



**Universidad**  
Zaragoza

TRABAJO DE FIN DE GRADO

---

**IMPACTO DE LA EVOLUCIÓN VIRAL EN  
LAS TRAYECTORIAS EPIDÉMICAS A  
TRAVÉS DE REDES DE GENOTIPOS**

---

*Autor:*

Hernán Gracia Lázaro

*Directores:*

Dr. Jesús Gómez Gardeñes

Santiago Lamata Otín

UNIVERSIDAD DE ZARAGOZA

FACULTAD DE CIENCIAS

Julio 2025

# Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Marco teórico</b>	<b>2</b>
2.1. Modelización de epidemias . . . . .	2
2.1.1. Modelos compartimentales . . . . .	2
2.1.2. Número reproductivo básico $R_0$ y umbral epidémico . . . . .	3
2.2. Mutación de patógenos . . . . .	3
2.3. Redes complejas . . . . .	4
2.3.1. Nociones teóricas y métricas estructurales . . . . .	4
2.3.2. Redes de genotipos reales . . . . .	6
<b>3. Modelo epidemiológico coevolutivo SIMS</b>	<b>8</b>
3.1. Dinámica compartimental . . . . .	8
3.2. Ecuaciones dinámicas . . . . .	9
3.3. Dinámica del modelo SIMS con una sola cepa . . . . .	11
3.4. Dinámica del modelo SIMS con múltiples cepas . . . . .	14
3.4.1. Dinámica bajo inmunidad permanente . . . . .	14
3.4.2. Resurgimiento de genotipos . . . . .	16
<b>4. Efecto de la topología del espacio antigénico en las trayectorias epidémicas</b>	<b>17</b>
4.1. Redes sintéticas: estructuras homogéneas vs. cúmulos de mutantes . . . . .	17
4.1.1. Endemicidad constante frente a epidemia estacional . . . . .	17
4.1.2. Inmunidad cruzada y cúmulos de mutantes . . . . .	18
4.2. Trayectorias epidémicas en redes de genotipos reales . . . . .	20
<b>5. Conclusiones</b>	<b>22</b>
<b>Referencias</b>	<b>24</b>
<b>Apéndices</b>	<b>25</b>
<b>A. Glosario de símbolos y parámetros</b>	<b>25</b>
<b>B. Propiedades estructurales de las redes de genotipos reales</b>	<b>26</b>
<b>C. Infectividad máxima en el modelo SIR</b>	<b>28</b>
<b>D. Desarrollo analítico: modelo SIMS con múltiples cepas</b>	<b>28</b>
<b>E. Métodos computacionales</b>	<b>30</b>

## 1. Introducción

Uno de los objetivos principales de la Física es comprender los fenómenos de la naturaleza a partir de sus componentes fundamentales. Para ello, la Física Estadística trata de encontrar una conexión entre las propiedades microscópicas individuales y los comportamientos macroscópicos observados. Esta perspectiva ha conseguido brindar un fundamento microscópico a conceptos como la temperatura y la entropía. En la misma línea, pero con el foco puesto en sistemas tradicionalmente fuera del ámbito físico, surge la Física de Sistemas Complejos, cuyas aplicaciones abarcan problemas como la sincronización de grupos de neuronas, la propagación de rumores o desinformación en redes sociales, y la estabilidad de ecosistemas ante perturbaciones [1]. Esta rama hace uso de la Física Estadística para explicar comportamientos colectivos derivados de la interacción de un gran número de elementos conectados entre sí. El comportamiento individual de cada elemento influye en los demás, dando lugar a fenómenos emergentes que no podrían explicarse estudiando cada uno de los elementos de manera aislada. En este contexto, las redes complejas o grafos son los objetos matemáticos que permiten definir la estructura de las interacciones entre elementos y estudiar el efecto de distintas topologías de conectividad sobre el comportamiento colectivo del sistema.

El estudio de la propagación de epidemias es uno de los campos principales de la Física de Sistemas Complejos, enfocado en entender cómo evoluciona la transmisión de patógenos en función de los patrones de contacto entre individuos para tener la capacidad de prevenir y actuar ante un eventual brote infeccioso. Las trayectorias epidémicas de algunos virus muestran un rango muy variado de comportamientos. Algunos cambios de tendencia pueden explicarse a través de la relación entre la dinámica de la epidemia y la evolución del virus, ya que los individuos infectados desarrollan una respuesta inmune basada en anticuerpos que neutralizan el virus. En consecuencia, para escapar de la presión inmune de los individuos, el virus muta a nuevas cepas que tienen propiedades antigénicas diferentes. Cada una de las cepas corresponde a un genotipo distinto del virus, y sus relaciones se pueden capturar mediante redes de genotipos. En estas redes, cada nodo representa un genotipo, y existen conexiones entre aquellos que solo se diferencian en una mutación puntual.

En este trabajo tomamos como referencia el modelo epidemiológico presentado en [2] para analizar las trayectorias epidémicas que tienen lugar al incluir mutaciones virales. Para ello, primero describimos el marco teórico de las redes de genotipos y después introducimos el modelo, que tiene en cuenta el contagio, las mutaciones entre cepas y la dinámica de inmunidad. Este modelo permite acoplar las distintas escalas temporales de estos procesos para analizar cómo afectan conjuntamente a la evolución de una epidemia. Además, el modelo tiene en cuenta tanto la memoria inmunológica que desarrollan los individuos que ya han superado una cepa como la adquisición de inmunidad cruzada a diferentes cepas. Una vez caracterizado el modelo, estudiaremos cómo la topología de la red de genotipos afecta a la evolución de las trayectorias epidémicas. En particular, mostraremos que cambiando únicamente la topología de la red, el estado endémico puede ser de prevalencia fija o con olas epidémicas. Finalmente, estudiaremos cómo son las trayectorias epidémicas utilizando redes de genotipos reales, y en particular caracterizaremos el recorrido a través de su mesoescala.

## 2. Marco teórico

### 2.1. Modelización de epidemias

La modelización matemática de epidemias es una herramienta esencial para comprender la evolución de una enfermedad infecciosa en una población y ser capaces de diseñar estrategias de control efectivas. El estudio de los procesos epidémicos busca describir y predecir la propagación de enfermedades infecciosas, teniendo en cuenta las características de los patógenos que las desencadenan y las de la población en la que se propagan. Entre las primeras, debemos considerar la capacidad que tienen los patógenos de mutar y cambiar su material genético para adaptarse a la inmunidad adquirida por la población.

Los primeros modelos desarrollados empleaban la hipótesis de campo medio, considerando poblaciones completamente mezcladas y pasando por alto las diferencias entre los individuos. Estos modelos sentaron las bases de la modelización de epidemias, pero no resultaban adecuados para describir la propagación de enfermedades en comunidades heterogéneas. Por ello, con el auge de la teoría de redes complejas, aparecieron modelos más sofisticados y realistas, incorporando la estructura y dinámica de la población. El desarrollo de estos modelos es hoy en día una herramienta crucial para la predicción y gestión de epidemias, como en el caso del virus SARS-CoV-2.

#### 2.1.1. Modelos compartimentales

Los modelos compartimentales en epidemiología permiten capturar la evolución de enfermedades infecciosas. Estos modelos dividen la población en varias categorías mutuamente excluyentes en función del estado de los individuos frente al patógeno, y describen cómo se producen las transiciones entre compartimentos [3].

Uno de los modelos más básicos es el modelo SIS, que divide a la población en dos grupos: susceptibles e infectados. En un instante de tiempo concreto, los individuos susceptibles ( $S$ ) son aquellos que pueden contraer el patógeno, mientras que los infectados ( $I$ ) poseen el patógeno y pueden infectar a los susceptibles. La evolución del estado de la población entre un instante  $t$  y otro  $t + \Delta t$  se obtiene a través de ciertas tasas de paso. Concretamente, un individuo susceptible puede infectarse según una tasa  $\beta$  por cada contacto con infectados, mientras que un infectado puede recuperarse espontáneamente y pasar a ser susceptible con una tasa  $\mu$  (véase la Fig. 2.1.a). Este modelo es útil para describir enfermedades que no otorgan inmunidad y en las que la población alcanza equilibrios dinámicos entre los nuevos contagios y las recuperaciones. El parámetro de orden de la dinámica es precisamente la fracción de infectados  $I^*$  en este equilibrio, también denominada nivel endémico.

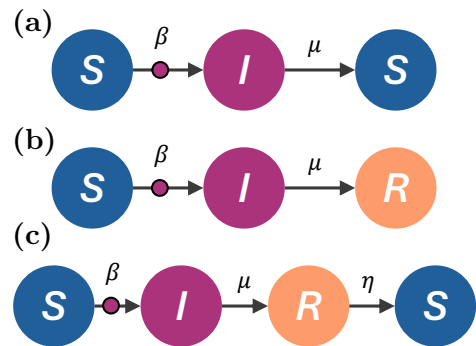


Fig. 2.1: **Modelos epidémicos compartimentales.** Se muestran los procesos epidemiológicos característicos de los modelos (a) SIS, (b) SIR, y (c) SIRS.

Otros modelos como el SIR o el SIRS incluyen otro compartimento para los individuos recuperados ( $R$ ), que han superado la enfermedad y desarrollado inmunidad ante ella (véase la Fig. 2.1.b y la Fig. 2.1.c). Los individuos se mantienen en este estado indefinidamente (modelo SIR) o vuelven a ser susceptibles tras un tiempo (modelo SIRS). Esta vuelta al compartimento de susceptibles representa la pérdida de inmunidad característica con el paso del tiempo. No obstante, estos modelos introducen la ganancia de inmunidad *ad hoc* mediante el nuevo compartimento, sin considerar la dinámica del proceso. En este trabajo, introduciremos la respuesta inmune de manera más realista.

### 2.1.2. Número reproductivo básico $R_0$ y umbral epidémico

El número reproductivo básico  $R_0$  es un concepto fundamental en epidemiología. Se define como el número promedio de infecciones secundarias que causa un individuo infectado en una población de individuos susceptibles durante toda su etapa infecciosa. En el caso del modelo SIS,  $R_0 = k\beta/\mu$ , con  $k$  la tasa de contactos promedio de los individuos. Su valor permite determinar si una enfermedad es capaz de establecerse en una población o no, y sirve para cuantificar cuán contagiosa es la enfermedad. Si  $R_0 > 1$ , la enfermedad tiende a propagarse y dar lugar a una epidemia en fase endémica, con una fracción de infectados en el equilibrio no nula,  $I^* \neq 0$ . Si  $R_0 < 1$ , las nuevas infecciones no logran compensar las recuperaciones, por lo que los infectados van disminuyendo con el tiempo hasta que la enfermedad desaparece, dando lugar a  $I^* = 0$ , en lo que se conoce como fase absorbente.

Se define el umbral epidémico como la infectividad mínima que debe tener el patógeno para desatar una epidemia. Es decir, el valor de  $\beta = \beta_c$  para el que se alcanza el punto crítico de transición de fase,  $R_0 = R_0^c = 1$ , como se muestra en la Fig. 2.2. En el caso del modelo SIS, el umbral epidémico se alcanza cuando la tasa de recuperación y la infectividad se igualan,  $\beta_c = \mu/k$ . En fase absorbente,  $\beta < \mu/k$ , el patógeno desaparece de la población y no hay epidemia en estado estacionario,  $I^* = 0$ . En cambio, cuando  $\beta > \mu/k$ , la enfermedad prolifera y alcanza una fracción de infectados estacionaria no nula.

## 2.2. Mutación de patógenos

Los virus de ARN presentan tasas de mutación elevadas, debido a los fallos en el proceso de replicación llevado a cabo por la proteína ARN polimerasa. No todas las mutaciones que sufre un patógeno resultan exitosas, pero algunas permiten al virus evadir la respuesta inmune adquirida por los huéspedes y, en consecuencia, prevalecer en la población. Esta capacidad de los virus para expandir su diversidad genética constituye un mecanismo fundamental para entender cómo se desarrollan las epidemias [4].

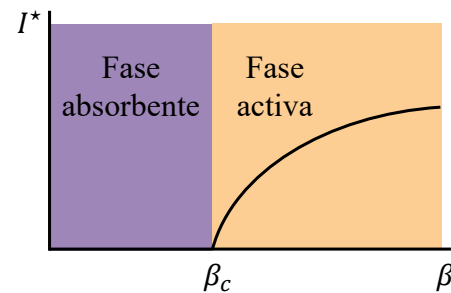


Fig. 2.2: **Diagrama de fases de la epidemia.** El umbral epidémico  $\beta_c$  separa las fases absorbente y activa.

Una cadena de ARN está compuesta por cuatro nucleótidos diferentes: adenina, uracilo, citosina y guanina. Los nucleótidos cumplen el criterio de complementariedad, emparejándose siempre la adenina con el uracilo y la citosina con la guanina. Esto ocurre de manera análoga al ADN, donde la timina sustituye al uracilo y la cadena es de doble hélice, frente a la cadena simple del ARN.

Las mutaciones puntuales son aquellas que se caracterizan por el cambio de un único nucleótido de una cadena de ARN. El espacio de mutaciones de una cadena de ARN comprende todas las secuencias que pueden formarse a partir de mutaciones puntuales de esa cadena. Por el criterio de complementariedad, al seleccionar un nucleótido en una posición de la cadena, queda determinado el nucleótido complementario. Por ello, si la cadena de ARN está constituida por  $N$  bases, existen  $2^N$  posibles secuencias de nucleótidos. En este trabajo estudiamos el fenómeno de las mutaciones y su importancia en el desarrollo de epidemias desde la perspectiva de las redes complejas, asociando cada secuencia con un nodo y las mutaciones puntuales con enlaces de la red.

## 2.3. Redes complejas

### 2.3.1. Nociones teóricas y métricas estructurales

Una red compleja comprende un conjunto de elementos o nodos y las conexiones que existen entre ellos. Se representa a través de un grafo,  $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ , donde  $\mathcal{N}$  es el conjunto de  $N$  nodos y  $\mathcal{L}$  es el conjunto de  $L$  enlaces o *links* entre ellos. Los nodos se etiquetan con un índice  $i = 1, 2, \dots, N$  y los enlaces son parejas  $(i, j)$  de nodos. En una red de genotipos, los nodos representan las  $N$  secuencias genéticas de un patógeno y cada enlace de la red conecta secuencias que difieren únicamente en una mutación puntual. Cada uno de estos genotipos se denota como  $s_i$ . Además, las redes de genotipos pueden construirse a distintas escalas genéticas: un nodo de la red puede representar secuencias de nucleótidos, de aminoácidos o hasta genes completos.

