



**Universidad**  
Zaragoza

TRABAJO DE FIN DE GRADO

---

**MUTACIÓN DE PATÓGENOS COMO  
PROCESO CO-EVOLUTIVO EN  
PROPAGACIÓN DE EPIDEMIAS**

---

*Autor:*

Octavian Constantin Rotita Ion

*Directores:*

Dr. Jesús Gómez Gardeñes  
Santiago Lamata Otín

UNIVERSIDAD DE ZARAGOZA

FACULTAD DE CIENCIAS

Junio 2024

# 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 y umbral epidémico . . . . .	3
2.2. Mutación de patógenos . . . . .	3
2.3. Redes complejas . . . . .	4
2.3.1. Redes de genotipos reales . . . . .	6
2.3.2. Redes de genotipos sintéticas . . . . .	7
<b>3. Modelo SIMS para la propagación y mutación de patógenos</b>	<b>9</b>
3.1. Formalismo matemático del modelo SIMS . . . . .	9
3.2. Umbral epidémico en el modelo SIMS . . . . .	10
3.3. Dinámica del modelo SIMS en ausencia de mutaciones . . . . .	14
3.4. Dinámica del modelo SIMS en ausencia de recuperación . . . . .	16
<b>4. Dinámica del modelo SIMS en redes de genotipos</b>	<b>17</b>
4.1. Propagación y mutación de patógenos como procesos co-evolutivos . . . . .	17
4.2. Escape antigénico a través de redes de genotipos . . . . .	19
4.3. Resurgimiento de variantes en redes de genotipos . . . . .	20
4.4. Dinámica del modelo SIMS en redes de genotipos reales . . . . .	23
<b>5. Conclusiones</b>	<b>24</b>
<b>Referencias</b>	<b>24</b>
<b>Apéndices</b>	<b>26</b>
<b>A. Colección de parámetros y siglas</b>	<b>26</b>
<b>B. Complementos de la dinámica del modelo SIMS en redes de genotipos sintéticas</b>	<b>28</b>
<b>C. Complementos de la dinámica del modelo SIMS en redes de genotipos reales</b>	<b>29</b>

## 1. Introducción

Según la tradición clásica de la Física, la motivación principal de esta ciencia es la búsqueda de las partes e interacciones fundamentales de la materia. La Física Estadística surgió con la idea de explorar las propiedades de la materia en su conjunto, proporcionando un marco para relacionar el comportamiento individual de las partículas con las propiedades macroscópicas de los sistemas que las contienen. Con esta misma filosofía aplicada a sistemas más allá del dominio tradicional, la física de Sistemas complejos busca describir la emergencia de propiedades colectivas a partir del análisis de los elementos del sistema y sus interacciones. Cada componente del sistema tiene un comportamiento individual que influye en los demás, dando lugar a la emergencia de fenómenos colectivos como la sincronización de un conjunto de neuronas, o la propagación de epidemias. Asimismo, y a diferencia de las teorías de campo medio habituales en física estadística, la física de sistemas complejos añade un componente esencial en el comportamiento del sistema complejo: la estructura de las interacciones entre los elementos que lo componen. En particular, las redes complejas actúan como el mapa de estas interacciones, catalogando tanto a los elementos del sistema como a sus relaciones.

El estudio de la propagación de epidemias mediante modelos matemáticos y redes complejas ofrece una comprensión detallada de cómo las enfermedades se propagan y evolucionan. Al entender el vínculo entre las relaciones entre individuos y las características de un patógeno, podemos prevenir y actuar en la propagación de estas enfermedades. Cuando la pandemia desatada por el virus SARS-CoV-2 sacudió al mundo, la primera respuesta humana fue la adaptación, tanto a nivel social, mediante medidas como la cuarentena, como a nivel biológico, desarrollando anticuerpos (bien a partir de la vacunación o bien como consecuencia de la exposición al virus) para combatir el patógeno en caso de infección. No obstante, esta adaptación no se limitó solo a los seres humanos; el propio virus también tuvo que adaptarse en respuesta a nuestros intentos de adquirir inmunidad frente a él. Su adaptación se manifestó a través de mutaciones, es decir, cambios en su material genético que dieron lugar a nuevas cepas. Las relaciones entre diferentes cepas de un virus se captan mediante las redes de genotipos, que esencialmente son conjuntos de genotipos conectados entre sí, donde cada conexión representa una mutación puntual del genoma viral.

En este trabajo, introducimos un modelo epidemiológico para explorar la propagación de patógenos que mutan a través de redes de genotipos. Además, presentamos un modelo de generación de redes sintéticas que permita imitar a las redes reales de genotipos observadas en la evolución de virus de ARN. Estudiando la dinámica del modelo apreciamos que se captura la localización de la infección en la red de genotipos, y podemos derivar una expresión analítica que determina las condiciones para detectar el inicio de una epidemia. Posteriormente, analizamos de manera aislada la influencia de la inmunidad en la dinámica de la epidemia, así como el proceso de difusión en las redes. Asimismo, examinamos la dinámica del modelo completo para comprender la correlación entre las evoluciones de las cepas para llegar a estudiar su capacidad de evadir la inmunidad. Finalmente, nos planteamos la pregunta de si es posible que una cepa pueda volver a constituir una amenaza sanitaria una vez ésta ha sido suprimida de la población y la inmunidad adquirida haya disminuido.

## 2. Marco teórico

### 2.1. Modelización de epidemias

El estudio de los procesos dinámicos de una epidemia incluye comprender cómo evolucionan y se propagan los patógenos a través de una población. Este enfoque requiere, entre otros aspectos, tener en cuenta que los patógenos poseen la capacidad de mutar y adaptarse a nuevas condiciones ambientales y a las respuestas inmunitarias de los huéspedes. Por lo tanto, comprender la naturaleza intrínseca de estos patógenos es fundamental para desarrollar estrategias de intervención efectivas que buscan generar inmunidad en la población.

Los primeros modelos de modelización de epidemias se basaban en la hipótesis de campo medio, ignorando las diferencias entre los individuos de una población y las complejas interacciones entre los mismos. Estos modelos iniciales proporcionaron una base teórica importante, pero eran limitados en su capacidad para representar con precisión la propagación de enfermedades en poblaciones heterogéneas. Con el tiempo, los modelos se han vuelto más sofisticados al incorporar la estructura y dinámica de las interacciones en una comunidad.

#### 2.1.1. Modelos compartimentales

Los modelos compartimentales son la base de la modelización de epidemias. Estos modelos dividen la población en compartimentos mutuamente excluyentes, dependiendo del estado de los individuos respecto al patógeno y, asimismo, definen las reglas de transición entre ellos.

Uno de los modelos compartimentales más sencillos es el modelo SIS que permite describir enfermedades que no confieren inmunidad. Los dos compartimentos del modelo son: el compartimento de susceptibles (S), que recoge a los individuos que no son infecciosos en el instante de tiempo  $t$ , y el compartimento de infecciosos (I), que denota el estado en que los individuos están infectados del patógeno y además son capaces

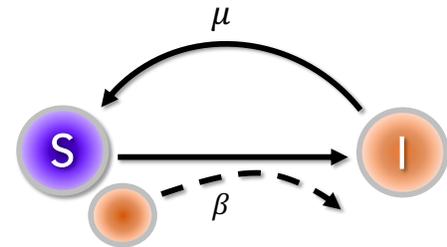


Figura 2.1: **Modelo compartimental SIS.** El contagio de los susceptibles se da según la tasa  $\beta$  por contacto con un infectado. La recuperación ocurre de manera espontánea con tasa  $\mu$ .

de contagiar a sus vecinos susceptibles. Las transiciones entre compartimentos de un tiempo  $t$  a un tiempo  $t + \Delta t$  se definen mediante tasas de paso. En el caso del modelo SIS, un individuo susceptible se convierte en infectado con tasa  $\beta$  por contacto con un individuo infectivo, y los infectados pueden volverse susceptibles de manera espontánea con tasa  $\mu$ . Para el modelo SIS en el estado estacionario todos los individuos son o susceptibles (S) o infectados (I), por lo que el parámetro de orden que plasma el impacto del patógeno en una población es la fracción de infectados<sup>1</sup> en el equilibrio dinámica, también llamado nivel endémico,  $I^\infty$ .

A modo de extensión del modelo SIS, el modelo SIRI es útil para describir enfermedades en

<sup>1</sup>En los modelos de ola epidémica, como el SIR, el parámetro que caracteriza el estado de la epidemia es la fracción de recuperados,  $R^\infty$

las que los individuos recuperados no son inmediatamente susceptibles, sino que tienen un periodo de inmunidad que, finalmente, se pierde. En este modelo los compartimentos disponibles son: el compartimento de susceptibles (S), el compartimento de infecciosos (I), y el compartimento de recuperados (R) al que pertenecen los individuos recuperados que tienen inmunidad y no son susceptibles otra vez de contraer la enfermedad. Al estar permitida la reinfección, el modelo SIRI presenta, al igual que el modelo SIS, un equilibrio endémico de recuperados, infectados y sanos en el estado activo de la epidemia [1]. Como veremos más adelante, el modelo que se presenta en este trabajo es una evolución del modelo SIS que incluye la adquisición de inmunidad efectiva sin necesidad de un compartimento extra de recuperados.

### 2.1.2. Número reproductivo básico y umbral epidémico

El número reproductivo básico  $R_0$  es el número promedio de infecciones secundarias causadas por un individuo infeccioso en una población de solo susceptibles durante el tiempo en que este individuo permanece infeccioso. Si  $R_0 > 1$ , cada individuo infectado es capaz de infectar a más de una persona sana. Por tanto, el patógeno se difunde entre la población y la epidemia se encuentra en fase activa caracterizada por un  $I^\infty \neq 0$ . Por el contrario, si  $R_0 < 1$ , un individuo infecta a menos de una persona, por lo que la epidemia no persiste y se halla en fase absorbente ( $I^\infty = 0$ ). En el caso del modelo SIS, este número reproductivo básico es  $R_0 = \frac{\beta}{\mu}$ .

El umbral epidémico es el valor mínimo de la tasa de contagio ( $\beta_C$ ) para el que tiene lugar un brote epidémico. En otras palabras, el valor de  $\beta_C$  será aquel que corresponde con  $R_0 = 1$  [2]. Por tanto, el umbral epidémico es el punto crítico de una transición de fase capturada con  $\beta$ , siendo el parámetro de control y  $I^\infty$  siendo el parámetro de orden. En el estado estacionario, el modelo SIS tiene dos posibles escenarios descritos por el parámetro de orden  $I^\infty$ . Si  $\mu > \beta$ , esto es, para una tasa de recuperación mayor que la tasa de infección, el patógeno desaparece de la población y se establece un estado libre de epidemia ( $I^\infty = 0$ ). En caso contrario,  $\mu < \beta$ , cuando la tasa de recuperación es menor que la tasa de infección, la enfermedad está presente continuamente y el parámetro de orden ( $I^\infty$ ) adquiere un valor constante [2], como se observa en la Figura 2.2.

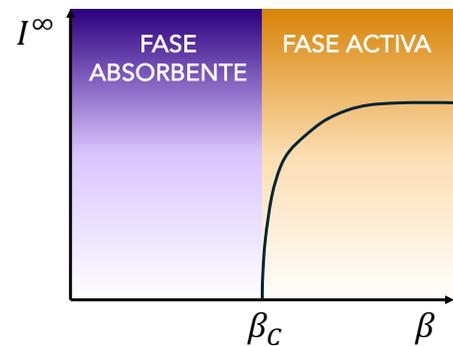


Figura 2.2: **Diagrama de fases de una epidemia.** Se muestra el umbral epidémico  $\beta_C$  diferenciando las fases absorbente y activa.

## 2.2. Mutación de patógenos

La evolución y propagación de epidemias virales son fenómenos complejos influenciados por diversos factores biológicos y ambientales. Un aspecto crucial es la capacidad de los virus para mutar y generar diversidad genética, lo cual les permite adaptarse rápidamente a nuevas situaciones, escapando de la respuesta inmune. Por ello, el estudio de las mutaciones es fundamental para comprender cómo los patógenos, especialmente los que contienen ARN, desencadenan epidemias.

Una mutación puntual es aquella en la que se altera solamente un nucleótido de una cadena de ARN. Las mutaciones se producen principalmente por errores durante la replicación del genoma viral causada por la baja fidelidad de la ARN-polimerasa, es decir, por la capacidad limitada de una enzima para replicar con precisión una cadena de ARN durante este proceso de replicación [3]. Los virus que presentan unas altas tasas de mutación tienden a evadir la inmunidad con más facilidad y pueden desarrollar resistencia a los tratamientos antivirales con mayor rapidez. Por ejemplo, el virus de la gripe experimenta cambios genéticos continuos, lo que dificulta la eficacia de las vacunas estacionales y requiere una vigilancia constante para adaptar las vacunas a las nuevas cepas predominantes [3].

Una cadena de ARN de  $N$  bases está compuesta por cuatro tipos de nucleótidos: adenina (A), uracilo (U), citosina (C) y guanina (G), que cumplen el criterio de complementariedad, de tal forma que la adenina se empareja con el uracilo, y la citosina con la guanina. El espacio de mutaciones abarca todas las posibles secuencias que se pueden formar a partir de esta cadena. Al fijar un nucleótido en una posición de la cadena, automáticamente se establece el nucleótido complementario. Por tanto, existen  $2^N$  posibles combinaciones para una cadena simple. En este trabajo se estudia la mutación de patógenos desde la perspectiva de las redes complejas, de forma que cada secuencia será un nodo que podrá estar conectado a otros siempre y cuando estos nodos representen secuencias que se relacionen con la primera a través de una mutación puntual.

