Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de
Biosistemas

Modelos estocásticos de ecología de bacterias y virus.

Tesis Doctoral

Rafael Ángel Vida Delgado

Máster Física Sistemas Complejos
Licenciado en Física

2019

Ingeniería Agroforestal
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de
Biosistemas

Modelos estocásticos de ecología de bacterias y virus.

Rafael Ángel Vida Delgado

Máster Física Sistemas Complejos
Licenciado en Física

Directores:

Javier Galeano

Doctor en Ciencias Físicas
UPM

Sara Cuenda

Doctora en Ciencias Físicas
UAM

2019

A la esencia de lo estocástico.

Agradecimientos

Hay tanto que agradecer, hay tantas personas que han participado, se han implicado, han ayudado, han guiado... ¿por dónde empezar? Siguiendo el orden aleatorio que proporciona la memoria, quiero agradecer a:

Mis tutores, Sara y Javi, que me han guiado, ayudado, enseñado y soportado, que han aceptado tener un alumno crecido, que da guerra y que no siempre tiene el tiempo que realmente se necesita. A Javi, que siempre se emociona con la física, siempre ve el vaso medio lleno, siempre tiene alguna propuesta y con paciencia infinita. Es un ejemplo a seguir. A Mario, con cuyas charlas recuperaba la ilusión por lo que estaba haciendo y comparte conmigo la pasión por aprender. A Lucía, que me abrió un mundo entero, me ha llevado a conocer a personas increíbles y nos preparó el terreno de investigación. A Rosa, que enseguida se embarcó en el proyecto y nos ha ayudado mucho más de lo que nadie sabrá nunca. A JuanPe, tranquilo, seguro de sí mismo, que en el momento de saturación, a punto de tirar la toalla, ayudó. A Sole y a Sofía, que han sido capaz de soportar un doctorado compartiendo mi poco tiempo con dos trabajos y una tesis. A Sofía, que es más lista que el hambre. Sofía, espero que esto te inspire en el futuro.

Índice general

I	Introducción	3
1.	Introducción	5
II	Modelos estocásticos de propagación de virus en redes	9
2.	Conceptos previos	11
2.1.	Sistemas Complejos	11
2.2.	Introducción a las redes complejas	13
3.	Propagación de virus en redes de comunicaciones	23
3.1.	Ciberguerra	23
3.2.	Propagación de epidemias en red	27
3.3.	Modelo de propagación SIS en redes	29
4.	Modelo de propagación de virus en redes multicapa	33
4.1.	Modelo de propagación multicapa	33
4.1.1.	Matriz efectiva de contagio	36
4.2.	Aplicación de los resultados a la red de colaboraciones de FisEs.	39
4.2.1.	Análisis topológico de la red	44
4.3.	Propagación de virus e inmunización	47
4.3.1.	Identificación de nodos críticos en la propagación de virus en redes multicapa	49

III Modelos estocásticos en Ecología del Microbioma y Aplicaciones	53
5. Conceptos previos	55
5.1. Interacciones entre especies	56
5.2. Modelos clásicos de depredación y competición	58
5.3. Introducción a agent-based modeling (ABM)	61
6. Fibrosis quística	65
6.1. Fibrosis quística y su relación con la microbiota	65
6.2. Modelo con dos tipos de bacterias para FQ	70
6.2.1. Estudio del equilibrio sin difusión	72
6.2.2. Estudio del equilibrio con difusión	75
6.3. Modelo con más de dos especies para FQ	76
6.3.1. Caso 1: una especie de depredador y una de presa. . .	77
6.3.2. Caso 2: un depredador en equilibrio dinámico con dos presas.	80
6.4. Modelo muchas especies con difusión para FQ	86
6.5. Efecto de la diversidad sobre la estabilidad en el modelo para FQ	90
7. <i>C. difficile</i> y transplante de heces	101
7.1. Modelo para microbiota intestinal y Clostridium	102
7.1.1. Datos experimentales	102
7.1.2. Análisis de estabilidad del modelo para Clostridium . .	106
7.2. Modelo orientado a agentes: <i>C. difficile</i>	112
IV Conclusiones y futuros pasos	119
8. Conclusiones y futuros pasos	121
8.1. Conclusiones	121
8.2. Artículos y colaboraciones	123
8.3. Futuros pasos	123

V	Anexos	125
A.	Procesos de Markov y ecuación maestra	127
A.1.	Estado del sistema	127
A.2.	Transiciones entre estados	129
A.3.	Procesos de Markov	129
B.	Análisis de estabilidad de puntos críticos	133
C.	Distribución de brechas	137
D.	Nociones de Microbiología	141
E.	Material y métodos	147
E.1.	Propagación de malware	147
E.2.	Microbiota humana	148
	Bibliografía	149
	Índice de figuras	161

Resumen

En la presente tesis se presentan los resultados del estudio de un conjunto de problemas relacionados con la propagación de virus en redes multicapa, ecología de poblaciones y dinámica de los ecosistemas, así como las técnicas y los modelos matemáticos utilizados en dichos estudios.

El primer problema de interés que se propone en la Tesis es el de propagación de virus informáticos en redes multicapa. Los virus actuales usan un método de propagación que se conoce como multivector. Esto quiere decir que usan, simultáneamente, diferentes protocolos de comunicaciones para infectar. Adicionalmente a la infección a través de la red de comunicaciones, usan técnicas de propagación que involucran a los dispositivos externos que estén conectados a los ordenadores. También aplicaciones de software, correo electrónico, etc. Estas formas de comunicación entre los nodos presentan topologías de red diferentes entre sí, aunque los nodos que componen las redes son los mismos en todas ellas, es decir, constituyen una única red multicapa. Se presentan en el presente documento de Tesis modelos matemáticos que permiten estudiar la propagación de la infección de virus multivector de manera más realista que los modelos existentes hasta ahora, que asumen que los nodos son diferentes en cada capa o analizan la propagación en las capas por separado. El modelo se basa en modelos anteriores de infección SIS en redes. Como principal resultado se ha determinado que la densidad de nodos infectados en redes multicapa es mayor que la esperada si se analizan por separado las propagaciones en cada capa. Además, el modelo propuesto permite diseñar estrategias de inmunización de los nodos con el fin de ralentizar o evitar la epidemia en redes informáticas. Esto tiene aplicaciones inmediatas en la protección de redes de ordenadores frente a infecciones que puedan afectar a la continuidad del negocio. Se ha usado el modelo con los datos experimentales de una red multicapa formada

por los miembros de una comunidad científica, sus colaboraciones y las redes de ordenadores existentes en sus centros de investigación o universidades.

Con el fin de estudiar diferentes tipos de interacciones interespecíficas se han modelizado y analizado dos situaciones biológicas relacionadas con los ecosistemas de bacterias existentes en el cuerpo humano. Uno de ellos es el tratamiento de síntomas de la enfermedad de Fibrosis Quística (FQ), que es una enfermedad genética cuya sintomatología está relacionada con la microbiota del pulmón. Para este problema se ha propuesto la introducción en esa microbiota de una bacteria bacteriófaga depredadora de ciertos tipos de bacterias como moduladora de factores de virulencia y como una estrategia ecológica para el control de las bacterias de los pulmones del enfermo de FQ. Se propone un modelo de depredador-presa sobre el que se analiza la estabilidad, además de un modelo orientado a agentes. Se ha determinado mediante estos modelos que el tratamiento propuesto tiene fundamento teórico y que hay un umbral mínimo de bacterias depredadoras que debe introducirse para conseguir el efecto deseado terapéuticamente. Esto permite diseñar estrategias de tratamiento de los síntomas de enfermos de FQ.

El tercer problema que se utiliza como marco de referencia en los modelos es el de la infección por *C. difficile* en el intestino humano, que, en ciertos casos, puede producir la muerte. Un tratamiento muy efectivo consiste en realizar un trasplante de heces de un donante sano, introduciendo una cantidad de heces con bacterias en el intestino del enfermo. En nuestro estudio se han realizado varios modelos de la microbiota, incluyendo un modelo orientado a agentes. Estos modelos han permitido determinar que existe un umbral relativo de proporción mínima de bacterias correspondientes al donante que deben introducirse para que se pueda alterar el ecosistema del intestino de forma que se desplace la población del patógeno. Se ha reproducido cualitativamente el comportamiento observado en la distribución de poblaciones en la microbiota tras un trasplante. También se ha analizado el efecto de las diferentes interacciones interespecíficas en la estabilidad del sistema de la microbiota, teniendo en cuenta que la estabilidad está relacionada con la salud del sujeto hospedante. Se ha determinado la importancia de la competición entre las bacterias de una misma especie en esa estabilidad.

Abstract

In the present Thesis we present the results of the study of a set of problems related to the propagation of viruses in multilayered networks, ecology of populations, dynamics of ecosystems, as well as the techniques and mathematical models used in these studies.

The first problem of interest is the propagation of computer viruses in multilayer networks. Current viruses use a propagation method known as multivector. This means that they use different communication standards in the infection process. In addition to the usual infection process using the communications network, they use propagation techniques that involve external devices connected to computers, software applications, email, etc. These forms of communication between the nodes have network topologies that are different from each other, although the nodes that constitute the networks are the same in all of them. That is, they constitute a single multilayer network. In this Thesis document we present mathematical models that allow us to study the propagation of multivector virus infection in a more realistic way than the models existing up to now, which assume that the nodes are different in each layer, or analyze the propagation in the layers separately. The model is based on previous models of SIS infection in networks. As main results it has been determined that the density of infected nodes in multilayer networks is higher than expected if the propagations in each layer are analyzed separately. In addition, the proposed model allows the design of immunization strategies for the nodes in order to slow down or prevent the epidemic in computer networks. This has immediate applications in the protection of computer networks against infections that may affect the continuity of the business. The model has been used with the experimental data of a multilayer network formed by the members of a scientific community, their collaborations and the existing computer networks in their research centers or universities.

In order to study different types of interspecific interaction, two biological situations related to the ecosystems of bacteria existing in the human body have been modeled and analyzed. One of them is the treatment of symptoms of Cystic Fibrosis (CF) disease, which is a genetic disease whose

symptomatology is related to the microbiota of the lung. For this problem has been proposed the introduction into that microbiota of a predatory bacteriophage of certain types of bacteria as a virulence factor modulator as an ecological strategy in order to control pathogenic bacteria in the lungs of the CF patient. We propose a predator-prey model on which stability is analyzed, as well as an agent-oriented model. It has been determined through these models that the proposed treatment has, in fact, a theoretical basis and that there is a minimum threshold of predatory bacteria that must be introduced to achieve the desired therapeutic effect. This makes it possible to design strategies for treating the symptoms of CF patients.

The third problem that is used as a framework in the models is that of infection by *C. difficile* in the human intestine, which, in certain cases, can cause death. A very effective treatment consists of performing a transplant of feces from a healthy donor, introducing a quantity of feces with bacteria in the intestine of the patient. In our study, several models of the microbiota have been made, including an agent-oriented model. These models have allowed to determine that there is a relative threshold of minimum proportion of bacteria corresponding to the donor that must be introduced so that the ecosystem of the intestine can be altered and therefore the population of the pathogen is displaced. The behavior observed in the distribution of populations in the microbiota after a transplant has been reproduced qualitatively. The effect of the different interspecific interactions on the stability of the microbiota system has also been analyzed, bearing in mind that stability is related to the health of the host subject. The importance of competition between bacteria of the same species in that stability has been determined.

Parte I

Introducción

Introducción

1

” *The Three Theorems of Psychohistorical Quantitivity: The population under scrutiny is oblivious to the existence of the science of Psychohistory. The time periods dealt with are in the region of 3 generations. The population must be in the billions (± 75 billions) for a statistical probability to have a psychohistorical validity.*

— Isaac Asimov
Hari Seldon, Foundation

Motivación

Los trabajos que se presentan en esta tesis doctoral tienen como origen el trabajo de fin de máster de Física de Sistemas Complejos¹ del autor. En este trabajo se analizó la red de ordenadores de Telefónica utilizando técnicas de redes complejas y se presentaron propuestas de análisis de propagación de virus informáticos en redes usando la información topológica y estructural de esa red. También se propusieron modelos de propagación de virus sobre otros existentes de propagación de enfermedades (SIS²).

Tras estos trabajos iniciales surgió la pregunta de si estos virus compiten entre sí, es decir si hay otro tipo de interacciones interespecíficas similares a las existentes en otros ecosistemas. O lo que es lo mismo: si hay una ecología de malware.

Para analizar esa y otras cuestiones se desarrolló el trabajo de análisis de propagación de virus en redes multicapa que se presenta en la presente me-

¹Facultad de Ciencias Físicas de la UNED.

²Para la definición del modelo SIS, véase el capítulo 3.2.

moria y que tuvo como fruto dos publicaciones científicas (Vida *et al.*, 2013; Vida *et al.*, 2015). Para estos trabajos se consideró que los virus informáticos y el *malware*³ en general usan varios vectores de propagación simultáneamente⁴. Esto supuso que propusiéramos un modelo basado en modelos de redes multicapa. El modelo describe de manera muy eficiente cómo se propagan estas piezas de software maligno infectando ordenadores y otros elementos informáticos (impresoras, móviles, tabletas). También permite determinar estrategias de protección de la red. Ha demostrado tener aplicaciones prácticas relevantes. Actualmente se usan algunos resultados de nuestro modelo en la estimación de probabilidad de ataque de malware en los análisis de riesgos de varias empresas españolas.

Sin embargo, fruto de una colaboración con el departamento de Microbiología de Ramón y Cajal, se decidió avanzar en otro marco de referencia y adaptar el modelo a la ecología de bacterias. Esto supone disponer de un marco de trabajo muy interesante, de una complejidad elevada y en el que hay varios problemas con aplicaciones prácticas con valor médico. El cambio permite, además, avanzar en otro tipo de modelos estocásticos y nos abrió un campo de investigación con mucho potencial.

Se presentan en el presente documento de tesis los resultados de algunas de estas investigaciones sobre la microbiota⁵ humana. En primer lugar se analiza la enfermedad de Fibrosis Quística, cuyos síntomas están relacionados con el equilibrio en la microbiota del pulmón. Se obtienen resultados prometedores para un tratamiento de los síntomas de dicha enfermedad. Este trabajo ha dado como fruto una publicación científica, en colaboración con médicos y microbiólogos del Ramón y Cajal (Dios *et al.*, 2017).

Además se ha estudiado el efecto de la competición en el equilibrio del conjunto de bacterias del intestino humano. Se ha observado que, tras la administración de antibióticos, la microbiota se ve afectada. Desaparece la diversidad de bacterias y, en ciertos casos, se produce una infección por una bacteria (*C. difficile*) que produce diarreas en el sujeto pudiendo llegar a

³Viene de *malicious software*, cualquier pieza de software diseñada para hacer daño a un ordenador.

⁴La red Ip de comunicaciones, redes de relaciones sociales mediante correo electrónico, etc.

⁵El conjunto de microorganismos que conviven simbióticamente en el cuerpo del hospedador. Véase anexo D.

matar al enfermo. Con estos modelos intentamos dar respuesta a cuestiones tales cómo: ¿Cómo puede perturbarse el estado de equilibrio de la microbiota que supone un problema de salud para el hospedante humano? ¿Qué elementos participan más en la recuperación de una microbiota determinada? ¿Cómo es de importante la diversidad para mantener el ecosistema en equilibrio? ¿Y para desequilibrarlo? ¿Y qué papel tiene la interacción entre los miembros de una misma especie para el sistema completo?

Estructura de la Tesis

En la **primera parte** se expone una introducción que contiene, además de este capítulo, la motivación y el desarrollo histórico de los trabajos que se presentan.

En la **segunda parte** tratamos los modelos relacionados con la propagación de virus informáticos en redes. En el primer capítulo de esta parte se presentan ciertos conceptos previos necesarios para entender los modelos y el marco matemático en el que se desarrollan. Hablamos de conceptos básicos de complejidad y de redes complejas, enunciando algunas métricas importantes a la hora de caracterizar esas redes. El segundo tema de esta parte es una introducción a la ciberguerra, la relevancia del malware como arma en la sociedad actual y sobre la propagación de virus informáticos en redes complejas. El tercer tema presenta nuestra contribución al estudio de la propagación de estos virus en redes multicapa mediante un modelo propio mucho más realista que los presentados hasta ahora y más adecuado a la forma de propagación del software maligno que usa formas de propagación multivector. Este modelo en redes multicapa se usa en una red real de colaboraciones científicas y se utiliza para determinar estrategias de inmunización como método de protección contra la infección.

La **tercera parte** es la relativa a los modelos de microbiota humana. El primer tema es, de nuevo, una introducción a los conceptos previos necesarios para entender la modelización y su uso en microbiota del pulmón y del intestino. Se repasan las interacciones interespecíficas entre las especies según la ecología de poblaciones, haciendo hincapié en dos tipos de interacciones en ecosistemas como elemento regulador de la estabilidad,

sobre todo en la competición y la depredación ya que son las interacciones que más efecto tienen en los ecosistemas de bacterias estudiados. Se basan en modelos clásicos de depredación y competición. También se realiza una introducción a los modelos orientados a agentes, que permitirá desarrollar los estudios que complementa al modelo matemático.

El segundo tema presenta la enfermedad de Fibrosis Quística y varios modelos de la microbiota basados en modelos de depredación, ya que esto permite estudiar un posible tratamiento de los principales síntomas de esa enfermedad. En ese tema presentamos un modelo con dos tipos de bacterias, con y sin difusión, un modelo propuesto para el caso muchas especies de bacterias y los resultados de un modelo orientado a agentes. El último tema de esta parte presenta la infección por una bacteria en el intestino, y se analiza la estabilidad de la microbiota usando para ello nuestro modelo orientado a agentes. El modelo da soporte al análisis del efecto de trasplante de heces en enfermos con infecciones de *Clostridium difficile* en sus tres fases: tratamiento con antibióticos, infección y la recuperación mediante repoblación de la microbiota con el trasplante de heces. Para ello, se analiza la estabilidad con matrices aleatorias y se utiliza el modelo orientado a agente. El análisis es prometedor ya que parece que permitirá diseñar una estrategia de recuperación de la microbiota en enfermos de enfermedad inflamatoria intestinal⁶.

La **cuarta parte** contiene el tema correspondiente a una compilación de las conclusiones y futuros pasos.

Finalmente hay varios **anexos**, que sirven de soporte para entender la gravedad de la situación del software malicioso en las redes de ordenadores actuales, los procesos de Markov y la ecuación maestra, análisis de estabilidad, nociones de microbiología y finalmente un capítulo de material y métodos.

⁶Se usará el término en inglés *Inflammatory Bowel Disease* o más comúnmente: IBD.

Parte II

Modelos estocásticos de propagación
de virus en redes

Conceptos previos

2

” *In retrospect, Euler’s unintended message is very simple: Graphs or networks have properties, hidden in their construction, that limit or enhance our ability to do things with them. For more than two centuries the layout of Königsberg’s graph limited its citizens’ ability to solve their coffeehouse problem. But a change in the layout, the addition of only one extra link, suddenly removed this constraint.*

— **Albert-László Barabási**
(Físico)

En el presente capítulo se repasan algunos conceptos previos necesarios para entender el modelo de propagación de virus informáticos en redes multicapa. Se presentan las técnicas de redes complejas y se enumeran las principales métricas que caracterizan las propiedades de las redes, haciendo foco en las usadas en el documento de Tesis.

2.1 Sistemas Complejos

Los sistemas sobre los que se trabaja en la presente Tesis de doctorado son sistemas complejos. Es interesante pararse en esta sección a definir éste y otros conceptos relacionados. De acuerdo con Manson (2001) hay tres tipos de complejidad que los investigadores suelen asociar, en ocasiones, a un mismo problema: algorítmica (Turing, 1950), determinista (Feigenbaum, 1983) y agregada. En lo que concierne a esta memoria de Tesis, los problemas sobre los que trabajamos pertenecen principalmente a este último tipo, por lo que nos detenemos a examinar más detenidamente la definición.

Así, un sistema es complejo si está compuesto por partes que interaccionan unas con otras de forma que esas relaciones producen propiedades nuevas que no pueden explicarse directamente a partir de las propiedades de los elementos aislados. Estas propiedades se conocen como propiedades emergentes (Newman, 2011). Típicamente estos sistemas se definen más por las relaciones que por sus partes constituyentes. Además, según Costanza *et al.* (1993), dado el número y variedad de estas relaciones, la deducción de comportamientos del sistema a partir de las de sus partes no es alcanzable usando técnicas tradicionales.

Los problemas que se plantean en la presente Tesis se desarrollan en varios sistemas complejos (entendidos principalmente como complejidad agregada):

1. Redes multicapa formadas por interacciones sociales y las redes de comunicaciones informáticas asociadas (cuya introducción puede verse en el capítulo 2), en las que se quiere estudiar cómo se propaga un virus.
2. Los que forman los ecosistemas (ver capítulo 5 como introducción) en los que diferentes especies compiten por recursos, colaboran y depredan. Particularmente, estudiamos el ecosistema formado por bacterias del sistema del microbioma humano, formado por poblaciones de bacterias interaccionando.

Para estudiar el comportamiento de estos sistemas hay varios conjuntos de técnicas. Por un lado, se usan modelos matemáticos simplificados, entendiendo por modelo el conjunto de ecuaciones diferenciales que permiten simular un sistema. Estos modelos usan herramientas de teoría de juegos, sistemas dinámicos, redes complejas y métodos numéricos. En los modelos de ecuaciones diferenciales, se utiliza una técnica de análisis que consiste en determinar los puntos de equilibrio de los mismos y analizar el comportamiento del sistema tanto en esos puntos como en zonas cercanas (May, 1972). Como toda ecuación diferencial de orden superior puede reducirse a un sistema de ecuaciones diferenciales de primer orden equivalente, el estudio de la estabilidad de las soluciones puede reducirse al estudio de la estabilidad de los sistemas de ecuaciones. De este modo, pueden localizarse los puntos de equilibrio de las soluciones y, estudiando el tipo de equilibrio,

se puede determinar el comportamiento del sistema y su respuesta a cambios de condiciones.

La Tesis que se presenta en este documento utiliza también esas técnicas de análisis de estabilidad lineal para estudiar el comportamiento dinámico de sistemas complejos relacionados con la microbiota humana y con la propagación de enfermedades o de virus informáticos. Se usan, además, técnicas de redes complejas¹ que permitirán analizar matemáticamente la relación entre las partes de los sistemas estudiados y propiedades emergentes (véase Boccaletti *et al.*, 2006; Newman, 2010 y más recientemente Newman, 2018).

Otro conjunto de técnicas de análisis de sistemas complejos es el asociado a sistemas discretos, basadas en las simulaciones computacionales, métodos de Montecarlo, etc. En concreto, en los trabajos realizados para la presente Tesis, se utilizan simulaciones basadas en agentes para sistemas que son discretos simultáneamente en tiempo y espacio. En este tipo de simulaciones se diseña un modelo en el que cada agente posee una serie de propiedades e interacciona con otros agentes. Esto se usa para modelar infecciones asociando agentes a los nodos susceptibles de ser infectados y funciones que cambian una de propiedades (el estado, que pasa de susceptible a infectado, como se verá en el apartado correspondiente).

Los diferentes modelos se complementan unos a otros para obtener una visión del comportamiento de la dinámica continua y de la dinámica discreta.

2.2 Introducción a las redes complejas

Una red es un conjunto de elementos (conocidos como nodos) que se relacionan entre sí mediante enlaces. Algunos ejemplos de estas redes son las que se forman cuando consideramos personas y sus relaciones interpersonales, ordenadores que mediante cables pueden intercambiar información,

¹Se definirá correctamente este concepto en el apartado correspondiente de la introducción. Simplificando muchísimo y, en aras de la claridad, se trata de modelos en los que los componentes del sistema complejo son nodos de una red y sus relaciones constituyen los enlaces.

ciudades que se comunican mediante carreteras, conceptos que aparecen relacionados en textos, etc.

Históricamente aparece la teoría de grafos² como la rama de la matemática que estudia un conjunto de problemas planteados sobre un escenario de redes y se inicia con el famoso ejemplo de los puentes de Königsberg (cuya solución fue propuesta por Euler en 1736). En 1920 resurge este concepto, gracias a la aparición de ramas del saber tales como las ciencias sociales, en las que las relaciones entre individuos y su caracterización cobran importancia. El marco teórico estaba bastante desarrollado, pero no fue hasta cerca del año 2000 que no se retoman los trabajos relacionados con la teoría de grafos, momento en el que aparecen dos artículos que lanzan el estudio de las redes complejas (Watts y Strogatz, 1998; Barabási *et al.*, 1999) gracias a que la capacidad de computación y el desarrollo de redes de ordenadores permitieron el desarrollo y sobre todo la aplicación del conjunto de herramientas estadísticas necesarias para analizar el comportamiento de redes complejas.

La teoría de redes constituye una herramienta importante en el estudio de sistemas complejos en los que el comportamiento global del sistema presenta propiedades que no se pueden deducir directamente del comportamiento individual de sus componentes. Proporciona un marco que permite, a partir de la definición de agentes y sus relaciones (nodos y enlaces), analizar esos comportamientos emergentes y del sistema. Se convierte, por lo tanto, en una herramienta importante en el estudio de sistemas complejos agregados. Esto es así porque existe una relación entre la estructura de la red y su función³. Las revisiones de Albert y Barabási (2002) y Newman (2003) constituyen un magnífico punto de referencia.

En este capítulo se presenta un ejemplo de la red de relaciones entre superhéroes para ilustrar la caracterización y análisis de las redes. Esta red se construye a partir de las interacciones de los superhéroes en cómic de Marvel, donde los nodos son los superhéroes y los enlaces se crean cuando

²Precursora de la teoría de redes complejas.

³Ha quedado demostrada esa relación en numerosas ocasiones.

Fig. 2.1.: Red social de Superhéroes de cómics Marvel, creada por Cesc Rosselló, Ricardo Alberich y Joe Miro de la Universidad de las Islas Baleares (Alberich *et al.*, 2002). Recopilada por Infochimps, transformada y mejorada por Kai Chang. La representación ha sido hecha por el autor de esta Tesis a partir de esos datos usando el paquete informático Gephi. Los colores van de rojos para nodos de mayor grado hacia azul, nodos de menos grado.

dos personajes aparecen en el mismo cómic⁴. En la en la figura 2.1 se representan los nodos de esa red y los enlaces tal y como se acaban de definir. En esa representación gráfica usada en esta figura la información que puede extraerse es muy poca: ¿hay muchas colaboraciones? ¿Qué héroes aparecen más en los cómics de otros? Para responder a estas preguntas vamos a definir una serie de conceptos y métricas asociadas a los nodos y enlaces de la

⁴Datos obtenidos de Alberich *et al.* (2002), cuyo trabajo tenía como objetivo analizar si una red inventada, tal y como es la de superhéroes de Marvel, tenía las mismas propiedades que una red social real.

red (para ampliar esta información véase Boccaletti *et al.*, 2006; Boccaletti *et al.*, 2014).

Como se ha dicho anteriormente, existen dos elementos constitutivos de la red:

1. **Nodos**, caracterizados por etiquetas y un conjunto de propiedades. Denotemos el nodo por el índice i .
2. **Enlaces**, que definen las relaciones entre los nodos, denotados por l_{ij} (es decir, el enlace entre los nodos i y j). Así mismo, puede tener un peso que indica qué conexiones son más fuertes o abundantes.

Redes dirigidas y no dirigidas. La red puede ser dirigida, si la relación entre los nodos es simétrica o, en caso contrario, no dirigida. Por ejemplo, la red de superhéroes de nuestro ejemplo es no dirigida, ya que sólo se ha tenido en cuenta que dos superhéroes aparezcan juntos en un cómic para que exista enlace entre ellos. Si por el contrario hubiésemos tenido en cuenta cuál es la colección a la que pertenece el cómic en el que aparece, podríamos hablar de superhéroes anfitriones y huéspedes, por lo que habría que definir enlaces dirigidos entre ellos.

Matriz de adyacencia. Dos nodos que están conectados se dice que son vecinos o que son adyacentes. Estas relaciones se caracterizan mediante una matriz de adyacencia, cuyos elementos A_{ij} son uno si hay conexión entre los nodos i y j y cero si no están conectados directamente. Se usa esta matriz de adyacencia en varias ocasiones durante la Tesis. El número total de enlaces de una red de N nodos será, entonces, 0 como mínimo y $N(N - 1)/2$ como máximo para una red dirigida, o $N(N - 1)$ como máximo para una no dirigida (suponiendo que no hay autoenlaces).

