



DEPARTAMENTO DE FÍSICA DE LA MATERIA CONDENSADA

TRABAJO DE FIN DE GRADO

Procesos de Contagio en Metapoblaciones y Modelos de Movilidad

Autor
PABLO VALGAÑÓN RUIZ

Supervisores
DR. JESÚS GÓMEZ GARDEÑES
DAVID SORIANO PAÑOS

14 de julio de 2020

Índice

1. Introducción	2
2. Marco teórico	4
2.1. Modelo SIS	4
2.2. Metapoblaciones y redes complejas	5
2.3. La matriz Origen-Destino	6
2.4. Indistinguibilidad y distinguibilidad	6
2.4.1. Individuos Indistinguibles	6
2.4.2. Individuos Distinguibles	7
3. Ecuaciones markovianas para individuos distinguibles	8
3.1. Obtención de las ecuaciones	8
3.2. Verificación de las ecuaciones	9
4. Estudio del umbral epidémico	10
4.1. Observación cualitativa. Variables λ , p_d	10
4.2. Linealización de las ecuaciones. Cálculo del umbral epidémico	11
4.3. Comparando los formalismos. Detrimento de la propagación en $p_d = 1$	14
5. Impacto de la epidemia para alta movilidad	16
6. Conclusiones	20
7. Anexos	23
7.1. Anexo I: Recuperación de las ecuaciones del formalismo de individuos indistinguibles	23

1. Introducción

De forma general, un sistema es un conjunto de elementos interactuantes que forman parte de un todo [1]. Sin embargo, para poder considerar que un sistema es complejo además de constar de un número grande de agentes interactuantes debe exhibir un comportamiento emergente, el cual no podemos explicar directamente a partir del comportamiento de los componentes individuales y además no está dirigido por un controlador central. Esto implica que necesitamos herramientas que tengan en cuenta las interacciones entre dichos componentes (o agentes) para observar estas propiedades macroscópicas que no son evidentes en un principio. Este es el campo de estudio de la Física de Sistemas Complejos. Sus aplicaciones abarcan áreas que no han tenido relación con la Física tradicionalmente, pero se pone en evidencia la necesidad de herramientas como la Física Estadística para abordarlas. Entre los usos para los que estas herramientas son de gran utilidad encontramos el estudio de sincronización [2], congestiones [3], epidemias [4], rumores [5], economía [6], etc.

Del interés que surge de explicar ciertos sistemas complejos a partir de las conexiones entre sus agentes nace la Física de Redes Complejas. Esto se debe a que dichas conexiones en muchas ocasiones siguen una estructura compleja y el acercamiento por teorías como campo medio, que consideran equivalentes a todos los agentes, no son suficientes y se necesita caracterizar cada agente individualmente. En el mundo real tenemos muchos ejemplos de redes complejas, especialmente en relación con Internet. Los enlaces entre páginas web o relaciones de amistad en redes sociales son los más evidentes, pero también se pueden encontrar en muchas otras disciplinas (biología, economía, etc). El caso que nos incumbe en este momento, y que será la base del trabajo, es la movilidad de personas entre distintas localizaciones. Al tratar de resolver problemas relacionados con el movimiento de individuos trataremos estos sistemas como redes complejas, y podremos obtener resultados que de otra forma no sería posible.

El objeto de estudio del trabajo son los fenómenos de contagio de enfermedades, y propagación de epidemias. Durante las últimas décadas ha empezado a hacerse evidente la necesidad de aumentar los esfuerzos en la investigación de contagios sociales debido a diversas enfermedades que han surgido y amenazan la salud pública a escala mundial. La razón más importante es el aumento de los viajes de larga distancia que realizan miles de personas diariamente, que hace muy difícil que las epidemias se mantengan aisladas y localizadas en su lugar de origen. Esto posibilita que cualquier enfermedad que surja en una parte del mundo llegue a lugares remotos en cuestión de muy poco tiempo. Casos como el SIDA, la gripe A o el ébola han servido como recordatorio de que es necesario elaborar formas de contención de una pandemia seria en caso de que se salga de control.

Finalmente, este año la llegada del COVID-19 ha asestado un duro golpe al mundo entero, poniendo en evidencia la falta de preparación a escala global para detener enfermedades extremadamente contagiosas. Con todos los problemas y las muertes que ha causado, gran cantidad de investigadores trabajan no solamente para obtener medicamentos o vacunas, sino para elaborar políticas de prevención que frenen la transmisión del virus y limiten su efecto. Es aquí donde entra en juego la modelización de contagio de enfermedades. Conocer y predecir la evolución de una epidemia es vital para lidiar con ella, a la vez que estudiar el efecto que distintas políticas de movilidad pueden tener a la hora de limitar su impacto.

Aquí podemos introducir la importancia de tratar la movilidad humana con detenimiento, ya que las propiedades que surjan de los modelos de contagio dependerán directamente de cómo tratamos estos viajes y contactos entre personas. Recientemente se ha observado un gran crecimiento en la cantidad de publicaciones científicas acerca de la movilidad humana, y de cómo puede esto afectar a muchos otros aspectos de la vida cotidiana. Varios de los motivos que

impulsaron éste área de investigación [7] son la predicción de congestión de tráfico [8], planificación urbana [9], seguridad interna [10] y modelos epidémicos [11, 12, 13], entre otros.

Además el aumento de la capacidad de computación y la resolución de la observación de la movilidad humana (seguimiento de recorridos de billetes individuales [14], antenas telefónicas, GPS [15], redes sociales, etc) permiten simular sistemas más sofisticados y compararlos con casos reales para obtener resultados cada vez más realistas.

La modelización de estos procesos de contagio ha sido llevada a cabo desde hace mucho tiempo, pero muchos modelos no pueden replicar comportamientos que emergen de las conexiones entre agentes. Los primeros intentos usan herramientas como la aproximación de campo medio, que asume que la población es homogénea y todos los individuos tienen contactos con el resto de la misma manera. Como es un sistema muy limitado, el siguiente paso consiste en tener en cuenta que los individuos son diferentes y tienen distintas conexiones entre ellos. Se puede considerar una red compleja, en la que los nodos son individuos, pero su limitación principal es que los contactos ocurren siempre entre los mismos agentes, por lo que perdemos muchas propiedades.

Nosotros utilizaremos metapoblaciones, en las que los nodos de la red ahora pasarán a ser subpoblaciones de individuos, en las que a la hora de realizar los contactos usamos teoría de campo medio. Sin embargo, en cada paso cada individuo se quedará en su nodo o se desplazará a otro distinto. Al principio, estos desplazamientos se consideraban caminatas aleatorias a lo largo de toda la red, pero en nuestro caso cada persona volverá a su nodo de origen después de cada paso temporal. La cantidad de gente que viaja de un nodo a otro serán las conexiones (links) de la red.

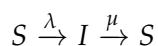
Tanto los fenómenos de contagio como de movilidad son procesos estocásticos. Es decir, a cada paso temporal hay una probabilidad de que dos personas en contacto se transmitan una enfermedad, de que una persona viaje o permanezca en su nodo, y de que elija un nodo u otro al que viajar. Estos procesos se conocen como procesos de reacción-difusión. Por lo tanto, la primera forma de trabajar con este modelo será la simulación de agentes individuales, o de tipo Montecarlo. En ellas tenemos en cuenta el estado de cada individuo y su comportamiento. Sin embargo, podemos hacer un Acercamiento Macroscópico de Cadena de Markov (MMCA), en el que asumimos que cada estado depende únicamente del anterior y podemos hallar unas ecuaciones que describan la evolución del sistema a nivel mesoscópico (es decir, nos dará el porcentaje de infectados en cada nodo sin tener en cuenta los agentes individualmente). Este método hace que se pierdan las propiedades estocásticas y disminuye mucho el tiempo de simulación, además de que nos permite estudiar de forma analítica otras características del sistema.

Introduciremos un nuevo modelo de movilidad y contagio dentro del marco existente. Obtendremos las ecuaciones markovianas y las pondremos a prueba frente a simulaciones de agentes individuales para verificarlas. Después de analizar los resultados estudiaremos el umbral epidémico (que definiremos más adelante) linealizando las ecuaciones. Nos centraremos en comparar nuestro modelo (al que nos referiremos como modelo de individuos distinguibles) con el propuesto en [16] (de individuos indistinguibles). Trataremos de explicar las diferencias entre ambos y cómo cada uno se adapta a distintos tipos de datos.

2. Marco teórico

2.1. Modelo SIS

El modelo SIS (Susceptible - Infectado - Susceptible) [4] propone un proceso en el cual una persona que se encuentra en un estado susceptible, al estar en contacto con una persona infectada tiene una probabilidad λ de infectarse. Al mismo tiempo, una persona infectada tiene una probabilidad μ de recuperarse y volver a ser susceptible. Una forma esquemática de representar esta dinámica es la siguiente:



Al escribir las ecuaciones las variables están en función del tiempo, el cual consideraremos como discreto. Si nos encontramos a tiempo t , diremos que el siguiente paso temporal será $t + 1$. Por lo tanto, las probabilidades que hemos descrito se definen para un paso temporal. Vamos a dejar claras varias definiciones que usaremos a lo largo del trabajo:

- $\lambda \equiv$ Probabilidad de que una persona susceptible se infecte debido al contacto con una persona infectada.
- $\mu \equiv$ Probabilidad de que un individuo infectado vuelva a ser susceptible.
- $\rho(t) \equiv$ Probabilidad de que una persona en una población se encuentre infectada. Es equivalente a la fracción de personas infectadas respecto al total.
- $n \equiv$ Cantidad de individuos que residen en una población.

Cuando hacemos una aproximación de campo medio, consideramos que todos los individuos son equivalentes, y cada uno tendrá la misma cantidad de contactos que el resto. Por simplicidad, suponemos que en una población cada agente contacta con todos los demás una vez en cada paso de tiempo. Cada persona con la que contacta tiene una probabilidad $\rho(t)$ de estar infectada. Si en caso de que lo esté la probabilidad de que al tener contacto se contagie la primera es λ , nos queda una probabilidad $\lambda \cdot \rho(t)$ de contagio con cada contacto.

