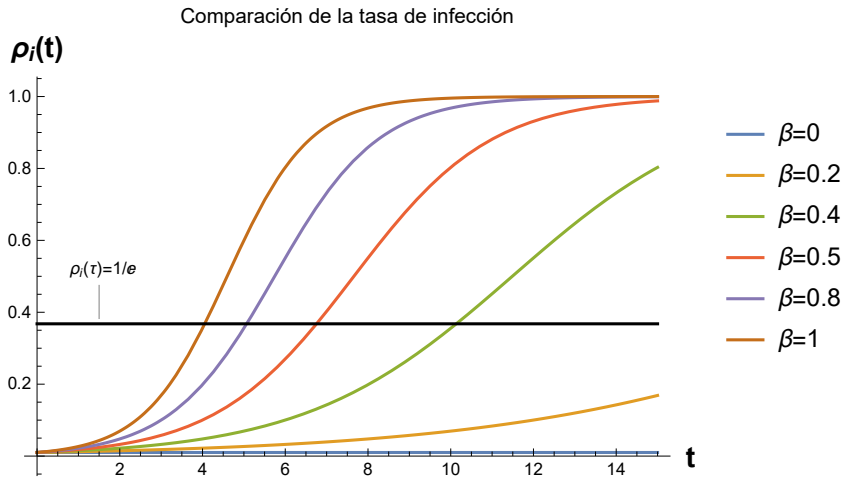


Figura 8: Comparación del número de infectados en función del tiempo para distintos valores de la tasa de contagio β en el modelo SI con $\rho_{I,0} = 1\%$.

Como se aprecia en la Figura 8, existen dos regiones. Al comienzo de la epidemia (tiempos pequeños), el crecimiento de la infección crece exponencialmente. En la segunda región se produce una saturación del número de infectados, llegando siempre al total de la población para tiempos suficientemente grandes.

El **tiempo característico**, τ , se define como el momento en el que la densidad de infectados alcanza el valor $1/e \approx 0,36$. Se utiliza como medida para predecir la duración de la epidemia en el tiempo y permite obtener información sobre el valor de β , resultando un parámetro clave para el estudio de la propagación. En la Figura 8 aparece indicado para cada β y se corresponde con el punto de corte de cada línea a β constante con $\rho_I(\tau) = 1/e$ (línea negra). Para el modelo SI resulta:

$$\tau = \frac{1}{\beta}.$$

Recuérdese que la tasa de contagio comprende la densidad de la red a través del grado medio y el coeficiente de infección. Así, cuanto más contactos existan en la red (aristas) y mejor adaptado esté el patógeno para poder infectar, menor será el tiempo característico y más rápido se producirá el contagio, como se puede ver en las comparaciones hechas para distintos valores de β de la Figura 8.

8.2. Modelo SIS

Se considera de nuevo $R(t) = 0 \forall t$ pero suponiendo que los infectados pueden volver a formar parte de la población susceptible tras un periodo de enfermedad, sin considerar que se ha producido inmunización, es decir, que podrán ser infectados de nuevo.

Aparece la **tasa de recuperación**, μ , representando la cantidad de infectados que dejan de estarlo y pasan a ser susceptibles. La variación de infectados aumenta en función del número de susceptibles y disminuye según la tasa de recuperación de los propios infectados (ley de acción de masas para el modelo SIS), quedando

$$\frac{d\rho_I(t)}{dt} = \beta\rho_I(t)(1 - \rho_I(t)) - \mu\rho_I(t). \quad (41)$$

Con la condición inicial $\rho_I(t = 0) = \rho_{I,0}$, la ecuación presenta dos soluciones

$$\rho_I(t) = \begin{cases} \left(1 - \frac{\mu}{\beta}\right) \frac{C e^{(\beta-\mu)t}}{1 + C e^{(\beta-\mu)t}} & \text{con } C = \frac{\rho_{I,0}}{(1 - \rho_{I,0} - \mu/\beta)} \text{ para } \beta \neq \mu \\ \frac{1}{\beta + \rho_{I,0}^{-1}} & \text{para } \beta = \mu \end{cases}. \quad (42)$$

En este modelo, se pueden observar tres finales posibles que dependen de la relación entre la tasa de contagio y la de recuperación. Para analizar estas tres fases, se considera el cociente de ambas, denominado **tasa efectiva de propagación** (λ) o **número reproductivo básico** (R_0), según la bibliografía que se consulte. En este trabajo se hablará de tasa de propagación, siendo esta

$$\lambda = \frac{\beta}{\mu}.$$

Nótese que para $\lambda \rightarrow \infty$ se recupera el modelo SI, pues para dicho caso $\mu \rightarrow 0$.

En la Figura 9 se representan tres posibles finales en función del valor que tome λ :

- Si $\lambda = 1$, el número de infectados y de población sana es constante para todo tiempo t , manteniéndose las condiciones iniciales.

- Si $\lambda > 1$, la tasa de contagio es mayor, por lo que aumenta el número de infectados. Al contrario que en el modelo SI, no toda la población acabará infectada, pues la posible recuperación de los individuos hace que se alcance un estado estacionario en el número de infectados y sanos. Este estado final se denomina **estado endémico**. En la Figura 10 se reproduce la tasa de infectados y susceptibles para este caso.
- Si $\lambda < 1$, la tasa de recuperación prevalece y la enfermedad tenderá a desaparecer con el tiempo, llegando a $\rho_I(\infty) = 0$.

Al igual que en el modelo anterior, se puede calcular el tiempo específico de la epidemia como

$$\tau = \frac{1}{\mu(\lambda - 1)},$$

donde se pueden observar los tres distintos estados finales que se han comentado. Como en el modelo SI, a mayor tasa de propagación, menor es el tiempo característico y más rápida propagará la epidemia.

Quede claro que, al igual que se puede considerar la tasa de infectados para modelizar la propagación de la epidemia, se puede calcular la tasa de individuos susceptibles como

$$\frac{d\rho_S(t)}{dt} = \mu\rho_S(t) - \beta\rho_S(t)(1 - \rho_S(t))$$

y obtener el número de individuos susceptibles en función del tiempo.