Grado de un nodo. La cantidad de enlaces que tiene con otros nodos se conoce como **grado** k_i del nodo i . Podemos calcular el grado de un nodo a partir de la matriz de adyacencia:

$$k_i = \sum_j A_{ij} \quad (2.1)$$

para el caso de una red no dirigida. En el caso de que la red sea dirigida, podremos definir el grado de salida, el grado de entrada, o el grado total de

Fig. 2.2.: Selección de los superhéroes con más relaciones con otros superhéroes a partir de la red de figura 2.1. El tamaño de los nodos es proporcional al grado, y el color va del rojo para los de mayor grado a los azules oscuro los de menor grado. Se representan los nodos de grado 350 a 2185.

cada nodo y su distribución. La distribución de grado nos da información relevante de la red. Usamos esta métrica para tener una mejor representación de cómo es la red en figura 2.2, donde hemos seleccionado sólo un tipo de nodos, los superhéroes mejor relacionados (o lo que es lo mismo, los nodos de mayor grado). En la figura 2.3 se ha presenta la distribución de grado de la misma red de superhéroes que estamos usando de ejemplo en este capítulo. La distribución sigue un patrón típico de una ley de potencias, en

Fig. 2.3.: Distribución de grado de la red de superhéroes de Marvel presentada como ejemplo en figura 2.1 en escala logarítmica. En este caso la distribución sigue aproximadamente una ley de potencias.

el que hay unos pocos nodos con muchas conexiones y muchos nodos con pocas conexiones.

Betweenness centrality, que relaciona el número de nodos con la importancia que tienen en los caminos más cortos que recorren la red. Cuantos más caminos más cortos pasen por un nodo, mayor será la *betweenness centrality*, reflejando lo céntrico o relevante que es un determinado nodo. Si n_{jk} es el número de rutas más cortas que conectan los nodos j y k y $n_{jk}(i)$ el número de rutas más cortas que van de j a k pasando por i , se define la *betweenness centrality* del nodo i como:

$$b_i = \sum_{j,k,j \neq k} \frac{n_{jk}(i)}{n_{jk}} \quad (2.2)$$

Uno de los nodos con mayor valor de esta métrica es Spiderman. Esto quiere decir que es un contacto interesante como intermediario entre cualesquiera

otros dos superhéroes. Es a la persona a la que deberías dirigirte para localizar un contacto cercano con otro superhéroe.

Closeness centrality o **cercanía**, también es una medida de la centralidad y se calcula como la media de los caminos más cortos existentes entre un determinado nodo y todos los demás en la red. Coincide con la idea intuitiva de lo que debe ser un nodo central: cuanto más central, más cercano estará respecto a los otros nodos. Es una medida similar a la de *betweenness centrality* en el sentido de que mide en cierta manera la influencia de un nodo sobre otros. En la figura 2.4 se representa la centralidad de la red de superhéroes frente al número de vecinos. Hay unos cuantos nodos que tienen un valor muy alto, a pesar de tener pocos vecinos. Estos superhéroes tienen un papel importante en la red social que se forma.

Eigenvector centrality, o **auto-centralidad**, es una medida de la influencia que tiene un nodo en la red. Para calcularla se asocian valores relativos a todos los nodos basándose en las conexiones, de forma que los nodos con mayor valor contribuyen en mayor medida a los valores de los nodos con los que están conectados. Este valor pretende determinar la importancia de los nodos. Se define de forma que depende de la auto-centralidad de los nodos adyacentes:

$$x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j \quad (2.3)$$

que es proporcional a la suma de auto-centralidades de sus vecinos. Si tomamos el mismo λ para todos los nodos, equivale a la ecuación de autovalores $A\vec{x} = \lambda\vec{x}$. El teorema de Perron-Frobenius nos asegura que las componentes de $\vec{x}$ son no negativas para el mayor autovalor asociado a A . Los superhéroes con mayor *eigenvector centrality* tendrán, en general, contactos más valiosos. A veces lo más importante no es conocer a muchos superhéroes, si no a los relevantes.

Coefficiente de clustering, que da una idea de la probabilidad de que si dos nodos están conectados con otro, estén también conectados entre sí. Por ejemplo, en tu red de amigos es muy posible que los amigos de tus amigos

Fig. 2.4.: *Closeness centrality* de la red de superhéroes frente al número de vecinos. Se representa para cada nodos su *closeness centrality* frente al grado.

sean también amigos tuyos. Para redes no dirigidas Watts y Strogatz (1998) lo definieron así:

$$c_i = \frac{n_{\text{triangulos},i}}{k_i(k_i - 1)/2} \quad (2.4)$$

siendo k_i el grado del nodo i y $n_{\text{triangulos},i}$ el número de triángulos que tienen un vértice en el nodo i . En la figura 2.5 se muestra el promedio de coeficiente de *clustering* en función del grado para el mismo ejemplo de la red de interacciones de superhéroes de cómics Marvel. Los superhéroes aparecen en publicaciones de otros superhéroes formando grupos. Este resultado es consistente con otras formas de analizar comportamiento social. Los grupos mas coexionados tienen un número de superhéroes de entre 10 y 100 miembros. Los grupos más grandes están menos interrelacionados.

Fig. 2.5.: Distribución de coeficiente de *Clustering* medio respecto al número de vecinos de la red presentada de apariciones en cómics, representada en escala logarítmica. Proporciona una medida del grado en el que los nodos tienden a formar una agrupación de racimo (*cluster*).

Propagación de virus en redes de comunicaciones

” *Why are experts so polite, patient, and forgiving when talking about cybersecurity and National Security? The drama of each script kiddie botnet attack and Nation State pilfering of our IP has been turned into a soap opera through press releases, sound bites and enforced absurdity of mainstream media.*

— **James Scott**

(Institute for Critical Infrastructure
Technology)

En este capítulo se analiza la situación actual de la ciberguerra, tipos de malware y métodos de control de la propagación de virus informáticos, además de analizar la propagación en redes de estos virus, el umbral de infección y las técnicas disponibles hasta ahora para anañizar ese tipo de infecciones.

3.1 Ciberguerra

La lucha contra el cibrecrimen está cobrando cada vez más importancia, tanto en empresas privadas como legislativamente. El reglamento europeo de la privacidad (GDPR¹), el Esquema Nacional de la seguridad² y la Ley

¹Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE.

²BOE-A-2010-1330 que persigue fundamentar la confianza en los sistemas de información.

de Infraestructuras Críticas³ son claros ejemplos de cómo la sociedad está muy preocupada por la ciberdelincuencia y, por otro lado, muestran que todavía nos movemos en un marco inmaduro de medidas de protección. El capítulo presente desarrolla un modelo de propagación de malware en redes multicapa que tiene como principal objetivo ayudar en la protección y mejorar la resiliencia de las redes informáticas, colaborando con los objetivos de control de seguridad que se presentan en esos marcos.

La Unión Europea inició varias iniciativas para intentar presentar una estrategia común contra lo que ahora se denomina ciberguerra. En concreto, la directiva NIS⁴ (Lütkebohle, 2016) pretende homogeneizar las respuestas frente a incidentes de seguridad que puedan suponer daño a la propia estructura de los estados y de focalizar los esfuerzos de protección en las infraestructuras críticas, es decir, aquellas cuyo fallo supone daño para sectores fundamentales de la sociedad: servicios médicos, telecomunicaciones en emergencias, suministro eléctrico, etc.

En respuesta a una petición de ministro de defensa de Gran Bretaña (Anderson *et al.*, 2012) se estimó que el coste mundial del malware en 2010 fue de unos 370M\$. En este informe se explica que una de las razones por las que la guerra contra el cibercrimen⁵ es tan ineficiente es porque que el malware es global y tiene fuertes extensiones. La estrategia tradicional de aislar mediante medidas de seguridad perimetral⁶ ciertos ordenadores o nodos ha quedado obsoleta y empieza a plantearse de manera generalizada la necesidad de actualizar la forma de defensa (Choo, 2011).

Si analizamos el número y tamaño de las brechas de seguridad a lo largo de 10 años, según se representan en la figura 3.1, puede apreciarse que el problema no está bajo control. En dicha gráfica cada punto representa una brecha de seguridad y la altura representa el número de nodos implicados en el suceso. Aunque hay menos ataques en los últimos años que en los

³Ley 8/2011, de 28 de abril, orientada a proteger de ciberataques las infraestructuras básicas para la supervivencia de la sociedad.

⁴Network and Information Systems.

⁵Acciones que tienen como objetivo dañar por medios electrónicos y redes de comunicaciones.

⁶Es decir, se conectan los ordenadores con una sensibilidad determinada o una función específica en una red de comunicaciones cuyo acceso está controlado.

primeros, el número de nodos afectados es similar. Aunque la tecnología de defensa y la tecnología asociada están en un nivel de madurez elevado, tanto el impacto en número de nodos afectados por las brechas de seguridad como el número de brechas no han disminuido a lo largo de los años como podría pensarse. Y esto es así a pesar de que las tecnologías de defensa y protección⁷ de sistemas informáticos y redes, así como las herramientas de detección se han generalizado y ya son parte del despliegue básico de infraestructura IT. En el anexo C se presentan datos de distribución de brechas de seguridad, por tipo de ataque y tipo de organización que las sufre de manera más detallada y respecto a los años 2016 y 2017 si desea ampliarse la visión de la situación actual en la lucha contra el cibercrimen.

Fig. 3.1.: Brechas de seguridad en un periodo de 10 años, fuente de datos: <https://www.privacyrights.org/data-breaches>, figura (Edwards *et al.*, 2016).

⁷Por ejemplo antivirus, sistemas de detección de intrusos (IDS), escáners de vulnerabilidades, *firewalls*, y un largo *etc.*

Pasemos ahora a entender la técnica usada por los virus durante los procesos de infección. Durante los últimos años el malware ha desarrollado mecanismos de propagación complejos que permite a un virus usar (simultáneamente) diferentes formas⁸ de infección. Es decir, la misma pieza de software dañina (malware) puede entrar en un sistema mediante la ejecución de una aplicación (oculta en el proceso de instalación, por ejemplo) y luego usar servicios de red para propagarse, infectar dispositivos de almacenamiento⁹ y usar el correo como forma de propagación. Un par de ejemplos claros de este tipo de comportamiento de infección multivector son el virus Conficker (Porrás *et al.*, 2009) y Stuxnet (Falliere *et al.*, 2011), o los recientes ataques de *Rasomware*,¹⁰ por citar algunos de los más famosos que tuvieron un impacto gigantesco en la red de Internet. En la figura 3.2 se representa la actividad el virus Conficker en los últimos años, que aún sigue estando activo e infectando nodos de las redes de comunicaciones.

Una vez penetran en una red de área local (LAN) la propagación es muy rápida si encuentran las vulnerabilidades adecuadas, ya que usan la red IP para localizar esas vulnerabilidades en los sistemas conectados a la red.

Debido al uso de múltiples vectores de infección y técnicas de adaptación de los virus, el proceso de re-infección es bastante común en epidemias informáticas. Técnicamente es complicado limpiar una red completamente para parar la infección antes de que el virus vuelva a extenderse. Por eso es crítico poder identificar los nodos cuya inmunización detienen de manera más eficiente la propagación y así poder diseñar una estrategia para ralentizar o incluso parar la propagación de una infección. Para defenderse de la propagación de virus en red existen medidas de control conocidas como medidas de seguridad perimetral (Antoine *et al.*, 2003). Estas medidas intentan evitar la primera infección dentro de una determinada red, limitando las conexiones desde nodos externos. Se asume que detener la infección en una red una vez se ha infectado un nodo es difícil. Por supuesto, esto ha hecho que los desarrolladores de virus (que saben que el escaneado que hace el

⁸Vectores de propagación, en terminología de seguridad de la información.

⁹Pendrives y discos usb externos.

¹⁰Un tipo de programa dañino que restringe el acceso a determinadas partes o archivos del sistema infectado y pide un rescate a cambio de quitar esta restricción, típicamente cifrando esos archivos.

malware para localizar vulnerabilidades se corta gracias a la existencia de esas medidas perimetrales¹¹) usa otros vectores de infección para saltar esas medidas de seguridad perimetral.

Volviendo a los virus con propagación multivector, el que es considerado el primer virus diseñado específicamente para la ciberguerra (Stuxnet), usa un proceso de propagación en tres estadios. El primero consiste en usar pendrives y discos portátiles infectados con el malware. Algunos autores relacionan el origen de la infección con una conferencia en el 2010 sobre sistemas SCADA (Supervisory Control and Data Acquisition), porque ese tipo de sistemas eran el principal objetivo del ataque. La infección estaba incrustada en software que se distribuyó en una conferencia de SCADA (Matrosov *et al.*, 2011). Una vez se infectaba un nodo, se propagaba dentro de la LAN, utilizando varios protocolos¹² de comunicaciones. Es decir, usó como vector de entrada la red social mediante la infección de software del congreso y usó una red IP diferente para propagarse. Varios años después del máximo de infección, sigue siendo uno de los malwares más extendidos (Asghari *et al.*, 2015).

Algunos tipos de malware, además, desarrollan también métodos de auto-protección y de ocultación a los sistemas de detección que se actualizan mediante redes Peer-To-Peer (Fox, 2001), de forma que frustra los intentos de borrado de herramientas de seguridad y las actualizaciones del antivirus y sistemas operativos. Esto hace que la eliminación de malware sea una tarea muy complicada.

3.2 Propagación de epidemias en red

En la literatura hay dos modelos fundamentales y que son usados ampliamente: el modelo Susceptible-Infectado-Recuperado (SIR) y el modelo Susceptible-Infectado-Susceptible (SIS). El modelo SIR es más adecuado

¹¹Los routers que las interconectan cortan ese tipo de tráfico IP, al igual que los cortafuegos (*Firewalls*).

¹²Es un lenguaje de intercambio de información entre elementos informáticos, un conjunto de reglas que permiten que dos o más entidades de una red de comunicación se comuniquen entre ellas para transmitir información. Pueden ser protocolos físicos (variando una magnitud física, tal y como la fase de una onda electromagnética), lógicos, de aplicación, etc.

Fig. 3.2.: Distribución de número de sistemas infectados durante el año anterior (fecha Febrero 2018). Es decir, 8 años después del máximo de infección de Conficker siguen detectándose infecciones. Fuente: Conficker Working Group, (2018).

para entender la propagación en enfermedades que proporcionan inmunización de por vida, mientras el modelo SIS se usa en enfermedades en las que el sujeto puede recaer en la enfermedad (Keeling y Eames, 2005), tal y como sucede en el caso de infecciones de virus informáticos que estamos analizando.

En ambos modelos los nodos se clasifican por su estado respecto a la infección y se definen las reglas que especifican las condiciones que debe cumplir un nodo para cambiar de estado. Estos estados son:

1. Susceptible. Estado que puede ser infectado si se ven expuestos a la enfermedad (virus en nuestro caso).
2. Infectado, si el virus ha conseguido entrar en el nodo y además el nodo puede infectar a otros nodos con los que tenga relación.
3. Recuperado, cuando el nodo ha eliminado la infección y, además, está inmunizado (no puede volver a infectarse).

Aquellos modelos se generalizaron para la propagación en redes en la literatura científica (Pastor-Satorras y Vespignani, 2001; Pastor-Satorras y Vespignani, 2001; May y Lloyd, 2001; Moreno *et al.*, 2003; Mieghem *et al.*, 2009; Gómez *et al.*, 2013; Klemm *et al.*, 2012) y más recientemente sobre varias capas

de redes interconectadas (Buldyrev *et al.*, 2010, Son *et al.*, 2012, Saumell-Mendiola *et al.*, 2012, Granell *et al.*, 2013, Cozzo *et al.*, 2013).

En esos trabajos previos se asumió que los nodos pertenecientes a diferentes redes son distintos en cada capa, de manera general. Sin embargo esta hipótesis no refleja lo que realmente sucede durante la propagación de malware. En realidad un mismo ordenador pertenece simultáneamente a diferentes redes: está recibiendo correos electrónicos, usándose para acceder a un aplicación online como *Facebook*, o moviéndose físicamente a la red de la universidad de un colaborador, con el que intercambia ficheros. El estado de infección de estos nodos es el mismo, independientemente de la red que se haya usado como vector de infección.

3.3 Modelo de propagación SIS en redes

Como se ha visto, en una red de ordenadores real los nodos infectados son los mismos en todas las capas aunque las conexiones entre ellos dependen de la capa asociada al vector de propagación del virus. La nueva contribución que proponemos para el estudio de la propagación de virus en redes complejas es el resultado de la investigación que permite definir una nueva forma de agrupar las redes multicapa y analizar de forma realista la propagación de malware. Además se pueden identificar nodos críticos en la propagación de la enfermedad. En esta sección se desarrollan brevemente las herramientas y expresiones útiles para explicar nuestro modelo de propagación de malware en redes. Está basado en modelos SIS en red, y se justificará porqué se ha usado un modelo estocástico y la razón del esfuerzo realizado para establecer y analizar la ecuación maestra como herramienta.

Empecemos analizando de manera teórica la propagación de la epidemia en una red simple, de forma que podamos generalizar y comparar en el estudio de redes multicapa. Definimos el estado de la red mediante un vector del estado de sus nodos. La red tiene N nodos y, como estamos usando un modelo SIS, el estado del nodo i puede ser susceptible (en cuyo caso asociamos al elemento del vector de estado el valor $x_i = 0$) o infectado (en cuyo caso es $x_i = 1$). Tendremos entonces que el vector $\vec{x} = (x_1, x_2, \dots, x_N)$ define el estado del sistema en un instante de tiempo (Gómez *et al.*, 2010).

La red estará definida por la matriz ω , de forma que ω_{ij} es el peso de las conexiones entre los nodos i y j , con $\omega_{ij} = 0$ si los nodos i y j no están conectados.

Un nodo puede estar únicamente en uno de esos dos estados, así que definimos el operador *flip* $f_i(\vec{x})$ que cambia el estado del nodo i de infectado a susceptible y viceversa. Esto implica que:

$$\begin{cases} x_i = 0 \implies f_i(\vec{x}) = 1 - x_i = 1, \\ x_i = 1 \implies f_i(\vec{x}) = 1 - x_i = 0, \end{cases} \quad (3.1)$$

Podemos escribir la ecuación maestra (En el anexo A se deriva la ecuación maestra de la siguiente forma:

$$\begin{aligned} \frac{\partial P(\vec{x}, t)}{\partial t} = \sum_{i=1}^N \left\{ q_i^+(f_i(\vec{x})) P(f_i(\vec{x}), t) + q_i^-(f_i(\vec{x})) P(f_i(\vec{x}), t) \right. \\ \left. - [q_i^+(\vec{x}) + q_i^-(\vec{x})] P(\vec{x}, t) \right\}, \end{aligned} \quad (3.2)$$

usando el operador flip y en la que $P(\vec{x}, t)$ es la distribución de probabilidad de que el sistema se encuentre en el estado $\vec{x}$ en el momento t ; $q_i^+(\vec{x})$ es la probabilidad de transición de que el nodo susceptible por i se infecte por al menos un nodo vecino y $q_i^-(\vec{x})$ es la probabilidad de la transición en la que el nodo infectado i pasa a ser susceptible. Además $q_i^+(\vec{x}) = 0$ si el nodo está infectado y $q_i^-(\vec{x}) = 0$ si el nodo es susceptible. El sumatorio recorre todos los nodos de la red que van de $i = 1, \dots, N$.

Las probabilidades de transición vienen dadas por:

$$q_i^-(\vec{x}) = \mu x_i, \quad (3.3)$$

siendo μ la tasa de recuperación de infectado a susceptible y

$$q_i^+(\vec{x}) = (1 - x_i) \left[1 - \prod_{j=1}^N (1 - C_{ij} x_j) \right], \quad (3.4)$$

donde C_{ij} es la matriz de contagio, que nos dice cómo es el proceso de infección. En un **proceso de contacto**¹³ (CP) cada nodo infectado reparte su capacidad de infectar entre todos sus vecinos, y la matriz de contagio C se define de forma que sus elementos vienen dados por $C_{ij} = \beta \frac{\omega_{ij}}{\omega_i}$, donde β es la tasa de infección, ω_{ij} es el peso del enlace entre el nodo i y el nodo j y $\omega_i = \sum_j \omega_{ij}$ es la fortaleza del nodo i . Otro tipo de proceso es el **proceso reactivo**¹⁴ (RP), donde la tasa de infección no disminuye con el número de vecinos, y por tanto sus elementos vienen dados por $C_{ij} = \beta$. Los dos casos los podemos expresar como (Barthélemy *et al.*, 2005):

$$C_{ij} = \beta \left(\frac{\omega_{ij}}{\omega_i} \right)^\gamma, \quad (3.5)$$

con $\gamma = 0$ en un proceso reactivo y $\gamma = 1$ en un proceso de contacto.

Si se asume que la probabilidad p_i de que el nodo i esté infectado, $p_i(t) = P(X_i = 1)$, no depende de la probabilidad de que esté infectado otro nodo j para $j \neq i$, en el caso de probabilidades pequeñas de infección, obtenemos:

$$\frac{\partial p_i(t)}{\partial t} \simeq -\mu p_i + \sum_{j=1}^N C_{ji} p_j, \quad (3.6)$$

que nos lleva (Mieghem *et al.*, 2009) a la solución general $\vec{p}(t) = \vec{p}(0)e^{(C-\mu I)t}$ con $\vec{p}(t) = (p_1(t), p_2(t), \dots, p_N(t))$.

Esta solución, aunque sea la aproximación para p_i pequeñas, puede darnos alguna visión sobre los primeros estadios de la dinámica de la epidemia en la red para un número pequeño de nodos infectados de media. El exponente de la solución determina si la enfermedad se dispara y se infectan todos los nodos de la red o por el contrario se mantiene controlada. En particular si $\lambda_{\max}$ es el autovalor mayor de C , entonces para $\lambda_{\max} < \mu$ se sigue que $p_i(t) \rightarrow 0$ cuando t crece para cualquier condición inicial $\vec{p}(0)$, ya que todos los autovalores de $C - \mu I$ son negativos: el sistema está por debajo del umbral de la epidemia (Gómez *et al.*, 2013; Klemm *et al.*, 2012). Por el contrario, si $\lambda_{\max} > \mu$ y $\vec{p}(0)$ tiene componente no nula en algún espacio

¹³En el que los infectados sólo pueden infectar con una cierta tasa como máximo a un vecino en cada momento.

¹⁴Cada nodo infectado puede infectar a todos los vecinos con una cierta tasa.

de autovectores con autovalores mayores que μ , entonces $\vec{p}(0)$ crece, por lo que estamos encima del umbral de infección. Los autovalores de C , por lo tanto, determinan si la enfermedad se propaga o termina desapareciendo. El umbral epidémico de la red corresponde al caso en el que $\lambda_{\max} = \mu$, en el que el mayor autovalor del exponente $C - \mu I$ es cero y el sistema converge hacia un estado no nulo.

El resultado que hemos obtenido coincide con análisis previos relacionados con el umbral de infección (Mieghem *et al.*, 2009), pero también tiene en cuenta la importancia del conjunto seleccionado de nodos infectados $\vec{p}(0)$ en el proceso de propagación y depende de la componente de $\vec{p}(0)$ en el espacio de autovectores de $\lambda_{\max}$, que puede ser distinto de cero en algunas ocasiones. Esta dependencia es importante para el estudio de la propagación final en términos de los nodos iniciales.

Modelo de propagación de virus en redes multicapa

” *Warning. It is usual to read into the law of large numbers things which it definitely does not imply. If Peter and Paul toss a perfect coin 10,000 times, it is customary to expect that Peter will be in the lead roughly half the time. This is not true. In a large number of different coin-tossing games it is reasonable to expect that any fixed moment heads will be in the lead in roughly half of all cases.*

— William Feller

(An Introduction To Probability Theory
And Its Applications)

En este capítulo se presenta un nuevo modelo de propagación de malware en sistemas informáticos, partiendo de modelos existentes de propagación de enfermedades en red y generalizando para virus informáticos en redes multicapa. El modelo refleja de manera realista la dinámica de propagación típica de este tipo de incidentes de seguridad. Además se analiza la propagación de malware en la red multicapa formada por la red de colaboraciones científicas de los miembros del FisEs¹ y las redes IP de las universidades e instituciones a las que están adscritos.

4.1 Modelo de propagación multicapa

Una vez se dispone de los resultados de la propagación de una enfermedad en una red formada por una sola capa y teniendo en cuenta el modelo epidémico SIS en red multicapa de Gómez *et al.* (2010) (en el que los autores

¹Congreso nacional de Física Estadística (<https://fises18.gefenol.es/>).

consideran que los nodos pueden ser diferentes en las distintas capas), hemos desarrollado un modelo con un nuevo formalismo que tiene en cuenta la propagación multivector del malware moderno. En el modelo innovador que presentamos se construye una red multicapa en la que cada capa es una red en la que los nodos susceptibles de ser infectados son los mismos que los que hay en cualquier otra capa. Como el vector de propagación es diferente en cada tipo de red, la infección se propaga en cada capa de forma diferente (salvo casos particulares). Esto se ajusta al comportamiento observado del malware multivector, que usa diferentes formas de propagación usando diferentes redes lógicas o sociales, siguiendo una estructura multicapa.

Nuestra contribución es que la descripción de la dinámica del sistema se hace en términos de una matriz efectiva de contagio (que engloba los procesos de contagio de todas las capas) de forma que tenemos una red equivalente². Esto permite calcular cualquier medida que dependa o bien del estado de los nodos o bien de los enlaces entre nodos de esa nueva red equivalente. Veremos cómo la *eigenvector centrality* permite estimar el efecto que produce en la propagación de malware eliminar de nodos o enlaces usando diferentes estrategias de inmunización para ralentizar o controlar la dinámica del contagio.

Consideremos M capas de redes formada cada una de ellas por N nodos, como ya hemos visto. N debe ser igual en todas las capas, porque los mismos nodos están presentes en cada una de ellas, aunque los enlaces sean, en general, distintos. Cada matriz de adyacencia es, en el caso general, diferente a las otras. Por ello, en lugar de la matriz de adyacencia que suele usarse para definir la estructura de red usamos un conjunto de matrices,

$$A^{(\alpha)} = \left(A_{ij}^{(\alpha)} \right)$$

con $\alpha = 1, \dots, M$, donde cada matriz A^α especifica los enlaces entre nodos en cada capa α . El estado de los nodos que tienen la misma etiqueta es el mismo en todas las redes y el cambio de estado de ese nodo en una capa cambia automáticamente el estado en todas (Fig. 4.1).

²La dinámica de la propagación de la enfermedad que se puede estudiar usando esa matriz efectiva de contagio es equivalente a la propagación en todas las capas de la red multicapa.

Fig. 4.1.: Red multi-capa en la que el estado de cada nodo es el mismo independientemente de en qué capa se esté analizando y las interacciones se realizan considerando las conexiones entre los nodos en cada capa, por lo que pueden ser diferentes. El color rojo indica que el nodo está infectado y el verde que es susceptible. Cada capa representa un tipo de relación entre nodos y por eso presenta una topología diferente.

En esta red multi-capa se estudiará un proceso de propagación epidémico SIS en el que el contagio entre nodos adyacentes en cada capa α puede propagarse de forma diferente. Se asume que la propagación epidémica en cada capa puede seguir un proceso de contacto, un proceso reactivo³ o algo intermedio (Gómez *et al.*, 2013). Para ello generalizamos la definición (3.5) a cada capa α (Barthélemy *et al.*, 2005):

$$C_{ij}^{(\alpha)} = \beta^{(\alpha)} \left(\frac{\omega_{ij}^{(\alpha)}}{\omega_i^{(\alpha)}} \right)^{\gamma_i^{(\alpha)}}, \quad (4.1)$$

donde $\omega_{ij}^{(\alpha)}$ es el peso del enlace entre el nodo i y j , y $\omega_i^{(\alpha)}$ es la fortaleza del nodo i en la capa α . La tasa $\beta_i^{(\alpha)}$ es la de contagio del nodo i en la misma capa α y $\gamma_i^{(\alpha)}$ es el parámetro que define el proceso de contagio para el nodo i , que varía de un proceso reactivo, en el que $\gamma_i^{(\alpha)} = 0$ a uno de contacto con $\gamma_i^{(\alpha)} = 1$.

Nótese que, como cada capa sigue un proceso distinto de contagio, las escalas de tiempo en cada capa (incluso en cada nodo) pueden variar considerablemente. Para tratar con este problema, se modela el sistema como un proceso de Markov microscópico en tiempo continuo, donde el

³La descripción de estos dos procesos de infección se ha realizado en el apartado 3.3.

estado del sistema se define a partir del estado de cada uno de los nodos $\vec{x} = (x_1, \dots, x_N)$, con $x_i = 0$ cuando el nodo i es susceptible y $x_i = 1$ si está infectado.