### 2.3. Redes complejas

Las redes complejas son directorios o catálogos de las componentes de un sistema, y sus interacciones. Se denomina como vértices o nodos a sus elementos, y como enlaces o *links* a las interacciones a pares entre ellos. El objeto matemático que describe una red compleja [4] es un grafo  $G(\mathcal{N}, \mathcal{L})$  formado por dos conjuntos de elementos,  $\mathcal{N} \neq \emptyset$  y  $\mathcal{L}$ . Cada uno de los  $N$  nodos del conjunto  $\mathcal{N}$  se etiqueta con un índice  $i = 1, 2, \dots, N$ , mientras que cada uno de los  $L$  enlaces son, a su vez, pares  $(i, j)$  de elementos del conjunto  $\mathcal{N}$ . Se distingue entre grafos dirigidos y no dirigidos; en grafos dirigidos cada enlace  $(i, j)$  tiene asignada una dirección del nodo  $i$  al nodo  $j$ , mientras que en grafos no dirigidos todos los enlaces son bidireccionales. Además, si los enlaces tienen un peso asignado,  $w_{ij}$ , se habla de grafos pesados.

Centrándonos en el caso de redes no pesadas y no dirigidas, como las que vamos a tratar en este estudio, se les asigna a cada enlace el mismo peso,  $a_{ij} = 1 \forall (i, j) \in \mathcal{L}$ . La información de los  $L$  enlaces se puede almacenar en una matriz  $\mathbf{A} = \{a_{ij}\}$  llamada matriz de adyacencia, cuyos elementos se definen tal que

$$a_{ij} = a_{ji} = \begin{cases} 1 & \text{si } (i, j) \in \mathcal{L}, \\ 0 & \text{si } (i, j) \notin \mathcal{L}. \end{cases} \quad (2.3.1)$$

A partir de la matriz de adyacencia, se pueden calcular diferentes métricas que caracterizan la red. En particular, el grado de los nodos que representa el número de enlaces en los que participa el

nodo  $i$  es:

$$k_i = \sum_{j=1}^N a_{ij} \quad (2.3.2)$$

Partiendo de la matriz de adyacencia y el de grado de las redes, se puede construir la matriz laplaciana  $\mathbf{L} = \{l_{ij}\}$ , cuyos elementos están definidos en función de la matriz de adyacencia y el grado de los nodos:

$$l_{ij} = \delta_{ij}k_j - a_{ij}. \quad (2.3.3)$$

La matriz laplaciana tiene propiedades matemáticas útiles. En particular, todas las filas y columnas de la matriz laplaciana suman cero, es decir,  $\sum_{i=1}^N l_{ij} = \sum_{j=1}^N l_{ji} = 0$ . Por tanto,  $\mathbf{L}$  tiene un autovalor igual a cero, cuya multiplicidad se corresponde con el número de componentes conexas de la red. Además, todos los autovalores son números no negativos dado que  $L$  es una matriz semi-definida positiva. Así, se pueden ordenar los valores de  $L$  en orden creciente como  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ . El segundo autovalor más pequeño ( $\lambda_2$ ) se conoce como conectividad algebraica, ya que contiene información importante sobre la conectividad de una red [5], y determina la velocidad de convergencia de los procesos que operan en dicha red [6], como los procesos de difusión [7].

La probabilidad de que dado un nodo aleatorio del grafo  $G$ , este tenga  $k$  vecinos se recoge en la distribución de grado  $P(k)$ . A partir de las distribuciones de grado  $P(k)$ , se define el  $n$ -ésimo momento de la distribución  $\langle k^n \rangle = \sum_{k=0}^{N-1} k^n P(k)$ , correspondiéndose el primer momento de la distribución  $\langle k \rangle$  con el valor medio del grado, mientras que el segundo momento  $\langle k^2 \rangle$  se relaciona con la varianza. Asociada a esta idea, se define la función de distribución acumulativa complementaria de grado (CCDF),  $P(K \geq k)$ , que es la probabilidad de que un nodo tenga al menos  $k$  vecinos:  $P(K \geq k) = \sum_{k'=k}^{\infty} P(k')$ . Esta función es útil porque resalta la cola de la distribución y permite visualizar cuántos nodos tienen grados extremadamente altos.

Otra métrica relevante es el coeficiente de *clustering*, que cuantifica la tendencia de los nodos a formar triángulos, es decir, que dos nodos conectados con un tercero, estén conectados entre sí. El *clustering* de un nodo  $i$  de la red queda determinado como

$$C_i = \frac{\text{triángulos conectados al nodo } i}{\text{posibles triángulos conectados}} = \frac{\sum_{j=1}^N \sum_{k=1}^N A_{ij} A_{jk} A_{ki}}{k_i(k_i - 1)}, \quad (2.3.4)$$

y el *clustering* promedio de se expresa como  $C = \frac{1}{N} \sum_{i=1}^N C_i$ .

La asortatividad es la tendencia de los nodos a conectarse con otros nodos que tienen un grado similar. En otras palabras, mide la correlación entre los grados de los nodos conectados. La asortatividad por grado  $r$  se calcula utilizando el coeficiente de correlación de Pearson para los grados de los nodos en cada arista dado por

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2} \quad (2.3.5)$$

El término  $q_k$  es la distribución del grado restante, es decir, del número de enlaces salientes del

nodo, distintas de las que conectan el par  $(j, k)$ . La distribución de este término se deriva de la distribución de grado  $p_k$  como  $q_k = \frac{(k+1)p_{k+1}}{\sum_{j \geq 1} j p_j}$ . En cuanto al término  $e_{jk}$  se refiere a la distribución conjunta de los grados restantes de ambos vértices. Esta cantidad es  $\sum_{jk} e_{jk} = 1$ , y  $\sum_j e_{jk} = q_k$ .

El valor de  $r$  varía entre -1 y 1 con diferente casuística: (i)  $r > 0$  indica una asortatividad positiva, es decir, los nodos tienden a conectarse con otros nodos de grado similar. (ii)  $r < 0$  indica una asortatividad negativa, es decir, los nodos tienden a conectarse con nodos de grados diferentes. (iii)  $r = 0$  indica que no hay correlación entre los grados de los nodos conectados.

Una vez establecidas las métricas y representaciones necesarias, las redes utilizadas en este estudio pueden ser definidas. En primer lugar, las redes *lattice* regulares son aquellas en las que todos los nodos tienen el mismo grado  $z$  y, por lo tanto, la distribución de grados está dada por  $P(k) = \delta(z - k)$ , y además siguen una estructura geométrica. En particular, se utilizará una red de cadena lineal unidimensional caracterizada por  $\langle k \rangle = 2$ , una *lattice* regular bidimensional con  $\langle k \rangle = 4$ , y una *lattice* regular tridimensional con nodos equiespaciados y  $\langle k \rangle = 6$ .

Si se rebaja la condición de *lattice*, se obtienen redes regulares aleatorias (*Random Regular Graph*, RRG). En la red RRG todos los nodos tienen el mismo grado  $z$ , pero las conexiones entre ellos se distribuyen de manera aleatoria. Ahora, si en vez de considerar  $P(k) = \delta(z - k)$  se considera una distribución de grado de tipo Poisson, se llega a la red de tipo Erdős-Rényi [4] Otra red utilizada es la *Random Geometric Graph* (RGG), que se genera al ubicar los nodos en un espacio métrico y, posteriormente, se crea una conexión entre dos nodos si la distancia entre ellos es inferior a un umbral preestablecido  $r$ . Estas redes también presentan una distribución de grado  $P(k)$  de tipo Poisson, pero a diferencia de las redes aleatorias se caracterizan por un elevado coeficiente de *clustering*.

A pesar de la importancia académica de los anterior tipos de redes homogéneas, la mayoría de redes reales como internet [8] son redes heterogéneas y libres de escala. Esto se traduce en que tienen una distribución de grado  $P(k)$  que obedece a una ley de potencias  $P(k) \sim k^{-\gamma}$  con  $\gamma \in (2, 3)$ . Para  $\gamma < 3$  el segundo momento de la distribución diverge, por lo que el grado promedio no es un buen descriptor en la topología de estas redes. El comportamiento libre de escala revela la coexistencia dentro de una misma red de muchos elementos con muy poca conectividad, y de unos pocos conectados a una gran cantidad de nodos, llamados *hubs*. El modelo más conocido que representa las redes libres de escala es el modelo de Barabási-Albert [4], que construye las redes agregando nodos de manera incremental, y cada nuevo nodo se conecta a nodos ya existentes con una probabilidad proporcional al grado de los nodos existentes. Asimismo, el modelo minimal de red que captura la naturaleza libre de escala es la red estrella. Su estructura consta de un nodo central que actúa como nexo, conectado a todos los demás nodos, los cuales no están interconectados entre sí.

### 2.3.1. Redes de genotipos reales

Como en este trabajo las redes van a representar la estructura de genotipos, cada nodo representa uno de estos genotipos, y cada enlaces conecta aquellos que difieren en una sola mutación puntual [9]. La topología de la red de genotipos refleja la estructura evolutiva y la diversidad genética. Por

ejemplo, si se establece la distancia entre dos genotipos como el número de enlaces del camino más corto entre ellos, los genotipos similares entre sí se encuentran más cerca en el espacio de mutaciones, mientras que los genotipos más divergentes están más alejados.

Teóricamente, en las redes de genotipos completas se representarían las  $2^N$  posibles secuencias de un patógeno (cuya secuencia genética consta de  $N$  bases) sin importar la viabilidad de cada cepa. No obstante, en la práctica solo un subconjunto de cepas emergen y son visibles, dejando lo que se conoce como la red de genotipos realizada. Además, el conjunto de datos empíricos es limitado y, por tanto, esta red se submuestra aún más, con lo que se obtiene la red de genotipos observada [10].

Un ejemplo de red de genotipos real es la red de INFV A (H3N2) [10] que captura las relaciones genéticas<sup>2</sup> de las cepas de INFV H3N2 entre el año 2010 y 2020. Esta estructura, representada en la Figura 2.3.(a) tiene 9714 nodos (cepas únicas), 7599 enlaces (posibles puntos de mutación entre cepas), y 3262 componentes conectadas, de las cuales 384 con más de un nodo. B. J. M. Williams et al. [10] observan que la cola de la función de distribución complementaria de grado (CCDF),  $P(K \geq k)$ , presenta una ley de potencias del tipo  $P(K \geq k) \propto k^{-\alpha_k}$  con un exponente estimado de  $\alpha_k = 2.29$ , véase Figura 2.3.(b). En el artículo presentado por B. J. M. Williams et al. [10] se estima que el grado de asortatividad, o la correlación entre el grado de un nodo y el de sus vecinos, exhibe una tendencia negativa. Esto indica la preferencia de los nodos con alto grado por unirse a nodos de bajo grado. Además, los autores observan que el número de triángulos, proporcional al coeficiente de *clustering*, no es insignificante.

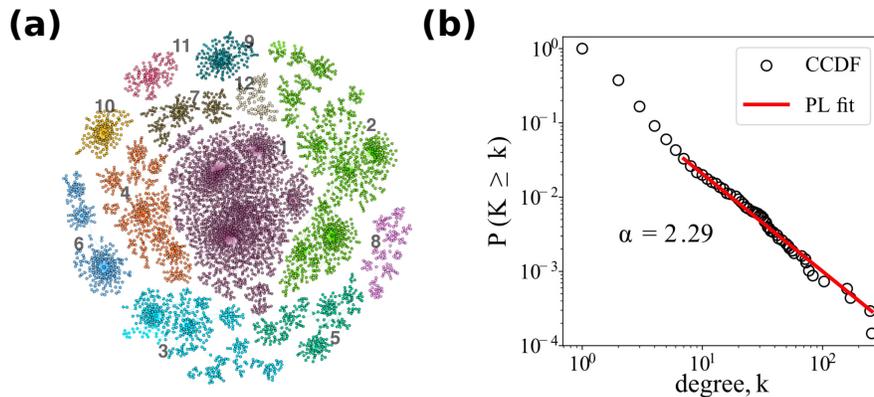


Figura 2.3: **Red de INFV A (H3N2)**. Imágenes tomadas del artículo de B. J. M. Williams et al. [10]. El panel (a) es la representación gráfica de la red de genotipos INFV A (H3N2). Cada nodo representa un genotipo. El panel (b) muestra la CCDF del grado para la red de INFV A (H3N2) con los puntos obtenidos por muestreo de la red y el ajuste a una ley de potencias.

### 2.3.2. Redes de genotipos sintéticas

Una vez conocidas las propiedades de las redes reales de genotipos, introducimos un modelo sintético que permite generar redes de características similares. Para ello, se distribuye las cepas en un espacio métrico 2D que aproxima el vasto espacio de genotipos, y se asigna a cada cepa un

<sup>2</sup>Nótese que no se considera la totalidad del código genético, sino solo la parte de la proteína HA del INFV A (H3N2).

*fitness* determinado, que se corresponde con su adecuación biológica, es decir, su capacidad relativa para sobrevivir, reproducirse y propagarse en un entorno determinado.

Para generar las redes, se siguen los siguientes pasos: (i) Se genera el espacio de genotipos como un paisaje 2D donde se van a incrustar las  $M \leq 2^N$  cepas. (ii) Se eligen ubicaciones específicas en el paisaje 2D para representar  $W_M$  cepas silvestres (*wild types*), es decir, las primeras variantes de un virus descubiertas en la naturaleza. (iii) Se distribuye la probabilidad de encontrar los  $M$  genotipos restantes alrededor de las posiciones de las cepas silvestres a través de una distribución gaussiana. (iv) Se realiza el muestreo de las  $M$  cepas siguiendo la distribución gaussiana, y así representar los nodos de la red de genotipos observada. (v) Se crean enlaces entre las distintas cepas siguiendo dos reglas: Primero se establece un umbral de distancia geométrica  $\theta_g$  para determinar qué pares de nodos están lo suficientemente cerca en el paisaje como para diferir en una sola mutación puntual  $|x^\sigma - x^\nu| < \theta_g$ . Segundo, se elige una regla de *fitness* según el trabajo previo de G. Caldarelli et al. [11]. Se establece un rango de valores posibles de *fitness*  $f \in [0, \infty)$  y se asignan a cada uno de los  $M$  nodos de acuerdo a la distribución de probabilidad  $P(f) \sim e^{-x}$ . Se crea una conexión entre dos nodos  $\sigma$  y  $\nu$  si se cumple la condición  $f^\sigma + f^\nu \geq \theta_f$  y, además, se cumple la condición anterior.