Para la aplicación en redes de genotipos, nos centramos en grafos no pesados y no dirigidos, por lo que todos los enlaces son bidireccionales y tienen el mismo peso. La matriz de adyacencia  $\mathbf{A} = \{a_{ij}\}$  recoge estas conexiones, de manera que  $a_{ij} = a_{ji} = 1$  si existe el enlace entre  $s_i$  y  $s_j$ , es decir,  $(i, j) \in \mathcal{L}$ , mientras que  $a_{ij} = a_{ji} = 0$  en caso contrario.

Otra matriz que captura la estructura de conexiones de la red y que es muy útil para el estudio de dinámicas difusivas es la matriz laplaciana normalizada por el grado de salida,  $\mathbf{L} = \{\ell_{ij}\}$ . Los elementos de esta matriz vienen dados por  $\ell_{ij} = \delta_{ij} - a_{ij}/k_j$ , donde  $k_i = \sum_{j=1}^N a_{ij}$  es el grado del genotipo  $i$ . Esta matriz cumple ciertas propiedades matemáticas relacionadas con sus autovalores  $\{\lambda_j\}$ , que ordenamos como  $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{N-1}$ . En primer lugar, la matriz laplaciana es semidefinida positiva, por lo que  $\lambda_j \geq 0 \forall j$ . Además, la suma de cualquier fila o columna se anula,  $\sum_{i=1}^N \ell_{ij} = \sum_{j=1}^N \ell_{ij} = 0$ , por lo que el primer autovalor siempre es nulo,  $\lambda_0 = 0$ , y su multiplicidad indica la cantidad de subgrafos conexos en la red. Finalmente, el autovalor  $\lambda_2$  se denomina conectividad algebraica y es una medida de la conectividad de un grafo: para valores menores de  $\lambda_2$ , el grafo presenta una estructura más modular. Este autovalor determina además la velocidad de convergencia de procesos como el de difusión que pueden tener lugar en la red.

También introducimos la matriz de distancia genética  $\mathbf{X} = \{x_{ij}\}$ , que recoge el número de enlaces entre dos secuencias genéticas. Suponiendo que el grafo es conexo, esta matriz mide la distancia del camino más corto entre cada par de genotipos. Como los enlaces de la red conectan secuencias que difieren en mutaciones puntuales, la matriz indica por cuántos nucleótidos difieren dos secuencias genéticas  $s_i$  y  $s_j$ . En la Fig. 2.3 se muestra un grafo y sus matrices, a modo de ejemplo.

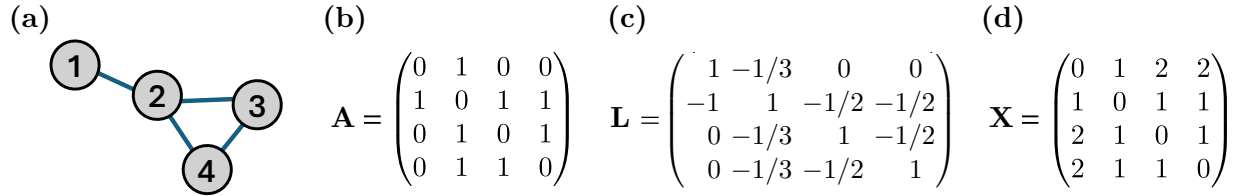


Fig. 2.3: **Matrices relacionadas con un grafo.** Se muestra el esquema del grafo (a), así como su matriz de adyacencia (b), matriz laplaciana normalizada (c) y matriz de distancias (d). El número en cada nodo del esquema del grafo indica la fila y columna correspondiente en las matrices.

La distribución de grado  $P(k)$  recoge la probabilidad de que un nodo del grafo escogido aleatoriamente tenga exactamente  $k$  vecinos. A partir de ello, se define el  $n$ -ésimo momento de la distribución como  $\langle k^n \rangle = \sum_{k=0}^{N-1} k^n P(k)$ . Cabe destacar que el primer momento  $\langle k \rangle$  es el valor medio del grado y el segundo momento  $\langle k^2 \rangle$  está relacionado con la varianza.

Por otra parte, en el estudio de redes complejas, se pueden definir comunidades como conjuntos de nodos más interconectados entre ellos que con el resto de la red. La existencia de comunidades en una red indica la riqueza en su mesoescala y puede influir en su dinámica. Para detectar estas regiones emplearemos el algoritmo *Infomap* [5] en Python, que emplea un modelo de caminata aleatoria para optimizar el flujo de información a través de la red, y de esta manera determina las comunidades.

También introducimos la longitud media del camino más corto entre dos nodos,

$$L(\mathcal{G}) = \frac{1}{N(N-1)} \sum_{i \neq j} x_{ij}, \quad (2.1)$$

donde  $x_{ij}$  es la distancia genética más corta entre las cepas  $s_i$  y  $s_j$ . En el caso de grafos no conexos,  $L(\mathcal{G}) = \infty$ . Por otro lado, el coeficiente de asortatividad se expresa como

$$r(\mathcal{G}) = \frac{\sum_{(i,j) \in \mathcal{L}} (k_i - \bar{k})(k_j - \bar{k})}{\sum_{(i,j) \in \mathcal{L}} (k_i - \bar{k})^2}. \quad (2.2)$$

Este coeficiente evalúa la similitud en los grados (número de conexiones) entre nodos, siendo  $\bar{k}$  el grado medio de todos los nodos en  $\mathcal{N}$ . Se define también el coeficiente de *clustering* o de agrupamiento, que cuantifica la tendencia de los nodos a formar triángulos. Es decir, mide la probabilidad de que dos nodos conectados a otro nodo  $i$  estén a su vez conectados entre sí, y viene dado por

$$C_i = \frac{\# \text{ triángulos conectados a } i}{\# \text{ posibles triángulos conectados a } i} = \frac{\sum_{j,k=1}^N a_{ij} a_{ik} a_{jk}}{k_i(k_i - 1)}, \quad (2.3)$$

con  $C_i = 0$  si  $k_i \leq 1$ . El *clustering* promedio de la red se calcula a partir de estos coeficientes como

$$C = \frac{1}{N} \sum_{i=1}^N C_i. \quad (2.4)$$

Por otro lado, definimos la transistividad  $T(\mathcal{G})$  mediante la expresión

$$T(\mathcal{G}) = 3 \frac{\# \text{ triángulos}}{\# \text{ triadas}}, \quad (2.5)$$

donde una triada es un conjunto de tres nodos con al menos dos de ellos conectados. Este coeficiente mide entonces la tendencia global de la red a formar tríos cerrados. Finalmente, se define la modularidad,

$$Q(\mathcal{G}) = \frac{1}{2L} \sum_{i,j \in \mathcal{N}} \left[ a_{ij} - \frac{k_i k_j}{2L} \right] \delta(g_i, g_j), \quad (2.6)$$

donde  $L = |\mathcal{L}|$  es el número de enlaces en la red y el factor  $\delta(g_i, g_j)$  toma el valor uno si los nodos  $i$  y  $j$  están en la misma comunidad, y cero en caso contrario. Por tanto, la modularidad  $Q(\mathcal{G})$  mide la densidad de enlaces intracomunitarios en relación con los enlaces intercomunitarios.

### 2.3.2. Redes de genotipos reales

Los avances en las técnicas de secuenciación genética han permitido identificar un gran número de genotipos relacionados con un mismo patógeno. Gracias a ello, y tras un tratamiento complejo de los datos disponibles, Williams et al. [6] fueron capaces de reconstruir la red de genotipos del virus de la influenza tipo A (representada en la Fig. 2.4), constituida a partir de las cepas muestreadas entre los años 2010 y 2020. A partir de esta red, podemos analizar la estructura evolutiva y la diversidad genética del virus, así como su impacto en las trayectorias epidémicas.

La red de genotipos del virus Influenza A (H3N2) aglutina una gran cantidad de secuencias genéticas distintas. En particular, la red se ha construido a partir de más de 25000 muestras del virus, que suponen unos 9700 nodos o cepas diferentes. Además, la estructura se compone de más de 3000 componentes conexas, la mayoría con tan solo uno o dos nodos. Entre estas componentes, las ocho más grandes contienen al menos 130 nodos. La Fig. 2.5.a muestra la segunda componente conexa más grande del virus, que presenta una topología compleja y muy diferente a la de las redes aleatorias. Su distribución de grado (véase Fig. 2.5.b) se caracteriza por una cola de muy pocos nodos con una conectividad muy elevada en la red, característica de las redes libres de escala.

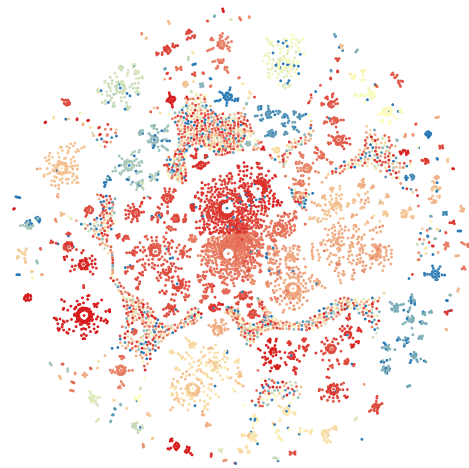


Fig. 2.4: **Red de genotipos del virus Influenza A.** Red real construida por Williams et al. [6]

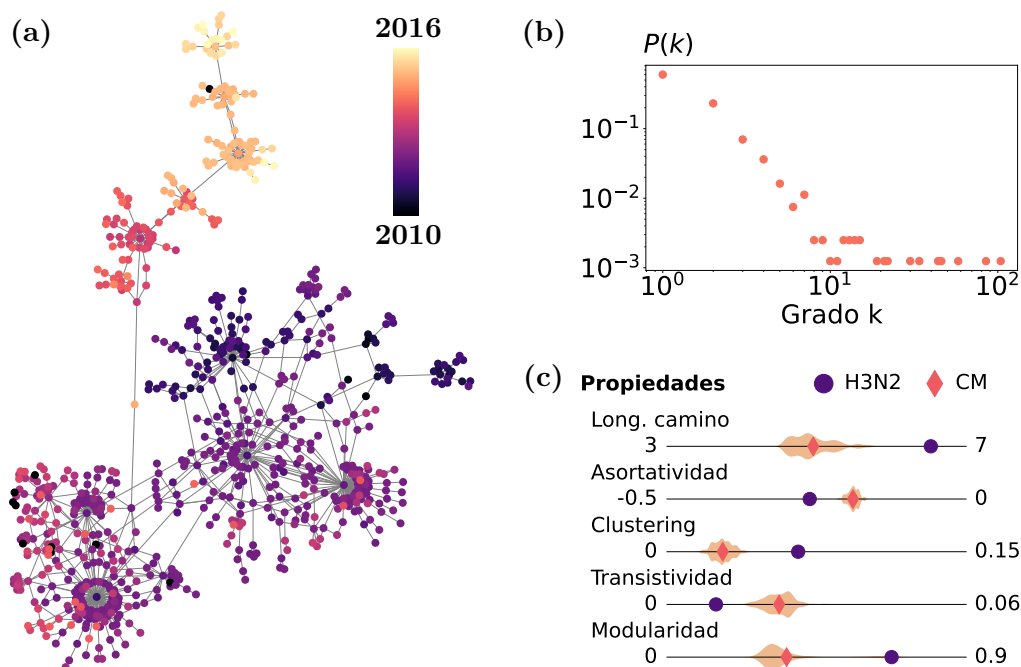


Fig. 2.5: **Propiedades estructurales de las redes de genotipos.** (a) Segunda componente conexas más grande de la red de genotipos de Influenza A, donde el color de los nodos representa la fecha de muestreo. (b) Distribución de grado de la segunda componente más grande de la red de genotipos de Influenza A. (c) Propiedades estructurales, los puntos se corresponden con la segunda componente de Influenza A, mientras que los gráficos de violín se corresponden a un conjunto de cien redes con la misma secuencia de grado.

Para comprender mejor si estas características estructurales son consecuencia de la distribución de grado o existen otros ingredientes involucrados, comparamos la red real del virus Influenza A con su equivalente aleatoria. En la Fig. 2.5.c se recogen las métricas estructurales de la segunda componente conexas más grande de la red real, junto con los valores correspondientes a un conjunto de cien redes aleatorias que conservan la distribución de grado (véase el Ap. E para más detalles sobre estas redes y el Ap. B para el análisis correspondiente a otras componentes de la red real).

Los resultados muestran diferencias significativas entre estos dos tipos de redes. En primer lugar, las redes de genotipos reales muestran una mayor longitud media del camino más corto que sus homólogas aleatorias, indicando una estructura de conexiones más enrevesada. Además, el coeficiente de agrupamiento elevado apunta a una mayor tendencia a formar triángulos, indicando la existencia de una organización jerárquica en el esqueleto de la red. Por otro lado, en la red real se obtienen valores superiores de modularidad (calculada a partir de las comunidades obtenidas con *Infomap*), lo que sugiere la organización de los nodos en escalas mesoscópicas. Por último, vemos que la red real de genotipos es disasortativa, por lo que los nodos con muchas conexiones tienden a estar unidos con aquellos de menor conectividad. Esto se explica por el proceso evolutivo de los virus, donde las cepas exitosas generan muchas variantes genéticamente similares. Todas estas diferencias estructurales motivan el análisis sobre cómo la topología de la red de genotipos afecta a las trayectorias epidémicas.

### 3. Modelo epidemiológico coevolutivo SIMS

Con el marco teórico establecido, nos centramos ahora en el modelo que pretende capturar la interacción entre las dinámicas de contagio, respuesta inmune y mutación. La dificultad de integrar todos los mecanismos en un mismo modelo reside en que estos procesos operan en diferentes escalas espaciales y temporales. El contagio surge de la interacción entre individuos, mientras que la respuesta inmune aparece dentro de un individuo como consecuencia de la interacción entre el virus y los anticuerpos, y la mutación depende de la dinámica interna de la población del virus.