La tasa de transición para que el nodo i pase de infectado a susceptible es, al igual que en redes monocapa (ec. 3.3), $q_i^-(\vec{x}) = \mu x_i$, donde μ es la tasa de recuperación (constante) que asumimos independiente del nodo y de la capa en la que esté siendo evaluado el nodo. Esto es así porque se trata de una propiedad intrínseca del nodo y los mismos mecanismos de curación están disponibles para todos los nodos. Esta suposición puede relajarse fácilmente.

Por otro lado, la tasa de transición de susceptible a infectado también es similar a la del caso monocapa,

$$q_i^+(\vec{x}) = (1 - x_i) \left[1 - \prod_{\alpha=1}^M \prod_{j=1}^N (1 - C_{ji}^{(\alpha)} x_j) \right]. \quad (4.2)$$

solo que en esta expresión están considerados todos los posibles contagios desde los vecinos del nodo i que estén en estado infectado en todas las capas.

Con esas tasas de transición, podemos escribir el proceso de Markov en tiempo continuo con la ecuación maestra que utilizamos para redes con una sola capa (ecuación 3.2), pero teniendo en cuenta las tasas de transición del sistema multicapa.

4.1.1 Matriz efectiva de contagio

Como las variables x_j son binarias podemos hacer:

$$\prod_{\alpha} \prod_j (1 - C_{ji}^{(\alpha)} x_j) = \prod_j \prod_{\alpha} (1 - C_{ji}^{(\alpha)} x_j) \quad (4.3)$$

y por lo tanto, se puede definir una matriz efectiva de contagio que agrupa las contribuciones de las M capas

$$\bar{C}_{ij} \equiv 1 - \prod_{\alpha=1}^M (1 - C_{ij}^{(\alpha)}). \quad (4.4)$$

Podemos así reescribir la tasa de transición $q_i^+(\vec{x})$ en función de la matriz efectiva de contagio como

$$q_i^+(\vec{x}) = (1 - x_i) \left[1 - \prod_{j=1}^N (1 - \bar{C}_{ji} x_j) \right]. \quad (4.5)$$

Nótese que la expresión (4.5) no tiene dependencia explícita con la capa, por lo que $\bar{C}$ puede interpretarse como la matriz de contagio que debería desarrollar la misma dinámica que posee el nodo globalmente en un sistema monocapa equivalente.

De la expresión (4.4), para el caso de dos capas, los elementos de la matriz de contagio efectiva son

$$\bar{C}_{ij} = C_{ij}^{(1)} + C_{ij}^{(2)} - C_{ij}^{(1)} C_{ij}^{(2)}.$$

y dado que $0 \leq C_{ij}^{(\alpha)} \leq 1$, se deduce que:

$$\max(C_{ij}^{(1)}, C_{ij}^{(2)}) \leq \bar{C}_{ij} \leq \min(1, C_{ij}^{(1)} + C_{ij}^{(2)}),$$

que puede ser extendida al caso general de red multi-capa por inducción. Como consecuencia, la matriz efectiva de contagio dada por la ecuación (4.4) puede ser escrita como en la ecuación (4.1), donde los nodos verifican que todos los elementos de la fila i son $\bar{C}_{ij} < 1$ o $\bar{C}_{ij} \in \{0, 1\}$.

Hay dos casos especiales que se pueden discutir para dar cierta visión sobre el juego entre las dos capas del sistema completo. En el caso de que ambas redes tengan la misma topología (es decir, la misma matriz de adyacencia A) y la capa 1 siga un proceso reactivo con $\beta^{(1)} = 1$, la matriz de contagio efectiva pasa a ser $\bar{C} = C^{(1)} = A$. Este ejemplo, en el que una capa no tiene influencia de ningún tipo en la dinámica del sistema completo multi-capa, permite ver que los efectos del sistema no son tan simples como la suma de los efectos de cada capa.

El segundo caso es el único en el que la matriz de contagio del sistema sí es la suma de las matrices de contagio de cada red independiente: es el caso en el que las dos capas en las que la intersección de los conjuntos de enlaces

de cada capa es el conjunto vacío. Es decir, no hay enlaces comunes en las dos capas. En este caso $C_{ij}^{(1)}C_{ij}^{(2)} = 0$ para cada par de nodos del sistema.

Como la dinámica del sistema multi-capa es la misma que la de una red normal usando la matriz efectiva de contagio (4.4), se pueden usar los resultados ya conocidos que relacionan las propiedades de esa matriz con la dinámica⁴ de la propagación de la enfermedad (Mieghem *et al.*, 2009). A partir de la estructura topológica podemos determinar propiedades dinámicas de propagación de la enfermedad.

Otro resultado que podemos usar para analizar la propagación y caracterizarla es que el autovector por la izquierda que está asociado al autovalor máximo⁵ $\bar{\lambda}_{\max}$, aproxima las probabilidades esperadas de que los nodos se infecten en el límite de probabilidades pequeñas (Gómez *et al.*, 2013) y está relacionado también con la influencia dinámica de cada nodo con el resto de la red durante el proceso de contagio (Klemm *et al.*, 2012). El autovector asociado al autovalor máximo, por lo tanto, es la relación entre la dinámica y la topología.

Hagamos que $\vec{q}$ sea el vector dual por la derecha de $\vec{p}$. Para cada capa α , $\lambda^{(\alpha)}$ es el máximo autovalor de $C^{(\alpha)}$, con el autovector por la izquierda $\vec{p}^{(\alpha)}$ y autovector por la derecha $\vec{q}^{(\alpha)}$. Puede deducirse, entonces, que

$$\lambda^{(\alpha)} \leq \frac{\vec{p}^{(\alpha)} \bar{C} \vec{q}^{(\alpha)}}{\vec{p}^{(\alpha)} \vec{q}^{(\alpha)}} \leq \bar{\lambda}_{\max} \leq \frac{\vec{p} \left(\sum_{\alpha} C^{(\alpha)} \right) \vec{q}}{\vec{p} \vec{q}} \leq \sum_{\alpha} \lambda^{(\alpha)} \quad (4.6)$$

para cada capa α .

El límite inferior de esta expresión indica que no puede ser que una capa simple pueda disparar de manera independiente la propagación de la epidemia si, considerando el sistema multi-capa como un todo, la enfermedad ha muerto. El límite superior en la expresión (4.6) produce un límite inferior para el arranque de la epidemia en el sistema completo multi-capa.

⁴Es decir, relacionan la topología de la red con la dinámica.

⁵Que ya sabemos que determina el límite de propagación de la enfermedad.

4.2 Aplicación de los resultados a la red de colaboraciones de FisEs.

En este apartado se analiza el tándem formado por las redes locales (LANes) de las instituciones en las que participan los científicos y la red de colaboraciones científicas realizadas en los años 2014 y 2015 por participantes en el FisEs como un ejemplo de redes multi-capa. Se analiza y estima el efecto de infecciones multivector usando el modelo presentado en el apartado anterior.

Se construye una red en la que los investigadores científicos que participan en el congreso de FisEs son los nodos de la misma y las relaciones entre esos nodos define una red con dos capas:

1. Capa de colaboraciones científicas. Cada uno de esos nodos puede trabajar con otros científicos de otras instituciones o de las mismas. Esas relaciones llevan consigo una interacción directa con otros nodos siguiendo una estructura muy diferente a la de la capa de comunicaciones. La red social tiene más que ver con la historia y la evolución profesional de cada nodo.
2. Capa de comunicaciones IP de los sistemas informáticos asociados a los investigadores. Típicamente están conectados a las redes de comunicaciones de las instituciones a las que pertenecen. Estas redes IP, dentro de una institución, consisten en áreas de comunicaciones que permiten la conexión interna de todos los nodos con todos⁶. Las conexiones (salientes desde esas redes, pero sobre todo las entrantes) se controlan mediante medidas de seguridad perimetral. Las redes de comunicaciones internas suelen construirse mediante la interconexión de varias redes locales más pequeñas, usando para ello unos dispositivos que se llaman switches o routers IP.

Los datos de las colaboraciones científicas se obtienen a partir de la lista de asistentes a dos congresos FisEs consecutivos⁷ y se ha complementado con una capa de colaboraciones científicas que se ha construido mediante

⁶Desde un punto de vista de red lógica son buses.

⁷ En concreto en esta investigación se han usado contribuciones científicas de los años 2014 y 2015.

la búsqueda de publicaciones y colaboraciones en *Google Scholar*. Se han obtenido mediante software desarrollado por el autor que realizaba consultas a las APIs de dicha web y su procesamiento automatizado. En esta red (en adelante la *red social*) dos autores están conectados si tienen al menos una contribución común en alguno de los dos encuentros.

En la red de LANes (en adelante *red física*) dos autores están conectados si están afiliados a la misma universidad o el mismo centro de investigación. Nótese que esta red física está formada por zonas no conectadas (puede verse en la figura 4.2), que es la forma más común de implementar controles de seguridad perimetral.

Los nodos internos de cada universidad se consideran fiables y los nodos con IPs externas como posiblemente peligrosos. Esto define una estructura perimetral IP en el que ciertos nodos IP son seguros y se aceptan las conexiones entre los mismos sin restricciones y el resto se consideran externos y potencialmente peligrosos. Se controla la conexión desde dichos nodos externos usando protocolos de comunicaciones TCP⁸ o UDP⁹, además de los intentos de conexión desde ellos a nodos internos.

Ese control se realiza mediante cortafuegos o Listas de Control de Acceso¹⁰ (ACLs). Existen, además, otras herramientas tales como los Sistemas de Detección de Intrusos (IDS) que analizan el tráfico y detectan posibles ataques e infecciones mediante la identificación de patrones o sistemas heurísticos, detección de anomalías o identificación de firmas o Sistemas de Prevención de Intrusiones (IPS por sus siglas en inglés) que permiten cortar o modular el tráfico además de detectarlo. Con estas técnicas, los organismos de investigación científica evitan la gran mayoría de ataques conocidos.

Este aislamiento, que está bien definido cuando vemos sólo una capa, se rompe cuando consideramos las interacciones sociales entre alumnos, profesores e investigadores. En particular, en las colaboraciones científicas intervienen un amplio espectro de interacciones sociales, tales como correos electrónicos, reuniones virtuales o presenciales, visitas de investigación, conferencias invitadas, etc. Estas interacciones tienen como consecuencia

⁸Transmission Control Protocol

⁹User Datagram Protocol

¹⁰En el caso de que el elemento separador de redes sea un *router*

elementos tales como intercambio de ficheros por dispositivos de almacenamiento, conexión de un portátil a una red que no es la original, etc.

Tras el proceso de extracción de información obtuvimos 345 contribuciones de 687 autores distribuidas en 105 afiliaciones, dando como resultado 105 zonas independientes en la red física, la mayor formada por 39 nodos. Con respecto a la red social, hay 73 componentes conectadas, la mayor formada por 188 nodos (figura 4.5).

Una de las características más impactantes de la red acoplada¹¹ es que el número de componentes conectadas se reduce dramáticamente a 8 y la componente mayor tiene 657 nodos, casi la totalidad de los nodos considerados. Nótese que acoplando dos redes poco conectadas se obtiene una red que reduce en un orden de magnitud el número de componentes conectadas y cuya componente mayor es del orden del número total de nodos. En la figura 4.3 se pueden ver las interconexiones que la red social añade a la red física. En esa representación el tamaño del nodo indica la cantidad de colaboraciones efectuadas (durante los años 2014 y 2015) por esos miembros, el color indica la red IP asociada al centro al que están adscritos (universidades, centros de investigación) y, por lo tanto, nodos del mismo color tienen sus ordenadores y otros dispositivos conectados a las mismas redes locales. Los enlaces indican qué miembros de la comunidad han colaborado con otros miembros; el tamaño del enlace indica la cantidad de esas colaboraciones. En dicha representación gráfica se agrupan dos redes que componen la red multicapa objeto del estudio, pero los enlaces que se representan son únicamente los de la red de relaciones científicas mediante colaboraciones de los miembros del FisEs. El color indica la afiliación al centro investigador, pero no aparecen esos enlaces explícitamente.

¹¹La red multicapa considerada como un todo.

Fig. 4.2.: Representación de todas las redes IP asociadas a las instituciones de los miembros del FisEs sobre los que se estudia la propagación de virus. 105 zonas independientes, número medio de vecinos 10,815, 687 nodos y 105 afiliaciones.

Fig. 4.3.: Red social de la comunidad de científicos de FisEs. Los enlaces representan colaboraciones entre autores, el tamaño del nodo es proporcional al número de enlaces de ese nodo (grado) y el color indica la afiliación. Nodos que están conectados entre sí, pero tienen color diferente son colaboraciones entre investigadores de diferentes instituciones. El algoritmo usado para la representación gráfica tiende a agrupar los nodos entre los que hay más conexiones.

Tal y como sucede para en caso de una red de una sola capa (capítulo 3.3), el autovalor máximo $\bar{\lambda}_{\max}$ de $\bar{C}$ (de la red equivalente) está relacionado con el comienzo de la enfermedad, ya que ese autovalor determina el límite a partir del cual la enfermedad o el virus informático consigue propagarse en esa red. En la figura 4.4 se representan los valores de la diferencia entre el autovalor máximo y la tasa de recuperación $\mu = 0,5$ en la red acoplada construida tal y como se ha descrito en éste capítulo frente a las tasas de infección en ambas redes capas, como una forma de visualizar los resultados. Los valores negativos corresponden al caso en el que la red completa está bajo el umbral de infección y, por lo tanto, no hay propagación del virus.

Los valores mayores que cero corresponden a los casos en los que la red se infecta.

Fig. 4.4.: Representación de $\lambda_{max} - \mu$ de la matriz equivalente de la red compuesta del FisEs en función de las tasas de infección de cada una de las capas (β_1 en la red social y β_2 en la red física). Los valores negativos indican que estamos por debajo del valor umbral de infección (denotado por una curva de nivel de valor 0) y los positivos que estamos por encima. La tasa de recuperación se ha escogido $\mu = 0,5$.

4.2.1 Análisis topológico de la red

A continuación se representan algunas distribuciones de métricas relevantes que permiten visualizar la topología de la capa de red social estudiada. Estas métricas y otras similares permiten caracterizar la red. En esta Tesis se usarán para analizar la dinámica de la propagación de virus informáticos en redes multicapa, estudiando la relación entre algunas de estas métricas, como la distribución de grado y la centralidad, con la velocidad y efectividad de la propagación de malware.

Fig. 4.5.: Representación de las principales componentes de la red social de colaboraciones científicas de la red española FisEs. En la red completa hay 73 componentes conectadas, la mayor formado por 188 nodos.

La primera distribución que se presenta es la de grado¹² (figura 4.6), en la que se representan cuántos nodos tienen cada grado, en escala logarítmica. Hay pocos nodos con muchas interacciones de colaboración y muchos nodos que tienen pocas colaboraciones totales, siguiendo una ley de potencias.

Fig. 4.6.: Distribución de grado de la red social usada en el análisis de la propagación de malware en la red multicapa formada por las redes IPs de las instituciones a las que pertenecen los miembros de FisEs y la red de colaboraciones científicas entre esos miembros. La línea roja representa una ley de potencias de exponente -2.11.

En la figura 4.7 tenemos la conectividad media de los vecinos que determina el número mínimo de elementos (nodos o enlaces) que se necesita eliminar para desconectar el resto de nodos unos de otros. Es una medida importante de la resiliencia como red. En la figura 4.8 se representa la cercanía de los nodos frente a su conectividad. Cuanto mayor sea la cercanía de un nodo, menos saltos hay desde los otros nodos. Esta métrica permite identificar los nodos importantes como controladores del traspaso de infor-

¹²Véase el capítulo 2 para ver la definición de ésta y otras métricas

mación en este tipo de redes Ibarra y Andrews (1993). En la gráfica se ve que hay unos cuantos nodos con la *closeness centrality* elevada en la línea correspondiente a 1,00 y, en particular, hay un nodo con cerca de 22 vecinos y la centralidad muy elevada. Este nodo no sólo está más cerca de todos los otros, sino que además está directamente conectado con muchos otros científicos de la red social.

Fig. 4.7.: Conectividad promedio de los vecinos de los nodos de la red social usada en el análisis de la propagación de malware en la red multicapa formada por las redes IPs de las instituciones a las que pertenecen los miembros de FisEs y la red de colaboraciones científicas entre esos miembros. La línea roja es la aproximación a una ley de potencias de exponente 0.328.

4.3 Propagación de virus e inmunización

Analizaremos la densidad de nodos infectados en la red completa en función de las tasas de infección de las dos capas que la componen. El comportamiento global, como hemos visto, no puede deducirse simplemente de la adición de las dos capas, es decir, una única red en la que cada nodo

Fig. 4.8.: Cercanía (*closeness centrality*) de los nodos de la red social en función de su conectividad usada en el análisis de la propagación de malware en la red multicapa formada por las redes IPs de las instituciones a las que pertenecen los miembros de FisEs y la red de colaboraciones científicas entre esos miembros.

tiene la suma de los enlaces totales. Debe hacerse a través de la matriz de contagio efectiva del sistema total calculada en el apartado 4.1.1.

En la figura 4.9 se representa la densidad de nodos infectados en el estado estacionario en función de las tasas de contagio en la red social y en la red física. En la capa correspondiente a las colaboraciones científicas se ha considerado que el mecanismo de propagación social puede modelarse mediante un proceso de contacto, dado que la capacidad de infección de cada nodo infectado se divide entre sus vecinos. En la red física, por otro lado, el virus está atacando constantemente todos los nodos vecinos con la misma intensidad, independientemente de su conectividad y por lo tanto se ha modelado esta capa como un proceso reactivo.

Fig. 4.9.: Densidad (ρ) de nodos infectados en el estado estacionario de la red acoplada en términos de las tasas de contagio entre nodos en la red social (β_1) y en la red física (β_2), con $\mu = 0,5$. Las β_i son las tasas de contagio en cada una de las capas que componen la red multicapa. La densidad de nodos infectados de la red total está relacionada con el autovalor máximo de la matriz de contagio y el autovector correspondiente.

Las simulaciones se han promediado sobre 200 ejecuciones empezando por el estado completamente infectado. Se seleccionó este estado inicial para intentar evitar más frecuentemente el estado absorbente del sistema en el que todos los nodos son susceptibles a la infección.

4.3.1 Identificación de nodos críticos en la propagación de virus en redes multicapa

Como aplicación de estos resultados se va a estudiar el proceso para determinar qué nodos son críticos desde el punto de vista de la propagación de infecciones en redes multicapa, típicamente la propagación de malware multivector. Mediante la identificación de esos nodos, se pueden diseñar

estrategias de monitorización, de inmunización, o incluso redistribuir la red de forma que la propagación se haga más lenta, proporcionando tiempo para defender la red para su recuperación, corrección de vulnerabilidades mediante el despliegue de parches o simplemente desconectando los nodos críticos que son susceptibles de ser infectados. Como ya hemos visto, el criterio que $\bar{C} - \mu I$ sea cero implica que el sistema converge a un estado no-nulo para las condiciones iniciales con componentes no nulas en el autoespacio de $\lambda_{\max}$.

Se estudia el efecto de la densidad de infección ρ sobre la fracción de nodos ϕ usando dos estrategias diferentes: inmunizando los nodos más fuertes (los que tienen mayor $s_i = \sum_j \bar{C}_{ij}$) o los que tienen mayor autovalor de centralidad por la izquierda en la componente mayor (Klemm *et al.*, 2012). El procedimiento es como sigue: en cada paso se selecciona el nodo ℓ con el mayor valor de fuerza o autovector de centralidad (dependiendo de la estrategia que estemos usando)¹³ y lo inmunizamos. En este caso inmunizar quiere decir que el nodo pasa a un estado que no puede ser infectado, o lo que es lo mismo, cambiando el valor de $\bar{C}_{ij} = 0$ para todos $i = \ell$ o $j = \ell$. Finalmente, calculamos la densidad de nodos infectados ρ en el estado estacionario y se procede con el siguiente nodo.

Los resultados de ambas estrategias son muy parecidos y pueden verse en la figura 4.10. En el panel principal se compara el efecto de la inmunización en la red acoplada (mediante la matriz efectiva de contagio) y en las redes desacopladas, seleccionando tasas de contagio tales que todos los sistemas tengan la misma densidad de nodos infectados ρ para $\phi = 0$. Es importante destacar las diferencias en la selección de tasas β_1 y β_2 con el fin de alcanzar esa condición. El inserto de la figura 4.10, en la que varios procesos de inmunización se ha simulado para un valor fijo de β_2 , permite ver que la red física, con *clusters* fuertemente conectados y proceso reactivo, confiere al virus una gran capacidad de propagación y la red social mejora esa robustez con la adición de nuevos enlaces. Es decir, se infravalora la propagación del virus si se consideran las capas por separado ya que la estrategia del malware de usar varios vectores de propagación en diferentes

¹³Si hay varios nodos que tengan el mismo valor mayor, elegimos uno aleatoriamente.

Fig. 4.10.: Densidad de nodos infectados ρ , vs. densidad de nodos inmunizados ϕ . Los triángulos ($\triangle$) indican una estrategia de inmunización basada en la fortaleza de los nodos en la matriz de contagio $\bar{C}$, mientras que los círculos ($\circ$) usan la centralidad del vector por la izquierda (*left eigenvector centrality*) de $\bar{C}$. Las tasas de contagio en este caso son $\beta_1 = 1,0, \beta_2 = 0,0$ (en azul); $\beta_1 = 0,0, \beta_2 = 0,027$ (verde) y $\beta_1 = 0,18, \beta_2 = 0,02$ (rojo). El inserto corresponde a la misma situación, pero con tasas de contagio dadas por $\beta_2 = 0,2$ y $\beta_1 = 0,0$ (magenta); $\beta_1 = 0,1$ (azul); $\beta_1 = 0,2$ (verde) y $\beta_1 = 0,5$ (rojo).

redes sobre los mismos nodos permite acelerar la infección respecto a la aproximación tradicional de uso de únicamente la red de comunicaciones IP.

Parte III

Modelos estocásticos en Ecología del
Microbioma y Aplicaciones

Conceptos previos

5

” *So right away I found out something about biology: it was very easy to find a question that was very interesting, and that nobody knew the answer to. In physics you had to go a little deeper before you could find an interesting question that people didn't know.*

— **Richard Feynman**
(Físico y premio Nobel)

En la memoria de Tesis que se presenta se estudian dos problemas relacionados con el ecosistema de bacterias existente en el ser humano mediante modelos matemáticos. Concretamente, analizamos algunos procesos dinámicos de la microbiota en el pulmón humano y en el intestino. Para ello se usarán técnicas de la ecología (que es la rama de la biología que estudia las relaciones de los diferentes seres vivos entre sí y con su entorno) y, en particular, la ecología de poblaciones. Ésta estudia las poblaciones formadas por los organismos de una misma especie desde el punto de vista del tamaño de las poblaciones (número de individuos), la estructura (sexo y edad) y la dinámica (variación en el tiempo). En este capítulo se repasan conceptos previos necesarios para entender el grupo de modelos correspondientes a este apartado de la memoria de Tesis, repasando los modelos sobre los que se basan los trabajos de investigación que se presentan a lo largo de los siguientes capítulos. Se parte del modelo clásico de Lotka-Volterra al que se le añade, en primer lugar, un crecimiento de poblaciones acotado y después un término de difusión espacial. Además se adapta el sistema de Lotka-Volterra a especies que compiten. Esto sirve de base para desarrollar el modelo orientado a agentes del capítulo siguiente y los análisis de estabilidad lineal y espectral de matrices aleatorias.

5.1 Interacciones entre especies

Habitualmente no se encuentran poblaciones de especies aisladas en la naturaleza, ya que comparten entorno. La interacción entre las diferentes especies participa de manera relevante en la dinámica de esas poblaciones de forma que no puede deducirse directamente a partir de las interacciones locales. Esto hace particularmente interesante cualquier problema relacionado con interacciones ecológicas.

Fig. 5.1.: Esquema de las principales interacciones interespecíficas en una comunidad ecológica. El signo + indica que una de las especies tiene una interacción tal que se ve favorecida por la otra especie. El signo – indica que la especie se ve perjudicada y 0 que tiene un efecto neutro.

Las interacciones entre los individuos de esas poblaciones pueden clasificarse de acuerdo a los siguientes tipos (considerando únicamente las que suceden entre especies diferentes), conocidas como interacciones interespecíficas (ver el esquema de la Fig. 5.1):

1. **Depredación.** Los individuos de una especie se alimentan de los de otra. Esto sirve de mecanismo de control de crecimiento de la especie presa y de los depredadores. Un ejemplo típico es la comunidad formada por zorros y conejos. El zorro es el depredador y el conejo la presa.
2. **Competición.** Los individuos de dos o más especies comparten un determinado recurso que es limitado. Como ejemplos, la hiena y el león que se alimentan de las mismas presas, dos niños pequeños compiten por la atención de la madre o especies herbívoras por el pasto.
3. **Cooperación.** Una asociación entre dos especies en la que ambas se ven beneficiadas. Es el caso de las abejas y las flores en el proceso de polinización o el de hormigas y pulgones donde los pulgones segregan ligamaza, que sirve de alimento a las hormigas y éstas protegen a los pulgones de depredadores. Se conoce también como mutualismo.
4. **Comensalismo.** Similar al mutualismo, pero en la asociación, una de las dos especies no se ve ni beneficiada ni dañada. Los pájaros usan los árboles como refugios y los árboles, en general, no obtienen nada de esa relación. Otro ejemplo es el tiburón y el pez rémora. El primero no obtiene ni es perjudicado en nada por la presencia del segundo, mientras que el pez rémora obtiene protección y restos de alimentos.

Este tipo de interacciones se han estudiado en ecología de animales y puede aplicarse al estudio de microorganismos porque existe relación entre la estructura de la microbiota y sus funciones fisiológicas (Guarner y Malagelada, 2003; Gentile *et al.*, 2007; Chu *et al.*, 2007; Cottrell y Kirchman, 2000; Smith *et al.*, 2007) y esa estructura puede derivarse de las reglas de interacción entre los individuos que componen la comunidad y en los procesos internos.

Además, se sabe que las diferentes especies de bacterias compiten por recursos (Faust y Raes, 2012; Walter y Ley, 2011), establecen relaciones mutualistas (Faust *et al.*, 2012; Thiele *et al.*, 2013) y, en general, tienen interacciones similares a las de otros sistemas macroscópicos (Darch *et al.*, 2017), lo que justifica esta aproximación.

Por eso se utilizarán técnicas de ecología de poblaciones para estudiar la dinámica de diferentes microbiotas además de usar modelos orientado a agentes cuya formulación parte de la misma manera de las interacciones interespecíficas.

5.2 Modelos clásicos de depredación y competición

A continuación se presentan los modelos clásicos de depredación y competición, que son las dos interacciones interespecíficas más relevantes en lo que concierne a esta memoria de Tesis.

Como ya se ha visto, en la depredación, un individuo de una de las especies involucradas (identificado como depredador) se alimenta de otro individuo de otra especie (que se conoce como presa). La interacción es beneficiosa para el depredador y perjudicial para la presa, que muere en el proceso. En los modelos clásicos, el depredador puede reproducirse si se alimenta. La interacción es, por lo tanto, de la forma $(+/-)$ y ambas especies regulan mutuamente las dinámicas poblacionales. Un exceso de consumo de presas hace que escaseen y, por lo tanto, el depredador no puede reproducirse porque le cuesta alimentarse. En Volterra (1926) se propuso un modelo dinámico de un depredador y una presa, en el que si $N(t)$ es la población de la presa y $P(t)$ la población del depredador:

$$\begin{cases} \frac{dN}{dt} = N(a - bP) \\ \frac{dP}{dt} = P(cN - e) \end{cases} \quad (5.1)$$

dónde a es una tasa que define el crecimiento de las presas en ausencia de depredadores, b la tasa de muertes cuando hay depredadores, c la de crecimiento de depredadores cuando depredan a una presa y e la tasa de muertes de depredadores cuando no hay presas. Definido así, en ausencia de depredadores la presa crece de manera exponencial (Malthus, 1888). Este crecimiento viene dado en el sistema de ecuaciones por el término aN . La población de presas disminuye con el número de depredadores de la forma

$-bNP$. La población de depredadores, por otro lado, decrece de manera proporcional al número de depredadores $-eP$ y crece cuantas más presas haya, cNP . En este modelo no hay tasa de mortalidad de las presas cuando no son depredadas y tampoco crecimiento de los depredadores si no hay presas

Sin embargo, este modelo sigue siendo demasiado simple. Que en ausencia de depredadores las presas crezcan sin límite no sucede en la naturaleza. Los recursos, el espacio y otros parámetros determinan cómo crece esa población. Existen formas más realistas de crecimiento como, por ejemplo, la dada por una ecuación logística (Verhulst, 1838) donde los individuos de una misma especie compiten por los recursos de su entorno. En este caso la población crece hasta alcanzar un valor relacionado con su capacidad de carga K o *carrying capacity*. Introduciendo la capacidad de carga en las ecuaciones, tenemos:

$$\begin{cases} \frac{dN}{dt} = aN \left(1 - \frac{N}{K}\right) - bNP \\ \frac{dP}{dt} = cNP - eP \end{cases} \quad (5.2)$$

Para que el modelo sea más realista se pueden incluir otros tipos de interacciones entre diferentes especies. De hecho una forma más general de formular modelos de esta índole es considerar que las interacciones pueden ser de varios tipos. Es decir, dos especies pueden competir por un recurso común limitado en ciertas condiciones, y en otras en las que el recurso es abundante, cooperar.

