

Pablo Valgañón Ruiz

Modelización y control de epidemias a través de la movilidad urbana

Director/es

Soriano Paños, David
Gómez Gardeñes, Jesús

<http://zaguan.unizar.es/collection/Tesis>



Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606



Universidad
Zaragoza

Tesis Doctoral

MODELIZACIÓN Y CONTROL DE EPIDEMIAS A TRAVÉS DE LA MOVILIDAD URBANA

Autor

Pablo Valgañón Ruiz

Director/es

Soriano Paños, David
Gómez Gardeñes, Jesús

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

2025

Tesis Doctoral

Modelización y control de epidemias a través
de la movilidad urbana

Autor

Pablo Valgañón Ruiz

Director/es

Jesús Gómez Gardeñes

David Soriano Paños

Universidad de Zaragoza

Escuela de Doctorado

Programa de Doctorado en Física

2025

Compendio de artículos

Esta tesis es un compendio formado por artículos publicados durante la duración del doctorado, en los cuales el candidato a doctor es coautor. Los cuatro artículos que componen el cuerpo de la tesis son los siguientes:

- P. Valgañón, D. Soriano-Paños, A. Arenas, and J. Gómez-Gardeñes, “Contagion–diffusion processes with recurrent mobility patterns of distinguishable agents”, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 32 no. 4, 2022.
- P. Valgañón, U. Lería, D. Soriano-Paños and J. Gómez-Gardeñes, “Socioeconomic determinants of stay-at-home policies during the first covid-19 wave”. *Frontiers in Public Health*, vol. 11, ISSN 2296-2565, 2023.
- P. Valgañón, A.F. Useche, D. Soriano-Paños, G. Ghoshal and J. Gómez-Gardeñes, “Quantifying the heterogeneous impact of lockdown policies on different socioeconomic classes during the first COVID-19 wave in Colombia”. *Scientific Reports*, vol. 13, no. 1, 2023.
- P. Valgañón, A.F. Useche, F. Montes, A. Arenas, D. Soriano-Paños and J. Gómez-Gardeñes, “Human behavior-driven epidemic surveillance in urban landscapes”. *npj Complexity* vol. 01, no. 21, 2024.

Agradecimientos

Nada de esto habría sido posible sin las personas que me han acompañado durante, antes y después de esta etapa. Por eso, en primer lugar quiero mostrar mi más sincera gratitud a mis directores de tesis. Jesús, dudo que mucha gente pueda fardar de haber creado un grupo como el nuestro. No solamente es extraordinaria tu capacidad como investigador, profesor y líder, sino también por el genial ambiente que has cultivado. Has conseguido que se sienta como una familia para todos los que hemos sido parte del grupo. David, no hay agradecimiento que pueda pagarte por todo el apoyo que me has dado desde el principio. Gracias por todo lo que me has ayudado a mejorar, como investigador y persona, y lo mucho que has estado pendiente de mí, realmente has sido como el hermano mayor del grupo. Espero volver a trabajar contigo e ir a muchos más congresos con todos. A ambos: no habéis hecho más que darme una oportunidad tras otra, no solamente durante el doctorado sino también después. Ojalá de aquí en adelante me traten y aprenda a tratar a otros la mitad de bien de lo que lo habéis hecho vosotros.

A mis padres, que habéis estado a mi lado y me habéis apoyado sin pedir absolutamente nada a cambio. Y cuando parece que no podéis hacer más, hacéis el doble. Si estoy aquí es porque os habéis volcado en darme la mejor educación posible, dentro de casa y fuera. Desde que era pequeño y mostré el más mínimo interés en la ciencia me habéis comprado libros, series, películas, me habéis llevado a exposiciones y museos... Os habéis esforzado tanto en que todo me fuera bien, y me habéis facilitado de tal forma conseguir todo lo que me he propuesto, que solo puedo pensar que me ha tocado la lotería. Doy gracias también a mis abuelos, que me han apoyado con todas sus fuerzas. No estaría aquí si no fuese por vosotros. A mi familia, a Tania y mis primos Omar, Ángela y Nora, y a José Luis, que me habéis tratado genial.

A todos mis amigos, que me habéis acompañado y soportado hasta ahora. Gracias, Álex, alias “*asainz*”, por estar a mi lado desde siempre. Sas, por las tardes de plática y *code review*. A mis cerbunos, Dennis, Raúl, Irene, Mónica, David, y las horas de cháchara con vosotros cada noche en la residencia. A los miembros

del Yugulé que conocí después, Carol, Cris, Sergio y Marco. Como no soy filólogo no puedo llegar a expresar lo divertido que ha sido vivir y pasar estos años con vosotros. Todo lo que hemos pasado juntos me ha hecho mejor persona, y me ha ayudado a escribir esta tesis. A mis compañeros de Taek y a Chabi, gracias por los entrenamientos que me han mantenido cuerdo durante años.

Finalmente, a los miembros de GOTHAM que me preceden, Alessio, Adriana, Franchó, Sergio, que me habéis inspirado y mostrado siempre apoyo, y a los que me estorban en el cluster, hugo, Santiago, y pagall, gracias por todas las discusiones, cafés y congresos en los que me habéis acompañado. También a profesores como Mario y Fernando, con quien también he compartido tantos cafés y conversaciones, o Álex Arenas, que ha guiado e informado en gran medida mi investigación, porque aunque no hayáis sido codirectores de esta tesis, es cierto que lleva parte de todos vosotros.

Resumen

La rápida proliferación de enfermedades de naturaleza contagiosa es objeto de un gran cuerpo de estudio, que ha crecido significativamente durante las últimas décadas y de forma más drástica tras la pandemia de COVID-19. Dos factores que cabe destacar son la creciente globalización, que conecta partes remotas de todo el mundo, y la creciente urbanización que permite la concentración de grandes cantidades de humanos en áreas reducidas. En ambos factores encontramos que la movilidad juega un papel crucial, por lo que entender cómo los desplazamientos de las personas afectan a la dispersión y crecimiento de brotes se vuelve prioritario en epidemiología. Esta tesis pone su atención en las dinámicas de contagio en paisajes urbanos, donde los viajes que realizan diariamente las personas pueden acelerar el contagio a niveles que no son capaces de predecir otros modelos metapoblacionales.

En la primera parte del cuerpo de la tesis se analiza el impacto de las medidas de contención a diferentes escalas, con el objetivo de comprender cómo las diferencias socioeconómicas de personas o comunidades afecta a la respuesta de las mismas. Utilizando datos obtenidos durante la primera ola de la pandemia de COVID-19, creamos un modelo que tiene en cuenta la movilidad registrada en diferentes países y replica el mecanismo de cuarentena para determinar la efectividad de los confinamientos en cada país. Por otro lado, estudiamos cómo los patrones de movilidad de personas en las ciudades de Bogotá, Medellín y Cali evolucionan, llegando a la conclusión de que tanto la reducción de movilidad alcanzada como la efectividad a la hora de frenar la epidemia están fuertemente correlacionadas con el nivel económico de la población.

En la segunda parte nos centramos en modelos dinámicos de propagación de enfermedades en metapoblaciones, probando que la mayor parte de los formalismos utilizados hasta ahora subestiman gravemente la capacidad de contagio de brotes en sistemas urbanos. El modelo presentado distingue los comportamientos propios de cada persona en lugar de suponer que todos comparten los mismos patrones de movilidad, lo que permite una mayor concentración de agentes infecciosos en los principales focos de contagio a lo largo del tiempo. Adicionalmente, el análisis de

las ecuaciones permite conocer la vulnerabilidad de las personas en función de sus lugares de destino, en lugar de exclusivamente sus residencias. Esto se utiliza más adelante a nuestro favor para poner a prueba estrategias de vigilancia epidémica que prioricen los grupos más susceptibles en caso de brote en diferentes ciudades, resultando en mayores detecciones tempranas y tiempos de alerta mucho más cortos. Finalmente, se pone a prueba la estrategia en simulaciones más realistas donde se utiliza la red de transporte del Transmilenio de Bogotá para fijar aquellas estaciones en las que focalizar recursos de vigilancia epidémica para la anticipación de brotes epidémicos.

Índice general

Agradecimientos	II
Resumen	IV
1. Introducción	1
1.1. Introducción general	1
1.1.1. Objetivos	6
1.2. Introducción a la epidemiología matemática	9
1.2.1. Epidemias en una población homogénea y estática	9
1.2.2. Etapa temprana de una epidemia	11
1.2.3. Modelos Compartimentales	17
1.2.4. Contagios en redes	24
2. Determinantes socioeconómicos del confinamiento durante la primera ola de COVID-19	40
3. El impacto de las políticas de cuarentena en las diferentes clases socioeconómicas a nivel urbano	57
4. Procesos de contagio y difusión con movilidad recurrente de agentes distinguibles	69
5. Control epidémico basado en el comportamiento humano en paisajes urbanos	82
6. Conclusiones	94
Apéndice	100
Bibliografía	102

Índice de figuras

1.1. Propagación exponencial de epidemias	10
1.2. Encuesta empírica de contactos	14
1.3. Tamaño final y máximo de infectados	22
1.4. Redes simétricas y asimétricas	25
1.5. Grafos aleatorios y libres de escala	29

Capítulo 1

Introducción

1.1. Introducción general

Es cierto que no podemos afirmar que el estudio sistemático y matemático de las epidemias haya sido llevado a cabo desde que existen las epidemias mismas. De hecho, durante la mayor parte de la historia las enfermedades contagiosas se consideraban una forma de castigo divino, sobre las que las acciones humanas tenían un efecto despreciable. Esta mentalidad dejó un vacío de unos dos mil años entre el trabajo de Hipócrates de Cos (460 a. C. - 370 a. C.) [1] y el nacimiento de la epidemiología moderna, fruto del asentamiento del método científico a partir del siglo XVII. Este gran vacío fue únicamente cubierto en parte durante la Edad de Oro del islam con los trabajos de intelectuales como el médico Ibn An-Nafis (1210 d. C. - 1288 d. C.) [2]. Sus esfuerzos establecieron que el impacto de las epidemias estaba determinado en parte por las variaciones del clima, la localización geográfica, las concentraciones de personas y las condiciones de higiene y pobreza. Sin embargo, no es hasta mucho más adelante que Anton van Leeuwenhoek [3,4] demuestra con el uso del microscopio la existencia de los microbios en 1675, a los que se les atribuiría, más adelante, la causa de diversas enfermedades infecciosas.

A pesar de estos avances de índole experimental, no fue hasta casi 100 años después cuando Daniel Bernoulli utiliza los primeros modelos matemáticos para estudiar el impacto de enfermedades. En particular, Bernoulli desarrolló en 1760 un modelo de la viruela que le permitiría estimar los efectos del equivalente a un precursor de la vacuna en la mortalidad de la población, calculando el cambio en la esperanza de vida en caso de proteger a la población frente a dicha enfermedad. Otro reconocido médico que ayudó a asentar los pilares de la epidemiología moderna fue John Snow, quien identificó el foco de un importante brote de cólera en Londres durante 1854 a partir de un análisis espacial en el cual relacionó

las infecciones con la localización de las diferentes fuentes de agua posiblemente contaminadas en la zona [5–7]. Además, Snow promovió la idea de que los brotes epidémicos terminaban extinguiéndose debido a la desaparición de las reservas de susceptibles, lo que contradecía la creencia de que la epidemia perdía fuerza de infección con el paso del tiempo. Finalmente, la teoría de los gérmenes, impulsada por figuras como Louis Paster y Robert Koch para explicar la transmisión de enfermedades contagiosas entre personas [8–10], permitió a Ronald Ross crear el primer modelo matemático que simula la dinámica de propagación de epidemias [11–15]. En él, Ross aplica la ley de acción de masas a las infecciones en poblaciones humanas a partir de ecuaciones diferenciales deterministas. La teoría propuesta por Ross no solamente describe la transmisión de la malaria entre mosquitos y humanos, sino que establece una nueva rama de la ciencia, denominada originalmente con el nombre de patometría *a priori*. Esta rama, actualmente conocida como *epidemiología matemática*, se basa metodológicamente en asumir un conocimiento causal previo, formular ecuaciones a partir de dichas suposiciones y contrastar los resultados obtenidos con los datos empíricos.

Todos los anteriores avances y, en particular, el nuevo enfoque empleado de Ross y Hamer, llevan finalmente a la contribución más influyente y citada de la temprana epidemiología: los modelos compartimentales de Kermack y McKendrick. A pesar de que su formalismo tiene muchas similitudes con el de Ross, y de que es probable que lo desarrollaran independientemente y de forma casi contemporánea, son estos los generalmente conocidos como los padres de la epidemiología matemática. Kermack y McKendrick publican varios artículos en los que generalizan las ecuaciones diferenciales a cualquier enfermedad que se transmite directamente a través del contacto entre individuos en una población y, asimismo, formalizan la definición del número básico de reproducción R_0 [16–20]. En estos estudios, se argumenta que, contrariamente a lo que se pensaba, las olas epidémicas no finalizan debido a una disminución en la fuerza de infección o al agotamiento de la población susceptible. En su lugar, se explica que los contagios crecen exponencialmente únicamente cuando la densidad de individuos susceptibles supera un umbral crítico y que, al reducirse esta densidad, los casos comienzan a decrecer [21]. Paralelamente, L. H. Reed y W. H. Frost [22] introdujeron modelos probabilísticos en tiempo discreto, sentando las bases de los enfoques analíticos.

Abierta la senda de epidemiología matemática, en la segunda mitad del siglo XX los científicos continúan expandiendo estos formalismos, desarrollando diferentes modelos compartimentales aplicados a un rango más amplio de enfermedades, y contrastando con resultados estadísticos, o *a posteriori*. De especial relevancia es

la introducción de modelos espaciales que tratan de capturar la propagación de enfermedades a lo largo de regiones geográficas, lo cual supuso un avance importante en la comprensión de cómo se expanden las epidemias a partir de brotes localizados [23–29]. A finales del siglo XX los avances en la capacidad computacional y su utilidad para las simulaciones de agentes [30–34] y modelos hacen resurgir la disciplina, y es a partir de los años noventa que el auge de otro campo revoluciona la epidemiología matemática.

En una disciplina tan fuertemente conectada a las matemáticas como es la Física siempre ha dominado el reduccionismo. El enfoque reduccionista considera que el objetivo prioritario de la ciencia debe ser el descubrimiento de las leyes fundamentales que gobiernan nuestro universo y defiende que, una vez encontradas, nos será posible extrapolarlas para explicar todos los fenómenos a escalas mayores como la adición de sus partes. Inevitablemente, una nueva filosofía científica gana momento a finales de siglo. Nuestro mundo está colmado de mecanismos y agentes que interactúan entre ellos. Desde los elementos que forman las proteínas en nuestras células, pasando por la intrincada red de neuronas en nuestro cerebro, la comunicación en las colonias de hormigas, los movimientos cohesivos de las bandadas de pájaros en vuelo o bancos de peces, hasta nuestra organización como humanos, nos encontramos mecanismos similares. Tenemos la certeza de que los sistemas que forman estos diversos elementos son más que simplemente la suma de sus partes [35, 36], debido a la emergencia de comportamientos, patrones o estructuras que no pueden ser explicadas de forma trivial a partir de las acciones individuales de sus componentes en ausencia de la intervención de un control central. Este paradigma, el emergentismo, defiende que estos fenómenos ontológicos que surgen a cada escala son resultado de la estructura de las interacciones entre objetos, más allá de las propiedades individuales de los mismos.

El alcance y precisión de enfoques tradicionales en estos contextos se ven limitados y la necesidad de recurrir a la ciencia de sistemas complejos se vuelve evidente. Este campo es altamente interdisciplinar debido a que atraviesa la frontera de lo que se considera física, entrando en áreas dentro de la biología [32, 37], medicina [38] o sociología [39, 40], utilizando un amplio repertorio de herramientas, incluyendo la mecánica estadística, teoría del caos, de la información y de grafos. Esta última da lugar a la ciencia de redes, que aplica esta teoría a redes de gran tamaño que podemos encontrar en contextos del mundo real. Fueron pioneros Watts y Strogatz, que encontraron el comportamiento de *mundo pequeño* [41, 42] en redes de contactos extensas, y destacan más tarde los trabajos de Barabási y Albert, que llegaron a la conclusión de que en el mundo real, las redes son extre-

madamente heterogéneas o *libres de escala* [43, 44].

La integración de esta ciencia de redes en la epidemiología matemática permitió a Vespignani y Pastor-Satorras demostrar que en estos tipos de redes del mundo real, el umbral epidémico se vuelve alarmantemente pequeño [45, 46]. Tras incontables avances en el estudio de contagios en redes de contactos [47–56], los modelos metapoblacionales [57–60] empiezan a incorporarse con el objetivo de llevar la epidemiología a una escala espacio-temporal mayor y proporcionarle mayor eficiencia en escenarios epidémicos reales. Es aquí cuando se empiezan a incorporar los efectos de los desplazamientos de personas en el estudio de cómo se propagan las epidemias a diferentes escalas [61–69], teniendo en cuenta la variabilidad y heterogeneidad características del comportamiento humano. Desde entonces, el crecimiento exponencial de capacidad computacional y la acumulación de grandes cantidades de datos de diversos tipos (dispositivos móviles, geolocalización, redes sociales o encuestas) han concedido a los científicos innumerables herramientas para el desempeño de su trabajo.

Es justo en este punto de desarrollo de la epidemiología matemática donde esta tesis encuentra su lugar y motivación. En particular, nos centraremos en el campo de la movilidad humana y su fuerte interconexión con la dinámica de propagación de epidemias que, lejos de ser un simple proceso biológico, forma un sistema complejo con profundas implicaciones para la salud pública, las dinámicas sociales y la economía a escala global. La humanidad siempre se ha definido por el movimiento, desde los cazadores-recolectores nómadas hasta los migrantes y viajeros actuales que hacen uso de los sistemas de transporte modernos. Es la movilidad la que da forma a las civilizaciones, ha permitido el comercio y ha impulsado el progreso de las culturas, pero también ha permitido la aparición y propagación de epidemias desde que la humanidad comenzó a concentrarse en núcleos urbanos. Por ejemplo, la peste negra que asoló Europa se propagó a través de las rutas comerciales marítimas y terrestres que la conectaban con Asia [70]. De forma similar, la viruela fue exportada desde Europa a América y diezmó las poblaciones indígenas, que no habían desarrollado inmunidad en su contra [71]. En la era contemporánea, incontables casos como el dengue [72–74], la malaria [75, 76], el ébola [77, 78] o la reciente pandemia de COVID-19 [79, 80] nos llevan a pensar que, con la creciente conectividad entre personas de cualquier lugar del mundo y otros factores como el cambio climático [81–84], es previsible e incluso inevitable que la frecuencia de brotes epidémicos que se extienden a escala global aumente con el tiempo.

La ausencia de tratamientos efectivos o vacunas contra el virus contagioso cau-

sante de la enfermedad, el SARS-CoV-2, puso el foco en otra clase de medidas e intervenciones no farmacológicas que tenían el objetivo de reducir el impacto de la crisis [85]. Estas respuestas se centran en diagnosticar el riesgo de brotes epidémicos y llevar a cabo cierres de fronteras, cuarentenas, confinamiento perimetral y otras formas de reducción de la movilidad o aislamiento de personas en riesgo. La mayor parte de los gobiernos se vieron entonces obligados a determinar qué medidas eran pertinentes para no solamente frenar el avance de la enfermedad, sino también evitar el impacto negativo que pueden suponer para las libertades individuales de las personas y el funcionamiento de la economía. Esto llevó a un amplio abanico de respuestas y, por tanto, resultados en diferentes partes del mundo [65, 86–90].

En retrospectiva, sería fácil afirmar que conocemos las decisiones que deberían haber sido tomadas durante la pandemia de COVID-19 y que, en consecuencia, no cometeremos los mismos errores en ocasiones posteriores. Sin embargo, no es tan sencillo. Tal y como hemos establecido anteriormente, la dinámica de propagación de enfermedades contagiosas está interconectada con el comportamiento humano, por lo que los efectos de diferentes medidas no son en absoluto triviales, y es necesario un estudio sistemático y en profundidad de cada caso. El reto de entender los mecanismos de propagación de epidemias y las consecuencias de diferentes medidas debe ser abordado desde el punto de vista de los sistemas complejos, en los que los patrones de comportamiento humano y contagio puedan ser estudiados conjuntamente [91]. Es especialmente importante resaltar la imposibilidad de predecir exactamente el curso de una curva epidémica, que está implícita en el factor humano. La razón no es necesariamente la misma por la que se vuelve más difícil predecir fenómenos como la meteorología con el tiempo, donde las propias ecuaciones son inherentemente caóticas. Por una parte, el proceso físico de contagio tiene una naturaleza altamente estocástica, haciendo que haya que recurrir a modelos puramente probabilísticos. Por otra parte, el propio hecho de publicar resultados en los que se predice una gran prevalencia puede llevar a una reacción que altere la evolución de la curva [92, 93]. En esta reacción distinguimos principalmente dos actores: las autoridades responsables, que establecen las leyes o medidas pertinentes para prevenir el aumento de casos, y los individuos, los cuales pueden o no cambiar su comportamiento (aislamiento, utilización de profilácticos como mascarillas, o priorización de la higiene), ya sea por cuenta propia o siguiendo las directrices de dichas autoridades [94–96]. Esto hace que en la mayor parte de los casos los modelos predictivos deban ser flexibles, y su principal objetivo pase a ser el estudio de los efectos de diferentes medidas y describir los escenarios más probables, con la finalidad última de informar, educar y aconsejar tanto a las

autoridades como al público.

En esta situación, es de esperar que la respuesta de la población no sea homogénea, sino que dependa en gran medida de la relación entre la clase socioeconómica a la que pertenece cada persona y su capacidad de cambiar su comportamiento. Por ejemplo, ciertos grupos y comunidades pueden ser más propensos a vacunarse y usar mascarillas, o tienen la posibilidad de trabajar desde casa y dejar de hacer viajes recurrentes. De la misma forma, las medidas que se aplican sobre una población tienen efectos muy diferentes en cada comunidad en otros aspectos. Las órdenes de aislamiento y cuarentena pueden hacer que problemas existentes de segregación y desigualdades socioeconómicas empeoren [97–99]. El análisis de los efectos de las políticas de confinamiento, tanto en la respuesta humana a diferentes escalas como en su éxito a la hora de combatir la epidemia de la COVID-19 es de esta manera uno de los principales ejes en los que se centra esta tesis. Al mismo tiempo, dada la importancia de abordar los brotes epidémicos con el mínimo impacto posible en la libertad de movimiento de las personas y el desarrollo económico de una región, está en nuestro interés procurar que las medidas que se tomen se centren en los grupos más vulnerables y en los lugares donde es más probable que se des controle la epidemia. Utilizando nuevos formalismos metapoblacionales, esta tesis muestra cómo analizar datos demográficos y de patrones de movilidad en entornos urbanos para descubrir qué flujos de personas contribuyen más a la proliferación de enfermedades y, a partir de este conocimiento, elaborar estrategias de control con precisión quirúrgica que detecten los brotes con antelación y les pongan freno.

1.1.1. Objetivos

Esta tesis fue concebida con la intención de crear y estudiar modelos teóricos capaces de caracterizar de una forma más exacta la movilidad humana, para replicar con mejor precisión la propagación de epidemias en sistemas urbanos. Dicho esfuerzo fue motivado por el hecho de que gran parte de los modelos metapoblacionales son imperfectos, hacen asunciones que llevan a conclusiones poco exactas o no tienen en cuenta características y cualidades esenciales del comportamiento humano. La llegada de la pandemia de COVID-19 y sus graves consecuencias, sin embargo, permitieron la obtención de cantidades impensables de información que trasciende la epidemiología matemática. Por ejemplo, la imposición de intervenciones no farmacológicas como confinamientos obligatorios puede haber tenido más o menos éxito dependiendo de muchos factores, pero gracias a la digitalización de

nuestro entorno y la facilidad de obtención de datos a grandes escalas, el análisis de la efectividad de dichas intervenciones es posible.

Por esta razón el primer objetivo de la tesis se centra en analizar las consecuencias de las medidas no farmacológicas, tanto en los cambios de comportamiento de las personas como en su éxito a la hora de frenar el aumento de casos, en función de factores socioeconómicos. El segundo objetivo se convierte entonces en aplicar los conocimientos obtenidos del análisis para crear un modelo metapoblacional de propagación epidémica que nos permita manufacturar estrategias de control precisas y efectivas, capaces de diagnosticar con rapidez los posibles focos de un brote epidémico y ponerle fin.

Estos dos objetivos generales se materializan en cuatro artículos que constituyen los capítulos centrales de esta tesis y cuyos objetivos específicos detallamos a continuación:

- En el Capítulo 2 se recogen datos de movilidad y defunciones a causa de COVID-19 de cuarenta y dos países. Utilizando un modelo epidémico sencillo en aproximación de campo medio e introduciendo un parámetro de permeabilidad, que representa la dificultad de la población para aislarse a pesar de reducir su actividad y movilidad, se relaciona la eficacia de las cuarentenas con el producto interior bruto de cada país.
- En el Capítulo 3 nos vamos a una escala intraurbana, utilizando información de movilidad de alta resolución en Bogotá, Medellín y Cali a lo largo del tiempo para comparar la evolución de los patrones de movilidad de personas pertenecientes a diferentes niveles socioeconómicos. El artículo tiene la intención de mostrar cómo las medidas de confinamiento tienen un efecto diferente dependiendo de la clase económica de la población en la que se aplica, y su relación a la segregación o inseguridad económica.
- En el Capítulo 4 se vuelve a los modelos teóricos, abordando un problema que tienen la mayoría de formalismos que replican los viajes de personas a nivel urbano: la recurrencia. Mientras otros trabajos consideran que el movimiento de las personas es aleatorio, nuestro formalismo asume que los viajes que realiza cada persona se centran en gran medida en dos lugares, la residencia y el lugar de destino habitual. El artículo desarrolla el formalismo matemáticamente y muestra los fenómenos que se pueden explicar gracias a

tener en cuenta la recurrencia del comportamiento humano.

- En el Capítulo 5, se utiliza el anterior formalismo matemático para asesorar posibles estrategias de control que puedan ser de utilidad en caso de nuevos brotes epidémicos en paisajes urbanos. Se hace evidente que ciertos flujos de movilidad, es decir, grupos de personas que viajan recurrentemente entre dos lugares, son responsables de la mayor parte del contagio. Nuestro objetivo es descubrir cuáles son los flujos críticos y centrar los recursos en los grupos de personas que pertenezcan a los mismos para detectar brotes epidémicos con antelación y frenarlos. Adicionalmente, buscamos una forma de distribuir los recursos a través de la red del Transmilenio de la ciudad de Bogotá, y demostrar que esta solución es plausible.

Además de estos capítulos, y como prolegómeno necesario a los resultados expuestos en los artículos, en el siguiente capítulo recapitularemos los métodos y formalismos esenciales para entender los capítulos 3, 4, 5 y 6. Asimismo, como colofón, tras la exposición de los artículos, añadimos un breve capítulo dedicado a discutir las conclusiones globales del trabajo desarrollado.

1.2. Introducción a la epidemiología matemática

Tanto la epidemiología matemática como la mayoría de otras aplicaciones de la ciencia de sistemas complejos lidian con situaciones demasiado complejas como para ser replicadas con exactitud. Esto obliga a realizar numerosas simplificaciones que reduzcan el sistema a su estructura esencial hasta dejarlo *en los huesos* (*bare-boned*) para, después, aumentar la sofisticación de forma gradual y, si es posible, controlada. Siguiendo esta estrategia, en esta sección se construirán gradualmente los formalismos usados en los trabajos realizados en esta tesis partiendo de los principios básicos de la epidemiología matemática y progresivamente incorporaremos detalles adicionales que nos acerquen a modelos más realistas.

1.2.1. Epidemias en una población homogénea y estática

Las poblaciones de personas susceptibles a enfermedades de carácter contagioso suelen tener un cierto tiempo característico de reemplazo, es decir, una generación muere y da lugar a la siguiente. Dependiendo de la naturaleza de la epidemia que estemos estudiando y la fase en la que nos encontremos, esto puede tener una relevancia clara o ser despreciable. Dado que en esta tesis nos centramos en la propagación de enfermedades causadas por patógenos cuyo tiempo de residencia en cada hospedador comprende una ventana temporal típica de días o semanas durante la que puede ser transmitido a otros hospedadores, podemos ignorar el relevo generacional que se da a lo largo de décadas. De esta forma, asumiremos que en nuestros modelos la población es constante y de tamaño N . También consideraremos que, en el inicio, todos los huéspedes son susceptibles y no han estado en contacto con el agente infeccioso, con la excepción del primer individuo (al que nos referimos como semilla) en introducirlo.

Uno de los pasos más importantes para la formulación de modelos matemáticos de epidemias es la elección del mecanismo de transmisión de las mismas. En nuestro contexto suponemos que el agente es un microorganismo que se transmite exclusivamente de humano a humano y que, además, lo hace principalmente de forma aérea. De esta forma, cada vez que una persona infecciosa tiene un contacto directo con otra susceptible, existe una probabilidad λ de que se infecte la segunda, la cual contraerá la enfermedad.