#### 3.1. Dinámica compartimental

En esta sección introducimos el modelo SIMS (Susceptible-Infectado-Mutación-Susceptible), que unifica los tres procesos mediante un enfoque compartimental. Se trata de una extensión del SIS que incorpora la respuesta inmune y la diversidad genética de los virus a través de sus diferentes cepas. En el modelo SIMS, cada individuo de la población puede encontrarse en el compartimento de susceptibles ( $S$ ) o en uno de los  $n$  compartimentos de infectados ( $I_i$ ), correspondientes con cada cepa del virus,  $s_i \in \mathcal{N}$ . Una de las hipótesis clave del modelo es que la dinámica interna del huésped ocurre en una escala temporal mucho más rápida que la propagación epidemiológica. Esto se traduce en que se producen suficientes ciclos de replicación viral durante una infección, de forma que una única cepa dominante determina por completo el estado infeccioso del individuo.

La estructura del modelo SIMS, representada en la Fig. 3.1.a, sirve de base para la descripción de sus principales procesos.

- Contagio: un individuo susceptible ( $S$ ) contrae una enfermedad por contacto directo con individuos infecciosos ( $I_i$ ), con una tasa de infección  $\beta$  independiente de la variante del virus. Tras la infección, el individuo susceptible transita al compartimento  $I_i$  de quien le contagia el virus, adquiriendo su misma cepa  $s_i$ . Permanece infectado de media una duración de  $\mu_i^{-1}$  unidades temporales antes de recuperarse y volver al compartimento  $S$ .
- Respuesta inmune: la tasa de recuperación  $\mu_i$  es distinta para cada cepa y no es constante en el tiempo, sino que depende del historial de contagio de los individuos (véase la Fig. 3.1.b). Este parámetro parte de un valor de referencia  $\mu_0$  y aumenta proporcionalmente a la carga de la cepa  $s_i$  en la población, modulada por la tasa de adquisición de inmunidad  $\alpha$  y la sensibilidad del sistema inmune  $\zeta$ . La inmunidad adquirida decae con el tiempo hacia  $\mu_0$  con un ritmo característico de pérdida de inmunidad  $\gamma$ . Además, los individuos consiguen inmunidad cruzada para cepas genéticamente similares a la de infección. Esta inmunidad decrece con la distancia genética entre variantes, con una longitud característica de decaimiento  $\Delta$ .
- Mutación: la evolución viral se modeliza como un proceso de difusión en la red de genotipos. Un individuo infectado de la cepa  $s_i$  puede experimentar una mutación del patógeno a una cepa vecina  $s_j$ , con una tasa  $D_x$ . Cuando esta mutación sucede, el individuo transita del compartimento  $I_i$  al  $I_j$ , creando un ciclo entre la mutación y la dinámica de contagio.

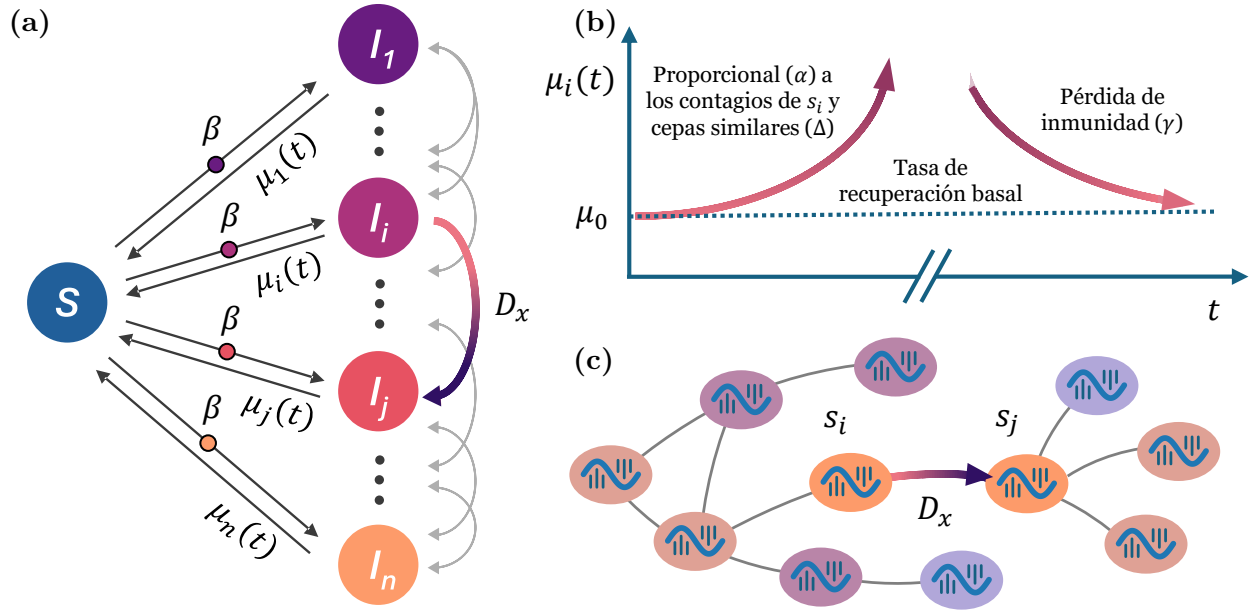


Fig. 3.1: **Esquemas del modelo dinámico SIMS.** El panel (a) muestra la estructura compartimental, donde los individuos susceptibles ( $S$ ) pueden infectarse con una tasa  $\beta$  ante el contacto con individuos infectados de la cepa  $s_i$ . Al hacerlo, transitan al compartimento correspondiente de entre los  $n$  compartimentos infecciosos ( $I_i$ ), cada uno relacionado con una variante del virus. Los individuos infectados se recuperan con una tasa  $\mu_i$  dependiente de cada cepa, volviendo al compartimento de susceptibles. También se indica la transición entre compartimentos infecciosos por la mutación del patógeno en los individuos. El panel (b) ilustra la respuesta inmune adaptativa de la población. La tasa de recuperación  $\mu_i$  aumenta proporcionalmente al peso de la cepa  $s_i$  en la población y al de las cepas similares genéticamente, cuya proximidad en la red de genotipos es modulada por el parámetro  $\Delta$ . Este crecimiento se contrarresta con el decaimiento de la inmunidad con el tiempo, con un ritmo  $\gamma$ , hacia su valor basal de referencia,  $\mu_0$ . El panel (c) representa la evolución viral a través de mutaciones en la red de genotipos. Los individuos infectados de una cepa  $s_i$  pueden transitar en este espacio a una cepa adyacente  $s_j$ , con una tasa de mutaciones  $D_x$ . Este mecanismo, indicado por las flechas coloreadas en los paneles (a) y (c), establece un ciclo entre las dinámicas de contagio y mutación.

A modo de resumen, considerando  $n$  genotipos de un patógeno, el modelo SIMS está constituido por  $n + 1$  compartimentos, de manera que los individuos susceptibles y los infectados por cada cepa pertenecen a un compartimento individual. Además, la evolución del sistema depende de seis parámetros epidemiológicos ( $\beta$ ,  $\alpha$ ,  $\zeta$ ,  $\gamma$ ,  $\Delta$ ,  $D_x$ ), permitiendo capturar la interacción entre los tres mecanismos de evolución en poblaciones virales diversas.

### 3.2. Ecuaciones dinámicas

La dinámica del modelo SIMS puede capturarse mediante un conjunto de ecuaciones diferenciales acopladas. Para simplificar el estudio de la epidemia, suponemos una población homogénea bien mezclada, donde todos los individuos son equivalentes desde el punto de vista microscópico frente al virus. Esta consideración conduce a una descripción de campo medio de la población, por lo que las variables relevantes en cada instante de tiempo son la probabilidad  $\rho^S(t)$  de ser susceptible y la probabilidad de estar infectado de cada una de las cepas virales,  $\rho_i^I(t)$ , con  $i = 1, \dots, n$ .

Teniendo en cuenta también que la tasa de recuperación de cada cepa  $\mu_i(t)$  evoluciona con el tiempo, podemos formular las ecuaciones diferenciales que describen la evolución del modelo SIMS,

$$\dot{\rho}_i^I(t) = \beta \rho_i^I(t) \rho^S(t) - \mu_i(t) \rho_i^I(t) - D_x \sum_{j=1}^n \ell_{ij} \rho_j^I(t), \quad (3.1)$$

$$\dot{\mu}_i(t) = \alpha \sum_{j=1}^n \left[ \frac{\rho_j^I(t)}{\zeta^{-1} + \rho_j^I(t)} e^{-x_{ij}/\Delta} \right] - \gamma (\mu_i(t) - \mu_0). \quad (3.2)$$

Como puede verse, las matrices y parámetros relevantes introducidos con anterioridad aparecen en las ecuaciones. Este conjunto de  $2n$  ecuaciones describe completamente la evolución del sistema, dado que la variable  $\rho^S(t)$  restante puede expresarse en función del resto. Concretamente, como la suma de las probabilidades de los estados de la población debe ser la unidad, se obtiene que

$$\rho^S(t) = 1 - \sum_{i=1}^n \rho_i^I(t). \quad (3.3)$$

Vamos ahora a explicar cada término de las Ecs. (3.1) y (3.2), y cómo se relacionan con los tres mecanismos fundamentales del modelo. Los dos primeros términos de la Ec. (3.1) describen, para cada variante del virus, el proceso de contagio típico de un modelo SIS con campo medio. La infección ocurre con una tasa efectiva  $\beta$ , y es proporcional a la cantidad de individuos infectados de la cepa y a los individuos susceptibles. La recuperación viene marcada por una tasa  $\mu_i(t)$  distinta para cada genotipo.

A diferencia del modelo SIS, las tasas de recuperación evolucionan de forma dinámica con el tiempo para incorporar el efecto de la respuesta inmune, siguiendo la Ec. (3.2). El primer término de la ecuación captura el aumento de la tasa de recuperación conforme se adquiere inmunidad. Concretamente, la ganancia es proporcional a la tasa de adquisición de inmunidad  $\alpha$  y a la carga del genotipo  $s_i$ , que se supone proporcional a la probabilidad  $\rho_i(t)$  de infección por el patógeno. Este término de la ecuación introduce la memoria inmune como un término de inercia, permitiendo al sistema de manera efectiva recordar infecciones superadas. Además, esta respuesta se regula a través de una función de Hill, donde se introduce el parámetro de sensibilidad  $\zeta$ . Cuando  $\zeta \rightarrow 0$ , la ganancia de inmunidad es despreciable, mientras que  $\zeta \rightarrow \infty$  conlleva el máximo valor,  $\alpha$ .

Además de la inmunidad obtenida por la infección directa, la Ec. (3.2) considera la adquisición de inmunidad cruzada entre cepas relacionadas. Este efecto está modulado a través de una exponencial que depende de la distancia genética entre las variantes, almacenada en la matriz  $\mathbf{X} = \{x_{ij}\}$ . Cuanto más parecidas son dos cepas en cuanto a su material genético, mayor es la inmunidad que ganan los individuos a una de ellas tras contagiarse de la otra. El alcance de la inmunidad cruzada se controla a través de la longitud característica  $\Delta$ . Cuando  $\Delta \rightarrow 0$ , la inmunidad actúa aisladamente para cada cepa, mientras que  $\Delta \rightarrow \infty$  conlleva máxima inmunidad cruzada entre todas las cepas. Por último, el segundo término de la Ec. (3.2) incorpora el efecto de la pérdida progresiva de inmunidad

a través de una aproximación de tiempo de relajación  $\tau$ . Así, la inmunidad decae hacia su valor de referencia  $\mu_0$  con un ritmo  $\gamma = 1/\tau$ .

Nos queda analizar el último término de la Ec. (3.1), que da cuenta de la dinámica de mutación, de forma que los individuos infectados pueden cambiar de compartimento infeccioso durante su periodo de contagio. Este proceso se captura mediante un término de difusión a través de la red de genotipos, gobernado por la matriz laplaciana normalizada  $\mathbf{L} = \{\ell_{ij}\}$ . Además, la importancia de este término viene determinada por la tasa de mutación  $D_x$ .

Una vez analizadas las ecuaciones que describen el modelo, pasamos a estudiar los fenómenos que emergen al considerar la dinámica SIMS sobre distintas redes de genotipos. Nos interesa especialmente la fracción total de individuos infectados en la población  $I(t)$  y su valor estacionario  $I^*$ , que vienen dados por

$$I(t) = \sum_{i=1}^n \rho_i(t), \quad (3.4)$$

$$I^* = \lim_{t \rightarrow \infty} I(t). \quad (3.5)$$

Además, en caso de que la red de genotipos presente diferentes comunidades, definimos la prevalencia relativa como la fracción de infectados debida a variantes pertenecientes a la comunidad ( $I_{com}$ ) en relación con la fracción de infectados total:

$$I^{rel}(t) = \frac{I_{com}(t)}{I(t)}. \quad (3.6)$$

### 3.3. Dinámica del modelo SIMS con una sola cepa

En primer lugar, consideramos el modelo SIMS con una sola cepa,  $n = 1$ . De esta manera podemos ver cómo se captura la dinámica epidemiológica básica. Fijamos los parámetros  $\mu_0 = 0.1$  y  $\beta = 0.3$ , correspondientes con un número reproductivo básico  $R_0 = \beta/\mu_0 = 3$ . Tras integrar las ecuaciones para varios casos, se obtienen las trayectorias de la Fig. 3.2.a. En ausencia de dinámica de inmunidad, es decir,  $\alpha = \gamma = 0$ , el modelo recupera el comportamiento del SIS, donde  $I(t)$  presenta un crecimiento inicial exponencial que se estabiliza a un estado endémico (línea azul). Si introducimos adquisición de inmunidad ( $\alpha \neq 0$ ) y mantenemos las pérdidas despreciables ( $\gamma = 0$ ), el sistema ya no mantiene el estado endémico. En lugar de ello se observa una ola que finaliza en un estado absorbente  $I^* = 0$  (línea naranja), igual que en un modelo SIR. Además, con los parámetros introducidos,  $\alpha = 0.036$ ,  $\zeta = 1$ , el pico alcanza una fracción de infectados de  $I_{max} \approx 0.3$ , que concuerda con los valores del SIR para  $R_0 = 3$  (véase el Ap. C). Por último, cuando incluimos la ganancia y la pérdida de inmunidad ( $\alpha, \gamma \neq 0$ ), en concreto  $\gamma = 0.02$ , se observa que la ola epidémica converge a un estado endémico (línea rosa), con una trayectoria característica del modelo SIRS.