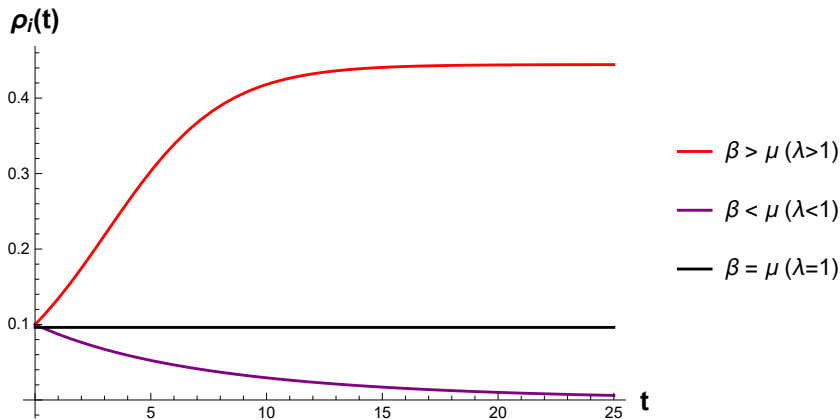


Figura 9: Comparación del número de infectados del modelo SIS en función de λ .

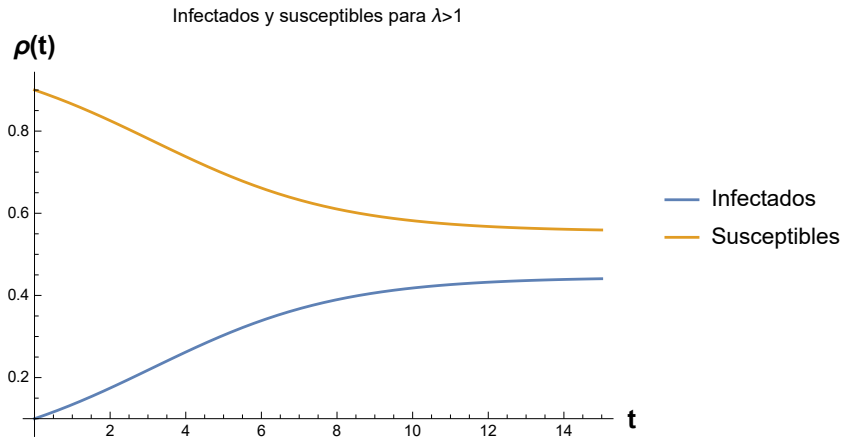


Figura 10: Densidad de infectados y susceptibles para $\lambda > 1$ y $\rho_I(0) = 0,1$.

8.3. Modelo SIR

Se establece la existencia de una población recuperada, ρ_R , que son aquellos que ya han sido infectados y que han superado la enfermedad, sin poder volver a contagiarse, lo que significa que, o bien se han inmunizado o han muerto a causa de la enfermedad. En todo caso, el número de muertos no disminuirá el total de la población total por no estar considerando la dinámica vital.

El sistema de ecuaciones diferenciales que resulta para este modelo se explica por la aplicación de ley de acción de masas en tres puntos:

- La variación del número de infectados aumenta según los individuos susceptibles y los infectados, y disminuye proporcionalmente con los infectados que pasan a estar recuperados.
- Los susceptibles disminuyen según los infectados y los propios susceptibles que se puedan infectar (igual que en el resto de modelos).
- El número de recuperados aumenta en función de los infectados.

De todo ello resulta el siguiente sistema de ecuaciones diferenciales:

$$\begin{aligned}
 \frac{d\rho_I(t)}{dt} &= \beta\rho_S(t)\rho_I(t) - \mu\rho_I(t) \\
 \frac{d\rho_S(t)}{dt} &= -\beta\rho_S(t)\rho_I(t) \\
 \frac{d\rho_R(t)}{dt} &= \mu\rho_I(t)
 \end{aligned}
 \tag{43}$$

siendo las condiciones iniciales $\rho_R(0) = \rho_{R,0}$, $\rho_S(0) = \rho_{S,0}$, $\rho_I(0) = 0$.

Proposición 18. La evolución de la epidemia siguiendo el modelo SIR conlleva la desaparición de los infectados para $t \rightarrow \infty$.

Demostración. Para que el número de recuperados deje de variar, se debe cumplir $\frac{d\rho_R(\infty)}{dt} = 0$. Como $\frac{d\rho_R(\infty)}{dt} = \mu\rho_I(\infty)$, entonces $\mu\rho_I(\infty) = 0$, por lo que $\rho_I(\infty) = 0$. \square

La Proposición 18 se puede observar en la Figura 11, donde se ha simulado una propagación siguiendo el modelo SIR.

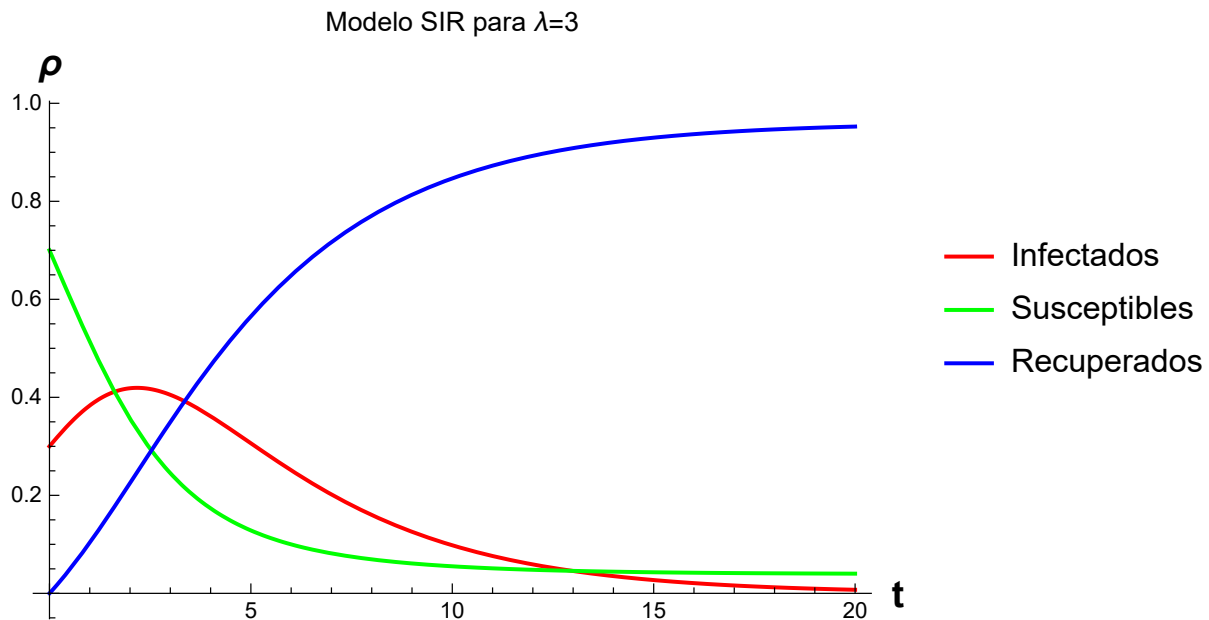


Figura 11: Modelo SIR con $\lambda = 3$ y un porcentaje inicial de infectados del 30%.