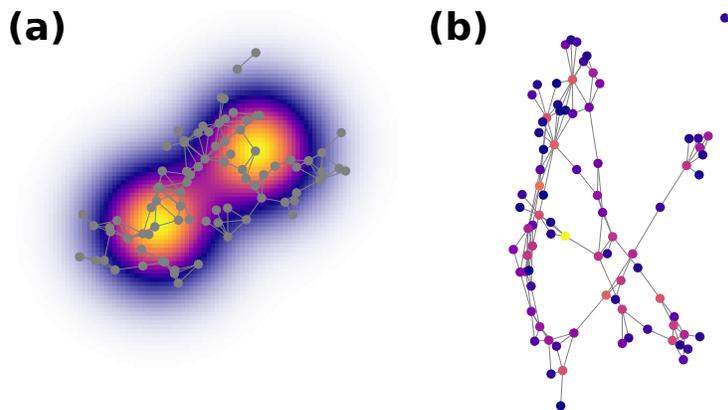


Figura 2.4: **Representación de una red generada a partir del modelo sintético.** En el panel (a) se presenta el espacio 2D de dimensión  $1 \times 1$  con  $M = 100$  cepas. Se han preestablecido  $W_M = 2$  cepas silvestres, cada una con desviación estándar igual a 0.015. Las reglas de creación de enlaces se rigen por el parámetro geométrico  $\theta_g = 0.07$  y el parámetro de *fitness*  $\theta_f = 2$ . En el panel (b) se observa solamente la red del panel (a) en la que el color de cada nodo indica el *fitness* asociado.

Las redes generadas por este modelo son dependientes de la elección de parámetros, por lo que en la Figura 2.5 se presentan el mapa de calor del grado promedio de los nodos  $\langle k \rangle$ , el coeficiente de clustering  $C$  y la asortatividad  $r$ .

La Figura 2.5.(a) muestra el mapa de calor de la conectividad promedio  $\langle k \rangle$  en función de  $\theta_g$  y  $\theta_f$ . A medida que aumenta  $\theta_g$  y disminuye  $\theta_f$ , el grado promedio también aumenta. Esto se debe a que un  $\theta_g$  mayor implica que los nodos pueden estar conectados a otros a mayores distancias, y un parámetro de *fitness*  $\theta_f$  pequeño permite que cepas menos favorecidas biológicamente se unan con las favorecidas. En la gráfica (b) de la Figura se percibe como el coeficiente de clustering  $C$  exhibe la misma tendencia que  $\langle k \rangle$  de ver incrementada su conectividad para valores altos de  $\theta_g$  y pequeños de  $\theta_f$ . Finalmente, en la gráfica (c) de la Figura se presenta el mapa de color de la asortatividad  $r$ .

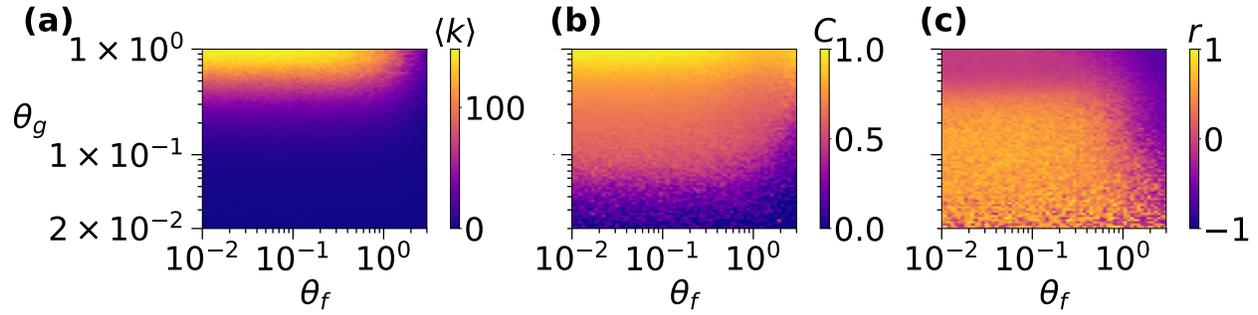


Figura 2.5: **Propiedades de las redes sintéticas.** El panel (a) representa el mapa de calor para el grado promedio  $\langle k \rangle$  en función del parámetro geométrico  $\theta_g$  y el parámetro de *fitness*  $\theta_f$ , el mapa de calor del panel (b) es para el coeficiente de clustering  $C$ . Finalmente, el panel (c) contiene el mapa de calor de la asortatividad  $r$ . Los parámetros de la simulación son  $M = 150$  cepas,  $W_M = 5$  cepas silvestres y  $\sigma = 0.015$ .

En general, la asortatividad parece positiva para valores pequeños tanto de  $\theta_g$  y  $\theta_f$ , así las cepas con un número similar de conexiones tienden a formar clústers. y negativa para valores elevados. Para valores elevados de  $\theta_g$  y  $\theta_f$ , la asortatividad parece negativa, lo que conduce a una mayor variabilidad genética, dado que los nodos muestran una tendencia a conectarse con otros nodos de grado diferente.

Las redes presentan valores más elevados tanto del grado promedio  $\langle k \rangle$  como del coeficiente de clustering  $C$  cuando el umbral geométrico  $\theta_g$  es más grande, ya que se permiten la creación de enlaces entre nodos más lejanos.

### 3. Modelo SIMS para la propagación y mutación de patógenos

#### 3.1. Formalismo matemático del modelo SIMS

Para estudiar la dinámica emergente de la propagación de una patógeno en una población, considerando que el patógeno muta de acuerdo a una red de genotipos subyacente, introducimos el modelo compartimental presentado en la Figura 3.1. Los individuos de la población pertenecen al compartimento de susceptibles ( $S$ ) si no son infecciosos de ninguna cepa de la enfermedad. Una vez infectados, se accede a uno de los  $M$  compartimentos de infecciosos según la cepa con la que han sido infectados. En este momento, los individuos infecciosos pueden recuperarse y volver al estado de susceptibles, o mutar y ser infecciosos de otra cepa. Además, el modelo introduce la generación de inmunidad efectiva, ya que la tasa de recuperación respecto a una cepa determinada aumenta conforme la población genera anticuerpos al infectarse. Para modelar matemáticamente esta dinámica, se utiliza el siguiente sistema de ecuaciones diferenciales:

$$\dot{\rho}^\sigma = \beta \left( 1 - \sum_{\nu=1}^M \rho^\nu \right) \rho^\sigma - \mu^\sigma \rho^\sigma + \gamma \sum_{\nu=1}^M \Gamma^{\sigma\nu} (\rho^\nu - \rho^\sigma), \quad (3.1.1)$$

$$\dot{\mu}^\sigma = \alpha \frac{\rho^\sigma}{K_0 + \rho^\sigma} + \Upsilon(\mu_0 - \mu^\sigma). \quad (3.1.2)$$

La Ecuación 3.1.1 modela cómo cambia la fracción de la población infectada por la cepa  $\sigma$  ( $\rho^\sigma$ ) a través del tiempo. Cuando un individuo susceptible entra en contacto con un individuo infeccioso de la cepa  $\sigma$ , el susceptible se convierte en infeccioso con una tasa de infectividad marcada por  $\beta$ . Una vez en este estado infeccioso hay dos alternativas : (i) recuperarse con una tasa  $\mu^\sigma$ , y volver al compartimento de susceptibles; o (ii) convertirse en infeccioso de la cepa  $\nu \neq \sigma$  a través de un proceso de mutación regulado por la tasa  $\gamma$ , y condicionado por el término  $\Gamma^{\sigma\nu}$  de la matriz de adyacencia.

En cuanto a la tasa de recuperación del genotipo  $\sigma$  ( $\mu^\sigma$ ) dada por la Ecuación 3.1.2, su valor depende del tiempo y está inspirada en la respuesta inmune de un virus que aumenta por la presencia del patógeno (ganancia de inmunidad) y se relaja hasta un valor estable cuando el patógeno ya no está presente (pérdida de inmunidad) [12].

El término de ganancia de inmunidad,  $\alpha \frac{\rho^\sigma}{K_0 + \rho^\sigma}$ , tiene forma de función de Hill e indica que la tasa de recuperación aumenta con la fracción  $\rho^\sigma$ , y proporcionalmente al coeficiente  $\alpha$ , que representa la influencia efectiva de la fracción de infectados en la recuperación.

Las funciones de Hill son utilizadas en sistemas biológicos con efectos de saturación.  $K_0^{-1}$  es la transcendencia de la inmunidad, de forma que valores bajos de este término conllevan a que la tasa de recuperación crezca más lentamente y los genotipos se acumulen más en la población antes de que su tasa de recuperación aumente significativamente, reflejando una menor respuesta inmunológica. En el lado contrario, valores altos de  $K_0^{-1}$  hacen que el término  $\frac{\rho^\sigma}{K_0 + \rho^\sigma}$  se aproxime rápidamente a 1 conforme  $\rho^\sigma$  aumenta, lo que implica que la tasa de recuperación  $\mu^\sigma$  se incrementa rápidamente, protegiendo a la población de la cepa  $\sigma$ . Por otro lado, el término de pérdida de inmunidad tiene una forma funcional tal que en ausencia de infectados de una cepa  $\sigma$  la tasa de recuperación regresa a su valor basal  $\mu_0$ . El parámetro  $\Upsilon$  actúa como un regulador de la escala temporal de la pérdida de memoria inmunológica.

### 3.2. Umbral epidémico en el modelo SIMS

Una vez introducido el modelo, pretendemos comprobar que la fenomenología reportada en la literatura [9, 10] para modelos de pérdida de inmunidad y mutación se recupera con el modelo SIMS. Para ello, en la Figura 3.2.(d)-(i) presentamos los diagrama de fases en función de la infectividad ( $\beta$ ) y la transcendencia de la inmunidad ( $K_0^{-1}$ ) para las mismas tres redes sintéticas utilizadas por B. J. M. Williams et al. [9]: una red estrella, una red regular *lattice* en dos dimensiones y una red regular de tipo cadena línea de una dimensión. Además de representar la fracción estacionaria de infectados en los paneles (g)-(i), en los paneles (d)-(f) cuantificamos la localización de infectados en el estado

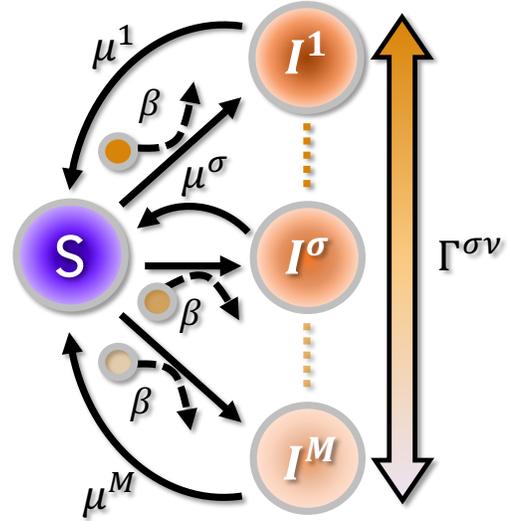


Figura 3.1: **Diagrama de flujo del modelo SIMS.** Los compartimentos epidemiológicos representados son para los susceptibles ( $S$ ) e infectados  $\rho^\sigma$  para cada una de las  $M$  cepas.

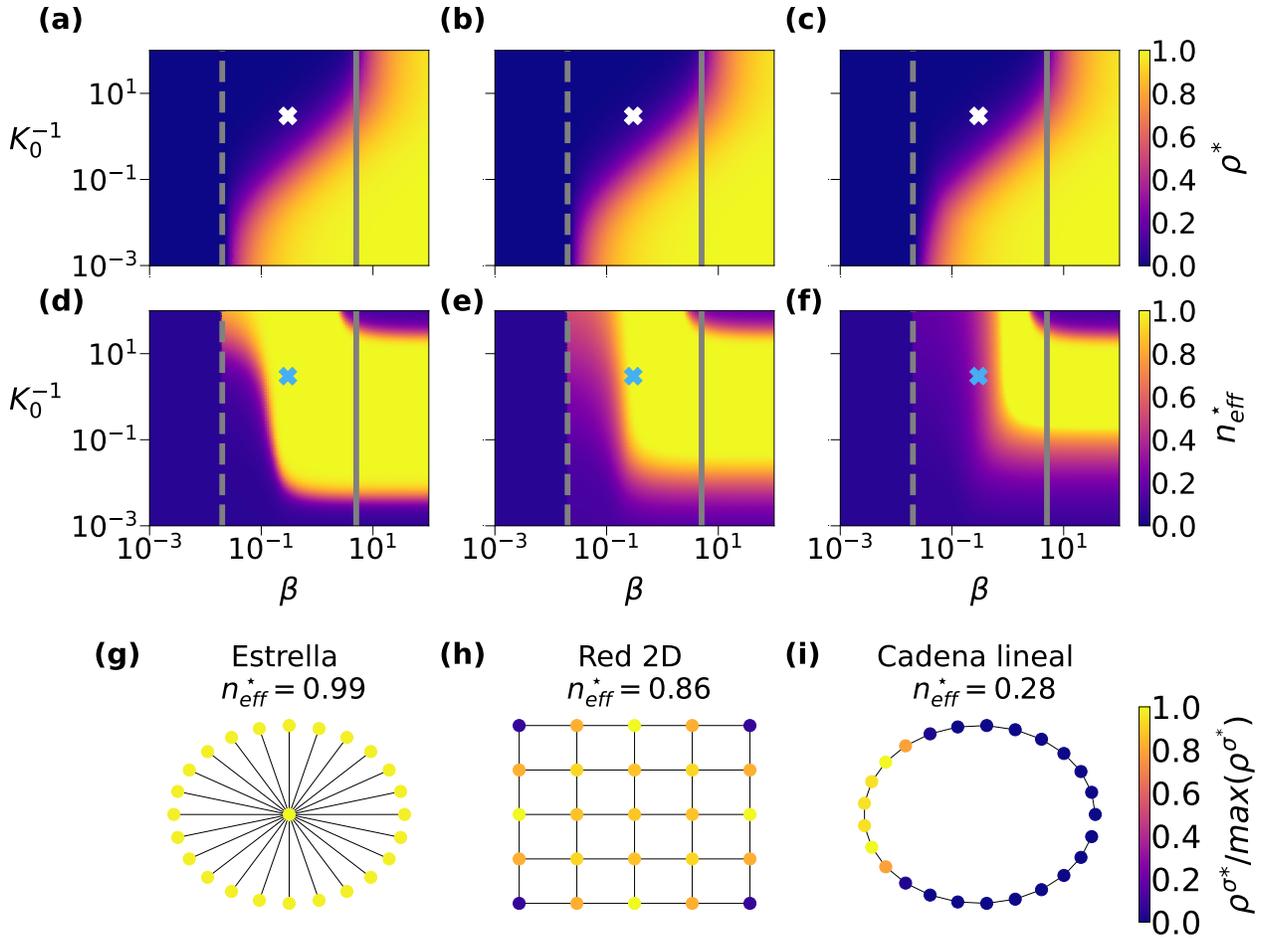


Figura 3.2: **Localización de la infección y características del estado de la epidemia.** Los paneles (a), (b) y (c) muestran la fracción de infectados en el estado estacionario para las redes estrella, red *lattice* 2D y cadena lineal respectivamente, con dos umbrales epidémicos bien diferenciados, calculados analíticamente con el desarrollo posterior. Los paneles (d), (e) y (f) presentan los regímenes de la localización de la infección a través de un mapa de calor del tamaño efectivo de la muestra. En estas seis gráficas los parámetros fijos son  $M = 25$ ,  $\gamma = 0.001$ ,  $\alpha = 1$  y  $\Upsilon = 0.2$ . Finalmente, en los paneles (g)-(i) se presenta la localización de la enfermedad en diferentes redes, para la combinación de parámetros marcados con una cruz en los paneles (a)-(f).