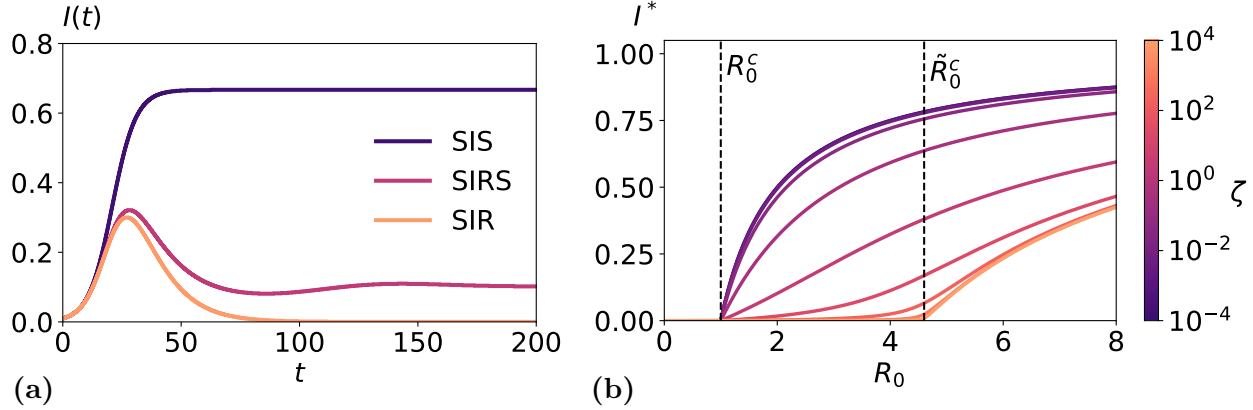


Fig. 3.2: **Dinámica del modelo SIMS con una sola cepa.** (a) Comportamientos dinámicos capturados por el modelo SIMS en ausencia de mutaciones ( $D_x = 0$ ). El régimen endémico tipo SIS se corresponde con  $\alpha = \gamma = 0$ , la ola epidémica tipo SIR se corresponde con  $\alpha = 0.036$  y  $\gamma = 0$ , y el comportamiento tipo SIRS, donde se alcanza la endemividad tras una primera ola, se ha obtenido con los valores  $\alpha = 0.036$  y  $\gamma = 0.02$ . (b) Diagramas epidémicos del modelo SIMS con una sola cepa, correspondientes a la Ec. (3.12). Se fijan los valores  $D_x = 0$ ,  $\alpha = 0.036$  y  $\gamma = 0.1$ , y se toman distintos valores de la sensibilidad inmune  $\zeta$ . Para ambas gráficas se ha tomado  $\mu_0 = 0.1$ , y en el caso del panel (a),  $\beta = 0.3$  y  $\zeta = 1$ .

Para caracterizar los estados endémicos que alcanzan las trayectorias de la Fig. 3.2.a, derivamos analíticamente las expresiones de las soluciones estacionarias. Como hay una sola cepa,  $\rho_i(t) = \rho(t) = I(t)$  y  $\mu_i(t) = \mu(t)$ , por lo que las Ecs. (3.1) y (3.2) se convierten en

$$\dot{I}(t) = \beta I(t) (1 - I(t)) - \mu(t)I(t), \quad (3.7)$$

$$\dot{\mu}(t) = \alpha \frac{I(t)}{\zeta^{-1} + I(t)} - \gamma (\mu(t) - \mu_0). \quad (3.8)$$

En estado estacionario,  $\dot{\mu} = \dot{I} = 0$ , siendo válida la solución trivial correspondiente a un estado absorbente de la epidemia:  $I_0^* = 0$ ,  $\mu^* = \mu_0$ . Si descartamos esta solución y tomamos  $I^* \neq 0$ , a partir de la Ec. (3.7) tenemos que

$$\mu^* = \beta (1 - I^*). \quad (3.9)$$

Nos queda introducir esta relación en la Ec. (3.8), de manera que

$$0 = \alpha \frac{I^*}{\zeta^{-1} + I^*} - \gamma (\beta (1 - I^*) - \mu_0). \quad (3.10)$$

Agrupando los términos en potencias de  $I^*$ , llegamos a la expresión

$$\gamma\beta (I^*)^2 + \left( \alpha + \gamma (\mu_0 - \beta (1 - \zeta^{-1})) \right) I^* + \gamma (\mu_0 - \beta) \zeta^{-1} = 0, \quad (3.11)$$

que se trata de una ecuación de segundo grado,  $A (I^*)^2 + BI^* + C = 0$ , cuyas soluciones son

$$I_{\pm}^* = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \quad (3.12)$$

y donde los coeficientes tienen las siguientes expresiones explícitas:

$$A = \gamma\beta, \quad (3.13)$$

$$B = \alpha + \gamma \left( \mu_0 - \beta \left( 1 - \zeta^{-1} \right) \right), \quad (3.14)$$

$$C = \gamma (\mu_0 - \beta) \zeta^{-1}. \quad (3.15)$$

La solución positiva de la ecuación de segundo orden toma valores positivos por encima del umbral epidémico, por lo que define el estado endémico del modelo. En la Fig. 3.2.b se representa la prevalencia estacionaria  $I^*$  en función del número reproductivo básico  $R_0$ , teniendo en cuenta que  $R_0 = \beta/\mu_0$ . Para ello, se escogen los parámetros  $\alpha = 0.036$  y  $\gamma = 0.1$ , junto con  $\mu_0 = 0.1$ . Cada curva se corresponde con un valor diferente de la sensibilidad del sistema inmunitario  $\zeta$ .

En la Fig. 3.2.b se observa que, cuando no hay sensibilidad ( $\zeta \rightarrow 0$ , línea azul), se recupera el comportamiento del modelo SIS. Este viene caracterizado por una transición continua en el umbral epidémico ( $R_0^c = 1$ ), desde un estado absorbente cuando  $R_0 < 1$  hacia un estado endémico cuando  $R_0 > 1$ . En el estado endémico se obtiene además la prevalencia característica del SIS,  $I^* = 1 - R_0^{-1}$ . Sin embargo, vemos que cuando la sensibilidad del sistema inmunitario es muy elevada ( $\zeta \rightarrow \infty$ , línea amarilla), el umbral epidémico se desplaza a un valor superior, dado por

$$\tilde{R}_0^c = 1 + \frac{\alpha}{\mu_0\gamma}, \quad (3.16)$$

y la prevalencia estacionaria pasa a estar dada por la relación

$$I^* = 1 - R_0^{-1} - \frac{\alpha}{\gamma\beta}. \quad (3.17)$$

En ambos casos ( $\zeta \rightarrow 0$  y  $\zeta \rightarrow \infty$ ), cuando  $R_0 = R_0^c$  y  $R_0 = \tilde{R}_0^c$ , respectivamente, se produce un cambio en la estabilidad que puede describirse como una bifurcación transcítica. En efecto, el punto fijo  $I_0^* = 0$  pasa de ser estable a inestable conforme  $R_0$  aumenta, y viceversa para  $I_+^*$ .

Para valores intermedios de la sensibilidad inmune, las curvas se sitúan entre los dos casos extremos. En el caso  $\zeta = 0$ , el sistema se encuentra en estado endémico para todos los valores de  $R_0$  por encima del umbral epidémico. Sin embargo, fijándonos en el caso  $\zeta = 1$ , el estado endémico tiende al estado absorbente conforme  $R_0 \rightarrow R_0^c$ . El cambio en el umbral epidémico mostrado por la Ec. (3.16) aumenta con la tasa de adquisición de inmunidad  $\alpha$  y disminuye con el ritmo de pérdida inmunitaria  $\gamma$ . Esto indica que el desplazamiento  $\alpha/(\mu_0\gamma)$  refleja el impacto relativo de la prevalencia inmunitaria ( $\alpha/\gamma$ ) respecto a la tasa de recuperación basal  $\mu_0$ . En el caso de inmunidad duradera ( $\gamma = 0$ ), el umbral  $\tilde{R}_0^c$  diverge al infinito, haciendo que las trayectorias epidémicas se sitúen en  $I^* \rightarrow 0$ . En consecuencia, sin evolución viral, el patógeno no es capaz de mantenerse en la población.

### 3.4. Dinámica del modelo SIMS con múltiples cepas

Una vez analizado el caso de una sola cepa, pasamos a considerar múltiples variantes del virus, de manera que el proceso de mutación se une a los de contagio y recuperación. Primero suponemos que la pérdida de inmunidad tiene lugar en una escala temporal mucho más lenta que la adquisición de inmunidad ( $\gamma \rightarrow 0$ ). Esto concuerda con lo observado en virus como Influenza y SARS-CoV-2, donde la inmunidad perdura mucho tiempo después de la recuperación [7]. Posteriormente estudiaremos cómo la pérdida de inmunidad ( $\gamma \neq 0$ ) puede llevar al resurgimiento de cepas. Durante toda la sección, establecemos para todas las simulaciones que la población es inicialmente susceptible salvo por una pequeña fracción infectada de la primera cepa ( $i = 1 \equiv$  cepa salvaje), concretamente  $\rho_1(t) = 0.01$  y  $\rho_{i \neq 1} = 0$ . Fijamos también la tasa de adquisición de inmunidad,  $\alpha = 0.036$ , y la sensibilidad del sistema inmune  $\zeta = 1$ .

#### 3.4.1. Dinámica bajo inmunidad permanente

Cuando un individuo adquiere inmunidad ante una cepa viral, el virus solo puede sobrevivir a través de mutaciones en el espacio antigénico. Tomamos como red de genotipos una cadena lineal (donde un nodo  $i$  está conectado a los dos adyacentes, el anterior  $i - 1$  y el posterior  $i + 1$ ) y simulamos la evolución del sistema, obteniendo las trayectorias epidémicas de la Fig. 3.3.a. Hemos fijado para ello una tasa de mutación  $D_x = 10^{-5}$ , en base a las estimaciones para la Influenza y otros virus ARN [8]; mientras que la inmunidad cruzada no se considera por el momento,  $\Delta = 0$ . A diferencia de la trayectoria tipo SIR de la Fig. 3.2.a, ahora el virus es capaz de sobrevivir más allá del primer brote, gracias a su capacidad de mutación. Como se observa en el cuadro interno de la Fig. 3.3.a, la inmunidad adquirida ante un genotipo hace que su número reproductivo básico caiga por debajo de la unidad. Sin embargo, el virus es capaz de persistir en la población mediante el escape antigénico, produciendo brotes sucesivos en las cepas adyacentes. A nivel macroscópico, esto se corresponde con una onda viajera en el espacio antigénico y en el tiempo, lo cual concuerda con los resultados de Rouzine et al. [9], quienes mostraron que este fenómeno se produce en espacios antigénicos unidimensionales.

Veamos ahora el efecto que tiene la inmunidad cruzada entre genotipos similares en la dinámica epidémica. Para ello, introducimos una longitud característica de inmunidad no nula,  $\Delta = 3$ , y observamos que únicamente las cepas suficientemente diferentes pueden presentar brotes (véase la Fig. 3.3.b). Esto es debido a que la inmunidad cruzada reduce el número reproductivo de las cepas cercanas por debajo del umbral epidémico. En el caso de la red de genotipos lineal, la cepa que presenta el segundo brote es la más cercana a la inicial que cumple  $\beta/\mu_i > 1$  (suponiendo que después de la primera ola  $\rho^S \approx 1$ ). Como  $\mu_i > \mu_0$ , este segundo brote tiene un impacto menor que el inicial y no genera una inmunidad tan fuerte para las siguientes cepas. Esto explica la onda viajera de menor prevalencia que se forma tras el segundo brote, similar a la trayectoria tipo SIRS observada en la Fig. 3.2.a.

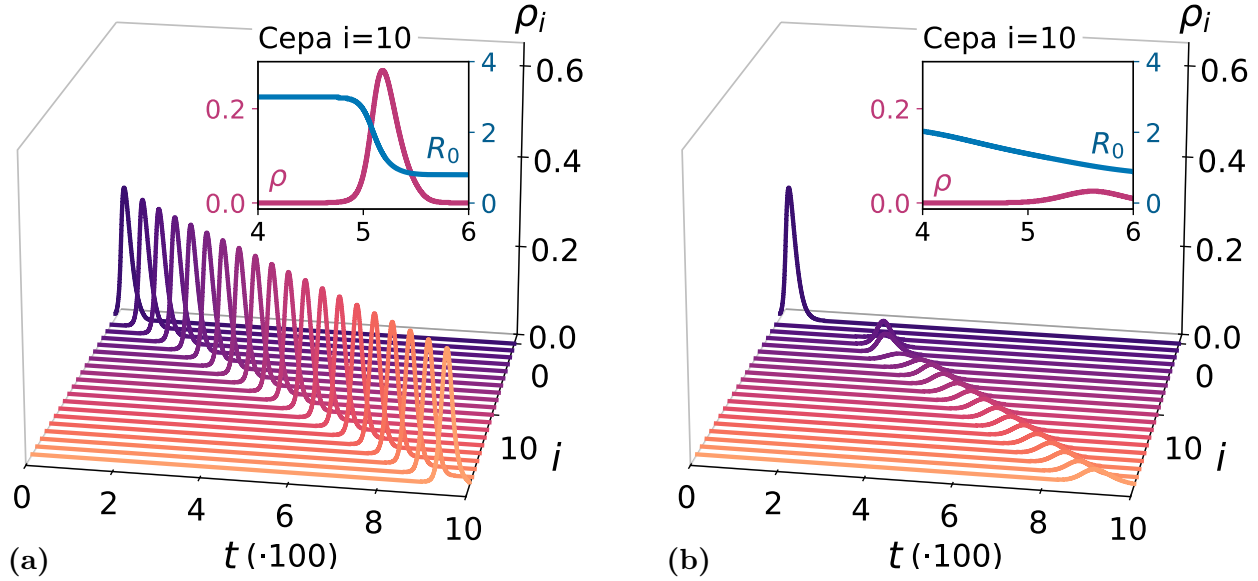


Fig. 3.3: **Dinámica del modelo SIMS con múltiples cepas.** Se emplea como red de genotipos una cadena lineal. En el panel (a) no se considera inmunidad cruzada,  $\Delta = 0$ , mientras que en el (b) se toma  $\Delta = 3$ . En ambos casos se observa el escape antigénico, con una dinámica epidémica diferente. Se han fijado además los parámetros  $\beta = 0.3$ ,  $\mu_0 = 0.1$ ,  $\alpha = 0.036$ ,  $\gamma = 0$ ,  $\zeta = 1$  y  $D_x = 10^{-5}$ .