Para completar la descripción matemática de una epidemia, se debe concretar la forma en que ocurren los contactos y los agentes involucrados en ellos. En la fase inicial de una epidemia, podemos considerar como despreciable el número de personas que han pasado por la enfermedad. De esta manera, todos los contactos que

una persona infecciosa tiene ocurrencias con susceptibles. Suponiendo una frecuencia de contactos constante dada por k , el individuo medio tendrá un contacto con kT susceptibles a lo largo de su periodo infeccioso, T . Se define el número reproductivo básico R_0 como el número de casos secundarios que un agente infeccioso produce antes de recuperarse cuando $S(0) = N$, y se puede calcular como:

$$R_0 = k\lambda T; . \quad (1.1)$$

Como se verá más adelante, no siempre se considera una infectividad λ constante a lo largo de un periodo T fijo, sino que varía en el tiempo para cada individuo y, en consecuencia, cabe introducir una definición más general de R_0 . Dejando de lado estas asunciones, es evidente cómo este factor es crítico en la propagación de epidemias, y que hay un valor umbral debajo del cual la epidemia desaparece de forma casi inmediata, al ser $R_0 \leq 1$. Por el contrario, si cada individuo consigue transmitir la enfermedad a más de una persona durante su periodo infeccioso, se cumple $R_0 > 1$ y existe la posibilidad de que se propague de forma exponencial con cada generación, tal y como se muestra en la figura 1.1. Llamamos *umbral epidémico* al valor de la infectividad λ_c para la cual $R_0 = 1$, lo que en este caso particular implica $\lambda_c = (kT)^{-1}$.

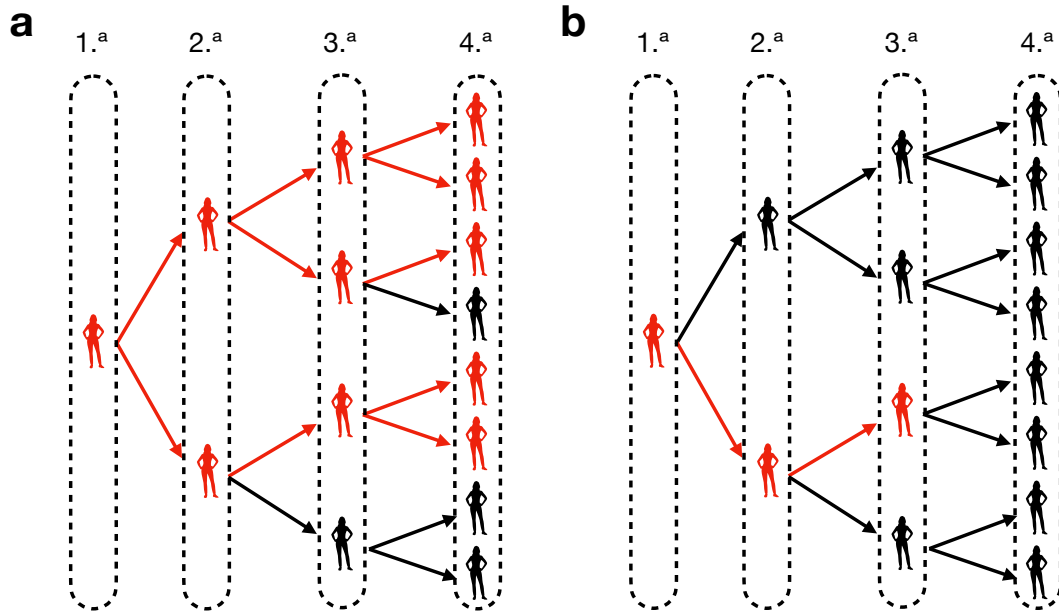


Figura 1.1: Representación de un proceso de contagio en el que, en un tiempo T , cada individuo tiene dos contactos. El umbral epidémico es en este caso $\lambda_c = 0.5$, según la ecuación 1.1. En cada generación, los individuos que no han contraído el agente infeccioso están representados en color negro, y los que sí lo han contraído en rojo. (a) Crecimiento exponencial de los contagios para una infectividad tal que $R_0 > 1$, y (b) extinción del agente infeccioso para $R_0 \leq 1$.

1.2.2. Etapa temprana de una epidemia

Crecimiento del brote

Es posible encontrar la evolución temporal del número de infectados en un brote epidémico entendiendo que una persona se puede contagiar al tener contactos con individuos que se han infectado con anterioridad. En el caso más general, la infectividad de los agentes depende del tiempo que ha pasado desde que se contagiaron, definiendo $A(\tau)$ como la fuerza de infección de los agentes tras un tiempo τ [100–102]. De esta forma, la incidencia se puede calcular como la contribución de todas las personas contagiadas anteriormente:

$$i(t) = \int_0^\infty A(\tau)i(t-\tau)d\tau . \quad (1.2)$$

Si, al igual que antes, se considera una infectividad λ constante durante un periodo de infección T tal que

$$A(\tau) = \begin{cases} k\lambda & \text{si } 0 \leq \tau < T \\ 0 & \text{en otro caso} \end{cases} . \quad (1.3)$$

la ecuación 1.2 queda

$$i(t) = k\lambda \int_0^T i(t-\tau)d\tau , \quad (1.4)$$

lo que permite una solución exponencial de la forma $i(t) = i(0)e^{rt}$, donde la tasa de crecimiento r debe cumplir la ecuación

$$r = k\lambda (1 - e^{-rT}) , \quad (1.5)$$

Esta solución no permite una expresión explícita de dicha tasa, pero es posible obtener su valor numéricamente. Aunque es común asumir que un R_0 alto conlleva un crecimiento muy rápido de la epidemia, este no es siempre el caso. Por ejemplo, enfermedades como el VIH lidian con escalas temporales de años o décadas, por lo que aunque se calcule un R_0 alto la tasa de crecimiento sigue siendo pequeña en intervalos de semanas o meses.

Efectos estocásticos

A pesar de su uso generalizado, las ecuaciones deterministas no tienen en cuenta la naturaleza estocástica de la propagación. Asimismo, estas ecuaciones ignoran que, de forma aleatoria, los infectados se pueden recuperar más rápido de lo esperado, haciendo desaparecer la epidemia a pesar de que la tasa de crecimiento sea

positiva.

Para abordar este problema, la cantidad de casos secundarios que cada agente puede desencadenar se considera una variable aleatoria que obedece una distribución de probabilidad dada. Así, definiendo m_i como la probabilidad de generar i casos nuevos, lo que debe cumplir $\sum_{i=0}^{\infty} m_i = 1$, tenemos que:

$$R_0 = \sum_{i=0}^{\infty} i m_i. \quad (1.6)$$

En este contexto se define además z como la probabilidad de que el brote creado por un único individuo se extinga antes de causar una epidemia a gran escala. En una población bien mezclada y asumiendo que los brotes causados por diferentes individuos son independientes, la probabilidad de que partiendo de i agentes contagiosos la epidemia no sobreviva es z^i . Esto permite calcular la probabilidad z como

$$z = \sum_{i=0}^{\infty} m_i z^i = g(z), \quad (1.7)$$

lo que define una función generatriz g que será útil más adelante dado que es fácil comprobar que $g'(1) = R_0$. Considerando los contagios como eventos aleatorios e independientes con frecuencia $k\lambda$, m tiene una distribución de probabilidad de Poisson para un tiempo infeccioso T , dada por

$$m_i(T) = \frac{(k\lambda T)^i}{i!} e^{-k\lambda T}, \quad (1.8)$$

que se cumple al asumir λ constante. Insertando este resultado en la ecuación 1.7, y resolviendo la serie se obtiene

$$g(z) = e^{R_0(z-1)}. \quad (1.9)$$

La ecuación $g(z) = z$ tiene como raíz la probabilidad de que se extinga el brote creado por una semilla de un individuo. La asunción de un periodo de incubación fijo T no lleva a una expresión explícita de este resultado, por lo que en la sección 1.2.3 se exploran otras asunciones que simplifican mucho el tratamiento analítico de las ecuaciones, especialmente para las fases más avanzadas de una epidemia.

Definición de R_0 y el umbral epidémico en una población estructurada

Para definir el número de reproducción básico R_0 , partimos de la ecuación 1.2, cuya versión completa tiene en cuenta que la cantidad de agentes susceptibles es finita:

$$i(t) = \frac{S(t)}{N} \int_0^\infty A(\tau) i(t - \tau) d\tau . \quad (1.10)$$

Cabe destacar que la fuerza de infección puede incluir de forma implícita tanto la infectividad de un individuo contagioso (por ejemplo, la evolución de la carga viral que expulsan) como la probabilidad de que sea eliminado del sistema (ya sea por recuperarse o fallecer). En estas ecuaciones deterministas, la métrica importante es la cantidad de nuevos casos que pueden generar los infectados. Debido a esto, es equivalente afirmar que tras un tiempo τ un 50 % de ellos haya fallecido, o que todos continúan estando infectados pero son un 50 % menos contagiosos. Es común separar esta fuerza de infección en

$$A(\tau) = k\lambda(\tau)\mathcal{F}(\tau) \quad (1.11)$$

siendo $\mathcal{F}(\tau)$ el porcentaje de infecciosos no eliminados y $\lambda(\tau)$ su infectividad. La ecuación 1.10 calcula la incidencia basada en las contribuciones de los individuos contagiados anteriormente. Los contagios del primer individuo infectado en particular, que definen el número de reproducción básico, se pueden calcular de forma sencilla tal que

$$R_0 = \int_0^\infty A(\tau) d\tau . \quad (1.12)$$

Para ciertos valores de $A(\tau)$, es sencillo encontrar los casos secundarios. Por ejemplo, con la definición de $A(\tau)$ en la expresión 1.3, se obtiene inmediatamente que $R_0 = k\lambda T$.

Las asunciones de población completamente homogénea y un estado epidémico único han permitido simplificar muchos cálculos, pero la mayoría de modelos deben poder ser generalizados para romper estas restricciones. Para ello, se declara que cada individuo puede encontrarse en un estado ξ de todo el espacio de posibles estados Ω , ya sean discretos o continuos (asumiremos continuos por el momento). Estos estados pueden representar diversos factores como la edad de las personas, el género, su localización geográfica, etcétera. Podemos entonces agrupar a la población de acuerdo a sus estados respectivos. Se define $S(\xi)$ como la densidad de individuos susceptibles con estado ξ (en el caso continuo, mientras que en el discreto sería la cantidad total de susceptibles en el estado ξ).

Al estimar el valor de R_0 , el primer paso lógico es tratar de calcular cuántos casos secundarios generará una persona. En una población heterogénea, esta persona no interactuará de la misma forma con cada uno de los grupos pertenecientes a los diferentes estados. Por ejemplo, en el caso de enfermedades de transmisión aérea es más probable que los niños contagien a otros niños en el colegio por encima de otros grupos de edad, habiendo una fuerza de interacción mayor entre estados similares tal y como se puede ver en la figura 1.2, dada por [103].

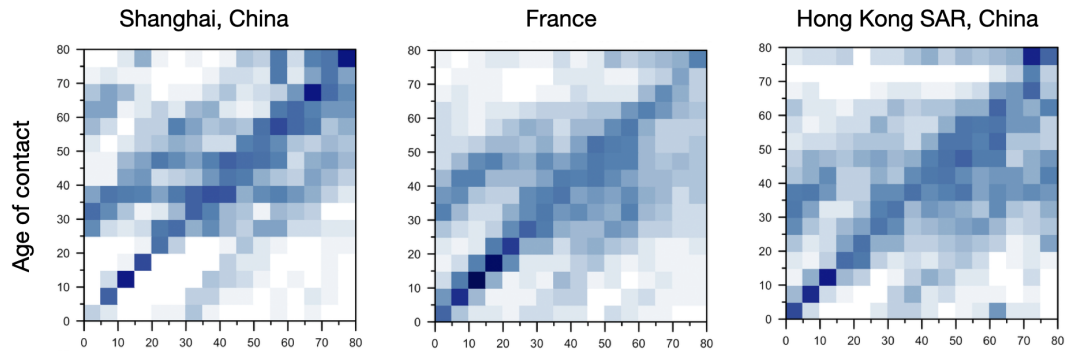


Figura 1.2: Encuesta empírica de la matriz de contactos entre diferentes grupos de edad en tres regiones geográficas. Colores más intensos (claros) denotan una mayor frecuencia de contactos. Imagen por [103], con datos de Zenodo, disponibles en https://zenodo.org/communities/social_contact_data/

Esto implica la necesidad de tratar con una fuerza de infección $A(\tau, \xi, \eta)$ entre un infectado en el estado η y un susceptible en el estado ξ . Aunque presentamos el tiempo de infección τ de forma explícita, en realidad puede ser parte del estado del individuo infeccioso. La cantidad de casos secundarios que podemos esperar entonces de un individuo infectado del estado η es

$$\int_{\Omega} \frac{S(\xi)}{N} d\xi \int_0^{\infty} A(\tau, \xi, \eta) d\tau . \quad (1.13)$$

El problema de esta expresión es obvio: no devuelve un valor de R_0 válido por dos razones. Primero, depende del estado del individuo del cual partimos. Segundo, este individuo generará casos con estados diferentes a sí mismo. Esto significa que cada generación no producirá casos secundarios con una distribución similar a la anterior. Se asumirá por este motivo que no partimos un individuo con un estado fijo, sino de un agente abstracto que sigue una distribución en el espacio de estados $\phi(\eta)$. Los casos generados por este individuo tendrán una distribución sobre los

estados ξ , y vienen dados por

$$(K(S)\phi)(\xi) = \frac{S(\xi)}{N} \int_{\Omega} \int_0^{\infty} A(\tau, \xi, \eta) \phi(\eta) d\tau d\eta, \quad (1.14)$$

definiendo así el operador de *next-generation* $K(S)$ [104], que actúa sobre una distribución (o vector) de estados $\phi(\eta)$ y devuelve la de la siguiente generación (al trabajar con estados discretos, la integral se interpretará como una suma).

Consideramos $\|\phi\| = 1$, al trabajar con un individuo con una distribución de probabilidad a lo largo de diferentes estados, pero físicamente se puede interpretar como la distribución de infectados en la población cuando $\phi(\eta)$ tenga un valor absoluto mayor que uno. Aún así, partiendo de un individuo, se tiene entonces que el número de casos secundarios total es $\|K(s)\phi\|$, o

$$\int_{\Omega} \frac{S(\xi)}{N} \int_{\Omega} \int_0^{\infty} A(\tau, \xi, \eta) \phi(\eta) d\tau d\eta d\xi, \quad (1.15)$$

llamado el factor de *next-generation*. De nuevo, R_0 puede tener diferentes valores dependiendo de la distribución de individuos de la que se ha partido, pero se cumple que su valor puede ser, como máximo, el elemento supremo del factor respecto de η , o la norma del operador $K(S)$. Es así razonable considerar definir de esa forma R_0 , como la mayor cantidad posible de casos secundarios causada por una distribución, pero queda por resolver el segundo problema: si los casos secundarios tienen una distribución con un factor diferente, este valor de R_0 no es consistente.

Por este motivo, la forma más exacta de calcular R_0 es tomar el número suficiente de generaciones tras el cual, en promedio, el número de casos se ve multiplicado por R_0 tras cada generación. En otros términos, si después de m generaciones la distribución de infectados es $K(S)^m \phi$, y en cada generación los infectados han crecido en promedio $\|K(S)^m\|^{1/m}$, se concluye:

$$R_0 = r(K(S)) = \lim_{m \rightarrow \infty} \|K(S)^m\|^{1/m}, \quad (1.16)$$

siendo $r(K(S))$ el radio espectral del operador K y considerando un número muy elevado de generaciones. El teorema de Perron-Frobenius establece que, para un operador irreducible y no negativo, el radio espectral es real y se corresponde con el mayor autovalor del operador, el cual tiene asociado un único autovector cuyas componentes son positivas. Es posible hacer que esto se cumpla, debido a que físicamente los estados que tenemos en cuenta son de cantidades positivas de personas (o probabilidades) y sus contribuciones siempre dan lugar a nuevos casos.

Si no se cumple la irreducibilidad, es posible separar nuestra población en varias subpoblaciones independientes unas de otras, con sus respectivos valores de R_0 , las cuales nunca llegan a interactuar.

Como un ejemplo práctico consideraremos estados discretos y finitos, en este caso la edad de las personas agrupadas en intervalos de un año, desde cero hasta noventa y nueve (dependerá de los datos disponibles, y como la edad no tiene un límite superior se pueden agrupar todas las personas que estén por encima en el último intervalo). Tenemos una matriz de contactos $C_{e,f}$ que determina cuánto interactúan los agentes de edad e con los de edad f . Se asume también que la carga vírica es constante, λ , durante un tiempo T fijo. Esto significa que

$$A(\tau, e, f) = \begin{cases} \lambda C_{e,f} & \text{si } 0 \leq \tau < T, \\ 0 & \text{en otro caso.} \end{cases} \quad (1.17)$$

Adicionalmente, la pirámide de población viene dada por S_e . Partiendo de esta información, la siguiente generación de infectados de cada grupo de edad, i_e , está dada por la contribución de los infectados de los demás grupos i_f , de la forma

$$i_e = \frac{S_e}{N} \int_0^T \lambda \sum_{f=0}^{99} C_{e,f} i_f d\tau = \frac{\lambda T}{N} \sum_{f=0}^{99} S_e C_{e,f} i_f. \quad (1.18)$$

Entonces, según la ecuación 1.16,

$$R_0 = \frac{\lambda T}{N} \Lambda^d(S_e C_{e,f}).$$

Las componentes del vector asociado al autovalor dominante, Λ^d , representan las proporciones de infectados por edad que esperamos encontrar al inicio de una epidemia en esta población, ya que tras varias iteraciones todos los vectores (sin componentes negativas) tienden a converger al del autovalor dominante.

Cabe recordar que el umbral epidémico λ_c , que es la infectividad mínima necesaria para que pueda existir una epidemia, se puede definir como el valor de la infectividad para el cual $R_0 = 1$. Más adelante se verá la aplicación de los operadores *next-generation* a modelos con los que es más sencillo trabajar analíticamente, además de otras formas de calcular el umbral epidémico independientemente de R_0 .

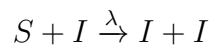
1.2.3. Modelos Compartimentales

Principales modelos

Los modelos compartimentales describen un conjunto de estados en los que se puede encontrar un hospedador y la forma en la que se transita de un estado a otro. La principal característica de las ecuaciones comúnmente utilizadas por estos modelos es que su dinámica es plenamente markoviana. Al asumir que los individuos no tienen memoria del tiempo que llevan infectados, o en el estado en el que se encuentran, se puede llegar a conclusiones con más facilidad. En la mayor parte de los casos, existe al menos un estado susceptible (S), libre de la enfermedad pero capaz de contraerla, y uno infeccioso (I), en el que el individuo hospeda al microorganismo y puede transmitirlo a otros. Los más comunes son los que se explican a continuación.

Modelo SI

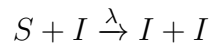
Este es el modelo más simple, donde los únicos dos estados son los descritos anteriormente: el susceptible (S) y el infeccioso (I). Dadas las asunciones anteriores, λ es de nuevo la probabilidad de que el susceptible contraiga la enfermedad al tener un contacto directo con un infeccioso.



Enfermedades como el sida, causado por el virus de inmunodeficiencia humana, pueden seguir este diagrama porque el virus reside permanentemente en el individuo contagiado.

Modelo SIR

Aquí el mecanismo de contagio es el mismo, pero tras contraer la enfermedad los individuos infecciosos se recuperan con frecuencia μ , siguiendo los diagramas:



En general, este modelo se aplica a un amplio rango de enfermedades, en las que se puede interpretar que el compartimento de los recuperados (R) representa tanto la defunción a causa de la infección, como el desarrollo de inmunidad a largo plazo por parte del individuo. En este modelo, existe un único estado estacionario en el que el número de agentes infecciosos $i(t) = 0$ cuando $t \rightarrow \infty$.

Modelo SIS

Se aplica esta descripción cuando los individuos infectados no desarrollan inmunidad a largo plazo al microorganismo responsable, y por lo tanto pueden contraer la enfermedad cualquier número de veces. De esta forma, las transiciones observadas en el modelo son:

$$S + I \xrightarrow{\lambda} I + I$$

$$I \xrightarrow{\mu} S$$

Dependiendo de la infectividad λ y la cantidad de contactos que tengan las personas, el estado estacionario puede ser de $i(t) = 0$ o $i(t) > 0$, al alcanzar el equilibrio entre infecciones y recuperaciones.

Modelos SEIR y SEIS

La mayoría de enfermedades tienen un periodo que sigue inmediatamente a la contracción del microorganismo, durante el cual el infectado es portador pero aún no es contagioso. Este estado tiene el nombre de expuesto (E) y una duración promedio η^{-1} , denominado periodo de incubación. De este modo, las transiciones de los eventos de contagio se modifican de la siguiente manera:

$$S + I \xrightarrow{\lambda} E + I$$

$$E \xrightarrow{\eta} I$$

En muchas ocasiones está en nuestro interés incluir este compartimento para replicar de forma más exacta el proceso por el que pasa la población en una epidemia, pero cualitativamente la dinámica es la misma que en los respectivos modelos SIR y SIS.

Ecuaciones diferenciales de los compartimentos

Estos modelos permiten escribir un conjunto de ecuaciones diferenciales que gobiernan una dinámica que sigue la ley de acción de masas, donde la tasa de generación de casos es proporcional a las concentraciones de susceptibles e infectados. A lo largo de la tesis nos centraremos principalmente en el SIR, con las

ecuaciones:

$$\dot{s} = -k\lambda si, \quad (1.19)$$

$$\dot{i} = k\lambda si - \mu i, \quad (1.20)$$

$$\dot{r} = \mu i. \quad (1.21)$$

En ellas, las variables s , i y r representan las proporciones de individuos susceptibles, infectados y recuperados, respectivamente, con respecto a la población total y cumpliendo $s + i + r = 1$. La variable k determina la frecuencia de contactos de cada individuo a lo largo del tiempo. Estas ecuaciones implican que la distribución del periodo de infección τ tiene la forma

$$Pr(\tau = T) = \mu e^{-\mu T}. \quad (1.22)$$

Es generalmente aceptado el hecho de que esta distribución no representa de forma fiel la evolución interna del virus en el cuerpo o el tiempo que tardan las personas en desarrollar inmunidad. Sin embargo, esta asunción es muy conveniente desde el punto de vista matemático, ya que la probabilidad de recuperarse en un intervalo temporal $d\tau$ no depende del instante temporal en el que el individuo contrajo la enfermedad, permitiendo que el tiempo no aparezca como variable explícita en las ecuaciones 1.19-1.21.

Crecimiento inicial y efectos estocásticos

Al inicio de la epidemia, asumiendo $s = 1$, hay una solución exponencial para la ecuación 1.20 de la forma $i(t) = i(0)e^{rt}$ donde $r = k\lambda - \mu$, es decir, tiene un valor explícito.

En los modelos compartimentales donde la probabilidad de que un agente se recupere en un tiempo T sigue una distribución exponencial, $Pr(t = T) = \mu e^{-\mu T}$, la probabilidad de infectar a i susceptibles debe tener en cuenta todos los posibles tiempos de recuperación. Generalizando la ecuación 1.8, que estimaba los casos secundarios generados durante un tiempo T , y utilizando la distribución exponencial queda:

$$m_i = \int_0^\infty Pr(t = T) m_i(T) dT = \int_0^\infty \mu e^{-\mu T} \frac{(k\lambda T)^i}{i!} e^{-k\lambda T} dT. \quad (1.23)$$

Intercambiando el orden de la integral y el sumatorio de la ecuación 1.7, se obtiene

la integral de una exponencial y, evaluándola, concluimos que

$$g(z) = \frac{\mu}{\mu - k\lambda(z - 1)} . \quad (1.24)$$

Esta expresión se puede derivar para obtener

$$g'(z) = \frac{\mu k\lambda}{(\mu - k\lambda(z - 1))^2} , \quad (1.25)$$

que se puede resolver debido a que $g'(1) = R_0$, obteniendo

$$R_0 = \frac{k\lambda}{\mu} . \quad (1.26)$$

Finalmente se calculan las raíces de $z = g(z)$, que son $z = 1$ y $z = \mu/k\lambda = 1/R_0$. En otras palabras, a pesar de que el número reproductivo básico sea mayor que uno, hay una probabilidad considerable de que una enfermedad se extinga si partimos de un solo individuo infeccioso, por ejemplo, un cincuenta por ciento para $R_0 = 2$. Esto es muy importante a la hora de contrastar los resultados de las simulaciones de agentes, en las que se tienen en cuenta los efectos estocásticos, con las ecuaciones deterministas en las que el tamaño final de la epidemia estimado es siempre el mismo y no tiene en cuenta la aleatoriedad. Por ello, la media en modelos compartimentales SIS de muchas simulaciones de agentes con semilla de i infectados será $1 - (1/R_0)^i$ veces la obtenida en las ecuaciones (en modelos SIR dependerá de los casos generados antes de extinguirse), y es conveniente empezar con una semilla lo suficientemente alta para poder ignorar este efecto.

Parámetros de orden e impacto

Una vez estudiada la primera fase de la epidemia, sigue analizar cómo va a ser su evolución y, en especial, su tamaño final. El impacto epidémico en el modelo SIR está definido como la fracción de la población que ha pasado por la enfermedad y se ha recuperado, $r(\infty)$. Partiendo de las ecuaciones diferenciales, es posible calcular cómo cambian los susceptibles en función de los recuperados como

$$\frac{ds}{dr} = -\frac{k\lambda}{\mu}s = -R_0s , \quad (1.27)$$

la cual se puede integrar para dr y ds en ambos lados

$$\int_{r(0)}^{r(\infty)} dr = -\frac{1}{R_0} \int_{s(0)}^{s(\infty)} \frac{ds}{s}, \quad (1.28)$$

lo que deja

$$s(\infty) = s(0)e^{-r(\infty)R_0} \quad (1.29)$$

si el sistema empieza con $r(0) = 0$.

De la expresión (1.29) es inmediato que, tras aproximar $s(0) \approx 1$ al empezar con una semilla muy pequeña y sabiendo que no quedan infectados en $t \rightarrow \infty$, se cumple que $r(\infty) = 1 - s(\infty)$ y, por lo tanto, queda la expresión

$$r(\infty)R_0 = -\ln(1 - r(\infty)). \quad (1.30)$$

Esta solución no tiene una forma explícita pero se puede resolver numéricamente para diferentes valores de R_0 , tal y como se muestra en el diagrama de fases de la Figura 1.3. En ella se pueden distinguir la fase libre de epidemia para $R_0 \leq 1$ y la epidémica para $R_0 > 1$. Estos diagramas son ampliamente utilizados para evaluar el impacto que tiene una epidemia en función de ciertos parámetros como la infectividad λ , y encontrar el umbral epidémico para el cual se extinguen los infectados antes de un brote mayor. Es frecuente el uso de diagramas de fases con modelos de tipo SIS, debido a que al llegar a un estado estacionario con $i(\infty) > 0$ es más sencillo obtener resultados con menor varianza sin ejecutar un gran número de simulaciones. En las ecuaciones diferenciales del SIS, basta con tomar el estado estacionario de $\dot{i} = 0 = k\lambda(1 - i)i - \mu i$, con el que se extrae inmediatamente $i = 1 - 1/R_0$.

Otro parámetro característico de las epidemias es la cantidad máxima de infectados concurrentes, a lo que nos referimos como el pico de la curva epidémica, porque su valor está asociado a la carga máxima que soportará el sistema sanitario de la región donde ocurre el brote. Conocer la expresión matemática de este indicador permite elaborar políticas de control centradas en evitar la saturación del sistema que pueden complementar aquellas diseñadas para reducir el impacto total del brote epidémico. Para obtener el tamaño del pico epidémico se puede reescribir la ecuación 1.29 como

$$i(t) = 1 - s(t) + \frac{1}{R_0} \ln(s(t)), \quad (1.31)$$

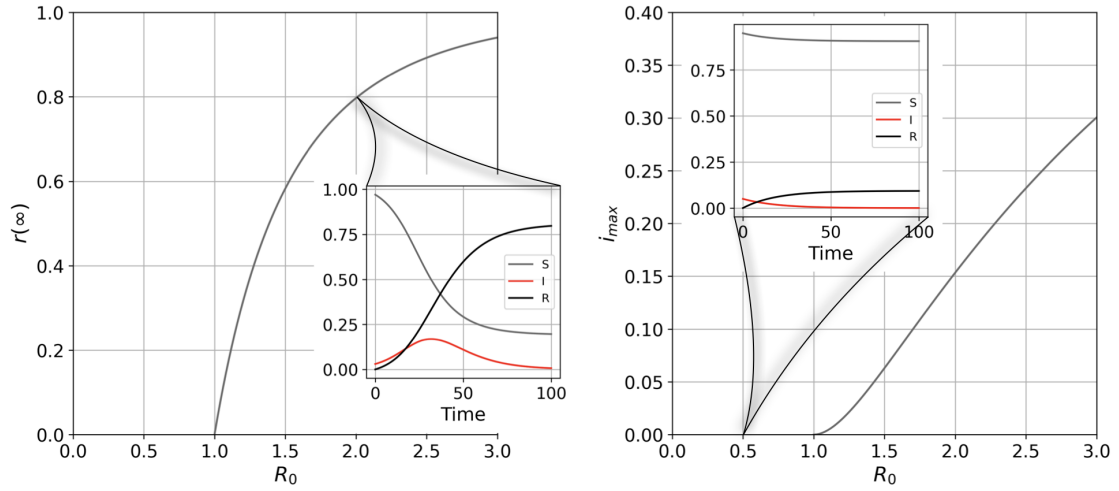


Figura 1.3: Tamaño final de la epidemia (izquierda) y número máximo de infecciosos (derecha) obtenidas en el modelo compartimental SIR, para diferentes valores de R_0 . La dinámica temporal de las proporciones de susceptibles, infecciosos y recuperados está representada en los plots interiores para $R_0 = 2$ (izquierda) y $R_0 = 0.5$ (derecha).

donde el máximo es calculado tras resolver $di/ds = 0$, obteniendo

$$i_{max} = 1 - \frac{1}{R_0} (1 + \ln(R_0)) . \quad (1.32)$$

R_0 y la matriz *next-generation*

En este apartado, se retoman los operadores *next-generation* para describir el crecimiento inicial de epidemias generadas por las ecuaciones diferenciales que gobiernan la evolución temporal de los modelos compartimentales anteriormente descritos. En ellos es común contar con varios estados infecciosos (expuestos, asintomáticos, sintomáticos, en cuarentena, etcétera) además de los característicos de toda la población que se han visto anteriormente (edad, sexo).