Para saber la probabilidad de contagiarte al tener contacto con todos los agentes hallamos primero la de no infectarse con ninguno. Como repetimos el contacto n veces, esa probabilidad queda como $(1 - \lambda \cdot \rho(t))^n$. Se lo restamos a 1 para saber finalmente la probabilidad de infectarse:

$$P(t) = 1 - (1 - \lambda \cdot \rho(t))^n$$

Queremos conocer la evolución de la cantidad de individuos en una población aislada. Para ello partimos de un ratio de infectados $\rho(t)$, y queremos conocer $\rho(t + 1)$. La cantidad se conservará respecto a la anterior, salvo dos términos: un porcentaje μ de ellos se recuperará, por lo que dejan de estar infectados. Además, los que son susceptibles se contagiarán con una probabilidad P . Queda entonces:

$$\rho(t + 1) = (1 - \mu)\rho(t) + P(t) \cdot (1 - \rho(t))$$

Donde P también depende de $\rho(t)$ como hemos visto antes. Este sería un ejemplo de campo medio para una población según el modelo SIS.

En estos casos, la cantidad de personas infectadas va a llegar a un estado estacionario en el que $\rho(t)$ tiende a una constante cuando t tiende a infinito. La forma principal de evaluar el impacto que una epidemia tiene en una población en el modelo SIS es obtener el valor al que tiende $\rho(t)$. En la figura 1 vemos un ejemplo de la forma que tiene el diagrama de fases de este sistema. Hay una transición de fase de segundo orden en λ_c , que llamaremos *umbral epidémico*.

El umbral epidémico es el mínimo valor de λ para el cual la epidemia no se extingue, y puede seguir contagiándose de un individuo a otro. Esto significa que, de media, cada agente que se infecte contagiará a otro. Disminuir el valor de λ hace que la epidemia desaparezca, y aumentarlo hace que el ρ estacionario aumente.

Es sencillo encontrar para este caso el valor del umbral analíticamente a partir de las ecuaciones mencionadas, dando $\lambda_c = \frac{\mu}{n}$. Ese valor nos servirá más adelante como referencia para estudiar cómo cambia el umbral en función de la movilidad.

El modelo SIS simplifica mucho la forma de estudiar y simular estos sistemas, y generalmente se asocia con enfermedades que una persona puede contraer varias veces, es decir, no se desarrolla inmunidad a largo plazo. El modelo que se adapta mejor en caso contrario es el SIR (Susceptible-Infectado-Recuperado), pero no lo trataremos en este trabajo.

Como hemos comentado en la introducción, esto no es suficiente para estudiar la propagación de una enfermedad en una población complicada. Incluso si tratamos a los agentes como los componentes de un sistema complejo, siempre tendrían los mismos contactos entre cada par de individuos, lo cual no es realista. Lo que observamos en el mundo real es que la gente se mueve, viaja, tiene contactos con distintas personas dependiendo de su localización, y por ello debemos tener esta movilidad en cuenta. Para ello seguiremos un modelo de metapoblaciones que introduciremos a continuación.

2.2. Metapoblaciones y redes complejas

Un modelo de metapoblaciones describe los contactos entre individuos a dos escalas distintas. La población total estará dividida en subpoblaciones. En cada subpoblación, los individuos interactúan entre ellos y tienen contacto unos con otros, permitiendo el contagio. A un nivel superior, tenemos los viajes entre subpoblaciones. Tal como se describe en [16], cada individuo tiene asociada una subpoblación a la que siempre retorna. Cada paso temporal tiene dos procesos. Primero uno de difusión (cada individuo decide viajar a un lugar de destino o quedarse) y otro de reacción (ocurren los contactos entre individuos dentro de cada subpoblación y se contagia la enfermedad).

Una red compleja es un grafo compuesto por N nodos y L enlaces (o links), los cuales se forman entre pares de nodos. La forma más utilizada para caracterizar estas redes es la *matriz de adyacencia* \mathbf{A} . Cada componente de esta matriz A_{ij} representa la fuerza del enlace que va del nodo i al j . Si los enlaces son no dirigidas (es decir, la conexión i a j es igual que la de j a i) la matriz de adyacencia es simétrica: $A_{ij} = A_{ji}$. Además, la red es no pesada si su matriz sólo se compone de unos y ceros. Un ejemplo de red no dirigida y no pesada es las relaciones de amistad en una red social. Por otro lado, la cantidad de mensajes de correo electrónico enviados entre usuarios es una red dirigida pesada.

La forma en la que aplicaremos redes complejas a este sistema es asociando cada subpoblación a un nodo de la red, y los enlaces entre nodos representarán la cantidad de personas que viajan de uno a otro. Dentro de cada subpoblación, sin embargo, no usaremos una red que describa a

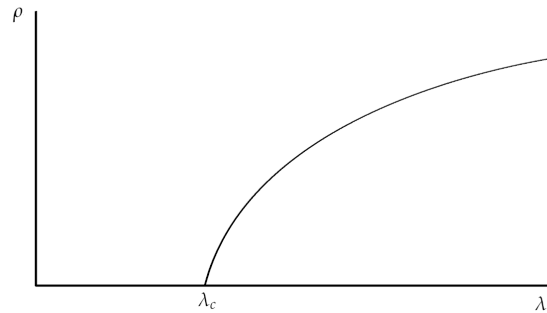


Figura 1: Diagrama de fases cualitativo de una epidemia en la aproximación MMCA, en el que se representa en el eje vertical la fracción de infectados en el estado estacionario ρ y en el eje horizontal la probabilidad de contagio por contacto λ , con λ_c separando ambos estados de la epidemia.

cada individuo en particular porque requeriría de una capacidad de computación desorbitada dada la cantidad de personas que pueden vivir en cada nodo. Usaremos una aproximación de campo medio (todos los individuos tienen un contacto con cada uno de los demás de la misma subpoblación).

2.3. La matriz Origen-Destino

Esta es la matriz que utilizamos para describir nuestra red [7], y por lo tanto está directamente relacionada con la matriz de adyacencia. A partir de esta matriz podemos obtener la cantidad de personas que viajan de un nodo i a otro j . Esta matriz entonces es pesada y dirigida, porque la cantidad de personas que viajan de un nodo i a otro j no tiene por qué ser la misma que las que viajan del j al i .

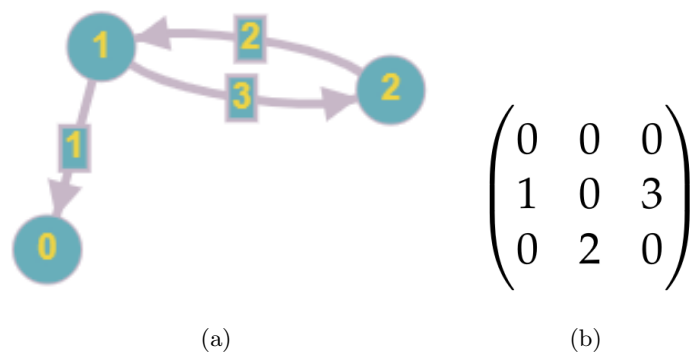


Figura 2: Ejemplo de un grafo (a) y su matriz de adyacencia (b)

Es aquí donde entra en juego la distinguibilidad de los individuos, que es el aspecto más importante de este trabajo.

2.4. Indistinguibilidad y distinguibilidad

Campos como la mecánica cuántica y la física estadística han mostrado en muchas ocasiones que las propiedades de un sistema pueden ser muy diferentes dependiendo de si consideramos sus agentes distinguibles entre ellos o no (paradoja de Gibbs). Esto es algo que al principio parece ser poco intuitivo, y sin embargo se hace muy aparente en este modelo.

2.4.1. Individuos Indistinguibles

En [16] además del campo medio se asume que todos los individuos de cada nodo son indistinguibles. Es decir, a cada nodo i se le asigna una probabilidad de que un individuo esté infectado $\rho_i(t)$, lo que hace que podamos definir el estado mesoscópico como el vector $(\rho_1(t), \dots, \rho_N(t))$, y la cantidad de individuos infectados en cada nodo será $n_i \cdot \rho_i(t)$ (donde n_i es la población de cada nodo).

Si todos los individuos son iguales, a la hora de decidir dónde van a viajar, cada uno tiene una probabilidad de desplazarse a cualquier otra subpoblación. La matriz OD se convierte entonces en una matriz de pesos de probabilidad, donde el elemento R_{ij} es la probabilidad de que un individuo del nodo i decida viajar al nodo j . Cumple:

$$\sum_{j=0}^{N-1} R_{ij} = 1 \quad \forall i \quad (1)$$

Para controlar la cantidad de movimiento entre nodos introducimos un factor global de la probabilidad de desplazamiento p_d . Tras haber elegido cada individuo un destino, p_d es la probabilidad de que decida finalmente viajar o quedarse en su nodo. Gran parte del estudio estará basada en analizar cómo cambia la propagación de la epidemia en función de dicha variable.

Las ecuaciones markovianas de este modelo se introducen en [16], y son las siguientes:

$$n_{j \rightarrow i} = \delta_{ij}(1 - p_d)n_i + p_d R_{ji}n_j \quad (2)$$

$$P_i(t) = 1 - \prod_{j=1}^N (1 - \lambda \rho_j(t))^{n_{j \rightarrow i}} \quad (3)$$

$$\Pi_i(t) = (1 - p_d)P_i(t) + p_d \sum_{j=1}^N R_{ij}P_j(t) \quad (4)$$

$$\rho_i(t+1) = (1 - \mu)\rho_i(t) + (1 - \rho_i(t))\Pi_i(t) \quad (5)$$

Además, se puede obtener una expresión analítica para el umbral epidémico. No vamos a entrar aún en detalle porque después seguiremos unos pasos similares para encontrar dicha expresión para el caso de individuos distinguibles. El umbral es:

$$\lambda_c = \frac{\mu}{\Lambda_{max}(\mathbf{M})} \quad (6)$$

Siendo $\Lambda_{max}(\mathbf{M})$ el autovalor máximo de la matriz:

$$M_{ij} = p(1 - p)(R_{ij} + R_{ji})n_j + (1 - p)^2 n_j \delta_{ji} + p^2 n_j \left(\sum_{l=1}^N R_{il}R_{jl} \right)$$

Este modelo asume que todos los individuos pertenecientes a un mismo nodo son equivalentes, pero esto no es algo que se ajusta a la realidad en la mayor parte de los casos. Cada persona tiene unos patrones propios a la hora de desplazarse. Debido al trabajo, la universidad, ocio u otros motivos una persona suele volver la mayor parte de las veces al mismo sitio, por lo tanto en un mismo nodo las personas que sigan un patrón tendrán diferente probabilidad de infectarse que las que sigan otros. En redes complejas, uno de los aspectos que más consecuencias tiene habitualmente en los resultados es la heterogeneidad de esa red. Al hacer las asunciones de este modelo estamos considerando un sistema mucho más homogéneo de lo que es en realidad. Debido a ello, en este trabajo introducimos un modelo en el que tenemos en cuenta esas diferencias.