estacionario mediante el número de Kish, que se define como la tasa efectiva de participación

$$n_{\text{eff}}^* = \frac{n_{\text{eff}}}{M} = \frac{(\sum_{\sigma} \rho^{\sigma})^2}{M \sum_{\sigma} \rho^{\sigma^2}}. \quad (3.2.1)$$

Cuando  $n_{\text{eff}}^* \rightarrow 1$ , todas las cepas contribuyen por igual a la infección. En cambio, cuando  $n_{\text{eff}}^* \rightarrow M^{-1}$ , solamente una cepa contribuye de manera efectiva a la infección.

Observando el estado estacionario en las Figura 3.2.(a)-(a) se aprecia que para valores bajos de  $\beta$  nos encontramos en la fase absorbente, incluso si  $K_0^{-1}$  es alto. El patógeno no tiene poder infectivo suficiente como para desencadenar una epidemia. A medida que  $\beta$  aumenta encontramos la fase activa de la dinámica, con una fracción estacionaria de infectados finita. Si  $K_0^{-1}$  es bajo, es decir, si el organismo adquiere poca inmunidad al patógeno, la recuperación ocurre como en el modelo SIS

original. Por lo contrario, si  $K_0^{-1}$  es alta, se necesita que la tasa de infectividad  $\beta$  sea mayor para alcanzar un estado activo de la epidemia. Teniendo en cuenta estas observaciones, parece que existe un cambio en el umbral epidémico  $\beta_C$  en función de la trascendencia de la inmunidad  $K_0^{-1}$ .

Asimismo, en las Figura 3.2.(d)-(f) se revelan diferentes regímenes de participación efectivos dentro de la fase activa de la epidemia. El número de Kish nos permite diferenciar entre el régimen localizado y el deslocalizado. Para la red estrella, panel (d), da lugar a una transición bastante abrupta entre un estado de baja y alta participación, ya que el nodo central, que se ha utilizado como semilla, facilita una propagación rápida y uniforme a través de toda la red. En cuanto a la red regular *lattice* 2D, panel (e), se observa una transición más gradual, lo que sugiere una propagación más uniforme de la infección, coincidiendo con que se ha reducido el número de vecinos primeros del nodo semilla. Finalmente, para la estructura de cadena lineal, imagen (f), la alta tasa de participación efectiva está en un rango de valores más restringido, resultando en una mayor localización y una difusión más lenta, ya que los nodos deben infectarse secuencialmente. Asimismo, en las tres redes, se observa que para valores muy elevados de la tasa de infectividad ( $\beta$ ) y de la trascendencia de la inmunidad ( $K_0^{-1}$ ) el número de Kish tiende a cero, lo que refleja la combinación de rápida propagación de la infección y una efectiva inmunización. Esta dinámica reduce participación efectiva porque la propagación de la infección y la adquisición de inmunidad ocurren casi simultáneamente en la red.

Las Figuras 3.2.(g)-(i) ejemplifican la distribución espacial de la fracción de infectados en el estado estacionario normalizado por el máximo de estos valores, para ciertas combinaciones de parámetros ( $\beta, K_0^{-1}$ ) señaladas por la cruz en los paneles (a)-(f). Efectivamente, se observa que en el ejemplo de la red estrella, panel (g), la participación efectiva es muy alta, ya que todos los nodos se hallan infectados. Por otro lado, los paneles (h) e (i), con la red *lattice* 2D y la cadena lineal respectivamente, son ejemplos del régimen localizado, donde la participación efectiva va decreciendo al reducir la conectividad promedio del nodo semilla.

Los resultados aquí presentados son consistentes con los obtenidos por B. J. M. Williams et al. [9]. En su artículo, los autores también obtenían que la trascendencia de la inmunidad reduce el impacto de la epidemia en el estado estacionario, y hallaban tanto regiones localizadas como deslocalizadas dentro del diagrama de fases.

A partir de las Ecuaciones (3.1.1)-(3.1.2) podemos derivar el umbral epidémico  $\beta_c$  para capturar la aparente dependencia con  $K_0^{-1}$  mostrada en las Figuras 3.2.(a)-(c). Se empieza recordando la relación 2.3.2, y extrapolando para el modelo SIMS se tiene que  $\sum_{\nu=1}^M \Gamma^{\sigma\nu} = k^\sigma$ . Por tanto, las ecuaciones del modelo se pueden reescribir como:

$$\dot{\rho}^\sigma = \beta \left( 1 - \sum_{\nu=1}^M \rho^\nu \right) \rho^\sigma - \mu^\sigma \rho^\sigma + \gamma \sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^\nu - \gamma k^\sigma \rho^\sigma, \quad (3.2.2)$$

$$\dot{\mu}^\sigma = \alpha \frac{\rho^\sigma}{K_0 + \rho^\sigma} + \Upsilon(\mu_0 - \mu^\sigma). \quad (3.2.3)$$

En el estado estacionario  $\dot{s} = \dot{\rho}^\sigma = 0$ ,  $\forall \sigma \in \mathcal{M}$ , por lo que se deriva una expresión explícita para la tasa de recuperación:

$$\mu^{\sigma^*} = \frac{1}{\rho^{\sigma^*}} \left\{ \beta \left( 1 - \sum_{\nu=1}^M \rho^{\nu^*} \right) \rho^\sigma + \gamma \sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^{\nu^*} - \gamma k^\sigma \rho^{\sigma^*} \right\} \quad (3.2.4)$$

$$= \beta \left( 1 - \sum_{\nu=1}^M \rho^{\nu^*} \right) + \frac{\gamma \sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^{\nu^*}}{\rho^{\sigma^*}} - \gamma k^\sigma. \quad (3.2.5)$$

Si ahora se introduce este resultado en la Ecuación 3.2.2 se obtiene la siguiente ecuación implícita:

$$\rho^{\sigma^*} = \frac{\gamma \Upsilon \sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^{\nu^*}}{\mu_0 \Upsilon - \beta \Upsilon \left( 1 - \sum_{\nu=1}^M \rho^{\nu^*} \right) + \gamma \Upsilon k^\sigma + \alpha \frac{\rho^{\sigma^*}}{K_0 + \rho^{\sigma^*}}}. \quad (3.2.6)$$

La iteración de esta ecuación mediante el método del punto fijo da la solución al estado estacionario del sistema dado un conjunto de parámetros.

Cerca del umbral epidémico  $\rho^{\sigma^*} \ll 1$  y por tanto podemos linearizar la ecuación. Los términos  $\sum_{\nu=1}^M \rho^{\nu^*}$  y  $\alpha \frac{\rho^{\sigma^*}}{K_0 + \rho^{\sigma^*}}$  quedan como términos cuadráticos si estuvieran en el numerador del lado izquierdo, y se pueden despreciar. Por ende, la expresión se convierte en lo siguiente:

$$\rho^{\sigma^*} = \frac{\gamma}{\mu_0 - \beta + \gamma \sum_{\nu=1}^M \Gamma^{\sigma\nu}} \sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^{\nu^*}. \quad (3.2.7)$$

Reorganizando términos y teniendo en cuenta que  $\sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^{\nu^*} = k^\sigma \rho^{\sigma^*}$ , se obtiene:

$$\frac{\mu_0 - \beta}{\gamma} \rho^{\sigma^*} = \sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^{\nu^*} - \sum_{\nu=1}^M \Gamma^{\sigma\nu} \rho^{\sigma^*} = \sum_{\nu=1}^M (\Gamma^{\sigma\nu} - \delta^{\sigma\nu} k^\nu) \rho^{\nu^*}, \quad (3.2.8)$$

esta ecuación se puede reescribir en términos de la matriz Laplaciana  $l^{\sigma\nu} = \delta^{\sigma\nu} k^\nu - \Gamma^{\sigma\nu}$  como una ecuación de autovalores

$$\frac{\mu_0 - \beta}{\gamma} \rho^{\sigma^*} = - \sum_{\nu=1}^M l^{\sigma\nu} \rho^{\nu^*} \iff \frac{\mu_0 - \beta}{\gamma} \boldsymbol{\rho}^* = -\mathbf{L} \boldsymbol{\rho}^*, \quad (3.2.9)$$

cuya solución para la tasa mínima de infección y poder dar lugar a un brote epidémico se corresponde con el menor autovalor de la matriz:

$$\beta_c = \mu_0 + \gamma \Lambda_{\min}(\mathbf{L}). \quad (3.2.10)$$

Para el caso particular de una red de genotipos conectada el mínimo autovalor  $\Lambda_{\min}(\mathbf{L})$  de la matriz Laplaciana es cero, lo que da como consecuencia el resultado:

$$\boxed{\beta_c = \mu_0} \quad (3.2.11)$$

Es decir, el umbral epidémico no depende de la trascendencia de la inmunidad  $K_0^{-1}$ , sino que es igual que en el modelo SIS original. No obstante, para el caso de máxima adquisición de inmunidad ( $K_0 = 0$ ), el término  $\alpha \frac{\rho^{\sigma^*}}{K_0 + \rho^{\sigma^*}} \rightarrow \alpha$  y la ecuación de autovalores se escribe en la forma

$$\frac{\mu_0 \Upsilon - \beta \Upsilon + \alpha}{\gamma \Upsilon} \rho^{\sigma^*} = -\mathbf{L} \rho^{\sigma^*}. \quad (3.2.12)$$

De esta forma, el umbral epidémico  $\beta_c$  ahora viene dado por la expresión:

$$\beta_c = \frac{\mu_0 \Upsilon + \alpha + \gamma \Upsilon \Lambda_{\min}(\mathbf{L})}{\Upsilon} \Rightarrow \boxed{\beta_c = \mu_0 + \frac{\alpha}{\Upsilon}} \quad (3.2.13)$$

En las Figuras 3.2.(g)-(i) representamos los valores del umbral epidémico según las Ecs. (3.2.11) y (3.2.13) con línea discontinua y continua respectivamente. Entre ambos valores del umbral epidémico la epidemia sigue en fase activa, pero con una incidencia muy baja, por lo que, de manera efectiva, el umbral es una interpolación entre ambos valores analíticos.

### 3.3. Dinámica del modelo SIMS en ausencia de mutaciones

Para comprender la dinámica del modelo completo aislamos cada uno de sus mecanismos. En particular, comenzamos explicando la influencia de la adquisición de inmunidad en la dinámica. Para ello, simplificamos el modelo considerando que todas las cepas son equivalentes. Bajo esa asunción de campo medio, el término de mutación se anula, y el conjunto de  $2M$  ecuaciones del modelo SIMS se convierten en únicamente

$$\dot{\rho} = \beta(1 - M\rho)\rho - \mu\rho, \quad (3.3.1)$$

$$\dot{\mu} = \alpha \frac{\rho}{K_0 + \rho} + \Upsilon(\mu_0 - \mu). \quad (3.3.2)$$

En el estado estacionario,  $\dot{\rho} = \dot{\mu} = 0$ , podemos despejar la tasa de recuperación:

$$\mu^* = \frac{\beta(1 - M\rho)\rho}{\rho} = \beta(1 - M\rho). \quad (3.3.3)$$

Cabe destacar que la simplificación de  $\rho$  en el denominador hace ignorar la solución trivial de la fase absorbente de la epidemia, dada por  $\rho^* = 0$  y  $\mu^* = \mu_0$ . Introduciendo la Ecuación 3.3.3 en la Ecuación 3.3.2 se obtiene la siguiente ecuación cuadrática

$$0 = \rho^2 A + \rho B + C, \quad (3.3.4)$$

donde

$$A = M\Upsilon\beta, \quad (3.3.5)$$

$$B = \alpha + \Upsilon[\mu_0 - \beta(1 - MK_0)], \quad (3.3.6)$$

$$C = \Upsilon K_0(\mu_0 - \beta). \quad (3.3.7)$$

Las dos soluciones de la ecuación de segundo grado se unen a la solución absorbente como posibles valores de  $\rho^*$ :

$$\rho^* = \begin{cases} 0 \\ \frac{-B + \sqrt{B^2 - 4AC}}{2A} \\ \frac{-B - \sqrt{B^2 - 4AC}}{2A} \end{cases} \quad (3.3.8)$$

Una vez en este punto, imponemos que  $B^2 - 4AC = 0$  para que exista al menos una solución no trivial. Dicha condición implica que  $\beta > \mu_0 \Leftrightarrow \frac{\beta}{\mu_0} > 1$ . Así, podemos definir el número reproductivo básico que controla la aparición de las soluciones no triviales como

$$R_0 = \frac{\beta}{\mu_0}. \quad (3.3.9)$$

No obstante, cuando  $K_0 = 0$  la condición  $B^2 - 4AC = 0$  implica que  $\Upsilon\beta > \alpha + \Upsilon\mu_0 \Leftrightarrow \frac{\Upsilon\beta}{\alpha + \Upsilon\mu_0} > 1$ , lo que permite definir un nuevo número reproductivo básico

$$\bar{R}_0 = \frac{\Upsilon\beta}{\alpha + \Upsilon\mu_0} \quad (3.3.10)$$

La Ecuación 3.3.10 se convierte automáticamente en la Ecuación 3.3.9 en el caso en que  $\alpha = 0$  y  $\Upsilon = 1$ . Asimismo, cabe destacar que para cada uno de los números reproductivos dados por la Ecuación 3.3.10 y la Ecuación 3.3.9 se obtienen los umbrales epidémicos  $\beta_c = \mu_0$  y  $\bar{\beta}_c = \mu_0 + \frac{\alpha}{\Upsilon}$  respectivamente, coincidiendo con el resultado obtenido en la Subsección 3.2.