Debido a la naturaleza markoviana de las ecuaciones de las que partimos ahora, no será necesario tratar con las integrales y la forma explícita de la probabilidad $P(\tau = T) = \mu e^{-\mu T}$. En su lugar, es posible trabajar con matrices construidas a partir de las ecuaciones diferenciales del modelo con el que estemos trabajando, tras haberlas linealizado (asumiendo una cantidad despreciable de infectados) [105]. En concreto, se parametrizan exclusivamente los estados de infección transitivos, es decir, se pueden excluir los compartimentos de susceptibles y de recuperados porque en él solo se acumularían agentes que no contribuirán a nuevos casos. Las

ecuaciones tales como las del modelo SIR (1.19-1.21) se representan de la forma:

$$\dot{\mathbf{x}} = (\mathbf{T} + \mathbf{\Sigma})\mathbf{x} \quad (1.33)$$

Estos vectores pueden incluir estados de infección o de heterogeneidad, pero en cualquier caso se debe cumplir que la matriz $\mathbf{T}$ incluye los nuevos casos creados por los individuos infectados y la matriz $\mathbf{\Sigma}$ rige las transiciones entre estados infectados, incluida la recuperación. Se argumenta que la matriz *next-generation* viene dada por

$$\mathbf{K} = \mathbf{T} \mathbf{\Sigma}^{-1}, \quad (1.34)$$

lo cual es razonable al interpretar $\mathbf{T}$ como la tasa de creación de infectados y el inverso de $\mathbf{\Sigma}$ como el tiempo promedio que se tarda en dejar los compartimentos contagiosos. Al igual que antes, R_0 va a ser el máximo autovalor de la matriz $\mathbf{K}$ mientras sea irreducible y no negativa.

Para visualizar concretamente el significado de las definiciones anteriores, a continuación se pone como ejemplo el modelo SEIR. Las ecuaciones que gobiernan los dos estados infectados (expuesto e infeccioso, que se agrupan en el vector $\mathbf{x} = (e, i)$) tras haberlas linealizado asumiendo $s \cdot i \simeq i$, son:

$$\dot{e} = -\eta e + k\lambda i, \quad (1.35)$$

$$\dot{i} = \eta e - \mu i \quad (1.36)$$

Separando las componentes de la matriz en función de si generan nuevos casos o redistribuyen los compartimentos existentes, queda

$$\mathbf{T} = \begin{pmatrix} 0 & k\lambda \\ 0 & 0 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} -\eta & 0 \\ \eta & -\mu \end{pmatrix}. \quad (1.37)$$

Se dice que el compartimento de expuestos (E) es el estado durante infección, es decir, el estado a través del cual se entra en el conjunto de compartimentos infectados. Aunque en este ejemplo el expuesto es el único estado durante infección, puede haber más dependiendo del modelo. A continuación se obtiene la expresión de la matriz *next-generation*

$$\mathbf{K} = \frac{k\lambda}{\eta\mu} \begin{pmatrix} \eta & -\mu \\ 0 & 0 \end{pmatrix}, \quad (1.38)$$

cuyo autovalor máximo es $k\lambda/\mu$, lo que implica que R_0 no es diferente al añadir el compartimento intermedio de expuestos, sin importar cuánto tiempo los agentes inviertan en él. Este proceso, por supuesto, se extrapola a casos más complejos como modelos compartimentales con más estados, o aquellos que incluyen heterogeneidades como el de la pirámide poblacional, donde cada uno de estos compartimentos está dividido en todos los posibles estados de los individuos, y cada elemento de la matriz debe ser sustituido por un bloque en sí mismo (como la matriz de contactos entre grupos de edades $C_{e,f}$). Con las herramientas que hemos visto y las que introduciremos a continuación, se dará paso a usar los formalismos de redes con el objetivo de intentar replicar la propagación de epidemias en redes del mundo real y teniendo en cuenta nuevos factores como la heterogeneidad o el desplazamiento de diversos grupos de personas.

1.2.4. Contagios en redes

Redes complejas

Grafos y su representación

Una red es, en términos generales, cualquier sistema que admite una representación matemática abstracta como un *grafo* cuyos nodos se identifican como los elementos del sistema y donde el conjunto de enlaces o conexiones representan la presencia de una relación o interacción entre ellos. Existen diferentes clases de grafos dependiendo de su estructura o lo que representan, en función de lo cual ciertas operaciones pueden ser simplificadas. Un grafo *no dirigido* consta de dos conjuntos $G = (\nu, \varepsilon)$, donde ν es un conjunto contable no nulo de elementos, que se corresponden con los nodos, y ε es un conjunto de pares *no ordenados* de distintos nodos, formando los enlaces. El enlace (i, j) une los nodos i y j , que se dice están conectados. El tamaño de un grafo viene dado por el número de nodos N que lo componen, o en su lugar del número de enlaces L . Para un grafo no dirigido de N nodos, la cantidad máxima de enlaces posible es de $L = \binom{N}{2}$. Por otro lado, un grafo *dirigido* consta de un conjunto de pares *ordenados*. En los grafos no dirigidos, una línea uniendo dos nodos los conecta en ambos sentidos, mientras que en los dirigidos puede conectarlos en cualquier sentido, o en los dos.

Desde un punto de vista matemático conviene definir un grafo a través de la

matriz de adyacencia, que es una matriz de dimensión $N \times N$ definida tal que:

$$A_{ij} = \begin{cases} 1 & \text{si } (i, j) \text{ conectados} \\ 0 & \text{si } (i, j) \text{ no conectados} \end{cases}$$

En la figura 1.4 se muestran ejemplos de grafos y sus matrices de adyacencia asociadas. Notar que los grafos no dirigidos tienen matrices de adyacencia simétricas.

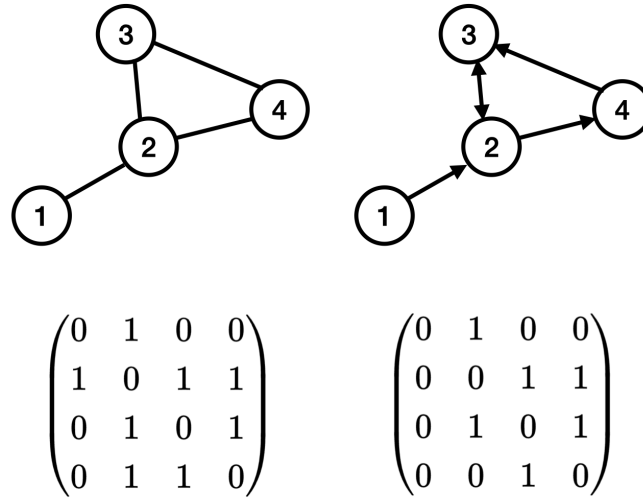


Figura 1.4: Red simétrica (izquierda) contra red asimétrica (derecha), junto con sus respectivas matrices de adyacencia.

Otra característica relevante de ciertas redes es que pueden ser pesadas. Esto significa que no todos los enlaces tienen la misma importancia. De esta forma, es posible representar muchas más clases de relaciones entre vértices. Un ejemplo puede ser la cantidad de personas que viajan de una ciudad a otra diariamente, lo cual está fuertemente conectado a la parte central de la tesis.

A la hora de trabajar con grafos es además conveniente tener en cuenta su dispersión (*sparseness*), o lo vacíos que se encuentran. Las redes pueden tener un amplio rango de enlaces por lo que a la hora de almacenar y trabajar con un grafo nos interesa saber cómo escala su cantidad de enlaces en función de sus nodos. Si $L \sim N^\alpha$, con $\alpha < 2$, decimos que es disperso. Es decir, conforme crece el valor de N , la cantidad L/N^2 va a tender a cero, y la matriz de adyacencia estará compuesta en casi su totalidad por elementos nulos. Por otro lado, si $\alpha = 2$, el grafo es denso. Cuando trabajamos con grafos dispersos, la matriz de adyacencia supone un gasto innecesario de memoria y computación, ya que casi todas las entradas son ceros, por lo que usamos otras formas de representar la red. Es conveniente representarlos

a partir de *listas de adyacencia*, dadas por $l(i, v \in \nu(i))$, donde $\nu(i)$ es el conjunto de vecinos del nodo i . Por ejemplo, la lista de adyacencia correspondiente al grafo no dirigido de la figura 1.4 es:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 2 & & & \\ 1 & 3 & 4 & \\ 2 & 4 & & \\ 2 & 3 & & \end{bmatrix}$$

En el caso de tratar con una red pesada, es suficiente con almacenar los pesos de cada enlace contiguamente. El ahorro de memoria no es la única ventaja de trabajar con listas de adyacencia. Muchas operaciones aplicadas sobre los enlaces a lo largo de la red serán además mucho más rápidas. Por ejemplo, si se desea sumar funciones dependientes de los vecinos del nodo i , tal que $f(M_{ij}) = 0$ si $M_{ij} = 0$, existen dos posibilidades:

$$\sum_{j=1}^N f(M_{ij}) \quad \sum_{j \in \nu(i)} f(M_{ij})$$

Si se aplicara esta operación para todos los nodos, llevaría un coste de N^2 pasos para la primera y L para la segunda, siendo esta última considerablemente menor en una red dispersa. Adicionalmente, existen gran cantidad de algoritmos (tales como la resolución de sistemas de ecuaciones lineales, multiplicación de matrices, u obtención de autovalores) que son significativamente más eficientes al tratar con redes dispersas.

Distribución de grado

El grado k de un vértice i se define como el número de enlaces del grafo que comparte con otros nodos. En el caso de redes dirigidas distinguimos los grados de entrada k_{in} y salida k_{out} , que vienen dados por:

$$k_{in,i} = \sum_j M_{ji} , \quad k_{out,i} = \sum_j M_{ij} .$$

Es habitual relacionar el grado de cada nodo con su relevancia, ya que en el estudio de la propagación de una epidemia los nodos de mayor grado tienden a tener un impacto mayor. En una red pesada es además posible definir la fuerza s de cada nodo de la misma manera que el grado, dado que la matriz de adyacencia incluye los pesos. Cuando los pesos son independientes de la topología, la fuerza de un nodo crecerá linealmente con su grado ($s \approx \langle w \rangle k$), donde $\langle w \rangle$ es el peso medio de las conexiones, y k es el grado del nodo.

Una de las formas más comunes de describir una red es a través de su **distribución de grado**, que se define como la probabilidad $P(k)$ de que un vértice elegido al azar tenga grado k . Esta distribución se puede construir a partir del histograma normalizado del grado de los nodos de la red y, a partir de su expresión, podemos conocer los momentos de la misma dados por:

$$\langle k^n \rangle = \sum_{k=1}^{\infty} k^n P(k).$$

El primer momento contiene la información del grado promedio a lo largo de la red:

$$\langle k \rangle = \sum_{k=1}^{\infty} k P(k) = \frac{1}{N} \sum_i k_i = \frac{2L}{N}.$$

En una red dispersa, $\langle k \rangle \ll N$. Por ejemplo, en el modelo de Ising en dos dimensiones, donde cada nodo de la red tiene un enlace con sus cuatro vecinos (exceptuando la frontera), el valor de $\langle k \rangle$ converge al aumentar la cantidad N indefinidamente.

Sistemas en el mundo real

Con los avances en tecnología y el acceso a una vasta cantidad de información, hemos logrado explorar y analizar redes a gran escala a través de datos sobre el comportamiento de sistemas de diferente naturaleza (como biológicos, sociales y tecnológicos). Estas redes presentan características únicas y complejas que ofrecen nuevas oportunidades para su estudio y aplicación. En concreto, casi todas estas redes muestran distribuciones de grado con colas pesadas o *libres de escala* [106]. Esta característica indica la presencia de una gran heterogeneidad entre los nodos de la red. Unos pocos, llamados *Hubs*, acumulan una gran cantidad de enlaces, mientras que el resto (la mayoría) tendrán muy pocas conexiones en comparación a la media.

Un tipo de distribuciones de grado que muestra estas características y que, comúnmente (aunque con importantes excepciones [107–109]), se encuentra al evaluar redes reales son las leyes de potencias. La estructura de redes libres de escala es, en cualquier caso, trascendental cuando se trata del estudio de sistemas dinámicos. Como se ha mencionado antes, el grado de los nodos de una red va a determinar el resultado de muchos procesos físicos pero la media del grado, que es uno de los indicadores más utilizados, pierde su valor descriptivo en las redes libres de escala. Es posible mostrar cómo para redes cuya distribución de grado siga una ley de potencias dada por $P(k) \propto k^{-\gamma}$, tal que $2 < \gamma \leq 3$, la media tiene un límite finito. Tratando, por simplicidad, la distribución $P(k)$ como una función

definida en el continuo podemos escribir:

$$\int_1^{k_c} kP(k)dk \propto \int_1^{k_c} k^{1-\gamma}dk = \frac{1}{2-\gamma} (k_c^{2-\gamma} - 1) . \quad (1.39)$$

En esta expresión la media converge para $\gamma > 2$ aunque el tamaño de la red crezca indefinidamente (y $k_c \rightarrow \infty$).

Sin embargo, es necesario estimar el error cometido al asumir que el grado de un vértice es $\langle k \rangle$, el cual se asocia al segundo momento, que es $\frac{1}{3-\gamma} (k_c^{3-\gamma} - 1)$, y diverge con el tamaño de la red. Por lo tanto, la media del grado de los nodos de una red no proporciona una información fidedigna sobre el grado de cada nodo y, como se verá más adelante, sobre la forma en la que procesos como el contagio epidémico se desarrolla en sistemas con esta característica. Vista esta propiedad y dada su presencia en las redes del mundo real, es importante conocer formas de crear redes sintéticas con mayor o menor heterogeneidad.

Redes Erdős-Renyi

El modelo de redes aleatorias de Erdős-Renyi [110] permite crear redes homogéneas de forma sencilla. En este método se parte de un número de nodos N y, para cada pareja de nodos, se puede aleatoriamente crear un enlace entre ellos con probabilidad p (alternativamente los nodos permanecen desconectados con probabilidad $(1-p)$). Con esta simple regla la red tendrá una distribución de grado $P(k)$ que cuando todas las parejas han sido sorteadas sigue una distribución binomial:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} .$$

Esta función $P(k)$ es una distribución de cola ligera, con una media $\langle k \rangle = p(N-1)$. Al aumentar el tamaño de la red, la distribución tenderá a una poissoniana, como se muestra en la figura 1.5, donde los nodos tienen grados similares. Este modelo, por tanto, es muy útil cuando queremos analizar redes homogéneas en las que los nodos tienen un grado igual o muy similar al grado medio.

Redes libres de escala

Para simular el comportamiento de redes más realistas (y por tanto más heterogéneas), A.L. Baràbasi y R. Albert introdujeron un algoritmo para generar redes libres de escala [43]. En su modelo, en vez de trabajar con un número fijo de nodos N , se van introduciendo vértices uno por uno y la red crece de forma dinámica. Este proceso de crecimiento se inicia con un pequeño grafo (semilla) de m_0 nodos y

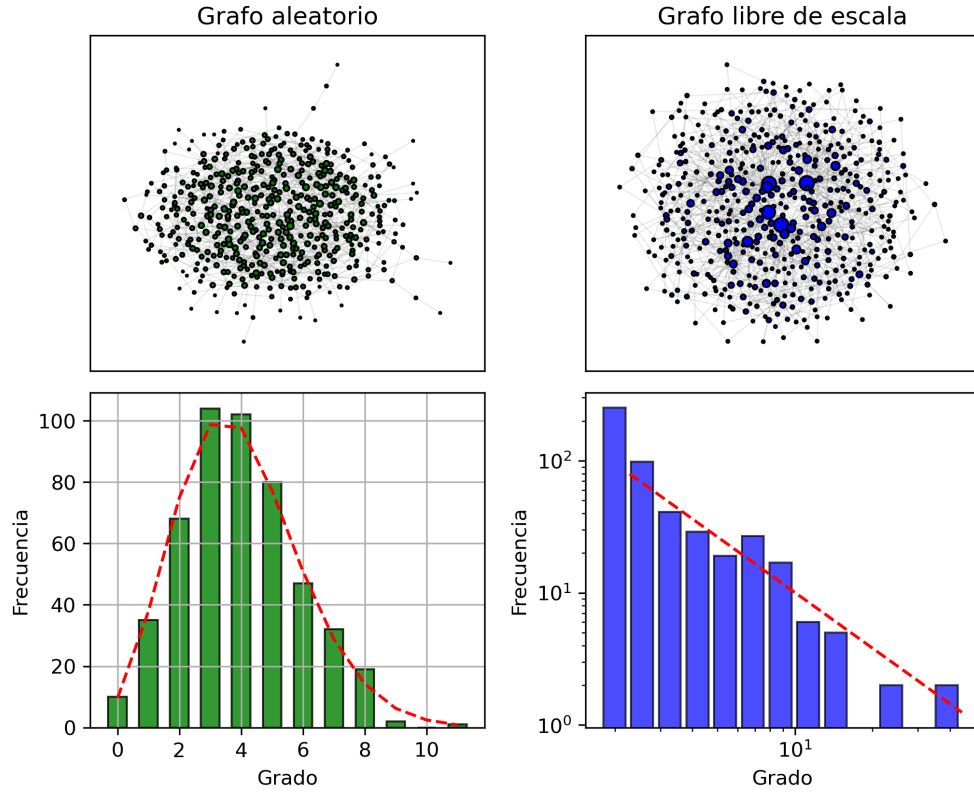


Figura 1.5: Representación de un grafo aleatorio (izquierda) frente a un grafo libre de escala (derecha), ambos con 500 nodos y conectividad media $\langle k \rangle = 4$. El histograma del grafo aleatorio se ajusta a una distribución de Poisson de media $\langle k \rangle = 4.05$, y el del grafo libre de escala se ajusta a una ley de potencias con exponente $\gamma = 2.80$.

compacto (todos los nodos unidos al resto). En cada paso de crecimiento, la introducción de un nuevo vértice conlleva la creación de m ($m < m_0$) nuevos links con los nodos ya presentes en la red. Además, la probabilidad de que el nodo entrante establezca un enlace con un nodo dado es proporcional al grado de dicho nodo. De esta forma, va a haber una preferencia por los nodos de mayor grado. Este tipo de enlazamiento preferencial resulta en la creación de hubs conforme la red crece los cuales, a su vez, aumentan su probabilidad de captar nuevos enlaces de nodos futuros, amplificando la desigualdad en el seno de la red. Cuando $N \rightarrow \infty$ se puede comprobar que el grado medio es $\langle k \rangle = 2m$, mientras que la distribución de grado se aproxima a una ley de potencias $P(k) \propto k^{-3}$, tal y como se muestra en la figura 1.5.

Modelos compartimentales en redes complejas

Las ecuaciones diferenciales descritas en los modelos compartimentales que se han estudiado en la sección anterior seguían un acercamiento de campo medio, es decir, asumían que cada persona interactúa con variables promediadas a lo largo de la población. Por ejemplo, si hay una fracción de infectados i , se considera dicha fracción como la probabilidad de que cualquier persona con la que se tiene un contacto esté infectada. Esto no es cierto en el mundo real, donde las interacciones entre personas tienen fuertes correlaciones, heterogeneidades y recurrencias. En esta sección se revisarán varios modelos que, de forma incremental, van incorporando esta complejidad en los sistemas que tratan de replicar.

Campo medio heterogéneo

En la aproximación de campo medio de los modelos compartimentales se toma el valor promedio k como la tasa de contactos que tiene cualquier persona a lo largo del tiempo, lo cual se aleja mucho de la realidad. Como se ha probado al construir las redes aleatorias y libres de escala, incluso en las redes homogéneas existe una desviación de la media para las conexiones de los individuos. Como se mostrará a continuación, la media $\langle k \rangle$ apenas ofrece información relevante a la hora de estimar si la propagación de un patógeno contagioso dará lugar a una epidemia. El modelo de campo medio heterogéneo (HMF) [45] agrupa los agentes suponiendo que, dependiendo del grado (su cantidad de contactos) tendrán una probabilidad mayor o menor de contagiarse. Aunque sigue siendo un modelo de campo medio ya que cada persona interactúa con el promedio de todos los agentes del mismo grado, añade un nivel de complejidad al sistema. Consideremos una distribución de probabilidad $P(k)$ de los agentes de una red de contactos, donde i_k representa la proporción de agentes con grado k que son infecciosos. Entonces, su evolución viene dada por:

$$\dot{i}_k = -\mu i_k + s_k k \lambda \sum_{k'} P(k|k') i_{k'} , \quad (1.40)$$

donde $P(k|k')$ es la probabilidad de que un individuo de grado k esté conectado a uno de grado k' . Si no hay correlaciones explícitas entre grados, cabe asumir que esa probabilidad solo depende del segundo individuo (su grado k' y la cantidad de ellos que hay, es decir, $P(k')$), por lo que $P(k|k') = k' P(k') / \langle k \rangle$. Esto simplifica mucho los resultados, ya que ahora todos los individuos *perciben* la misma proporción de infectados, que es

$$\Phi = \sum_{k'} \frac{k' P(k')}{\langle k \rangle} i_{k'} , \quad (1.41)$$

y la única diferencia será la cantidad de veces que cada uno interactúa con este individuo promedio.

A continuación queda calcular el umbral epidémico y analizar su dependencia en la distribución de grado. Las asunciones de aquí en adelante servirán como precedente a la hora de calcular la infectividad crítica en numerosos casos en los que el cálculo de R_0 resulta ser más complejo. Las dos condiciones principales son: el sistema se encuentra en el estado estacionario, donde la cantidad de infectados es constante y por tanto cada infectado genera de media un caso, y la proporción de infectados es mucho menor que uno. Estableciendo que $\dot{i}_k = 0$ para cualquier k , y que $s_k = 1$, resulta que para $\lambda = \lambda_c$:

$$i_k = \frac{k\lambda_c}{\mu} \Phi. \quad (1.42)$$

Recordando la definición de Φ en la ecuación 1.41, se aplica ese sumatorio en ambos lados de la ecuación anterior (para cada valor de k), llegando a

$$\Phi = \sum_k \frac{kP(k)}{\langle k \rangle} \frac{k\lambda_c}{\mu} \Phi, \quad (1.43)$$

que se puede simplificar, y tomando $\sum_k k^2 P(k) = \langle k^2 \rangle$, queda

$$\lambda_c = \mu \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (1.44)$$

Las consecuencias de este resultado son evidentes: para un número de nodos suficientemente grande, el umbral epidémico de las redes libres de escala tiende a cero (debido a que la varianza de las distribuciones que escalan con k^a , siendo $2 \leq a < 3$, diverge). Esto significa que no importa lo contagioso que sea un virus; si una red libre de escala es lo suficientemente grande, se propagará de forma exponencial al menos durante un corto tiempo. Se explica de esta forma la razón por la cual muchos virus informáticos conseguían propagarse con tal velocidad cuando gran parte de la población empezó a usar internet, siendo este uno de los mayores ejemplos de redes extremadamente heterogéneas.

Campo medio en redes estáticas

Las ecuaciones de HMF asumen implícitamente la existencia de una estructura de contactos dinámica en la que el número de interacciones de cada individuo se mantiene constante pero los individuos con los que contacta cambian con el tiempo. Sin embargo, en la mayor parte de las redes empíricas se observa que estos

contactos son recurrentes y cada persona tiene un grupo de vecinos que, aunque sufre variaciones con el tiempo, tiene cierta continuidad.

Para incorporar la recurrencia de los contactos, es necesario revisar la ley de acción de masas que se usa al calcular los nuevos casos, la cual decía que son proporcionales a $s \cdot i$. En su lugar, es más correcto determinar que los nuevos casos crecerán con la cantidad de parejas SI . Esto tiene en cuenta que si los infecciosos interactúan exclusivamente entre ellos, y no con los susceptibles, no habrá nuevos contagios. Se denomina ρ_i a la probabilidad de que el individuo i esté infectado, es decir, $\rho_i^I = Pr(I_i = 1) = \langle I_i \rangle$, siendo I_i uno cuando el individuo i está infectado y cero cuando no lo está. Para saber la tasa con la que el individuo i se puede infectar a partir de sus vecinos, hay tener en cuenta la probabilidad conjunta de que i sea susceptible y su contacto j sea infeccioso, es decir, $P(S_i = 1, I_j = 1)$ y que se escribe como $\langle S_i I_j \rangle$. Las ecuaciones que describen esta dinámica son:

$$\langle \dot{S}_i \rangle = -k\lambda \sum_j A_{ij} \langle S_i I_j \rangle, \quad (1.45)$$

$$\langle \dot{I}_i \rangle = k\lambda \sum_j A_{ij} \langle S_i I_j \rangle - \mu \langle I_i \rangle, \quad (1.46)$$

donde la matriz de adyacencia A_{ij} permite que haya contactos entre i y j (puede ser dirigida). La forma más sencilla de cerrar estas ecuaciones es considerar que la probabilidad de que cualquier miembro de una pareja esté infectado es independiente del otro, es decir, $\langle S_i I_j \rangle = \langle S_i \rangle \langle I_j \rangle$, con lo que volvemos de cierta forma a la ley de acción de masas. A este modelo se le denomina campo medio basado en individuos y el umbral epidémico se puede obtener de nuevo asumiendo que las derivadas con el tiempo son nulas y la proporción de susceptibles tiende a uno. El resultado está en el autovalor máximo de la matriz de adyacencia:

$$\lambda_c = \frac{\mu}{\Lambda_{max}(\mathbf{A})}. \quad (1.47)$$

Esta aproximación (suponer que no hay correlación entre el estado de dos vecinos) tiene sus problemas [111, 112], ya que requiere trabajar con términos que no tienen sentido físico o biológico: al eliminar dicha correlación el modelo está incluyendo el efecto de que un individuo sea susceptible e infeccioso al mismo tiempo, lo que permite que dos agentes retroalimenten sus probabilidades de infección de forma circular. La consecuencia es que este umbral, que es equivalente para los modelos SIS y SIR, se desvanece en redes con distribuciones de grados que siguen leyes de potencias con cualquier γ [51]. Esto se debe a que el autovalor máximo escala con el grado mayor de la red, de la forma $\sqrt{k_{max}}$ o $\langle k^2 \rangle / \langle k \rangle$ y tomando el más alto de

estos dos valores. Para distribuciones que siguen ley de potencias, queda

$$\lambda_c = \begin{cases} 1/\sqrt{k_{max}} & \gamma > 5/2 , \\ \langle k \rangle / \langle k^2 \rangle & 2 < \gamma < 5/2 . \end{cases} \quad (1.48)$$

Esto significa que, para redes estáticas, encontramos estados epidémicos con cualquier $\lambda > 0$ para prácticamente cualquier red aleatoria no uniforme, incluídas las Erdős-Renyi (aunque el autovalor máximo escala logarítmicamente, al contrario que en las de ley de potencias) [50, 113]. Contradiendo los resultados del HMF, la enfermedad no sobrevive por la heterogeneidad de grado sino exclusivamente gracias al hub, que actúa como un centro de reinfección, retroalimentándose de sus conexiones. Sin embargo, la aproximación de vecinos independientes empieza a fallar para el modelo SIR [50], donde la recuperación de un nodo cierra caminos de infección.

Esta discrepancia se puede solventar con modelos de campo medio basados en pares, en lugar de individuos. En lugar de cerrar las ecuaciones asumiendo independencia, se pueden escribir las ecuaciones que gobiernan la dinámica de las probabilidades conjuntas de los estados de parejas de individuos [52, 55, 114]. Sin embargo, esto conlleva incluir las probabilidades conjuntas de tríos de individuos. Por ejemplo, la probabilidad de que uno de ellos sea infeccioso y el otro susceptible evoluciona de la forma

$$\langle \dot{I_j S_i} \rangle = k\lambda \sum_{k \neq i} A_{kj} \langle I_k S_j S_i \rangle - k\lambda \sum_{k \neq j} A_{ki} \langle I_j S_i I_k \rangle - k\lambda A_{ji} \langle I_j S_i \rangle - \mu \langle I_j S_i \rangle , \quad (1.49)$$

donde el primer término representa la contribución positiva de un tercer vecino contribuyendo a que se infecte j , pero no i . El segundo término, por el contrario, representa a dicho vecino infectando a i , lo cual reduce la probabilidad conjunta de ese estado de la pareja. De la misma forma que en el campo medio basado en individuos, no es posible cerrar las ecuaciones a no ser que asumamos cierto nivel de independencia, por ejemplo [55] suponiendo

$$\langle A_i, B_j, C_k \rangle \approx \frac{\langle A_i B_j \rangle \langle B_j C_k \rangle}{\langle B_j \rangle} . \quad (1.50)$$

De esta forma, es posible también obtener la expresión del umbral epidémico a través del mayor autovalor de una matriz Jacobiana que se construye a partir de la matriz de adyacencia, pero es más complejo de resolver. Sin embargo, para el caso de matrices homogéneas donde todos los nodos tienen conectividad k , se puede

reducir a

$$\lambda_c = \frac{\mu}{k-1} \quad (1.51)$$

si la matriz de adyacencia es no pesada. Esto significa que el modelo tiene en cuenta que cuando un individuo es infectado por un vecino, este no podrá ser infectado de vuelta, por lo que la cantidad posible de nodos a los que podrá contagiar será como máximo $k - 1$.