2.4.2. Individuos Distinguibles

Si nos situamos en el otro extremo, en este caso cada individuo de una subpoblación tiene asignado un destino fijo, de tal forma que individuos distintos van a estar en contacto con subpoblaciones afectadas de forma distinta por la epidemia. Entonces, no podemos asumir que habrá la misma proporción $\rho_i(t)$ de infectados entre individuos que viajan a diferentes destinos. Tenemos que separarlos e introducimos $\rho_{ij}(t)$ como la probabilidad de que un individuo del nodo i que tiene como destino el nodo j esté infectado en tiempo t . Notar que anteriormente nuestro estado mesoscópico se caracterizaba por un vector de N componentes, pero ahora tenemos tantas como links haya en la red, así que el coste computacional va a ser considerablemente mayor en este caso.

La matriz OD que usaremos será la cantidad de individuos del nodo i que viajan al j , llamada n_{ij} . Se relaciona con el formalismo anterior de la siguiente forma:

$$n_i \cdot R_{ij} = n_{ij} \Rightarrow \sum_{j=1}^N n_{ij} = n_i \quad \forall i \quad (7)$$

Esto significa que, teniendo la matriz de adyacencia de la red, podemos obtener las matrices origen-destino directamente para ambos formalismos.

Vamos a desarrollar este modelo y compararlo con el anterior. En la vida real, cada persona tiene un patrón característico con jerarquías de los sitios a los que prefieren ir (en [17] se observa que la repetitividad de cada lugar sigue una ley de Zipf). Hay que tener presente entonces que un modelo realista caería en un lugar medio entre ambos formalismos, y cuando queramos decidir cuál usar tenemos que tener siempre en cuenta las fuentes de nuestros datos.

Cuando los datos que se nos proporcionan representan viajes recurrentes, tiene más sentido utilizar el formalismo de individuos distinguibles.

3. Ecuaciones markovianas para individuos distinguibles

3.1. Obtención de las ecuaciones

Para escribir estas ecuaciones, pasamos de considerar probabilidades a proporción de individuos que realizan una acción.

Por ejemplo, a cada paso la probabilidad de que un individuo viaje a otro nodo es p_d , por lo que ahora la cantidad de personas que deciden viajar del nodo i al j es: $p_d \cdot n_{ij}$.

Queremos hallar la probabilidad de que un individuo susceptible se infecte estando en un nodo (ya sea el de origen si no nos hemos movido, o el destino en caso contrario). Para ello hallamos antes la probabilidad de que no se infecte con ningún contacto que tenga con el resto de agentes en ese nodo. Esa es una probabilidad que multiplica por $(1 - \lambda\rho_{ij})$ por cada contacto. Si estamos en el nodo i :

La cantidad de personas que se quedan en el nodo i y tienen asignado el destino j es $(1 - p_d)n_{ij}$.

La cantidad de personas que viajan del nodo j al i es $p_d n_{ij}$.

Si tenemos en cuenta todos los individuos del nodo i que se quedan y todos los del resto de nodos que viajan al i , tenemos la probabilidad de que no se infecte con ninguno. Se lo restamos a uno para tener la *probabilidad de infectarse en el nodo i* :

$$P_i(t) = 1 - \prod_{j=1}^N (1 - \lambda\rho_{ji}(t))^{p_d n_{ji}} \cdot \prod_{j=1}^N (1 - \lambda\rho_{ij}(t))^{(1-p_d)n_{ij}} \quad (8)$$

Ahora bien, queremos saber la evolución de la cantidad de contagiados en cualquier nodo para el siguiente paso temporal. En cada paso, hay una probabilidad μ de que una persona infectada se recupere. También, si una persona susceptible se queda en su nodo de origen, tiene una probabilidad P_i de contagiarse, y si viaja a otro nodo habrá una probabilidad P_j de contagio. Podemos escribir entonces un sistema de L ecuaciones:

$$\rho_{ij}(t+1) = (1 - \mu)\rho_{ij}(t) + (1 - \rho_{ij}(t))[(1 - p_d)P_i(t) + p_d P_j(t)] \quad (9)$$

De esta forma hemos escrito $\rho_{ij}(t+1)$ en función de $\rho_{ij}(t)$, y podemos iterar para simular su evolución.

3.2. Verificación de las ecuaciones

Cuando queremos calcular la evolución de un sistema lo primero que nos proponemos es realizar una simulación de agentes individuales (Montecarlo), que nos da el estado microscópico del sistema en función del tiempo. Las ecuaciones markovianas son un acercamiento mesoscópico a este problema, que no tiene en cuenta el estado de cada individuo, sino únicamente la situación de cada nodo. Vemos que la simulación de agentes tiene naturaleza estocástica, porque cada individuo tiene una probabilidad en cada paso temporal de viajar a otro nodo o quedarse en el suyo, de infectarse y de recuperarse en caso de estar infectado. Las ecuaciones markovianas por otra parte son deterministas.

Para que las dos formas de calcular el impacto de la epidemia estén de acuerdo, debemos eliminar rastros de la aleatoriedad característica de la simulación de agentes, y esto lo conseguimos promediando a lo largo de grandes periodos de tiempo. Una vez el sistema alcanza el estado estacionario (termalización) oscilará alrededor del mismo y haciendo la media eventualmente nos dará un resultado preciso. Por lo tanto está claro que esta clase de simulación es muy costosa. No solamente estamos tratando el sistema microscópico y por tanto tenemos gran cantidad de variables, sino que además necesitamos muchos pasos temporales para tener una buena estadística. Además, cerca del umbral epidémico cabe la posibilidad de que el número de infectados baje a cero y no vuelva a recuperarse, aunque el equilibrio esté por encima. Esto causa una mala estadística y debemos compensarlo haciendo la simulación varias veces.

Por otro lado, al iterar las ecuaciones markovianas directamente, tenemos L variables y el tiempo necesario dependerá de la precisión con la que necesitamos hallar el estado estacionario, ya que tiende asintóticamente hacia el mismo.

Para estas simulaciones usaremos redes de ciudades de Estados Unidos, en las cuales cada nodo se corresponde a un código postal, y la matriz origen-destino se ha obtenido a partir de encuestas a los residentes, teniendo así una representación de la cantidad de personas que viajan diariamente de una zona a otra. Estos datos se encuentran públicos actualmente en Internet ¹. Por lo tanto tenemos redes reales y en una escala razonable, ya que los pasos temporales corresponden a un día, que es el periodo de tiempo que las mayoría de personas toman para desplazarse al trabajo y volver. Una gran ventaja de este modelo es que incorpora inmediatamente datos de movilidad reales y se pueden aplicar a cualquier ciudad o a diferentes escalas.

En la Figura 3 mostramos los resultados de este punto. Se representa la fracción de individuos infectados en función de la probabilidad de contagio λ . El eje horizontal está normalizado respecto a λ_0 , que es el umbral epidémico cuando $p_d = 0$. Su expresión es la misma que cuando introducimos el modelo SIS: $\lambda_0 = \mu/n_{max}$.

Podemos observar que, a partir de cierto punto, la curva crece con una pendiente que eventualmente se reducirá. Todas estas curvas tienden asintóticamente a $1 - \mu$, que en nuestra simulación hemos tomado como $\mu = 0,2$.

Hemos obtenido un ajuste excelente, por lo que las ecuaciones han quedado validadas y no volveremos a utilizar las simulaciones de Montecarlo de ahora en adelante, porque requieren tiempos de computación excesivos.

¹Fuente de datos de movilidad de ciudades de Estados Unidos: <https://lehd.ces.census.gov/data/lodes/LODES7/>

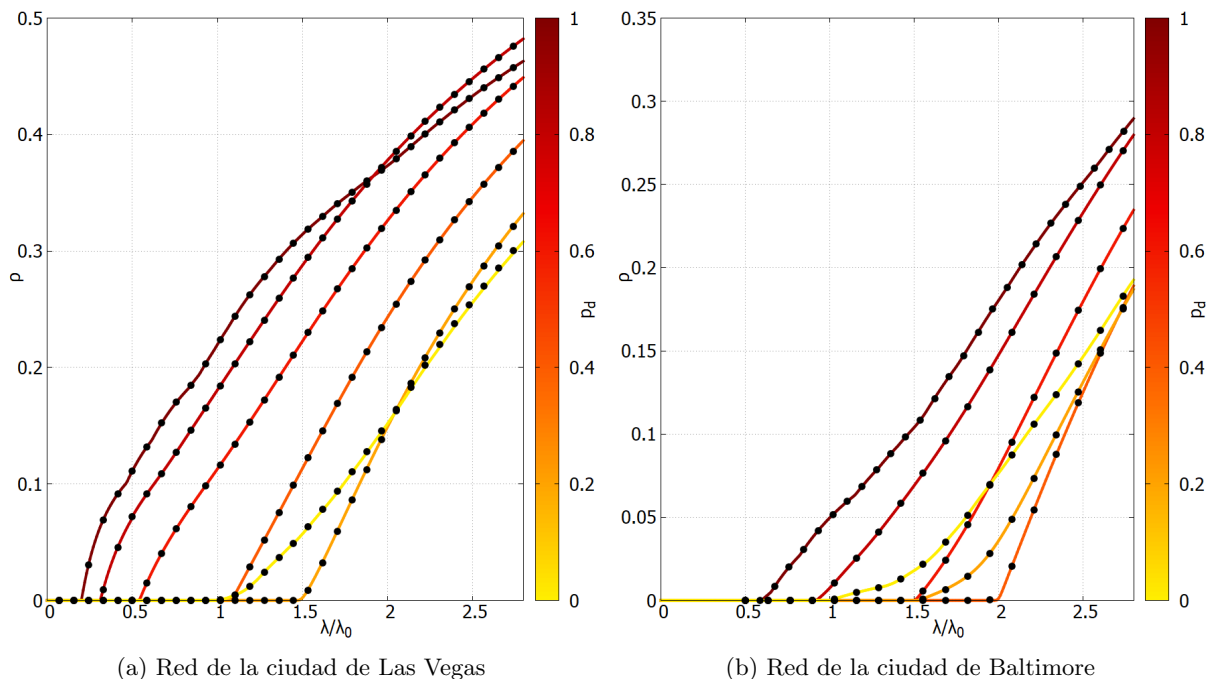


Figura 3: Validación de las ecuaciones de Markov para (a) la ciudad de Las Vegas y (b) Baltimore. En el eje vertical se representa la fracción de infectados ρ en el estado estacionario para toda la población, y en el eje horizontal el parámetro λ normalizado respecto al umbral epidémico para $p_d = 0$. El color se corresponde a distintos valores de p_d con los que se han llevado a cabo las simulaciones. Los puntos representan el promedio temporal de las simulaciones de agentes (Montecarlo) y las líneas continuas muestran los valores que resultan de iterar las ecuaciones markovianas derivadas en este punto.