9. El papel de las redes

Una vez estudiados distintos tipos de redes complejas y varios modelos de propagación de epidemias, se pueden aunar ambos campos para establecer el análisis de la influencia de la topología de la red en la propagación de un patógeno que siga alguno de los modelos observados.

Cuanto mayor sea el grado del nodo mayor será la probabilidad de que acabe infectado, pues es más posible que uno del resto de nodos con los que está unido se encuentre infectado y lo contagie. Una vez infectado el *hub*, el patógeno se propagará muy rápido hacia el resto de nodos de la red al tener un número de contactos tan alto.

Para considerar la variación de contactos debido a la distribución de grado de la red, se introduce la probabilidad de que un nodo esté conectado a otro ya contagiado, $\Theta(t)$, siguiendo

$$\Theta(t) = \frac{1}{\langle k \rangle} \sum_{k=1}^N k p_k \rho_{I,k}(t), \quad (44)$$

apareciendo la distribución de grado p_k . La demostración de esta expresión se puede encontrar en [16].

Mientras que para el modelo SIS, un nodo infectado puede infectar a otro nodo de grado k , volviéndose después susceptible y pudiendo volver a ser infectado por el nodo de grado k , esto no puede ocurrir en el modelo SIR, pues el primer nodo infectado, al recuperarse, ya no puede volver a contraer la infección, luego el nodo de grado k pasa a poder infectar a $k - 1$ nodos. Esto se reflejará como

$$\Theta(t) = \frac{1}{\langle k \rangle} \sum_{k=1}^N (k - 1) p_k \rho_{I,k}(t). \quad (45)$$

$\Theta(t)$ depende del modelo por la dependencia de la densidad de infectados, debido a la ecuación (44), pues al derivar ambos miembros con respecto del tiempo, se obtiene

$$\frac{d\Theta(t)}{dt} = \sum_{k=1}^N \frac{(k-1)}{\langle k \rangle} p_k \frac{d\rho_i}{dt}. \quad (46)$$

Se puede definir un valor límite que será clave en la propagación o desaparición de una epidemia y que recibe el nombre de **umbral epidémico**, λ_c . El proceso para obtenerlo en cada modelo se puede encontrar completamente explicado en [2] y consiste en sustituir el incremento de infectados en cada modelo dentro de la ecuación (46), resolviendo la ecuación diferencial y relacionando las variables con el tiempo característico (distinto de los que aparecen en las secciones anteriores debido a la dependencia en $\Theta(t)$). Por tanto, será un parámetro que dependa del tipo de red. Las expresiones que resultan se recogen en la Tabla 3.

| Modelo | τ | λ_c |
|--------|---|---|
| SI | $\frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$ | 0 |
| SIS | $\frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}$ | $\frac{\langle k \rangle}{\langle k^2 \rangle}$ |
| SIR | $\frac{\langle k \rangle}{\beta \langle k^2 \rangle - (\mu + \beta) \langle k \rangle}$ | $\frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$ |

Tabla 3: Comparación de los tiempos característicos y los umbrales epidémicos para los modelos estudiados.

Se relaciona entonces la topología de la red con las características del patógeno mediante λ_c y λ , encontrando dos estados posibles.

- Si $\lambda > \lambda_c$ la epidemia se propagará y tenderá a alcanzar el futuro que le corresponda según el modelo y características que se estudien.
- Si $\lambda < \lambda_c$ la epidemia no se propaga.

Redes homogéneas

En las redes homogéneas, todos los nodos tienen el mismo grado, que coincide con el grado medio de la red ($k_i = \langle k \rangle$, con $i = 1, 2, \dots, N$). Por tanto, no se hará uso de $\Theta(t)$. En este caso, la Tabla 3 no sirve y bastarán las ecuaciones introducidas al presentar cada modelo (véanse las ecuaciones (39), (41) y (43)).

El **modelo SI** presenta el mismo final para todo tipo de red, la completa infección de la población, por lo que no existe un umbral epidémico. La variación de infectados para el **modelo SIS** será

$$\frac{d\rho_i}{dt} = \beta \langle k \rangle \rho_i (1 - \rho_i) - \mu \rho_i.$$

Este modelo auguraba dos futuros ($t \rightarrow \infty$): un estado endémico y uno sin infectados, ambos caracterizados por ser estables, es decir, no tener variación de individuos infectados. Matemáticamente supone que $d\rho_i/dt = 0$, obteniendo como resultado

$$\rho_I(\infty) = \begin{cases} 0, & \text{para } \lambda < \lambda_c \\ 1 - \frac{\lambda_c}{\lambda}, & \text{para } \lambda \geq \lambda_c \end{cases},$$

siendo

$$\lambda_c = \frac{1}{\langle k \rangle},$$

permitiendo las dos fases del modelo.

Redes aleatorias

En la ecuación (8) se había determinado el grado medio de una red aleatoria que seguía la distribución binomial. Si el tamaño de la red era muy grande, la función de distribución pasaba a ser la de Poisson. Aunque la obtención de esta magnitud se había obtenido directamente a partir del número de aristas de la red, la demostración formal se basa en calcular

$$\langle k \rangle = \sum_{k=0}^N k p_k,$$

con p_k siguiendo una distribución de Poisson. La demostración de que el resultado coincide con la ecuación (8) se puede encontrar en cualquier manual de estadística. El valor $\langle k \rangle$ toma importancia para poder establecer los umbrales epidémicos en este tipo de redes.

Por otro lado, el momento de orden dos, $\langle k^2 \rangle$, se obtendría como

$$\langle k^2 \rangle = \sum_{k=0}^N k^2 p_k = \langle k \rangle (1 + \langle k \rangle).$$

Ya se ha mencionado que, con el **modelo SI**, el número de infectados siempre acabará siendo el total de la población ($\lambda_c = 0$), pero es relevante conocer la velocidad de propagación, cuyo marcador es el tiempo característico, que queda como

$$\tau = \frac{1}{\beta \langle k \rangle}.$$

En el **modelo SIS**, el umbral epidémico toma un valor de

$$\lambda_c = \frac{1}{\langle k \rangle + 1},$$

resultando la propagación de la epidemia para $\lambda > \lambda_c$ y su extinción para $\lambda < \lambda_c$, consecuencia directa de que en este tipo de redes, el umbral epidémico sea siempre distinto de cero. Otra forma de llegar a esta conclusión se puede encontrar en [16].