Los resultados en el campo medio basado en pares están en mejor acuerdo con las simulaciones de agentes, y siempre se puede implementar orden mayor en las tuplas de vecinos. Un problema recurrente en cualquier acercamiento de orden q es que siempre aparecen términos de orden $q + 1$, por lo que es necesario romper alguna dependencia. Aún así, cuanto mayor es el orden q de las ecuaciones, mayores son las correlaciones que se tienen en cuenta y, por lo tanto, más precisa será la aproximación. El coste, sin embargo, es alto, ya que para una red de N nodos el número de ecuaciones escala con $\binom{N}{q}$.

El acercamiento microscópico de cadena de Markov (MMCA)

Hasta ahora se han descrito modelos markovianos de tiempo continuo, en los que las transiciones entre los diferentes compartimentos están gobernadas por tasas constantes. Los modelos que se utilizarán en la tesis, sin embargo, están basados en tiempo discreto. La extrapolación de los resultados obtenidos anteriormente se puede realizar de forma simple si se consideran ciertas implicaciones que conlleva la discretización. Esto permitirá el uso de escalas temporales mayores con menor coste computacional, pero conservando la estabilidad dinámica incluso para intervalos grandes.

En el acercamiento microscópico de cadena de Markov [54], en lugar de producirse contactos de forma continua, se asume que todos los contactos de un individuo ocurren de forma simultánea en cada instante temporal. Para conocer la probabilidad de infectarse, se calcula primero la de que un agente no se infecte en ninguno de los contactos como

$$q_i(t) = \prod_j (1 - \lambda A_{ij} I_j(t)) , \quad (1.52)$$

y este término se puede incorporar en la dinámica de forma similar al campo medio

basado en individuos:

$$I_i(t+1) = (1 - q_i(t))(1 - I_i(t)) - \mu I_i(t) . \quad (1.53)$$

La linealización de este formalismo lleva de nuevo a un umbral epidémico $\lambda_c = \mu/\Lambda_{max}(\mathbf{A})$, que es consistente con los anteriores. Con una complejidad lineal con el número de enlaces en la red y N ecuaciones, este modelo replica de forma precisa la dinámica de contagio y los parámetros de orden.

Un modelo de propagación de epidemias debe describir la forma en la que los individuos se relacionan y tienen contactos entre ellos. Una primera forma de simular estos sistemas ha sido utilizando las redes de contactos entre individuos. Sin embargo, esto no tiene en cuenta la difusión que se observa en la vida real: las personas no se relacionan siempre con los mismos grupos de vecinos. Aquí es donde entran en juego las metapoblaciones.

Movilidad y epidemias

Modelos metapoblacionales

Los modelos de metapoblaciones permiten capturar cómo los movimientos de los agentes entre diferentes localizaciones determinan sus interacciones. Una metapoblación es una red compleja compuesta de subpoblaciones, que son conjuntos de personas reunidas en un lugar y que interactúan entre ellas. Estas subpoblaciones constituyen los nodos o vértices de una red, y los enlaces representan la movilidad entre las mismas. Por ejemplo, dentro de una ciudad, nuestra red puede describir la cantidad de personas que viajan del barrio en el que residen a otro donde esté localizado su lugar de trabajo.

Las metapoblaciones han sido utilizadas extensivamente en el estudio de epidemias, ya que nos permiten analizar la evolución de epidemias considerando diferentes escalas espaciales y tener en cuenta los efectos de los movimientos de las personas. A pesar del gran avance en el uso de metapoblaciones, todavía no hay un consenso a la hora de elegir la naturaleza de los movimientos incluidos en los enlaces de la red para estudiar un escenario epidémico. Asimismo, teóricamente, no se ha explorado exhaustivamente cómo dicha naturaleza afecta a las trayectorias epidémicas generadas. Esta elección suele depender de la clase de datos con la que se esté trabajando, que pueden proceder de fuentes como censos, encuestas, seguimiento GPS, o transporte público. Al reducirse generalmente estos datos a cantidades de personas que viajan de un parche a otro durante un intervalo tempo-

ral, es de especial importancia conocer la naturaleza de estos viajes y comprobar que se adhieren al contexto que queremos modelizar. A grandes escalas espacio-temporales, es común tener en cuenta la migración o las nuevas generaciones de susceptibles, mientras que en escalas más granulares los viajes al trabajo, escuela u otros destinos frecuentes ganan importancia.

A la hora de elegir el formalismo con el que se van a tratar los movimientos se tienen en cuenta dos principales descripciones: la **euleriana** y la **lagrangiana** [115, 116]. La decisión entre ambas puede afectar considerablemente a los resultados [117]. Los modelos de movimiento euleriano no siguen el comportamiento de agentes individuales, sino que únicamente describen la tasa con la que personas de una subpoblación se desplazan a otra, sin que un agente en concreto recuerde su posición anterior a haberse movido. Por otro lado, un modelo de movimiento lagrangiano incluye a cualquiera que describa la cantidad de tiempo que los agentes pasan en otros parches antes de volver a su lugar de origen, así que poseen memoria temporal y permiten introducir movimientos recurrentes.

Un ejemplo de movilidad euleriana es el modelo de flujo, que describe los viajes como una difusión de la población de un lugar a otro. Si la cantidad de personas que se encuentran en un nodo i es N_i , se puede describir el cambio de las poblaciones en tiempo continuo como

$$\frac{dN_i}{dt} = - \sum_{j=1}^K f_{ij} N_i + \sum_{j=1}^K f_{ji} N_j , \quad (1.54)$$

donde la población total de la metapoblación se mantiene constante con el tiempo, pero la de cada parche puede variar hasta que no se alcance un estado estacionario. En un modelo de portadores de una enfermedad, la difusión de los infectados se describirá también con esta ecuación.

Por otro lado, el modelo del viaje recurrente es un ejemplo de movilidad lagrangiana. Con el objetivo de representar los viajes diarios que realizan los habitantes de una población a destinos como la escuela o el trabajo, cada viajante mantendrá la información de su residencia original, a la cual siempre volverá antes de moverse de nuevo. Se puede asociar a los agentes a un lugar de residencia de forma

permanente, y sus ecuaciones de la dinámica en tiempo continuo son

$$\frac{dN_{ii}}{dt} = - \sum_{j=1}^K \phi_{ij} N_{ii} + \sum_{j=1}^K \tau_{ij} N_{ij}, \quad (1.55)$$

$$\frac{dN_{ij}}{dt} = -\tau_{ij} N_{ij} + \phi_{ij} N_{ii}. \quad (1.56)$$

Este modelo es entonces ideal para modelizar el contagio de epidemias en entornos intraurbanos, donde la mayor parte de los viajes son cortos y repetitivos.

Matriz Origen-Destino

Se denomina matriz origen-destino (OD) a la matriz de adyacencia cuyos pesos representan los viajes entre subpoblaciones dentro de una metapoblación. Existen diferentes maneras de almacenar esta información, dependiendo de cómo se va a aplicar en cada modelo. Una forma consiste en almacenar en un vector la cantidad de personas n_i que reside en cada población, y la matriz OD representa el porcentaje de gente que se desplaza a otros nodos, dado como R_{ij} . Obviamente debe cumplir

$$\sum_j R_{ij} = 1. \quad (1.57)$$

Una segunda forma se reduce a simplemente representar en la matriz OD la cantidad total de personas que se mueven entre nodos n_{ij} . Se relaciona con el formalismo anterior con las igualdades

$$n_i \cdot R_{ij} = n_{ij}, \quad (1.58)$$

$$\sum_j n_{ij} = n_i. \quad (1.59)$$

En los modelos que aparecen a continuación, se trabajará con procesos de reacción-difusión similares a los estudiados en otras áreas como la química, donde ciertas reacciones entre elementos pueden estar favorecidas (o limitadas) por la capacidad de transporte del medio que carga con ellas.

Epidemias en metapoblaciones: el formalismo MIR

El formalismo Movimiento-Interacción-Retorno, o MIR [69], es fundamental en el desarrollo de esta tesis, extrapolando el acercamiento microscópico de cadena de Markov (MMCA) a redes de movilidad que describen los viajes recurrentes en paisajes urbanos a tiempo discreto. En este modelo se define una probabilidad de desplazamiento, p , con la que cada agente decide si viajar a un parche de destino al principio de cada paso temporal o no. En caso de viajar, escogerá su destino

entre los posibles vecinos de su residencia, cada uno con probabilidad dada por la matriz origen-destino R_{ij} . Tras el viaje, ocurre la etapa de interacción, en la que se asume que cada subpoblación es homogénea y se simula la dinámica de contagio a través de modelos compartimentales asumiendo que cada individuo interactúa con todos los individuos que se encuentran en su parche. Como consecuencia, cada individuo tiene una probabilidad de contagiarse dependiendo de dónde se encuentre durante la fase de interacción. A diferencia de otros modelos que incorporan una descripción lagrangiana, este modelo fuerza a todos los agentes a retornar a su parche original tras el periodo de interacción.

En el caso en que el modelo compartimental subyacente sea un modelo SIS, se puede describir el estado del sistema a partir de la proporción de infectados en cada parche i a tiempo t , $\rho_i(t)$. Las ecuaciones que gobiernan la dinámica de estas variables son

$$\rho_i(t+1) = (1-\mu)\rho_i(t) + (1-\rho_i(t))\Pi_i(t) , \quad (1.60)$$

donde $\Pi_i(t)$ es la probabilidad de que un agente susceptible residente en i se infecte durante el paso temporal t . Esta probabilidad se calcula teniendo en cuenta que puede quedarse en su subpoblación o viajar a otra. Si la probabilidad de infectarse estando en el parche i durante la interacción es $P_i(t)$, entonces:

$$\Pi_i(t) = (1-p)P_i(t) + p \sum_j R_{ij}P_j(t) . \quad (1.61)$$

Teniendo en cuenta que cada agente establece tantos contactos como individuos se encuentren en ese momento en el nodo, se obtiene :

$$P_i(t) = 1 - \prod_j (1 - \lambda \rho_j(t))^{n_{j \rightarrow i}} , \quad (1.62)$$

donde $n_{i \rightarrow j}$ es la cantidad de población que se desplaza de j a i , dada por

$$n_{j \rightarrow i} = \delta_{ij}(1-p)n_i + pR_{ji}n_j . \quad (1.63)$$

Aunque esta cantidad de contactos pueda ser poco realista (especialmente en sistemas con grandes números de individuos), en su aplicación se compensa con la probabilidad λ de infectarse en cada contacto, que se escogerá o estimará dependiendo de cada caso particular. Tal y como muestran las ecuaciones, en este modelo la cantidad de contactos por persona depende del parche en el que se encuentran, y es proporcional a su población (o densidad si así se especificara). Esto permite estudiar el efecto de movilidad en situaciones en las que el contagio de una enfermedad es mayor en ciertos lugares, lo que da lugar a una representación

más cercana a la realidad e introduce fenómenos nuevos en los modelos.

Adicionalmente, este modelo nos permite calcular una expresión analítica del umbral epidémico λ_c a partir de lo que se denomina la matriz de mezcla M , que se puede calcular linealizando la probabilidad de infectarse $\Pi_i(t)$. Asumiendo que la semilla de infectados es muy pequeña y constante en el tiempo, es decir, $\rho_i(t) = \varepsilon_i^* \ll 1$ para todos los parches i , se obtiene [69]

$$\Pi_i \approx \lambda \sum_{j=1}^N \underbrace{\left[(1-p_d)^2 \delta_{ij} n_j + p_d(1-p_d) n_j (\mathbf{R} + \mathbf{R}^\top)_{ij} + p_d^2 n_j (\mathbf{R} \cdot \mathbf{R}^\top)_{ij} \right]}_{M_{ij}} \varepsilon_j^* \quad (1.64)$$

$$= \lambda (\mathbf{M} \varepsilon^*)_i . \quad (1.65)$$

La matriz resultante es el equivalente a la matriz de adyacencia en estos modelos metapoblacionales, ya que representa la cantidad de contactos que tienen las personas residentes en el parche i con las de cualquier otro parche j , y su radio espectral permite calcular el umbral como

$$\lambda_c = \frac{\mu}{\Lambda_{max}(\mathbf{M})} . \quad (1.66)$$

Como se muestra más adelante en el cuerpo de la tesis, esta no es la única utilidad de la matriz de mezcla. Conforme los modelos se vuelven más complejos, la información aportada por su descomposición espectral va a ser más rica y puede ayudarnos a determinar cuáles son los individuos o patrones en la red de movilidad más críticos para la supervivencia y expansión de una epidemia.

Adicionalmente, este formalismo tiene la ventaja de caracterizar poblaciones donde, en ausencia de movilidad, pueden existir comunidades en régimen supercrítico al mismo tiempo que otras en el subcrítico; lo que permite que ocurran fenómenos como el *detrimento epidémico* al incorporar los viajes. Este detrimento se caracteriza por el aumento del umbral epidémico (o la disminución de R_0) con valores pequeños de p . El efecto, aunque poco intuitivo, puede observarse en cualquier red de movilidad multiescala suficientemente heterogénea con un rango de R_0 locales que se ve homogeneizado por la mezcla causada en los viajes. En concreto, la razón entre la escala temporal característica del estado infeccioso y la del movimiento espacial de la población es la principal métrica asociada al detrimento, ya que una mayor frecuencia de viajes impide que los infectados en el parche más denso pasen el tiempo suficiente en él para continuar contagiando a otros [118].

Capítulo 2

Determinantes socioeconómicos del confinamiento durante la primera ola de COVID-19



OPEN ACCESS

EDITED BY

Pierpaolo Ferrante,
National Institute for Insurance against
Accidents at Work (INAIL), Italy

REVIEWED BY

Pei Yuan,
York University, Canada
Syed Muzamil Basha,
REVA University, India

*CORRESPONDENCE

David Soriano-Paños
✉ sorianopanos@gmail.com
Jesús Gómez-Gardeñes
✉ gardenes@unizar.es

RECEIVED 24 March 2023

ACCEPTED 19 June 2023

PUBLISHED 05 July 2023

CITATION

Valgañón P, Lería U, Soriano-Paños D and
Gómez-Gardeñes J (2023) Socioeconomic
determinants of stay-at-home policies during
the first COVID-19 wave.
Front. Public Health 11:1193100.
doi: 10.3389/fpubh.2023.1193100

COPYRIGHT

© 2023 Valgañón, Lería, Soriano-Paños and
Gómez-Gardeñes. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Socioeconomic determinants of stay-at-home policies during the first COVID-19 wave

Pablo Valgañón^{1,2}, Unai Lería¹, David Soriano-Paños^{2,3*} and
Jesús Gómez-Gardeñes^{1,2*}

¹Department of Condensed Matter Physics, University of Zaragoza, Zaragoza, Spain, ²GOTHAM Lab -
Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza,
Spain, ³Institute Gulbenkian of Science (IGC), Oeiras, Portugal

Introduction: The COVID-19 pandemic has had a significant impact on public health and social systems worldwide. This study aims to evaluate the efficacy of various policies and restrictions implemented by different countries to control the spread of the virus.

Methods: To achieve this objective, a compartmental model is used to quantify the “social permeability” of a population, which reflects the inability of individuals to remain in confinement and continue social mixing allowing the spread of the virus. The model is calibrated to fit and recreate the dynamics of the epidemic spreading of 42 countries, mainly taking into account reported deaths and mobility across the populations.

Results: The results indicate that low-income countries have a harder time slowing the advance of the pandemic, even if the virus did not initially propagate as fast as in wealthier countries, showing the disparities between countries in their ability to mitigate the spread of the disease and its impact on vulnerable populations.

Discussion: This research contributes to a better understanding of the socioeconomic and environmental factors that affect the spread of the virus and the need for equitable policy measures to address the disparities in the global response to the pandemic.

KEYWORDS

COVID-19, epidemic modeling, Bayesian inference, compartmental models, non-pharmaceutical containment policies

1. Introduction

The COVID-19 disease, caused by the novel coronavirus SARS-CoV-2, is a highly contagious respiratory illness that was first reported in Wuhan, China in December 2019 (1). It was declared a pandemic by the World Health Organization on March 11th 2020 and by that time, the disease had spread globally, resulting in an international public health crisis that impacted all aspects of life for those affected. As it continued to spread, countries worldwide implemented a range of strict measures to contain the virus with varying degrees of success, including lockdowns, travel restrictions, and social distancing measures (2). Since the early stages of the pandemic, experts have studied the uneven spatial spread of the virus, which is associated with socioeconomic and environmental factors (3, 4). These studies revealed that minorities, low-income areas, and vulnerable populations (5, 6) have been disproportionately affected by the situation, exacerbating existing inequalities.

Numerous studies have examined the factors that contribute to the reproduction number or the speed at which a disease propagates in a population (7). Most of these studies have found positive correlations between the transmissibility of a pathogen and factors such as population density, income inequality, urban areas, and household size, among others (8–11). At the onset of the COVID-19 pandemic, an important metric that also

showed a positive correlation with the reproduction number of SARS-CoV-2 was the Gross Domestic Product (GDP) per capita (12–14). This suggests that the virus spread more rapidly in affluent countries due to a large percentage of the population living in densely populated cities and the significant inflow of air traffic facilitating the initial importation of a large number of cases. However, the data on this correlation was inconclusive in the subsequent phases of the COVID-19 pandemic.

Despite the implementation of measures to mitigate the spread of the virus, there are significant disparities between countries in their ability to contain the pandemic. While wealthier countries have been able to enforce stay-at-home policies by taking the appropriate measures and ensuring the safety of the most vulnerable populations, lower income countries are unable to prevent the loss of income and jobs, leading to food insecurity and reduced access to healthcare. As a result, even with mobility restrictions in place, people in these countries are still exposed to the risk of the pandemic due to the need to work and maintain their income.

The goal of this study is to highlight the connection between the different efficacy of lockdown policies observed across countries and their socioeconomic features. To this aim, we will attempt to measure the *social permeability* of these nations, which accounts for the inability of the population to remain in confinement and thus continuing social mixing that allows the disease to spread. The study of this variable allow us to distinguish unique scenarios that appear for each studied country and contribute to the already existing efforts (7) to show the varied relationships between the spread of epidemics and economic indicators.

2. Methods

2.1. Modeling

2.1.1. Discrete time compartmental model

The main core of this study is the development of a compartmental model to capture COVID-19 epidemic trajectories and how they were impacted by non-pharmaceutical interventions implemented in various countries. To this aim, the model should be sufficiently complex to provide an accurate representation of the epidemic process and the primary mechanisms behind virus transmission, yet simple and adaptable enough to be applied to different countries.

The proposed compartmental framework is an extended version of the *Susceptible-Exposed-Infected-Recovered* (SEIR) model (15), which allows for monitoring both the number of deaths over time for each country and the effects of non-pharmaceutical interventions. Our framework is a discrete-time model, being each time step a day. The model consists of six compartments: *Susceptible* (S), *Exposed* (E), *Infectious* (I), *Recovered* (R), *Pre-deceased* (P_d), and *Deceased* (D). The flows diagram connecting these compartments is represented in Figure 1. This diagram depicts the following sequence of events: In the absence of interventions, *Susceptible* individuals (S) have a likelihood of contracting the virus (β) for each contact with an infected person, leading them to move to the *Exposed* compartment (E), meaning that they are carriers of the virus but

not yet contagious. Once in compartment E , individuals can move to the *Infectious* compartment (I) with a probability η , where they can infect *Susceptible* individuals. *Infectious* individuals leave their compartment with probability μ , either recovering and entering the *Recovered* compartment (R) with a probability of $1 - \Upsilon$, or dying due to the disease with a probability of Υ . In the latter case, they enter in the *Pre-deceased* compartment (P_d) and, eventually, move to the *Deceased* compartment (D) with probability ξ .

Following the previous assumptions for the compartmental model, we can propose the dynamical equations driving the evolution of the individuals. In particular, the evolution from the *Infectious* stage can be straightforwardly derived, yielding:

$$I(t+1) = \eta E(t) + (1 - \mu)I(t), \quad (1)$$

$$R(t+1) = \mu(1 - \Upsilon)I(t) + R(t), \quad (2)$$

$$P_d(t+1) = \mu\Upsilon I(t) + (1 - \xi)P_d(t), \quad (3)$$

$$D(t+1) = \xi P_d(t) + D(t). \quad (4)$$

The remaining equations of the model account for the contagion of the *Susceptible* individuals and are shaped by the non-pharmaceutical interventions, which manifest in a reduction in mobility and the formation of *social bubbles* throughout the populations. The impact of non-pharmaceutical interventions will be encapsulated in a fraction of the *Susceptible* population gathering the number of individuals staying at their households, not being reachable by their infectious counterparts. To model the evolution of this confined population, we assume that mobility is governed by a time-dependent parameter $p_{act}(t)$. In particular, for each day, a fraction $p_{act}(t)$ of the *Susceptible* compartment remains active (S_{active}) while the rest, a fraction $(1 - p_{act}(t))$ of the pool of *Susceptibles*, reduce their mobility and social interactions. Of these individuals, a fraction $(1 - \phi)$ stays at home and forms a social bubble with the rest of the members of the household ($S_{confined}$). The rest of the individuals that became inactive but did not successfully isolate themselves completely ($S_{inactive}$) mix with members of other households due to their *social permeability*. These three fractions of the *Susceptible* compartment can be represented as:

$$S_{active}(t) = S(t)p_{act}(t) \quad (5)$$

$$S_{inactive}(t) = S(t)(1 - p_{act}(t))\phi \quad (6)$$

$$S_{confined}(t) = S(t)(1 - p_{act}(t))(1 - \phi) \quad (7)$$

Taking into account the aforementioned policies and every group in the *Susceptible* population, the equation governing the time evolution of the occupation of the *Exposed* compartment reads:

$$E(t+1) = S_{active}(t)P_{active}(t) + S_{inactive}(t)P_{inactive}(t) + S_{confined}(t)P_{confined}(t) + (1 - \eta)E(t), \quad (8)$$

where $P_{active}(t)$, $P_{inactive}(t)$ and $P_{confined}(t)$ account for the probability that *Susceptible* individuals belonging to each of these groups contract the disease at time t . We assume that the number of contacts made by each group ($k_{active}(t)$ and $k_{inactive}(t)$) depends on the mobility levels in their respective settings, yielding:

$$k_{active}(t) = \langle k_{active} \rangle p_{act}(t), \quad (9)$$

$$k_{inactive}(t) = \langle k_{inactive} \rangle p_{pres}(t), \quad (10)$$

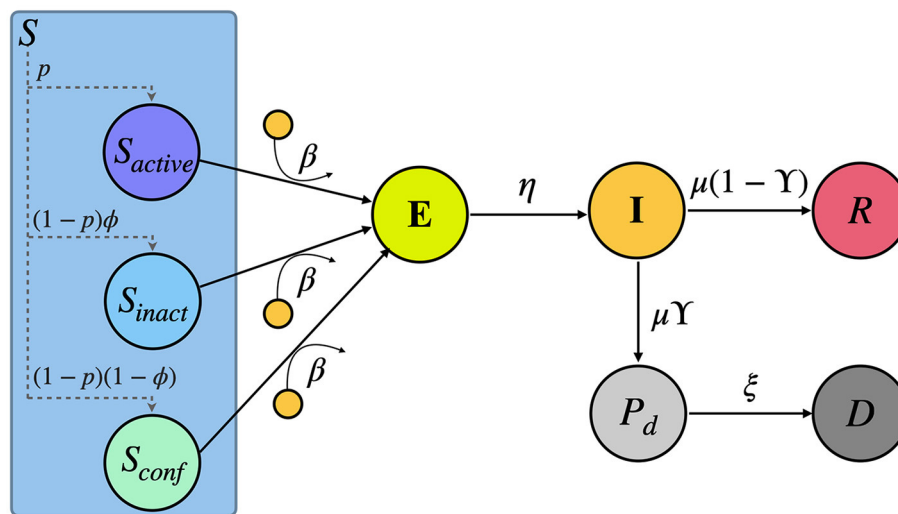


FIGURE 1

Scheme of the compartmental model here proposed. The model comprises six compartments: *Susceptible S*, *Exposed E*, *Infectious I*, *Recovered R*, *Pre-deceased Pd* and *Deceased D*. Note that, as a result of the non-pharmaceutical interventions, the *Susceptible* compartment is divided into three sub-compartments: S_{active} , $S_{inactive}$, and S_{conf} representing a fraction $p(t)$, $(1-p(t))\phi$ and $(1-p(t))(1-\phi)$ of the total number of *Susceptible* individuals, respectively. A detailed explanation of the flows connecting these compartments can be found in the literature 2.

where $p_{act}(t)$ represents the observed mobility of the people that travel to their daily destinations and $p_{res}(t)$ is the mobility of those who remain inactive in their residential areas. Likewise, $\langle k_{active} \rangle$ corresponds to all contacts made by individuals in the baseline scenario whereas $\langle k_{inactive} \rangle$ constitutes their interactions at home. Both can be estimated from social data existing in the literature (16).

Assuming a well-mixed population, the probabilities of contracting the disease can be calculated as:

$$P_{active}(t) = 1 - \left(1 - \beta \frac{I(t)}{N}\right)^{k_{active}(t)}, \quad (11)$$

$$P_{inactive}(t) = 1 - \left(1 - \beta \frac{I(t)}{N}\right)^{k_{inactive}(t)}, \quad (12)$$

for the active and inactive population. In addition, there is a chance that confined individuals contract the disease from other infectious members of their social bubble. These agents make $k_{inactive}$ contacts with others in their own household. The probability of getting infected for this group of individuals depends on the number of infected people in their households as:

$$P_{confined}(t) = 1 - \sum_{i=0}^{\sigma-1} p(i) \left(1 - \beta \frac{i}{\sigma-1}\right)^{k_{inactive}(t)}, \quad (13)$$

where σ is the number of people in each household (meaning that a susceptible member is able to make contacts with the other $\sigma-1$ residents) and the probability of finding i infected individuals in a household is

$$p(i) = \binom{\sigma-1}{i} \left(\frac{I(t)}{N}\right)^i \left(1 - \frac{I(t)}{N}\right)^{\sigma-1-i}. \quad (14)$$

To round off, we assume a closed population so that the occupation of the *S* compartment changes with:

$$S(t+1) = N - E(t+1) - I(t+1) - R(t+1) - P_d(t+1) - D(t+1). \quad (15)$$

2.2. Data sources

2.2.1. COVID-19 deaths

To calibrate our model, we rely on data regarding the daily number of fatalities in each country. As consistency across populations is crucial, the number of detected cases is not a suitable metric due to surveillance issues that affect countries differently (17, 18). There are several contributing factors to the possible discrepancies on the reported cases. Firstly, testing strategies can vary across populations, with some focusing on high-risk groups or areas, while not accounting for asymptomatic cases. Additionally, limited testing capacity can lead to an underestimation of the true number of infections, particularly in areas with high community transmission. Furthermore, differences in the definition and reporting of cases, as well as demographic variations such as age, gender, and underlying health conditions, can make it challenging to compare trends between different countries or regions.

The accuracy of reported deaths is also not guaranteed for similar reasons, including possible changes in definition and reporting delays. To address these reporting issues, we specifically choose countries with continuous and consistent records of this information, which are listed in Table 1. The selection of the countries was performed according to two criteria: there must be a peak of at least 10 deaths, and the daily mobility and number of deaths have to be available and consistent for this entire duration.

TABLE 1 Average contacts of active, $\langle k_{active} \rangle$, and inactive, $\langle k_{inactive} \rangle$, average household size σ , GDP per capita and minimum level of mobility during lockdown for each country.