En la Figura 3.3 se representan las curvas dadas por las expresiones analíticas 3.3.8. El panel (a) de la Figura 3.3 representa la evolución de la fracción de infectados en el estado estacionario de la epidemia en función del número reproductivo básico  $R_0 = \frac{\beta}{\mu_0}$  para diferentes valores de  $K_0$ , junto a los dos umbrales epidémicos  $\beta_c$  y  $\bar{\beta}_c$ . Cuando la transcendencia de la inmunidad es máxima,  $K_0 = 0$ , el umbral epidémico se sitúa en el valor del número reproductivo básico  $\bar{R}_0$ , y conforme se va aumentando el valor de  $K_0$  la incidencia en el estado estacionario entre  $R_0$  y  $\bar{R}_0$  se reduce cada vez más. Por tanto, la representación gráfica de las expresiones analíticas confirman lo observado en la región de valores de  $K_0$  entre ambos umbrales de los diagramas de fase de los paneles (g)-(i) de la Figura 3.2, donde la incidencia de la fase activa es baja, por lo que parece que se crea un umbral epidémico “efectivo”. El panel (b) de la misma figura presenta el impacto de  $\alpha$  en la fracción de infectados en el caso  $K_0 = 0$ . Cuando aumenta el valor de  $\alpha$  el umbral se desplaza hacia valores más altos del número reproductivo básico y la fracción de infectados  $\rho^*$  se reduce.

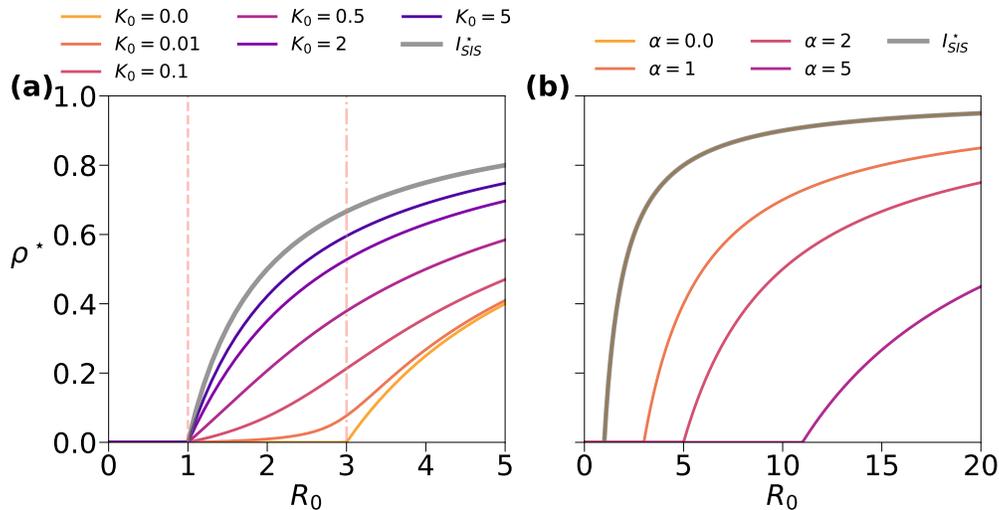


Figura 3.3: **Evolución de la infección en el estado estacionario.** El panel (a) representa las curvas de fracción de infectados en el estado estacionario dadas por la Ecuación 3.3.8 para varios valores de  $K_0$ , junto a los dos umbrales epidémicos  $\beta_c$  y  $\bar{\beta}_c$  representados con líneas discontinuas. Los parámetros empleados son  $mu_0 = 1$ ,  $\Upsilon = 0.5$  y  $alpha = 1$ . El panel (b) muestra la fracción de infectados al final del brote epidémico para varios valores de  $\alpha$ . Los parámetros empleados son  $mu_0 = 1$ ,  $\Upsilon = 0.5$  y  $K_0 = 0$ . En ambos paneles, (a) y (b), la curva gris representa la fracción de infectados en el estado estacionario para el modelo SIS.

### 3.4. Dinámica del modelo SIMS en ausencia de recuperación

Una vez comprendida la influencia en la dinámica de la adquisición de inmunidad, en esta sección pretendemos analizar de manera aislada la dinámica de mutaciones. Esto permite estudiar la mutación desde el punto de vista de un proceso de difusión entre los  $M$  nodos de la red de genotipos. El segundo autovalor de la matriz laplaciana de una red es un indicador de la velocidad de convergencia de los procesos en la red. Por ello, en el panel (a) de la Figura 3.4 se presenta el segundo autovalor de la matriz laplaciana  $\lambda_2$  frente al valor medio del grado de la red  $\langle k \rangle$  para distintos modelos de redes sintéticas.

Las redes regulares de tipo *lattice* exhiben una dependencia directa entre la rapidez del proceso de difusión y la dimensionalidad de la *lattice*. Esto sucede debido que, al aumentar la dimensionalidad de estas redes regulares, aumentará también la conectividad de cada nodo, y en consecuencia aumentará  $\lambda_2$ . En el panel (b) presentamos la evolución temporal de cada una de las cepas de la red regular en una dimensión. Al estar cada nodo conectado únicamente a sus vecinos adyacentes en la cadena, su conectividad es  $\langle k \rangle = 2$ , lo que se traduce en que la transmisión a todas las cepas será muy lenta, tal y como se observa que para que esto se produzca se tiene que llegar al orden de las decenas de millar. En cuanto a la conectividad promedio de las redes aleatorias, se ha decidido imponer la misma que la red regular en tres dimensiones de estructura cúbica simple, es decir,  $\langle k \rangle = 6$ . Tal y como se muestra en el panel (a), la red más óptima para la difusión es la red RRG porque todos los nodos tienen la misma conectividad  $\langle k \rangle$  y la difusión se puede hacer de manera equitativa desde cualquier nodo. Las redes BA y ER, la red ER difunden peor la información entre sus vecinos, ya que, al no ser regulares, existen nodos con poca conectividad. No obstante, en las redes BA hay nodos con mucha conectividad (*hubs*) que facilitan la difusión a los nodos poco conectados. El panel

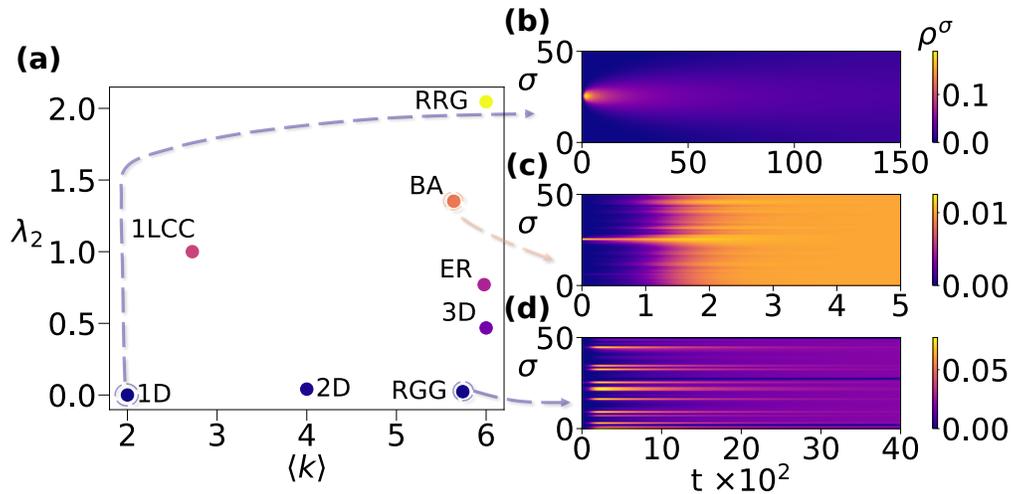


Figura 3.4: **Velocidad de convergencia de los procesos para diferentes redes.** El panel (a) representa el segundo autovalor de la matriz laplaciana  $\lambda_2$  de distintas redes en función del valor medio de la red  $\langle k \rangle$ . Para las redes aleatorias se ha preestablecido  $\langle k \rangle = 6$ , y  $\lambda_2$  de cada tipo se ha determinado a partir de la iteración de 100 redes aleatorias. La red "1LCC" hace referencia a la componente más grande de la red de INFV A (H3N2) presentada en el artículo de B. J. M. Williams et al. [10]. Los paneles (b), (c) y (d) son las evoluciones temporales de la red RRG, cadena lineal y BA respectivamente. Los parámetros empleados en la simulación son  $M = 50$  nodos,  $\beta = 3$  y  $\gamma = 1$ .

(d) de la Figura muestra la rapidez del proceso de difusión en una red BA, en comparación con las otras, ya que se llega al estado estacionario en un orden temporal de las centenas. Se observa que la red más lenta para el proceso de difusión es la red RGG. Esto es debido a que los nodos solamente se conectan con aquellos dentro de su vecindad geométrica. En consecuencia, la información debe recorrer el espacio poco a poco. Finalmente, la componente más grande de la red real de INFV A (H3N2) [10] presenta un conectividad algebraica  $\lambda_2$  parecida a la red BA, pero con una conectividad promedio mucho menor.

## 4. Dinámica del modelo SIMS en redes de genotipos

### 4.1. Propagación y mutación de patógenos como procesos co-evolutivos

Una vez comprendidos los elementos del modelo de forma aislada, queremos comprender el efecto en la dinámica de todos los ingredientes. Permitiendo tanto el proceso de mutación,  $\gamma \neq 0$ , como el proceso de recuperación,  $\mu \neq 0$ , el objetivo es entender tanto la dinámica temporal como la estacionaria del sistema.

Primero, estudiamos la competición entre las escalas temporales del proceso de contagio y el de mutación. Para ello usamos el coeficiente de correlación de Spearman  $r_s$ , que es una medida de la correlación entre dos variables aleatorias  $X$  e  $Y$  que se cuantifica como

$$r_s = \frac{\text{cov}(R(X), R(Y))}{\sigma(R(X))\sigma(R(Y))}, \quad (4.1.1)$$

donde  $R(X)$  y  $R(Y)$  es el orden de las variables  $X$  e  $Y$  respectivamente,  $\text{cov}(R(X), R(Y))$  la cova-

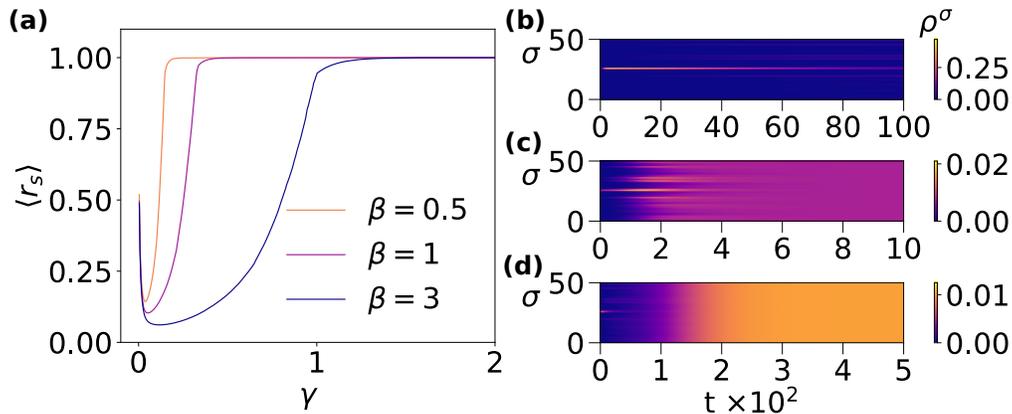


Figura 4.1: **Correlación promedio de la fracción de infecciosos de las cepas de una red RRG con  $\langle \mathbf{k} \rangle = 6$ , junto a las evoluciones temporales.** El panel (a) se corresponde con la correlación promedio de la evolución temporal de las cepas a través del método de Spearman para una red RRG frente a la tasa de mutación  $\gamma$  para distintas tasas de infectividad. Los parámetros fijos empleados en la gráfica (a) son  $\alpha = 1$ ,  $\mu_0 = 0.065$ ,  $K_0 = 1$ ,  $\Upsilon = 10$ . Evolución temporal de los  $M = 50$  nodos de una red RRG. Los paneles (b)-(d) muestran la evolución desde un perfil transversal de las trayectorias. En el panel (b) se ha establecido una tasa de mutación  $\gamma = 0.001$ , el panel (c) cuenta con  $\gamma = 0.40$ , y para el panel (d)  $\gamma = 2.0$ . Para los paneles (b)-(d) los parámetros fijos son  $\beta = 3$ ,  $\alpha = 1$ ,  $\mu_0 = 0.065$ ,  $K_0 = 1$ ,  $\Upsilon = 10$ .

rianza de los órdenes de estas variables aleatorias, y  $\sigma(R(X))$  y  $\sigma(R(Y))$  la desviación estándar de los órdenes de  $X$  e  $Y$ .