Cabe destacar que la prevalencia endémica en la Fig. 3.3.b, donde  $\Delta \neq 0$ , es menor que en la Fig. 3.3.a, donde  $\Delta = 0$ . Esto se puede entender como una consecuencia del desplazamiento de los umbrales epidémicos por la inmunidad cruzada, lo cual afecta a la prevalencia endémica. Suponiendo que  $x_{ij} = 0$  si  $i = j$  y  $x_{ij} = \langle x \rangle$  en el resto de casos, puede obtenerse (véase el Ap. D) que, para  $\zeta \rightarrow \infty$ , el umbral epidémico viene dado por

$$\tilde{R}_0^c = 1 + \frac{\alpha}{\mu_0 \gamma} \left[ 1 + (N - 1) e^{-\langle x \rangle / \Delta} \right], \quad (3.18)$$

mientras que la prevalencia de la fase endémica, bajo aproximación de campo medio, resulta

$$I^* = 1 - R_0^{-1} - \frac{\alpha}{\gamma \beta} \left[ 1 + (N - 1) e^{-\langle x \rangle / \Delta} \right]. \quad (3.19)$$

Por lo tanto, desde un punto de vista macroscópico, la inmunidad cruzada potencia la respuesta inmune, ya que actúa como un incremento efectivo de  $\alpha$  y conduce a un estado de menor prevalencia endémica y mayor umbral epidémico efectivo. Comparando las gráficas de la Fig. 3.3 observamos que, conforme la longitud característica de inmunidad cruzada aumenta, la prevalencia tras el brote inicial disminuye. Esto implica que solo unos pocos individuos de la población sufren la propagación de variantes genéticamente diferentes, lo cual sugiere que la extinción viral es muy probable en entornos estocásticos realistas. A pesar de ello, en virus persistentes es habitual observar resurgimientos en distintas regiones del espacio antigénico, planteando nuevas cuestiones sobre el rol de los individuos inmunodeprimidos en la evolución viral, ya que mantienen la infección durante periodos extensos.

### 3.4.2. Resurgimiento de genotipos

En el análisis anterior, hemos partido de la suposición de que la pérdida de inmunidad opera en tiempos mucho mayores que el resto de procesos. En este apartado analizamos cómo afecta a la dinámica de la epidemia considerar una relación menos restrictiva entre estas escalas temporales. Concretamente, se observa que esta relación determina la reaparición de los genotipos. En la Fig. 3.4.a presentamos la evolución de la fracción total de infectados para distintos valores del ritmo de pérdida de inmunidad, donde se ha empleado una red tipo cadena lineal de 20 nodos para representar el espacio antigénico. Si la escala temporal de pérdida de inmunidad es similar a la de adquisición (en este caso  $\alpha = 1$ ), la población no obtiene inmunidad efectiva, por lo que los genotipos coexisten sin presión inmune.

En el caso contrario, si el proceso de pérdida inmune tiene lugar en periodos mucho más lentos, la fracción de infectados de las cepas decae tras un primer brote, ya que el sistema inmune de la población todavía recuerda el patógeno. Sin embargo, tras un tiempo del orden de  $\tau = 1/\gamma$ , la población ha perdido la inmunidad. En consecuencia, el genotipo resurge y es capaz de contagiar de nuevo a los individuos. Además, tras un periodo transitorio con varias reapariciones, la prevalencia alcanza un valor estacionario estable, que es menor cuanto mayor sea la escala temporal de pérdida de inmunidad (en concordancia con la Ec. (3.19)).

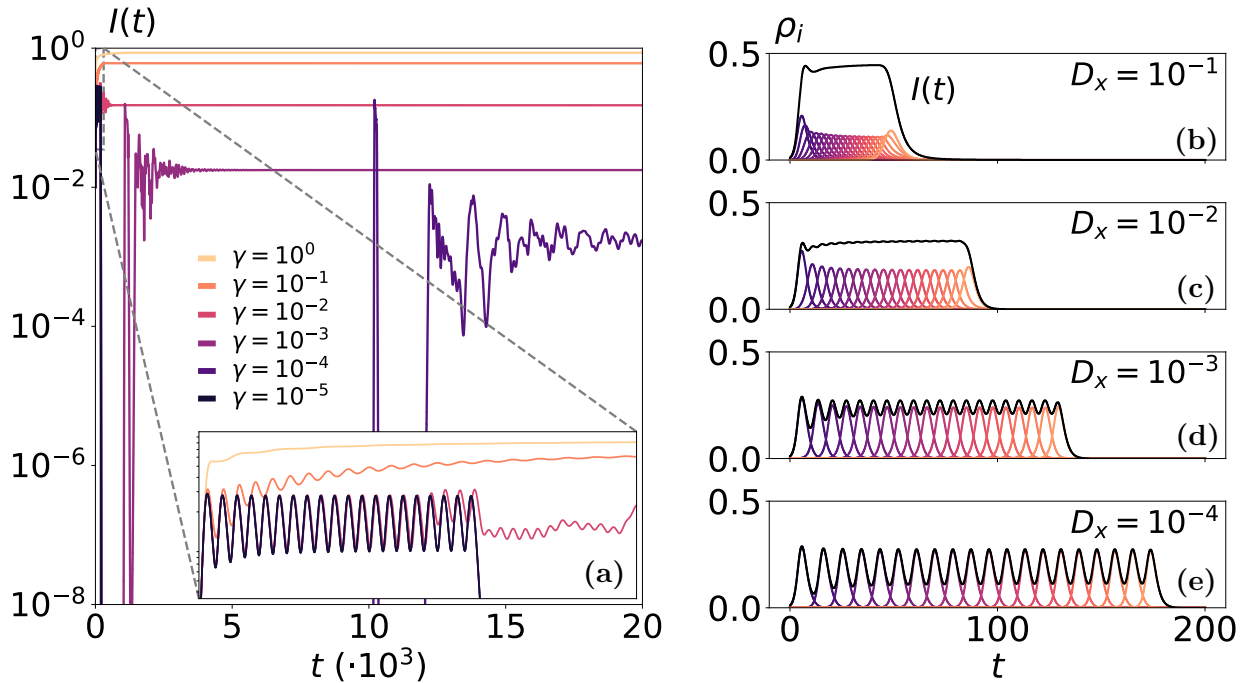


Fig. 3.4: **Resurgimiento de genotipos y exploración del espacio antigénico.** (a) Trayectorias epidémicas para distintos valores del ritmo de pérdida inmunitaria, donde se observa el fenómeno del resurgimiento de genotipos. Se han utilizado los valores  $\beta = 1$ ,  $\mu_0 = 0.1$ ,  $\alpha = 1$ ,  $\zeta = 1$ ,  $\Delta = 0$  y  $D_x = 10^{-4}$ . (b) Las trayectorias para distintos valores de la tasa de mutación muestran la diferencia entre la difusión y el escape antigénico como mecanismos de perpetuación de un patógeno. Los parámetros son  $\beta = 1$ ,  $\mu_0 = 0.1$ ,  $\alpha = 1$ ,  $\zeta = 1$ ,  $\Delta = 0$  y  $\gamma = 0$ . En ambos paneles se emplea una red lineal de 20 nodos como red de genotipos.

Por otro lado, las Figs. 3.4.b-e muestran cómo la tasa de mutación afecta a la trayectoria epidémica. Observamos que cuando la tasa de mutación es alta, la dinámica de la epidemia viene marcada por la difusión del virus por el espacio antigénico. Por contra, si la tasa de mutación es baja, es la inmunidad adquirida ante una cepa lo que provoca el brote de cepas adyacentes en el espacio antigénico. Esto revela que el orden de magnitud de la tasa de mutación supone la diferencia entre una dinámica dirigida por la difusión o por el escape antigénico a la obtención de inmunidad.

## 4. Efecto de la topología del espacio antigénico en las trayectorias epidémicas

Una vez caracterizado el modelo SIMS y la interacción entre los procesos de contagio, respuesta inmune y mutación en casos sencillos, continuamos estudiando cómo la estructura de la red de genotipos afecta a las trayectorias epidémicas.

### 4.1. Redes sintéticas: estructuras homogéneas vs. cúmulos de mutantes

#### 4.1.1. Endemicidad constante frente a epidemia estacional

Las enfermedades endémicas exhiben un amplio rango de patrones temporales de incidencia en las poblaciones. Por ejemplo, afecciones como la Influenza suelen presentar fluctuaciones estacionales casi periódicas, mientras que enfermedades como la tuberculosis [10] y la malaria [11] mantienen niveles de incidencia estables en muchas regiones africanas. Estas tendencias tan diferentes se deben a diversos factores, entre los que destaca la estructura del espacio antigénico. En particular, la topología de este espacio resulta esencial para explicar la evolución viral de patógenos emergentes como el SARS-CoV-2. Mientras que sus primeras variantes reflejan un espacio antigénico de agrupaciones de genotipos cercanos genéticamente (cúmulos de mutantes), variantes posteriores como Ómicron se distribuyen en el espacio antigénico de forma más homogénea. Estos hallazgos empíricos incitan al análisis sobre el efecto de la topología del espacio antigénico en la evolución viral y en las trayectorias epidémicas.

En primer lugar, empleamos dos redes sintéticas diferentes que capturan los comportamientos observados en algunas regiones de las redes de genotipos reales. Para emular los cúmulos de genotipos, construimos la red como una secuencia de grafos estrella, en los que un nodo central se conecta con el resto, denominados hojas. Estos núcleos se unen mediante nodos intermedios que conectan las hojas de cada par de estrellas consecutivas (esquema azul de la Fig. 4.1.a). Esta estructura genera trayectorias epidémicas caracterizadas por brotes periódicos que llevan al sistema a un patrón estacionario de ciclos estacionales (línea azul de la Fig. 4.1.b). Por contra, si representamos el espacio antigénico de manera homogénea, a través de una red regular (esquema rosa de la Fig. 4.1.a), la trayectoria epidémica se caracteriza por alcanzar un equilibrio endémico (línea rosa de la Fig. 4.1.b). A nivel microscópico, este estado se corresponde con una onda viajera en el

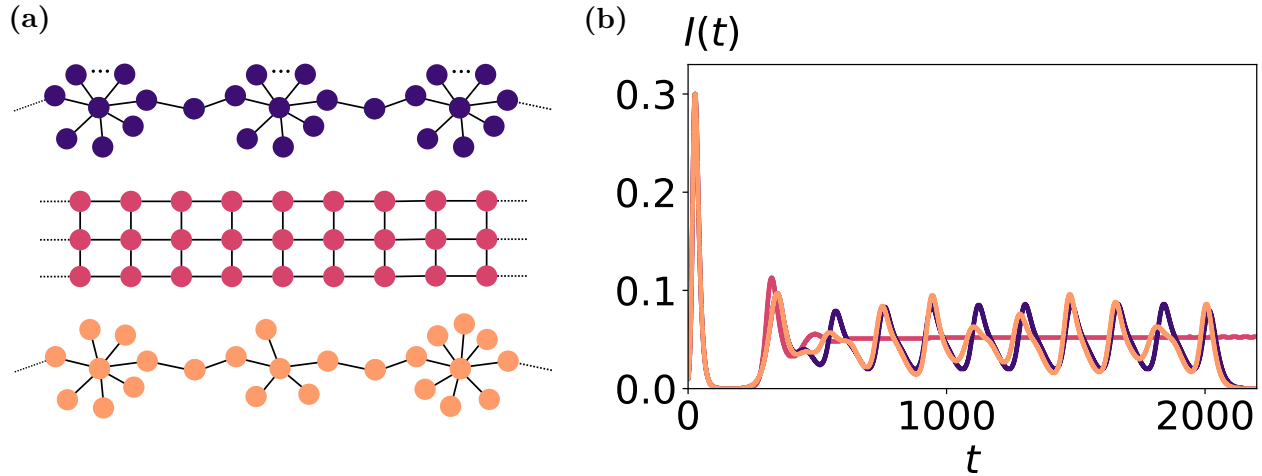


Fig. 4.1: **Endemicidad y estacionalidad.** (a) Esquemas de las redes sintéticas empleadas: estrellas consecutivas (azul), red regular (rosa) y estrellas heterogéneas (naranja). (b) Trayectorias epidémicas: la línea azul se corresponde con la red de estrellas homogéneas, construida con 10 estrellas de 40 hojas cada una, unidas mediante un nodo conector; la línea rosa indica la trayectoria epidémica en la red periódica, de  $3 \times 100$  nodos dispuestos en cuadrícula; y la línea naranja muestra la evolución correspondiente a la red de estrellas heterogéneas. En todos los casos,  $\beta = 0.3$ ,  $\mu_0 = 0.1$ ,  $\alpha = 0.036$ ,  $\gamma = 0$ ,  $\zeta = 1$ ,  $D_x = 10^{-5}$  y  $\Delta = 3$ .

espacio antigénico, ya que varía la cepa presente en la población. Notablemente, estas diferencias en la trayectoria epidémica residen únicamente en la estructura de la red de genotipos empleada para modelizar el espacio antigénico.