Redes de escala libre

En la Sección 4 se introdujeron tres regímenes para este tipo de redes en función del valor del exponente de grado. Para $\gamma < 3$, la red seguía una distribución de escala libre, donde el valor $\langle k^2 \rangle$ divergía para redes de gran tamaño. En cambio, para $\gamma > 3$, la red tomaba la apariencia de una red aleatoria con $\langle k^2 \rangle$ finito. El comportamiento de estas últimas se puede comparar con lo explicado en el apartado anterior, por lo que el análisis se centrará en aquellas que siguen un comportamiento de escala libre.

Debido a la dependencia con $1/\langle k^2 \rangle$ del tiempo característico (τ) en los **modelos SI y SIS**, éste tiende a cero. El significado de $\tau \rightarrow 0$ no es otro que la propagación inmediata de la infección en cuanto aparece en la red, sea cual sea su tasa efectiva de propagación.

El hecho de que la propagación se produzca siempre está condicionado por la existencia de los *hubs*. Incluso el patógeno con menor posibilidad de contagio, debido a su baja tasa efectiva, si infecta a uno de los supernodos, podrá propagarse por la red. El desenlace final (tiempos largos), dependerá ya de λ , pero obligará a la existencia de algún infectado en cualquiera de los dos modelos.

En el modelo SIS se había hablado de un posible estado endémico, condicionado por el coeficiente λ . A continuación, se desarrolla el número de infectados final para una población que siga una red de escala libre:

$$\rho_I(\infty) = \frac{2}{e^{\frac{1}{m\lambda}} - 1} \left(1 - \frac{1}{m\lambda \left(e^{\frac{1}{m\lambda}} - 1 \right)} \right). \quad (47)$$

Para la demostración ver [16].

El estado final del modelo SIR conlleva la desaparición de los infectados y una población repartida entre recuperados y susceptibles. Dicho estado se corresponde con

$$\begin{cases} \rho_S = \sum_{k=1}^N p_k e^{-\alpha k \phi_\infty} \\ \rho_R = \sum_{k=1}^N p_k (1 - e^{-\alpha k \phi_\infty}) \end{cases},$$

con

$$\phi_\infty = \frac{1}{\mu} - \frac{1}{\mu \langle k \rangle} \sum_{k=1}^N k p_k e^{-\alpha k \phi_\infty}.$$

La demostración se puede encontrar en [16].

10. La dificultad de frenar una epidemia

La modelización de epidemias permite conocer cómo se van a desarrollar y propagar en función de los parámetros que se vayan conociendo según progresen. Ante situaciones de alarma sanitaria, la importancia de erradicar la enfermedad es vital. A lo largo la historia de la humanidad se han sucedido numerosas epidemias, algunas de ellas llevando consigo consecuencias demográficas, políticas y sociales muy destacables (como la *Peste Negra*, a mediados del siglo XIV). En la Figura 12 se muestra el número de fallecimientos de algunas epidemias surgidas en la historia.⁹

La transmisión del patógeno causante de la enfermedad varía en función de la época. Así, la *Peste Negra*, surgida en Asia Menor (aunque no se puede afirmar con seguridad), se propagó a través de la Europa Medieval durante unos 5 años. Sin embargó, la enfermedad por *COVID-19* surgida en China en diciembre de 2019 llegó a Italia y España entre enero y febrero de 2020. Los puentes aéreos entre ambos continentes promovieron el avance rápido de la enfermedad, al igual que lo hicieron las rutas comerciales en el siglo XIV de manera más lenta. Los viajes aéreos o las rutas de comercio se pueden estudiar como redes cuyos nodos son aeropuertos o puntos clave comerciales y las aristas el camino entre ellos, por lo que se puede modelizar la propagación [17].

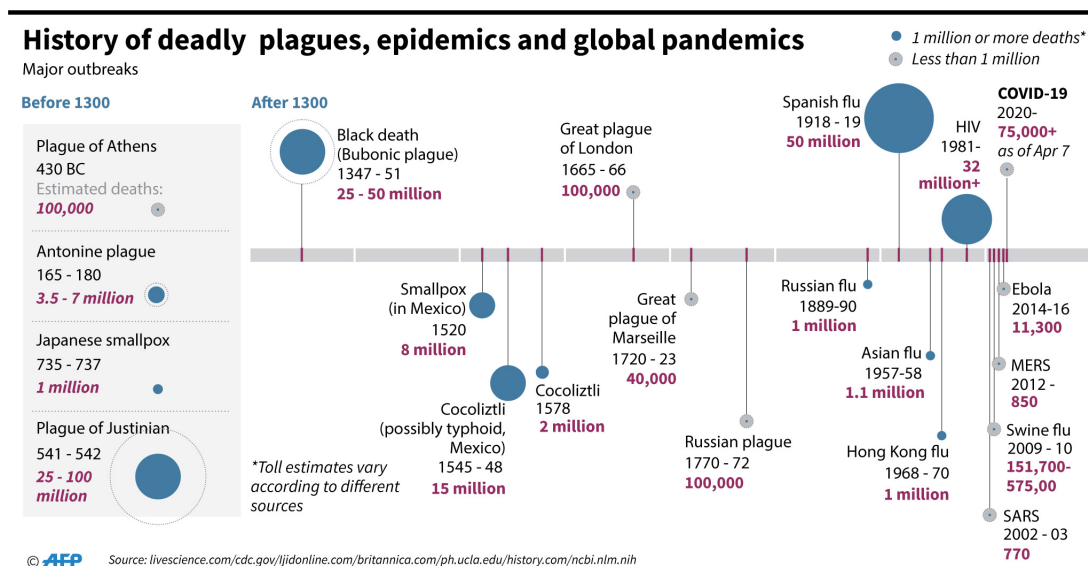


Figura 12: Número de muertes por las principales epidemias a lo largo de la historia. Agence France-Presse (AFP).

⁹El número de muertes por COVID-19 se sitúa entorno a los 3 millones a 27 de abril de 2020.

Durante la crisis del *COVID-19* numerosos países decretaron el confinamiento obligatorio de los ciudadanos para evitar el contagio masivo, además de restringir el tráfico aéreo o a motor y medidas de higiene extremas para evitar la infección. El cumplimiento de estas medidas busca evitar la saturación de los recursos sanitarios: al disminuir el número de infectados, por la ley de acción de masas, la tasa de infección disminuirá con el tiempo, suavizando todo lo posible el pico de la curva de infección (*#FrenaLaCurva*), ver Figura 11. Métodos como la predicción en tiempo real o las simulaciones generadas por ordenador, permitieron a los gobiernos tomar medidas con antelación según se iban desarrollando las distintas fases de la epidemia.