En nuestro estudio lo que queremos analizar es la correlación entre las trayectorias temporales que siguen la fracción de infectados de las cepas. A través del coeficiente de Spearman, se calcula la correlación de cada pareja posible de trayectorias y, posteriormente, se promedia entre todos los coeficientes obtenidos, lo que nos da  $\langle r_s \rangle$ . Este proceso lo repetimos en un rango de tasas de mutación  $\gamma$ , y así divisamos la diferencia entre el proceso de contagio y el de mutación. Este análisis se elige hacer sobre una red de tipo RRG con  $M = 50$  nodos, es decir, 50 trayectorias temporales y, en consecuencia, 50 variables a tener en cuenta a la hora de calcular la correlación. Se ha elegido una red de tipo RRG porque todos los nodos tienen la misma conectividad, en este caso  $k = 6$ , lo que permite considerar todas las cepas como equivalentes.

En la Figura 4.1.(a) representamos la correlación promedio en función de la tasa de mutación. Se diferencian dos regiones: para valores bajos de la tasa de mutación  $\gamma$  existe poca correlación entre la evolución temporal de las cepas ya las condiciones iniciales del proceso de contagio las diferencia, véase el panel (b). Para valores más altos de la tasa de mutación, como en la situación mostrada en el panel (d), nos encontramos en el régimen predominado por la mutación. En este punto, la difusión a todas las cepas de la red de genotipos es más rápida que el proceso de contagio, por lo que se pierde la influencia de las condiciones iniciales y se produce la evolución de la dinámica pareja en todas las cepas. Además, es notable que cuando  $\gamma = 0$  la ola epidémica únicamente ocurre en la cepa que inicialmente tenía una fracción inicial de infectados no nula, como se muestra en el panel (b). Al ser la única serie temporal con evolución diferente al resto, la correlación repunta hasta valores más elevados.

Asimismo, en el panel (a), se presentan las correlaciones para varias tasas de infectividad  $\beta$ .

Cuando esta tiene valores pequeños, la correlación promedio de las trayectorias temporales se desplaza favoreciendo el proceso de mutación, ya que se necesita una tasa de mutación menor para que se difunda la infectividad en el espacio de mutaciones antes de que despegue la ola epidémica. Por lo contrario, valores altos de  $\beta$  favorecerán el proceso de contagio, retrasando la transición.

#### 4.2. Escape antigénico a través de redes de genotipos

Una vez entendida la relación entre el proceso de contagio y el de mutación, pasamos a estudiar la dinámica emergente de considerar a la vez adquisición de inmunidad y mutación. Esperamos que ocurra escape antigénico, el mecanismo por el que una cepa del virus evade la inmunidad a través de una mutación puntual. Vamos a intentar caracterizar esto en la red de tipo cadena lineal, debido a que todos los nodos tienen solamente dos vecinos directos, por lo que se espera que el escape antigénico se produzca de manera progresiva.

El panel (a) de la Figura 4.2 muestra la evolución temporal de los nodos de la red de cadena lineal con  $M = 25$  nodos. La cepa semilla es el nodo número cero. A medida que una cepa acumula incidencia, la inmunidad actúa haciendo incrementar la tasa de recuperación, para reducir la carga viral. Si comparamos la Figura

4.2.(a) con la Figura 4.2.(b), donde se representa la evolución de la tasa de recuperación para cada cepa, se percibe que la tasa de recuperación empieza a subir después de un incremento en la fracción de infectados. Cuando la tasa de recuperación,  $\mu^\sigma$ , aumenta, debería reducirse la incidencia de la cepa  $\rho^\sigma$ . Sin embargo, a medida que la inmunidad se adapta a la cepa, la presión sobre el virus se incrementa. En respuesta a esta presión, la cepa puede sufrir una mutación puntual a uno de sus dos vecinos adyacentes de la red de cadena lineal. Esto significa que la inmunidad desarrollada contra la cepa original no es efectiva contra la cepa mutante, y vuelve a iniciarse el proceso de contagio y la respuesta de la inmunidad. Esto se visualiza mejor en la Figura 4.2.(c) donde se representa tanto la evolución de la fracción de infectados, como la tasa de recuperación para la cepa número

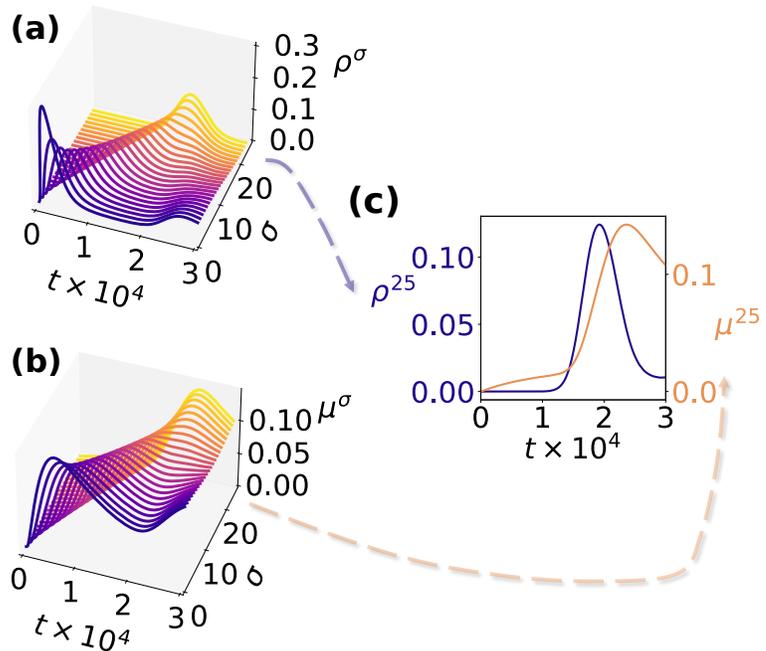


Figura 4.2: **Evolución de temporal de la fracción de infectados y la tasa de recuperación de una red tipo cadena lineal de  $M = 25$  nodos.** El panel (a) representa la evolución temporal de la fracción de infectados de todas las cepas de la red. El panel (b) es la evolución temporal de la tasa de recuperación. El panel (c) es la evolución temporal de la fracción de infectados del nodo número veinticinco de la cadena, junto a la evolución de la tasa de recuperación de dicha cepa. Los parámetros fijos son  $\alpha = 0.01$ ,  $\beta = 1$ ,  $\mu_0 = 0.02$ ,  $K_0 = 0.33$ ,  $\Upsilon = 0.01$  y  $\gamma = 0.1$ .

veinticinco. Se observa que, efectivamente, la recuperación actúa con cierto retraso a la incidencia de la cepa, pero finalmente produce una respuesta y reduce su presencia.

### 4.3. Resurgimiento de variantes en redes de genotipos

La intuición permite anticipar que el modelo SIMS da paso al resurgimiento de cepas aparentemente ya extintas en la población dando lugar a nuevas olas epidémicas. Por tanto, en esta sección queremos caracterizar dicho resurgimiento, encontrando las regiones de parámetros para las que ocurre.

Como queremos examinar este fenómeno en situaciones realistas y contraponerlo con las situaciones aleatorias utilizamos redes sintéticas de genotipos, dado el control que nos permiten ejercer sobre su creación al variar los parámetros del modelo que las genera. La red de la Figura 4.3.(a) está generada a partir de unos parámetros geométrico y de *fitness* más restrictivos,  $\theta_g = 0.01$  y  $\theta_f = 2$  respectivamente. El valor de estos parámetros nos permite generar una red caracterizada por un grado promedio  $\langle k \rangle = 6.03$ , una asortatividad negativa,  $r = -0.15$ , correspondiente a la tendencia de nodos de bajo grado a unirse a nodos de alto grado, y un coeficiente de *clustering*  $C = 0.43$ . Por otro lado, la red de la Figura 4.3.(b) está creada a partir de unos parámetros que favorecen la generación de enlaces,  $\theta_f = 0.3$  y  $\theta_g = 0.2$ . El valor bajo del umbral de *fitness* permite que cepas menos favorecidas biológicamente puedan establecer enlaces. Asimismo, el umbral geométrico más elevado que en el caso anterior conlleva a que nodos más distantes en el espacio geométrico creen enlace. Esta red está caracterizada por un grado promedio  $\langle k \rangle = 26.68$ , una asortatividad positiva  $r = 0.27$  y un y un coeficiente de *clustering*  $C = 0.65$ . Las métricas de estas dos gráficas están en consonancia con lo que se observa en las redes de la Figura 4.3. La conectividad promedio es una diferencia muy notable entre ambas redes, la red (a) apenas muestra conexiones, salvo algunos nodos, mientras que la red (b) está altamente conectada. Asimismo, la tendencia negativa de la asortatividad en la red (a) está presente, dado que se observan los nodos altamente conectados con aquellos que carecen de tantas conexiones.

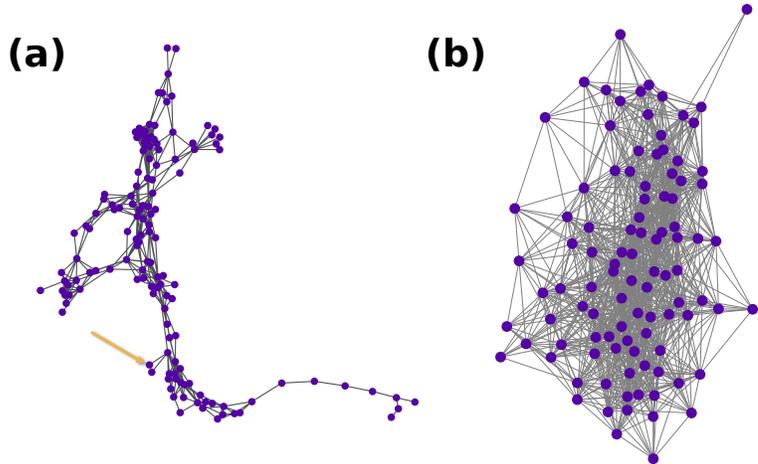


Figura 4.3: **Componentes conexas más grandes de dos redes sintéticas.** El panel (a) utiliza los parámetros son:  $M = 150$ ,  $\sigma = 0.015$ ,  $\theta_f = 2$  y  $\theta_g = 0.01$ . La flecha muestra el nodo usado como semilla inicial para la simulaciones sobre esta red. Los parámetros del panel (b) son:  $M = 150$ ,  $\sigma = 0.015$ ,  $\theta_f = 0.3$  y  $\theta_g = 0.2$ .

Como queremos examinar este fenómeno en situaciones realistas y contraponerlo con las situaciones aleatorias utilizamos redes sintéticas de genotipos, dado el control que nos permiten ejercer sobre su creación al variar los parámetros del modelo que las genera. La red de la Figura 4.3.(a) está generada a partir de unos parámetros geométrico y de *fitness* más restrictivos,  $\theta_g = 0.01$  y  $\theta_f = 2$  respectivamente. El valor de estos parámetros nos permite generar una red caracterizada por un grado promedio  $\langle k \rangle = 6.03$ , una asortatividad negativa,  $r = -0.15$ , correspondiente a la tendencia de nodos de bajo grado a unirse a nodos de alto grado, y un coeficiente de *clustering*  $C = 0.43$ . Por otro lado, la red de la Figura 4.3.(b) está creada a partir de unos parámetros que favorecen la generación de enlaces,  $\theta_f = 0.3$  y  $\theta_g = 0.2$ . El valor bajo del umbral de *fitness* permite que cepas menos favorecidas biológicamente puedan establecer enlaces. Asimismo, el umbral geométrico más elevado que en el caso anterior conlleva a que nodos más distantes en el espacio geométrico creen enlace. Esta red está caracterizada por un grado promedio  $\langle k \rangle = 26.68$ , una asortatividad positiva  $r = 0.27$  y un y un coeficiente de *clustering*  $C = 0.65$ . Las métricas de estas dos gráficas están en consonancia con lo que se observa en las redes de la Figura 4.3. La conectividad promedio es una diferencia muy notable entre ambas redes, la red (a) apenas muestra conexiones, salvo algunos nodos, mientras que la red (b) está altamente conectada. Asimismo, la tendencia negativa de la asortatividad en la red (a) está presente, dado que se observan los nodos altamente conectados con aquellos que carecen de tantas conexiones.