Algo que llama la atención es que el umbral epidémico parece cambiar para distintos valores de p_d , y nos preguntamos ahora cuál es su dependencia con la cantidad de desplazamientos. A primera vista nos damos cuenta de que aumentar el valor de p_d por encima de cero puede hacer que el umbral epidémico crezca en algunos casos, lo cual es sorprendente. Sin embargo, gracias a haber obtenido las ecuaciones de Markov, podremos obtener también una expresión analítica de $\lambda_c(p_d)$ más adelante.

4. Estudio del umbral epidémico

Este es uno de los objetivos principales de la investigación de la propagación de epidemias en poblaciones: observar cómo cambia el umbral epidémico con la movilidad. Recordemos que el umbral epidémico nos da el valor que debe tener la probabilidad de contagio de una enfermedad para que la epidemia no se extinga. Es de gran interés a la hora de tomar medidas en una epidemia alcanzar valores altos del umbral.

4.1. Observación cualitativa. Variables λ , p_d

Al igual que en las gráficas anteriores, en las que representábamos la media de infectados respecto a λ para cierto p_d , calcularemos dicha media en función de ambas variables y lo dibujaremos en un mapa.

Lo primero que vemos en la Figura 4 es que hay dos regiones claramente diferenciadas. En la región inferior (en negro) tenemos todos los puntos en los que la enfermedad desaparece y sólo hay individuos sanos. En la región superior la enfermedad consigue mantener un número

determinado de infectados, alcanzando un equilibrio, y su aumento respecto a λ lo hemos visto ya en la Figura 3.

Lo siguiente que podemos ver es algo que también aparece en [16], y es que hay un detrimento en el contagio debido a la movilidad. El umbral epidémico aumenta para valores bajos de p_d y llega a un máximo generalmente en $0,2 < p_d < 0,4$ tras el cual vuelve a decrecer a un valor menor incluso que el inicial.

Es algo anti-intuitivo que haya un detrimento de la propagación de la epidemia al aumentar la movilidad, ya que tradicionalmente se piensa que al haber más movimiento entre poblaciones una enfermedad se contagia más fácilmente. Sin embargo, tiene sentido que exista ese detrimento porque en $p_d = 0$ la enfermedad está concentrada en el nodo de mayor población. Al aumentar p_d esparcimos los portadores de la enfermedad a otros nodos más pequeños donde no se va a propagar bien, mientras que los individuos que llegan desde el resto de nodos pequeños son mayoritariamente sanos. Esto homogeneiza la epidemia, de forma que tenemos menos infectados en el nodo más poblado y más en otros nodos pequeños. Hay que tener en cuenta que, en general, a mayor heterogeneidad en una red menor será el umbral epidémico.

Todo esto no cambia el hecho de que si partimos del otro extremo, $p_d = 1$, también podemos reducir el impacto de la enfermedad e incluso hacerla desaparecer al disminuir los viajes. En el siguiente paso vamos a obtener las ecuaciones con las que podremos calcular de forma mucho más rápida el umbral epidémico $\lambda_c(p_d)$ y con las que tendremos una visión más significativa del proceso en sí.

4.2. Linealización de las ecuaciones. Cálculo del umbral epidémico

En los algoritmos de Montecarlo y las iteraciones de las ecuaciones de Markov partimos de estadística o de iterar ecuaciones que se acercan asintóticamente al valor correcto. En esta parte tendremos por primera vez un método que nos permitirá llegar directamente al resultado exacto. Pero antes tenemos que linealizar las ecuaciones.

- Estamos en el estado estacionario. La cantidad de individuos que se infectan cada turno debe ser igual a la de individuos que recuperan en cada nodo, y la cantidad de infectados permanece constante:

$$\rho_{ij}(t+1) = \rho_{ij}(t) \equiv \rho_{ij}^* \quad \forall i, j \quad (10)$$

- La cantidad de infectados debe tender a cero, por lo que definimos:

$$\rho_{ij}^* \equiv \varepsilon_{ij}^* \ll 1 \quad \forall i, j \quad (11)$$

Lo mismo se aplica para $P_i(\boldsymbol{\varepsilon}(t)) = P_i(\boldsymbol{\varepsilon}^*)$.

Entonces las ecuaciones (8) y (9) tienen la siguiente forma:

$$P_i(\boldsymbol{\varepsilon}^*) = 1 - \prod_{j=1}^N (1 - \lambda \varepsilon_{ji}^*)^{p_d n_{ji}} \cdot \prod_{j=1}^N (1 - \lambda \varepsilon_{ij}^*)^{(1-p_d) n_{ij}}$$

$$\varepsilon_{ij}^* = (1 - \mu) \varepsilon_{ij}^* + (1 - \varepsilon_{ij}^*) [(1 - p_d) P_i(\boldsymbol{\varepsilon}^*) + p_d P_j(\boldsymbol{\varepsilon}^*)]$$

Podemos linealizar la primera ecuación tomando sólo los términos hasta orden 1 de su desarrollo en serie:

$$\prod_{k=1}^N (1 - \lambda \varepsilon_{ki}^*)^{p_d n_{ki}} \sim \prod_{k=1}^N (1 - \lambda p_d n_{ki} \varepsilon_{ki}^*) \sim 1 - \sum_{k=1}^N \lambda p_d n_{ki} \varepsilon_{ki}^*$$

$$\prod_{k=1}^N (1 - \lambda \varepsilon_{ik}^*)^{p_d n_{ik}} \sim \prod_{k=1}^N (1 - \lambda p_d n_{ik} \varepsilon_{ik}^*) \sim 1 - \sum_{k=1}^N \lambda p_d n_{ik} \varepsilon_{ik}^*$$

Multiplicando ambos términos, la primera ecuación queda así:

$$P_i(\varepsilon^*) = 1 - \left(1 - \sum_{k=1}^N \lambda p_d \varepsilon_{ki}^* n_{ki}\right) \cdot \left(1 - \sum_{k=1}^N \lambda (1 - p_d) \varepsilon_{ik}^* n_{ik}\right) \sim \sum_{k=1}^N (\lambda p_d \varepsilon_{ki}^* n_{ki} + \lambda (1 - p_d) \varepsilon_{ik}^* n_{ik})$$

Ahora sustituimos P_i y P_j en la otra ecuación y tenemos el resultado final:

$$\frac{\mu}{\lambda} \varepsilon_{ij}^* = \sum_{k=1}^N p_d (1 - p_d) n_{ki} \varepsilon_{ki}^* + (1 - p_d)^2 n_{ik} \varepsilon_{ik}^* + p_d^2 n_{kj} \varepsilon_{kj}^* + p_d (1 - p_d) n_{jk} \varepsilon_{jk}^* \quad (12)$$

Esto es evidentemente una ecuación de autovalores, en la que el elemento ε_{ij}^* ocupa la posición $i \cdot N + j$ en el autovector ε^* . Reescribimos los términos de la siguiente manera para que queden los conceptos más claros:

$$\frac{\mu}{\lambda} \varepsilon^i_j = \sum_{k,l=1}^N M_{jk}^{il} \cdot \varepsilon^l_k \quad (13)$$

Podemos conocer cada término de la matriz estudiando cada caso a partir de la ecuación (12):

$\forall i, j, k$ tal que $1 \leq i, j, k \leq N$

- $M_{ji}^{ik} = p_d (1 - p_d) n_{ki}$
- $M_{jk}^{ii} = (1 - p_d)^2 n_{ik}$
 - $M_{jj}^{ik} = p_d^2 n_{kj}$
- $M_{jk}^{ij} = p_d (1 - p_d) n_{jk}$

Lo interpretamos de la siguiente manera. Cada término es la cantidad de contactos que tiene un individuo que suele viajar del nodo i al j con:

- El primer término es con los individuos que se desplazan del k al i .
- El segundo es con los que se desplazan del i al k .
- El tercero es con los que van del k al j .
- Por último, este es el de los que van del j al k .

Como un mismo término de la matriz puede aparecer en el sumatorio varias veces, hay ciertas excepciones donde hay que sumar términos cuando $k = i$ o $k = j$:

- $M_{jj}^{ii} = (1 + 2p_d(p_d - 1)) n_{ij}$
- $M_{ji}^{jj} = 2p_d(1 - p_d) n_{ji}$

Podemos representar la matriz \mathbf{M} de la forma:

$$\frac{\mu}{\lambda} \begin{pmatrix} \varepsilon_1^1 \\ \dots \\ \varepsilon_N^1 \\ \dots \\ \varepsilon_1^N \\ \dots \\ \varepsilon_N^N \end{pmatrix} = \begin{pmatrix} M_{11}^{11} & \dots & M_{1N}^{11} & \dots & \dots & | & M_{11}^{1N} & \dots & M_{1N}^{1N} \\ \dots & \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ M_{N1}^{11} & \dots & M_{NN}^{11} & \dots & \dots & | & M_{N1}^{1N} & \dots & M_{NN}^{1N} \\ \dots & \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ M_{11}^{N1} & \dots & M_{1N}^{N1} & \dots & \dots & | & M_{11}^{NN} & \dots & M_{1N}^{NN} \\ \dots & \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ M_{N1}^{N1} & \dots & M_{NN}^{N1} & \dots & \dots & | & M_{N1}^{NN} & \dots & M_{NN}^{NN} \end{pmatrix} \begin{pmatrix} \varepsilon_1^1 \\ \dots \\ \varepsilon_N^1 \\ \dots \\ \varepsilon_1^N \\ \dots \\ \varepsilon_N^N \end{pmatrix} \quad (14)$$

Cada bloque M^{ab} contiene a los elementos que relacionan al individuo que sale del nodo a con el que sale del nodo b .