Inmunización

La topología de la red no solo sirve para estudiar los casos en los que la epidemia se propaga según el tipo de patógeno (λ), sino que también se pueden emplear para intentar erradicar la enfermedad. La primera idea sería eliminar la posibilidad de que cualquier individuo se volviera infectado, lo que supondría una inmunización total de la población, algo que se escapa de la realidad. Desde este límite, se declaran una serie de propuestas para volver inmune a la población:

- **Inmunización aleatoria.** Se selecciona al azar un grupo de individuos que quedarán inmunizados (g), disminuyendo el número de susceptibles hasta $(1 - g)$. La tasa de propagación cambiará entonces a $\lambda' = \lambda(1 - g)$. Para redes aleatorias se puede escoger un valor de g suficiente para que la epidemia desaparezca. En el caso de redes de escala libre este método no resulta eficiente [2].
- **Inmunización de *hubs*.** Si en una red de tipo escala libre se inmuniza a los supernodos, se eliminará una gran fuente de contagio, pues son los más propensos en resultar infectados y los que más contagios van a producir. De esta forma, los *hubs* serán invisibles para el patógeno, quien verá una topología de la red totalmente distinta, apareciendo pequeñas islas de nodos (véase Figura 13). El problema de este tipo de inmunización está en la dificultad de encontrar los *hubs* en las redes de contagio.

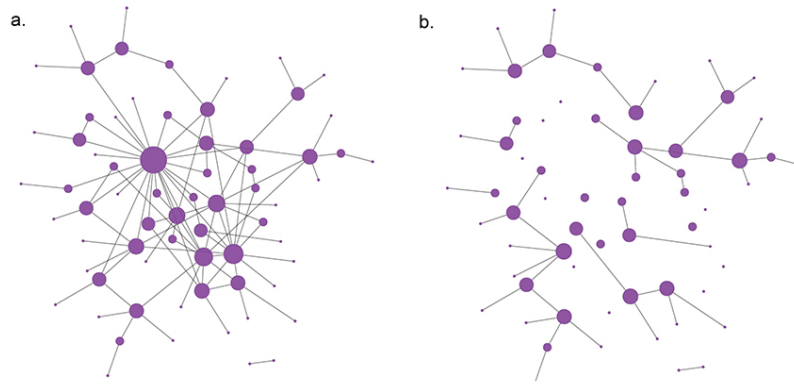


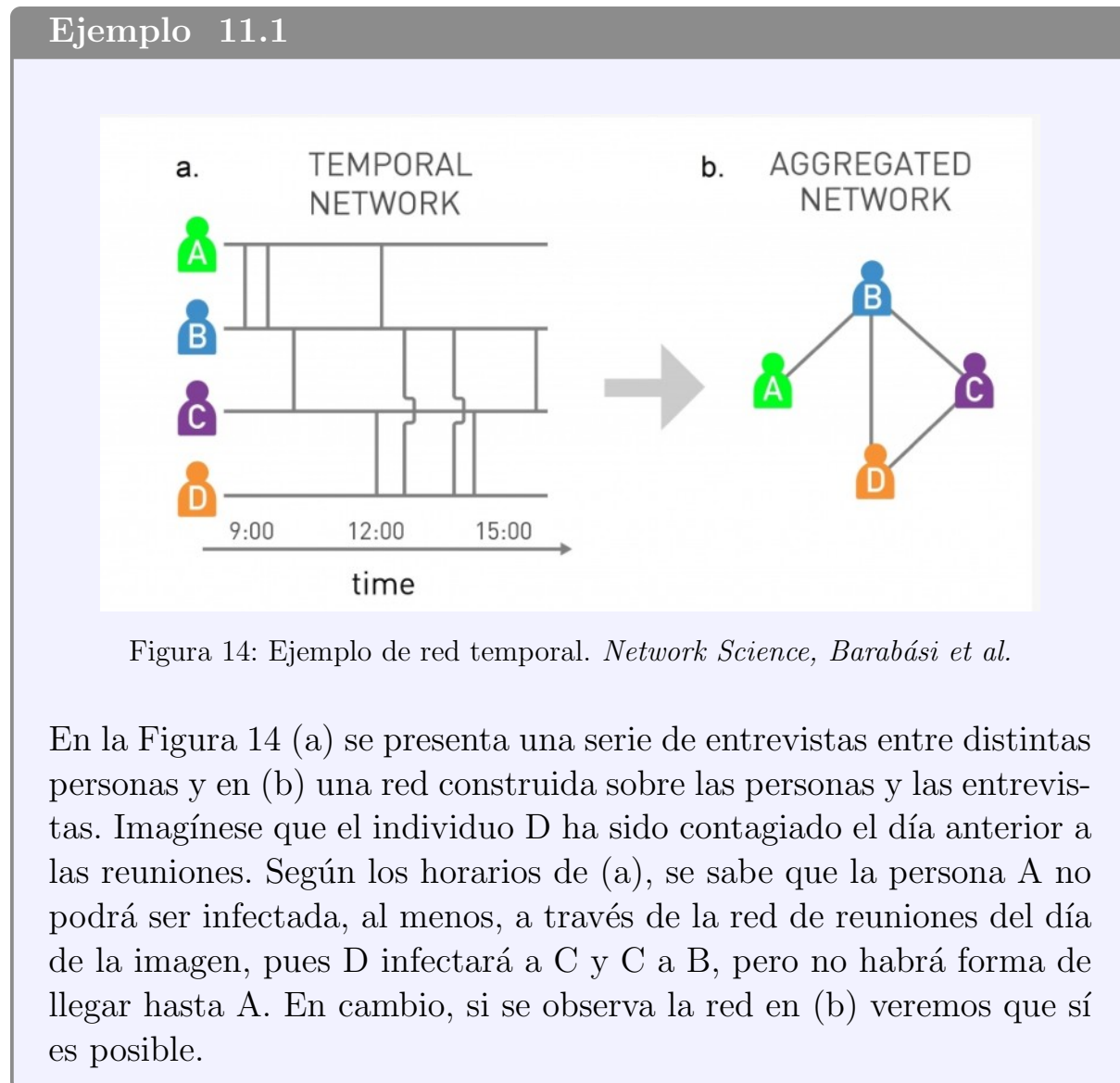
Figura 13: En (a) se presenta una red de escala libre. En (b), la misma red pero sin los *hubs*, que han sido inmunizados. De esta forma, el patógeno solo podrá ver conjuntos pequeños de nodos en lugar de la red entera. Imagen extraída de [2].

- Inmunización conocida.** Se selecciona aleatoriamente un conjunto de individuos y de entre sus nodos vecinos se inmunizan algunos al azar. De esta forma, es muy probable que alguno de estos vecinos sean *hubs* que queden inmunizados. Este método resulta muy eficiente en ambos tipos de redes.