Para seguir avanzando en la construcción del modelo introducimos una interacción interespecífica distinta a la depredación. En particular hablamos de la competición. La competición entre especies aparece cuando uno de los tipos de individuos sufre una reducción en su capacidad de reproducción o de crecimiento ya sea como efecto directo de una limitación de recursos finitos que necesita para reproducirse o bien porque hay interferencia directa entre especies. Si la cantidad de recursos disponibles es suficientemente grande,

los efectos de competición no son relevantes¹. Cuando los recursos son escasos, la competición en la comunidad ecológica resulta ser el mecanismo controlador de las poblaciones.

Volterra añadió a su sistema de ecuaciones la competición interespecífica mediante la adición de un término nuevo que depende del coeficiente conocido como de competición a_{ij} . Este coeficiente² define el efecto de la especie j sobre el crecimiento de la especie i (Volterra, 1927). De esta manera podemos escribir el sistema de Lotka-Volterra que incluye competición de una manera muy compacta:

$$\frac{dN_i}{dt} = N_i r_i \left(1 - \frac{N_i}{K_i} - a_{ij} \frac{N_j}{K_i} \right). \quad (5.3)$$

donde K_i define la capacidad de carga de la especie i y establece el límite de crecimiento de su población. Las distintas N_i son las poblaciones de las especies dadas por i y finalmente r_i es la constante de crecimiento de la especie i en ausencia de la otra especie.

Para describir una interacción genérica, podemos escribir:

$$\frac{dX_i}{dt} = X_i \left(r_i - s_i X_i + \sum_{j=1, i \neq j}^S a_{ij} X_j \right) \quad (5.4)$$

donde X_i es la población de la especie i (en nuestro caso la bacteria de tipo i), r_i es la tasa de crecimiento de la especie dada por i , a_{ij} nos da la interacción entre las especies i y j y, por lo tanto, s_i es la interacción de la especie consigo misma. Mediante la selección adecuada del signo de los a_{ij} podemos describir interacciones de varios tipos. Cuando $a_{ij} = -a_{ji}$ tenemos una interacción de depredación. Si ambos son negativos tenemos competición y en el caso de que ambos sean positivos será cooperación.

¹Esto es lo que se observa en los modelos durante los primeros instantes de tiempo, ya que hay recursos suficientes.

²El término de interacción a_{ij} puede interpretarse como un factor de conversión. Los individuos de la especie j se comportan como los de la especie i en un grado dado por a_{ij} , expresando los individuos de la segunda especie en unidades de la primera.

Con este sistema de ecuaciones se recoge tanto la interacción de tipo depredador-presa como la competición, tanto entre diferentes especies como la existente entre individuos de la misma especie. Es el punto de partida de los modelos que se presentan en la Tesis y se completarán incluyendo los términos de difusión espacial.

5.3 Introducción a agent-based modeling (ABM)

Muchos problemas biológicos consisten en controlar un sistema complejo de forma que, dadas ciertas condiciones de contorno, se desea encontrar una solución que optimice alguna variable del problema: minimizar el impacto de una medicina en el tejido humano, maximizando el efecto sobre cierta bacteria; controlar un ecosistema de forma que se maximice la población de una cierta especie; minimizar el efecto sobre la tensión usando medicamentos para tratar dolores de cabeza; maximizar el efecto de un pesticida, de forma que se minimice el daño a la planta que están protegiendo, o al sujeto que luego ingiere el alimento. La necesidad de un control óptimo aparece casi en cada problema.

La forma más común de aproximación a este tipo de problemas es mediante el uso de modelos matemáticos, en los que se definen una o más ecuaciones diferenciales (ordinarias o parciales). En estas ecuaciones aparecen las variables más relevantes del problema, así como los parámetros de control que caracterizan el conjunto de soluciones. El modelo matemático es una simplificación o abstracción del sistema real y evidentemente no es posible representar todas esas variables reales en las ecuaciones, por lo que la correcta elección del modelo y las variables es crítica.

Una de las limitaciones de los modelos basados en ecuaciones diferenciales es que éstas son continuas, mientras que ciertos procesos biológicos son intrínsecamente discretos. Por ejemplo, en los procesos en los que se analiza el crecimiento de poblaciones el número de individuos no puede ser una variable continua (no tiene sentido hablar de 1.3 individuos). En el caso del estudio de procesos de depredador-presa las interacciones son discretas:

un depredador come a un número entero de presas. También es cierto que para grandes números y en ciertos procesos la aproximación continua es suficiente. Esto sucede, por ejemplo, si nos interesan tasas de crecimiento o si el número de individuos es lo suficientemente grande como para que pueda ser considerado un continuo. Pero no siempre es posible asumir que hay un gran número de aquéllos. En el caso de que se usen muchas especies diferentes, la única forma de resolver las ecuaciones es integrando numéricamente, que requiere adicionalmente un proceso de discretización.

Por otro lado, este tipo de modelos permiten capturar el hecho de que las propiedades emergentes de sistema completo aparecen a partir de reglas locales de interacción. La definición y presentación de los modelos puede ser más intuitiva para perfiles científicos menos matemáticos (Hinkelmann *et al.*, 2011) pero que conocen profundamente el comportamiento de los agentes en estudio. Esto facilita la colaboración en grupos interdisciplinarios, como ha sido el caso de varios modelos que se presentan en esta Tesis. Tradicionalmente ha sido utilizado en ecología y biología en estudios de propagación de enfermedades (Situngkir, 2004) y más recientemente integrando además redes complejas (Niazi, 2017; Kurve *et al.*, 2013). Sin embargo es difícil establecer un marco de referencia analítico para los modelos estocásticos orientados a agente. Los modelos orientados a agentes tienen una dureza computacional intrínseca y simular un amplio espectro de valores para extraer conclusiones es muy costoso y a veces simplemente inabarcable con los recursos de simulación disponibles.

También hay que tener en cuenta que en muchos modelos biológicos es más sencillo definir las interacciones desde el punto de vista del individuo.

El modelado orientado a agentes se está extendiendo como herramienta de análisis de comportamiento de diversos sistemas biológicos (Grimm y Railsback, 2005) como alternativa al sistema de ecuaciones diferenciales. Son modelos computacionales que permiten simular el comportamiento de unas unidades básicas conocidas como agentes. Éstos se definen, típicamente, a partir de un conjunto de propiedades y unas funciones que modifican esas propiedades. Las propiedades pueden ser (como en este caso de estudio) el tipo de bacteria, su capacidad para desplazarse, tamaño, posición y el estado anterior del agente o de otros agentes.

Las interacciones y la evolución dinámica de los agentes se modelan usando funciones propias que modifican sus propiedades mediante reglas lógicas, que pueden ser deterministas o probabilísticas. Cada agente tiene, pues, un número finito de posibles estados y las reglas deciden en qué estado está. En sistemas complejos y en particular en los modelos biológicos estas reglas son típicamente estocásticas (están gobernadas por una colección de probabilidades).

Los primeros modelos de agentes aparecen en Von Neumann (1951). En 1970 aparecen los autómatas celulares (Gardner, 1970). Uno de los más famosos es el juego de la vida (Conway, 1970), del que se presenta una imagen en la figura 5.2, que apareció en la columna de juegos matemáticos de *Scientific American* en 1970. Es un ejemplo de emergencia y autoorganización en el que aparecen patrones complejos a partir de reglas sencillas. En la versión original se crea un tablero dividido espacialmente en cuadrícula. Cada cuadrado representa una célula. Si una célula está rodeada por tres vecinos inmediatos vivos, pasa al estado vivo, y si mantiene dos o tres vecinos vivos se mantiene viva en cada paso. Si tiene menos, muere. Todas las células se actualizan simultáneamente en cada paso. La razón por la que estos modelos aparecen tan tarde en la historia de la ciencia es porque se requieren capacidades de computación muy elevadas, capacidades que están ahora disponibles generalizadamente.

Fig. 5.2.: Juego de la Vida de John Conway <https://bitstorm.org/gameoflife/>.

Fibrosis quística

6

” *La naturaleza nos es hostil porque no la conocemos: sus crueldades representan la venganza contra nuestra indiferencia.*

— **Ramón y Cajal**
(Médico y premio Nobel de Medicina)

El estudio que se presenta en este capítulo de la Tesis es fruto de una colaboración con miembros del Servicio de Microbiología, Hospital Universitario Ramón y Cajal y con el Instituto Ramón y Cajal de Investigación (Dios *et al.*, 2017). En dicho estudio proponemos la introducción en la microbiota del pulmón de depredadores específicos que regulen la población de las bacterias asociados a la enfermedad de Fibrosis Quística (FQ) como una posible aproximación terapéutica de las reagudizaciones respiratorias asociadas. En dicha investigación, también, se destaca la existencia de depredadores procariontes como parte de la microbiota pulmonar de la FQ y su contribución a la regulación de la población bacteriana. Desarrollamos modelos matemáticos basados en modelos clásicos de depredador-presa además de un modelo orientado a agentes para analizar la posibilidad de uso de esos bacteriófagos como depredadores de ciertas especies de bacterias relacionadas con la sintomatología de la FQ. Estos

6.1 Fibrosis quística y su relación con la microbiota

La FQ es una enfermedad genética que aparece como consecuencia de la mutación en el gen CFTR. Este gen tiene como función sintetizar la proteína llamada reguladora de la conductancia de la transmembrana de la fibrosis

quística¹. Esta proteína funciona como un canal de iones a través de la membrana de las células mucosas² (producen moco, sudor, saliva, lágrimas y enzimas digestivas). El canal transporta iones de cloruro a través de la membrana de las células regulando a su vez por ósmosis la cantidad de agua que permanece en los tejidos. La alteración del funcionamiento normal del canal CFTR hace que enfermos de FQ tengan un desequilibrio electrolítico a que el moco sea más deshidratado y espeso, lo que es responsable de la sintomatología de la enfermedad. La enfermedad respiratoria causa la muerte de al menos el 80 % de los pacientes de FQ por diversas complicaciones, casi todas ellas asociadas a procesos infecciosos.

La evolución natural de esta enfermedad es una disminución progresiva de la función pulmonar causada por un círculo vicioso de inflamación y destrucción tisular, desencadenado y mantenido por la colonización bacteriana crónica del tracto respiratorio inferior (Ramsey *et al.*, 2011). En particular, los datos clínicos³ reflejan esa colonización en los pacientes de FQ de *Pseudomonas aeruginosa*. Su adquisición determina la supervivencia final y la calidad de vida del paciente (Gilligan, 2014), ya que esta colonización es el principal motor de la inflamación pulmonar.

Además de la *P. aeruginosa* se han encontrado otras bacterias clínicamente cultivables y relevantes, tales como las especies complejas de *Burkholderia cepacia*, *Staphylococcus aureus* (resistente a la meticilina) y *Mycobacterium abscessus*. La introducción de herramientas independientes de cultivo, en particular las basadas en la secuenciación de próxima generación (NGS), ha permitido identificar una microbiota pulmonar más compleja, diversa y abundante que incluye no sólo esos patógenos, sino también una amplia comunidad de taxones bacterianos, siendo la mayoría de ellos incultivables (Bacci *et al.*, 2016; Cameron *et al.*, 2016; Carmody *et al.*, 2013; Rudkjøbing *et al.*, 2011). Aunque el papel de la microbiota en el curso clínico de la enfermedad aún no ha sido aclarado, se ha sugerido su implicación en el inicio de las exacerbaciones⁴.

¹Cystic Fibrosis Transmembrane Conductance Regulator.

²Véase el anexo D en el que se presenta una introducción a conceptos de microbiología.

³Utilizando técnicas clásicas de cultivo.

⁴La exacerbación es el aumento transitorio de la gravedad de un síntoma o de una enfermedad.

Fig. 6.1.: Distribución de *phyla* en los 15 paciente de FQ. La última columna (*) representa la media de los valores de las 56 muestras (Dios *et al.*, 2017).

Fig. 6.2.: Distribución de *phyla* en los 15 paciente de FQ. La última columna (*) representa la media de los valores de las 56 muestras (Dios *et al.*, 2017)

La composición de la microbiota es específica de cada uno de los pacientes y la reducción de su complejidad y diversidad se asocia a menudo con la progresión de la enfermedad. Las alteraciones de la microbiota pulmonar de los pacientes de FQ están relacionadas, principalmente, con la acumulación de múltiples tratamientos antibióticos (véase el anexo D), aunque existen otros factores como el sistema inmunológico, los metabolitos humanos, las moléculas de detección de *quórum sensing* bacteriano⁵ (véase, por ejemplo, Huang y LiPuma, 2016). En la figura 6.1 se representa la distribución de *phylum*⁶ en los paciente de FQ estudiados durante dicho trabajo y que se ha usado en el diseño de los modelos matemáticos y de agentes.

Proponemos la introducción de depredadores de *P. aeruginosa* (Sherrard *et al.*, 2016; Caverly *et al.*, 2015; Bernasconi *et al.*, 2016; Rolain *et al.*, 2011) en el ecosistema del pulmón de ciertas bacterias como moduladoras de factores de virulencia⁷ como una estrategia ecológica para el control de las bacterias patógenas de los pulmones del enfermo de FQ.

Como posible depredador, la especie más estudiada es *Bdellovibrio bacteriovorus*, ampliamente distribuida en ecosistemas ambientales y particularmente vinculada a entornos de agua salada, para la que se ha sugerido un papel ecológico para la regulación de comunidades bacterianas. También se ha descrito recientemente la presencia de *B. bacteriovorus* en la microbiota intestinal de sujetos sanos y pacientes con FQ, así como su capacidad para depredar los patógenos clásicos de FQ como *P. aeruginosa*, *S. aureus* o *S. maltophilia* y formación tanto a nivel planctónico como en formación de biofilm⁸ (Iebba *et al.*, 2013). La *Bdellovibrio spp.* (gramnegativa δ -*Proteobacteria* aerobia, móvil, muy pequeña con un diámetro de 0,6 μ) es depredador obli-

⁵Es el mecanismo de regulación de la expresión genética en respuesta a la densidad de población celular, una forma de comunicación bacteriana.

⁶Término que se refiere a la clasificación taxonómica, véase el anexo D en el que se describen éste y otros términos propios de la Microbiología.

⁷Los factores de virulencia son los mecanismos con los que cuenta un microorganismo para poder entrar al organismo, invadir los tejidos y provocar enfermedad.

⁸Un biofilm es una estructura colectiva de microorganismos que se adhiere a superficies vivas o inertes. Suelen tener características funcionales y estructuras complejas. Las bacterias, cuando crecen en formación de biofilm, son capaces de resistir a los biocidas y a los antibióticos de un modo más eficaz que aquellas que viven como organismos libres y soportan dosis considerablemente mayores de productos antimicrobianos.

gado de bacterias gramnegativas y grampositivas (Jurkevitch, 2006, Iebba *et al.*, 2014).

El proceso biológico consiste en que, tras el contacto inicial, el depredador se adhiere a la envoltura bacteriana y entra en el espacio periplásmico donde forma sus estructuras replicativas llamadas bdelloplastos usando el citoplasma de la presa como fuente de energía (Guerrero *et al.*, 1986). Una vez que el citoplasma se ha agotado, la lisis de la célula remanente permite la salida de múltiples células de *B. bacteriovorus* al medio. El desplazamiento espacial juega un papel importante en el modelado ya que el depredador busca y localiza las presas y se desplaza a gran velocidad siguiendo las trazas de las presas.

Otra bacteria de interés es el *Vampirovibrio chlorellavorus*, un depredador bacteriano cuya única presa conocida es la microalga *Chlorella vulgaris* (Ganuza *et al.*, 2016 y Soo *et al.*, 2015).

La distribución de estas dos bacterias *B. bacteriovorus* y *Vampirovibrio* en los 15 pacientes estudiados se representa en la figura 6.2, en la que se compara en diferentes estadios del nivel de avance de la enfermedad en esos pacientes, ordenados de leve a grave.

6.2 Modelo con dos tipos de bacterias para FQ

El uso de bacterias como una estrategia ecológica para el control de las bacterias dañinas puede ser formulado como modelo de Lotka-Volterra para depredación. El modelo se construye partiendo de las ecuaciones (5.2) y considerando, además, que las especies se desplazan espacialmente. Para ello incluimos un nuevo término espacial en el sistema de ecuaciones planteado en los conceptos previos, pero teniendo en cuenta el proceso de difusión espacial de las especies con lo que se obtiene un sistema típico de reacción-difusión:

$$\begin{cases} \frac{dN}{dt} = aN \left(1 - \frac{N}{K}\right) - bNP - D_N \nabla^2 N \\ \frac{dP}{dt} = cNP - dP + D_P \nabla^2 P \end{cases} \quad (6.1)$$

considerando que los términos de difusión se han limitado a dos dimensiones, es decir $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

Para analizar la ecuación (6.1) y realizar las simulaciones es conveniente reducir el número de parámetros de los que dependen las ecuaciones. Para ello se hacen sustituciones en las variables re-escalando las mismas. Si hay dos o más dimensiones, las variables deben re-escalarse de manera independiente. Asumimos que el espacio es isotrópo (no hay diferencia entre las diferentes direcciones espaciales) lo que nos permite reescribir las ecuaciones como:

$$\begin{cases} \frac{dN}{dt} = N(1 - N - P) + D\nabla^2 N \\ \frac{dP}{dt} = aP(N - b) + \nabla^2 P \end{cases} \quad (6.2)$$

En el proceso de análisis de la dinámica del sistema se estudia la estabilidad lineal⁹ del sistema. La estabilidad se considera crítica para la salud del hospedante (Gill *et al.*, 2006; Lozupone *et al.*, 2012) y, adicionalmente, se sabe que grandes cambios en la composición de la comunidad de bacterias está relacionada con enfermedades (Coyte *et al.*, 2015; De Cruz *et al.*, 2015). Aunque la estructura de esa microbiota es distinta para diferentes personas, los grandes cambios en las poblaciones son los que se relacionan la disbiosis.

Es conveniente estudiar el equilibrio de la ecuación (6.2) en dos fases, en la primera se parte de las ecuaciones sin difusión (sin la parte espacial, que se corresponde con las ecuaciones (5.2)) y en la segunda se analizará el efecto de la difusión sobre la estabilidad de dichos puntos, siguiendo el método expuesto en el anexo de estabilidad (ver apéndice B).

⁹Véase el anexo B en el que definen estos términos, además del proceso de análisis de estabilidad lineal.

6.2.1 Estudio del equilibrio sin difusión

Los puntos de equilibrio son aquellos que hacen que, en nuestro caso:

$$\begin{cases} 0 = N(-P - N + 1) \\ 0 = aP(-b + N) \end{cases} \quad (6.3)$$

Cuyas soluciones son los siguientes puntos de equilibrio:

$$(N, P) = \begin{cases} (0, 0) \\ (1, 0) \\ (b, -b + 1) \end{cases} \quad (6.4)$$

Para determinar el tipo de estabilidad de estos puntos, analizamos los autovalores del determinante del jacobiano evaluado en esos puntos, que da una visión de cómo, en el equilibrio, reacciona el sistema frente a perturbaciones pequeñas. Si la parte real de los autovalores es menor que cero, la comunidad es capaz de volver al equilibrio cuando hay una perturbación pequeña. Esto es más rápido cuanto más negativa es esa parte real. El jacobiano correspondiente a las ecuaciones (5.2) es

$$J(N, P) = \begin{pmatrix} -P - 2N + 1 & -N \\ aP & a(-b + N) \end{pmatrix}. \quad (6.5)$$

Los puntos que cumplen la condición (6.3) son:

- $(N, P) = (0, 0)$. Los autovalores del jacobiano son $(1, -ab)$. Son ambos reales y, como a y b son siempre positivos, uno de los autovalores es positivo y el otro negativo. Esto quiere decir que el punto $(0, 0)$ es inestable.
- $(N, P) = (1, 0)$. Los autovalores del jacobiano evaluado en este punto de equilibrio son $(a(a - b), -1)$. Si $b > a$ es un punto de equilibrio inestable (un autovalor real es positivo y el otro negativo). Es el caso en el que la población de las presas alcanza su máximo dado por la capacidad de carga. Se representa el espacio de fases y algunas trayectorias asociadas a diferentes valores iniciales en la figura 6.3.

Fig. 6.3.: Espacio de fases con trayectorias para varias condiciones iniciales al rededor del punto de equilibrio $(1, 0)$. El estado final es tal que se alcanza la *carrying-capacity* de la población de presas y los depredadores tienden a cero.

- $(N, P) = (b, 1 - b)$. Punto mucho más interesante que los anteriores. En este caso los autovalores correspondientes al punto de equilibrio son

$$\frac{-b - \sqrt{b(4ab - 4a + b)}}{2}, \frac{-b + \sqrt{b(4ab - 4a + b)}}{2}$$

En este caso hay dos posibilidades. Si el radicando es menor que cero, tenemos dos autovalores complejos conjugados, cuya parte real es negativa, por lo que el punto de equilibrio sería un foco estable. O en el caso contrario, tendremos dos autovalores reales negativos que indican que se trata de un nodo estable. En ambos casos el punto es local y asintóticamente estable. Se representan varias trayectorias en el espacio de fases en la figura 6.4 para el caso de autovalores complejos conjugados y se representan las trayectorias en el espacio de fases asociadas a un conjunto de valores iniciales de poblaciones tienden al punto estable.

Los casos de equilibrio se pueden resumir tres situaciones:

Fig. 6.4.: Espacio de fases con trayectorias para varias condiciones iniciales al rededor del punto de equilibrio $(b, 1 - b)$. En este caso buscamos la estabilidad asintótica, que se obtiene al hacer que se cumpla la condición de estabilidad, es decir, $4a \leq \frac{b}{1-b}$.

1. Desaparecen las dos poblaciones
2. La población de presas alcanza su máximo dado por su capacidad de carga y los depredadores desaparecen.
3. El sistema tiende a una solución en la que las poblaciones de ambas especies oscila de forma cada una controlada por la otra, de forma que coexisten, pero disminuyendo progresivamente sus poblaciones.

6.2.2 Estudio del equilibrio con difusión

Un vez hemos determinado el tipo de puntos de equilibrio estables en el sistema de ecuaciones (6.3), podemos pasar a la segunda fase del análisis, considerando el efecto que produce añadir un término de difusión¹⁰, con lo que el jacobiano queda:

$$\tilde{J} = \begin{bmatrix} -D\omega^2 - f - 2r + 1 & -r \\ af & a(-b + r) - \omega^2 \end{bmatrix} \quad (6.6)$$

- $(N, P) = (0, 0)$. En este punto de equilibrio, los autovalores de (6.6) evaluada en el punto $(0, 0)$ son $(D\omega^2 + 1, D\omega^2 - ab)$. Como ω^2 es siempre positivo, así como D , uno de los autovalores es real y positivo y el otro será real y positivo si $D\omega^2 > ab$ y negativo en caso contrario. Es decir, al introducir la difusión espacial, el punto que era inestable sigue siendo inestable, pero puede pasar a ser un punto de silla en ciertas condiciones.
- $(N, P) = (1, 0)$. Los autovalores son ahora:

$$(D\omega^2 - 1, D\omega^2 - ab + a).$$

Sigue siendo inestable, salvo en el caso particular en el que se den simultáneamente las condiciones $D\omega^2 < 1$, $ab < 1$ y $D\omega^2 < a(1 - ab)$, en cuyo caso ambos autovalores son reales y negativos. Es decir, existe un conjunto de condiciones con las que el punto puede ser estable al añadir difusión espacial.

¹⁰El desarrollo matemático que justifica esta técnica se desarrolla en el anexo de estabilidad, capítulo B.

- $(N, P) = (b, 1 - b)$. En este punto los autovalores son

$$\left(D\omega^2 - \frac{b}{2} - \frac{1}{2}\sqrt{b(4ab - 4a + b)}, \quad D\omega^2 - \frac{b}{2} + \frac{1}{2}\sqrt{b(4ab - 4a + b)} \right).$$

Recuperamos las dos mismas situaciones que para el sistema sin difusión: si el radicando es negativo obtenemos dos autovalores complejos conjugados o bien si el radicando es positivo hay que tener en cuenta que aparecen dos subcasos. Si

$$D\omega^2 < -\frac{b}{2} - \frac{1}{2}\sqrt{b(4ab - 4a + b)}$$

y

$$D\omega^2 < -\frac{b}{2} + \frac{1}{2}\sqrt{b(4ab - 4a + b)}$$

ambos autovalores serán negativos y el sistema es estable, e inestable en caso contrario.

6.3 Modelo con más de dos especies para FQ

La generalización a más de dos especies proporciona un nivel de realismo mayor al modelo. Para ello usamos la ecuación (5.4), que podemos escribir de manera más compacta (debido a la complejidad de la solución, no se tiene en cuenta el proceso de difusión en este modelo):

$$\frac{dX_i}{dt} = X_i \left(r_i + \sum_{j=1}^S a_{ij} X_j \right), \quad (6.7)$$

en la que se definen las a_{ij} de forma que la interacción sea de depredación presa, es decir, que $a_{ij} = -a_{ji}$. En el modelo que se presenta se usan cinco especies de bacterias, dos que actúan como depredadores y tres como presas, que se corresponden con *Bdellovibrio* y *Vampirovibrio* (depredadores) y *Pseudomonas*, *Staphylococcus* y *Haemophilus* (presas). Representan los tipos de bacterias que se han identificado como posibles bacteriófagos en el pulmón y las que aparecen experimentalmente en mayor proporción en los pacientes de FQ (ver figuras 6.1 y 6.2) y cuyas poblaciones vienen dadas por X_i . El resto de parámetros son:

- a_i es la tasa de reproducción de la presa i ,
- b_{ij} la interacción de la presa de tipo i con el depredador de tipo j ,
- c_{ij} la interacción del depredador i con la presa j
- d_i la tasa de muerte del depredador cuando no hay presas.

Estudiando analíticamente el equilibrio del sistema de ecuaciones para cinco bacterias, aparecen dos grandes grupos de soluciones, que se presentan a continuación como caso 1 y caso 2.

6.3.1 Caso 1: una especie de depredador y una de presa.

El sistema tiende a una situación en la que sólo queda una especie de depredador y una de presa. Para ilustrar la dinámica se representan una sección del espacio de fases en la fig. 6.5. En esta figura se representan dos trayectorias en el espacio de fases de las poblaciones de un depredador y una de las tres presas ($i = 1$), para dos valores distintos de condiciones iniciales. Podemos observar como una de las trayectorias termina en el caso trivial en el que mueren tanto las presas como el depredador (trayectoria en azul). La otra trayectoria (en rojo) corresponde al caso en el que el depredador sobrevive y su población aumenta y disminuye cíclicamente. Esto es posible porque hay otras especies de presas y el depredador continúa alimentándose de una de las especies restante. Para ver cómo el depredador se alimenta de otras presas se representa la población de ese mismo depredador respecto a otra de las especies de presas ($i = 2$) en la figura 6.6, donde puede verse el ciclo límite hacia el que tiende la población del depredador, recuperando el ciclo de oscilaciones típico de un sistema de Lotka-Volterra de dos especies.