Esto significa que podemos obtener los autovalores de la matriz, que se corresponderán con:

$$\Lambda(\mathbf{M}) = \frac{\mu}{\lambda} \quad (15)$$

Como queremos conocer el umbral epidémico, que es el mínimo valor de λ para el cual se mantiene la enfermedad, vamos a tomar únicamente el autovalor máximo de la matriz:

$$\lambda_c = \frac{\mu}{\Lambda_{max}(\mathbf{M})} \quad (16)$$

La matriz \mathbf{M} es característica de la matriz origen-destino y el valor de p_d , así que podemos obtener el umbral epidémico de la red en función de esa variable introduciendo los valores en un programa que nos devuelva los autovalores.

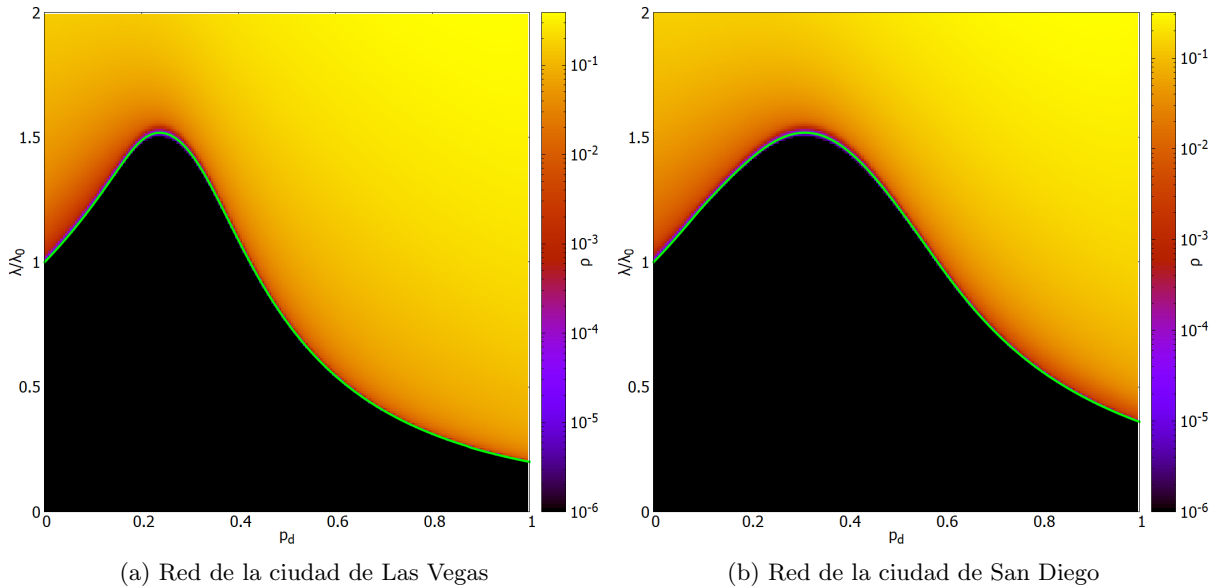


Figura 4: Mapas de calor para las ciudades de (a) Las Vegas y (b) San Diego. En el eje vertical se encuentra la probabilidad de infección por contacto λ normalizado de la misma forma que anteriormente, y en el eje horizontal tenemos la probabilidad de desplazamiento p_d . La escala de color representa la fracción de infectados ρ estacionaria obtenida iterando las ecuaciones markovianas, y la línea continua verde es el umbral epidémico, resultado de linealizar las ecuaciones, que marca la frontera entre las dos fases (en negro no existe epidemia, en amarillo se propaga).

Comprobamos que, en efecto, este método es válido superponiendo la curva $\lambda_c(p_d)$ sobre el mapa de calor que hemos obtenido anteriormente. En la Figura 4 se muestra como la línea verde.

4.3. Comparando los formalismos. Detrimento de la propagación en $p_d = 1$

Mostramos en la Figura 5 los umbrales epidémicos para varias redes, correspondientes a diversas ciudades estadounidenses. Las curvas del caso distinguible se han obtenido a partir de la expresión (16) resultante de linealizar las ecuaciones de Markov, y las del caso indistinguible vienen de la ecuación (6) que se ha desarrollado en [16].

Para cada red, el umbral epidémico es siempre menor en el caso distinguible que en el indistinguible. Esto de nuevo tiene que ver con la diferencia de heterogeneidad entre ambos modelos. Mientras que en el caso distinguible los individuos que mantienen el foco principal (en el nodo con más población efectiva) siempre regresan al mismo, manteniendo activa la epidemia. En el caso indistinguible, por el contrario, una persona que haya estado en el foco puede acabar en otro lugar distinto al siguiente paso, homogeneizando el sistema y aumentando de esta forma el umbral. Estas diferencias se acentúan conforme el valor de p_d crece.

En el otro extremo, el de $p_d = 0$ los dos casos dan la misma pendiente de λ respecto a p_d . Es razonable pensar que, ya que el detrimento es causado por el hecho de que los residentes del foco de contagio empiezan a salir de ese nodo, a primer orden no importa a qué otro nodo vayan. Sólo es relevante el hecho de que salgan, y por eso las diferencias entre los formalismos sólo se hacen evidentes cuando los individuos que entran (no los que salen) al foco juegan un papel importante.

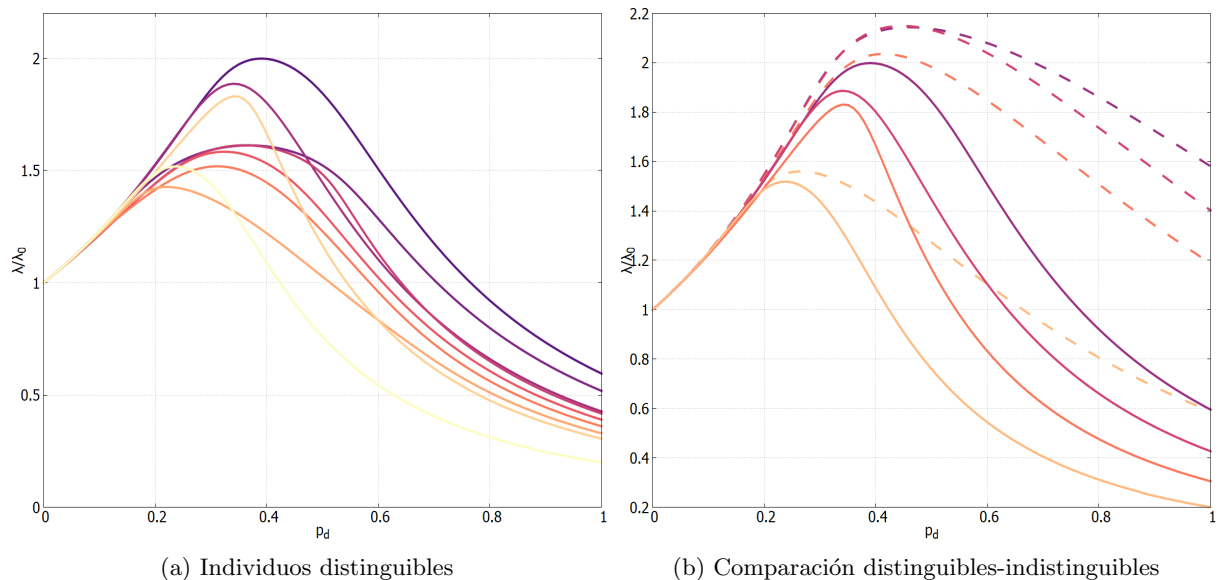


Figura 5: En la figura (a) representamos los umbrales epidémicos obtenidos en el caso distinguible para nueve ciudades, en orden de oscuro a claro: Baltimore, Minneapolis, San Francisco, Atlanta, Pittsburgh, San Diego, Seattle, Houston y Las Vegas. En el eje vertical está el umbral epidémico y en el horizontal la probabilidad de desplazamiento. En la figura (b) comparamos los umbrales de cuatro ciudades para el caso distinguible (línea continua) con el indistinguible (línea discontinua), de las ciudades Baltimore, San Francisco, Houston y Las Vegas de oscuro a claro.

Notar que si tomáramos sistemas de los cuales hemos obtenido, por ejemplo, datos de viajes diarios (regulares) y los estudiáramos con el formalismo de individuos indistinguibles, podríamos estar sobreestimando el valor del umbral epidémico en muchos casos.

Debemos preguntarnos cómo afecta disminuir la movilidad en cada caso, por lo que medimos las pendientes de todas las curvas en $p_d = 1$ y las comparamos entre ellas (Figura 6).

No se ve una clara relación entre las pendientes en ambos casos para cada red, los puntos están bastante dispersos. Lo que sí que podemos notar es que en casi todos los casos la pendiente en el caso distinguible es menor. Esto es debido a que el valor máximo de la pendiente se alcanza antes (alrededor de $p_d = 0,6$) y ya se ha frenado, mientras que las curvas en el caso indistinguible llegan al punto de inflexión muy cerca de $p_d = 1$ o no llegan en ningún momento.

En situaciones en las que estemos estudiando una epidemia existente y queramos hacer un modelo que se acerque lo máximo posible a la realidad, esta parte cobra mucha importancia ya que nos da mucha información de cuán prioritario es reducir la movilidad entre poblaciones. Mientras que no es el caso en este trabajo, en el que estudiamos un caso muy general y no nos centramos en obtener valores numéricos, podemos extraer algunas conclusiones.