Hemos visto que el escape antigénico es el mecanismo que permite un aumento de la fracción de infectados de una cepa, y su posterior disminución como una lucha entre la infección y la recuperación. Este ciclo de infección, respuesta inmunológica y mutación puede ocurrir una vez por

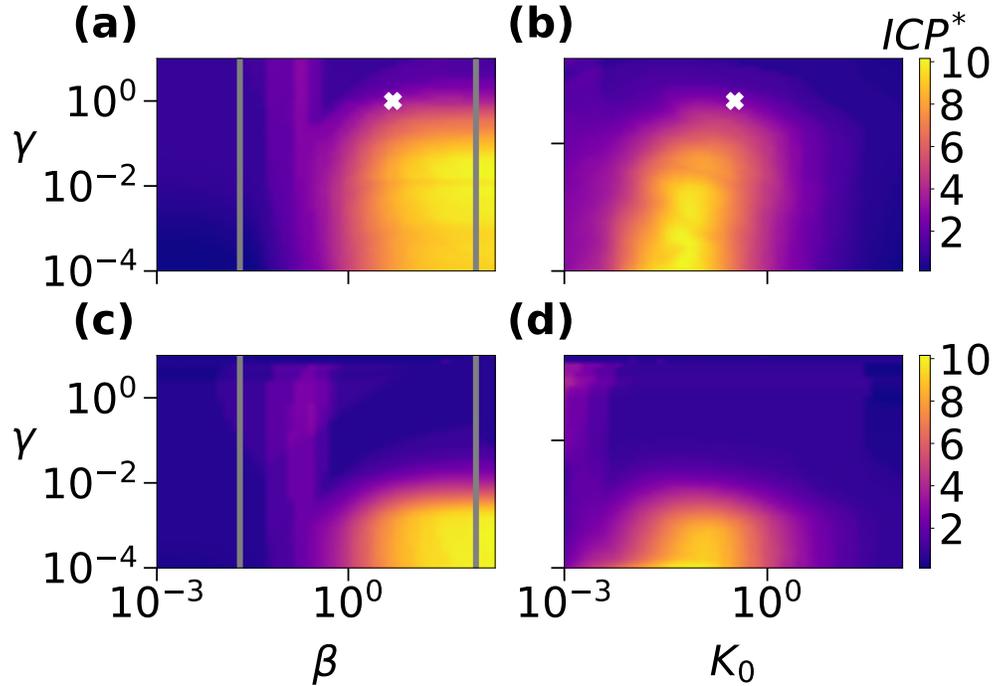


Figura 4.4: Mapas de color para como función de  $\gamma$ ,  $\beta$  y  $K_0$  para resaltar la región de resurgimiento. La gráfica (a) representa el mapa de calor del número de picos encontrados para la red de la Figura 4.3.(a) de la tasa de mutación  $\gamma$  frente a la tasa de infectividad  $\beta$  con parámetros  $\alpha = 1$ ,  $\mu_0 = 0.02$ ,  $K_0 = 0.33$ ,  $\Upsilon = 0.01$ . La gráfica (b) es la representación es de  $\gamma$  frente al parámetro  $K_0$  en una fase activa de resurgimiento de la epidemia fijando  $\alpha = 1$ ,  $\mu_0 = 0.02$ ,  $\beta = 5$ ,  $\Upsilon = 0.01$  con la simulación sobre la misma red de la Figura 4.3.(a). Estas dos gráficas, (a) y (b), muestran los dos umbrales epidémicos correspondientes a  $\beta$  y  $\bar{\beta}_c$ . La gráfica (c) tiene como diferencia con respecto a la gráfica (a) que la red empleada es la mostrada en la Figura 4.3.(b). La gráfica (d) es el equivalente de la gráfica (b) empleando la red de la Figura 4.3.(b). En las gráficas (a) y (b), la cruz indica el conjunto de parámetros que se han utilizado para la evolución temporal de las Figuras 4.5.(a) y (b) respectivamente.

cepa, llevando a un estado endémico donde la prevalencia de las cepas se mantiene estable o, en algunos casos, este ciclo puede repetirse varias veces, permitiendo la circulación continua del virus y la aparición de múltiples olas epidémicas. Para conocer los regímenes de existencia de estas olas se estudia el índice de conteo de picos ( $ICP^*$ ) al variar los parámetros del modelo. Este índice cuenta el número de máximos de estas fluctuaciones en la evolución del patógeno, definido como:

$$ICP^* = \frac{1}{M} \sum_{\sigma=1}^M \max(\rho^\sigma) \quad (4.3.1)$$

donde  $\max(\rho^\sigma)$  es el número de máximos encontrados en la trayectoria temporal de la cepa  $\sigma$ . No obstante, en el proceso de detección de estos máximos se establece un umbral de detección para evitar que se contabilicen variaciones “residuales” del sistema sin información física. Se decide que este umbral se corresponda con la centésima parte del valor del primer pico observado para cada trayectoria.

En las Figuras 4.4.(a) y (c) se representa el  $ICP^*$  en función de  $\gamma$  y  $\beta$ . Se observa que dentro de la fase activa de la epidemia se diferencian dos regiones: para valores bajos de la tasa de mutación

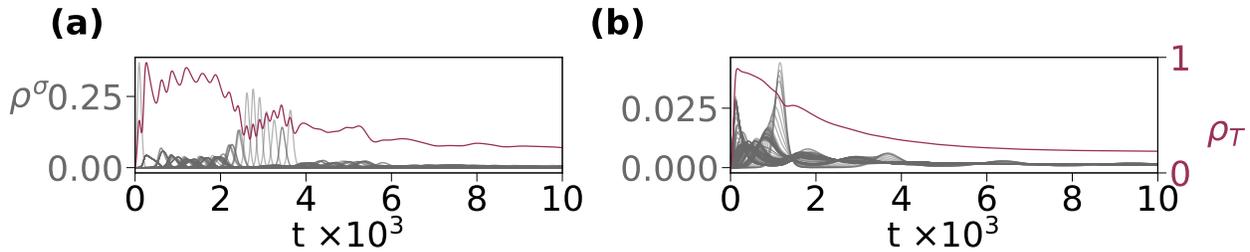


Figura 4.5: **Evolución temporal de los  $M = 150$  nodos de la red sintética de la Figura 4.3.(a) al variar la tasa de mutación.** Ambas simulaciones se han realizado con parámetros  $\alpha = 1$ ,  $\mu_0 = 0.02$ ,  $K_0 = 0.33$ ,  $\beta = 5$  y  $\Upsilon = 0.01$ . La simulación del panel (a) emplea una tasa de mutación  $\gamma = 0.001$ , mientras que en el panel (b) la tasa de mutación es  $\gamma = 1$ .

existe un resurgimiento intenso. Por contra, para valores elevados de la tasa de mutación, no se aprecia resurgimiento. Además, para la red con mayor conectividad promedio, correspondiente a la Figura 4.3.(b) correspondiéndose con la Figura 4.3.(c), la región de parámetros está más limitada que para la red con menor conectividad promedio correspondiente a la Figura 4.3.(a). En cuanto a la variación de  $\gamma$  y  $K_0$ , se vuelve a hacer un barrido de parámetros para la red con menor conectividad, panel (b), y para la red con mayor grado promedio, panel (d). La aparición de estos picos epidémicos aparecen más frecuentemente al disminuir el valor de  $\gamma$ , tanto en la red del panel (b), como en la red del panel (d). Asimismo, en ambas redes se observa una región de parámetros en un entorno localizado de  $K_0$ , y conforme nos alejamos de este valor central, la cantidad de picos epidémicos disminuye. Al comparar los paneles (b) y (d) de la Figura, se vuelve a observar que el fenómeno del resurgimiento está más limitado hacia valores más pequeños de  $\gamma$  para la red con mayor conectividad.

Una vez conocida la región de parámetros que permite hallarnos en fase de resurgimiento, es interesante situarnos en una de las zonas para ver la evolución temporal al variar el parámetro  $\gamma$ . Las Figuras 4.5.(a) y (b) reproducen la evolución temporal de las cepas de la red de la Figura 4.3.(a) para los parámetros  $\alpha = 1$ ,  $\mu_0 = 0.02$ ,  $K_0 = 0.03$ ,  $\beta = 5$  y  $\Upsilon = 0.01$ , junto a la suma de la fracción de infectados de todas las cepas  $\rho_T = \sum_{\sigma=1}^M \rho^\sigma(t)$ . En particular el panel (a) muestra la evolución para un valor de  $\gamma = 0.001$ , y el panel (b) para  $\gamma = 1$ .

En la Figura 4.5.(a) nos encontramos en la fase de contagio, y se observa un pico muy elevado nada más empezar la dinámica. A medida que avanza el tiempo, se observa que la evolución de las cepas presenta múltiples secuencias de picos desplazándose en un “equilibrio” dinámico, a la vez que su amplitud decae. En esta situación, como la tasa de mutación es baja, da tiempo a que se pierda la inmunidad adquirida hasta que la cepa vuelva a tener prevalencia. En contraste, cuando  $\gamma$  es mayor, como en la Figura 4.5.(b), la dinámica se halla en una fase de mutación en la que la evolución de las cepas presenta una disminución rápida, mientras se absorben en un *damping* en el estado estacionario. Esto indica que, con una alta tasa de mutación, la recuperación se enfrenta constantemente a nuevas cepas, creando una memoria inmunológica asociada a una rápida respuesta de la tasa de recuperación frente a infecciones. En el Apéndice B presentamos la versión de esta Figura en escala logarítmica.

#### 4.4. Dinámica del modelo SIMS en redes de genotipos reales

Una vez caracterizada la dinámica del modelo en las redes sintéticas, estudiamos la dinámica en la componente más grande de la red real de genotipos de INFV A (H3N2) [10]. Para ello, en la Figura 4.6 se presenta la evolución de 150 de los 1000 nodos de la red.

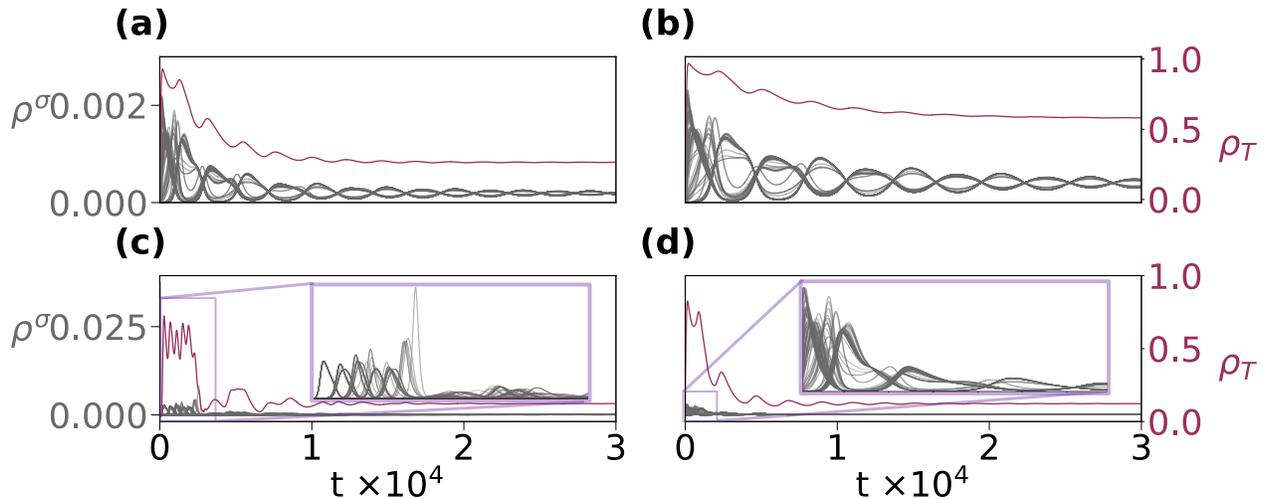


Figura 4.6: **Evolución temporal de la incidencia de 150 nodos de la componente más grande de la red real del INFV A (H3N2) [10], junto a la suma aportada por todas las cepas.** Los paneles (a) y (b) comparten los parámetros  $\alpha = 1$ ,  $\beta = 3$ ,  $\mu_0 = 0.05$ ,  $\gamma = 1$  y  $\Upsilon = 0.01$ . La diferencia es que el panel (a) utiliza  $K_0 = 0.09$ , mientras que el panel (b)  $K_0 = 0.33$ . Los paneles (c) y (d) comparten los parámetros  $\alpha = 100$ ,  $\beta = 3$ ,  $\mu_0 = 0.05$ ,  $K_0 = 0.33$ ,  $\Upsilon = 0.01$ , y su diferencia es que en el panel (c) la tasa de mutación es  $\gamma = 0.001$ , y en el panel (d)  $\gamma = 1$ . Ambas incluyen un *insight* de la zona de interés.

Las Figuras 4.6.(a) y (b) muestran la dinámica en la red para distintos valores del parámetro  $K_0$ . En particular, el panel (a) de la Figura emplea  $K_0 = 0.09$ , y el panel (b)  $K_0 = 0.33$ . En el caso de esta última gráfica, un valor más alto de  $K_0$  implica que la capacidad de prevenir futuras infecciones está más limitada, por lo que la memoria inmunológica adquirida se pierde con mayor facilidad y las infecciones serán más pronunciadas hasta que la tasa de recuperación sea lo suficientemente fuerte como para ejercer una presión sobre el virus y este escape a través de una mutación puntual. Además, en el caso en que  $K_0$  es mayor, la incidencia en el estado estacionario será superior.

En cuanto a las Figuras 4.6.(c) y (d), hemos establecido la situación  $\alpha \gg \Upsilon$  con el fin de contraer la escala temporal y que la cantidad de olas que aparecen sea mayor. La explicación de que la combinación de valores altos de  $\alpha$ , y bajos de  $\Upsilon$  favorezcan la aparición de olas es que en esta situación la ganancia de la inmunidad se acelera rápidamente, lo que, a su vez, favorece el escape antigénico hacia otras cepas. En el Apéndice C se presentan las gráficas con la diferencia entre situaciones marcadas por  $\alpha \gg \Upsilon$  y  $\alpha \ll \Upsilon$ . La Figura 4.6.(c) se ha generado con una tasa de mutación muy baja,  $\gamma = 0.001$ , mientras que la Figura 4.6.(d) se ha generado con una tasa de mutación  $\gamma = 1$ . Para el panel (c) de la Figura la amplitud de las olas es mayor que para las observadas en el panel (d). Esto es debido a que con una tasa de mutación pequeña, el virus no puede mutar fácilmente, lo que se traduce en que su incidencia será mayor antes de que se produzca una mutación puntual. Como estas olas tienen una amplitud pequeña y se producen en un periodo

relativamente corto de tiempo, se presenta un *insight* de la zona temporal para el conjunto de olas de mayor amplitud que aparecen. Además, en el Apéndice C también se presenta la versión logarítmica de esta Figura.

## 5. Conclusiones

En este trabajo hemos estudiado la dinámica epidemiológica con mutación de patógenos, integrando ambos procesos co-evolutivos en un modelo compartimental que incluye las diferentes variantes de un mismo patógeno.

Primero, hemos realizado simulaciones en redes simples para descubrir las diferentes regiones del retrato de fases. Para lograr un mejor entendimiento, hemos obtenido la expresión analítica que separa la fase absorbente de la epidemia de la fase activa. Además, dentro de la fase activa de la epidemia existen regiones en las que las infecciones existentes se corresponden con unas pocas cepas, y otras en las que intervienen muchas cepas.