Considerando de nuevo la primera red, podemos introducir heterogeneidad en la estructura al variar el número de hojas en cada estrella (esquema naranja de la Fig. 4.1.a). Al hacer esto, se obtiene que los ciclos epidémicos varían en su forma y en la magnitud de la prevalencia. Además, observamos que el tamaño del cúmulo de cepas es proporcional a la magnitud del brote correspondiente. Esto induce a pensar que la aparición de estos cúmulos de mutantes en torno a variantes exitosas del espacio antigénico sirve al virus como mecanismo para contrarrestar la presión de la inmunidad cruzada, ya que la aparición simultánea de genotipos similares da lugar a brotes conjuntos, que por tanto tienen mayor prevalencia global.

#### 4.1.2. Inmunidad cruzada y cúmulos de mutantes

Veamos ahora en mayor detalle el equilibrio dinámico que existe entre la inmunidad cruzada y la existencia de cúmulos de mutantes. Para ello, en esta parte consideramos una red de genotipos sintética sencilla, con dos estrellas conectadas, la cual representamos en la Fig. 4.2.a. La adquisición de inmunidad cruzada es un mecanismo complementario a la inmunidad natural que contribuye a la neutralización de la carga viral. La importancia de esta inmunidad cruzada depende de la relación entre su longitud característica y el tamaño y forma del espacio antigénico.

La simulación llevada a cabo sobre la red de doble estrella da lugar a la trayectoria epidémica mostrada en la Fig. 4.2.b, donde cada color se corresponde con el cúmulo marcado de igual manera

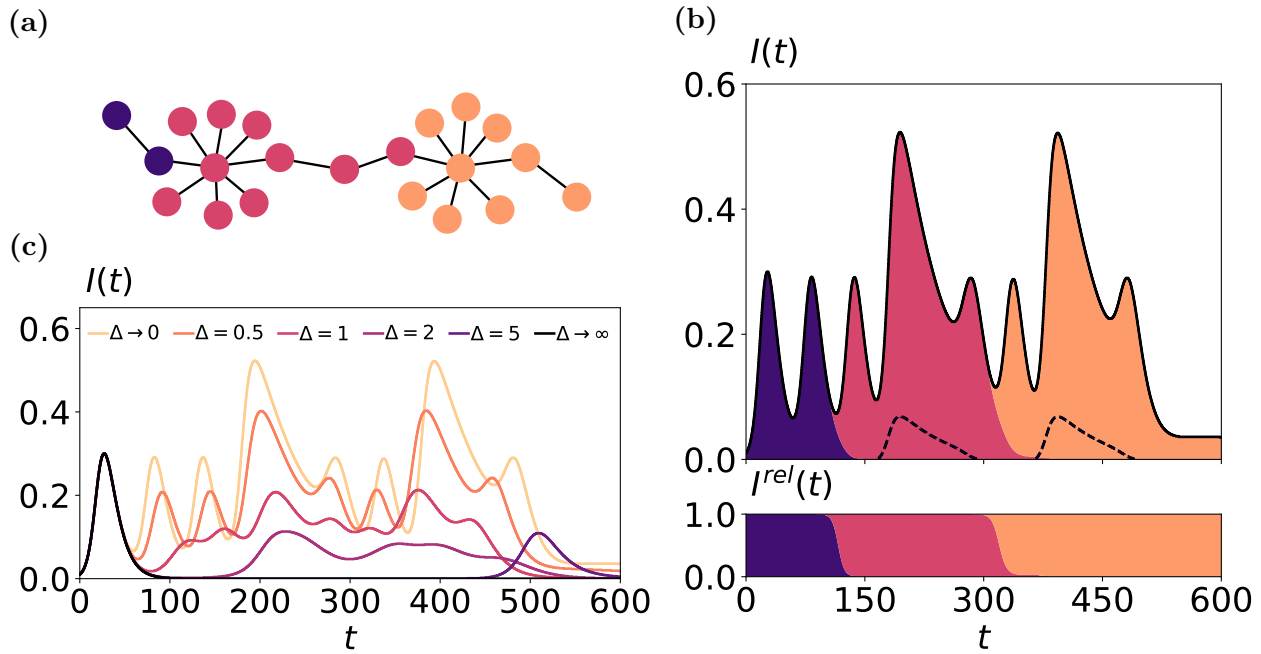


Fig. 4.2: **Relación entre la inmunidad cruzada y los cúmulos de mutantes en el espacio antigénico.** El panel (a) presenta un esquema de la red de genotipos empleada y el panel (b) muestra la trayectoria epidémica resultante. Los colores representan las regiones de la red y las líneas discontinuas marcan la evolución de una de las hojas en cada cúmulo. El gráfico inferior del panel (b) muestra la prevalencia relativa para las tres zonas de la red. Finalmente, el panel (c) recoge la evolución de la fracción de infectados considerando la misma red de genotipos, en función de la longitud característica  $\Delta$  de la inmunidad cruzada. En ambos casos se han utilizado  $\beta = 0.3$ ,  $\mu_0 = 0.1$ ,  $\alpha = 0.036$ ,  $\gamma = 0$ ,  $\zeta = 1$  y  $D_x = 10^{-5}$ , además de  $\Delta = 0$  para el panel (b).

en el esquema de la red. En esta trayectoria epidémica pueden distinguirse fácilmente los brotes debidos a cada cepa aislada, mientras que el aumento en la prevalencia de los picos más altos y anchos se corresponde con la incidencia conjunta de las hojas de cada estrella. Esto concuerda con el comportamiento de la Fig. 3.3.a, que muestra que los genotipos iguales desde el punto de vista estructural generan brotes epidémicos de magnitud y forma similares. Además, las líneas discontinuas representan la fracción de infectados de una hoja de cada estrella. Esto muestra no solo cómo la mutación simultánea a todas las hojas hace que se dispare la fracción total de contagiados, sino también la disminución de la presión de cada cepa individual sobre la población, reflejada en el menor número de infectados de las líneas discontinuas. Esto se debe a que el patógeno central muta a todas las cepas adyacentes, haciendo que la carga se reparta entre los nuevos genotipos y que cada uno por separado infecte a una menor fracción de la población. Además, el panel inferior de la Fig. 4.2.b muestra la prevalencia relativa de cada cúmulo para facilitar la visualización de la evolución en el espacio antigénico.

Por otro lado, la Fig. 3.3.b muestra que el incremento de la longitud característica de inmunidad cruzada hace que se reduzca la prevalencia de los genotipos en la población. Sin embargo, del mismo modo que la mutación es una herramienta que ayuda al virus a evadir la respuesta inmune del huésped, la presencia de cúmulos de mutantes en el espacio antigénico puede hacer que el virus

alcance niveles comparables al del primer brote de la epidemia, como se aprecia en la Fig. 4.2.c. Esto pone de manifiesto que las agrupaciones de genotipos en torno a cepas exitosas son un mecanismo del virus que contrarresta la inmunidad cruzada, al conseguir una contribución conjunta de todas las cepas significativa. Es precisamente el compromiso entre la magnitud de los cúmulos y la longitud característica de inmunidad cruzada lo que determina la fracción de población infectada a lo largo de la trayectoria epidémica.

#### 4.2. Trayectorias epidémicas en redes de genotipos reales

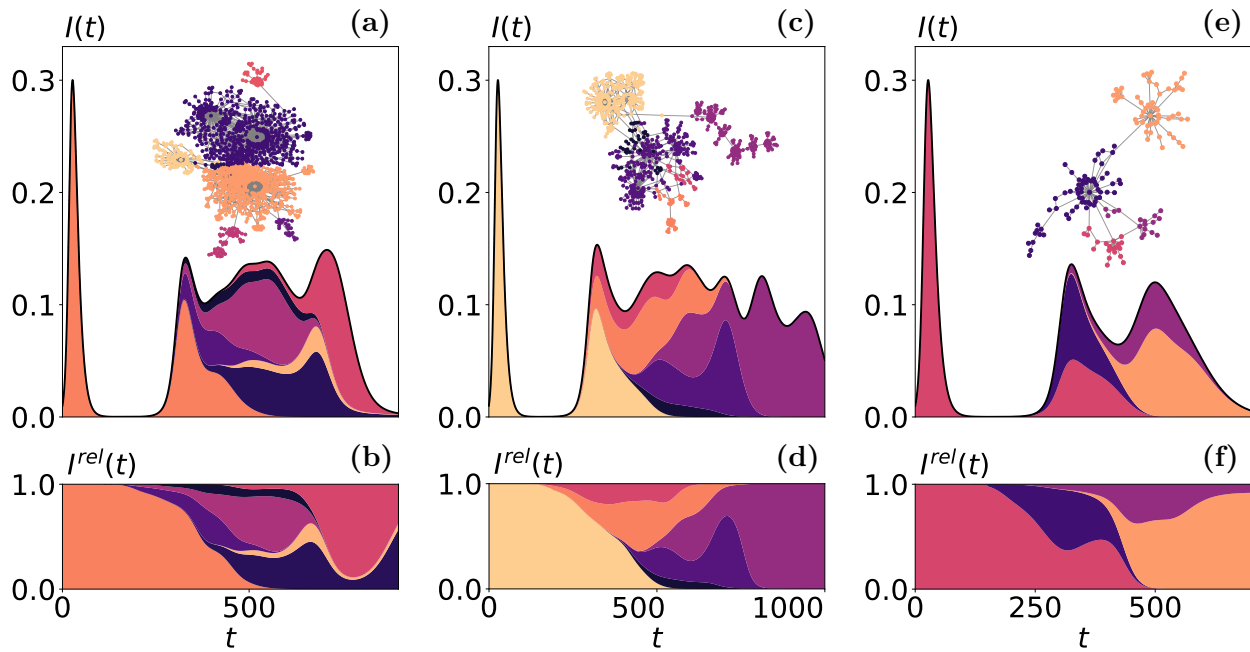


Fig. 4.3: **Trayectorias epidémicas en redes de genotipos reales.** Los paneles (a), (c), (e) recogen la trayectoria epidémica correspondiente a la primera, segunda y séptima componente conexa más grande de la red de Influenza A, respectivamente. Los colores indican la incidencia relativa de cada comunidad, indicadas sobre la red en el esquema interior. Los paneles (b), (d), (f) muestran la prevalencia relativa de cada cúmulo respecto al total de infectados. Se han utilizado los parámetros  $\beta = 0.3$ ,  $\mu_0 = 0.1$ ,  $\alpha = 0.036$ ,  $\gamma = 0$ ,  $\zeta = 1$ ,  $D_x = 10^{-5}$  y  $\Delta = 3$ .

Una vez comprobado el rol de los cúmulos de genotipos y la capacidad de la estructura de la red para dictar la evolución epidémica, pasamos a considerar una red de genotipos real. En concreto, utilizamos la red construida por Williams et al. [6] del virus Influenza A, descrita en la Subsec. 2.3.2. Para ser capaces de monitorizar la evolución del virus en el espacio antigénico, utilizamos la detección de comunidades en la red mediante *Infomap*, identificando los cúmulos de cepas dentro de cada componente conexa. El esquema interno en la Fig. 4.3.a muestra la componente conexa más grande de la red de Influenza A, donde los distintos colores de los nodos representan los distintos cúmulos de mutantes. La gráfica externa muestra la trayectoria epidémica  $I(t)$  para esta componente de la red, considerando que la primera cepa registrada se corresponde con el primer genotipo de la componente detectado empíricamente. Las regiones coloreadas bajo la curva de prevalencia indican

la fracción de la población infectada por cada cúmulo de mutantes del virus. Observamos que, en un inicio, los contagios se deben al cúmulo representado en naranja, al que pertenece la cepa inicial. Sin embargo, tras el primer brote, el escape antigénico modifica la dinámica hacia los demás cúmulos conforme se adquiere inmunidad. La Fig. 4.3.b indica la prevalencia relativa, es decir, la fracción de infecciones debidas a cada cúmulo de cepas. En esta representación se observa mejor la transición de prevalencia entre la comunidad naranja y las azul oscuro y magenta.

El esquema interno de la Fig. 4.3.c presenta la segunda componente más grande de la red de Influenza A, con seis comunidades principales. Dado que la topología de esta componente es más compleja, la trayectoria epidémica correspondiente, mostrada en el mismo panel, también lo es. Presenta varios cúmulos que evolucionan casi al mismo tiempo en la región más densa de la red de genotipos. Sin embargo, se observa que cerca del final de la trayectoria domina la comunidad morada, lo cual se debe a su mayor distancia genética con el resto de regiones. El paisaje de cúmulos mutantes mostrado por  $I^{rel}(t)$  para esta componente es más elaborado (véase la Fig. 4.3.d), con diversas comunidades que aparecen y desaparecen con el tiempo.

Por último, la Fig. 4.3.e muestra la trayectoria epidémica y el esquema de la séptima componente conexa más grande de la red, que exhibe una estructura más jerárquica que las otras, similar a la red sintética de estrellas consecutivas vista antes. En consecuencia, resulta más sencillo identificar los cúmulos mutantes correspondientes a cada pico de la trayectoria epidémica. De nuevo, la Fig. 4.3.f recoge la prevalencia relativa de los cúmulos.

En todos los casos mostrados en la Fig. 4.3, las trayectorias epidémicas son notablemente más complejas que las derivadas de representaciones de menor dimensión, como la cadena lineal mostrada en la Fig. 3.3. Mesoscópicamente, nuestros resultados muestran cómo la deriva antigénica a través de redes de genotipos produce complejos paisajes dinámicos de variantes. Se observa la emergencia y desaparición de linajes debido a las propiedades antigénicas o la interferencia por dominar la población de las diferentes comunidades, lo cual recuerda a los comportamientos observados en brotes reales. Esta correspondencia surge de la interacción entre la historia de las infecciones en la población y la trayectoria del virus a través de la red antigénica.

## 5. Conclusiones

En este trabajo hemos utilizado el modelo SIMS (Susceptible-Infectado-Mutación-Susceptible) y las redes de genotipos para estudiar cómo la evolución viral influye en el desarrollo de las epidemias. Este modelo, a través de su estructura compartimental, considera los procesos de contagio, inmunidad y mutación del patógeno, además de tener en cuenta la memoria inmune y la posible adquisición de inmunidad cruzada para genotipos similares. Estos procesos se reflejan directamente en las ecuaciones dinámicas del modelo, que se integran a lo largo del trabajo para obtener los diferentes resultados.