Country	Country code	$\langle k_{active} \rangle$	$\langle k_{inactive} \rangle$	σ	GDP per capita (US\$)	p_{min}
Argentina	AR	14.12	3.84	2.95	8475.73	0.15
Austria	AT	12.48	3.13	2.27	48105.63	0.17
Bangladesh	BD	16.34	3.55	4.26	2000.64	0.25
Belgium	BE	11.38	2.88	2.36	45028.32	0.21
Bolivia (Plurinational State of)	BO	17.04	3.03	3.53	3133.1	0.12
Bulgaria	BG	12.73	4.06	2.34	10058.08	0.35
Canada	CA	12.57	3.1	2.45	43559.71	0.43
Chile	CL	13.71	3.66	3.04	13231.71	0.31
Colombia	CO	15.26	3.6	3.53	5332.77	0.18
Egypt	EG	15.67	3.64	4.13	3608.84	0.38
France	FR	11.78	3.1	2.22	38958.6	0.12
Germany	DE	6.86	1.79	2.05	45908.72	0.38
Greece	GR	11.79	3.18	2.44	18117.07	0.21
Guatemala	GT	18.98	4.08	4.81	4331.69	0.31
Honduras	HN	18.27	4.23	3.87	2405.73	0.17
Hungary	HU	12.07	3.46	2.6	16128.65	0.39
Indonesia	ID	15.26	3.14	3.86	3869.59	0.54
Iraq	IQ	20.64	4.35	6.35	4145.86	0.33
Ireland	IE	12.47	3.43	2.83	86250.99	0.15
Israel	IL	13.6	3.84	3.14	47033.59	0.14
Italy	IT	14.37	2.94	2.4	31238.05	0.11
Kuwait	KW	16.42	4.16	5.8	24809.04	0.12
Luxembourg	LU	16.37	3.46	2.41	117181.7	0.17
Malaysia	MY	15.4	3.59	4.56	10401.79	0.2
Mexico	MX	15.42	4.04	3.75	8325.57	0.4
Morocco	MA	14.3	3.72	4.58	3108.18	0.18
Nigeria	NG	20.47	4.16	4.66	2085.47	0.44
Pakistan	PK	18.65	4.18	6.8	1167.22	0.31
Panama	PA	14.62	3.45	3.64	12269.05	0.14
Philippines	PH	17.04	3.72	4.23	3298.83	0.17
Poland	PL	13.93	4.42	2.81	15764.11	0.32
Portugal	PT	11.88	3.03	2.66	22413.04	0.23
Romania	RO	11.97	3.41	2.88	12928.58	0.27
Russian Federation	RU	12.88	3.42	2.58	10165.51	0.45
Saudi Arabia	SA	15.64	4.1	5.6	20110.32	0.27
South Africa	ZA	15.92	3.94	3.36	5094.38	0.24
Spain	ES	12.02	3.19	2.69	27408.63	0.08
Switzerland	CH	13.14	3.11	2.21	86918.65	0.19
Turkiye	TR	13.72	3.72	4.07	8538.13	0.27
Ukraine	UA	12.65	3.48	2.53	3557.48	0.47
The United Kingdom	GB	9.48	2.25	2.27	40718.22	0.22
United States of America	US	12.6	3.24	2.49	63122.59	0.54

The data used in this study is extracted from the official daily counts of COVID-19 deaths reported for countries by the [World Health Organization](#) and smoothed using a 7-day rolling average.

2.2.2. Mobility reduction data

We extract the level of mobility inside each country at a certain time, $p_{act}(t)$ and $p_{res}(t)$, from the [Google COVID-19 Community Mobility Reports](#). Among the different types of movements included in this study, we focus on the *Retail and Recreation* category for the active individuals and the *Residential* category for the inactive ones. For each day, the level of mobility is computed by comparing the amount of flows recorded for this day with their median values measured in a pre-pandemic baseline scenario, spanning 5 weeks from January 3 to February 6, 2020. To reduce data noise, we smooth the curves using a 7-day rolling average.

2.2.3. Socioeconomic data

A key component of the epidemiological model is the average number of people an individual encounters (contacts) throughout their day, which varies from one country to another and has been measured in a certain number of them taking into account the heterogeneities in the populations. The dataset used for the simulation comes from a study (16) where the authors extrapolate the known data to 152 countries and provide contact matrices representing the number of contacts a person of each age group has with the others in different settings. For each country, we obtain the average number of contacts from an active individual (k_{active}) as the weighted average of the total number of contacts made by each individual of each group in all the settings, taking into account the population age pyramid of this country (19). To compute the same quantity for inactive (controlled) individuals ($k_{inactive}$), we repeat the same process by just accounting for the contacts made at home.

The average number of residents in a single home is also an important parameter of the model, and was reported by the United Nations (20) for most countries in the world. The household size σ is available at <https://www.un.org/development/desa/pd/data/household-size-and-composition>, and in the model has been rounded to the nearest integer in order to follow the equations.

Lastly, the Gross Domestic Product (GDP) per capita is taking into account to find a relation between the wealth of different populations and the success of their confinement policies. It is available at www.worldbank.org.

These country-dependent parameters (k_{active}), ($k_{inactive}$) and the GDP per capita) are summarized in [Table 1](#).

2.2.4. Epidemiological parameters

Some of the parameters in relation to the compartmental model have already been determined and are fixed based on the literature:

- η : Probability of leaving the *E* compartment. It is related to the average duration of the incubation period. We fix its value to $\eta = 1.0/5.2$ (21).
- μ : Probability of leaving the *I* compartment. It is related to the average duration of the infectious windows after contracting and incubating the virus. We fix its value to $\mu = 1.0/4.2$ (21).

- Υ : Infection fatality rate which, as reported in (22) and (23), is estimated to be $\Upsilon = 0.01$.

2.3. Model calibration

2.3.1. Approximate Bayesian Computation (ABC)

The Approximate Bayesian Computation (ABC) method, as described in (24) and (25), provides a solution to Bayesian inference problems where computing the likelihood function and its further exploration becomes cumbersome. ABC works by generating synthetic trajectories using a set of parameters and then accepting or rejecting those parameters based on how well the synthetic trajectories match real data. This approach allows the construction of approximate posterior distributions.

There are several ways of exploring the posterior distribution of the parameters, one of the simplest being the ABC rejection algorithm (26), which is used in our case. To quantify the goodness of a given trajectory generated by a set of parameters $\vec{\theta}$, we use a logarithmic distance function $\rho(\vec{\theta})$ defined as:

$$\rho(\vec{\theta}) \equiv \sum_t \log[|D_{obs}(t) - D_{\vec{\theta}}(t)| + 1], \quad (16)$$

where $D_{obs}(t)$ represents the observed daily fatalities at time t and $D_{\vec{\theta}}(t)$ its value predicted by the synthetic trajectory. Note that, among all possible choices for this goodness function, we have chosen a logarithmic function not to under-represent the initial stage of the epidemics, where there are fewer deaths.

The ABC rejection algorithm builds the posterior distribution for the model parameters by sampling them from the trajectories fulfilling that $\rho(\vec{\theta}) < \epsilon$, where ϵ is a tolerance threshold. In our case, we run two rounds of the ABC rejection algorithm. In the first round, we draw $20 \cdot 10^6$ random samples of variables from the prior distributions and compute the distance between the synthetic trajectories and the real data by taking into account the period between February 20 and May 20, 2020. We set a dynamical threshold ϵ_i to accept those 1,000 trajectories providing the best fits for the data in each country. We construct the prior distributions for the second from the accepted trajectories in the first one and the process is repeated. This second iteration allows the algorithm to give more accurate results for each country due to the intrinsic variability of the parameters across countries.

2.3.2. Model free parameters

The numerical iteration of Equation (1) allows one to obtain synthetic trajectories capturing the evolution of individuals in each of the compartments. The parameters of the model that are not fixed from the literature will be left for calibration via the ABC method. These parameters are:

- β : This parameter represents the probability of infection, which varies from one country to another due to factors such as population density, urbanization, and use of masks. The prior distribution of this parameter is $\beta \sim U(0.01, 0.3)$.
- ϕ : The permeability of the confinement, which is the main objective of the study to fit. A low permeability means a high

effectiveness of the non-pharmaceutical policies due to a good compliance from the population. The prior distribution of this parameter is $\phi \sim U(0, 1)$.

- ξ : The probability of leaving the P_d compartment to die because of the disease. The prior distribution of this parameter is $\xi \sim U(1/18, 1/6)$.
- T : The estimated number of days that have elapsed since the first case of the disease in the country and the day chosen as the starting point for comparison with observed deaths, which for all countries corresponds to 2020-02-20. The prior distribution of this parameter reads $T \sim U(0, 100)$.
- δ : The delay in death notification. The prior distribution of this parameter is set at $\delta \sim U(2, 20)$.

Note that δ is not a strictly needed parameter to run the model but becomes essential to make synthetic trajectories compatible with real data (27). In all the cases, the prior distributions chosen are broad enough to avoid biasing the inference of the posterior distributions.

2.4. Relationship between GDP per capita and countries permeability

In this section, we explain how we link the inferred permeability distributions of individual countries with certain economic indicators, such as GDP per capita. To obtain a meaningful association, we should exclude countries for which the model does not provide a reasonable fit, as well as those where mobility limitations did not substantially impact the control of the epidemic.

On the one hand, to determine which countries have been correctly modeled, we calculate the relative area between the estimated curve and the real data using the following equation:

$$\varepsilon(\vec{\theta}) = \frac{\sum_t |D_{obs}(t) - D_{\vec{\theta}}(t)|}{\sum_t D_{obs}(t)}. \quad (17)$$

For the subsequent analysis, we discard those countries for which $\min(\varepsilon(\vec{\theta})) > 0.4$ as the model does not fit well the epidemic trajectories there.

Once these countries are discarded, we quantify the relationship between permeability and GDP per capita by performing a non-linear regression fitting the permeability to the following function:

$$\phi(x) = ax^{-b} \quad (18)$$

where x is the GDP per capita. As our information about permeability comes from posterior distributions, we conduct 1,000 independent fits by sampling diverse sets of permeability values from these distributions. The confidence interval of the regression curve is calculated as the percentile 2.5 to the percentile 97.5 of the individuals fits obtained.

3. Results

We calibrate our model to real data using the Approximate Bayesian Computation (ABC) scheme described in the Methods section. The results of the calibration for each considered country

can be found in [Supplementary Figure 1](#). One important finding is that, despite its simplicity, our simple model accurately captures the time evolution of reported deaths for most countries and confirms the assumption that the mobility reduction has a direct effect on the number of daily contagions, which decreases as stricter policies are put in place. [Figure 2](#) illustrates this by showing the real and simulated epidemic trajectories for Spain, Colombia, and Ukraine. The selected countries represent three distinct types of behavior observed in our study. While Spain and Colombia implemented similar lockdown policies resulting in comparable reductions in average mobility, their outcomes were vastly different: Spain managed to stop the spread and bend the curve, whereas Colombia experienced a steady growth in casualties. This pattern of steady increase is also observed in Ukraine, which had a milder reduction in mobility compared to Spain and Colombia.

The unequal impact of mobility reduction on epidemic containment can be captured by the social permeability ϕ parameter in our model, which modulates the effective reproductive number of the circulating virus, as explained in (28). Namely, low permeability values significantly reduce the pool of susceptible individuals exposed to the virus due to the lower household mixing, whereas high permeability values means that all the individuals remain vulnerable to the virus but with a reduced exposure due to their hampered social activity.

In [Figure 3](#), we present the posterior distributions obtained of social permeability for each country analyzed in this study. Focusing on the specific case of Colombia and Spain, we confirm that lower efficiency of mobility reductions in Colombia translates to higher permeability values compared to those inferred for Spain. While not the focus of our manuscript, other model parameters also provide insightful information about the impact of the first COVID-19 epidemic wave and the associated contention measures across countries. For instance, the inferred values of parameter T enable to reconstruct the time of onset of the outbreak in each country, whereas parameter δ accounts for the heterogeneous delay in reporting deaths. Nonetheless, conclusions on these parameters should be drawn with caution because of the correlations between their posterior distributions, as illustrated in [Supplementary Figures 2–4](#) for the case of Spain, Colombia and Ukraine, respectively.

To round off, we check whether we can connect the heterogeneous permeability values inferred for each country with their corresponding socioeconomic features. In order to establish a meaningful link, we narrow our focus to countries where the model accurately predicts the course of the disease. We determine this accuracy by calculating the normalized distance $\varepsilon(\vec{\theta})$ between the data and the model trajectories, which enables us to establish a threshold and exclude countries where the model does not perform well. This procedure is described in more detail in the Section 2 and the distribution of the minimum normalized distances $\varepsilon^{\min}$ observed across countries is represented in [Supplementary Figure 5](#).

[Figure 4](#) represents the posterior distribution of the social permeability against the gross domestic product (GDP) per capita of the selected countries. The tendency showcased in the figure indicates that there is in fact a negative statistically significant correlation between the wealth of a country and the ability of its inhabitants to properly follow the restrictions and stay in lockdown.

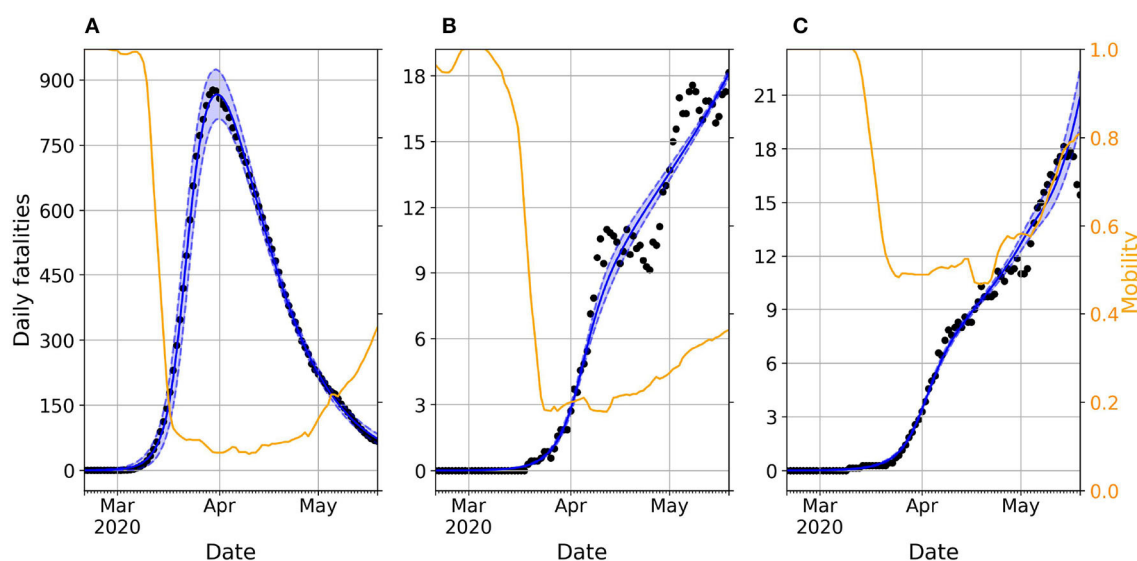


FIGURE 2

Daily evolution of the number of deaths in Spain (A), Colombia (B), and Ukraine (C). In all the panels, dots represent real reported data whereas the blue shadowed region corresponds to the 95% prediction interval of the accepted trajectories after calibrating the model. The blue solid line represents the median trajectory whereas the orange line corresponds to the time variation of mobility compared with a baseline pre-pandemic scenario spanning from January 3 to February 6, 2020.

This negative correlation between the permeability and GDP per capita is further supported by the non-linear regression of the data described in the Section 2.

For the sake of completeness, we study the influence of possible confounding factors on this correlation such as mobility reduction and deaths caused by the disease. [Supplementary Figure 6](#) shows that the permeability values have no correlation with the minimum observed mobility for each one of them, meaning the model can separate the level of mobility reduction and the effectiveness of the confinement without one depending on the other. Regarding the relationship between permeability and death toll per capita in each country, we observe in [Supplementary Figure 7](#) low permeability values for those countries with higher number of fatalities but a large variability without any clear relationship in those less severely affected.

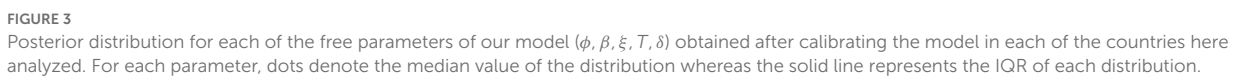
4. Discussion

The COVID-19 pandemic has had an undeniable impact on the world and has exposed the existing social and economic inequalities within many countries (29, 30). To address this issue, various countries have implemented policies and non-pharmaceutical interventions to control the virus's spread and reduce the number of casualties. However, the pandemic's impact has not been evenly distributed across society, with certain groups suffering more severe consequences than others (31, 32). This inequality is not limited to individual countries, but also occurs across nations due to various challenges that low income countries face in implementing measures to prevent transmission. Low income countries encounter numerous obstacles, such as inadequate infrastructure, a lack

of public trust, and a high percentage of individuals working in the informal sector, who cannot work remotely from home and lose their source of income (33–35). These factors have resulted in significant challenges in controlling the virus in many low income countries, underscoring the pressing need for a global effort to address the pandemic equitably and effectively.

The focus of our research has been to explore the impact of socioeconomic determinants on the efficacy of stay-at-home measures in controlling the spread of COVID-19. By using the change in mobility as a metric for the strictness of the restrictions, without the assumption that they are an accurate quantitative representation of the real level of confinement that the population went through, we have been able to replicate the epidemic trajectory in 42 countries. Our findings indicate that a reduction in mobility is strongly associated with a decrease in virus transmission. However, we recognize that this metric may not always be an accurate representation of the true level of confinement experienced by the population. To address this, we introduced the concept of social permeability, which was estimated using Approximate Bayesian Computation. Our results suggest that low-income countries tend to have a higher permeability, indicating that restrictions were less effective in achieving an efficient population confinement.

Finally, the framework here proposed constitutes a minimal approach to capture the evolution of COVID-19 pandemic under mobility restrictions and presents different limitations. First, the model assumes a well-mixed population inside each country, neglecting possible spatial heterogeneities existing among its different regions and resulting in the aggregated values for the variables not representing a fair indicator of the evolution of the pandemic. In this case, the aggregated values for these variables do not represent a fair indicator of the evolution of the pandemic.



from smartphone users that have opted in to Google's Location History feature, which is off by default. Because of this, the results are based on the assumption that these users represent the behavior of the entire population in their respective countries. Despite all these limitations, we hope that our model paves the way to the elaboration of more sophisticated frameworks addressing the relevance of the interplay socioeconomic features and mobility reductions during epidemic outbreaks.

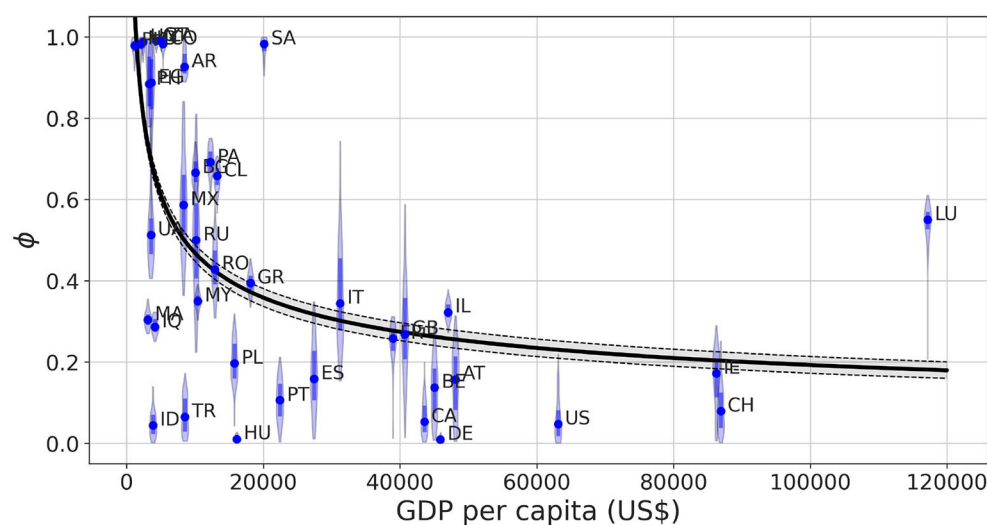


FIGURE 4

Posterior distribution obtained for the permeability parameter ϕ as a function of the GDP per capita of the country in which the model is calibrated. The shadowed region of the fit shows the 95% prediction interval of the trajectories obtained via non-linear regression $\phi(x) = ax^{-b}$, where x stands for GDP per capita and the parameters result in $a = 16 \pm 4.12$, $b = 0.39 \pm 0.03$. The solid line represents the average value of the fitted trajectories for each x . The Spearman correlation coefficient ρ_S between both variables is $\rho_S = -0.590$ with $p < 10^{-4}$.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DS-P and JG-G conceived the research project. PV and UL performed the data analysis and calibration of the compartmental model. PV, DS-P, and JG-G drafted the manuscript. All authors contributed to the study design and methodology. All authors have read and agree with the published version of the manuscript.

Funding

PV, DS-P, and JG-G acknowledge financial support from the Departamento de Industria e Innovación del Gobierno de Aragón y Fondo Social Europeo (FENOL Group Grant E36-20R) and from Grant PID2020-113582GB-I00 funded by MCIN/AEI/10.13039/501100011033.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2023.1193100/full#supplementary-material>

References

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. (2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5
- Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: a review. *Phys Rep*. (2021) 913:1–52. doi: 10.1016/j.physrep.2021.02.001
- Kapitsinis N. The underlying factors of the COVID-19 spatially uneven spread. Initial evidence from regions in nine EU countries. *Region Sci Policy Pract*. (2020) 12:1027–45. doi: 10.1111/rsp3.12340
- Kang D, Choi H, Kim JH, Choi J. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int J Infect Dis*. (2020) 94:96–102. doi: 10.1016/j.ijid.2020.03.076

5. Raharja A, Tamara A, Kok LT. Association between ethnicity and severe COVID-19 disease: a systematic review and meta-analysis. *J Racial Ethnic Health Disparities*. (2020) 8:1563–72. doi: 10.1007/s40615-020-00921-5
6. Magesh S, John D, Li WT, Li Y, Mattingly-App A, Jain S, et al. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic-review and meta-analysis. *JAMA Netw Open*. (2021) 4:e2134147. doi: 10.1001/jamanetworkopen.2021.34147
7. Benita F, Rebollar-Ruelas L, Gayton-Alfaro ED. What have we learned about socioeconomic inequalities in the spread of COVID-19? A systematic review. *Sustain Cities Soc*. (2022) 86:104158. doi: 10.1016/j.scs.2022.104158
8. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. *Math Biosci*. (2013) 244:125–34. doi: 10.1016/j.mbs.2013.04.013
9. House T, Keeling MJ. Household structure and infectious disease transmission. *Epidemiol Infect*. (2009) 137:654–61. doi: 10.1017/S0950268808001416
10. Liu P, McQuarrie L, Song Y, Colijn C. Modelling the impact of household size distribution on the transmission dynamics of COVID-19. *J R Soc Interface*. (2021) 18:20210036. doi: 10.1098/rsif.2021.0036
11. Esseau-Thomas C, Galarraga O, Khalifa S. Epidemics, pandemics and income inequality. *Health Econ Rev*. (2022) 12:7. doi: 10.1186/s13561-022-00355-1
12. Libório MP, Ekel PY, de Abreu JF, Laudares S. Factors that most expose countries to COVID-19: a composite indicators-based approach. *GeoJournal*. (2022) 87:5435–49. doi: 10.1007/s10708-021-10557-5
13. Varkey RS, Joy J, Sarmah G, Panda PK. Socioeconomic determinants of COVID-19 in Asian countries: an empirical analysis. *J Public Aff*. (2021) 21:e2532. doi: 10.1002/pa.2532
14. Baser O. Population density index and its use for distribution of COVID-19: a case study using Turkish data. *Health Policy*. (2021) 125:148–54. doi: 10.1016/j.healthpol.2020.10.003
15. Keeling M, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press (2008).
16. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol*. (2017) 13:e1005697. doi: 10.1371/journal.pcbi.1005697
17. Reese H, Iuliano AD, Patel NN, Garg S, Kim L, Silk BJ, et al. Estimated incidence of coronavirus disease 2019 (COVID-19) illness and hospitalization—United States, February–September 2020. *Clin Infect Dis*. (2020) 72:e1010–7. doi: 10.1093/cid/ciaa1780
18. Iuliano AD, Chang HH, Patel NN, Threlkel R, Kniss K, Reich J, et al. Estimating under-recognized COVID-19 deaths, United States, March 2020–May 2021 using an excess mortality modelling approach. *Lancet Region Health Am*. (2021) 1:100019. doi: 10.1016/j.lana.2021.100019
19. Nations U. *World Population Prospects, The 2019 Revision - Volume I: Comprehensive Tables*. United Nations (2019). Available online at: <https://www.un-ilibrary.org/content/books/9789210046428>
20. United Nations. *Household Size and Composition Around the World*. Economic and Social Affairs (2017).
21. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Internal Med*. (2020) 172:577–82. doi: 10.7326/M20-0504
22. Brazeau NF, Verity R, Jenks S, Fu H, Whittaker C, Winskill P, et al. Estimating the COVID-19 infection fatality ratio accounting for seroreversion using statistical modelling. *Commun Med*. (2022) 2:54. doi: 10.1038/s43856-022-00106-7
23. Bar-On YM, Flamholz A, Phillips R, Milo R. Science Forum: SARS-CoV-2 (COVID-19) by the numbers. *eLife*. (2020) 9:e57309. doi: 10.7554/eLife.57309
24. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian computation. *PLoS Comput Biol*. (2013) 9:e1002803. doi: 10.1371/journal.pcbi.1002803
25. Csilléry K, Blum MG, Gaggiotti OE, François O. Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol*. (2010) 25:410–8. doi: 10.1016/j.tree.2010.04.001
26. Gelman A. A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *Int Stat Rev*. (2003) 71:369–82. doi: 10.1111/j.1751-5823.2003.tb00203.x
27. Gutierrez E, Rubli A, Tavares T. Delays in death reports and their implications for tracking the evolution of COVID-19. *Covid Econom*. (2020) 1:116–44. Available online at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3645304
28. Arenas A, Cota W, Gómez-Gardeñes J, Gómez S, Granell C, Matamalas JT, et al. Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Phys Rev X*. (2020) 10:041055. doi: 10.1103/PhysRevX.10.041055
29. Van Dorn A, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *Lancet*. (2020) 395:1243–4. doi: 10.1016/S0140-6736(20)30893-X
30. Wachtler B, Michalski N, Nowossadeck E, Diercke M, Wahrendorf M, Santos-Hövenier C, et al. Socioeconomic inequalities and COVID-19—A review of the current international literature. *J Health Monitor*. (2020) 5(Suppl 7):3. doi: 10.25646/7059
31. Arceo-Gomez EO, Campos-Vazquez RM, Esquivel G, Alcaraz E, Martinez LA, Lopez NG. The income gradient in COVID-19 mortality and hospitalisation: an observational study with social security administrative records in Mexico. *Lancet Region Health Am*. (2022) 6:100115. doi: 10.1016/j.lana.2021.100115
32. Drefahl S, Wallace M, Mussino E, Aradhya S, Kolk M, Brandén M, et al. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nat Commun*. (2020) 11:1–7. doi: 10.1038/s41467-020-18926-3
33. Lou J, Shen X, Niemeier D. Are stay-at-home orders more difficult to follow for low-income groups? *J Transport Geogr*. (2020) 89:102894. doi: 10.1016/j.jtrangeo.2020.102894
34. Brodeur A, Grigoryeva I, Kattan L. Stay-at-home orders, social distancing, and trust. *J Popul Econ*. (2021) 34:1321–54. doi: 10.1007/s00148-021-00848-z
35. Huang X, Lu J, Gao S, Wang S, Liu Z, Wei H. Staying at home is a privilege: evidence from fine-grained mobile phone location data in the United States during the COVID-19 pandemic. *Ann Am Assoc Geograph*. (2022) 112:286–305. doi: 10.1080/24694452.2021.1904819

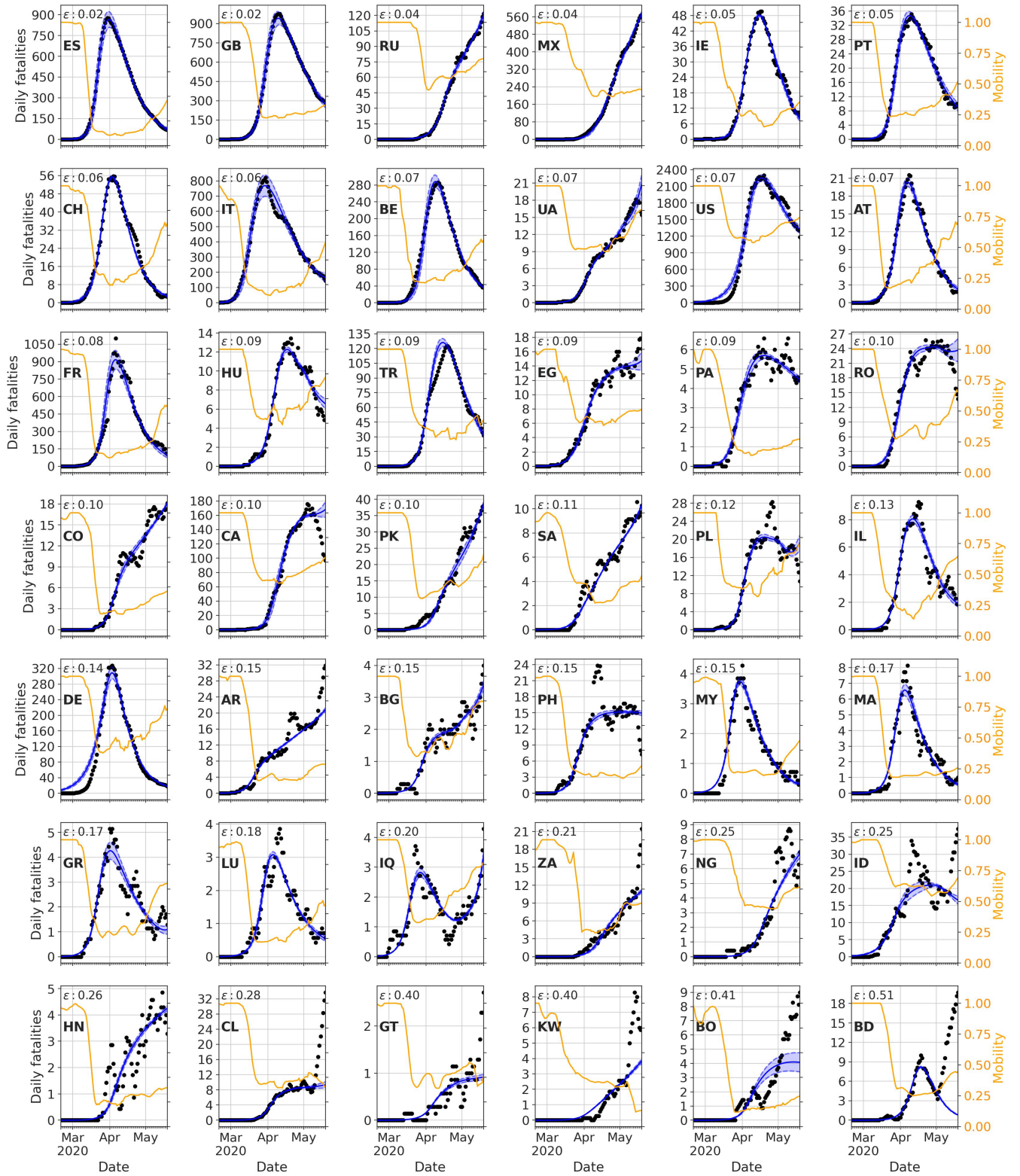


Figure S1. Daily evolution of the number of deaths in each of the 42 countries here analyzed. In all the panels, dots represent real reported data whereas the blue shadowed region corresponds to the 95% prediction interval of the accepted trajectories after calibrating the model. The blue solid line represents the median trajectory whereas the orange line corresponds to the time variation of mobility compared with a baseline pre-pandemic scenario spanning from January 3 to February 6, 2020.