En lo relacionado con el tipo de modelos estudiados, la inmunización para el modelo SI o SIS no tiene interés, pues no existen nodos que se puedan recuperar.

11. Redes temporales

En general, la red de la interacción social suele presentar cambios en función del tiempo. En un conjunto real de nodos, las uniones entre ellos no son constantes, sino que aparecen y desaparecen.



Además del orden temporal de los contactos entre nodos, se destacan las *ráfagas*, periodos de actividad muy alta con tiempos entre contactos de longitud variable. Al considerar una red estática, se puede suponer que el contacto se mantiene durante todo el estudio de la red. Pero si se analiza una red temporal, existe un ritmo en las conexiones, es decir, una frecuencia y duración. Si se parte relaciones aleatorias, la actividad seguirá una distribución de Poisson. Pero se ha observado que las relaciones personales

tienden a seguir una ley de potencias [18], mostrando periodos de no actividad largos mezclados con otros más cortos y duraciones generalmente más largas. Ambos tipos de relaciones (aleatorias y reales) se muestran en la Figura 15.

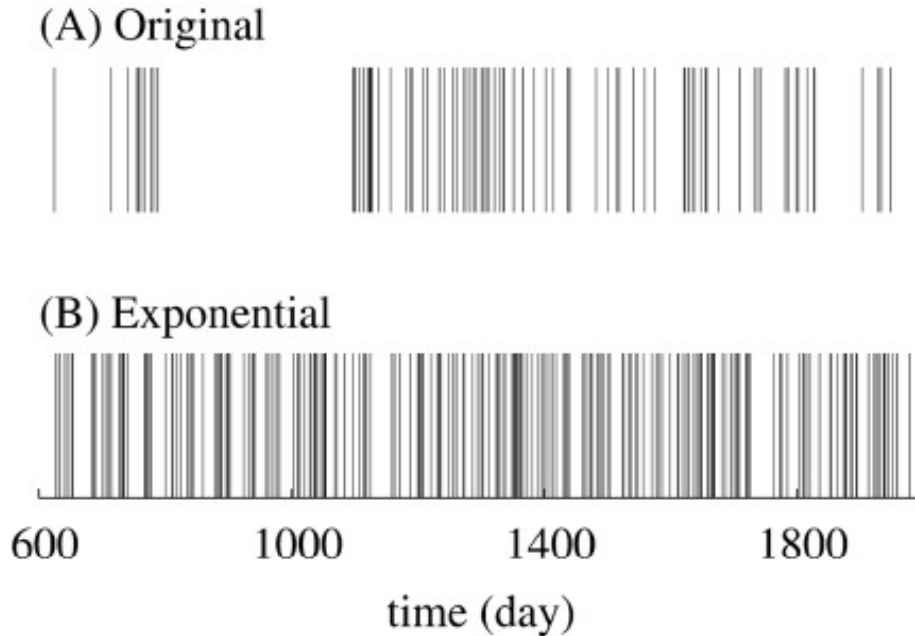


Figura 15: Ejemplo de distribución de ráfagas: ley de potencias en (a) y Poisson en (b). En ambas aparece el mismo número de contactos. Imagen extraída de [19].

Al estudiar únicamente la topología de la red no se considera el tiempo, sino si existe o ha existido relación entre nodos. Esto hará que se sobreestime [20] o subestime [21] la velocidad de la propagación.

Otros parámetros que afectan a la propagación

Una de las características sobre redes que no se ha tratado a lo largo del trabajo ha sido la tendencia de los nodos a agruparse en **comunidades**. Estas se caracterizan por una gran cantidad de nodos comunicados entre sí dentro de ellas pero con escasas uniones exteriores. Ligado al concepto de comunidad aparece el **peso** de la unión ($w_{i,j}$), un factor que representa la frecuencia de conexión entre nodos unidos por una arista. De esta forma, el contagio no sólo será función del factor β , sino que la probabilidad de contagio será $p_{i,j} \sim \beta w_{i,j}$.

En el fenómeno social, se considera que gran parte de los miembros de una comunidad se encuentran unidos entre sí, diferenciando dos tipos de aristas (llamados lazos): fuertes y débiles. Si el individuo A conoce a B y C, establece un lazo fuerte con ellos. Aunque B y C no se conozcan, tendrán posibilidad de compartir información a través de A, lo que establece un lazo débil.

Por otro lado, es necesario estudiar el tiempo para que se produzca el contagio entre nodos, que puede variar en función de cada patógeno y que se relaciona directamente con el tiempo que se mantiene contacto entre un individuo infectado y otro susceptible, que deriva de las redes temporales.

12. Desarrollos futuros

Durante el desarrollo del trabajo, se ha comenzado a estudiar un modelo distinto. Se propone la existencia de un nuevo factor, llamado de **coeficiente de desplazamiento**, γ , cuando se estudia la propagación de una epidemia existiendo distintas poblaciones.

Se estudiará en función del modelo SIS para comenzar el desarrollo de este modelo, suponiendo varias comunidades de individuos con distintas configuraciones de red en las que una fracción de población de cada una se desplaza a otras de las comunidades a estudio. Este modelo puede servir de guía para modelizar propagaciones reales entre ciudades distintas que tengan una población que realice trayectos entre ellas. La población desplazada puede ser de ambos tipos (susceptible o infectada).

Por el momento, se ha recreado el nuevo modelo para dos comunidades sin considerar aún su topología, que siguen el siguiente sistema de ecuaciones diferenciales:

$$\begin{cases} \frac{dI_1}{dt} = \beta I_1 S_1 - \mu I_1 - \gamma_1 I_1 + \gamma_2 I_2 \\ \frac{dS_1}{dt} = -\beta I_1 S_1 + \mu I_1 - \gamma_1 S_1 + \gamma_2 S_2 \end{cases} \quad (48)$$

$$\begin{cases} \frac{dI_2}{dt} = \beta I_2 S_2 - \mu I_2 + 1 - \gamma_2 I_2 + \gamma_1 I_1 \\ \frac{dS_2}{dt} = -\beta I_2 S_2 + \mu I_2 - 1 - \gamma_2 S_2 + \gamma_1 S_1 \end{cases} \quad (49)$$

Comparando este modelo con el SIS para una única comunidad, se ha observado que hay una variación de resultados, lo que permite pensar que el nuevo modelo puede servir para un verdadero estudio de la propagación con desplazamiento.

Al encontrarse este nuevo modelo en desarrollo, se anima al lector a mantenerse informado de los avances que pudieran darse relacionado con él.