En un primer análisis, donde el patógeno presenta solamente una cepa, se ha visto que el modelo es capaz de reproducir los resultados característicos de modelos más sencillos. En particular, las trayectorias epidémicas pueden replicar la endemidad del modelo SIS, el patrón de ola epidémica (brote seguido de extinción) del modelo SIR, y el brote seguido de estado endémico del modelo SIRS. Además, hemos derivado analíticamente cómo la relación entre la adquisición de inmunidad, su pérdida con el tiempo y la mutación del patógeno desplazan el umbral epidémico y modifican la prevalencia en el estado estacionario de la epidemia.

En el caso más complejo de considerar múltiples cepas, hemos observado cómo el proceso de mutación permite al patógeno evadir la respuesta inmune del huésped mediante el escape antigénico [9, 12]. Además, en este contexto hemos analizado la importancia del desarrollo de la inmunidad cruzada en la población. Cuando este tipo de inmunidad no existe, la trayectoria epidémica da lugar a olas en el espacio antigénico de prevalencia constante, permitiendo que el virus mantenga su incidencia global en la población a lo largo del tiempo. Sin embargo, al introducir inmunidad cruzada, la epidemia prácticamente desaparece tras el primer brote, hasta que una cepa suficientemente distinta genéticamente inicia una nueva ola en el espacio antigénico, pero esta vez con una incidencia mucho menor. Estas diferencias, unidas a las ecuaciones analíticas presentadas en el texto, indican que la inmunidad cruzada potencia la respuesta inmune, llevando al sistema a un estado de menor prevalencia.

Para el caso general de múltiples cepas, hemos estudiado cambios en la topología de la red de genotipos y observado su efecto en el desarrollo de una epidemia. Hemos comprobado que la estructura de la red por sí misma es capaz de capturar comportamientos complejos, dando lugar a trayectorias epidémicas diferentes. Concretamente, hemos visto que las redes homogéneas capturan la endemidad estable propia de algunas enfermedades, mientras que las redes con cúmulos de variantes, habituales en el paisaje antigénico de virus como la influenza, dan lugar a olas estacionales epidémicas. Además, hemos comprobado que los cúmulos de mutantes en el espacio antigénico pueden actuar como un mecanismo del virus para contrarrestar el efecto de la inmunidad cruzada. Estas diferencias resaltan la importancia de la exploración del espacio antigénico en el transcurso de una enfermedad.

Este análisis mediante redes complejas puede también aplicarse para redes del mundo real, dado que el avance del patógeno a través de la red de genotipos se asemeja a la secuencia temporal

de muestreo de las cepas. El análisis de la red de genotipos real del virus Influenza A (H3N2) confirma que la estructura de la red afecta de manera importante a la dinámica de una epidemia. Las características de estas redes, como la cola en la distribución de grado y una fuerte disasortatividad, dan lugar a comportamientos heterogéneos y reflejan la diversidad en la evolución viral. Estudiando el sistema a escalas mesoscópicas, se observa que los cúmulos compiten por el control de la población susceptible, facilitando la proliferación del patógeno. Además, la trayectoria evolutiva del virus a través de estas redes reales se asemeja a los patrones complejos observados para los virus en el mundo real, aunque evidentemente resulta complicado establecer una conexión completa entre estos modelos y los datos empíricos.

Como perspectiva de futuro, un siguiente paso natural sería refinar el modelo SIMS para incluir una red de interacciones entre individuos, e incluso incorporar patrones de movilidad. Esto permitiría analizar cómo varían las trayectorias epidémicas al considerar diferentes estructuras sociales y restricciones no farmacológicas. También podrían incluirse características más complejas en la dinámica, como una infectividad variable para las cepas [12] o la consideración de individuos inmunodeprimidos, con el objeto de determinar su rol en el desarrollo epidémico.

En conclusión, los resultados de este trabajo ilustran el papel tan destacado que tiene la estructura de la red de genotipos sobre las trayectorias epidémicas virales. El modelo empleado permite incorporar la forma en que los patógenos exploran el espacio antigénico y determinar el efecto que esto tiene sobre el desarrollo de una epidemia. Esto podría permitir, en última instancia, mejorar el diseño de estrategias de control sobre la población para mitigar las consecuencias de una enfermedad, ya que generalmente no se tiene en cuenta la coevolución viral.

## Referencias

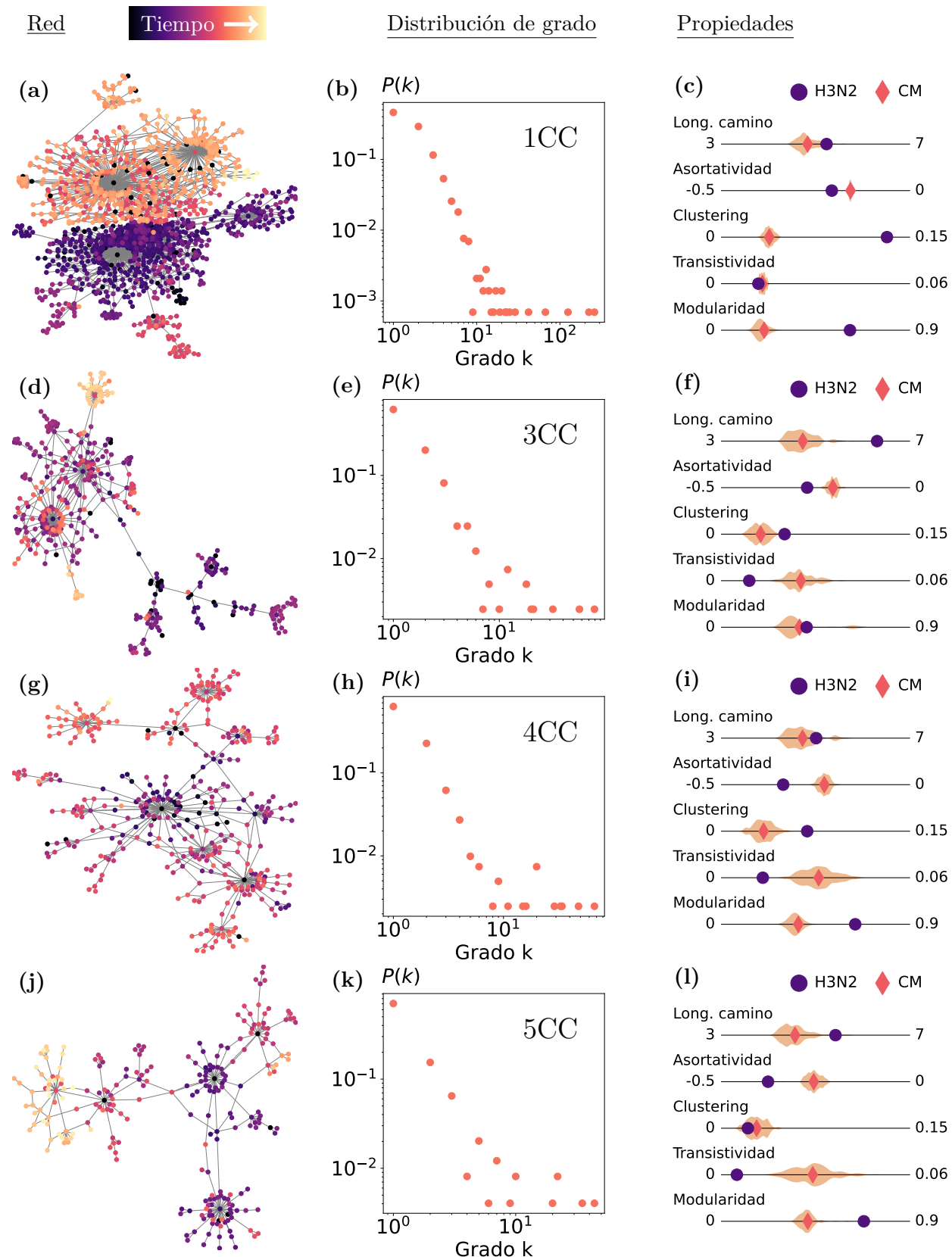
- [1] S. Boccaletti et al. «Complex networks: Structure and dynamics». En: *Physics Reports* 424.4 (2006), págs. 175-308. DOI: 10.1016/j.physrep.2005.10.009.
- [2] S. Lamata-Otín et al. *Genotype networks drive oscillating endemicity and epidemic trajectories in viral evolution*. 2025.
- [3] F. Brauer. «Compartmental Models in Epidemiology». En: *Mathematical Epidemiology*. Ed. por F. Brauer, P. van den Driessche y J. Wu. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, págs. 19-79. DOI: 10.1007/978-3-540-78911-6\_2.
- [4] R. Sanjuán y P. Domingo-Calap. «Mechanisms of viral mutation». En: *Cellular and Molecular Life Sciences* 73.23 (2016), págs. 4433-4448. DOI: 10.1007/s00018-016-2299-6.
- [5] A. H. Daniel Edler y M. Rosvall. *The MapEquation software package*. <https://mapequation.org>. Accessed: 07-07-2025. 2025.
- [6] B. J. M. Williams et al. «Immunity-induced criticality of the genotype network of influenza A (H3N2) hemagglutinin». En: *PNAS Nexus* 1.4 (ago. de 2022), pgac143. DOI: 10.1093/pnasnexus/pgac143.
- [7] Saranya Sridhar et al. «Longevity and determinants of protective humoral immunity after pandemic influenza infection». En: *American Journal of Respiratory and Critical Care Medicine* 191.3 (feb. de 2015), págs. 325-332. DOI: 10.1164/rccm.201410-17980C.
- [8] E. Nobusawa y K. Sato. «Comparison of the Mutation Rates of Human Influenza A and B Viruses». En: *Journal of Virology* 80.7 (2006), págs. 3675-3678. DOI: 10.1128/jvi.80.7.3675-3678.2006.
- [9] G. Rouzine Igor M. AND Rozhnova. «Antigenic evolution of viruses in host populations». En: *PLOS Pathogens* 14.9 (sep. de 2018), págs. 1-16. DOI: 10.1371/journal.ppat.1007291.
- [10] F. G. Cobelens et al. «Risk of infection with Mycobacterium tuberculosis in travellers to areas of high tuberculosis endemicity». En: *The Lancet* 356.9228 (2000), págs. 461-465. DOI: 10.1016/S0140-6736(00)02554-X.
- [11] S. I. Hay, D. L. Smith y R. W. Snow. «Measuring malaria endemicity from intense to interrupted transmission». En: *The Lancet Infectious Diseases* 8.6 (2008), págs. 369-378. DOI: 10.1016/S1473-3099(08)70069-0.
- [12] D. Soriano-Paños. «Eco-evolutionary constraints for the endemicity of rapidly evolving viruses». En: (2024).

# Apéndices

## A. Glosario de símbolos y parámetros

- $\mathcal{G}$ : Grafo.
- $\mathcal{N}$ : Conjunto de nodos.
- $\mathcal{L}$ : Conjunto de enlaces.
- $N$ : Cardinal del conjunto de nodos.
- $L$ : Cardinal del conjunto de enlaces.
- $\mathbf{A} = \{a_{ij}\}$ : Matriz de adyacencia de un grafo.
- $\mathbf{L} = \{\ell_{ij}\}$ : Matriz laplaciana normalizada de un grafo.
- $\{\lambda_j\}$ : Conjunto de autovalores de la matriz laplaciana.
- $\mathbf{X} = \{x_{ij}\}$ : Matriz de distancia de caminos más cortos entre nodos.
- $k_i$ : Grado del nodo  $i$  de un grafo.
- $P(k)$ : Distribución de grado de los nodos de un grafo.
- $L(\mathcal{G})$ : Longitud media del camino más corto.
- $r(\mathcal{G})$ : Coeficiente de asortatividad.
- $C_i$ : Coeficiente de *clustering* del nodo  $i$
- $C = \sum C_i/N$ : *Clustering* promedio de la red.
- $T(\mathcal{G})$ : Transistividad de la red.
- $Q(\mathcal{G})$ : Modularidad de la red.
- $s_i$ : genotipo o cepa  $i$  del patógeno.
- $S$ : compartimento de susceptibles.
- $I_i$ : compartimento de infectados por la cepa  $s_i$ .
- $\rho^S$ : fracción de individuos susceptibles.
- $\rho_i^I(t) \equiv \rho_i(t)$ : fracción de individuos infectados por la cepa  $s_i$ .
- $I(t) = \sum \rho_i(t)$ : fracción total de individuos infectados.
- $I^* = \lim_{t \rightarrow \infty} I(t)$ : nivel endémico estacionario, fracción de infectados en el equilibrio.
- $\mu_i(t)$ : tasa de recuperación de la cepa  $s_i$ .
- $R_0$ : número reproductivo básico.
- $R_0^c$ : número reproductivo básico crítico.
- $\beta$ : tasa de infectividad del patógeno.
- $\beta_c$ : umbral epidémico.
- $\mu_0$ : tasa de recuperación basal.
- $\alpha$ : tasa de adquisición de inmunidad.
- $\zeta$ : sensibilidad del sistema inmune.
- $\gamma$ : ritmo de pérdida de inmunidad.
- $\Delta$ : longitud característica de inmunidad cruzada.
- $D_x$ : tasa de mutación entre las cepas del patógeno.