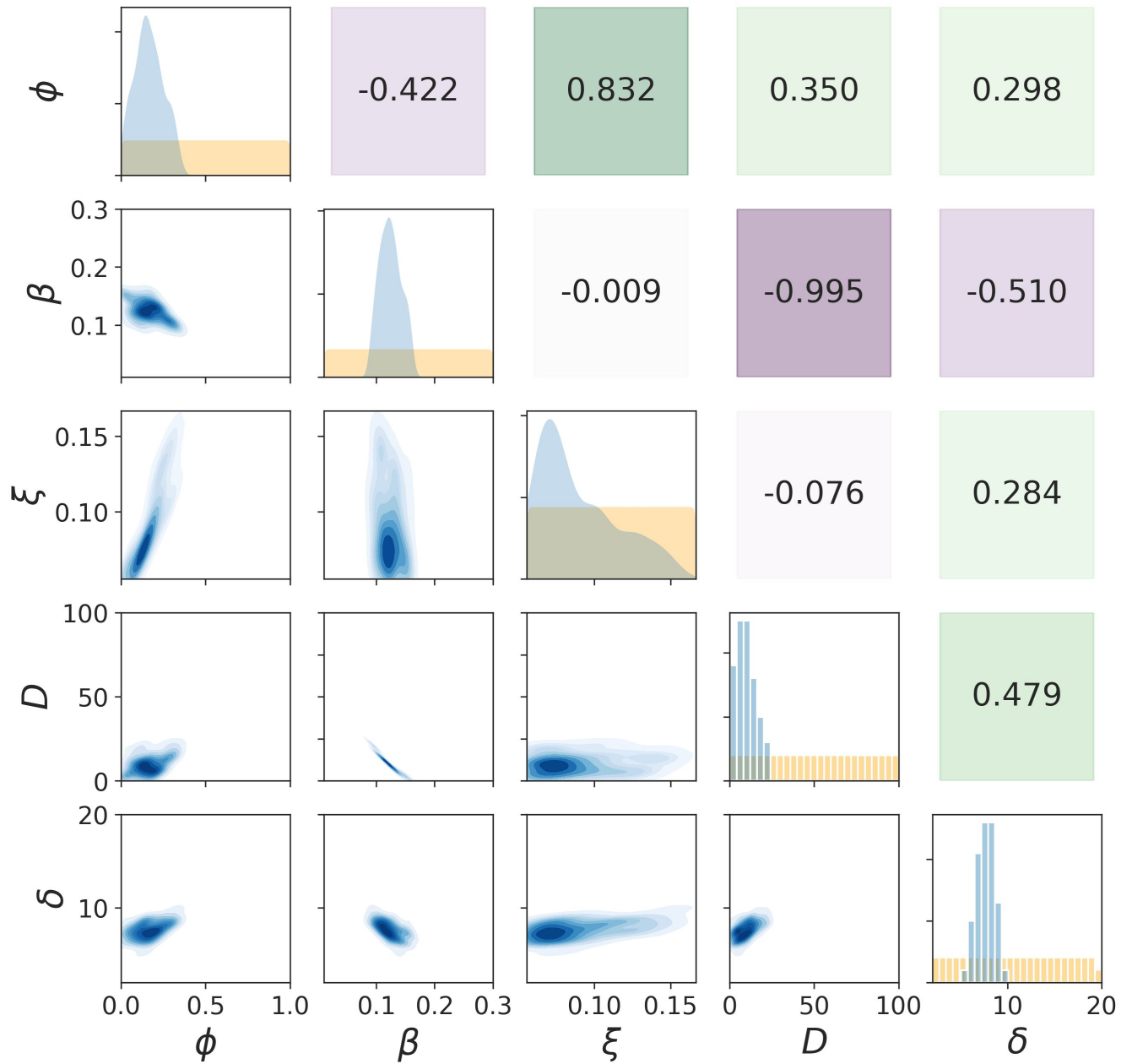


Figure S2. Model calibration in Spain. Panels in the diagonal show the prior distribution for each parameters (orange) and its posterior distribution after calibration (blue). Lower diagonal panels represent the scatter plots relating the pairwise combinations of values in the accepted trajectories whereas the upper diagonal panels contains the Spearman correlation coefficient between each pair of set of parameters found in the accepted trajectories.

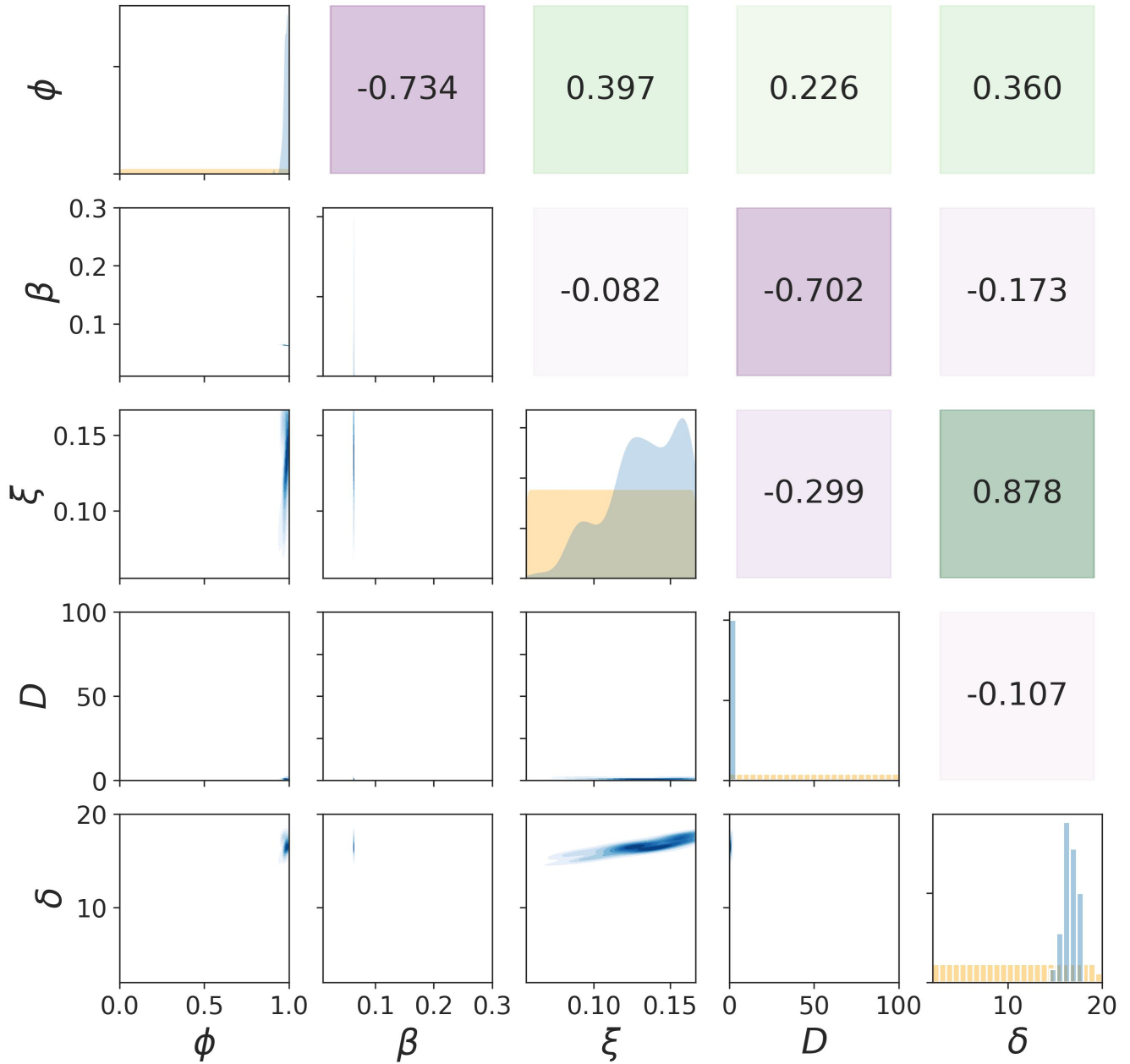


Figure S3. Model calibration in Colombia. Panels in the diagonal show the prior distribution for each parameters (orange) and its posterior distribution after calibration (blue). Lower diagonal panels represent the scatter plots relating the pairwise combinations of values in the accepted trajectories whereas the upper diagonal panels contains the Spearman correlation coefficient between each pair of set of parameters found in the accepted trajectories.

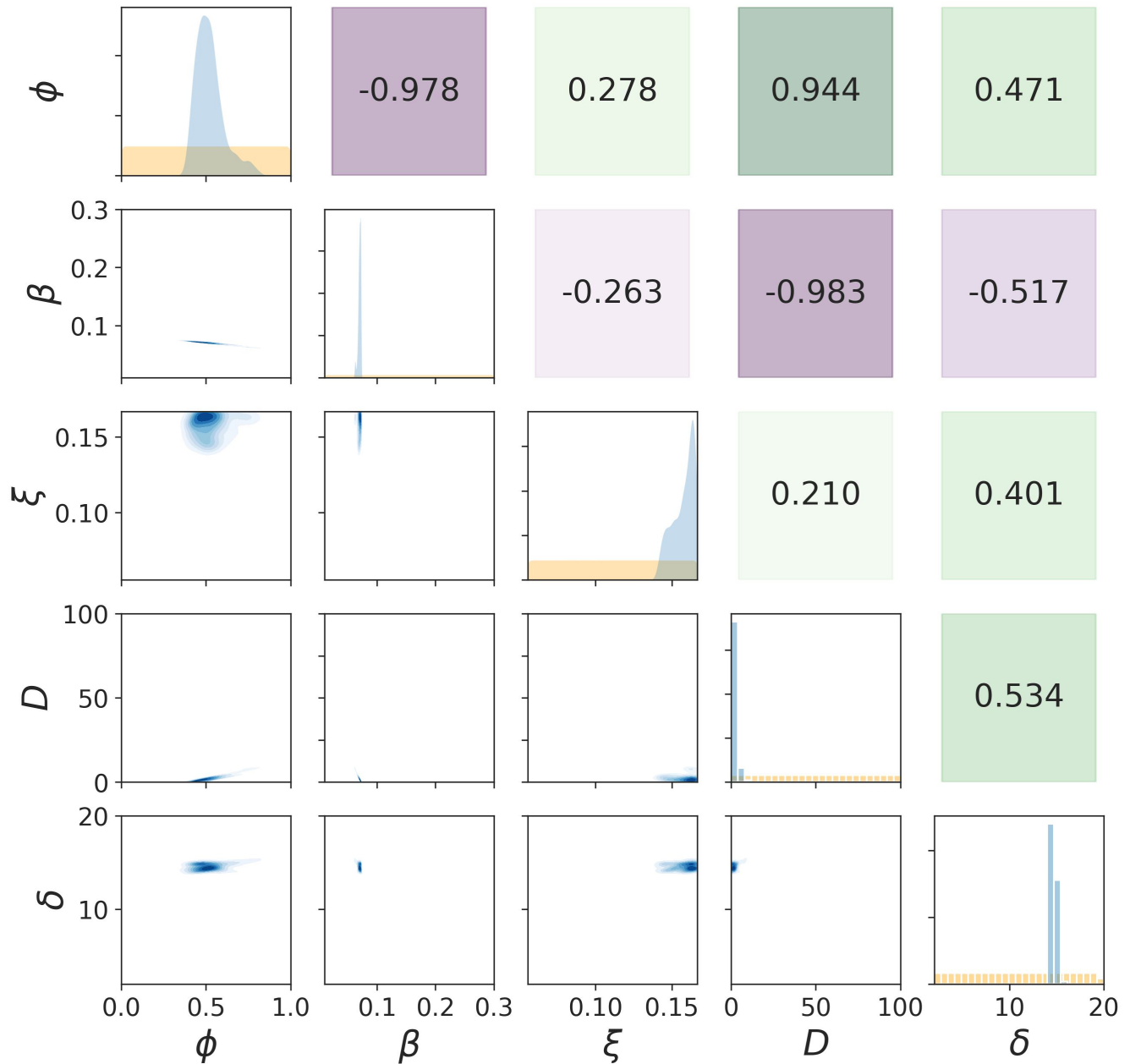


Figure S4. Model calibration in Ukraine. Panels in the diagonal show the prior distribution for each parameters (orange) and its posterior distribution after calibration (blue). Lower diagonal panels represent the scatter plots relating the pairwise combinations of values in the accepted trajectories whereas the upper diagonal panels contains the Spearman correlation coefficient between each pair of set of parameters found in the accepted trajectories.

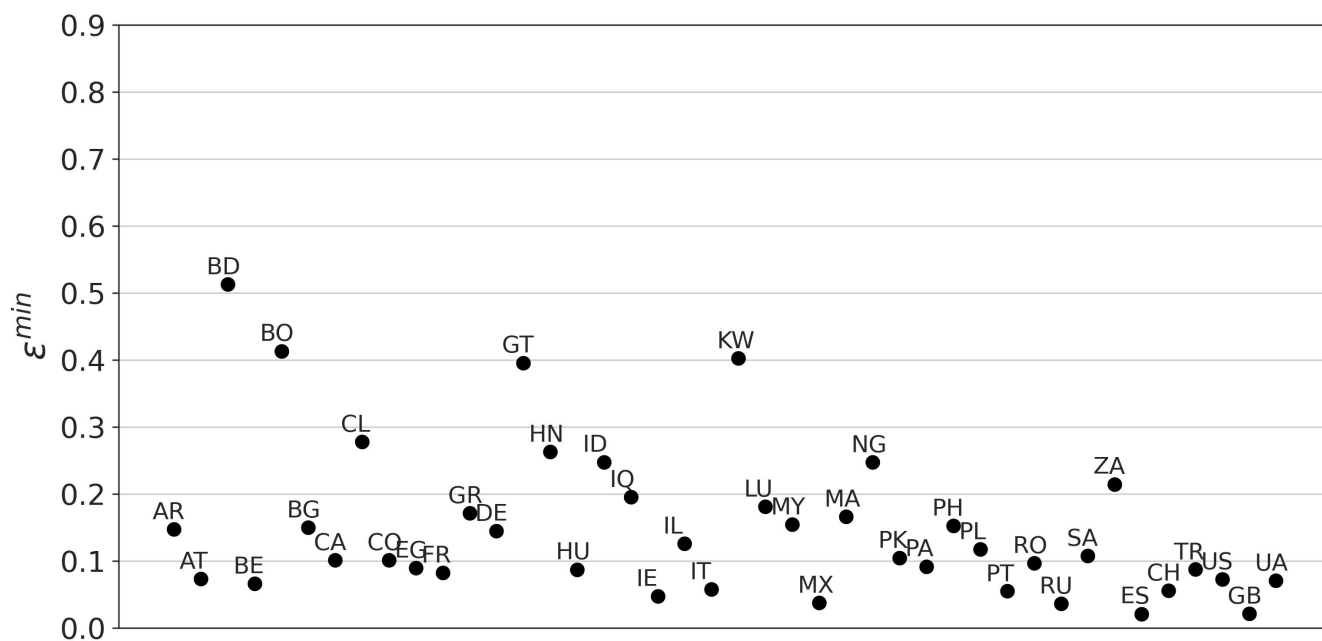


Figure S5. Minimum relative error of the fit $\varepsilon^{\min}$ for every country as defined in equation (13).

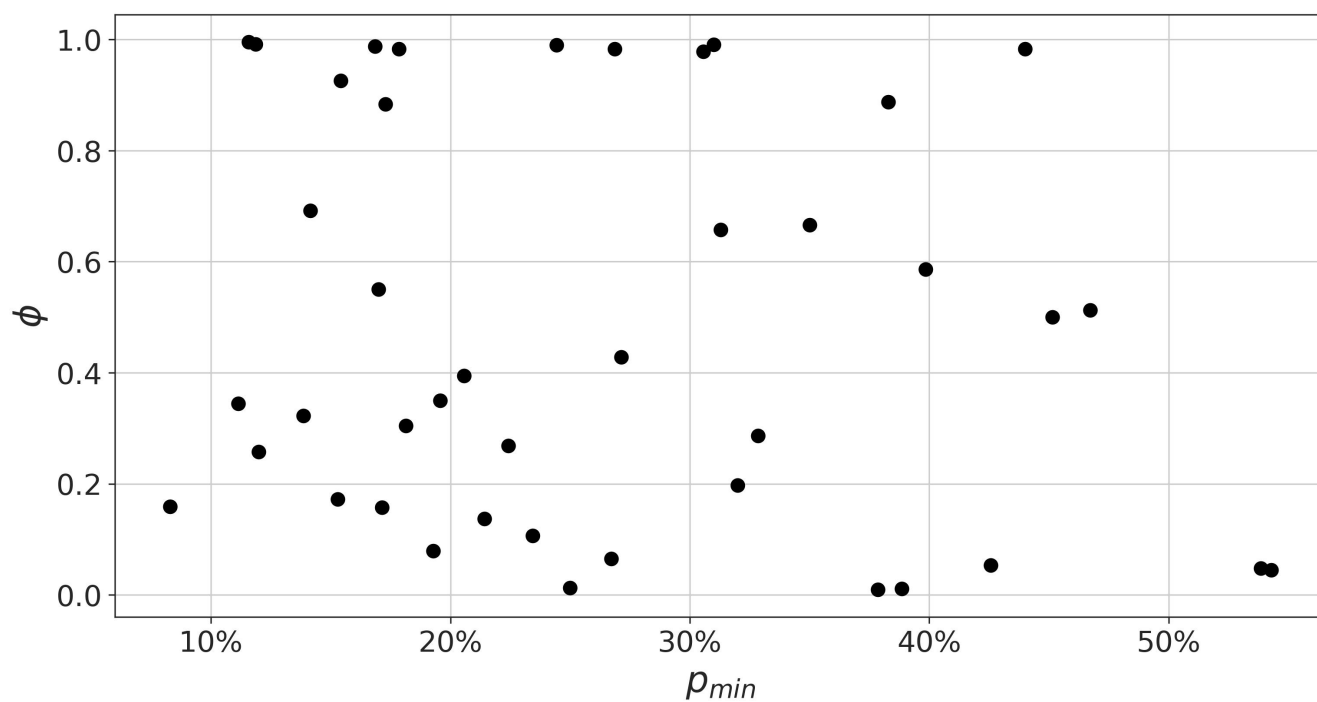


Figure S6. Median of the permeability distribution from the fit of each country as a function of the minimum mobility reached by its population.

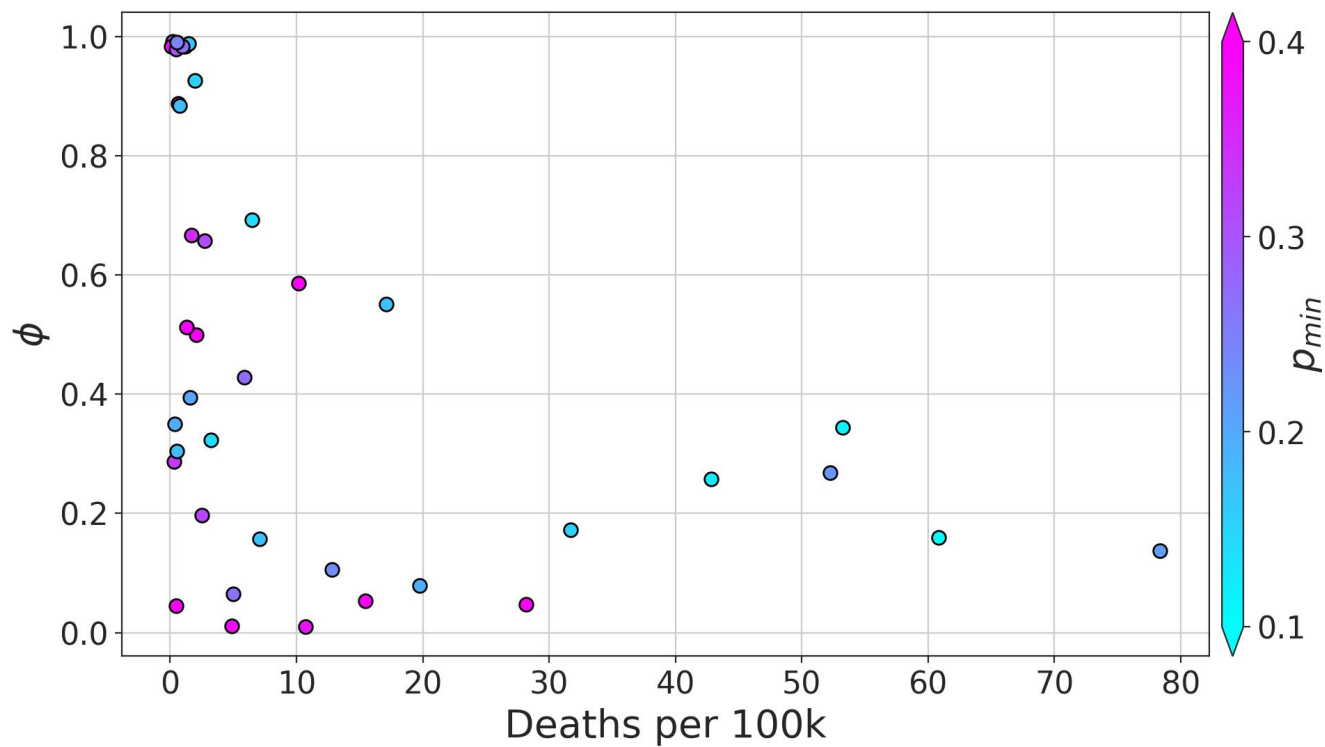


Figure S7. Median of the permeability distribution from the fit of each country as a function of the number of reported deaths during the studied period. The color code represents the minimum mobility achieved by the population of each country in relation to its original pre-pandemic value.

Capítulo 3

El impacto de las políticas de cuarentena en las diferentes clases socioeconómicas a nivel urbano



OPEN Quantifying the heterogeneous impact of lockdown policies on different socioeconomic classes during the first COVID-19 wave in Colombia

Pablo Valgañón^{1,2}, Andrés F. Useche³, David Soriano-Paños^{2,4}✉, Gourab Ghoshal⁵ & Jesús Gómez-Gardeñes^{1,2}

In the absence of vaccines, the most widespread reaction to curb the COVID-19 pandemic worldwide was the implementation of lockdowns or stay-at-home policies. Despite the reported usefulness of such policies, their efficiency was highly constrained by socioeconomic factors determining their feasibility and their associated outcome in terms of mobility reduction and the subsequent limitation of social activity. Here we investigate the impact of lockdown policies on the mobility patterns of different socioeconomic classes in the three major cities of Colombia during the first wave of the COVID-19 pandemic. In global terms, we find a consistent positive correlation between the reduction in mobility levels and the socioeconomic stratum of the population in the three cities, implying that those with lower incomes were less capable of adopting the aforementioned policies. Our analysis also suggests a strong restructuring of the mobility network of lowest socioeconomic strata during COVID-19 lockdown, increasing their endogenous mixing while hampering their connections with wealthiest areas due to a sharp reduction in long-distance trips.

During the first half of 2020, the lack of approved vaccines turned non-pharmaceutical interventions^{1–4} (NPIs) into our best ally to fight the first wave of the COVID-19 pandemic worldwide. NPIs encompass a wide variety of policies implemented to reshape the interaction patterns of the population to hamper the potential contagion pathways of infectious individuals. This goal can be accomplished by targeting different factors involved in virus transmission. At the individual level, the transmissibility of a virus in contacts between the infected and susceptible population is reduced by promoting prophylactic measures such as the use of masks^{5,6}. At the community level, the exposure of susceptible individuals to a virus can be reduced by testing-trace-isolate-quarantine (TTIQ) policies^{7–9}, shortening the effective infectious window of infectious individuals and isolating their contacts, or by population wide lockdowns and stay-at-home policies, decreasing the number of acquaintances of the entire population by controlling their mobility^{10–12}.

The success of community NPIs, such as lockdown or stay-at-home policies, in reducing social contacts is not universal but strongly depends on the complex relationship existing between the socioeconomic characteristics of the population and their mobility¹³. In this sense, several studies have found remarkably different mobility patterns across socioeconomic classes classified according to several criteria such as race, ethnicity or income level^{14–19}. In the context of epidemics, it is worth remarking that, even in uncontrolled scenarios, the coexistence of multiple mobility networks might be highly relevant for the evolution of epidemic outbreaks, leading to heterogeneous epidemic trajectories across the different socioeconomic classes^{20,21}. In addition to the socioeconomic flavor of daily recurrent mobility patterns, the feasibility of the aforementioned policies also varies across different socioeconomic groups. For example, the adoption of stay-at-home policies is clearly related to the possibility of working at home, an option that is much more accessible to those individuals with higher income levels^{22–24}.

¹Departament of Condensed Matter Physics, University of Zaragoza, 50009 Zaragoza, Spain. ²GOTHAM lab, Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza, Spain. ³Department of Industrial Engineering, School of Engineering, Universidad de Los Andes, 111711 Bogotá, Colombia. ⁴Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal. ⁵Department of Physics and Astronomy, University of Rochester, Rochester, NY 14627, USA. ✉email: dpanos@igc.gulbenkian.pt

Consequently, a positive correlation between the income level of the population and the levels of mobility reduction achieved through lockdown interventions has been measured in different cities^{25,26}.

Here we aim at further investigating the interplay between socioeconomic features and the mobility reduction observed as a consequence of lockdown policies in the three major cities of Colombia: Bogotá, Medellín and Santiago de Cali. As usual in Colombian cities, the population is divided into six different socioeconomic strata according to the quality of their households and neighborhoods. This classification is used to determine the cost of public services and access to government aid. Although there is not a one-to-one correspondence between strata and socioeconomic classes, stratum 1 typically gathers those individuals with less economic resources whereas stratum 6 usually correspond to the wealthiest population. The distribution of the population across strata is not homogeneous. For instance, local surveys carried out in Bogotá in 2017²⁷ revealed that 86% of the population belongs to the strata 1–3 whereas the highest strata, i.e., strata 5–6 only gather 4.5% of its population. Similarly, the spatial distribution of these strata is not homogeneous either. For instance, in Bogotá, socioeconomic strata are highly spatially segregated, being the lowest one located to the south of the city^{28,29}, a poorly connected area with an inadequate service infrastructure and deficiencies in access to health systems. This pattern is consistent across the other cities here studied, where lower socioeconomic strata are located far from the city centers, with conditions similar to those described in the case of Bogotá^{29–31}.

Previous studies have already studied the heterogeneous impact of lockdown policies on the mobility of different socioeconomic strata in Bogotá. In particular, Dueñas et al.³² analyze time variations in the use of public transport, finding higher mobility reductions for the wealthiest socioeconomic strata and a generalized shrinkage of trips distance. Here we use another source for mobility data, which are estimated from the travel patterns of mobile phone users, and extend the analysis to more Colombian cities. The spatio-temporal resolution of this dataset allows us to construct a time-varying multiplex network containing weekly mobility patterns for each stratum. We analyze the properties of such network, providing further evidence of how socioeconomic constraints shaped the outcome of lockdown policies in Colombia. Finally, we go one step further and analyze how lockdown policies altered the structure of the mobility patterns of different strata and their social mixing. By integrating both socioeconomic and mobility data, we reveal that lockdown policies also shaped the social structure of cities, exacerbating the segregation of lowest socioeconomic strata while barely affecting the mobility patterns of the highest ones.