Para lograr un mejor entendimiento de la casuística observada, hemos caracterizado la dinámica del modelo SIMS de manera detallada, analizando cada uno de sus componentes de forma aislada. Primero, hemos mostrado que dentro de la fase activa de la epidemia, para valores elevados de la trascendencia de la inmunidad, hay una región de inactividades en la que la fracción estacionaria de infectados es muy reducida. En función de la trascendencia de la inmunidad, el modelo SIMS presenta una transición desde un comportamiento similar al del modelo SIRI, donde se proporciona una protección temporal contra la reinfección, a otro compatible con el modelo SIS cuando reducimos el efecto de la trascendencia de la inmunidad. Por otro lado, hemos constatado la relación entre la topología de las redes y la velocidad del proceso difusivo en el espacio de mutaciones, tanto en redes regulares como en redes aleatorias. En redes regulares, la velocidad de difusión está ligada a la dimensionalidad de la red, mientras que en estructuras aleatorias, la velocidad varía de acuerdo a la estructura. En particular, la existencia de nodos más aislados, ya sea por restricciones geométricas o por tener baja conectividad, dificultan el proceso de difusión.

Considerando el modelo completo, hemos estudiado como la variación de la tasa de mutación permite distinguir entre una fase dominada por el contagio y otra dominada por la difusión a través de la red de mutaciones. Asimismo, hemos caracterizado el escape antigénico, mostrando la difusión de la ola epidémica a través de diferentes cepas adyacentes conforme aumenta la respuesta inmune. Finalmente, hemos caracterizado el resurgimiento de las cepas a través de un índice de conteo de picos epidémicos, mostrando que dentro de la fase activa de la epidemia existe una zona de resurgimiento para valores bajos de la tasa de mutación. Para completar el trabajo, hemos comprobado este resurgimiento en las trayectorias temporales tanto en una red real de genotipos como en un modelo sintético capaz de simular las relaciones entre las distintas cepas en función de su adaptación biológica, o su posición en el espacio de mutaciones.

## Referencias

- [1] J. Gómez-Gardeñes, A. S. De Barros, S. T. R. Pinho, and R. F. S. Andrade, “Abrupt transitions from reinfections in social contagions,” *EPL (Europhysics Letters)*, vol. 110, p. 58006, June 2015.
- [2] A.-L. Barabási and M. Pósfai, *Network science*. Cambridge, United Kingdom: Cambridge University Press, 2016. OCLC: ocn910772793.
- [3] R. Sanjuán and P. Domingo-Calap, “Mechanisms of viral mutation,” *Cellular and Molecular Life Sciences*, vol. 73, pp. 4433–4448, Dec. 2016.
- [4] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: principles, methods and applications*. Cambridge, United Kingdom: Cambridge University Press, 2017. OCLC: 1004620089.
- [5] M. Fiedler, “Algebraic Connectivity of Graphs,” *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.
- [6] R. Olfati-Saber and R. Murray, “Consensus Problems in Networks of Agents With Switching Topology and Time-Delays,” *IEEE Transactions on Automatic Control*, vol. 49, pp. 1520–1533, Sept. 2004.
- [7] A. Bertrand and M. Moonen, “Distributed computation of the Fiedler vector with application to topology inference in ad hoc networks,” *Signal Processing*, vol. 93, pp. 1106–1117, May 2013.
- [8] A.-L. Barabási, R. Albert, and H. Jeong, “Scale-free characteristics of random networks: the topology of the world-wide web,” *Physica A: Statistical Mechanics and its Applications*, vol. 281, pp. 69–77, June 2000.
- [9] B. J. M. Williams, G. St-Onge, and L. Hébert-Dufresne, “Localization, epidemic transitions, and unpredictability of multistrain epidemics with an underlying genotype network,” *PLOS Computational Biology*, vol. 17, p. e1008606, Feb. 2021.
- [10] B. J. M. Williams, C. B. Ogbunugafor, B. M. Althouse, and L. Hébert-Dufresne, “Immunity-induced criticality of the genotype network of influenza A (H3N2) hemagglutinin,” *PNAS Nexus*, vol. 1, p. pgac143, Sept. 2022.
- [11] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, “Scale-Free Networks from Varying Vertex Intrinsic Fitness,” *Physical Review Letters*, vol. 89, p. 258702, Dec. 2002.
- [12] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, Sept. 2006. Google-Books-ID: YXrIRDuAbE0C.

# Apéndices

## A. Colección de parámetros y siglas

Aquí presentamos una recopilación de los parámetros y las siglas mencionadas durante el trabajo.

- $\mathcal{G}$ : Grafo
- $\mathcal{N}$ : Conjunto de nodos
- $N$ : Cardinal del conjunto de nodos.
- $M$ : Número de genotipos total de una red.
- $\mathcal{L}$ : Conjunto de enlaces.
- $\mathbf{A}$ : Matriz de adyacencia asociada a un grafo no pesado.
- $k$ : Grado de un nodo en un grafo no dirigido.
- $P(k)$ : Distribución de grado de los nodos de un grafo.
- $P(K \geq k)$ : Distribución acumulativa complementaria de grado.
- $C_i$ : Coeficiente de *clustering* de un nodo  $i$ .
- $r$ : Coeficiente de asortatividad.
- $\mathbf{L}$ : Matriz laplaciana asociada a un grafo.
- $\lambda_2$ : Conectividad algebraica de una red.
- 1D: Red regular *lattice* en una dimensión.
- 2D: Red regular *lattice* en dos dimensiones.
- 3D: Red regular *lattice* en tres dimensiones.
- ER: Red Erdős-Rényi.
- BA: Red Barabási-Albert.
- RRG: Red *Random Regular Graph*.
- RGG: Red *Random Geometric Graph*.
- 1LCC: Componente más grande de la red real de INFV A (H3N2) [10].
- $W_M$ : Número de cepas silvestres de una red de genotipos sintética.
- $\theta_g$ : Umbral geométrico de una red de genotipos sintética.

- $S$ : Susceptibles.
- $\rho^\sigma$ : Fracción de infectados por la cepa  $\sigma$ .
- $\rho^*$ : Fracción de infectados total en el estado estacionario,  $\lim_{t \rightarrow \infty} \sum_{\sigma=1}^M \rho^\sigma(t)$ .
- $\rho_T$ : Suma de todas las cepas a tiempo  $t$ ,  $\sum_{\sigma=1}^M \rho^\sigma(t)$ .
- $\mu^\sigma$ : Tasa de recuperación asociada a la cepa  $\sigma$ .
- $\Gamma^{\sigma\nu}$ : Elemento  $(\sigma, \nu)$  de la matriz de adyacencia de la red de genotipos.
- $\alpha$ : Coeficiente de ganancia de inmunidad.
- $K_0^{-1}$ : Transcendencia de la inmunidad.
- $\Upsilon$ : Coeficiente de pérdida de inmunidad.
- $\mu_0$ : Tasa de recuperación basal.
- $n_{eff}^*$ : Número de Kish, definido como la tasa efectiva de participación.
- $R_0$ : Número reproductivo básico.
- $\bar{R}_0$ : Segundo número reproductivo básico al imponer  $K_0 = 0$ .
- $\beta_c$ : Umbral epidémico asociado al número reproductivo básico  $R_0$ .
- $\bar{\beta}_c$ : Umbral epidémico asociado al número reproductivo básico  $\bar{R}_0$ .
- $r_s$ : Coeficiente de correlación de Spearman.
- $ICP$ : Índice de conteo de picos.

## B. Complementos de la dinámica del modelo SIMS en redes de genotipos sintéticas

Presentamos el correspondiente a la Figura 4.5 en escala logarítmica para apreciar mejor los cambios en la evolución temporal.

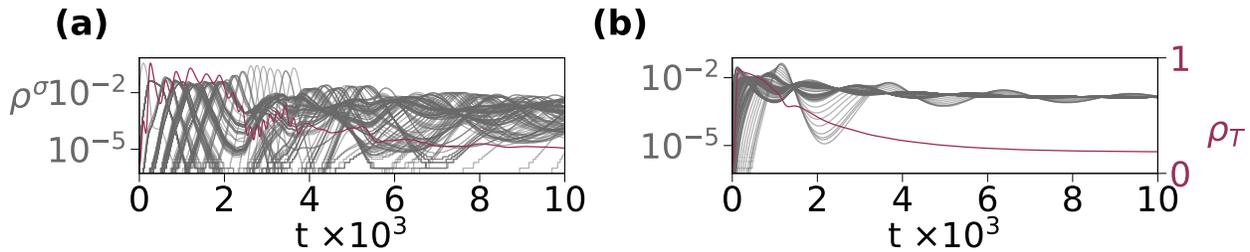


Figura B.1: **Evolución temporal en escala logarítmica de los  $M = 150$  nodos de la red sintética de la Figura 4.3.(a) al variar la tasa de mutación.** Ambas simulaciones se han realizado con parámetros  $\alpha = 1$ ,  $\mu_0 = 0.02$ ,  $K_0 = 0.03$ ,  $\beta = 5$  y  $\Upsilon = 0.01$ . La simulación del panel (a) emplea una tasa de mutación  $\gamma = 0.001$ , mientras que en el panel (b) la tasa de mutación es  $\gamma = 1$ .

## C. Complementos de la dinámica del modelo SIMS en redes de genotipos reales

En las Figuras C.1.(a) y (b) se presenta la evolución temporal de la casuística  $\alpha \gg \Upsilon$  y  $\alpha \ll \Upsilon$  respectivamente en escala logarítmica para 150 nodos de los 1440 de la red. En el panel (a) se aprecia el carácter oscilatorio de las distintas cepas mientras se van absorbiendo. En el caso contrario, el panel (b) nos muestra cómo la prevalencia de la pérdida de inmunidad imposibilita la existencia de unas olas tan marcadas y casi inmediatamente la totalidad de los individuos se vuelve infecciosa.

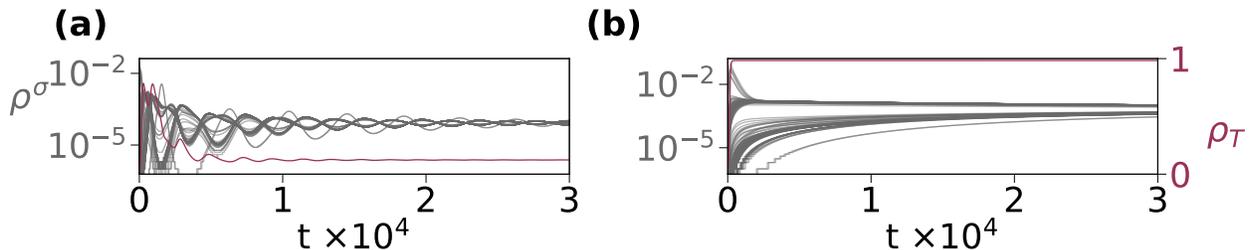


Figura C.1: **Evolución temporal de la incidencia de 150 nodos de la componente más grande de la red real del INFV A (H3N2) [10] en escala logarítmica, junto a la suma aportada por todas las cepas.** El panel (a) se caracteriza por  $\alpha = 100$  y  $\Upsilon = 0.01$ , mientras que el panel (b) se caracteriza por  $\alpha = 0.01$  y  $\Upsilon = 100$ . Los parámetros fijos en ambas gráficas son  $\beta = 3$   $\mu_0 = 0.05$ ,  $K_0 = 0.33$  y  $\gamma = 1$ .

A continuación se presenta la versión logarítmica de la Figura 4.6.

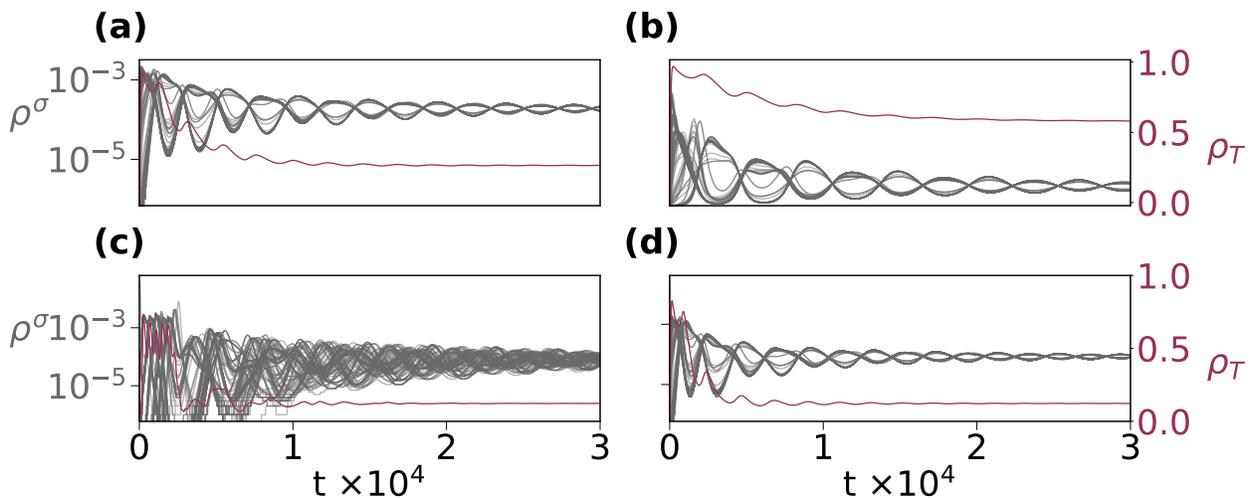


Figura C.2: **Evolución temporal de la incidencia de 150 nodos de la componente más grande de la red real del INFV A (H3N2) [10], junto a la suma aportada por todas las cepas.** Los paneles (a) y (b) comparten los parámetros  $\alpha = 1$ ,  $\beta = 3$   $\mu_0 = 0.05$ ,  $\gamma = 1$  y  $\Upsilon = 0.01$ . La diferencia es que el panel (a) utiliza  $K_0 = 0.09$ , mientras que el panel (b)  $K_0 = 0.33$ . Los paneles (c) y (d) comparten los parámetros  $\alpha = 100$ ,  $\beta = 3$   $\mu_0 = 0.05$ ,  $K_0 = 0.33$ ,  $\Upsilon = 0.01$ , y su diferencia es que en el panel (c) la tasa de mutación es  $\gamma = 0.001$ , y en el panel (d)  $\gamma = 1$ . Ambas incluyen un *insight* de la zona de interés.