Parte IV

Conclusiones/Conclusions

El estudio de las redes complejas que se ha presentado a lo largo del trabajo, nos ha permitido tratar varias características de estas:

- **Topología de la red.** En la Parte II ha sido estudiada la estructura de los distintos tipos de redes complejas: aleatoria, de mundo pequeño y de escala libre. Los modelos de construcción que permiten la existencia de estas redes, nos han acercado a estudiar sus principales características clave, como son: distancia, agrupamiento, y, sobre todo, la distribución de grado. La diferencia entre estos parámetros, se resumieron ya en la Sección 6, que puede servir como conclusión para la topología.

Si el lector prefiere no regresar a la mencionada sección, se recuerdan brevemente los resultados que se mostraban:

- Las redes homogéneas no suelen ser una representación cercana de redes reales.
 - El modelo de Erdős-Rényi, que se ha tratado como constructor de redes aleatorias, era también bastante limitado en cuanto a su proyección sobre la realidad. Su distribución de grado (de Poisson) no permitía la existencia de nodos hiperconectados.
 - Las redes de mundo pequeño, generadas por el algoritmo de Watts-Strogatz, se acercan a la representación de agrupamiento propia de redes reales.
 - El método de Barabási-Albert para generar las redes de escala libre, donde aparece la preferencia de unión, permite acercar la distribución de grado de las redes reales.
- **Modelos epidemiológicos.** La compartimentalización de la sociedad en susceptibles, infectados y recuperados nos permite estudiar el comportamiento y desarrollo de la propagación de una enfermedad. Esta no es siempre la misma:
 - El modelo SI predice el contagio total de la población susceptible y es el que seguiría el virus del SIDA.

- El modelo SIS, muestra dos escenarios finales: la erradicación de la epidemia o un estado endémico (parte de la población queda infectada y otra susceptible), todo ello en función de la relación que exista entre la recuperación de infectados y su paso a susceptibles, y la infección de susceptibles. Este modelo suele ser válido para algunas enfermedades de transmisión sexual, ETS, que permiten volverse a contagiar a pesar de haberla sufrido.
- Por último, el modelo SIR refleja la aparición del tercer estamento, una población recuperada que aparece tras haber pasado la enfermedad y que deja de ser contagiosa. Este modelo describe enfermedades como el sarampión, que presentan inmunidad una vez se han producido.

En la Figura 16 se presenta un esquema con los modelos que se han tratado en este trabajo. Es necesario remarcar la existencia de muchos otros modelos en los que se tienen en cuenta distintos parámetros y relaciones entre los compartimentos, que pueden ser distintos a los aquí mencionados.

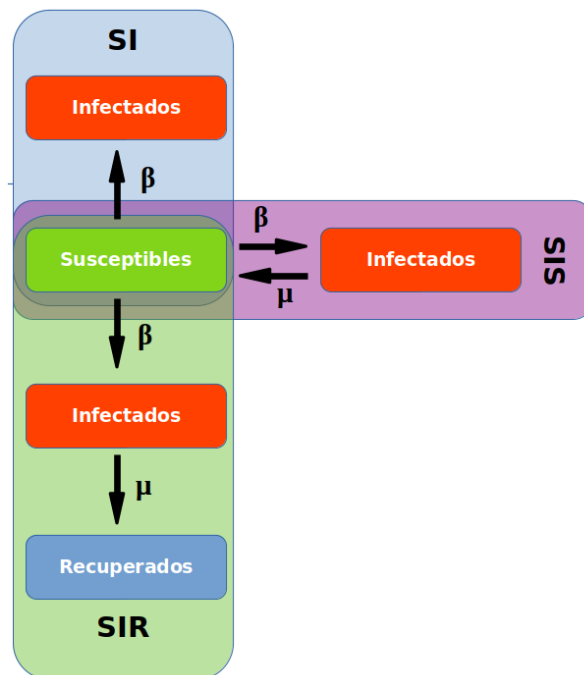


Figura 16: Esquema de los modelos SI, SIS y SIR.

- **Redes complejas y epidemias.** El estudio de los modelos de propagación de epidemias puede resultar, a primera vista, un elemento

alejado de la teoría de grafos. En la Sección 9 se ha podido estudiar qué relación existe entre ambos campos. El desarrollo de esta penetración bien podría ser considerado para futuros trabajos, pues resulta un terreno muy amplio. Cada modelo reacciona de una forma distinta según el tipo de red. Las más interesantes a estudiar por su relación con una enfermedad humana son las de escala libre.

El conocimiento de las redes no resulta útil únicamente para estudiar la propagación de la epidemia sino también para combatirla, donde entra en juego la inmunización de la población, como se ha mostrado en la Sección 10.

Una idea que debe de quedar marcada en el lector al finalizar la lectura de este trabajo es que se ha tratado continuamente sobre propagación de epidemias identificando a individuos con personas. El motivo de esta selección es poder aplicar los modelos de propagación a las redes sociales personales que se desarrollan entre seres humanos. Aún así, estas epidemias se pueden aplicar a otros entes, como pueden ser virus informáticos (tanto en ordenadores como en teléfonos móviles), cadenas de mensajes de *spam* enviados a través de plataformas sociales virtuales, etc. Más allá de las epidemias, también se pueden tratar la propagación y evolución de pequeñas ideas realizadas por un número limitado de individuos que, por contacto, se van desarrollando hasta completar nuevas teorías.

Por último, mencionar que el desarrollo del trabajo en torno a la distribución de grado, p_k , permite estudiar la propagación entendiendo las redes como un continuo, sin buscar la individualización de los nodos. En cambio, si el tamaño de la red fuera pequeño, el estudio de la propagación se podría discretizar, permitiendo el análisis de manera directa de los contagios. Esto ocurre, por ejemplo, con los llamados “equipos de rastreo”: cuando aparece un individuo contagiado, se observan contactos con otros individuos que ha podido mantener en los últimos días. De este rastreo, aparecerán posibles susceptibles de haber sido contagiados, además de un paciente anterior que infectara al caso detectado. De esta forma, se puede aislar el brote epidémico para evitar la propagación a lo largo del entramado completo de la red, así como encontrar al *paciente cero*.

Conclusions

We can summarize some complex networks' features which has been study along the project:

- **Networks' topology.** In Part II we have made a study about the structure of different complex networks: random, small world and scale-free networks. The construction models that allow the existence of these networks have been used to study some key features as: distance, clustering and the degree distribution, especially the last one. The difference between this parameters was still summarized at Section 6, that could be as a conclusion for this item.