Methods

Mobility dataset

The mobility patterns of the population are sourced from The Google COVID-19 Aggregated Mobility Research Dataset. The dataset contains anonymized mobility flows aggregated over mobile phone users who have turned on the Location History setting, which is off by default. This is similar to the data used to show how busy certain types of places are in Google Maps—helping identify when a local business tends to be the most crowded. The data set aggregates weekly flows of people from region to region, with a spatial resolution of level 12 s2cells areas. These correspond to an area between 3.04 and 6.38 squared kilometres. By default, all metrics defined in this section correspond to weekly values, but can be generalized to quantify averages over any arbitrary number of weeks.

To produce this data set, machine learning is applied to logs data to automatically segment it into semantic trips³³. To provide strong privacy guarantees, all trips were anonymized and aggregated using a differentially private mechanism³⁴ to aggregate flows over time (see <https://policies.google.com/technologies/anonymization>). This research is done on the resulting heavily aggregated and differentially private data. No individual user data was ever manually inspected, only heavily aggregated flows of large populations were handled.

All anonymized trips are processed in aggregate to extract their origin and destination location and time. For example, if users traveled from location *a* to location *b* within time interval *t*, the corresponding cell (*a*,*b*,*t*) in the tensor would be $n \pm \text{err}$, where *err* is Laplacian noise. The automated Laplace mechanism adds random noise drawn from a zero mean Laplace distribution and yields (ϵ , δ)-differential privacy guarantee of $\epsilon = 0.66$ and $\delta = 2.1 \times 10^{-29}$ per metric. Specifically, for each week *W* and each location pair (*A*,*B*), the number of unique users who took a trip from location *A* to location *B* during week *W* is computed. To each of these metrics, Laplace noise is added from a zero-mean distribution of scale $1/0.66$. All metrics for which the noisy number of users is lower than 100 are then removed, following the process described in³⁴, and the rest published. This yields that each published metric satisfies (ϵ , δ)-differential privacy with values defined above. The parameter ϵ controls the noise intensity in terms of its variance, while δ represents the deviation from pure ϵ -privacy. The closer they are to zero, the stronger the privacy guarantees.

Demographic and socioeconomic data

We extract demographic and socioeconomic data from the 2018 Colombian census. Specifically, we use the spatial distribution of the population and the economic stratum of households aggregated to the level of census block. This way, for each census block *k*, we obtain the number of residents $\tilde{n}_k$ and the fraction of households belonging to each stratum *s*, E_k^s .

Construction of the mobility multiplex network

To construct the mobility networks associated to each stratum, we should aggregate both socioeconomic and demographic information to match the spatial resolution of mobility data, corresponding to level 12 s2cells areas as discussed above. To do this, we associate each census sector with the cell containing its centroid. Let us here remark that Colombian cities typically display high levels of segregation among socioeconomic classes¹⁵, which allows us not to lose detail or aggregate disparate population when merging mobility and socioeconomic data. After this process, the number of residents inside cell *i* from stratum *s* is computed as:

$$n_i^s = \sum_{k \in i} E_k^s \tilde{n}_k . \quad (1)$$

From these quantities, we can straightforwardly obtain the proportion of individuals belonging to a given stratum s in a cell i , R_i^s , which reads:

$$R_i^s = \frac{n_i^s}{\sum_{s=1}^6 n_i^s} . \quad (2)$$

Finally, we assume that there are no significant differences in the representativeness of each stratum in the mobility data set. The latter assumption allows us to express the flow of residents from stratum s moving from cell i to cell j at a given week t , $f_{ij}^s(t)$, as

$$f_{ij}^s(t) = F_{ij}(t) R_i^s , \quad (3)$$

where $F_{ij}(t)$ represents the origin-destination fluxes estimated from mobility data during week t . In what follows, we omit the time index in the explanation to ease notation. The sum of these trips across origins and destinations gives us the overall mobility of the population of stratum s during a week, which reads as follows:

$$f^s = \sum_{i,j} f_{ij}^s . \quad (4)$$

Likewise, for the analysis of the outward mobility at each patch, it is only necessary to sum over the destinations as

$$f_i^s = \sum_j f_{ij}^s . \quad (5)$$

Conceptually, for each week analyzed here, each city is thus represented by a multiplex mobility network $\mathbf{f}$ with $L = 6$ layers, each one associated with a different socioeconomic stratum, N patches, corresponding to the different cells in which each city is partitioned in the mobility data set. For the baseline scenario used to compute the effects of lockdown policies, we consider the mobility patterns recorded for the week starting on 02-02-2020.

Estimating mixing patterns among strata

Once we have embedded the mobility patterns of the different strata in a multiplex network, we can estimate how mixing between different strata changed over time as a consequence of the implemented lockdown policies. To this aim, we define the entries of the (6×6) mixing matrix $\mathbf{M}$, for which each entry $M_{s_1 s_2}$ quantifies the average proportion of individuals belonging to stratum s_2 in all destinations visited by individuals from stratum s_1 . Mathematically we compute these entries as:

$$M_{s_1 s_2} = \frac{\sum_{i,j} f_{ij}^{s_1} R_j^{s_2}}{\sum_{ij} f_{ij}^{s_1}} . \quad (6)$$

To compare this metric with the mean-field baseline scenario, in which the population would be well-mixed, we divide these values by the fraction of the population belonging to each stratum. This is a more representative way of analyzing the segregation of the different strata without the bias of the underlying heterogeneous distribution of the population across strata. The entries of the re-normalized mixing matrix, $\mathbf{M}_N$, are thus defined as:

$$M_{N s_1 s_2} = \frac{M_{s_1 s_2}}{n^{s_2}} \quad (7)$$

where n^{s_2} is the total population that belongs to stratum s_2 such that

$$n^s = \frac{\sum_i n_i^s}{\sum_{i,s'} n_i^{s'}} . \quad (8)$$

Each row of the mixing matrix provides information about the segregation of the stratum s_1 , which we can quantify in a single metric using Shannon entropy formula. This value, denoted by $S(s_1)$ is larger the less segregated the stratum is, as the components of the row become more homogeneous. We compute it as:

$$S(s_1) = - \sum_{s_2} M_{s_1 s_2} \ln (M_{s_1 s_2}) . \quad (9)$$

Another metric providing useful information about how people from different strata mix is the average distance that any person needs to travel to reach her destination. This distance can change over time and depends on the stratum of the individual. In particular, the average distance an individual from stratum s travels, denoted in what follows by $\langle d^s \rangle$ is calculated as:

$$\langle d^s \rangle = \frac{\sum_{ij} d_{ij} f_{ij}^s}{\sum_{ij} f_{ij}^s}, \quad (10)$$

where d_{ij} is the geometric distance between the centroids of geographical areas i and j . Analogously, the distance a person from stratum s_1 travels on average to arrive at a stratum s_2 destination, hereinafter denoted by $D^{s_1 s_2}$, is computed as

$$D^{s_1 s_2} = \frac{\sum_{ij} d_{ij} f_{ij}^{s_1} R_j^{s_2}}{\sum_{ij} f_{ij}^{s_1} R_j^{s_2}}. \quad (11)$$

Results

Mobility, lockdown and socioeconomic classes

We first analyze the socioeconomic determinants shaping the relationship between lockdown policies and mobility reduction. To do so, we construct a time-varying mobility network associated with each stratum by merging weekly mobility patterns estimated from mobile phone users with socioeconomic information provided by census surveys as detailed in the “Methods” section. For the sake of illustration, we represent in Fig. 1 the spatial distribution of: the number of residents (a–c), the household stratum (d–f), and the aggregated mobility patterns (g–i) for each of the three cities here analyzed.

Once constructed the multiplex mobility network, we can focus on the reduction in the level of mobility of each stratum, which we can calculate by aggregating all the flows in each layer as detailed in “Methods” section. Figure 2a–c showcase the effectiveness of the lockdown measures, as the three cities achieved a significant reduction in the overall mobility of their populations. This, however, was not homogeneous across strata, being much more significant for the wealthiest strata typically gathering socioeconomic classes with highest income. To further quantify this phenomenon, we compute the reduction in the volume of flows by comparing mobility data corresponding to the first week of April and the first of February in each geographical area (patch) as explained in “Methods” section. Figure 2d–f confirm the positive correlation between the average economic stratum in each patch and the reduction in mobility achieved by its residents. This correlation is consistent across the three cities here analyzed, being the Spearman correlation coefficients between both variables $\rho_S = 0.832$, $\rho_S = 0.878$ and $\rho_S = 0.924$ for Bogotá, Medellín and Santiago de Cali respectively.

Lockdown restructures mobility networks

So far, we have tackled how socioeconomic features shape the impact of lockdown policies on human mobility by studying how the macroscopic volume of movements associated with each stratum varied as these policies were enforced. In this section, we are interested in addressing whether, beyond the overall reduction in the level of mobility across the city, these interventions modified substantially the microscopic structure of the mobility network of each stratum. To do so, we focus on two different indicators: the distance involved in each stratum’s trips and the socioeconomic composition of their destinations.

Before studying the impact of lockdown policies, let us characterize the structure of the mobility networks in the baseline scenario (February 2020) before the arrival of COVID-19 pandemic in the three cities here analyzed. For this purpose, we compute the matrix $\mathbf{D}$, where each element $D^{s_1 s_2}$ of this matrix is the average distance of the flows from the zones associated with stratum s_1 to those associated with stratum s_2 (see “Methods” for the computation of the matrix). Figure 3a–c show that the different strata are systematically segregated in the cities studied, since the distance between two strata increases as they become less similar from a socioeconomic point of view. These results, computed from the merged mobility networks, are consistent with the clear segregation observed in Fig. 1d–f showing the spatial distribution of strata across these cities with the finest spatial resolution of census blocks.

To compute the impact of lockdown policies, we represent in Fig. 3d–f the time evolution of the average distance of trips departing from areas associated with each stratum s_1 , denoted by $\langle d^{s_1} \rangle$ (see “Methods” for the derivation of this quantity). In these panels, we observe a robust trend across the three cities consisting in a drastic reduction of trip distances as soon as NPIs were implemented. Note, however, that this effect is not homogeneous across each city, being more pronounced for those trips departing from areas associated with lowest strata.

To shed light into how the heterogeneous trends observed in Fig. 3d–f affect social mixing, we investigate how lockdown policies altered the socioeconomic composition of the destinations visited by each stratum. This information is encoded in the mixing matrix $\mathbf{M}$ in which each element $M_{s_1 s_2}$ denotes the average fraction of individuals belonging to stratum s_2 in the destinations of the trips departing from stratum s_1 areas. The mathematical expression of these elements can be found in the “Methods” sections. As for the distance analysis, let us first analyze the mixing matrix $\mathbf{M}$ in the baseline scenario, when no lockdown policies were at play. Figure 4a–c confirm the strong socioeconomic impact on mixing patterns in Colombian cities. Specifically, we observe that trips connecting areas populated by similar strata are overrepresented whereas connections among distant strata are hampered. By removing the bias related to the unequal population size of the different strata and computing the re-normalized mixing matrix $\mathbf{M}_N$ (see “Methods” for details on its calculation), the former result becomes more evident, as illustrated in Fig. 4d–f.

In Fig. 4g–i we show how lockdown policies altered the elements of the mixing matrices. There, by comparing the structure of the matrices corresponding to the baseline pre-pandemic scenario and the week starting from 29th March (during lockdown), we observe that the impact of the implemented policies on the structure

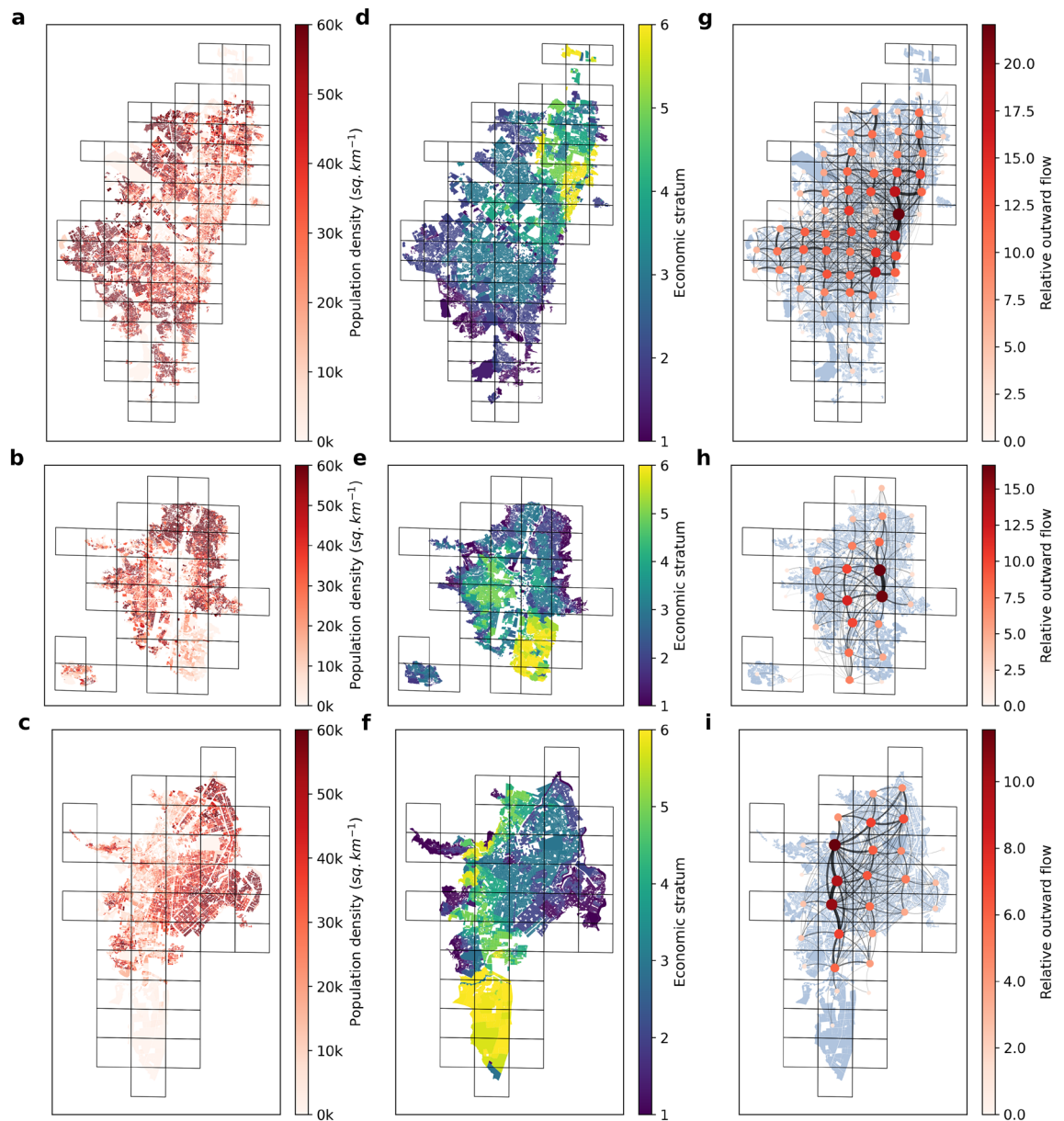


Figure 1. (a–c) Population density maps at the level of census block, corresponding to the 2018 census. (d–f) Average economic strata of the different households at the level of census block, ranging from stratum 1 typically gathering those individuals with lowest economic income to stratum 6 associated with wealthiest population. (g–i) Schematic representation of the mobility network of each city with a spatial resolution of level 12 s2cells (see “Methods” for further explanations). Both color and size of nodes are proportional to the total number of trips departing from a given area whereas the edge thickness reflects the the number of trips recorded between two locations. From top to bottom, the information shown corresponds to Bogotá, Medellín and Santiago de Cali respectively.

of mobility patterns is quite heterogeneous across strata. Once again, these heterogeneities across strata are consistent in the three cities here studied. In the first place, confinement policies provoked important modifications in the mobility patterns of lowest strata, exacerbating their endogenous nature while making it difficult for them to mix with those belonging to higher strata. Taking this result together with the average trip distance connecting different strata shown in Fig. 3, we realize that the pronounced drop in the trip distance for the lowest strata is explained by the huge reduction in long-distance trips connecting them to areas populated by the higher strata. In contrast, lockdown policies seemed to have a less relevant impact on the structure of contacts of the wealthiest strata in Colombia. However, note that, despite this overall small variation, lockdown policies appeared to increase the diversification of their mobility patterns, as they moved more evenly to places belonging to socially distant individuals.

To round off our study, we compute the time evolution of the Shannon entropy of the entries of the mixing matrices corresponding to each stratum and represent them in Fig. 5. The entropy values shown in these figures

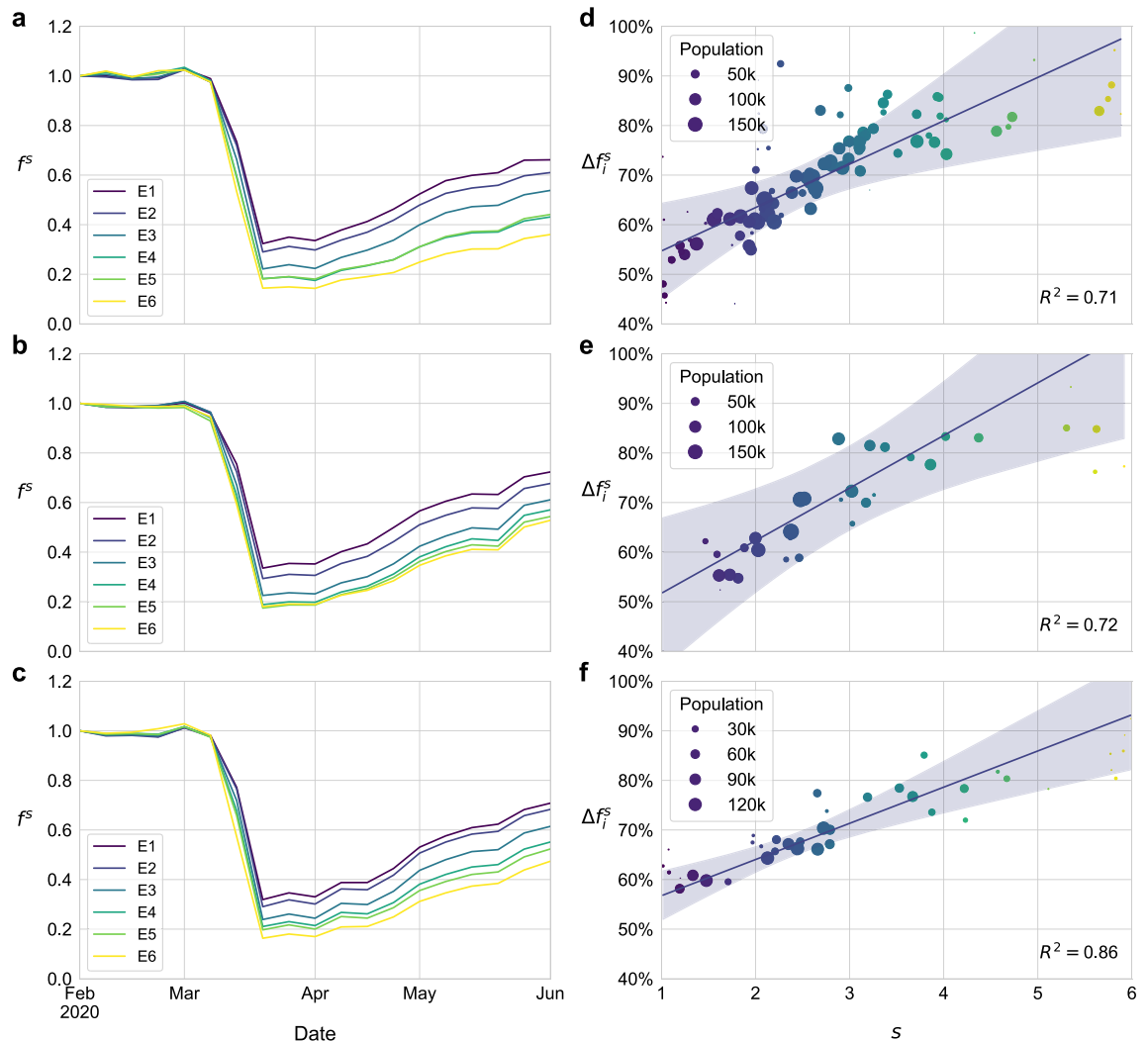


Figure 2. (a–c) Time evolution of the aggregated number of trips for each stratum (color code), re-scaled to a reference value set on 2020-02-02, corresponding to the pre-pandemic scenario. (d–f) Reduction in the trips departing from each patch on the week starting from 2020-03-29 compared to a baseline scenario (2020-02-02) according to the average economic stratum of its residents. The size of the dots denotes the number of residents in the corresponding patch whereas the color encodes the stratum information ranging from poorest areas (blue) to the richest ones (yellow). The shaded area is the prediction interval of the linear regression, and it represents the standard error of the predictions from the model, obtained with the member *var_pred_mean* from the Python library *statsmodels*. From top to bottom, the information shown corresponds to Bogotá, Medellín and Santiago de Cali respectively.

are normalized by their values as of 02-02-2020 (before the NPIs came into force). The three panels confirm the previous results: lockdown policies reduced considerably the social variability of the destinations visited by lowest strata whereas it left almost unaltered those corresponding to highest ones. However we find an exception in Medellín, where population from stratum 6 tend to isolate more as also shown by Fig. 4h.

Discussion

The extremely rapid unfolding of COVID-19 pandemic at the beginning of 2020 arose from the subsequent combination of the international exportation of infectious cases^{35,36} through the airport mobility network and the local exponential growth in the number of cases driven by community transmission of the virus. Since vaccines were not available at the time, many countries deployed NPIs to reduce the impact of the pandemic. These containment measures included mobility restrictions to curb the importation of cases from high-incidence areas and local control policies that reshaped the contact patterns of the population responsible for community transmission of the virus.

In this work, we have studied how NPIs altered the mobility patterns of different socioeconomic classes in the three main cities of Colombia. Our analysis reveals a strongly unequal impact of control policies on the different strata into which Colombian society is divided. First, we have reported a consistent positive correlation between the reduction in the number of trips registered for a given area and the average stratum of its residents, implying

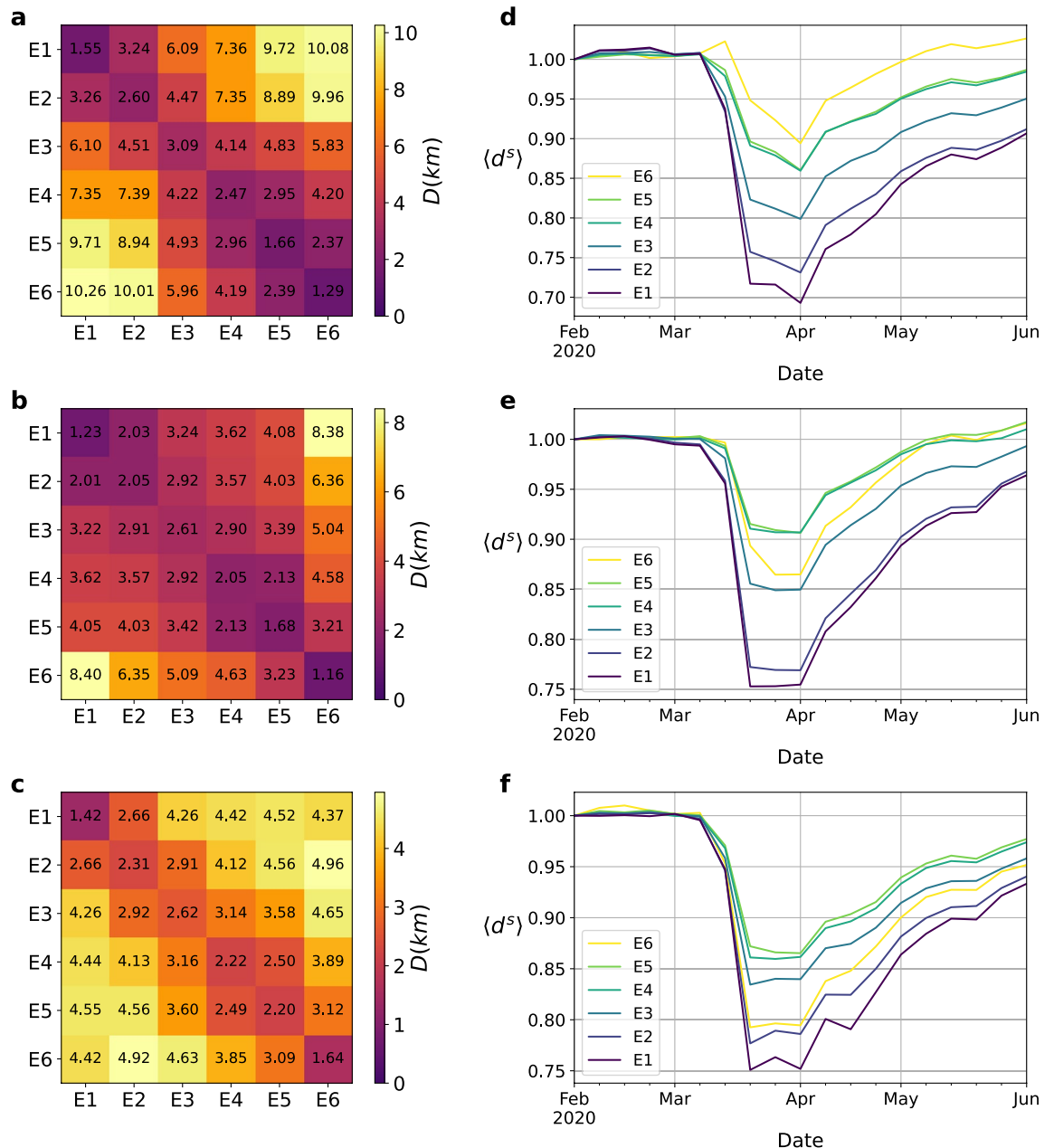


Figure 3. (a–c) Distance matrices D encoding the average distance of trips departing from areas associated with one stratum (rows) and arriving in going to stratum s_2 areas (see “Methods” for an explanation of their computation). These values are computed using mobility patterns from the week starting on 02-02-2020, which corresponds to a pre-pandemic scenario. (d–f) Time evolution of the average distance made by individuals belonging to each stratum s , $\langle d^s \rangle$ (color code). Note that the values represented have been re-scaled by those corresponding to the pre-pandemic scenario. From top to bottom, the information shown corresponds to Bogotá, Medellín and Cali respectively.

that those individuals with more economic resources, typically belonging to the highest strata, were more able to reduce their mobility and, therefore, their social activity. Secondly, we have revealed that lockdown policies, in addition to affecting the volume of movements recorded in each city, modified their architecture, reducing the distance involved in the movements and increasing the social similarity in the composition of the origin and destination of the flows recorded. However, this phenomenon was not homogeneous in all strata. The blocking policies hindered long-distance urban trips connecting lowest socioeconomic classes with wealthier areas, leading to a significant increase in the endogenous interactions with geographically proximate individuals and a reduction in the distance of their trips. In contrast, our analysis reveals that the implemented policies hardly affected the mobility patterns of the highest strata.