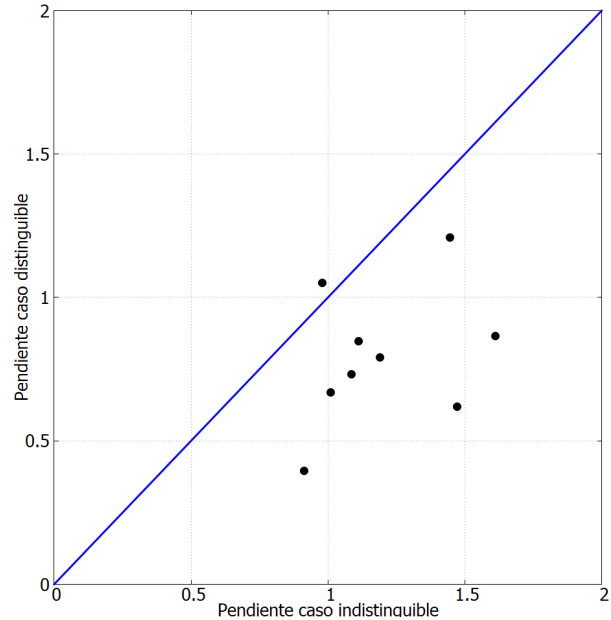


Figura 6: Representación de la pendiente de la curva del umbral epidémico en $p_d = 1$ en el caso distinguible respecto al indistinguible para cada red. La línea azul es de pendiente $m = 1$

En los casos en los que las personas viajen a un lugar repetidamente va a ser más difícil que la epidemia se extinga, mientras que si cada persona viaja a diversos sitios ocurre lo contrario. Además, reducir la movilidad va a ser a primer orden más efectivo en el caso indistinguible, pero para una reducción grande de p_d la epidemia se frena más bruscamente en el caso distinguible.

Como hemos comentado en la introducción, el estudio de la movilidad humana nos ayuda mucho a conocer cuánto se acerca nuestro modelo a la realidad y, si es necesario, elegir qué formalismo utilizar dependiendo del sistema que estudiemos. El comportamiento de las personas puede ser muy diferente dependiendo de la escala (barrios, ciudades o incluso países), y mientras que los viajes diarios son repetitivos (el modelo de individuos distinguibles sería más adecuado) otra clase de viajes como los internacionales pueden no ser recurrentes (por lo tanto siendo conveniente usar el formalismo de individuos indistinguibles).

Un resultado muy interesante que se obtiene es que se pueden recuperar las ecuaciones linealizadas del formalismo de individuos indistinguibles a partir de las del de distinguibles. Las cuentas se encuentran en el anexo 7.1. El procedimiento será imponer una serie de condiciones sobre las ecuaciones (12) a partir de las cuales, tras un desarrollo, llegamos a las del caso indistinguible (17). A continuación veremos por qué obtener estas ecuaciones es tan importante para comprender las diferencias entre ambos formalismos y cómo la heterogeneidad influye no solamente en el umbral epidémico, sino en su comportamiento a altos valores de λ .

5. Impacto de la epidemia para alta movilidad

En el caso en el que estamos cerca de $p_d = 1$ hemos visto cómo cambia el umbral epidémico, y las diferencias que tenemos entre los dos formalismos. Sin embargo, no siempre nos vamos a encontrar cerca del umbral, sino que podemos tener un caso en el que cierta enfermedad sea muy contagiosa y queramos medir el impacto que va a tener en la población. Nos topamos con algo inesperado que vemos en la Figura 7: la curva de infección en el caso de individuos indistinguibles, a pesar de que tiene un umbral mayor y empieza con una infección mucho menor, acaba alcanzando y superando con creces a la del caso de individuos distinguibles. Es decir, para enfermedades muy contagiosas, parece que la enfermedad se propaga más fácilmente si los individuos están dispersos y viajan a diferentes sitios. A primera vista esto contradice lo que habíamos dicho en un principio.

Sin embargo, vamos a aprovechar esto para entender por fin las diferencias principales entre los formalismos. Volvemos a las ecuaciones linealizadas de ambos: (12) y (17).

Como hemos visto, en el caso distinguible (12) cada término nos relaciona los individuos susceptibles con los infectados (ejemplo: el primer término relacionaba a los individuos susceptibles que viajan del nodo i al j con los infectados que van del k al i). Nos damos cuenta de que ese valor no es más que el número de contactos que va a tener el susceptible con otros individuos

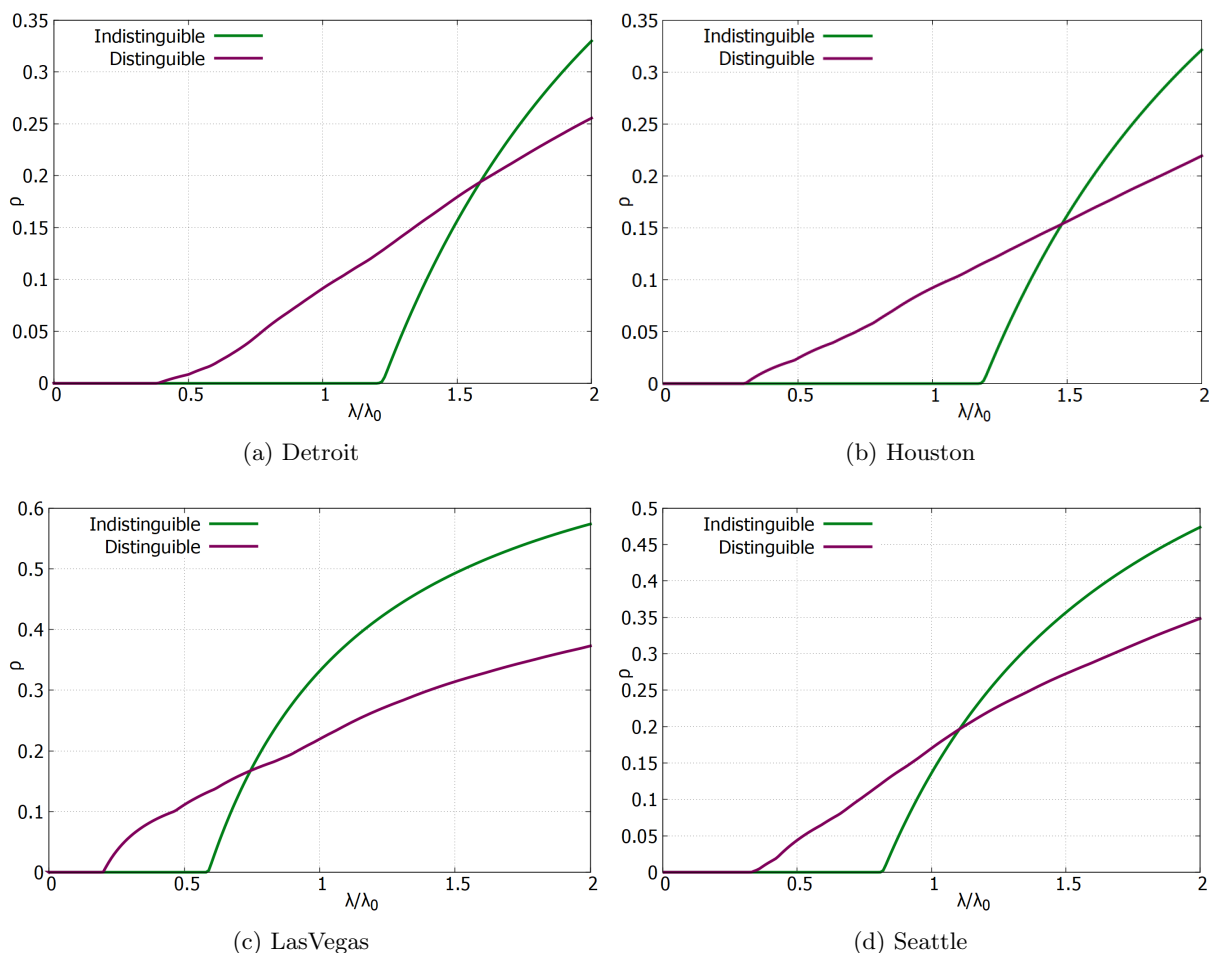


Figura 7: Impacto de la epidemia en función de la probabilidad de infección λ para $p_d = 1$ en las ciudades de (a) Detroit, (b) Houston, (c) Las Vegas y (d) Seattle. La curva verde corresponde al caso indistinguible y la morada al caso distinguible.

infectados. Si no tenemos en cuenta el término ε , lo que queda es el número *total* de contactos que un individuo tiene con los de otros nodos, y es lo que contiene la matriz \mathbf{M} .

La forma en la que vamos a abordar la pregunta es a partir de la cantidad de contactos que ocurren en cada paso. Lo primero que podemos observar es que, en total, hay la misma cantidad de contactos para ambos formalismos. Esto tiene sentido, ya que en ambos casos hay la misma cantidad de gente viajando de un nodo a otro, y las personas que hay en cada nodo a la hora de infectarse (ese número se conoce como la población efectiva de los nodos) son las mismas.

Calcular la cantidad de contactos es sencillo: quitando los términos ε de la parte derecha de la ecuación (12) obtenemos el número de contactos de una persona que va del nodo i al j .

Explícitamente eso es:

$$N_{\text{contactos}}(i \rightarrow j) = \sum_{k=1}^N p_d(1 - p_d)n_{ki} + (1 - p_d)^2 n_{ik} + p_d^2 n_{kj} + p_d(1 - p_d)n_{jk}$$

Si recordamos cómo hemos definido la matriz \mathbf{M} en (14) nos damos cuenta de que esto se corresponde a sumar todos los términos en la fila $i \cdot N + j$ de dicha matriz:

$$N_{\text{contactos}}(i \rightarrow j) = \sum_{k,l=1}^N M_{jl}^{ik}$$

Si queremos calcular los contactos totales que hay en toda la red a cada turno, multiplicamos los contactos de cada individuo por la cantidad de individuos n_{ij} , y sumamos todo:

$$N_{\text{contactos}} = \sum_{i,j=1}^N \left(n_{ij} \sum_{k,l=1}^N M_{jl}^{ik} \right)$$

Si solamente nos interesa hacer esto para $p_d = 1$ es más sencillo. La ecuación (12) pasa a ser:

$$\varepsilon_{ij} = \sum_{k=1}^N n_{kj} \varepsilon_{kj}^* \Rightarrow N_{\text{contactos}}(i \rightarrow j) = \sum_{k=1}^N n_{kj}$$

Que es la suma de los elementos de la columna j de la matriz \mathbf{n} . De nuevo:

$$N_{\text{contactos}} = \sum_{i,j=1}^N \left(n_{ij} \sum_{k=1}^N n_{kj} \right)$$

Hasta aquí el caso distinguible. Si queremos hacer lo mismo para el caso de individuos indistinguibles, volvemos a las ecuaciones (17) y hacemos lo mismo. Quitando el término ε_j de la ecuación tenemos el número de contactos que tiene un individuo perteneciente al nodo i con individuos del nodo j . Esto es explícitamente:

$$N_{\text{contactos}}(i) = \sum_{j=1}^N p_d(1 - p_d)(R_{ij} + R_{ji})n_j + (1 - p_d)^2 n_j \delta_{ji} + p_d^2 n_j \left(\sum_{l=1}^N R_{il} R_{jl} \right) = M_{ij}^{\text{indist}}$$

Si tomamos $p_d = 1$ se simplifica a:

$$N_{\text{contactos}}(i) = \sum_{j=1}^N n_j \left(\sum_{l=1}^N R_{il} R_{jl} \right)$$

Este resultado no es tan intuitivo como en el caso de individuos distinguibles, donde cada individuo que se desplaza simplemente va a su destino y tiene contactos con los que haya allí. Ahora, un individuo tiene muchos destinos, por lo que podemos imaginarnos que va a todos al

mismo tiempo. De ahí viene el término R_{il} , que nos da con qué 'peso' va a cada lugar.