Fig. 6.5.: Espacio de fases con trayectorias para la presa dada por $i = 1$ y una de las especies de depredadores $j = 2$ dado por la ecuación (6.7), caso 1, $a_1 = 0,52$, $a_2 = a_3 = 0,5$, $b_{11} = b_{12} = b_{31} = 0,03$, $b_{21} = b_{22} = b_{32} = 0,02$, $d_1 = 0,4$, $d_2 = 0,42$, $c_{11} = 0,006$, $c_{21} = 0,0042$, $c_{21} = c_{22} = c_{23} = 0,006$ con 400 individuos como población total, siguiendo una distribución de valores iniciales con proporciones (0,7, 0,2, 0,05, 0,04, 0,01) respectivamente.

Fig. 6.6.: Espacio de fases con trayectorias para la presa dada por $i = 2$ y una de las especies de depredadores $j = 2$ en la ecuación (6.7), caso 1, $a_1 = 0,52$, $a_2 = a_3 = 0,5$, $b_{11} = b_{12} = b_{31} = 0,03$, $b_{21} = b_{22} = b_{32} = 0,02$, $d_1 = 0,4$, $d_2 = 0,42$, $c_{11} = 0,006$, $c_{21} = 0,0042$, $c_{21} = c_{22} = c_{23} = 0,006$ con 400 individuos como población total inicial.

Podemos visualizar el comportamiento global del sistema en la figura 6.7, en la que se ha representado la evolución temporal de las poblaciones de todas las especies del modelo mediante simulación numérica. Las poblaciones de todas las presas desaparecen, salvo una (y lo mismo sucede para los depredadores). Es decir, en este ejemplo y para estas condiciones iniciales, se recupera la situación del ciclo típico de Lotka-Volterra, en el que las poblaciones de una presa y de un depredador están en equilibrio dinámico, oscilando desfasadas entre sus máximos y mínimos. El equilibrio se alcanza para un gran número de condiciones iniciales cuando sólo queda una especie de presa y una de depredador y el resto mueren, así que es un fenómeno muy común en las simulaciones. Precisamente este comportamiento es el que se quiere reproducir en el pulmón en el caso de enfermos de FQ para reducir la población de la bacteria asociada a los síntomas de FQ introduciendo un depredador natural de ésta.

Fig. 6.7.: Población de especies de presas y depredadores a lo largo del tiempo, pasos temporales de la simulación, caso 1, $a_1 = 0,52$, $a_2 = a_3 = 0,5$, $b_{11} = b_{12} = b_{31} = 0,03$, $b_{21} = b_{22} = b_{32} = 0,02$, $d_1 = 0,4$, $d_2 = 0,42$, $c_{11} = 0,0042$, $d_2 = 0,006$ con 400 individuos como población total, siguiendo una distribución de valores iniciales (0,7, 0,2, 0,05, 0,04, 0,01) siendo los tres primeros valores los correspondientes a las especies presa etiquetados con los valores de i de 0 a 2 y los dos últimos los depredadores con $j = 0$ o $j = 1$.

6.3.2 Caso 2: un depredador en equilibrio dinámico con dos presas.

El segundo caso que se va a estudiar es el que un depredador permanece en equilibrio dinámico con dos presas. De nuevo se representan dos trayectorias en el espacio de fases (en la figura 6.8), correspondientes a las poblaciones de un depredador y de la presa dada por el índice $i = 1$. Podemos observar que ambas trayectorias terminan en los casos en los que muere la presa y el depredador muere o bien muere la presa y el depredador sigue consumiendo otras presas. El resultado es similar al del caso 1. Ahora bien, si miramos las trayectorias correspondientes al mismo depredador respecto a las otras dos presas (figura 6.9) la población tiende a un ciclo límite con

cada una de las especies restantes de forma que el máximo poblacional de las presas tiene distinto tamaño (en el caso general).

Fig. 6.8.: Espacio de fases con trayectorias para la presa dada por $i = 0$ y una de las especies de depredadores $j = 4$ dado por el conjunto de ecuaciones ??, caso 2: $a_1 = 0,52, a_2 = a_3 = 0,5, b_{11} = b_{22} = b_{32} = 0,03, b_{21}, b_{12} = b_{31} = 0,02, d_1 = 0,4, d_2 = 0,42, c_{11} = c_{21} = 0,0042, c_{12} = c_{13} = 0,03, c_{21} = c_{22} = 0,0032$ para los depredadores.

Fig. 6.9.: Espacio de fases en el caso 2, representando las trayectorias para las presas dada por $i = 2$ (arriba) e $i = 3$ (abajo) y una de las especies de depredadores $j = 4$. Valores de los parámetros para las ecuaciones (6.7): $a_1 = 0,52, a_2 = a_3 = 0,5, b_{11} = b_{22} = b_{32} = 0,03, b_{21}, b_{12} = b_{31} = 0,02, d_1 = 0,4, d_2 = 0,42, c_{11} = c_{21} = 0,0042, c_{12} = c_{13} = 0,03, c_{21} = c_{22} = 0,0032$. 400 individuos. Cada trayectoria tiene asociados unos valores iniciales de población indicados en las gráficas.

Para entender mejor la dinámica se representan las poblaciones de todas las especies respecto al tiempo (en unidades arbitrarias) en la figura 6.10. Estos resultados han sido obtenidos numéricamente. Y con el fin de complementar la imagen visual en ambos casos de estudio, se ha representado en la figura 6.11 el espacio de fases para un depredador y dos pares de presas en una proyección 3D, desde dos puntos de vista diferentes. En naranja y azul se representan trayectorias en el espacio de fases correspondientes a dos pares de depredador-presa. La trayectoria naranja tiende a una situación en la que no hay presas y la población de depredador oscila entre su máximo y su mínimo. Esto puede suceder porque ese depredador está siendo alimentado de otra presa distinta. Cuando representamos ese mismo depredador respecto a otra presa, en azul, vemos como converge a un ciclo límite, en el que se mantiene en equilibrio con la presa en un ciclo típico de Lotka-Volterra de dos especies. Obsérvese que tiene proyección sobre las poblaciones de dos presas distintas de cero (en un eje va aproximadamente entre 0 y 100 y en el otro entre 0 y 30). Si el depredador dependiera de una sola presa, el ciclo estaría en un plano cuya proyección sería cero sobre los ejes correspondientes a las otras presas.

Fig. 6.10.: Población de especies de presas y depredadores a lo largo del tiempo en el caso 2, pasos temporales de la simulación, $a_1 = 0,52$, $a_2 = a_3 = 0,5$, $b_{11} = b_{22} = b_{32} = 0,03$, $b_{21}, b_{12} = b_{31} = 0,02$, $d_1 = 0,4$, $d_2 = 0,42$, $c_{11} = c_{21} = 0,0042$, $c_{12} = c_{13} = 0,03$, $c_{21} = c_{22} = 0,0032$ y 400 individuos. Valores iniciales (0,7, 0,2, 0,05, 0,04, 0,01) siendo los tres primeros valores los correspondientes a las especies presa (i de 0 a 2) y dos depredadores ($j = 0$ o $j = 1$).

Fig. 6.11.: Espacio de fases para el caso 2, pares de presas dada por $(i = 2, i = 3)$ y $(i = 1, i = 3)$ y un depredador $j = 1$ caso 2, $a_1 = 0,52, a_2 = a_3 = 0,5, b_{11} = b_{22} = b_{32} = 0,03, b_{21}, b_{12} = b_{31} = 0,02, d_1 = 0,4, d_2 = 0,42, c_{11} = c_{21} = 0,0042, c_{12} = c_{13} = 0,03, c_{21} = c_{22} = 0,0032$ con 400 individuos como población inicial. Varias vistas.

6.4 Modelo muchas especies con difusión para FQ

Se ha realizado un modelo orientado a agentes con el fin de complementar los resultados sobre puntos de equilibrio que se han encontrado en el modelo anterior. Este tipo de modelos proporcionan un marco con ventajas para estudiar un conjunto de muchos tipos de especies en los que además hay difusión. En el modelo que se propone los agentes son la abstracción de las bacterias, estas bacterias pueden ser de diferentes especies y por lo tanto poseen unas características que están asociadas a los tipos de agente. Un conjunto de valores de esas propiedades define un tipo de bacteria. Los agentes del mismo tipo comparten unas propiedades pero poseen otras propias como individuos, tales como la ubicación espacial y un identificador único que permite distinguirlas. En el modelo propuesto podemos introducir ciertas especies cuya interacción sea distinta a la de depredación (por ejemplo cooperación), variar la interacción entre las especies presas, además de permitir definir movimientos espaciales de bacterias de manera simple (desde un punto de vista computacional).

En el caso del ecosistema formado en los pulmones, el modelo se ha configurado como un proceso de depredador-presa. Para ello, los agentes se han diseñado de la siguiente forma:

1. Todos los individuos de una misma especie comparten los parámetros típicos de esa especie, velocidad de desplazamiento, radio de detección, tasas de reproducción, valores de la matriz de interacción (que sólo depende de los tipos de bacterias involucradas en una interacción i y j).
2. Las proporciones de las poblaciones de depredadores y presas se basan en los resultados obtenidos mediante técnicas de NGS a partir de datos experimentales, al igual que en los modelos anteriores y en Dios *et al.* (2017).
3. En la simulación las bacterias están situadas espacialmente en una superficie bidimensional, cuyos bordes están conectados, de forma que las bacterias no perciben límites.

4. Las bacterias pueden desplazarse espacialmente y cada especie posee su velocidad de desplazamiento. El movimiento es aleatorio, es decir, en cada paso de la simulación una determinada bacteria puede desplazarse una distancia aleatoria dentro de su rango de acción, con una probabilidad aleatoria. En este modelo las bacterias no son capaces de buscar presas, o lo que es lo mismo, no hay mecanismos que permitan identificar que hay una posible presa cercana y desplazarse hacia ella, alterando la aleatoriedad de su movimiento.
5. Las presas poseen un radio dentro del cual pueden ser detectadas. Si se encuentran depredadores dentro de dicho radio, se elige aleatoriamente si esa bacteria es depredada o no. El radio es propio de la especie presa.
6. Se decide, en función de la tasa de reproducción, si el individuo se reproduce o no. Si se reproduce, la bacteria resultante es del mismo tipo que la progenitora y permanece en un área cercana (en términos de su tamaño) a su lugar de nacimiento.

En cada paso temporal de duración arbitraria se evalúa al menos un cambio de estado de media a cada una de las bacterias de la simulación. El algoritmo es el siguiente:

1. Se elige una bacteria aleatoriamente.
2. Se determina aleatoriamente si se desplaza o no y qué distancia se desplaza considerando su velocidad de desplazamiento v_i que depende exclusivamente del tipo de bacteria.
3. Una vez se ha desplazado y si es una presa, el agente busca depredadores dentro de su radio de detección propio de la especie a la que pertenece.
4. Si hay depredadores, se selecciona uno aleatoriamente y se decide si es devorada o no en función de la tasa correspondiente de interacción entre ambas, a partir de una matriz de interacción A , cuyos elementos son las tasas de depredación a_{ij} entre la bacteria tipo i y la bacteria cuyo tipo es j .

La distribución espacial inicial de bacterias es aleatoria. En la figura 6.12 se representan las distribuciones espaciales inicial, un paso intermedio

y la distribución final de una simulación. El color indica el tipo de bacteria. Además se ha representado la evolución temporal de una simulación en la figura 6.13 y los resultados globales obtenidos con este modelo en los puntos temporales arbitrarios reproducen la solución oscilatoria clásica de las ecuaciones de Lotka-Volterra y fueron consistentes con la extinción de todas las poblaciones excepto un depredador y una presa, que finalmente coexisten en equilibrio.

Fig. 6.12.: Evolución temporal de la distribución espacial. La primera figura es la distribución inicial de 5,000 bacterias de diferentes especies (3,500 *Pseudomonas* [azul], 1000 *Staphylococcus* [verde], 350 *Haemophilus* [azul claro], 100 *Bdellovibrio* [rojo] y 50 *SPP* [amarillo]). La segunda imagen es la distribución tras 2,500 pasos arbitrarios y la tercera a 5,000. Nótese que las dos últimas gráficas sólo presentan dos especies, un depredador y una presa (Dios *et al.*, 2017).

Para comprender la influencia de las poblaciones iniciales de depredadores, realizamos 50 repeticiones de cada simulación estudiando si las poblaciones sobreviven o mueren.

La figura 6.14 muestra que las poblaciones con un porcentaje inicial de depredadores del 1 % del total de bacterias en las simulaciones siempre sobrevive, mientras que las poblaciones con un porcentaje de depredador inicial de 20 % siempre muere. Este comportamiento refleja la existencia de un umbral en las simulaciones a partir del cual se alcanza un punto estable en el que sólo sobrevive una especie depredadora y una especie depredada. La existencia del umbral es relevante ya que el objetivo del modelo es determinar si es posible desplazar a la especie que causa la enfermedad

Fig. 6.13.: Evolución temporal en 5000 pasos de tiempo arbitrarios. El modelo tiene dos depredadores, tres presas y las proporciones iniciales de depredadores son 0,04 y 0,01 (Dios, Vida et al. 2017)

mediante la introducción de un depredador con el que compita por las presas.

Fig. 6.14.: Tasa de supervivencia respecto al porcentaje inicial de depredadores, tras 50 simulaciones. Debajo del 8 % la población siempre sobrevive. Con poblaciones iniciales mayores del 20 % del total de bacterias en las simulaciones, siempre muere (Dios *et al.*, 2017).

6.5 Efecto de la diversidad sobre la estabilidad en el modelo para FQ

Con el fin de disponer de una visión más realista de la dinámica de la microbiota en el pulmón humano, diseñamos un modelo de muchas especies de bacterias. Ya hemos visto, mediante modelos específicos, el efecto que tiene un tipo determinado de interacción sobre una comunidad con pocos tipos de especies. Sin embargo, cuando analizamos un conjunto grande de diferentes especies hay que tener en cuenta varios fenómenos. Una misma bacteria puede ser depredadora de otra, pero cooperar con una tercera especie, por ejemplo. Además en un conjunto grande podremos encontrar toda clase de interacciones.

Partimos del sistema de ecuaciones 6.7, donde ahora $a_{ii} = s_i$. En el punto de equilibrio, se cumple que la derivada temporal de la población es nula (la población no varía):

$$X_i \left(r_i + \sum_{j=1}^S a_{ij} X_j \right) = 0, \quad i = 1, 2 \dots S, \quad (6.8)$$

para cada i . Esto puede darse por dos razones, o bien porque $X_i = 0$ o por que $\sum_{j=1}^S a_{ij} X_j = -r_i$. Los valores de X_i que cumplan esas condición son puntos estables. Puesto que estamos interesados en soluciones con todos los X_i positivos (puntos interiores), la condición de equilibrio en estos casos se puede escribir de forma matricial:

$$A\vec{x} = -\vec{r} \quad (6.9)$$

con $A = (a_{ij})$ y $\vec{r} = (r_1, r_2, \dots, r_s)$. La solución de este sistema viene dada por

$$\vec{x}^* = A^{-1}\vec{r}. \quad (6.10)$$

Sin embargo, no todos los $\vec{x}^*$ que verifican la ecuación (6.9) son puntos interiores. Eso depende de los valores de a_{ij} y de r_i . Es decir, es una condición necesaria, pero no suficiente, cumplir la ecuación (6.9) para que $\vec{x}^*$ sea un punto de equilibrio interior de (6.7). La estabilidad se determina con el análisis del determinante del jacobiano de dicho sistema evaluado en el punto de equilibrio $\vec{x}^*$. Recordemos que si la parte real de los autovalores es menor que cero, la comunidad es capaz de volver al equilibrio cuando hay una perturbación pequeña. Y esto es más rápido cuanto más negativa es esa parte real. Como además la matriz es real, los autovalores aparecerán en pares de complejos conjugados.

Para estudiar un conjunto de microbios, cuyos tipos de especies sigan una distribución determinada, estudiamos un conjunto de sistemas de ecuaciones cuyos parámetros sean compatibles con el modelo que se quiere estudiar. En particular construimos matrices A (en la ec. 6.10) de interacción que reproducen una distribución aleatoria de poblaciones centrada en los valores que son consistentes con el sistema biológico que representan. De esta forma los resultados que obtengamos no dependen directamente de los

parámetros particulares de una simulación concreta. Podemos obtener una relación entre la estabilidad y lo que influye en esa estabilidad¹¹, además de una estimación de la capacidad de recuperación tras una perturbación y el efecto de la complejidad de las relaciones entre las especies que componen la microbiota del pulmón. Nos basamos principalmente en la teoría de Wigner (1958) y en trabajos posteriores (May, 1974; Sommers *et al.*, 1988; McCann, 2000; Foster y Bell, 2012; Allesina y Tang, 2012 y Coyte *et al.*, 2015). Como se garantiza que los autovalores permanecen en una zona localizada del espacio complejo dentro de una elipse (salvo quizás un autovalor dominante que aparece fuera de esa elipse), podemos analizar la estabilidad de un gran número de sistemas compatibles con el modelo propuesto.

La parte más delicada del proceso es la construcción de las matrices aleatorias de forma que sean consistentes con (6.8). El procedimiento que hemos diseñado es el siguiente:

1. Creamos una matriz M cuyos elementos indican la conectividad entre especies. Es decir, cada elemento de la matriz indica si la especie i interacciona con la especie j . Para generar esta matriz se define un valor C (llamador de conectividad) y por cada interacción posible (M_{ij}, M_{ji}) creamos una variable p_1 aleatoria siguiendo una distribución uniforme $U([0, 1])$.
 - a) Si $p_1 \leq C$ entonces $M_{ij} = M_{ji}$. y asignamos a ese elemento de matriz el valor 1
 - b) En caso contrario $M_{ij} = M_{ji}$. Asignamos el valor 0.

La matriz es simétrica porque si la especie i interacciona con la j , el elemento correspondiente a la interacción de j con i debe ser 1 también. Podemos ver que C nos dice la proporción de especies que interaccionan con otras.

2. Determinamos el tipo de interacción mediante la selección de los coeficientes del jacobiano. Si hay depredación, por ejemplo, uno de los valores del jacobiano será positivo y el simétrico negativo (una de las especies muere y la otra se reproduce, porque se ha producido la depredación). Esto permite diseñar modelos en los que, por ejemplo,

¹¹Ya ha quedado establecido que la relación entre estabilidad de la microbiota y salud del hospedante.

una bacteria es depredadora de otra, pero presa de otro tipo. Para ello P_m es la proporción de interacciones tipo $(+/+)$, P_c las de tipo $(-/-)$ y, en este modelo, $1 - P_m - P_c$ las de tipo $(+/-)$. Esto permite crear ecosistemas en los que hay varios tipos de interacciones: competición, mutualismo, depredación o mezcla. La suma de todas las proporciones ha de ser 1¹².

3. Se construye la matriz A , con la que se resolverá el sistema de ecuaciones (6.10). Cada elemento i, j de la matriz es distinto de cero para el que caso en que se cumpla que $M_{ij} = M_{ji} = 1$. Cuando eso sucede, escogemos una variable p_2 de una distribución uniforme $U([0, 1])$ y el valor de esta nueva variable se compara con los valores que determinan la distribución de interacciones de la siguiente forma:
 - a) Si $p_2 \geq 1 - P_m$ quiere decir que M_{ij} representa una interacción cooperativa y seleccionamos los valores de las entradas de J_{ij} y J_{ji} con una distribución normal $-N(0, \sigma^2)$.
 - b) Si $p_2 \leq P_c$ quiere decir que M_{ij} representa una interacción competitiva y definimos las entradas con una *half-normal* $N(0, \sigma^2)$.
 - c) Si $P_c < p_2 < 1 - P_m$, entonces M_{ij} es una interacción de explotación. Elegimos un p_3 de una distribución uniforme $U([0, 1])$. Si $p_3 \leq 0,5$ i se beneficia de j y establecemos J_{ij} a partir de una distribución normal $-N(0, \sigma^2)$ y J_{ji} de $N(0, \sigma^2)$. En caso contrario, cuando $p_3 > 0,5$, tenemos de una distribución normal $-N(0, \sigma^2)$ y J_{ij} de $N(0, \sigma^2)$.
4. Se resuelve el sistema de ecuaciones matricialmente, usando la matriz A diseñada en el apartado anterior. Si todos los valores del vector solución $\vec{x}^*$ son positivos, se continúa con el siguiente paso, si no se abandona la matriz. Esto garantiza que sea un punto interior.
5. Se construye el Jacobiano del sistema, usando el valor del punto estable encontrado en el paso anterior. Como

$$\frac{dx_i}{dt} = f_i(\vec{x})$$

¹²Podemos considerar otro tipo de interacciones, tales como $(+/0)$ o $(-/0)$, pero no en este modelo.

donde $f_i(\vec{x})$ es cada una de las ecuaciones del sistema de ecuaciones dado por (6.7):

$$J_{ij}(\vec{x}^*) = \frac{\partial f_i(\vec{x})}{\partial x_j} = \begin{cases} r_i + 2a_{ii}x_i^*, & \text{para } i = j \\ x_i^* a_{ij}, & \text{para } i \neq j \end{cases} \quad (6.11)$$

Si el valor de $M_{ij} = M_{ji} = 0$ el del Jacobiano es cero también. La interacción intraespecífica, la de una especie consigo misma, está relacionada con el valor de las a_{ii} .

Usando esta construcción de matrices, el modelo de depredador en FQ se puede definir mediante la siguiente particularización:

- $C = 0,6$. Esta constante es la proporción de bacterias que de media interaccionan con otras especies de bacterias. Para comparar los resultados con los modelos anteriores usamos 5 bacterias, de las cuales 2 bacterias interaccionan con 3. Así que ponemos un valor cercano a $2/3$.
- $P_c = 0$. No hay competición entre las bacterias de diferente especie, inicialmente. Sin embargo se analizará el efecto que tiene sobre la estabilidad del modelo añadir la competición entre especies distintas.
- $P_m = 0$. No hay cooperación.
- La única interacción posible en el modelo es la de depredación presa, interacciones del tipo $+/-$. Tal y como se construye el modelo, la probabilidad de que la interacción entre dos especies tengan una interacción de tipo depredación presa viene dada por $1 - P_c - P_m$. Como los valores que hemos seleccionado para esas probabilidades es cero, como punto de partida todas serán del tipo depredación.
- $a_{ii} = s_i \neq 0$. Hay interacción entre los miembros de la misma especie. También se estudia el efecto de variar este tipo de interacción dentro de los miembros de una misma especie. Veremos, además, que la elección del parámetro $s_i = 0$ no es realista.

Los resultados de varias simulaciones numéricas pueden verse en la figura 6.16, en la que podemos ver el espectro de autovalores para los parámetros definidos en el apartado de estabilidad lineal. Los valores de partida del resto de parámetros son $C = 0,6, \sigma = 0,05, P_m = 0, P_c = 0, S = 5$.

Se representan gráficamente los autovalores encontrados numéricamente en 50 simulaciones de matrices aleatorias. Como puede verse en esa figura, la elipse que contiene a los autovalores no tiene parte positiva en el eje real, por lo que el sistema es estable.

Mediante esta técnica podemos hacer el análisis de un sistema con un mayor número de especies de bacterias, sin dificultad computacional. Con 20 especies definidas ($S = 20$), se muestran los resultados en la figura 6.17. Los autovalores se acercan a la parte positiva del eje real, pero siguen siendo negativos, por lo que el sistema sigue siendo estable.

Fig. 6.16.: Autovalores de 50 matrices aleatorias y 5 especies del modelo de depredador-presa para FQ. ($\sigma = 0,05, s_i = 0,1$).

Fig. 6.17.: Autovalores de 50 matrices aleatorias y 20 especies del modelo de depredador-presa para FQ. ($\sigma = 0,05$, $s_i = 0,1$).

Podemos analizar cómo cambia la estabilidad haciendo que parte de las especies de bacterias tengan tipos de interacciones distintas. Se varía la proporción de bacterias que compiten, que cooperan o mezcla de ambas. Estos casos se analizan:

1. Variando la competición. El efecto obtenido es tal que los autovalores del sistema se alejan hacia la parte positiva del eje real (ver Fig. 6.18).
2. Modificando la cooperación. Casi no tiene efecto en la estabilidad, si nos fijamos en el autovalor más a la derecha (ver Fig. 6.19).
3. Variando la proporción de especies que interaccionan una con la otra. Los cambios realizados en la variable C que da la proporción de especies de bacterias que interaccionan con otras (figura 6.20), tiene poco efecto sobre la estabilidad del sistema, tal y como ya observa Coyte *et al.* (2015).

Fig. 6.18.: Autovalores de 50 matrices aleatorias y 20 especies del modelo de depredador-presa para FQ. Se puede ver el efecto de la competición, variando el parámetro P_c en 0,2 en cada gráfica ($\sigma = 0,05$, $s_i = 0,1$).

Fig. 6.19.: Autovalores de 50 matrices aleatorias y 20 especies del modelo de depredador-presa para FQ. Se puede ver el efecto de la cooperación, variando el parámetro P_m en 0,2 en cada gráfica ($\sigma = 0,05$, $s_i = 0,1$).

Fig. 6.20.: Autovalores de 50 matrices aleatorias y 20 especies del modelo de depredador-presa para FQ. Se puede ver el efecto de la proporción de interacciones, variando el parámetro C en 0,1 en cada gráfica ($\sigma = 0,05$, $s_i = 0,1$).

De esta manera, a modo de resumen, podemos decir que desde el punto de vista de la modelización, la propuesta de partida de introducir un depredador natural de las bacterias que participan en la agudización respiratoria asociada a la FQ es una aproximación realista como posible tratamiento terapéutico. Sin embargo hay que considerar que, según nuestro modelo, existe un umbral mínimo de proporción de bacterias que deben introducirse para que el procesado de depredación deseado se produzca. Por otro lado, los análisis del modelo para muchas especies apuntan a que las

interacciones existentes distintas a las de depredación pueden jugar un papel crítico en el equilibrio del sistema. La microbiota debe ser estudiada en su conjunto. No es suficiente introducir el depredador adecuado, el resto de interacciones deben ser adecuadas.

C. difficile y transplante de heces

” *Lots of people think, well, we’re humans; we’re the most intelligent and accomplished species; we’re in charge. Bacteria may have a different outlook: more bacteria live and work in one linear centimeter of your lower colon than all the humans who have ever lived.*

— Neil deGrasse Tyson

(Space Chronicles: Facing the Ultimate Frontier)

Los antibióticos alteran la microbiota intestinal y esto tiene a veces consecuencias, no deseadas, asociadas al cambio de la estructura de la microbiota (Manichanh *et al.*, 2012; Sokol *et al.*, 2008 y Manichanh *et al.*, 2006). Normalmente entre el 3 % y el 29 % de los pacientes tratados con antibióticos¹ (Högenauer *et al.*, 1998; Turck *et al.*, 2003) desarrollan diarrea no complicada, colitis pseudomembranosa² o incluso megacolon tóxico³ (Monge *et al.*, 2008) y, en ciertos casos, puede originar sepsis y producir la muerte (Fernández y Aladrén, 2011). La asociación más común a los casos de colitis pseudomembranosa es con la bacteria *Clostridium difficile*⁴ (*C. difficile*).

Uno de los tratamientos más efectivos para la infección consiste en realizar un transplante de la microbiota de las heces de un donante sano, normalmente familiar (Kassam *et al.*, 2013). Se produce una repoblación de

¹Típicamente cefalosporinas, clindamicina y quinolonas.

²Inflamación del colón, producida por la toxina que genera el *C. Difficile*. Produce unas placas blanquecinas en el interior del colón, que se conocen como pseudomembranas.

³Se presenta cuando la inflamación se propaga a capas profundas del colón. El colón se ensancha y deja de funcionar.

⁴Para ampliar información, véase el apéndice D.

la microbiota y los pacientes se recuperan con un alto grado de éxito, mayor que los tratamientos con antibióticos (Brandt *et al.*, 2012; Petrof *et al.*, 2013).

Esa infección y el proceso de repoblación y recuperación de la microbiota sirve de marco experimental para este modelo.

7.1 Modelo para microbiota intestinal y *Clostridium*

7.1.1 Datos experimentales

Fruto de una colaboración con el Servicio de Microbiología, Hospital Universitario Ramón y Cajal y el Instituto Ramón y Cajal de Investigación de Madrid, disponemos de datos de trasplantes de heces realizados a enfermos, así como datos sobre la evolución posterior de la microbiota. Además disponemos de la composición de la microbiota de muestras de los donantes del trasplante. Se representan estos datos en las figuras 7.1, 7.2 y 7.3, dónde puede verse la distribución de la población de los principales *phylum*⁵ en tres momentos diferentes durante el proceso de recuperación de la microbiota en los pacientes tratados con trasplante de heces. Esto permite ver la evolución temporal del efecto del trasplante. En la primera figura (7.1), por ejemplo, se puede apreciar que la cantidad de *Fimicutes* que aparece en las muestras del paciente 1 y del paciente 2 es muy similar y el paciente 3 tiene una proporción mayor que los otros dos.