## B. Propiedades estructurales de las redes de genotipos reales



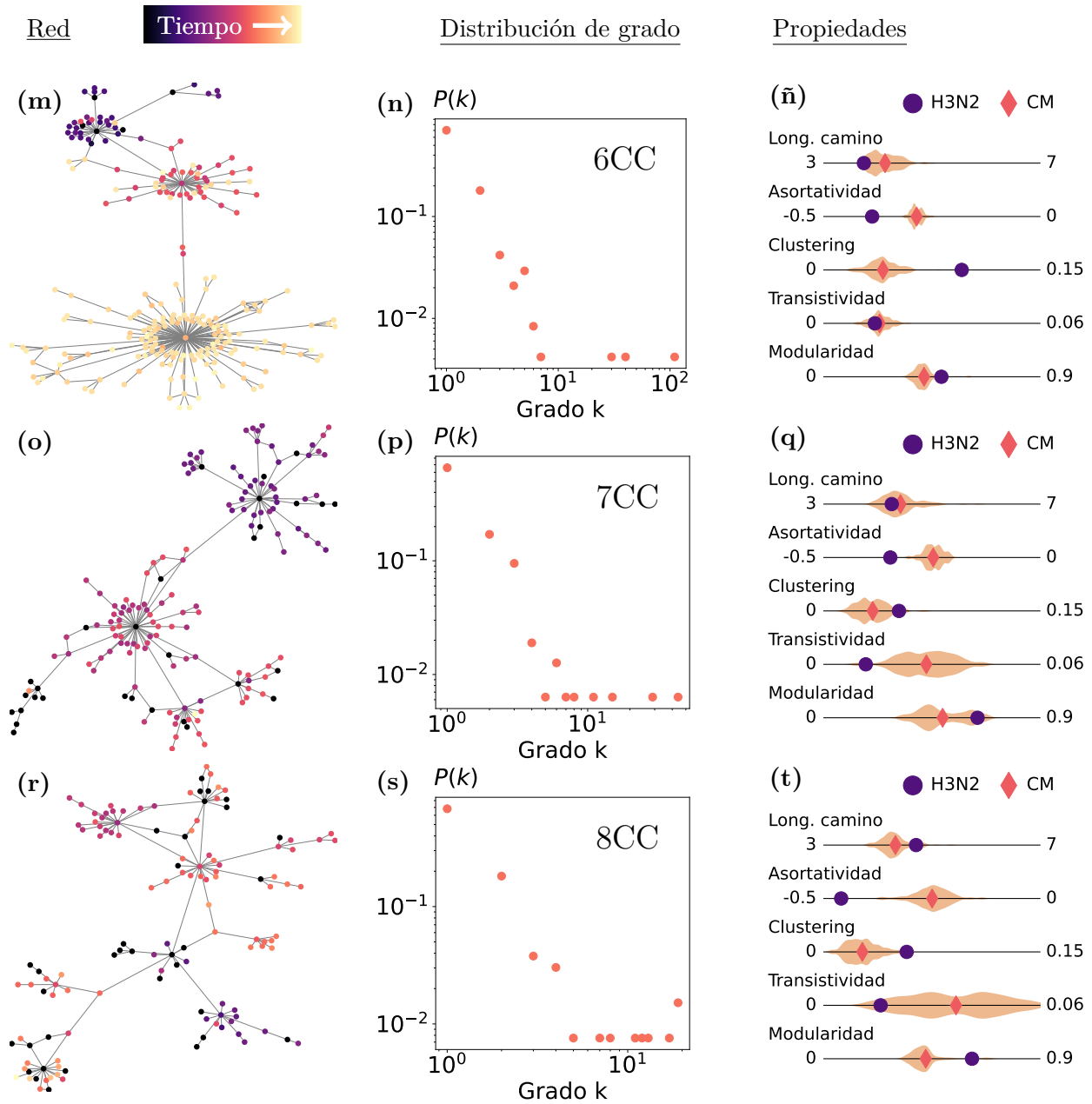


Fig. B.1: **Propiedades estructurales de las redes de genotipos.** Extensión del análisis llevado a cabo en la Fig. 2.5 para la segunda componente conexa más grande de la red de Influenza A. Se muestra la misma información para las otras siete componentes de mayor tamaño, correspondientes con cada fila de la figura. La primera columna muestra cada red, donde el color de los nodos denota la antigüedad de muestreo de la cepa. La columna central se corresponde con la distribución de grado de cada componente de la red, y la columna derecha muestra los valores de algunas propiedades estructurales de cada componente, comparándolos con los obtenidos para un conjunto de cien redes sintéticas con la misma secuencia de grado. Las redes reales analizadas provienen de Williams et al. [6].

## C. Infectividad máxima en el modelo SIR

Supuesto un patógeno con número reproductivo básico  $R_0$ , la incidencia máxima en el brote epidémico del modelo SIR viene dada por

$$I_{max} = 1 - \frac{1}{R_0} [1 + \ln(R_0)]. \quad (\text{C.1})$$

En el caso de  $R_0 = 3$  mencionado en la Subsec. 3.3 del texto, se obtiene  $I_{max} = 0.3005$ . Comprobamos entonces que coincide con el valor máximo del pico de la curva naranja tipo SIR de la Fig. 3.2.a.

## D. Desarrollo analítico: modelo SIMS con múltiples cepas

### i) Expresión del umbral epidémico

En estado estacionario,  $\dot{\mu}_i = \dot{\rho}_i = 0$ , y a partir de la Ec. (3.1) se obtiene la relación  $\mu_i^* = \mu_i^*(\rho_i^*)$ ,

$$\mu_i^* = \beta \left( 1 - \sum_{j=1}^N \rho_j^* \right) - \frac{D_x}{\rho_i^*} \sum_{j=1}^N \ell_{ij} \rho_j^*. \quad (\text{D.1})$$

Introduciendo esto en la Ec. (3.2), se obtiene que

$$\rho_i^* = \frac{-D_x \sum_{j=1}^N \ell_{ij} \rho_j^*}{\mu_0 - \beta \left( 1 - \sum_{j=1}^N \rho_j^* \right) + \frac{\alpha}{\gamma} \sum_{j=1}^N \left( \frac{\rho_j^*}{\zeta^{-1} + \rho_j^*} e^{-x_{ij}/\Delta} \right)}. \quad (\text{D.2})$$

Situándonos cerca del umbral epidémico, la enfermedad se vuelve endémica con una incidencia pequeña,  $\rho_i^* = \varepsilon_i^* \ll 1 \forall i$ . Teniendo esto en cuenta, aproximamos a primer orden la siguiente fracción,

$$\frac{\rho_j^*}{\zeta^{-1} + \rho_j^*} = \zeta \rho_j^* \frac{1}{1 + \zeta \rho_j^*} \approx \zeta \rho_j^* (1 - \zeta \rho_j^*) \approx \zeta \rho_j^*. \quad (\text{D.3})$$

Entonces el denominador de la Ec. (D.2) puede expresarse como

$$\mu_0 - \beta \left( 1 - \sum_j \varepsilon_j^* \right) + \frac{\alpha}{\gamma} \sum_j \left( \zeta \varepsilon_j^* e^{-x_{ij}/\Delta} \right) = (\mu_0 - \beta) + \sum_j \varepsilon_j^* \left( \beta + \frac{\alpha \zeta}{\gamma} e^{-x_{ij}/\Delta} \right). \quad (\text{D.4})$$

Como  $\varepsilon_j^* \rightarrow 0$ , el término dominante es  $\mu_0 - \beta$ . Volviendo entonces a la Ec. (D.2), tenemos que

$$\varepsilon_i^* \approx \frac{-D_x \sum_j \ell_{ij} \varepsilon_j^*}{\mu_0 - \beta} \quad (\text{D.5})$$

De esta forma, obtenemos la siguiente solución para la Ec. (D.2) linealizada,

$$\frac{\mu_0 - \beta}{D_x} \varepsilon_i^* = - \sum_{j=1}^N l_{ij} \varepsilon_j^*. \quad (\text{D.6})$$

El umbral epidémico  $\beta_c$  es el mínimo valor de  $\beta$  para el que se verifican estas ecuaciones, comenzando el estado endémico. La Ec. (D.6) define un problema de autovalores, donde el umbral epidémico está relacionado con el autovalor mínimo  $\Lambda_{min}$  de la matriz laplaciana normalizada por el grado de salida, de manera que

$$\beta_c = \mu_0 + D_x \Lambda_{min}(\mathbf{L}). \quad (\text{D.7})$$

Por definición, los autovalores de la matriz laplaciana son semipositivos, con al menos uno de ellos igual a cero. Entonces  $\Lambda_{min}(\mathbf{L}) = 0$  y se obtiene el mismo umbral epidémico del modelo SIS:

$$\beta_c = \mu_0 \iff R_0^c = 1. \quad (\text{D.8})$$

## ii) Expresión del umbral epidémico en el caso $\zeta \rightarrow \infty$

Si la sensibilidad del sistema inmune es máxima, el término  $\frac{\alpha}{\gamma} \sum_{j=1}^N \left( \frac{\rho_j}{\zeta^{-1} + \rho_j} e^{-x_{ij}/\Delta} \right)$ , responsable de la adquisición de inmunidad, se vuelve independiente del estado de la población,

$$\frac{\alpha}{\gamma} \sum_{j=1}^N \left( \frac{\rho_j}{\zeta^{-1} + \rho_j} e^{-x_{ij}/\Delta} \right) \rightarrow \frac{\alpha}{\gamma} \sum_{j=1}^N e^{-x_{ij}/\Delta}. \quad (\text{D.9})$$

Bajo esta suposición, y considerando de nuevo que  $\rho_i^* = \varepsilon_j^* \ll 1$ , linealizamos la Ec. (D.2):

$$\begin{aligned} \varepsilon_i^* &= \frac{-D_x \sum_j l_{ij} \varepsilon_j^*}{\mu_0 - \beta + \frac{\alpha}{\gamma} \sum_j e^{-x_{ij}/\Delta}} \left( 1 + \frac{\beta}{\mu_0 - \beta + \frac{\alpha}{\gamma} \sum_j e^{-x_{ij}/\Delta}} \sum_j \varepsilon_j^* \right)^{-1} \\ &\approx \frac{-D_x \sum_j l_{ij} \varepsilon_j^*}{\mu_0 - \beta + \frac{\alpha}{\gamma} \sum_j e^{-x_{ij}/\Delta}} \left( 1 - \frac{\beta}{\mu_0 - \beta + \frac{\alpha}{\gamma} \sum_j e^{-x_{ij}/\Delta}} \sum_j \varepsilon_j^* \right) \approx \frac{-D_x \sum_j l_{ij} \varepsilon_j^*}{\mu_0 - \beta + \frac{\alpha}{\gamma} \sum_j e^{-x_{ij}/\Delta}}. \end{aligned}$$

Obtenemos así la solución

$$\frac{\mu_0 - \beta + \frac{\alpha}{\gamma} \sum_j e^{-x_{ij}/\Delta}}{D_x} \varepsilon_i^* = - \sum_j l_{ij} \varepsilon_j^*, \quad (\text{D.10})$$

que es formalmente similar a la Ec. (D.6). Para dar una solución analítica, suponemos que  $x_{ij} = 0$  si  $i = j$ , mientras que  $x_{ij} = \langle x \rangle$  en caso contrario. Se define entonces un problema de autovalores,

$$\frac{\mu_0 - \beta + \frac{\alpha}{\gamma} \left[ 1 + (N-1) e^{-\langle x \rangle / \Delta} \right]}{D_x} \varepsilon_i^* = - \sum_j l_{ij} \varepsilon_j^*, \quad (\text{D.11})$$

donde el umbral epidémico ahora también depende de los parámetros que determinan la dinámica inmune del sistema. Usando que  $\Lambda_{min}(\mathbf{L}) = 0$ , igual que en el caso anterior, obtenemos que el umbral epidémico aumenta, y viene dado por

$$\tilde{\beta}_c = \mu_0 + \frac{\alpha}{\gamma} \left[ 1 + (N-1) e^{-\langle x \rangle / \Delta} \right] \iff \tilde{R}_0^c = 1 + \frac{\alpha}{\mu_0 \gamma} \left[ 1 + (N-1) e^{-\langle x \rangle / \Delta} \right]. \quad (\text{D.12})$$

### iii) Prevalencia de la fase endémica en el caso $\zeta \rightarrow \infty$

Consideramos una aproximación de campo medio,  $\rho_i(t) = \rho(t)$  y  $\mu_i(t) = \mu(t)$ , y suponemos que  $x_{ij} = 0$  si  $i = j$ , y  $x_{ij} = \langle x \rangle$  en caso contrario. En el caso de  $\zeta \rightarrow \infty$ , podemos escribir las Ecs. 3.1 y 3.2 de la forma

$$\dot{\rho}(t) = \beta \rho(t) [1 - N \rho(t)] - \mu(t) \rho(t), \quad (\text{D.13})$$

$$\dot{\mu}(t) = \alpha \left[ 1 + (N-1) e^{-\langle x \rangle / \Delta} \right] - \gamma (\mu(t) - \mu_0). \quad (\text{D.14})$$

En el estado estacionario,  $\dot{\mu} = \dot{\rho} = 0$ , y la Ec. (D.13) presenta una solución trivial absorbente,  $\rho_0^* = 0$ . Tomando la solución  $\rho^* \neq 0$  se obtiene la relación

$$\mu^* = \beta (1 - N \rho^*). \quad (\text{D.15})$$

Introduciendo esto en la Ec. (D.14) y recordando que  $I^* = \sum_i \rho_i^* = N \rho^*$ , se obtiene la prevalencia en fase endémica,

$$I^* = 1 - R_0^{-1} - \frac{\alpha}{\gamma \beta} \left[ 1 + (N-1) e^{-\langle x \rangle / \Delta} \right]. \quad (\text{D.16})$$

## E. Métodos computacionales

**Simulaciones numéricas:** las trayectorias epidémicas mostradas a lo largo del trabajo (Figs. 3.2.a, 3.3, 3.4, 4.1, 4.2 y 4.3) se obtienen mediante la integración de las ecuaciones del modelo. Esta integración se lleva a cabo mediante un algoritmo de Runge-Kutta de orden 4, con paso temporal  $h = 0.01$ .

**Modelo configuracional:** para la Fig. 2.5 y el Ap. B se utiliza un conjunto de cien redes sintéticas, cuyas propiedades se comparan con la red real de Influenza A. Este conjunto de redes se construye utilizando la librería *networkx* de Python, concretamente la función *configuration\_model*. Para cada componente conexas de la red real, se crean cien redes con la misma secuencia de grado. Es decir, cada nodo tiene un número fijo de conexiones, pero aleatorizadas. Para determinar los valores de las métricas de las redes reales y sintéticas, se utiliza también la librería *networkx*.

**Detección de comunidades:** el algoritmo de detección de comunidades se emplea en la Fig. 4.3 sobre la red real del virus Influenza A. Para determinar las comunidades se ha utilizado el algoritmo de agrupación de redes *Infomap*, como también se explica en la Subsec. 2.3.1.