If the reader doesn't want to go back to this section, we can remember briefly the results. On one hand we got heterogeneous networks and random networks, that are not a good approximation to real networks. The first one presents a beauty geometry and the Erdős-Rényi algorithm, that has been studied to construct the second one, fails at the degree distribution (Poisson distribution) that doesn't allow the existence of hubs. On the other hand, we found small world networks (generated by the Watts-Strogatz algorithm) and scale-free networks, which its construction is based on the Barabasi-Albert model. These have a good approximation to reality. Random networks represent fairly good clustering; and the scale-free networks, the degree distribution.

- **Epidemiological models.** The compartmentalization of society in susceptible, infected or recovered allows us to study an epidemic spread's evolution. This is not always the same: the SI model predicts the total infection of the population, like what AIDS virus will do if we consider homogeneous mixture (i.e., if the contagion will be between each connected individual). SIS model shows us two possible final scenarios: the removal of the epidemic or an endemic state, in which a fraction of the population stays infected and, the other, susceptible. These futures depend on the relation between the recovery coefficient and the infection rate. The last model, the SIR, shows the third kind of population: the recovered, that appears once the infected stop spread. At Figure 16, the reader can see a scheme of these models. It's important to say that there are more models than the presented here, with

different parameters and population's compartments.

- **Complex networks and epidemics.** The reader can say that there is no relation between these areas at a first look. In Section 9 we can study this relation. Each model has a different reaction depending on the complex network that we are treating. The most interesting because of their relation with human disease are scale free.

The knowledge about networks is not only applicable to the epidemic spread, but we can use it to combat it thanks to immunization, as shown in Section 10.

It is interesting to say that along this project we have used the epidemic spread in humans. This is because we can implement personal networks between them. But we can also consider studying networks and spread in malware or spam, in example. In addition to epidemics, we can treat the spread of other items, like the evolution of little ideas made by individuals that are increased by other people and finalized in new theories.

Lastly, to mention that having focused on degree distribution, p_k , allows us to study the spread considering the networks as a continuous, without the study of individual nodes. Instead of it, if the network size was small we could discretize it, allowing the direct analysis of contagions. This is what "tracking team" make: when an infected individual appears, they look at the contacts that he has made with other people. By this tracking, appear possibles susceptible that can be infected by the previous patient. So, this is a way to isolate the outbreak, avoiding the spread in all the network and finding the "zero patient".

Índice de figuras

| | | |
|-----|---|----|
| 1. | Red homogénea y su distribución de grado | 5 |
| 2. | Algunos tipos de redes homogéneas | 6 |
| 3. | Tres redes aleatorias con igual N y p | 7 |
| 4. | Red aleatoria y su distribución de grado | 9 |
| 5. | Redes aleatorias por conservación de grado y aleatoriedad total | 17 |
| 6. | Red de escala libre y su distribución de grado | 28 |
| 7. | Red de escala libre por modelo de selección de arista | 30 |
| 8. | Evolución del modelo SIS según β | 44 |
| 9. | Evolución del modelo SIS según λ | 46 |
| 10. | Estado endémico para modelo SIS | 47 |
| 11. | Evolución de individuos para modelo SIR | 48 |
| 12. | Comparación del número de muertes en distintas epidemias | 54 |
| 13. | Red con inmunización de <i>hubs</i> | 56 |
| 14. | Red temporal | 57 |
| 15. | Distribución de ráfagas | 58 |
| 16. | Esquema modelos SI, SIS y SIR | 62 |

Índice de tablas

| | | |
|----|---|----|
| 1. | Valores de diferentes parámetros para algunos tipos de redes reales | 14 |
| 2. | Distribución de grado y coeficiente de agrupamiento para distintos tipos de redes | 39 |
| 3. | Tiempos característicos y umbrales epidémicos para los modelos SI, SIS y SIR | 50 |

Referencias

- [1] Alsina C. *Mapas del metro y redes neuronales*. RBA Coleccionables.
- [2] Barabási AL et al. *Network science*. Cambridge university press, 2016.
- [3] Newman M. *Networks: An Introduction*. Oxford University Press, Inc., USA, 2010.
- [4] Maslov S. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, May 2002.
- [5] Milgram S, Diaz-Guilera A, Arenas A, Guimerà R, Molina JL, Casas F, and Lozares C. Comentarios a el problema del mundo pequeño de stanley milgram. *Araucaria*, 4, 01 2003.
- [6] Backstrom L, Boldi P, Rosa M, Ugander J, and Vigna S. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference, WebSci '12*, page 33–42, New York, NY, USA, 2012. Association for Computing Machinery.
- [7] Fernholz D and Ramachandran V. The diameter of sparse random graphs. *Random Structures & Algorithms*, 31(4):482–516, 2007.
- [8] Cohen R and Havlin S. Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5), Feb 2003.
- [9] Bianconi G and Barabasi AL. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54:436, 05 2001.
- [10] Dorogovtsev SN, Mendes JF, and Samukhin AN. Structure of growing networks with preferential linking. *Physical review letters*, 85(21):4633, 2000.
- [11] Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, and Vicsek T. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):590–614, Aug 2002.
- [12] Bauke H, Moore C, Rouquier JB, and Sherrington D. Topological phase transition in a network model with preferential attachment and node removal. *The European Physical Journal B*, 83(4):519–524, 2011.

- [13] Dorogovtsev SN and Mendes JFF. Effect of the accelerating growth of communications networks on their structure. *Physical Review E*, 63(2):025101, 2001.
- [14] Dorogovtsev SN and Mendes JFF. Evolution of networks with aging of sites. *Physical Review E*, 62(2):1842, 2000.
- [15] Pastor-Satorras R, Vázquez A, and Vespignani A. Dynamical and correlation properties of the internet. *Physical review letters*, 87(25):258701, 2001.
- [16] Martín del Rey A and Peinado A. Mathematical models for malware propagation in wireless sensor networks: An analysis. In *Computer and Network Security Essentials*, pages 299–313. Springer International Publishing, aug 2017.
- [17] Zviovich D. [Geo-spatial-temporal COVID-19 simulations and visualizations over USA](#), 2020.
- [18] Vazquez A, Rácz B, Lukács A, and Barabási AL. Impact of non-poissonian activity patterns on spreading processes. *Phys. Rev. Lett.*, 98:158702, Apr 2007.
- [19] Masuda N and Holme P. Predicting and controlling infectious disease epidemics using temporal networks. *F1000prime reports*, 5:6, 03 2013.
- [20] Karsai M, Kivela M, Pan RK, Kaski K, Kertész J, Barabási AL, and Saramäki J. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83(2), Feb 2011.
- [21] Rocha LEC, Liljeros F, and Holme P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLOS Computational Biology*, 7(3):1–9, 03 2011.