⁵Descripción del término *phylum* en el apéndice D.

Fig. 7.1.: Distribución de bacterias de tres pacientes, en la fase previa al trasplante. Se representan los tres pacientes para comparar las distribuciones de familias de bacterias en la microbiota del intestino. La gráfica conecta a cada paciente con el *phylum* más abundante. El paciente 1 se representa en morado, el 2 en verde y el 3 en azul. El grosor de la conexión es proporcional a la cantidad de bacterias encontrada en las muestras de cada paciente.

Fig. 7.2.: Distribución de bacterias relacionadas con tres pacientes, en la **primera fase**, inmediatamente después del trasplante. Se representan los tres pacientes para comparar las distribuciones de familias de bacterias en la microbiota del intestino. La gráfica conecta a cada paciente con el *phylum* más abundante. El paciente 1 se representa en morado, el 2 en verde y el 3 en azul. El grosor de la conexión es proporcional a la cantidad de bacterias encontrada en las muestras de cada paciente.

Fig. 7.3.: Distribución de bacterias relacionadas con tres pacientes, en **fase posterior** al trasplante (dos semanas a un mes). Se representan los tres pacientes para comparar las distribuciones de familias de bacterias en la microbiota del intestino. La gráfica conecta a cada paciente con el *phylum* más abundante. El paciente 1 se representa en morado, el 2 en verde y el 3 en azul.

La repoblación de la microbiota se consigue a partir de bacterias del donante y se observa que, una vez se alcanza el equilibrio la distribución final de bacterias en el enfermo sigue una distribución similar a la microbiota del donante. Es una de las razones por las que se seleccionan donantes que compartan el mismo ambiente y si es posible entorno familiar (Kassam *et al.*, 2013)

Se estudia este proceso de repoblación asociado al trasplante de heces mediante dos modelos. El primero basado en un modelo de Lotka-Volterra con competición similar al definido para FQ en el capítulo anterior. El segundo modelo es un modelo orientado a agentes, en el que también tiene un papel fundamental la interacción interespecífica de tipo competición.

7.1.2 Análisis de estabilidad del modelo para *Clostridium*

En el modelo que proponemos las bacterias compiten de manera intran-
sitiva (Levine *et al.*, 2017) de forma que se reflejan de forma más realista
las interacciones entre diferentes tipos de bacterias. Se sabe que el papel de
las interacciones entre especies es importante en la estabilidad (Hairston
et al., 1968) y también que la limitación de recursos puede llevar a estados
estables de coexistencia entre diferentes especies. Se permite también que
algunas bacterias puedan formar una unidad funcional, es decir, que formen
una agrupación de bacterias que tienen una función diferente a las que lo
componen y, en nuestro modelo, la unidad funcional es más resistente a
la competición contra las bacterias que consideramos dañinas, tal y como
apunta el estado del arte en la investigación de la resiliencia y resistencia de
las comunidades de bacterias (Allison y Martiny, 2008; Shade *et al.*, 2012).

Fig. 7.5.: Representación gráfica de la matriz de interacción entre diferentes especies. Cada nodo representa un tipo de bacteria y las conexiones describen la interacción. Si hay conexión, hay interacción y la flecha indica que tipo tiene interacción con qué tipo.

La representación gráfica de las interacciones entre tipos de bacterias puede verse en la figura 7.5. En este gráfico (en el que se proponen interacciones de ejemplo) la bacteria 5 interacciona con la 2, pero no al contrario. En nuestro modelo esto se traduce en que si la bacteria de tipo 5 se encuentra dentro de su radio de acción con una del tipo 2, hay cierta probabilidad de que esta muera y aquella se reproduzca. Es una representación gráfica de la matriz de interacciones usada en el modelo orientado a agentes. Una forma de entender este tipo de interacciones es el juego⁶ (Reichenbach *et al.*, 2007).

Construimos el modelo a partir del sistema de ecuaciones (5.4). A partir de este sistema de ecuaciones se estudia cómo afectan los distintos tipos de interacciones (competición, cooperación y depredación) entre bacterias al equilibrio del sistema. Hemos supuesto que el principal tipo de interacción interespecífica es la de competición y se estudian los puntos estables soluciones del sistema de ecuaciones, mediante el análisis de estabilidad espectral⁷, con los siguientes parámetros y valores iniciales:

1. $C = 0,5$. Como punto de partida establecemos que de media la mitad de los tipos de bacterias interacciona unas con otras. Además se estudiará el efecto del cambio en el valor de conectividad entre especies sobre la estabilidad.
2. $P_c = 0,2$. Tasa de competición, las bacterias compiten especie contra especie determinada, no hay bacterias que compitan con todas en el modelo inicial. Uno de los esquemas de relación usados se representa en figura 7.5
3. $P_m = 0,7$. Hay cooperación, las unidades funcionales pueden formarse en grupos de tres bacterias.
4. $s_i = 0$. Es decir, no hay interacción entre los miembros de la misma especie. El papel de esta interacción es relevante para que el sistema sea más estable frente a perturbaciones. Como se verá en este capítulo, la competición entre los miembros de una misma especie contribuye a aumentar la estabilidad. Veremos que el caso $s_i \neq 0$ es más realista.

⁶O algo más complejo: piedra, papel tijera, lagarto, spock (Levine *et al.*, 2017).

⁷Seguendo el método expuesto en 6.5.

En este modelo, a diferencia con el modelo de FQ, se parte de una matriz de a_{ij} en la que existe competición (hay términos tales que $a_{ij} = -a_{ji}$), hay cooperación de forma que se refleje la posibilidad de que existan unidades funcionales (véase Little *et al.*, 2008 como una buena descripción de estos fenómenos) e, inicialmente, no hay interacción entre los individuos de una misma especie ($a_{ii} = s_i = 0$).

En la figura 7.6 se representan el espectro de autovalores de un conjunto de 50 simulaciones de matrices correspondientes al modelo propuesto. En cada una de las subfiguras (de izquierda a derecha y de arriba abajo) se aumenta la proporción de bacterias cooperativas. Para ello se va variando P_m a partir de los valores iniciales que corresponden a la gráfica de arriba a la izquierda. Puede observarse que los autovalores se van extendiendo hacia la parte negativa del eje real, así que los puntos de equilibrio son estables. La cooperación extiende la elipse que envuelve los autovalores hacia zonas más negativas. Nótese que se están simulando situaciones para 7 especies distintas y así poder comparar con los modelos orientado a agentes del siguiente apartado. Esa situación pocas veces se da en ecosistemas reales, en los que hay un gran número de especies.

En la Fig. 7.7 se simulan esas variaciones para el caso en el que hay 20 especies distintas ($S = 20$) de forma que el modelo sea más realista y se varía la proporción de bacterias que compiten⁸. Cuando además se introduce competición entre los miembros de una misma especie el sistema llega a ser inestable. Esto puede verse porque los autovalores más a la derecha llegan a cortar el eje de la parte real igual a cero.

⁸De nuevo las condiciones iniciales las correspondientes a la primera gráfica arriba a la izquierda.

Fig. 7.6.: Autovalores de 50 matrices aleatorias y 7 especies del modelo de competición y unidades funcionales diseñado para el tratamiento del caso de infección por *C. difficile*. Se puede ver el efecto de la cooperación, variando el parámetro P_m en 0,2 en cada gráfica ($\sigma = 0,05$, $s_i = 0,1$).

Fig. 7.7.: Autovalores de 100 matrices aleatorias y 20 especies del modelo de competición y unidades funcionales diseñado para el tratamiento del caso de infección por *C. difficile*. Se puede ver el efecto de la competición, variando el parámetro P_c en 0,2 en cada gráfica ($\sigma = 0,05$, $s_i = 0,1$).

Introducir competición interna ayuda a entender cómo la microbiota llega al equilibrio. Esto se consigue haciendo mayor s_i . En la figura 7.8 se representa el espectro de autovalores de un sistema con 20 tipos de bacterias y 50 simulaciones (50 matrices aleatorias) en los que esa competición interna viene dada por $s_i = 0,2$. Cada gráfica varía respecto a la anterior añadiendo competición entre especies mediante la adición de 0,2 a la tasa P_c de la gráfica anterior, siendo la correspondiente a los valores iniciales la de gráfica de la esquina superior izquierda. Puede verse que la competición

intraespecífica desplaza el centro de la elipse hacia las zonas negativas del eje real. El efecto de añadir competición entre miembros de la misma especie hace más estable el sistema, y la competición interespecífica desplaza en dirección contraria la elipse.

Fig. 7.8.: Autovalores de 50 matrices aleatorias y 20 especies del modelo de competición y unidades funcionales diseñado para el tratamiento del caso de infección por *C. difficile*. Se puede ver el efecto de la competición, variando el parámetro P_c en 0,2 en cada gráfica y se ha aumentado la competición intraespecífica respecto a la fig 7.7 ($\sigma = 0,05$, $s_i = 0,2$).

7.2 Modelo orientado a agentes: *C. difficile*

También se ha definido un modelo orientado a agentes que permita complementar y analizar comportamientos emergentes del sistema para el caso de la infección de *C. difficile*. Aunque el análisis de estabilidad nos ha dado una visión de cómo afecta el tipo de interacción a la estabilidad tras una perturbación del equilibrio⁹, para entender de mejor forma la dinámica del sistema es conveniente analizar el modelo orientado a agentes. Como ya se ha descrito, este tipo de modelos parte de las reglas locales de interacción entre las bacterias. Estas reglas pueden ser mucho más complejas que las que se definen en un sistema analítico. Podemos incluir términos espaciales de desplazamiento modificación de los parámetros dinámicamente. Por otro lado, la definición de reglas locales de interacción y el modelo de abstracción de los agentes facilitan la comunicación con los colaboradores médicos y biólogos.

El modelo que utilizamos es discreto y tiene las siguientes características:

1. En la simulación las bacterias están situadas espacialmente en una superficie bidimensional. Los extremos de la superficie están conectados. Es decir, no hay bordes para las bacterias.
2. Las bacterias pueden desplazarse espacialmente. Cada especie i posee su velocidad de desplazamiento v_i que puede ser diferente de la velocidad de desplazamiento de las otras. El movimiento es aleatorio, de forma que en cada paso de la simulación una determinada bacteria puede desplazarse una distancia aleatoria dentro de su rango de acción, con una probabilidad aleatoria. Es decir, los tipos de bacterias cuya interacción pueda terminar en la muerte de una de ellas, mediante la competición, no se buscan voluntariamente.
3. Las bacterias poseen un radio dentro del cual pueden ser detectadas y ese radio depende de la especie de bacteria que sea.
4. Todos los individuos de una misma especie comparten los parámetros típicos de esa especie, velocidad de desplazamiento, radio de detección, tasas de reproducción, valores de la matriz de interacción. Estos últimos

⁹mediante la introducción de ciertas bacterias o la eliminación de otras.

valores sólo dependen de los tipos de bacterias involucradas en una interacción i y j .

5. Los individuos poseen valores propios de algunas propiedades, tales como la ubicación, identificador único, etc.
6. Se define una matriz de interacción entre diferentes especies:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

donde los elementos a_{ij} representan la probabilidad de interacción entre la especie i y la especie j y hay n especies distintas.

7. Cada tipo de bacteria tiene además un radio de alcance, dentro del cual es capaz de detectar otras bacterias con la que puede competir. Tras cada desplazamiento, la bacteria selecciona a otra dentro de su radio de acción, si hay alguna. A partir de la matriz de interacción, usando la tasa dada por a_{ij} , se determina aleatoriamente el resultado de esa interacción. Los posibles resultados son muerte del individuo o bien supervivencia. La forma de determinar el resultado de la interacción es mediante la selección de un número aleatorio de una distribución uniforme $U[0, 1]$ y si el valor es menor que el elemento correspondiente de la matriz, la bacteria muere en la competición. En caso contrario sobrevive.
8. Si se gana en la interacción, la bacteria tiene cierta probabilidad de reproducción, teniendo en cuenta además que se ha establecido un límite de crecimiento.
9. El espacio sobre el que pueden moverse las bacterias está limitado es bidimensional y el contorno sobre el que se mueven están conectados, es decir, el lado derecho es el izquierdo y el superior es el inferior, de forma que no hay bordes.

En cada paso computacional que representa una unidad de tiempo arbitrario de forma que, de media, todas las bacterias han evaluado un posible cambio de estado. En las simulaciones que se presentan en las figuras figura 7.9 y figura 7.10 aparecen las tres fases del proceso de simulación de un trasplante de heces.

Fig. 7.9.: Evolución temporal del sistema dinámico con siete especies de bacterias, se pueden observar las tres fases de la simulación: inicio, antibiótico y repoblación mediante trasplante de heces. La tasa de repoblación es de 0,01. Los diferentes tipos de bacterias se denotan por letras diferentes. La primera perturbación se introduce en $t = 150$, eliminando aleatoriamente bacterias de varios tipos menos uno. En $t = 300$ se introduce el trasplante de heces, repoblando bacterias.

Fig. 7.10.: Evolución temporal del sistema dinámico con siete especies de bacterias, se pueden observar las tres fases de la simulación: inicio, antibiótico y repoblación mediante trasplante de heces. La tasa de repoblación es de 0,06. La primera perturbación se introduce en $t = 150$, eliminando aleatoriamente bacterias de varios tipos salvo uno. En $t = 300$ se introduce el trasplante de heces, repoblando bacterias.

La primera fase consiste en dejar evolucionar el sistema desde unos valores iniciales hasta que se alcanza el equilibrio dinámico. Estos valores iniciales de población están relacionados con las proporciones de ciertas especies encontradas en el intestino de sujetos sanos que se presentan en éste mismo capítulo (concretamente en las figuras 7.1, 7.2 y 7.3).

En una segunda fase de las simulaciones, cuando ha pasado un determinado número de pasos de tiempo arbitrario, se perturba el sistema simulando el efecto de un antibiótico. Es decir, se matan bacterias seleccionadas aleatoriamente en una proporción determinada a cada especie y en menor medida a una especie en concreto. En el caso del *C. difficile* es una justificación fundamentada ya que la bacteria es resistente a antibióticos (Sebahia *et al.*, 2006).

Tras un tiempo para que alcance de nuevo el equilibrio (y en algunas simulaciones antes de llegar a él), se incluyen bacterias que repueblan la microbiota, siguiendo una distribución similar a la original encontrada en el sujeto sano o en los donantes sanos.

Se observan comportamientos en cascada de forma que, en ciertas ocasiones, la desaparición de un tipo de bacteria tiene como efecto que la población de otra desaparece, aunque no haya interacción directa entre ellas definida en la matriz de interacción, debido a relaciones intransitivas entre estas (como se explica en Levine *et al.*, 2017). Este tipo de comportamientos emergentes quedan excluidos del análisis de estabilidad, ya que solo se analizan los puntos interiores y evidencian la utilidad del modelo orientado a agentes.

En la figura 7.11 se presenta la densidad de población de la bacteria seleccionada como dañina respecto al total, en función de la cantidad de bacterias que se implantan durante el proceso de transplante de heces. En las simulaciones se aprecia que existe un umbral mínimo de proporción de nuevas especies de bacterias necesarias para que la repoblación y recuperación de la flora original se produzca.

Hemos comparado el efecto de la cooperación entre bacterias en este modelo, creando de manera aleatoria unidades funcionales entre los agentes de la simulación. En el caso del modelo orientado a agentes se construyen

esas unidades funcionales como agrupaciones de tres tipos de bacterias que están situadas dentro de un determinado radio y por lo tanto están próximas físicamente. Además deben pertenecer cada una de ellas a un tipo determinado, no se forman con tres bacterias del mismo tipo. En la simulación la pertenencia a una unidad funcional altera su tasa de supervivencia, de forma tal que son más resistentes a la depredación o a la competición. La formación de unidades funcionales es aleatoria, en el momento de la competición se observa si hay vecinos que formen unidad funcional con la bacteria simulada, y se modula la tasa de competición dada por la matriz de interacciones.

Fig. 7.11.: Densidad final de *Clostridium* en función de la proporción de bacterias añadidas durante el transplante, simulaciones realizadas mediante un modelo orientado a agentes y los parámetros definidos en la sección 7.2. Se comparan los resultados cuando se considera cooperación entre especies de bacterias mediante la inclusión o no de unidades funcionales.

Parte IV

Conclusiones y futuros pasos

Conclusiones y futuros pasos

8

” *To understand the difficulty of predicting the next 100 years, we have to appreciate the difficulty that the people of 1900 had in predicting the world of 2000.*”

— **Physics of the Future: How Science Will Shape Human Destiny and Our Daily Lives by the Year 2100**
(Michio Kaku)

8.1 Conclusiones

Como resultado del uso del método innovador de estudio de la dinámica de propagación de enfermedades y virus informáticos en redes multicapas se concluye que:

1. Se puede analizar la dinámica de la propagación de virus en redes multicapa de manera eficiente usando una red de una sola capa construida de forma que los nodos son los mismos en todas las capas y siguiendo nuestro método. Esto implica que mediante una matriz efectiva de contagio se desarrolla la dinámica equivalente a la red con todas las capas. El modelo describe de manera más realista el proceso de infección de malware multivector que los modelos existentes hasta ahora.
2. Considerando la matriz efectiva de contagio, el autovector por la izquierda que está asociado al autovalor máximo ($\bar{\lambda}_{\max}$) aproxima las probabilidades esperadas de que los nodos se infecten en el límite de probabilidades pequeñas y está relacionado también con la influencia dinámica de cada nodo con el resto de la red durante el proceso de contagio. El autovector asociado al autovalor máximo, por lo tanto, muestra la relación entre la dinámica y la topología.

3. Una capa simple no puede disparar de manera independiente la propagación de la epidemia y, considerando la red multicapa como un todo, la enfermedad está por debajo del umbral de infección.
4. Como principal resultado del modelo se ha determinado que la densidad de nodos infectados en redes multicapa es mayor que la esperada si se analizan por separado las propagaciones en cada capa. En una red multicapa que modela redes sociales y redes de comunicaciones la propagación de virus es más eficiente que si consideramos las capas que la componen por separado. Esto explica porqué que los virus modernos usan un método de propagación multivector.
5. La estrategia de inmunización más eficiente para ese tipo de redes consiste en seleccionar los nodos con mayor fortaleza.

Además los estudios de modelos biológicos permiten concluir que:

1. Desde el punto de vista de la modelización, la hipótesis de introducir un depredador natural de las bacterias que participan en la agudización respiratoria asociada a la FQ es una aproximación realista como posible tratamiento terapéutico, siendo una posible opción médica para disminuir la reagudización de los síntomas.
2. También asociado a la FQ, se concluye que la microbiota debe ser estudiada en su conjunto. No es suficiente introducir el depredador adecuado, el resto de interacciones inter e intraespecíficas deben ser tenidas en cuenta. Los análisis del modelo para muchas especies apuntan a que las interacciones existentes distintas a las de depredación pueden jugar un papel crítico en el equilibrio del sistema.
3. Existe un umbral mínimo de proporción de bacterias que deben introducirse para que el procesado de depredación deseado en el tratamiento para los síntomas de la FQ se produzca.
4. Existe un umbral mínimo de número de bacterias necesarias en la repoblación de la microbiota en el intestino humano mediante el trasplante de heces. Y la diversidad de la microbiota introducida juega un papel muy importante para recuperar la estabilidad de la microbiota del intestino.
5. Se puede reproducir de manera cualitativa la dinámica observada en el proceso de repoblación de la microbiota intestinal mediante modelos

orientados a agentes en lo que la principal interacción entre bacterias es la competición directa.

6. Se ha determinado la importancia de la competición entre las bacterias de una misma especie para alcanzar la estabilidad de las poblaciones. La competición intraespecífica juega un papel fundamental en la estabilidad de los ecosistemas de bacterias.

8.2 Artículos y colaboraciones

Se ha publicado el trabajo científico sobre la propagación de malware (Vida *et al.*, 2015) y además hay varias contribuciones en congresos científicos (Fises, Int. J. Complex Systems in Science, Vida *et al.*, 2013).

Se ha establecido un marco de colaboración multidisciplinar con expertos del Servicio de Microbiología, Hospital Universitario Ramón y Cajal, e Instituto Ramón y Cajal de Investigación en varios proyectos de investigación relacionados con los sistemas complejos del microbioma humano, de cuya colaboración surge una publicación científica (Dios *et al.*, 2017), además de varias contribuciones a varios congresos.

8.3 Futuros pasos

Como futuros pasos en la investigación se presentan varias oportunidades. En primer lugar la generalización del modelo estocástico de propagación de malware de forma que permita estimar niveles de riesgo en términos de probabilidad. Esto tiene aplicaciones inmediatas en la industria, y es una necesidad propia de grandes empresas. Se pretende, además, estudiar este tipo de procesos en redes evolutivas de forma que se refleje el comportamiento de malware que altera la estructura topológica de la red según se extiende, mediante alteración de protocolos de red o la implementación de protocolos TCP inexistentes previamente a la infección.

Se pretende mejorar el modelo orientado a agentes de forma que soporte comportamientos espaciales más realistas. Además se pretende generalizar la forma en la que se introducen procesos como el de la creación de biofilms. Se ha planteado crear un modelo físico 3D de un pulmón y estudiar

la distribución espacial, propagación y dinámica de las bacterias depredadoras, formando un biofilm sobre la superficie del pulmón. Esto permitirá avanzar en el diseño del tratamiento efectivo de la enfermedad de FQ. Se complementará con simulaciones computacionales.

Se está trabajando en un estudio de los criterios de estabilidad de comunidades ecológicas, basados en los presentados en esta Tesis.

Y finalmente se desea avanzar en un marco matemático de análisis de modelos orientados a agente, de forma que pueda explorarse la relación entre ciertos parámetros propios del modelo y su efecto si necesidad de realizar simulaciones costosas computacionalmente recorriendo el espacio de parámetros.

Parte V

Anexos

Procesos de Markov y ecuación maestra

Se denomina estocástico al proceso cuyo comportamiento evoluciona probabilísticamente en el tiempo. Según Nelson y Nelson (1985) y Montroll (2012), cualquier desarrollo temporal que pueda ser analizable en términos de probabilidad merece ser denominado proceso estocástico. En un proceso estocástico los posibles resultados de un experimento no se conocen con certidumbre, por lo que puede entenderse como una distribución de probabilidad sobre un espacio de rutas.

Algunos ejemplos de procesos estocásticos son bastante conocidos, por ejemplo el estudio del movimiento browniano por Einstein (1905) se usa ampliamente en biología (Beebe, 2012), finanzas (Cont, 2004), computación (Baron, 2013), criptografía (Katz y Yehuda, 2014), interacciones sociales (Wasserman, 1980) y un largo etc. Varios temas de investigación de la presente tesis se estudian mediante técnicas propias de sistemas estocásticos (como el comportamiento de la microbiota humana en el intestino, o la propagación de virus en redes de ordenadores) por lo que es necesario introducir algunos conceptos básicos de ellos.

A.1 Estado del sistema

Para estudiar este tipo de procesos¹ se plantean modelos estocásticos que predicen un conjunto de posibles resultados ponderados por sus probabilidades. Estos modelos pueden basarse en ecuaciones estocásticas que describen la dinámica del sistema. El espacio de posibles resultados de un experimento o de un proceso se conoce como espacio de muestra. Se define una variable aleatoria X como una función real definida en el espacio de probabilidades asociado a un experimento aleatorio. Esta variable aleatoria mapea un resultado de la muestra a un valor numérico. Cada una de las

¹La mayor parte de este apéndice está basado en Gardiner (1985).

variables aleatorias del proceso tiene su propia función de distribución de probabilidad sobre un conjunto distinto de posibles valores y pueden o no estar correlacionadas entre ellas. En adelante consideraremos que las variables aleatorias toman valores discretos, aunque la generalización de los resultados a variables continuas no supone dificultad. Esto permite definir un sistema estocástico como una colección de variables aleatorias $(x, y, z, \dots)$ indexada en el tiempo.

Definimos el estado del sistema en un instante t como una realización concreta de dichas variables, $\vec{a} = (x, y, z, \dots)$. De esta forma la dinámica del sistema se puede resumir en una jerarquía de probabilidades, que escribimos como $P_r(\vec{a}_1, t_1; \vec{a}_2, t_2; \dots; \vec{a}_r, t_r)$ y que denotan la probabilidad conjunta de que el sistema haya estado en el estado $\vec{a}_1$ en el momento t_1 , en el estado $\vec{a}_2$ en $t_2, \dots$ y en el estado $\vec{a}_r$ en el momento t_r .

Esta probabilidad está normalizada,

$$\sum_{\vec{a}_1} \cdots \sum_{\vec{a}_r} P_r(\vec{a}_1, t_1; \vec{a}_2, t_2; \dots; \vec{a}_r, t_r) = 1 \quad (\text{A.1})$$

y a partir de ella se pueden calcular probabilidades marginales en subconjuntos del histórico de estados,

$$\sum_{\vec{a}_j} \sum_{\vec{a}_{j+1}} \cdots \sum_{\vec{a}_r} P_r(\vec{a}_1, t_1; \vec{a}_2, t_2; \dots; \vec{a}_r, t_r) = P_{j-1}(\vec{a}_1, t_1; \vec{a}_2, t_2; \dots; \vec{a}_{j-1}, t_{j-1}). \quad (\text{A.2})$$

Además se cumple que si dos tiempos son iguales en los argumentos, la probabilidad de tener dos conjuntos distintos de valores es nula.

$$P_r(\vec{a}_1, t_1; \vec{a}_2, t_1; \dots; \vec{a}_r, t_r) = P_r(\vec{a}_1, t_1; \vec{a}_3, t_3; \dots; \vec{a}_r, t_r) \delta_{\vec{a}_1, \vec{a}_2}, \quad (\text{A.3})$$

donde $\delta_{\vec{a}_1, \vec{a}_2}$ es igual a 1 si $\vec{a}_1 = \vec{a}_2$ y cero en caso contrario. Si se está estudiando un caso de variable aleatoria continua sólo hay que tener en cuenta que las probabilidades pasan a ser densidades de probabilidad, las delta de Kronecker pasan a ser deltas de Dirac, los sumatorios integran y, puesto que las densidades de probabilidad son magnitudes con dimensiones,

normalmente conviene obtener un conjunto de variables adimensionales equivalentes, típicamente multiplicando por los diferenciales.

A.2 Transiciones entre estados

La probabilidad de transición a cierto estado desde un historial de estados anteriores, que escribimos

$$W_r(\vec{a}_r, t_r \mid \vec{a}_{r-1}, t_{r-1}; \dots; \vec{a}_1, t_1), \quad (\text{A.4})$$

no es más que la probabilidad condicionada de que el sistema esté en un determinado estado $\vec{a}_r$ en t_r teniendo en cuenta que estuvo en el estado $\vec{a}_{r-1}$ en t_{r-1} , en el estado $\vec{a}_{r-2}$ en t_{r-2} , $\dots$ y así hasta el estado $\vec{a}_1$ en el momento t_1 , con $t_1 < t_2 < \dots < t_{r-1} < t_r$.

La W_r está normalizada:

$$\sum_{\vec{a}_r} W_r(\vec{a}_r, t_r \mid \vec{a}_{r-1}, t_{r-1}; \dots; \vec{a}_1, t_1) = 1. \quad (\text{A.5})$$

Además, por ser una probabilidad condicionada, la relación entre P_r y W_r es:

$$P_r(\vec{a}_1, t_1; \vec{a}_2, t_2; \dots; \vec{a}_r, t_r) = W_r(\vec{a}_r, t_r \mid \vec{a}_{r-1}, t_{r-1}; \dots; \vec{a}_1, t_1) P_{r-1}(\vec{a}_1, t_1; \vec{a}_2, t_2; \dots; \vec{a}_{r-1}, t_{r-1}). \quad (\text{A.6})$$

Es decir, la probabilidad de que lleguemos a cierto estado es la probabilidad de que se produzca una transición a dicho estado desde un estado anterior por la probabilidad de que el sistema haya estado en ese estado anterior en concreto.