De nuevo multiplicamos esto por la cantidad de individuos en cada nodo, los sumamos y tenemos la cantidad total de contactos:

$$N_{\text{contactos}} = \sum_{i,j=1}^N n_i n_j \left(\sum_{l=1}^N R_{il} R_{jl} \right)$$

Haciendo esto obtenemos el mismo número total de contactos para distinguibles e indistinguibles. Pero esto no es toda la información. Es evidente que las diferencias entre los formalismos no se encuentran en la cantidad total de contactos, sino la localización de los mismos. Aquí es donde aparece la 'desigualdad' entre individuos. Mientras que unas personas tengan muy pocos contactos, otras tendrán muchos, y son las últimas las que mantienen la enfermedad transmitiéndose.

Así que lo que nos queda es figurar en qué formalismo hay más 'desigualdad' entre individuos. Una manera de verlo es calculando la varianza.

Primero calculamos la media, para lo que simplemente dividimos la cantidad total de contactos entre la población total, donde:

$$n_T = \sum_{i,j=1}^N n_{ij}$$

La varianza la podemos calcular como:

$$V = \langle x^2 \rangle - \langle x \rangle^2$$

Donde x es la cantidad de contactos de cada individuo. Como la media ya la sabemos, sólo queda calcular la media de x^2 :

$$x^2 = \left(\sum_{k=1}^N n_{kj} \right)^2 \Rightarrow \langle x^2 \rangle = \frac{1}{n_T} \cdot \sum_{i,j=1}^N n_{ij} \left(\sum_{k=1}^N n_{kj} \right)^2$$

Podemos ver los resultados en la tabla 1

Red	Media	Varianza (Indist)	Varianza (Dist)
Detroit	17 846	1 880	13 838
Houston	31 246	6 131	31 536
Las Vegas	43 270	6 039	44 980
Seattle	28 820	7 098	22 800

Cuadro 1: Medias y varianzas de los contactos por individuo para cada red y formalismo.

Esto nos da una pista muy importante y la diferencia es clara. La varianza en el caso de individuos distinguibles es mucho mayor que en el otro. Una varianza tan alta significa que hay mucha desigualdad, es decir, unos individuos tendrán muchos contactos en comparación a otros.

Esto explica el hecho de que el umbral epidémico del caso distinguishible sea mucho menor, y que la enfermedad tienda a no extinguirse con facilidad. Como hemos comentado varias veces a lo largo del trabajo, el umbral epidémico es más bajo cuanto más heterogénea es la red, y hemos visto que esta regla se verifica en cada paso. Pero todavía no hemos explicado por qué en el caso indistinguishible hay un impacto mayor cuando λ es alta. Para arrojar luz .

En el histograma contaremos la cantidad de personas que tienen ciertos contactos entre dos valores. Recordar que cada elemento que vamos a contar tiene un peso distinto (el elemento asociado a ε_{ij} debe contarse n_{ij} veces). De nuevo, compararemos ambos formalismos para cada red. Vemos el resultado en la Figura 8.

En efecto, los histogramas de casos distinguibles nos dan mucha información. Vemos que hay unos pocos nodos que acaparan muchas personas, y es gracias a ellos que la epidemia puede mantenerse incluso para una λ muy baja. Conforme aumenta la λ , otros nodos menos poblados empiezan a permitir contagios y por ello la curva parece que aumenta a 'saltos'. Cada nodo es independiente, y se van sumando consecutivamente. Podemos ver que hay también claras diferencias entre diferentes redes para individuos distinguibles. En el caso de Las Vegas, hay un nodo que atrae a una gran cantidad de individuos, dejando el resto de nodos con poca gente. Por eso la curva aumenta muy pronto, pero no llega a alcanzar un valor grande. En Detroit, por otro lado, al ser el histograma más continuado, la curva aumenta de forma constante.

En el caso de individuos indistinguibles vemos una sorprendente homogeneidad. Esto es porque aunque los nodos grandes acumulan los contagios, los individuos los dispersan en sus respectivos nodos de origen. Las personas tendrán una cantidad de contactos que media respecto a los que se dan en los nodos de destino. En redes con muchas conexiones, esto significa que esa media será similar para toda la red.

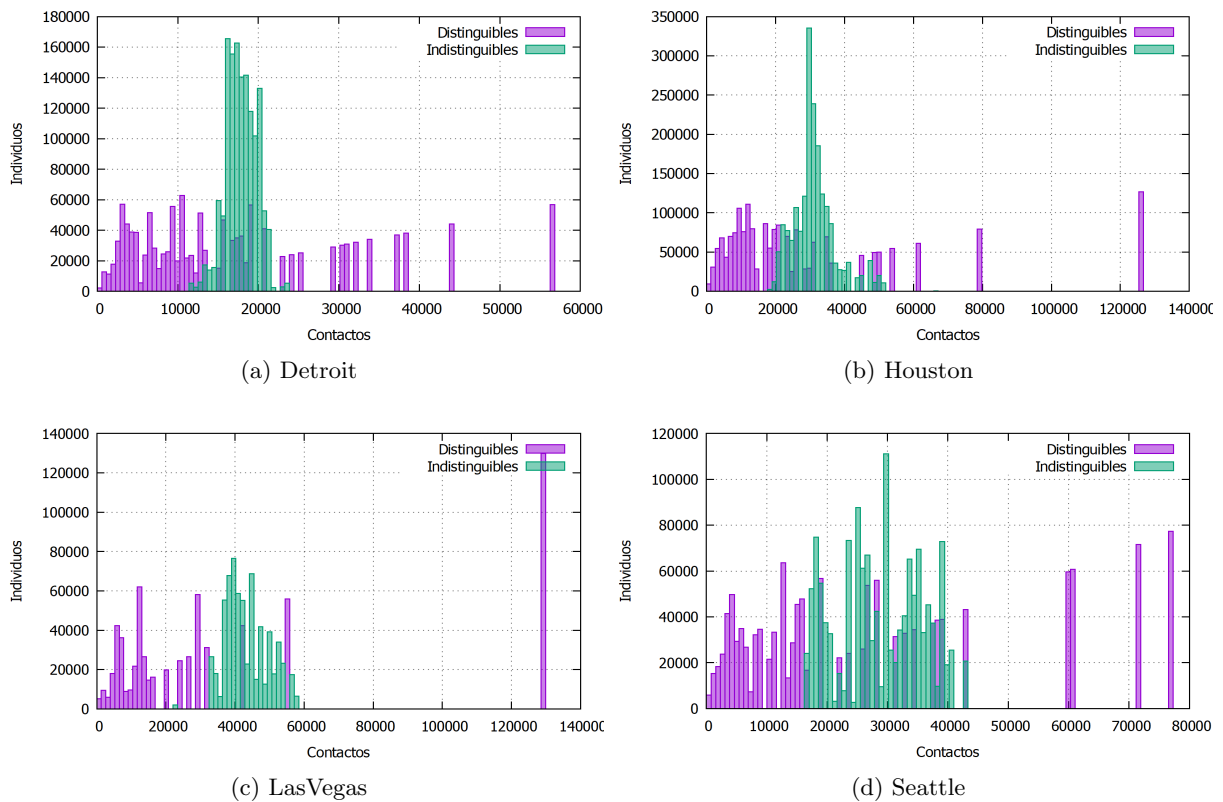


Figura 8: Histogramas que representan la cantidad de individuos que tienen un número de contactos que representamos en el eje horizontal, para las ciudades de (a) Detroit, (b) Houston, (c) Las Vegas y (d) Seattle. En color morado se representa el histograma para el caso distinguible y en verde para el caso indistinguible.

6. Conclusiones

A lo largo del trabajo se ha mostrado un marco existente en la investigación de la propagación de enfermedades, se ha visto por qué ha supuesto una mejora a los modelos anteriores y se ha incidido en sus limitaciones, introduciendo de esta forma un nivel más de complejidad en el modelo que permite representar de forma más fiable muchos casos reales. Hemos concluido en que el caso de individuos indistinguibles se adapta mejor a datos de movilidad no recurrentes, que ocurren a escalas mayores de espacio y tiempo (viajes internacionales o turismo) frente a los datos que obtenemos de viajes diarios o repetitivos (al trabajo, a la universidad, etc) para los cuales el modelo anterior puede subestimar el impacto de la epidemia (o sobreestimarlos en algunos casos), requiriendo de esta forma usar el formalismo de individuos distinguibles que hemos desarrollado en este trabajo.

Se han visto las ventajas de haber obtenido ecuaciones a partir del acercamiento de la cadena de Markov, con las que no solo podemos derivar una expresión analítica para el umbral epidémico sino que nos permiten un entendimiento mayor de cómo se contagia la enfermedad a nivel de la cantidad de contactos entre individuos y de cómo la heterogeneidad de la red juega un papel crucial.

A nivel de computación, las ecuaciones suponen una reducción enorme del tiempo y la memoria necesarios para llevar a cabo una simulación. El hecho de que al iterar las ecuaciones markovianas no se tenga en cuenta a todos los agentes individualmente y no haya que obtener estadística para desprendernos del ruido reduce los tiempos de computación en dos o tres órdenes de magnitud, dependiendo de la precisión que necesitemos. El modelo que hemos introducido, sin embargo, requiere de más tiempo y memoria que el anterior, ya que consta de L ecuaciones en vez de N . A la hora de hallar los autovalores de la matriz para obtener el umbral, el número de elementos no nulos de dicha matriz crece con L en el caso de individuos indistinguibles y con $L^{3/2}$ en el de distinguibles.