A.3 Procesos de Markov

En los problemas que se estudian en la presente tesis, los valores de las variables que definen el estado del sistema $\vec{a}_r, t_r$ dependen únicamente del estado del sistema en el momento anterior y hay total independencia entre los valores de las variables estocásticas con sus valores en instantes

anteriores $(t_{r-2}, t_{r-3}, \dots)$. Esto es lo que se conoce como un proceso de Markov, en el que:

$$W_r(\vec{a}_r, t_r \mid \vec{a}_{r-1}, t_{r-1}; \dots; \vec{a}_1, t_1) = W(\vec{a}_r, t_r \mid \vec{a}_{r-1}, t_{r-1}) \quad (\text{A.7})$$

En este tipo de procesos la expresión A.6, junto con la ecuación A.2, nos permiten escribir:

$$W(\vec{a}_r, t_r \mid \vec{a}_{r-2}, t_{r-2}) = \sum_{\vec{a}_{r-1}} W(\vec{a}_r, t_r \mid \vec{a}_{r-1}, t_{r-1}) W(\vec{a}_{r-1}, t_{r-1} \mid \vec{a}_{r-2}, t_{r-2}), \quad (\text{A.8})$$

que es la ecuación de **Chapman-Kolmogorov** (Gardiner, 1985). La condición de Markov dada por la ecuación (A.7) es suficiente para que se cumpla la expresión de Chapman-Kolmogorov, pero no es necesaria. Hay otros procesos que cumplen ésta ecuación sin ser procesos de Markov.

El objetivo ahora es conseguir una ecuación dinámica que nos permita conocer la evolución temporal del sistema a partir de las probabilidades de transición entre estados, usando las ecuaciones para procesos estocásticos vistas hasta ahora, y en particular las de los procesos de Markov. La ecuación maestra relaciona la derivada temporal de la probabilidad $P(\vec{a}, t) \equiv P_1(\vec{a}, t)$ con las propbabilidades en todos los estados en tiempo t . Por tanto, nos interesa calcular $\frac{\partial P(\vec{a}, t)}{\partial t}$. Para ello partimos de la ecuación de Chapman-Kolmogorov A.8. Multiplicando a ambos lados por $P(\vec{a}_{r=2}, t_{r-2})$ obtenemos:

$$P(\vec{a}, t + \Delta t) = \sum_{\vec{b}} W(\vec{a}, t + \Delta t \mid \vec{b}, t) P(\vec{b}, t), \quad (\text{A.9})$$

donde hemos reemplazado $t_{r-1} \rightarrow t$, $t_r \rightarrow t + \Delta t$, $\vec{a}_{r-1} \rightarrow \vec{b}$, $\vec{a}_r \rightarrow \vec{a}$. Separando en la suma de la expresión anterior el término con $\vec{b} = \vec{a}$ y restando en ambos lados $P(\vec{a}, t)$, obtenemos

$$P(\vec{a}, t + \Delta t) - P(\vec{a}, t) = \sum_{\vec{b} \neq \vec{a}} W(\vec{a}, t + \Delta t \mid \vec{b}, t) P(\vec{b}, t) - (1 - W(\vec{a}, t + \Delta t \mid \vec{a}, t)) P(\vec{a}, t). \quad (\text{A.10})$$

En la expresión anterior $1 - W(\vec{a}, t + \Delta t \mid \vec{a}, t)$ es la probabilidad de que el sistema no esté en el estado $\vec{a}$ en el momento $t + \Delta t$ condicionada a que

estaba en ese estado en t . Usando la ecuación A.5, dividiendo la expresión A.10 entre Δt y tomando el límite $\Delta t \rightarrow 0$ obtenemos finalmente la ecuación maestra:

$$\frac{\partial P(\vec{a}, t)}{\partial t} = \sum_{\vec{b} \neq \vec{a}} \left[\omega(\vec{a} | \vec{b}, t) P(\vec{b}, t) - \omega(\vec{b} | \vec{a}, t) P(\vec{a}, t) \right], \quad (\text{A.11})$$

donde

$$\omega(\vec{a} | \vec{b}, t) \equiv \lim_{\Delta t \rightarrow 0} \frac{W(\vec{a}, t + \Delta t | \vec{b}, t)}{\Delta t},$$

son las tasas de transición suponiendo que dicho límite existe. La ecuación A.11 es la ecuación de Kolmogorov hacia delante para procesos de Markov en tiempo continuo.

Análisis de estabilidad de puntos críticos

En este capítulo se repasan conceptos básicos sobre el análisis de estabilidad en puntos críticos para sistemas de reacción-difusión, cuya expresión matemática se encuentra en la ecuación (6.2). En dicha ecuación la parte espacial y la parte temporal están separadas. Esto permite realizar el análisis de estabilidad de forma sencilla y similar al que se realiza cuando no hay parte espacial (ver cap. ??). Sin embargo, hay que tener en cuenta que la forma de realizar esta linealización se hace de forma distinta para la parte temporal y la parte espacial. Para que las perturbaciones espaciales sean pequeñas, suponemos que son de la forma $A \sin(\omega x + \phi)$, es decir una amplitud por un seno. Cualquier condición inicial puede expresarse en una base de funciones trigonométricas y, además, esta selección de base asegura que está acotada entre -1 y 1 .

$$\begin{cases} \frac{dN}{dt} = N(1 - N - P) + D\nabla^2 N \\ \frac{dP}{dt} = aP(N - b) + \nabla^2 P \end{cases} \quad (\text{B.1})$$

Las ecuaciones del sistema (B.1) tienen la forma

$$\frac{\partial f_i}{\partial t} = R_i(f_1, f_2, \dots, f_n)$$

donde ahora las f_i son funciones de x y de t . El estado de equilibrio del sistema de EDPs anterior viene dado por $(f_1^*, f_2^*, \dots, f_n^*)$, siendo las f_i^* las soluciones del sistema homogéneo:

$$\begin{aligned}
0 &= R_1(f_1^*, f_2^*, \dots, f_n^*) \\
0 &= R_2(f_1^*, f_2^*, \dots, f_n^*) \\
&\vdots \\
0 &= R_n(f_1^*, f_2^*, \dots, f_n^*)
\end{aligned} \tag{B.2}$$

Para realizar el análisis de estabilidad lineal hacemos el siguiente cambio:

$$f_i(x, t) \implies f_i^* + \Delta f_i(x, t) \tag{B.3}$$

y tal como hemos dicho, asumimos que la perturbación es pequeña. Se desarrolla en serie en la base de senos y cosenos y nos quedamos con el primer término, que luego se linealiza:

$$f_i^* + \Delta f_i(x, t) = f_i^* + \sin(\omega x + \phi) \Delta f_i(t) \tag{B.4}$$

El sistema de ecuaciones queda, haciendo esa sustitución y escribiendo $S = \sin(\omega x + \phi)$ para simplificar la notación:

$$\begin{aligned}
S \frac{\partial \Delta f_1}{\partial t} &= R_1(f_1^* + S \Delta f_1^*, f_2^* + S \Delta f_2^*, \dots, f_n^* + S \Delta f_n^*) - D_1 \omega^2 S \Delta f_1 \\
S \frac{\partial \Delta f_2}{\partial t} &= R_2(f_1^* + S \Delta f_1^*, f_2^* + S \Delta f_2^*, \dots, f_n^* + S \Delta f_n^*) - D_2 \omega^2 S \Delta f_2 \\
&\vdots \\
S \frac{\partial \Delta f_n}{\partial t} &= R_n(f_1^* + S \Delta f_1^*, f_2^* + S \Delta f_2^*, \dots, f_n^* + S \Delta f_n^*) - D_n \omega^2 S \Delta f_n
\end{aligned} \tag{B.5}$$

Escrito en forma matricial, si R es el vector que representa los términos de la parte de reacción y D es una matriz Diagonal,

$$\sin(\omega x + \phi) \frac{\partial \Delta f}{\partial t} = R(f^* + \sin(\omega x + \phi) \Delta f) - D \omega^2 \sin(\omega x + \phi) \Delta f \tag{B.6}$$

Si ahora realizamos la operación de linearización en la parte temporal, tenemos el sistema linealizado completamente.

$$R(f^* + \sin(\omega x + \phi) \Delta f) \approx R(f^*) + \left(\begin{array}{cccc} \frac{\partial R_1}{\partial f_1} & \frac{\partial R_1}{\partial f_2} & \cdots & \frac{\partial R_1}{\partial f_n} \\ \frac{\partial R_2}{\partial f_1} & \frac{\partial R_2}{\partial f_2} & \cdots & \frac{\partial R_2}{\partial f_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial R_n}{\partial f_1} & \frac{\partial R_n}{\partial f_2} & \cdots & \frac{\partial R_n}{\partial f_n} \end{array} \right) \bigg|_{f=f^*} \sin(\omega x + \phi) \Delta f = \quad (\text{B.7})$$

$$= \sin(\omega x + \phi) \left(\begin{array}{cccc} \frac{\partial R_1}{\partial f_1} & \frac{\partial R_1}{\partial f_2} & \cdots & \frac{\partial R_1}{\partial f_n} \\ \frac{\partial R_2}{\partial f_1} & \frac{\partial R_2}{\partial f_2} & \cdots & \frac{\partial R_2}{\partial f_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial R_n}{\partial f_1} & \frac{\partial R_n}{\partial f_2} & \cdots & \frac{\partial R_n}{\partial f_n} \end{array} \right) \bigg|_{f=f^*} \Delta f \quad (\text{B.8})$$

donde hemos usado que $R(f^*)$ porque es la condición de estabilidad dada en (B.2), así que

$$\sin(\omega x + \phi) \frac{\partial \Delta f}{\partial t} = \sin(\omega x + \phi) J|_{f=f^*} \Delta f - D \omega^2 \sin(\omega x + \phi) \Delta f, \quad (\text{B.9})$$

lo que implica que

$$\frac{\Delta f}{\Delta t} = (J - D \omega^2) \big|_{f=f^*} \Delta f. \quad (\text{B.10})$$

Por lo tanto, la estabilidad de la ecuación de reacción-difusión (B.1) puede estudiarse calculando los autovalores de $(J - D \omega^2)|_{f=f^*}$, donde J es el jacobiano de la parte temporal del sistema de ecuaciones y D es la matriz diagonal construida con las constantes de difusión y ω es el parámetro que determina la frecuencia espacial de las perturbaciones.

Distribución de brechas

En este anexo se presentan algunos datos relacionados con la ciberguerra y los ataques informáticos asociados que hayan sido identificados. En concreto, en la figura C.1 se detallan las principales brechas de seguridad de los años 2016 y 2017, agrupadas por tipo de organización que las han reportado. Las áreas de cada rectángulo son proporcionales al número de incidentes detectados y el color corresponde a los diferentes tipos de organizaciones. Los códigos usados para identificar tipos de organización se presentan en la tabla C.1, y los tipos de ataque se identifican por los códigos de la tabla C.2. En esa gráfica se ve que las organizaciones más atacadas son las relacionadas con salud médica y con empresas no relacionadas con servicios financieros ni distribución online.

Tipo de organización	Significado
BSF	Negocio. Financiero o seguros.
BSO	Negocio. Otros.
BSR	Negocio. Venta, distribución, incluida venta online.
EDU	Instituciones educativas.
GOV	Gobierno y militar.
MED	Medicina, salud, proveedores médicos, etc.
NGO	Sin ánimo de lucro.

Tabla C.1.: Notación usada en las brechas de seguridad. Usado en las figuras 3.1, figura C.1 y figura C.2

Fig. C.1.: Distribución de las principales brechas de seguridad de los años 2016 y 2017, por tipo de organización que sufre el ataque y por tipo de ataque. El área es proporcional al número de brechas reportadas. Datos obtenidos de Privacy Rights Clearinghouse (<https://www.privacyrights.org/data-breaches>).

Tipos de brechas	Significado
Payment Card Fraud (CARD)	Fraude de tarjetas de crédito y de débito (no hacking).
Hacking or Malware (HACK)	Hackeadas por terceras partes.
Insider (INSD)	Acceso típicamente legítimo con intención de atacar.
Physical Loss (PHYS)	Pérdidas de documentos en papel, documentos descartados o robados.
Portable Device (PORT)	Pérdida o robo de pendrives, portátiles, CDs, discos duros, etc.
Stationary Device (STAT)	Pérdida de ordenadores fijos, robo, apropiación indebida.
Unintended Disclosure (DISC)	Publicación involuntaria que no incluye hacking.
Unknown (UNKN)	Desconocido.

Tabla C.2.: Notación de tipo de brechas de seguridad sobre el dato. Usado en las figuras C.1 y C.2

Para poder comparar de manera relativa el tipo de brechas por organización se ha representado en la figura C.2 en columnas normalizadas. La leyenda se presenta en la tabla asociada (C.2). Los ataques relacionados con las tarjetas de crédito aparecen más frecuentemente en organizaciones que venden productos online o las gubernamentales. Las de publicación de datos confidenciales, sin embargo, suponen la mayoría de los ataques detectados en organizaciones no gubernamentales.

Fig. C.2.: Distribución de las principales brechas de seguridad de los años 2016 y 2017, por tipo de organización que sufre el ataque y por tipo de ataque. Datos obtenidos de Privacy Rights Clearinghouse (<https://www.privacyrights.org/data-breaches>).

Nociones de Microbiología

La Microbiología es la ciencia que se dedica al estudio de microorganismos microscópicos, en particular los virus, las bacterias, los hongos y los parásitos. O, al menos, así se consideraba en sus inicios como ciencia porque, actualmente, estudia también organismos macroscópicos, tales como parásitos multicelulares de longitud de orden de metros.

En concreto, para lo que concierne a esta memoria de Tesis, nos centraremos en describir las principales características de las bacterias y, más concretamente, el conjunto de bacterias existente en diferentes partes del cuerpo humano, conformando la microbiota¹ humana. Las bacterias son células procariotas, es decir, que poseen un núcleo primitivo, en contraposición con las eucariotas, que presentan una membrana interna alrededor del núcleo de la célula. Ese núcleo contiene la información genética codificada mediante ADN² y ARN³.

La forma más común de clasificar estos organismos es mediante sus características morfológicas. En particular, hay una división entre bacterias grampositivas y gramnegativas, en referencia a su respuesta respecto a la tinción con cristal violeta⁴. La absorción de dicho tinte, según el método descrito por Hans Christian Gram (1853), es distinta en la pared de las bacterias en función del espesor de su peptidoglicano⁵. Si en la tinción de Gram las bacterias retienen el colorante azul se denominan grampositivas, mientras que las gramnegativas son aquellas que no retienen dicho colorante. Las primeras presentan un peptidoglicano de mayor grosor y las segundas no.

¹Más adelante daremos una definición más rigurosa de microbiota.

²Ácido desoxirribonucleico.

³Ácido ribonucleico.

⁴Hexametilo o violeta de metilo 10B, cloruro de pararosanilina.

⁵El peptidoglucano es un copolímero muy resistente que protege a las bacterias de una ruptura osmótica en ambientes acuáticos siendo la estructura básica de la pared celular de las bacterias y de las Prochlorophyta.

Las bacterias grampositivas tienen una pared celular rica en polímeros hidrosolubles de fosfatos de poliol⁶. Esto es relevante porque está relacionado con la capacidad de fijarse a otras bacterias y a receptores de las células de mamíferos, por lo que influye en su virulencia.

Algunas bacterias grampositivas pueden formar esporas bajo condiciones ambientales difíciles. Esto sucede, por ejemplo, en el género *Clostridium*, microorganismo que ha sido estudiado en la presente Tesis Doctoral. La espora es, básicamente, una partícula deshidratada y se crea mediante la formación de cubiertas adicionales en un proceso de captación de calcio. Estas esporas son altamente resistentes a los antibióticos y a las condiciones ambientales adversas, incluyendo la limpieza típica de los hospitales. La espora contiene toda la maquinaria celular necesaria para que, una vez rehidratada cuando se dan las condiciones necesarias, pueda producir una nueva célula vegetativa idéntica a la que formó esa espora. La capacidad del *C. difficile* para producir daño en el intestino viene dada por la producción de tres toxinas (toxina A, toxina B y toxina binaria⁷) que tienen la capacidad de producir inflamación en el colon y daño en la superficie mucosa epitelial (Monge *et al.*, 2008).

La microbiota humana es el conjunto de microorganismos (principalmente bacterias, pero hay también hongos, virus, etc) que conviven simbióticamente en el cuerpo del hospedador. Suele asociarse el término principalmente al intestino, pero existen otros ecosistemas como el pulmonar, el vaginal o la misma piel. Su composición depende de factores tales como la dieta, la edad, la salud, el estado hormonal, la higiene y ciertos medicamentos tales como antibióticos, etc. En la figura D.1 se representa los diferentes ecosistemas de la microbiota en el cuerpo humano y los principales componentes de la misma.

La microbiota controla la proliferación de microorganismos patógenos por métodos que incluyen competencia por nutrientes, producción de bacteriocinas, estimulación de la respuesta inmune del cuerpo humano anfitrión (Di Bella *et al.*, 2013), entre otras cosas. Si el equilibrio en dicha microbiota se altera mediante la administración de antibióticos (tal y como analizamos

⁶Ácidos teicoicos.

⁷También llamada C. difficile transferasa, CDT.

en este documento), puede producirse una enfermedad (Blumberg y Powrie, 2012; Claesson *et al.*, 2012) y en ocasiones alguna enfermedad cuyos síntomas están asociados a la colonización de algún tipo de bacteria (*Streptococcus pneumoniae*, *Staphylococcus aureus* y *Clostridium difficile*, entre otros). Estas bacterias forman parte de la distribución normal de poblaciones en la microbiota hasta que sucede este cambio en su población relativa.

Algunos ejemplos de esas enfermedades relacionadas con los casos de estudio de las investigaciones de esta memoria de Tesis es la enfermedad inflamatoria intestinal, en particular enfermedades de Crohn y colitis ulcerosa (Manichanh *et al.*, 2012; Sokol *et al.*, 2008 y Manichanh *et al.*, 2006) y la obesidad (Bäckhed *et al.*, 2004; Cani *et al.*, 2008 y Vajro *et al.*, 2013). También se ha demostrado que la disbiosis de la microbiota intestinal tiene efecto en la respuesta inmune fuera del intestino (Noverr y Huffnagle, 2004; Round y Mazmanian, 2009) y que, además, posee relación con el cáncer de mama (Xuan *et al.*, 2014) e incluso se ha estudiado la relación con alteraciones psiquiátricas, por ejemplo, psicosis afectiva y desórdenes bipolares (Eaton *et al.*, 2010) así como depresión y ansiedad (Filipović *et al.*, 2007) y se relaciona con alergias (Hanski *et al.*, 2012).

Entender, por tanto, las bases de la dinámica del ecosistema de bacterias y los mecanismos por los que se recupera la homeostasis⁸ y la resistencia simbiótica a la colonización es crítico para diseñar futuras terapias relacionadas con la manipulación de la microbiota.

Para clasificar e identificar cada bacteria se usa el método de clasificación basado en el de Linné (1767), de forma que actualmente el sistema de clasificación más aceptado tiene tres dominios y está basado en la filogenia molecular. En ese sistema, las bacterias son miembros del dominio Bacteria, diferenciándose taxonómicamente según su *phylum*, clase, orden, familia, género y especie.

Gracias a los métodos de secuenciación masiva, se ha establecido que en la microbiota existen muchas más especies de las que se pensaba originalmente (Qin *et al.*, 2010) y que los *phyla Firmicutes* y *Bacteroidetes* constitu-

⁸La capacidad de un organismo de mantener una condición interna estable compensando los cambios en su entorno mediante el intercambio regulado de materia y energía con el exterior.

Fig. D.1.: Tipos y localización de los ecosistemas de la microbiota humana (Ackerman, 2012).

yen la gran mayoría de la microbiota en el intestino humano (Arumugam *et al.*, 2011), figura D.2.

Existen varias aproximaciones recientes al tratamiento de esas enfermedades relacionadas con la microbiota. Por ejemplo, la repoblación o alteración del equilibrio asociado a la disbiosis en el intestino mediante trasplante fecal (Di Bella *et al.*, 2013; Anderson *et al.*, 2012) o la introducción de bacterias depredadoras del patógeno causante de la reagudización de la enfermedad, tal y como se propone en esta Tesis para modular el ecosistema pulmonar y prevenir el deterioro respiratorio. Este tipo de tratamientos alteran la situación de equilibrio o cambian la estructura de la microbiota de forma que la nueva situación sea más funcional para el paciente.

Fig. D.2.: Perfiles funcionales y filogénicos de la microbiota intestinal humana, obtenido de (Arumugam *et al.*, 2011).

Material y métodos

E.1 Propagación de malware

Para desarrollar la investigación presentada en la parte II, distinguimos dos grupos de actividades: el proceso de recogida de información, creación y abstracción de las redes y simulaciones estocásticas y por otro lado las fases de modelado, simulación y análisis.

En la fase de recogida se usaron programas que hemos desarrollado en *Python* para extraer la información de fuentes tales como *Google Scholar*, la propia web del *FISES*¹ y otras bases de datos de publicaciones científicas. La información se procesa en mi portátil con Linux y en ciertas ocasiones en servidores departamentales, para generar diccionarios de autores. Estos diccionarios permiten identificar autores independientemente del formato de la firma en los documentos que analizamos. Por ejemplo: *Rafael Vida*, *Rafael A. Vida*, *R. Vida* son el mismo autor y en ocasiones aparecen en distintas colaboraciones con esas posibles variaciones de denominación.

Para la construcción de la red multicapa se extrajeron las relaciones entre los autores y las universidades, considerando que un mismo autor puede haber participado en varias instituciones. La red de comunicaciones usa la información disponible en los registros DNS accesibles públicamente mediante consultas *nslookup*, asumiendo que todas las redes basaban su protocolo de comunicaciones en *IP*. Las redes formadas se analizaron y representaron usando librerías de redes complejas, tales como *Networkx* y *Graph-tool* para PYTHON además del usar los programas *Gephi* y *Cytoscape* para esas funciones.

El modelo de propagación de virus en redes multicapa es un modelo estocástico y para su evaluación se efectúan muchas simulaciones numéricas usando programas realizados en PYTHON y en C++, así como pequeños

¹<http://fisesXX.gefenol.es/>, donde XX es el año

programas de análisis simbólico realizados usando la librería de *Python* *Sympy*.

E.2 Microbiota humana

Para analizar la estabilidad de los sistemas dinámicos estudiados en este apartado, se han usado diferentes programas en *Python* con librerías de cálculo simbólico (*SymPy*) y librerías optimizadas para el uso de matrices, tales como *Numpy*. Se han realizado decenas de miles de simulaciones a lo largo de varios meses.

Los modelos de agentes de competición (*C. difficile*) y los de depredación (FQ) se han realizado en lenguaje PYTHON, con un programa orientado a objetos en el que las clases son la abstracción del agente y los métodos y propiedades están relacionadas directamente con las propiedades de los agentes y las funciones de cambio de estado. Para las simulaciones estocásticas de ambos modelos se han hecho programas en PYTHON y en C++, mediante el uso de técnicas de cadenas de Markov.

La representación gráfica de los resultados de las simulaciones se crean con *matplotlib* y la de los datos de pacientes con *PowerBi* y *Gephi*.

En el modelo orientado a agente de depredador-presa para FQ, se usaron datos de 15 pacientes adultos que han sido atendido en el hospital Ramón y Cajal, cada uno de ellos contribuyendo con 3-4 esputos inducidos durante un año de manera continua. Estos datos fueron proporcionados por colaboradores del servicio de Microbiología del Ramón y Cajal, como parte del estudio del modelo de bacteriófagos en la microbiota del pulmón. Tras la recolección, las muestras se separan en dos aliquots, uno que se usa en un proceso de cultivo convencional y el otro para realizar un análisis de secuenciación de nueva generación (NGS) del gen 16S rDNA, donde son congelados a -80°C.

Bibliografía

- Ackerman, Jennifer (2012). «The ultimate social network». En: *Scientific American* 306.6, págs. 36-43 (vid. pág. 144).
- Alberich, Ricardo, Joe Miro-Julia y Francesc Rosselló (2002). «Marvel Universe looks almost like a real social network». En: *arXiv preprint cond-mat/0202174* (vid. pág. 15).
- Albert, Réka y Albert-László Barabási (2002). «Statistical mechanics of complex networks». En: *Reviews of modern physics* 74.1, pág. 47 (vid. pág. 14).
- Allesina, Stefano y Si Tang (2012). «Stability criteria for complex ecosystems». En: *Nature* 483.7388, pág. 205 (vid. pág. 92).
- Allison, Steven D y Jennifer BH Martiny (2008). «Resistance, resilience, and redundancy in microbial communities». En: *Proceedings of the National Academy of Sciences* 105.Supplement 1, págs. 11512-11519 (vid. pág. 106).
- Anderson, JL, RJ Edney y K Whelan (2012). «Systematic review: faecal microbiota transplantation in the management of inflammatory bowel disease». En: *Alimentary pharmacology & therapeutics* 36.6, págs. 503-516 (vid. pág. 144).
- Anderson, R., C. Barton, R. Böhme *et al.* «Measuring the cost of cybercrime». En: *Paper in 11th Annual Workshop on the Economics of Information Security 2012* () (vid. pág. 24).
- Antoine, V., R. Bongiorno, A. Borza *et al.* (2003). *et al.* Technical Report. Router security configuration guide, version 1.1 b (vid. pág. 26).
- Arumugam, Manimozhiyan, Jeroen Raes, Eric Pelletier *et al.* (2011). «Enterotypes of the human gut microbiome». En: *nature* 473.7346, pág. 174 (vid. págs. 144, 145).
- Asghari, Hadi, Michael Ciere y Michel JG Van Eeten (2015). «Post-Mortem of a Zombie: Conficker Cleanup After Six Years.» En: *USENIX Security*, págs. 1-16 (vid. pág. 27).

- Bacci, Giovanni, Patrizia Paganin, Loredana Lopez *et al.* (2016). «Pyrosequencing unveils cystic fibrosis lung microbiome differences associated with a severe lung function decline». En: *PLoS One* 11.6, e0156807 (vid. pág. 66).
- Bäckhed, Fredrik, Hao Ding, Ting Wang *et al.* (2004). «The gut microbiota as an environmental factor that regulates fat storage». En: *Proceedings of the National Academy of Sciences of the United States of America* 101.44, págs. 15718-15723 (vid. pág. 143).
- Barabási, Albert-László, Réka Albert y Hawoong Jeong (1999). «Mean-field theory for scale-free random networks». En: *Physica A: Statistical Mechanics and its Applications* 272.1, págs. 173-187 (vid. pág. 14).
- Baron, Michael (2013). *Probability and statistics for computer scientists*. CRC Press (vid. pág. 127).
- Barthélemy, M., A. Barrat, R. Pastor-Satorras y A. Vespignani (2005). «Characterization and modeling of weighted networks». En: *Physica A* 346, págs. 34-43 (vid. págs. 31, 35).
- Bastian, Mathieu, Sebastien Heymann y Mathieu Jacomy (2009). *Gephi: An Open Source Software for Exploring and Manipulating Networks* (vid. pág. 163).
- Beebe, Nelson HF (2012). «A Bibliography of Publications of Alan Mathison Turing». En: (vid. pág. 127).
- Bernasconi, Eric, Céline Pattaroni, Angela Koutsokera *et al.* (2016). «Airway microbiota determines innate cell inflammatory or tissue remodeling profiles in lung transplantation». En: *American journal of respiratory and critical care medicine* 194.10, págs. 1252-1263 (vid. pág. 69).
- Blumberg, Richard y Fiona Powrie (2012). «Microbiota, disease, and back to health: a metastable journey». En: *Science translational medicine* 4.137, 137rv7-137rv7 (vid. pág. 143).
- Boccaletti, S., V. Latora, Moreno y, M. Chavez y D.-U. Hwang (2006). «Complex networks: Structure and dynamics». En: *Physics Reports* 424.4-5, págs. 175-308 (vid. págs. 13, 16).
- Boccaletti, Stefano, Ginestra Bianconi, Regino Criado *et al.* (2014). «The structure and dynamics of multilayer networks». En: *Physics Reports* 544.1, págs. 1-122 (vid. pág. 16).
- Brandt, Lawrence J, Olga C Aroniadis, Mark Mellow *et al.* (2012). «Long-term follow-up of colonoscopic fecal microbiota transplant for recurrent *Clostridium difficile* infection». En: *The American journal of gastroenterology* 107.7, pág. 1079 (vid. pág. 102).