Hemos visto que introducir este formalismo supone una mejora sobre el anterior, añadiendo más realismo y adaptándose mejor a muchos datos de movilidad. Sin embargo, hay muchas variables adicionales que pueden tenerse en cuenta y no hemos introducido aún en el modelo. Además, todavía no se ha llevado a cabo un estudio del formalismo para el modelo SIR (susceptible-infectado-recuperado), aunque la mayor parte de los resultados de los SIS y los SIR son equivalentes. Una gran ventaja del formalismo es que varias mejoras y adaptaciones llevadas a cabo en el modelo existente (de individuos indistinguibles) también se pueden aplicar de la misma forma al de individuos distinguibles (multiplexes [18], cambio de comportamiento de individuos infectados, vectores de transmisión [19], etc). Por otra parte, se abren caminos de investigación específicamente alrededor de este formalismo. Por ejemplo, distinguir personas que viajen a diario de las que siempre permanezcan en su nodo, añadiendo otro grado de distinguibilidad, o añadiendo un parámetro que relacione ambos formalismos (parámetro de distinguibilidad, decidiendo la proporción de viajes recurrentes frente a los aleatorios). De cualquier manera, queda mucho trabajo por delante y la necesidad de desarrollar estos modelos y seguir investigando los efectos de la movilidad en el contagio de enfermedades es en la actualidad mayor que nunca.

Referencias

- [1] N. Boccarda, *Modeling complex systems*. Springer Science & Business Media, 2010.
- [2] S. H. Strogatz, *Sync: How order emerges from chaos in the universe, nature, and daily life*. Hachette UK, 2012.
- [3] M. Resnick, *Turtles, termites, and traffic jams: Explorations in massively parallel micro-worlds*. Mit Press, 1997.
- [4] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.
- [5] B. Gonçalves and N. Perra, *Social phenomena: From data analysis to models*. Springer, 2015.
- [6] W. Souma, Y. Fujiwara, and H. Aoyama, “Complex networks and economics,” *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 1-2, pp. 396–401, 2003.
- [7] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, “Human mobility: Models and applications,” *Physics Reports*, vol. 734, pp. 1–74, 2018.
- [8] D. E. Boyce and H. C. Williams, *Forecasting urban travel: Past, present and future*. Edward Elgar Publishing, 2015.
- [9] B. Hillier, A. Turner, T. Yang, and H.-T. Park, “Metric and topo-geometric properties of urban street networks: some convergences, divergences and new results,” *Journal of Space Syntax Studies*, 2009.
- [10] V. E. Krebs, “Mapping networks of terrorist cells,” *Connections*, vol. 24, no. 3, pp. 43–52, 2002.
- [11] V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani, “Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions,” *PLoS medicine*, vol. 4, no. 1, 2007.
- [12] A. Vespignani, “Modelling dynamical processes in complex socio-technical systems,” *Nature physics*, vol. 8, no. 1, pp. 32–39, 2012.
- [13] M. Tizzoni, P. Bajardi, A. Decuyper, G. K. K. King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza, “On the use of human mobility proxies for modeling epidemics,” *PLoS computational biology*, vol. 10, no. 7, 2014.
- [14] D. Brockmann, L. Hufnagel, and T. Geisel, “The scaling laws of human travel,” *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.
- [15] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, “On the levy-walk nature of human mobility,” *IEEE/ACM transactions on networking*, vol. 19, no. 3, pp. 630–643, 2011.
- [16] J. Gómez-Gardeñes, D. Soriano-Paños, and A. Arenas, “Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks,” *Nature Physics*, vol. 14, no. 4, pp. 391–395, 2018.
- [17] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, “Understanding individual human mobility patterns,” *nature*, vol. 453, no. 7196, pp. 779–782, 2008.

- [18] D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, “Spreading Processes in Multiplex Metapopulations Containing Different Mobility Networks,” *Physical Review X*, vol. 8, no. 3, p. 031039, aug 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevX.8.031039>
- [19] D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, H. J. Martínez, S. Meloni, and J. Gómez-Gardeñes, “Vector-borne epidemics driven by human mobility,” *Phys. Rev. Research*, vol. 2, p. 013312, Mar 2020. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevResearch.2.013312>

7. Anexos

7.1. Anexo I: Recuperación de las ecuaciones del formalismo de individuos indistinguibles

Las ecuaciones linealizadas para el caso indistinguible, que han sido derivadas en [16], son:

$$\frac{\mu}{\lambda} \varepsilon_i^* = \sum_{j=1}^N (p_d(1-p_d)(R_{ij} + R_{ji})n_j + (1-p_d)^2 n_j \delta_{ji} + p_d^2 n_j (\sum_{l=1}^N R_{il} R_{jl})) \cdot \varepsilon_j^* \quad (17)$$

Para recuperar la indistinguibilidad vamos a escribir las siguientes condiciones:

- 1) $\varepsilon_{ij} = \varepsilon_i \forall j$
- 2) $n_{ij} = n_i \cdot R_{ij}$
- 3) $\sum_{j=1}^N R_{ij} = 1$

En realidad, únicamente la primera es una verdadera condición, ya que obligamos a que todas las ε_{ij} sean iguales para una misma i . La segunda solamente relaciona ambos formalismos, y la tercera es una igualdad que ya cumplían las R_{ij}

Vamos a partir de la ecuación (12), la cual vamos a reescribir como:

$$\frac{\mu}{\lambda} \varepsilon_{ij}^* = \sum_{m,l=1}^N p_d(1-p_d)n_{km} \varepsilon_{km}^* \delta_{mi} + (1-p_d)^2 n_{mk} \varepsilon_{mk}^* \delta_{mi} + p_d^2 n_{km} \varepsilon_{km} \delta_{mj} + p_d(1-p_d)n_{mk} \varepsilon_{mk} \delta_{mj} \quad (18)$$

De las condiciones 1) y 3) podemos obtener la siguiente igualdad:

$$\sum_{j=1}^N \frac{\mu}{\lambda} \varepsilon_{ij}^* R_{ij} = \sum_{j=1}^N \frac{\mu}{\lambda} \varepsilon_i^* R_{ij} = \frac{\mu}{\lambda} \varepsilon_i^* \sum_{j=1}^N R_{ij} = \frac{\mu}{\lambda} \varepsilon_i^* \quad (19)$$

Esto significa que si cogemos todos los términos ε_{ij} para un mismo i y los sumamos de esa forma, se nos simplifica y podemos hacer lo mismo con la parte derecha de la ecuación (18). Entonces:

$$\begin{aligned} \sum_{j=1}^N \frac{\mu}{\lambda} \varepsilon_{ij}^* R_{ij} &= \sum_{m,l,j=1}^N p_d(1-p_d)n_{km} \varepsilon_{km}^* \delta_{mi} R_{ij} + (1-p_d)^2 n_{mk} \varepsilon_{mk}^* \delta_{mi} R_{ij} + \\ &+ p_d^2 n_{km} \varepsilon_{km} \delta_{mj} R_{ij} + p_d(1-p_d)n_{mk} \varepsilon_{mk} \delta_{mj} R_{ij} \end{aligned}$$

Usando las condiciones anteriores nos queda:

$$\begin{aligned} \frac{\mu}{\lambda} \varepsilon_i^* &= \sum_{m,l,j=1}^N p_d(1-p_d)n_k R_{km} \varepsilon_k^* \delta_{mi} R_{ij} + (1-p_d)^2 n_m R_{mk} \varepsilon_m^* \delta_{mi} R_{ij} + \\ &+ p_d^2 n_k R_{km} \varepsilon_k \delta_{mj} R_{ij} + p_d(1-p_d)n_m R_{mk} \varepsilon_m \delta_{mj} R_{ij} \end{aligned}$$

Teniendo cuatro términos en la parte derecha, vamos a desarrollarlos uno a uno.

- Primer término:

$$\begin{aligned}
\sum_{m,l,j=1}^N p_d(1-p_d)n_k R_{km} \varepsilon_k^* \delta_{mi} R_{ij} &= p(1-p) \sum_k n_k \varepsilon_k \sum_{j,m=1}^N \delta_{mi} R_{km} R_{ij} = \\
&= p(1-p) \sum_k n_k \varepsilon_k \sum_{j=1}^N R_{ki} R_{ij} = p(1-p) \sum_{k=1}^N R_{ki} n_k \varepsilon_k \sum_{j=1}^N R_{ij} = \\
&= \sum_{k=1}^N p(1-p) R_{ki} n_k \varepsilon_k = \sum_{j=1}^N p(1-p) R_{ji} n_j \varepsilon_j
\end{aligned}$$

- Segundo término:

$$\begin{aligned}
\sum_{m,l,j=1}^N (1-p_d)^2 n_m R_{mk} \varepsilon_m^* \delta_{mi} R_{ij} &= (1-p)^2 \sum_{m=1}^N n_m \varepsilon_m \delta_{mi} \sum_{k=1}^N R_{mk} \sum_{j=1}^N R_{ij} = \\
&= \sum_{m=1}^N (1-p)^2 n_m \varepsilon_m \delta_{mi} = \sum_{j=1}^N (1-p)^2 n_j \delta_{ji} \varepsilon_j
\end{aligned}$$

- Tercer término:

$$\begin{aligned}
\sum_{j=1}^N p_d^2 n_k R_{km} \varepsilon_k \delta_{mj} R_{ij} &= p^2 \sum_{j,k=1}^N n_k R_{ij} \varepsilon_k \sum_{m=1}^N R_{km} \delta_{mj} = \\
&= p^2 \sum_{k=1}^N n_k \left(\sum_{j=1}^N R_{ij} R_{kj} \right) \varepsilon_k = \sum_{j=1}^N p^2 n_j \left(\sum_{l=1}^N R_{il} R_{jl} \right) \varepsilon_j
\end{aligned}$$

- Cuarto término:

$$\begin{aligned}
\sum_{m,l,j=1}^N p_d(1-p_d)n_m R_{mk} \varepsilon_m \delta_{mj} R_{ij} &= p(1-p) \sum_{j,k=1}^N n_j R_{jk} R_{ij} \varepsilon_j = \\
&= \sum_{j=1}^N p(1-p) n_j R_{ij} \varepsilon_j \sum_{k=1}^N R_{jk} = \sum_{j=1}^N p(1-p) n_j R_{ij} \varepsilon_j
\end{aligned}$$

Tras juntar el primer término con el cuarto, hemos acabado y recuperado las ecuaciones del caso indistinguible (17).