- Buldyrev, S. V., R. Parshani, G. Paul, H. E. Stanley y S. Havlin (2010). «Catastrophic cascade of failures in interdependent networks». En: *Nature* 464 (vid. pág. 29).
- Cameron, Simon JS, Keir E Lewis, Sharon A Huws *et al.* (2016). «Metagenomic sequencing of the chronic obstructive pulmonary disease upper bronchial tract microbiome reveals functional changes associated with disease severity». En: *PloS one* 11.2, e0149095 (vid. pág. 66).
- Cani, Patrice D, Rodrigo Bibiloni, Claude Knauf *et al.* (2008). «Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice». En: *Diabetes* 57.6, págs. 1470-1481 (vid. pág. 143).
- Carmody, Lisa A, Jiangchao Zhao, Patrick D Schloss *et al.* (2013). «Changes in cystic fibrosis airway microbiota at pulmonary exacerbation». En: *Annals of the American Thoracic Society* 10.3, págs. 179-187 (vid. pág. 66).
- Caverly, Lindsay J, Jiangchao Zhao y John J LiPuma (2015). «Cystic fibrosis lung microbiome: opportunities to reconsider management of airway infection». En: *Pediatric pulmonology* 50.S40, S31-S38 (vid. pág. 69).
- Choo, Kim-Kwang Raymond (2011). «The cyber threat landscape: Challenges and future research directions». En: *Computers & Security* 30.8, págs. 719-731 (vid. pág. 24).
- Chu, Haiyan, Takeshi Fujii, Sho Morimoto *et al.* (2007). «Community structure of ammonia-oxidizing bacteria under long-term application of mineral fertilizer and organic manure in a sandy loam soil.» En: *Applied and environmental microbiology* 73.2, págs. 485-91 (vid. pág. 57).
- Claesson, Marcus J, Ian B Jeffery, Susana Conde *et al.* (2012). «Gut microbiota composition correlates with diet and health in the elderly». En: *Nature* 488.7410, pág. 178 (vid. pág. 143).
- Cont, Rama (2004). *Financial modelling with jump processes*. Chapman y Hall (vid. pág. 127).
- Conway, John (1970). «The game of life». En: *Scientific American* 223.4, pág. 4 (vid. pág. 63).
- Costanza, Robert, Lisa Wainger, Carl Folke y Karl-Göran Mäler (1993). «Modeling complex ecological economic systems: toward an evolutionary, dynamic understanding of people and nature». En: *Ecosystem Management*. Springer, págs. 148-163 (vid. pág. 12).

- Cottrell, M T y D L Kirchman (2000). «Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter». En: *Applied and environmental microbiology* 66.4, págs. 1692-7 (vid. pág. 57).
- Coyte, Katharine Z, Jonas Schluter y Kevin R Foster (2015). «The ecology of the microbiome: networks, competition, and stability». En: *Science* 350.6261, págs. 663-666 (vid. págs. 71, 92, 96).
- Cozzo, Emanuele, Raquel A Banos, Sandro Meloni y Yamir Moreno (2013). «Contact-based social contagion in multiplex networks». En: *Physical Review E* 88.5, pág. 050801 (vid. pág. 29).
- Darch, Sophie E, Carolyn B Ibberson y Marvin Whiteley (2017). «Evolution of Bacterial “Frenemies”». En: *mBio* 8.3, e00675-17 (vid. pág. 57).
- De Cruz, Peter, Michael A Kamm, Amy L Hamilton *et al.* (2015). «Crohn’s disease management after intestinal resection: a randomised trial». En: *The Lancet* 385.9976, págs. 1406-1417 (vid. pág. 71).
- Di Bella, Stefano, Cecilia Drapeau, Esther García-Almodóvar y Nicola Petrosillo (2013). «Fecal microbiota transplantation: the state of the art». En: *Infectious disease reports* 5.2 (vid. págs. 142, 144).
- Dios, de, Rafael Vida y col. (2017). «Individual Patterns of Complexity in Cystic Fibrosis Lung Microbiota, Including Predator Bacteria, over a 1-Year Period». En: *mBio* 8.5, e00959-17 (vid. págs. 6, 65, 67, 68, 86, 88-90, 123).
- Eaton, William W, Marianne G Pedersen, Philip R Nielsen y Preben Bo Mortensen (2010). «Autoimmune diseases, bipolar disorder, and non-affective psychosis». En: *Bipolar disorders* 12.6, págs. 638-646 (vid. pág. 143).
- Edwards, Benjamin, Steven Hofmeyr y Stephanie Forrest (2016). «Hype and heavy tails: A closer look at data breaches». En: *Journal of Cybersecurity* 2.1, págs. 3-14 (vid. pág. 25).
- Einstein, Albert (1905). «On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat». En: *Annalen der physik* 17, págs. 549-560 (vid. pág. 127).
- Falliere, N., L. O Murchu y E. Chien (2011). «W32». Stuxnet dossier. Technical report, Symantec Corp., Security Response (vid. pág. 26).
- Faust, Karoline y Jeroen Raes (2012). «Microbial interactions: from networks to models». En: *Nature Reviews Microbiology* 10.8, págs. 538-550 (vid. pág. 57).

- Faust, Karoline, J. Fah Sathirapongsasuti, Jacques Izard *et al.* (2012). «Microbial Co-occurrence Relationships in the Human Microbiome». En: *PLoS Computational Biology* 8.7. Ed. por Christos A. Ouzounis, e1002606 (vid. pág. 57).
- Feigenbaum, Mitchell J (1983). «Universal behavior in nonlinear systems». En: *Physica D: Nonlinear Phenomena* 7.1-3, págs. 16-39 (vid. pág. 11).
- Fernández, O Alarcón y B Sicilia Aladrén (2011). *Diarrea postantibiótica. Colitis por Clostridium difficile. Tratamiento de las enfermedades gastroenterológicas* (vid. pág. 101).
- Filipović, Branislav R, Branka F Filipović, Mirko Kerkez, Nikola Milinić y Tomislav Randelović (2007). «Depression and anxiety levels in therapy-naive patients with inflammatory bowel disease and cancer of the colon». En: *World Journal of Gastroenterology: WJG* 13.3, pág. 438 (vid. pág. 143).
- Foster, Kevin R y Thomas Bell (2012). «Competition, not cooperation, dominates interactions among culturable microbial species». En: *Current biology* 22.19, págs. 1845-1850 (vid. pág. 92).
- Fox, Geoffrey (2001). «Peer-to-peer networks». En: *Computing in Science & Engineering* 3.3, págs. 75-77 (vid. pág. 27).
- Ganuza, Eneko, Charles E Sellers, Braden W Bennett, Eric M Lyons y Laura T Carney (2016). «A Novel Treatment Protects Chlorella at Commercial Scale from the Predatory Bacterium *Vampirovibrio chlorellavorus*». En: *Frontiers in Microbiology* 7 (vid. pág. 70).
- Gardiner, Crispin W (1985). *Stochastic methods*. Springer-Verlag, Berlin–Heidelberg–New York–Tokyo (vid. págs. 127, 130).
- Gardner, Martin (1970). «Mathematical games: The fantastic combinations of John Conway's new solitaire game "life"». En: *Scientific American* 223.4, págs. 120-123 (vid. pág. 63).
- Gentile, Margaret Elizabeth, Jennifer Lynn Nyman y Craig S Criddle (2007). «Correlation of patterns of denitrification instability in replicated bioreactor communities with shifts in the relative abundance and the denitrification patterns of specific populations». En: *The ISME Journal* 1, págs. 714-728 (vid. pág. 57).
- Gill, Steven R, Mihai Pop, Robert T DeBoy *et al.* (2006). «Metagenomic analysis of the human distal gut microbiome». En: *science* 312.5778, págs. 1355-1359 (vid. pág. 71).
- Gilligan, Peter H (2014). «Infections in patients with cystic fibrosis: diagnostic microbiology update». En: *Clinics in laboratory medicine* 34.2, págs. 197-217 (vid. pág. 66).

- Gómez, S., A. Arenas, J. Borge-Holthoefer, S. Meloni e Y. Moreno (2013). «Discrete-time Markov chain approach to contact-based disease spreading in complex networks». En: *Europhys. Lett* 89 (vid. págs. 28, 31, 35, 38).
- Gómez, Sergio, Alexandre Arenas, J Borge-Holthoefer, Sandro Meloni y Moreno yamir (2010). «Discrete-time Markov chain approach to contact-based disease spreading in complex networks». En: *EPL (Europhysics Letters)* 89.3, pág. 38009 (vid. págs. 29, 33).
- Granell, C., S. Gómez y A. Arenas (2013). «Dynamical interplay between awareness and epidemic spreading in multiplex networks». En: *Phys. Rev. Lett* 111 (vid. pág. 29).
- Grimm, Volker y Steven F Railsback (2005). *Individual-based modeling and ecology*. Vol. 2005. BioOne (vid. pág. 62).
- Guarner, Francisco y Juan-R Malagelada (2003). «Gut flora in health and disease». En: *The Lancet* 361.9356, págs. 512-519 (vid. pág. 57).
- Guerrero, Ricardo, Carlos Pedrós-Alió, Isabel Esteve *et al.* (1986). «Predatory prokaryotes: predation and primary consumption evolved in bacteria». En: *Proceedings of the National Academy of Sciences* 83.7, págs. 2138-2142 (vid. pág. 70).
- Hairston, NG, JD Allan, RK Colwell *et al.* (1968). «The relationship between species diversity and stability: an experimental approach with protozoa and bacteria». En: *Ecology* 49.6, págs. 1091-1101 (vid. pág. 106).
- Hanski, Ilkka, Leena von Hertzen, Nanna Fyhrquist *et al.* (2012). «Environmental biodiversity, human microbiota, and allergy are interrelated». En: *Proceedings of the National Academy of Sciences* 109.21, págs. 8334-8339 (vid. pág. 143).
- Hinkelmann, Franziska, David Murrugarra, Abdul Salam Jarrah y Reinhard Laubacher (2011). «A mathematical framework for agent based models of complex biological networks». En: *Bulletin of mathematical biology* 73.7, págs. 1583-1602 (vid. pág. 62).
- Högenauer, Christoph, Heinz F Hammer, Guenter J Krejs y C Reisinger (1998). «Mechanisms and management of antibiotic-associated diarrhea». En: *Clinical infectious diseases* 27.4, págs. 702-710 (vid. pág. 101).
- Huang, yvonne y John LiPuma (2016). «The microbiome in cystic fibrosis». En: *Clinics in chest medicine* 37.1, págs. 59-67 (vid. pág. 69).
- Ibarra, Herminia y Steven B Andrews (1993). «Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions». En: *Administrative science quarterly*, págs. 277-303 (vid. pág. 47).

- Iebba, Valerio, Floriana Santangelo, Valentina Totino *et al.* (2013). «Higher prevalence and abundance of *Bdellovibrio bacteriovorus* in the human gut of healthy subjects». En: *PLoS One* 8.4, e61608 (vid. pág. 69).
- Iebba, Valerio, Valentina Totino, Floriana Santangelo *et al.* (2014). «*Bdellovibrio bacteriovorus* directly attacks *Pseudomonas aeruginosa* and *Staphylococcus aureus* Cystic fibrosis isolates». En: *Frontiers in microbiology* 5, pág. 280 (vid. pág. 70).
- Jurkevitch, Edouard (2006). *Predatory prokaryotes: biology, ecology and evolution*. Vol. 4. Springer Science & Business Media (vid. pág. 70).
- Kassam, Zain, Christine H Lee, Yuan yuhong y Richard H Hunt (2013). «Fecal microbiota transplantation for *Clostridium difficile* infection: systematic review and meta-analysis». En: *The American journal of gastroenterology* 108.4, pág. 500 (vid. págs. 101, 105).
- Katz, Jonathan y Lindell Yehuda (2014). *Introduction to modern cryptography*. CRC press (vid. pág. 127).
- Keeling, Matt J y Ken TD Eames (2005). «Networks and epidemic models». En: *Journal of the Royal Society Interface* 2.4, págs. 295-307 (vid. pág. 28).
- Klemm, K., M. A. Serrano, V. M. Eguíluz y M. San Miguel (2012). «A measure of individual role in collective dynamics». En: *Sci. Rep* 2 (vid. págs. 28, 31, 38, 50).
- Kurve, Aditya, Khashayar Kotobi y George Kesidis (2013). «An agent-based framework for performance modeling of an optimistic parallel discrete event simulator». En: *Complex Adaptive Systems Modeling* 1.1, pág. 12 (vid. pág. 62).
- Levine, Jonathan M, Jordi Bascompte, Peter B Adler y Stefano Allesina (2017). «Beyond pairwise mechanisms of species coexistence in complex communities». En: *Nature* 546.7656, pág. 56 (vid. págs. 106, 107, 116).
- Linné, Carl von (1767). *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Vol. 1. Typis Ioannis Thomae (vid. pág. 143).
- Little, Ainslie EF, Courtney J Robinson, S Brook Peterson, Kenneth F Raffa y Jo Handelsman (2008). «Rules of engagement: interspecies interactions that regulate microbial communities». En: *Annu. Rev. Microbiol.* 62, págs. 375-401 (vid. pág. 108).
- Lozupone, Catherine A, Jesse I Stombaugh, Jeffrey I Gordon, Janet K Jansson y Rob Knight (2012). «Diversity, stability and resilience of the human gut microbiota». En: *Nature* 489.7415, pág. 220 (vid. pág. 71).

- Lütkebohle, Ingo (2016). *Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union*. <http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ:L:2016:194:TOC> (vid. pág. 24).
- Malthus, Thomas Robert (1888). *An essay on the principle of population: or, A view of its past and present effects on human happiness*. Reeves & Turner (vid. pág. 58).
- Manichanh, Chaysavanh, Lionel Rigottier-Gois, Elian Bonnaud *et al.* (2006). «Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach». En: *Gut* 55.2, págs. 205-211 (vid. págs. 101, 143).
- Manichanh, Chaysavanh, Natalia Borruel, Francesc Casellas y Francisco Guarner (2012). «The gut microbiota in IBD». En: *Nature Reviews Gastroenterology and Hepatology* 9.10, pág. 599 (vid. págs. 101, 143).
- Manson, Steven M (2001). «Simplifying complexity: a review of complexity theory». En: *Geoforum* 32.3, págs. 405-414 (vid. pág. 11).
- Matrosov, A., E. Rodionov, D. Harley y J. Malcho (2011). «Stuxnet under the microscope». En: *ESSET*, págs. 1-85 (vid. pág. 27).
- May, Robert M (1972). «Will a large complex system be stable?» En: *Nature* 238.5364, págs. 413-414 (vid. pág. 12).
- May, Robert M. y Alun L. Lloyd (2001). «Infection dynamics on scale-free networks». En: *Phys. Rev. E* 64 (vid. pág. 28).
- May, Robert McCredie (1974). *Model ecosystems*. Princeton University Press Princeton, NJ (vid. pág. 92).
- McCann, Kevin Shear (2000). «The diversity–stability debate». En: *Nature* 405.6783, pág. 228 (vid. pág. 92).
- Mieghem, P. Van, J. Omic y R. Kooij (2009). «Virus spread in networks». En: *IEEE/ACM Trans. Net* 17, págs. 1-14 (vid. págs. 28, 31, 32, 38).
- Monge, D, I Sánchez, C Pérez *et al.* (2008). «Brote hospitalario de enfermedad asociada a *Clostridium difficile* (EACD)». En: *Medicina Preventiva* 14.4, págs. 15-21 (vid. págs. 101, 142).
- Montroll, E (2012). *Fluctuation phenomena*. Vol. 7. Elsevier (vid. pág. 127).
- Moreno, Y., J. Gómez y A. Pacheco (2003). «Epidemic incidence in correlated complex networks». En: *Phys. Rev. E* 68 (vid. pág. 28).
- Nelson, Edward y E Nelson (1985). *Quantum fluctuations*. Princeton University Press Princeton (vid. pág. 127).

- Newman, M. E. J. (2003). «The Structure and Function of Complex Networks». En: *SIAM Review* 45.2, págs. 167-256 (vid. pág. 14).
- (2011). «Complex Systems: A Survey». En: arXiv: [1112.1440](#) (vid. pág. 12).
- Newman, M. E. J. (Mark E. J.) (2010). *Networks : an introduction*. Oxford University Press, pág. 772 (vid. pág. 13).
- Newman, Mark (2018). *Networks*. Oxford university press (vid. pág. 13).
- Niazi, Muaz A. (2017). «Towards A Novel Unified Framework for Developing Formal, Network and Validated Agent-Based Simulation Models of Complex Adaptive Systems». En: (vid. pág. 62).
- Noverr, Mairi C y Gary B Huffnagle (2004). «Does the microbiota regulate immune responses outside the gut?» En: *Trends in microbiology* 12.12, págs. 562-568 (vid. pág. 143).
- Pastor-Satorras, R. y A. Vespignani (2001a). «Epidemic dynamics and endemic states in complex networks». En: *Phys. Rev. E* 63 (vid. pág. 28).
- (2001b). «Epidemic spreading in scale-free networks». En: *Phys. Rev. Lett* 86, págs. 3200-3203 (vid. pág. 28).
- Petrof, Elaine O, Gregory B Gloor, Stephen J Vanner *et al.* (2013). «Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: ‘RePOOPulating’ the gut». En: *Microbiome* 1.1, pág. 3 (vid. pág. 102).
- Porras, P., H. Saidi y V. Yegneswaran (2009). «Conficker c analysis». Technical report, Computer Science Laboratory, SRI International (vid. pág. 26).
- Qin, Junjie, Ruiqiang Li, Jeroen Raes *et al.* (2010). «A human gut microbial gene catalogue established by metagenomic sequencing». En: *nature* 464.7285, pág. 59 (vid. pág. 143).
- Ramsey, Bonnie W, Jane Davies, N Gerard McElvaney *et al.* (2011). «A CFTR potentiator in patients with cystic fibrosis and the G551D mutation». En: *New England Journal of Medicine* 365.18, págs. 1663-1672 (vid. pág. 66).
- Reichenbach, Tobias, Mauro Mobilia y Erwin Frey (2007). «Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games». En: *Nature* 448.7157, pág. 1046 (vid. pág. 107).
- Rolain, Jean Marc, Laura Fancello, Christelle Desnues y Didier Raoult (2011). «Bacteriophages as vehicles of the resistome in cystic fibrosis». En: *Journal of antimicrobial chemotherapy* 66.11, págs. 2444-2447 (vid. pág. 69).

- Round, June L y Sarkis K Mazmanian (2009). «The gut microbiota shapes intestinal immune responses during health and disease». En: *Nature Reviews Immunology* 9.5, pág. 313 (vid. pág. 143).
- Rudkjøbing, Vibeke B, Trine R Thomsen, Morten Alhede *et al.* (2011). «True microbiota involved in chronic lung infection of cystic fibrosis patients found by culturing and 16S rRNA gene analysis». En: *Journal of clinical microbiology* 49.12, págs. 4352-4355 (vid. pág. 66).
- Saumell-Mendiola, A., M. A. Serrano y M. Bogu ná (2012). «Epidemic spreading on interconnected networks». En: *Phys. Rev. E* 86 (vid. pág. 29).
- Sebaihia, Mohammed, Brendan W Wren, Peter Mullany *et al.* (2006). «The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome». En: *Nature genetics* 38.7, pág. 779 (vid. pág. 116).
- Shade, Ashley, Hannes Peter, Steven D Allison *et al.* (2012). «Fundamentals of microbial community resistance and resilience». En: *Frontiers in microbiology* 3, pág. 417 (vid. pág. 106).
- Sherrard, Laura J, Scott C Bell y Michael M Tunney (2016). «The role of anaerobic bacteria in the cystic fibrosis airway». En: *Current Opinion in Pulmonary Medicine* 22.6, págs. 637-643 (vid. pág. 69).
- Situngkir, Hokky (2004). «Epidemiology Through Cellular Automata: Case of Study Avian Influenza in Indonesia». En: pág. 10 (vid. pág. 62).
- Smith, Jason M, Hector Castro y Andrew Ogram (2007). «Structure and function of methanogens along a short-term restoration chronosequence in the Florida Everglades». En: *Applied and environmental microbiology* 73.13, págs. 4135-41 (vid. pág. 57).
- Sokol, Harry, Bénédicte Pigneur, Laurie Watterlot *et al.* (2008). «*Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients». En: *Proceedings of the National Academy of Sciences* 105.43, págs. 16731-16736 (vid. págs. 101, 143).
- Sommers, HJ, A Crisanti, Haim Sompolinsky y Stein y (1988). «Spectrum of large random asymmetric matrices». En: *Physical review letters* 60.19, pág. 1895 (vid. pág. 92).
- Son, S w., G. Bizhani, C. Christensen, P. Grassberger y M. Paczuski (2012). «Percolation theory on interdependent networks based on epidemic spreading». En: *EPL* 97 (vid. pág. 29).
- Soo, Rochelle M, Ben J Woodcroft, Donovan H Parks, Gene W Tyson y Philip Hugenholtz (2015). «Back from the dead; the curious tale of the predatory cyanobacterium *Vampirovibrio chlorellavorus*». En: *PeerJ* 3, e968 (vid. pág. 70).

- Thiele, Ines, Almut Heinken y Ronan MT Fleming (2013). «A systems biology approach to studying the role of microbes in human health». En: *Current Opinion in Biotechnology* 24.1, págs. 4-12 (vid. pág. 57).
- Turck, Dominique, Jean-Paul Bernet, Jacques Marx *et al.* (2003). «Incidence and risk factors of oral antibiotic-associated diarrhea in an outpatient pediatric population». En: *Journal of pediatric gastroenterology and nutrition* 37.1, págs. 22-26 (vid. pág. 101).
- Turing, Alan M (1950). «Computing machinery and intelligence». En: *Mind* 59.236, págs. 433-460 (vid. pág. 11).
- Vajro, Pietro, Giulia Paoletta y Alessio Fasano (2013). «Microbiota and gut-liver axis: a mini-review on their influences on obesity and obesity related liver disease». En: *Journal of pediatric gastroenterology and nutrition* 56.5, pág. 461 (vid. pág. 143).
- Verhulst, Pierre-François (1838). «Notice sur la loi que la population suit dans son accroissement. correspondance mathématique et physique publiée par a». En: *Quetelet* 10, págs. 113-121 (vid. pág. 59).
- Vida, R, S Cuenda y J Galeano (2013). «Identifying critical nodes in multi-layered networks under multi-vector malware attack». En: *Int. J. Complex Systems in Science* 3.1, págs. 97-105 (vid. págs. 6, 123).
- Vida, Rafael, Javier Galeano y Sara Cuenda (2015). «Vulnerability of state-interdependent networks under malware spreading». En: *Physica A: Statistical Mechanics and its Applications* 421, págs. 134-140 (vid. págs. 6, 123).
- Volterra, Vito (1926). *Fluctuations in the abundance of species mathematically* (vid. pág. 58).
- (1927). *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*. C. Ferrari (vid. pág. 60).
- Von Neumann, John (1951). «The general and logical theory of automata». En: *Cerebral mechanisms in behavior* 1.41, págs. 1-2 (vid. pág. 63).
- Walter, Jens y Ruth Ley (2011). «The Human Gut Microbiome: Ecology and Recent Evolutionary Changes». En: *Annual Review of Microbiology* 65.1, págs. 411-429 (vid. pág. 57).
- Wasserman, Stanley (1980). «Analyzing social networks as stochastic processes». En: *Journal of the American statistical association* 75.370, págs. 280-294 (vid. pág. 127).
- Watts, Duncan J y Steven H Strogatz (1998). «Collective dynamics of 'small-world' networks». En: *nature* 393.6684, págs. 440-442 (vid. págs. 14, 20).

Wigner, Eugene P (1958). «On the distribution of the roots of certain symmetric matrices». En: *Annals of Mathematics*, págs. 325-327 (vid. pág. 92).

Xuan, Caiyun, Jaime M Shamonki, Alice Chung *et al.* (2014). «Microbial dysbiosis is associated with human breast cancer». En: *PloS one* 9.1, e83744 (vid. pág. 143).

Índice de figuras

2.1.	Red social de superhéroes de cómics Marvel	15
2.2.	Selección de los superhéroes	17
2.3.	Distribución de grado de la red de superhéroes	18
2.4.	Closeness	20
2.5.	Clustering.	21
3.1.	Brechas de Seguridad	25
3.2.	Número de sistemas infectados por Conficker	28
4.1.	Red multicapa, representación	35
4.2.	Redes IP en FisEs	42
4.3.	Red social del FisEs	43
4.4.	Autovalores máximos	44
4.5.	Componentes de las redes sociales en FisEs	45
4.6.	Distribución de grado de la red social de FisEs	46
4.7.	Distribución de conectividad en vecinos del FisEs	47
4.8.	Distribución de cercanía (closeness centrality), FisEs	48
4.9.	Densidad de nodos infectados en red multicapa	49
4.10.	Densidad de nodos infectados ρ , vs. densidad de nodos inmuni- zados ϕ	51
5.1.	Interacciones interespecíficas	56
5.2.	Juego de la Vida	63
6.1.	Distribución de <i>phyla</i> en los 15 paciente de FQ.	67
6.2.	Distribución de <i>phyla</i> en los 15 paciente de FQ (II)	68
6.3.	Espacio de fases, (1, 0)	73

6.4.	Espacio de fases, $(b, 1 - b)$	74
6.5.	Espacio de fases, $i = 1$, caso 1	78
6.6.	Espacio de fases, $i = 2$, caso 1	79
6.7.	Evolución temporal presas y depredadores	80
6.8.	Espacios de fases, presa y un depredador	81
6.9.	Espacio de fases con trayectorias, caso 2	82
6.10.	Evolución temporal, caso 2	84
6.11.	Espacio de fases con trayectorias, caso 2	85
6.12.	Evolución temporal de la distribución espacial en simulaciones pulmón	88
6.13.	Evolución temporal de una simulación de modelo orientado agentes de dos depredadores y tres presas, modelo orientado a agente	89
6.14.	Tasa de supervivencia de depredadores en simulaciones del pulmón	90
6.16.	Espectro, 5 especies modelo de FQ	95
6.17.	Espectro, 20 especies modelo de FQ	96
6.18.	Competición en el modelo de FQ	97
6.19.	Cooperación en el modelo de FQ	98
6.20.	Competición interna en el modelo de FQ	99
7.1.	Fase previa, distribución de bacterias de tres pacientes, <i>C. difficile</i>	103
7.2.	Inmediatamente después del trasplante, distribución de bacte- rias, <i>C. difficile</i>	104
7.3.	Fase posterior al trasplante, distribución de bacterias, <i>C. difficile</i>	105
7.5.	Representación gráfica de la matriz de interacción	106
7.6.	Competición en modelo de <i>C. difficile</i>	109
7.7.	Cooperación en modelo para <i>C. difficile</i>	110
7.8.	Competición interna en modelo para <i>C. difficile</i>	111
7.9.	Evolución temporal, <i>C. difficile</i> y trasplante	114
7.10.	Evolución temporal, <i>C. difficile</i> y trasplante	115
7.11.	Densidad final de bacterias y umbral	117
C.1.	Distribución de brechas de seguridad, por tipos de organización	138
C.2.	Distribución de brechas de seguridad, por tipos de brechas . . .	140

D.1.	Tipos de microbiota humana	144
D.2.	Perfiles de la microbiota humana	145

Notas finales

Ésta tesis ha sido escrita usando $\text{\LaTeX}$ y el estilo *Clean Thesis* desarrollado por Ricardo Langner. Se puede descargar el estilo *Clean Thesis* en <http://cleanthesis.der-ric.de/>.

Ha sido editado en *overleaf* <http://www.overleaf.com>, editor de $\text{\LaTeX}$ online.

La bibliografía se ha gestionado con *Mendeley*

Gran parte del desarrollo de programas se ha realizado con *Python*, de la PythonSoftware Foundation. Python Language Reference, version 2.7. Disponible en <http://www.python.org>, además de en lenguajes de programación *C* y *C++*.

Se ha usado *Gephi* Bastian *et al.*, 2009 y *Cytoscape* <http://cytoscapeweb.cytoscape.org> como herramientas de visualización y análisis de redes complejas.

Así mismo, se ha usado *Power Bi* de *Microsoft* como herramienta de análisis estadístico de datos y sobre todo representación gráfica de datos, así como correlaciones.

Declaración

El presente trabajo es fruto de la investigación y redacción del autor, salvo los extractos y citas de documentos científicos y libros, que se van indicando mediante referencias a lo largo del texto. Los trabajos de investigación realizados se han hecho en colaboración con los directores del doctorado y con miembros de otros grupos de investigación, tales como el *grupo de investigación del Servicio de Microbiología, Hospital Universitario Ramón y Cajal, e Instituto Ramón y Cajal de Investigación*, en Madrid. Las imágenes, gráficas e ilustraciones son propias y generadas por el autor, extractos de los artículos científicos publicados por el autor y colaboradores. Salvo que se indice lo contrario, citando en ese caso la fuente original de la misma.

Madrid, 2019

Rafael Ángel Vida Delgado