Our results suggest the need for accounting for socioeconomic variables when assessing the feasibility or the expected impact of community NPIs to mitigate an epidemic outbreak. In this sense, despite the clear

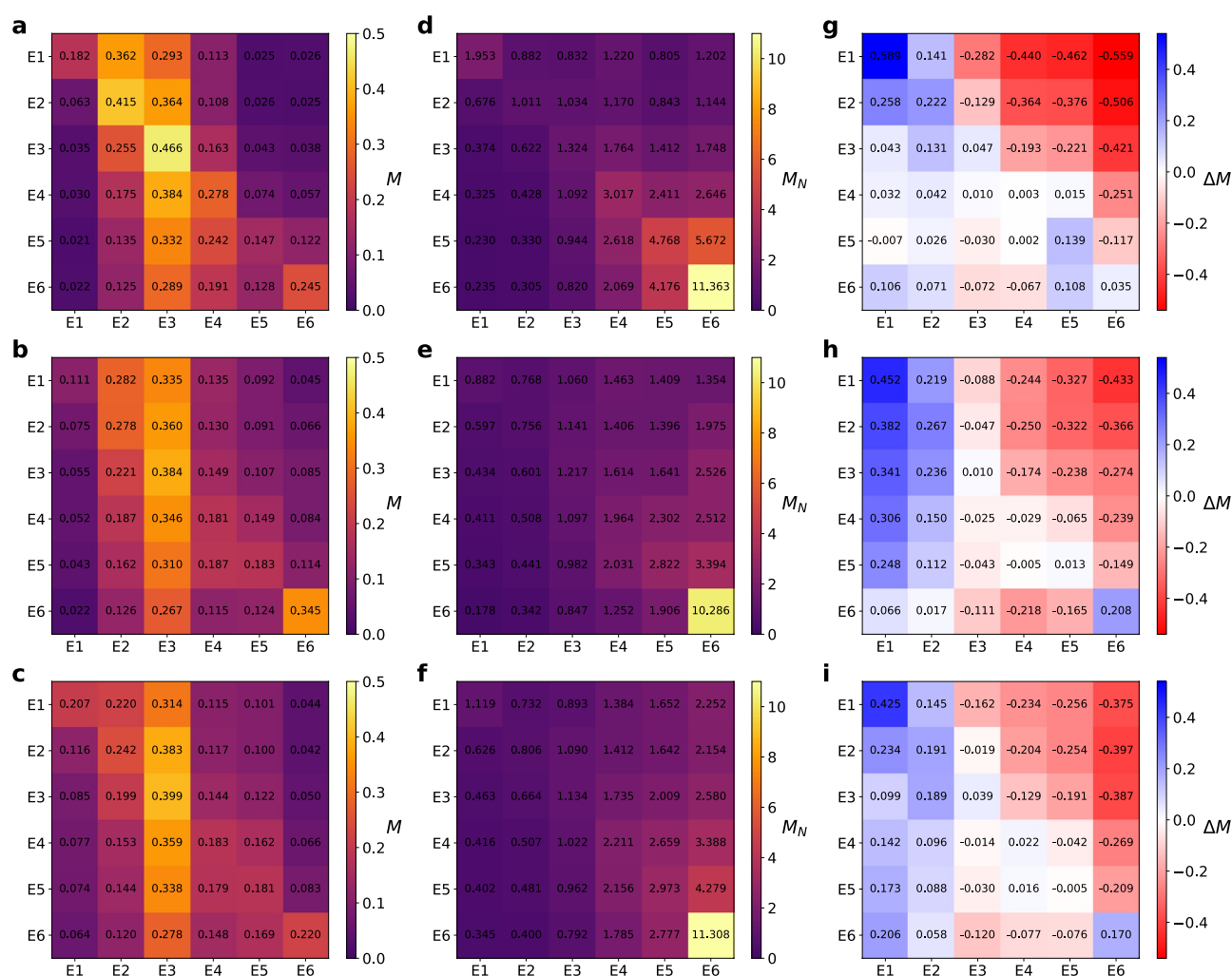


Figure 4. (a–c) Mixing matrices M encoding the fraction of trips corresponding to a certain stratum (rows) arriving in areas associated with another stratum (columns) on 2020-02-02 (pre-pandemic scenario). The sum of the elements on each row is equal to one. (d–f) Re-normalized mixing matrices M_N based on the proportion of the population belonging to each stratum, as detailed in the “Methods” section. In these matrices, entries above (below) one denote a higher (lower) tendency to interact with a given stratum than the one expected in a well-mixed population. (g–i) Relative change of these matrices ΔM comparing a lockdown scenario, corresponding to the week starting on 2020-03-29, with a pre-pandemic scenario, corresponding to the week starting on 2020-02-02. From top to bottom, the information shown corresponds to the cities of Bogotá, Medellín and Santiago de Cali respectively.

improvement of the health situation while gaining time for vaccine development, the socioeconomic dimension of NPIs entailed undesired collateral effects such as economic crisis^{37,38}, mental health issues^{39–45}, or the increase of social inequalities in different countries as a result of lockdown policies^{46–48}. Beyond social implications, the exacerbation of social inequalities might also have important implications for the evolution of epidemic outbreaks, as proven by the influence of income gradients on COVID-19 associated mortality^{49–51}. Apart from NPIs, socioeconomic determinants are also crucial to understand other important factors during COVID-19 pandemic such as the access to health services^{52–54}, the unequal vaccination willingness observed in the population^{55–57} or the heterogeneous occurrence of mental health issues across social and age groups⁵⁸. We hope that our study will shed light into the interplay between socioeconomic information and epidemic spreading and will pave the way to the design of control policies optimizing the trade-off between their health outcome and the damage of the socioeconomic fabric derived from their implementation.

Limitations

Our results should be interpreted in light of several limitations. First, the Google mobility data is limited to smartphone users who have opted in to Google’s Location History feature, which is off by default. Therefore, we assume the mobility patterns of this set of individuals to be representative of the movements of the entire population. Importantly, these limited data are only viewed through the lens of differential privacy algorithms,

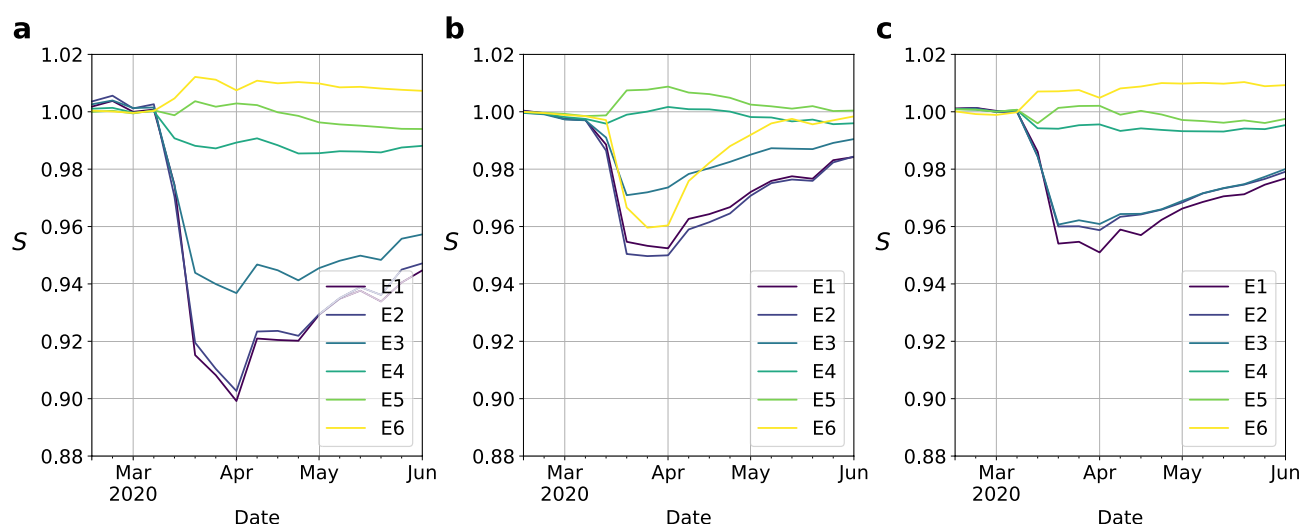


Figure 5. Time evolution of the relative entropy S for each row of the mixing matrix, which encodes the socioeconomic structure of the destinations visited by each stratum (color code), in Bogotá (a), Medellín (b) and Cali (c) respectively. Note that the values represented have been re-scaled to those corresponding to the week starting on 02-02-2020, corresponding to a pre-pandemic scenario. Let us note that lower entropy values reflect more concentrated flows towards a small number of economic strata.

specifically designed to protect user anonymity and obscure fine detail. Furthermore, comparisons across rather than within locations are only descriptive since these regions can differ in substantial ways.

Another limitation of this data set is that the origin of one movement recorded for one individual does not necessarily coincide with its residence, which might introduce some noise when computing the reduction of the mobility associated to each stratum. Nonetheless, previous studies leveraging this kind of data sets have found comparable results to those obtained when analyzing mobility patterns coming from census surveys^{18,33}.

Moreover, to construct the multiplex mobility networks, we have assumed that all socioeconomic strata are proportionally represented in the data set, thus neglecting socioeconomic biases in the use of mobile phone devices, which may vary by location. While this assumption could remain controversial, our results are consistent with those obtained by analyzing variations in public transport usage as a result of lockdown policies³².

Finally, we have restricted our analysis to cities in Colombia, where socioeconomic classes are highly segregated. This segregation is partly explained by the spatial distribution of these populations across cities, where poor populations tend to be located on informal settlements along the cities' boundaries. Frequently, these outskirts tend to be more disconnected from city centers and economical hubs due to scarce access to transport services. The generality of our findings when extending our analysis to other societies deserves further investigation and remains as future work.

Data availability

Mobility data are extracted from the Google SARS-CoV-2 Aggregated Mobility Research Dataset and are available with permission from Google LLC. Data regarding the demographic and socioeconomic information are publicly available and provided by the national statistics department in Colombia (accessed 2023-07-21, <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018>). Data about the time evolution of COVID-19 cases are extracted from official data reported by the authorities in Colombia (accessed 2023-02-14, <https://www.datos.gov.co/Salud-y-Proteccion-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data>). The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability

Codes used to generate the figures of the manuscript are available from the corresponding author on reasonable request.

Received: 27 February 2023; Accepted: 27 September 2023

Published online: 30 September 2023

References

1. Estrada, E. Covid-19 and sars-cov-2, modeling the present, looking at the future. *Phys. Rep.* **869**, 1–51 (2020).
2. Perra, N. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Phys. Rep.* **913**, 1–52 (2021).
3. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on Covid-19 in Europe. *Nature* **584**, 257–261 (2020).
4. Haug, N. *et al.* Ranking the effectiveness of worldwide covid-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).
5. Cheng, Y. *et al.* Face masks effectively limit the probability of sars-cov-2 transmission. *Science* **372**, 1439–1443 (2021).
6. Rader, B. *et al.* Mask-wearing and control of sars-cov-2 transmission in the USA: A cross-sectional study. *Lancet Digit. Health* **3**, e148–e157 (2021).

7. Reyna-Lara, A. *et al.* Virus spread versus contact tracing: Two competing contagion processes. *Phys. Rev. Res.* **3**, 013163 (2021).
8. Aleta, A. *et al.* Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
9. Kojaku, S., Hébert-Dufresne, L., Mones, E., Lehmann, S. & Ahn, Y.-Y. The effectiveness of backward contact tracing in networks. *Nat. Phys.* **17**, 652–658 (2021).
10. Arenas, A. *et al.* Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions. *Phys. Rev. X* **10**, 041055 (2020).
11. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Impact of lockdown on covid-19 epidemic in ile-de-france and possible exit strategies. *BMC Med.* **18**, 1–13 (2020).
12. Marziano, V. *et al.* Retrospective analysis of the Italian exit strategy from covid-19 lockdown. *Proc. Natl. Acad. Sci. USA* **118**, e2019617118 (2021).
13. Valgañón, P., Leria, U., Soriano-Paños, D. & Gómez-Gardeñes, J. Socioeconomic determinants of stay-at-home policies during the first covid-19 wave. *Front. Public Health* **11**, 1193100 (2023).
14. Lotero, L., Hurtado, R. G., Floría, L. M. & Gómez-Gardeñes, J. Rich do not rise early: Spatio-temporal patterns in the mobility networks of different socio-economic classes. *R. Soc. Open Sci.* **3**, 150654 (2016).
15. Sabatini, F. *The Social Spatial Segregation in the Cities of Latin America* (Inter-American Development Bank, 2006).
16. Reme, B.-A., Kotsadam, A., Bjelland, J., Sundsøy, P. R. & Lind, J. T. Quantifying social segregation in large-scale networks. *Sci. Rep.* **12**, 1–7 (2022).
17. Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large us cities. *Nat. Commun.* **12**, 1–10 (2021).
18. Barbosa, H. *et al.* Uncovering the socioeconomic facets of human mobility. *Sci. Rep.* **11**, 1–13 (2021).
19. Hilman, R. M., Iniguez, G. & Karsai, M. Socioeconomic biases in urban mixing patterns of us metropolitan areas. *EPJ Data Sci.* **11**, 32 (2022).
20. Soriano-Paños, D., Lotero, L., Arenas, A. & Gómez-Gardeñes, J. Spreading processes in multiplex metapopulations containing different mobility networks. *Phys. Rev. X* **8**, 031039 (2018).
21. Bosetti, P. *et al.* Heterogeneity in social and epidemiological factors determines the risk of measles outbreaks. *Proc. Natl. Acad. Sci. USA* **117**, 30118–30125 (2020).
22. Castro, N. R. & Moreira, G. C. Who worked from home in Brazil? Inequalities highlighted by the pandemic. *Nova Econ.* **31**, 899–927 (2022).
23. Irlacher, M. & Koch, M. Working from home, wages, and regional inequality in the light of covid-19. *Jahrbücher für Nationalökonomie Stat.* **241**, 373–404 (2021).
24. Bonacini, L., Gallo, G. & Scicchitano, S. Working from home and income inequality: Risks of a ‘new normal’ with covid-19. *J. Popul. Econ.* **34**, 303–360 (2021).
25. Castells-Quintana, D., Herrera-Idárraga, P., Quintero, L. & Sinisterra, G. *Unequal Response to Mobility Restrictions: Evidence from Covid-19 Lockdown in the City of Bogotá. UNDP COVID-19 Policy Documents Series* (2021).
26. Aromí, J. D., Bonel, M. P., Cristia, J., Llada, M. & Palomino, L. *Socioeconomic Status and Mobility During the Covid-19 Pandemic: An Analysis of Eight Large Latin American Cities. Tech. Rep., institutionIDB Working Paper Series* (2021).
27. Secretaría distrital de planeación. encuesta multipropósito (em) (2017).
28. Mejía-Arbelaiz, C. *et al.* Social inclusion and physical activity in ciclovía recreativa programs in Latin America. *Int. J. Environ. Res. Public Health* **18**, 180206 (2021).
29. Vernez, G. Residential movements of low-income families: The case of bogotá, Colombia. *Land Econ.* **50**, 421–428 (1974).
30. Castro-Aristizabal, G., Giménez-Esteban, G., Arango-Londoño, D., Moreno-Cediel, E. & Castillo-Cacedo, M. Spatial variation in educational quality in Colombia based on the phenomena of agglomeration and academic segregation. *Int. J. Environ. Res. Public Health* **12**, 180800 (2022).
31. Moya-Gómez, B., Stepniak, M., García-Palomares, J., Frías-Martínez, E. & Gutiérrez, J. *Exploring Night and Day Socio-spatial Segregation Based on Mobile Phone Data: The Case Of Medellín (colombia)* (Computers, Environment and Urban Systems, 2021).
32. Dueñas, M., Campi, M. & Olmos, L. E. Changes in mobility and socioeconomic conditions during the covid-19 outbreak. *Hum. Soc. Sci. Commun.* **8**, 1–10 (2021).
33. Bassolas, A. *et al.* Hierarchical organization of urban mobility and its connection with city livability. *Nat. Commun.* **10**, 1–10 (2019).
34. Wilson, R. J. *et al.* Differentially private sql with bounded user contribution. arXiv preprint [arXiv:1909.01917](https://arxiv.org/abs/1909.01917) (2019).
35. Gilbert, M. *et al.* Preparedness and vulnerability of African countries against importations of covid-19: A modelling study. *The Lancet* **395**, 871–877 (2020).
36. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* **368**, 395–400 (2020).
37. Maital, S. & Barzani, E. The global economic impact of covid-19: A summary of research. *Samuel Neaman Inst. Natl. Policy Res.* **2020**, 1–12 (2020).
38. Albanesi, S. & Kim, J. Effects of the covid-19 recession on the us labor market: Occupation, family, and gender. *J. Econ. Perspect.* **35**, 3–24 (2021).
39. Le, H. T. *et al.* Anxiety and depression among people under the nationwide partial lockdown in Vietnam. *Front. Public Health* **8**, 589359 (2020).
40. Tran, B. X. *et al.* Impact of covid-19 on economic well-being and quality of life of the Vietnamese during the national social distancing. *Front. Psychol.* **11**, 565153 (2020).
41. Wang, C. *et al.* The association between physical and mental health and face mask use during the covid-19 pandemic: A comparison of two countries with different views and practices. *Front. psychiatry* **11**, 569981 (2020).
42. Lee, Y. *et al.* Government response moderates the mental health impact of covid-19: A systematic review and meta-analysis of depression outcomes across countries. *J. Affect. Disord.* **290**, 364–377 (2021).
43. Lau, S. S. *et al.* Measurement of burnout during the prolonged pandemic in the Chinese zero-covid context: Covid-19 burnout views scale. *Front. Public Health* **10**, 1039450 (2022).
44. McIntyre, R. S. *et al.* Suicide reduction in Canada during the covid-19 pandemic: Lessons informing national prevention strategies for suicide reduction. *J. R. Soc. Med.* **114**, 473–479 (2021).
45. Tan, W. *et al.* Is returning to work during the covid-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of chinese workforce. *Brain Behav. Immunity* **87**, 84–92 (2020).
46. Van Dorn, A., Cooney, R. E. & Sabin, M. L. Covid-19 exacerbating inequalities in the US. *The Lancet* **395**, 1243–1244 (2020).
47. Wachtler, B. *et al.* Socioeconomic inequalities and covid-19-a review of the current international literature. *J. Health Monit.* **5**, 3 (2020).
48. Bajos, N. *et al.* When lockdown policies amplify social inequalities in covid-19 infections: Evidence from a cross-sectional population-based survey in france. *BMC Public Health* **21**, 1–10 (2021).
49. Arceo-Gomez, E. O. *et al.* The income gradient in covid-19 mortality and hospitalisation: An observational study with social security administrative records in mexico. *The Lancet Reg. Health-Am.* **6**, 100115 (2022).
50. Drefahl, S. *et al.* A population-based cohort study of socio-demographic risk factors for covid-19 deaths in Sweden. *Nat. Commun.* **11**, 1–7 (2020).

51. Decoster, A., Minten, T. & Spinnewijn, J. The income gradient in mortality during the covid-19 crisis: Evidence from Belgium. *J. Econ. Inequal.* **19**, 551–570 (2021).
52. Tran, B. X. *et al.* Capacity of local authority and community on epidemic response in Vietnam: Implication for covid-19 preparedness. *Saf. Sci.* **130**, 104867 (2020).
53. Tran, B. X. *et al.* The operational readiness capacities of the grassroots health system in responses to epidemics: Implications for covid-19 control in Vietnam. *J. Glob. Health* **10**, 011006 (2020).
54. Tran, B. X. *et al.* Reaching further by village health collaborators: The informal health taskforce of Vietnam for covid-19 responses. *J. Glob. Health* **10**, 010354 (2020).
55. Sitarz, R. *et al.* To vaccinate or not to vaccinate—reasons of willingness and reluctance of students against sars-cov-2 vaccination—an international experience. *Int. J. Environ. Res. Public Health* **19**, 14012 (2022).
56. Chew, N. W. *et al.* An Asia-Pacific study on healthcare workers' perceptions of, and willingness to receive, the covid-19 vaccination. *Int. J. Infect. Dis.* **106**, 52–60 (2021).
57. Hao, F. *et al.* Attitudes toward covid-19 vaccination and willingness to pay: Comparison of people with and without mental disorders in china. *BJPsych Open* **7**, e146 (2021).
58. Wang, C. *et al.* The impact of the covid-19 pandemic on physical and mental health in the two largest economies in the world: A comparison between the united states and china. *J. Behav. Med.* **44**, 741–759 (2021).

Acknowledgements

J.G.-G. and P.V. acknowledge funding from grant PID2020-113582GB-I00 funded by MCIN/AEI/10.13039/501100011033; and by the Departamento de Industria e Innovación del Gobierno de Aragón y Fondo Social Europeo through Grant No. E36-20R (FENOL group). D.S.-P acknowledges financial support from Instituto Gulbenkian de Ciência through the PONTE program. GG acknowledges support from the US Army Research Office through grant W911NF-22-1-0182.

Author contributions

D.S.P., G.G. and J.G.G. conceived the study. P.V. and A.U. analyzed the data. All the authors discussed the results and wrote the manuscript.

Competing interests.

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.S.-P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Capítulo 4

Procesos de contagio y difusión con movilidad recurrente de agentes distinguibles

Contagion-diffusion processes with recurrent mobility patterns of distinguishable agents

P. Valgañón,¹ D. Soriano-Paños,^{2,3} A. Arenas,⁴ and J. Gómez-Gardeñes^{1,3,5,a)}

¹⁾Departament of Condensed Matter Physics, University of Zaragoza, 50009 Zaragoza (Spain).

²⁾Instituto Gulbenkian de Ciência (IGC), 2780-156 Oeiras (Portugal)

³⁾GOTHAM lab, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza (Spain).

⁴⁾Departament de Matemàtiques i Enginyeria Informàtica, Universitat Rovira i Virgili, 43007 Tarragona (Spain).

⁵⁾Center for Computational Social Science, University of Kobe, 657-8501 Kobe (Japan).

(Dated: March 15, 2022)

The analysis of contagion-diffusion processes in metapopulations is a powerful theoretical tool to study how mobility influences the spread of communicable diseases. Nevertheless, many metapopulation approaches use indistinguishable agents to alleviate analytical difficulties. Here, we address the impact that recurrent mobility patterns, and the spatial distribution of distinguishable agents, have on the unfolding of epidemics in large urban areas. We incorporate the distinguishable nature of agents regarding both, their residence, and their usual destination. The proposed model allows both a fast computation of the spatio-temporal pattern of the epidemic trajectory and the analytical calculation of the epidemic threshold. This threshold is found as the spectral radius of a mixing matrix encapsulating the residential distribution, and the specific commuting patterns of agents. We prove that the simplification of indistinguishable individuals overestimates the value of the epidemic threshold.

Unraveling the influence that different aspects of human behavior have on how communicable diseases spread through populations is one of the most intriguing challenges in computational and theoretical epidemiology. Although it is now possible to include multiple types of human behavioral data for agent-based simulations, incorporating this entire arsenal of information into mathematical models to derive new analytical tools is a major challenge in epidemiology. In this article, we aim to go one step further on the road of increasing the realism of metapopulation-based epidemic models. In particular, we propose a theoretical framework that allows the inclusion of data on both the spatial distribution of populations and the distinction of agents according to their origin and destination. With this information at hand, we can derive the value of the epidemic threshold and its roots on the social mixing patterns that characterize the population under study providing, as a byproduct, a powerful tool to assess control strategies aimed at increasing its value under scenarios of epidemiological risk.

I. INTRODUCTION

About 100 years ago, two physicians, Ronald Ross and Anderson G. McKendrick, and a chemist, William O. Kermack laid the foundations of epidemic modeling. McKendrick and Kermack formulated the celebrated SIR (Susceptible-Infectious-Recognized) compartmental model in 1927^{1,2}, a framework that remains nowadays as the cornerstone of most theoretical works in epidemiology. Around the same time, in 1922, Lewis Fry Richardson³ proposed a set of differential

equations in an attempt to mathematically describe the evolution of the atmosphere and to go beyond qualitative weather forecasting to use quantitative and objective forms of prediction.

Meteorological models quickly confirmed their practical usefulness as soon as they could be implemented in the first computing machines, seeing how their reliability improved as they were refined and both the quantity and quality of meteorological data increased. In contrast, despite theoretical advances in their refinement^{4,5}, the usefulness of epidemiological models remained limited for many decades, finding their fundamental utility in the qualitative understanding of the phenomena observed in different epidemic waves. In this sense, the collection of epidemiological data for the validation and improvement of compartmental models was much more elusive than in the case of meteorology. In particular, apart from the biological features of the spreading pathogen, the main bottleneck for the development of reliable epidemic models was to accurately describe the human behavior underlying the observed infection patterns.

This barrier to the development of epidemiological models with predictive capacity was broken down with the advent of the 21st century, the internet era and the digitization of our daily lives. The new digital era represents a paradigm shift for the study of human behavior on a large scale, allowing us, among other things, to access, and describe the skeleton of interactions through which infectious diseases are transmitted. Thus, in the last two decades, a great deal of effort has been invested in incorporating into epidemic models aspects such as the complexity of contact networks, patterns of human mobility at different geographic scales, and the time scales associated with human interactions⁶. Equipped with this information, theoretical, and computational tools can be used to analyze epidemiological problems, and develop forecasting frameworks that integrate both advanced epidemiological

^{a)}Electronic mail: gardenes@unizar.es

models and massive amounts of real demographic, mobility, and socioeconomic data at multiple scales⁷⁻⁹.

Most of these approaches rely on agent-based models that allow recreating synthetic populations mimicking the relevant social attributes that shape the unfolding of epidemic outbreaks. Despite their relevance to forecast the evolution of real epidemics¹⁰ or to design non-pharmaceutical interventions with minimal socioeconomic impact¹¹, agent-based simulations do not offer the possibility of obtaining transparent and analytical information about the importance of human behavior in the transmission of infectious diseases. To fill the gap between agent-based mechanistic simulations and theoretical frameworks, epidemic modeling relies on reaction-diffusion dynamics in metapopulations^{12,13}. This framework incorporates coarse-grained information of several features intervening in disease spreading such as the mobility patterns between different areas, the demographic distribution of a population and information about social mixing.

In the last decade, the study of metapopulation dynamics has faced the challenge of approaching the realism of mechanistic simulations¹⁴ by incorporating more and more aspects of human behavior and mobility¹⁵. From the first works, including the complexity of human mobility networks¹⁶⁻¹⁹ the focus has been put on the recurrent nature of human mobility, being it of special importance in urban and regional scales²⁰⁻²³. Recently²⁴, we introduced a Markovian framework, the Movement-Interaction-Return (MIR) model, that allowed the study of real populations incorporating the demographic distribution and the network of commutes. This approach revealed that these two aspects are essential to assess the advisability of contention measures based on the restriction of mobility. This approach has been further generalized to include networks with multiple types of mobility^{25,26}, the study of vector-borne diseases^{27,28}, different permanence times on the destination²⁹, the heterogeneous of different contact patterns³⁰. Importantly, this Markovian framework has been used, after accounting for the particularities of SARS-CoV-2 transmission, to evaluate the evolution, and health systems impact, of COVID-19 in different countries³¹⁻³³. Likewise, the MIR model has been used to optimize resource allocation to control epidemics³⁴ or to evaluate the role that individual awareness plays in hampering the spread of diseases³⁵.

In this work, we go one step further in the formulation of the MIR epidemic model to better capture the recurrent mobility patterns of most human movements. To this aim, we get rid of one of the main hypotheses behind the former approach: indistinguishable agents, residents of a patch, according to their possible destinations. Including distinguishable agents allows us to analyze particular human commuting flows between different locations, and to identify those that are critical for the dissemination of infectious pathogens. This paves the path to inform about surgical interventions on the mobility patterns of a population to increase its resilience against the spread of a pathogen, in contrast to crude lockdowns spatially isolating one area.

II. RESULTS

A. Basic metapopulation framework

Let us first introduce the basic MIR metapopulation framework that allows capturing the specific individual commuting patterns in generic populations in which agents display recurrent mobility patterns. In the following we will focus on the simple but paradigmatic Susceptible-Infected-Susceptible (SIS) compartmental model as the process underlying microscopic contagions. However, the formalism can be straightforward generalized to more sophisticated models, as is the case for the MIR metapopulation model with indistinguishable agents^{25,27,29,31}.

The SIS model describes a process in which a person who is in a susceptible state, upon contact with an infected person, becomes infected with probability λ . At the same time, an infected person recovers with probability μ and, at variance with the SIR model, becomes susceptible again. When writing the equations for a single population of n interacting agents, one typically considers the fraction of infected individuals in the population $\rho(t)$. The simplest approach to model the evolution of this order parameter is to consider a mean-field scenario where all individuals are statistically equivalent, implying that all of them make $\langle k \rangle$ contacts and have a homogeneous probability of interacting with each other. An usual assumption when modelling the discrete-time evolution of the mean-field equations of the SIS model is to consider the number of interactions negligible with respect to the population size, i.e. $\langle k \rangle \ll n$. This assumption neglects the possibility of interacting twice with the same agent and allows us to estimate the probability that a susceptible individual chooses i infected individuals as:

$$p(i) = \binom{\langle k \rangle}{i} \rho^i (1 - \rho)^{\langle k \rangle - i}. \quad (1)$$

Since we have a probability $(1 - \lambda)^i$ of not getting infected when contacting i infected individuals, the total probability of getting infected reads:

$$P(t) = 1 - \sum_{i=0}^{\langle k \rangle} \binom{\langle k \rangle}{i} \rho^i (1 - \rho)^{\langle k \rangle - i} (1 - \lambda)^i, \quad (2)$$

which can be finally simplified to:

$$P(t) = 1 - (1 - \lambda \cdot \rho(t))^{\langle k \rangle}. \quad (3)$$

Note that this expression has been obtained assuming that all nodes have connectivity equal to the mean, $\langle k \rangle$. A more elaborate way to obtain $P(t)$ would be to consider each infection probability associated with those nodes with degree k and perform a weighted average with the degree distribution $P(k)$ of the population. The expression of $P(t)$ allows us to write the mean-field evolution for a single population of n agents. Considering a time-discrete version, the fraction of infected individuals at time $t + 1$ is given by:

$$\rho(t + 1) = (1 - \mu)\rho(t) + P(t) \cdot (1 - \rho(t)). \quad (4)$$

Yet simple, the SIS model allows getting analytical insights into a pivotal indicator in mathematical epidemiology: the epidemic threshold λ_c . The epidemic threshold is the minimum value of λ yielding an epidemic scenario in which the epidemic is not extinguished but keeps circulating from one individual to another. For the case of the mean-field SIS model the epidemic threshold can be easily derived, $\lambda_c = \frac{\mu}{\langle k \rangle}$, and implies that when $\lambda > \lambda_c$ the system reaches a steady-state in which the stationary value of $\rho(t)$ is constant and non-zero since new-infections are balanced by the recovery of infected agents.

In epidemiological terms, the epidemic threshold λ_c is closely related to the so-called basic reproduction number R_0 , defined as the number of secondary infections a single infectious individual would make in a population of fully susceptible agents. The basic reproduction number in the mean-field SIS model is $R_0 = \lambda \langle k \rangle / \mu$. Thus, when the infectivity per contact $\lambda = \lambda_c = \frac{\mu}{\langle k \rangle}$ the system has a reproduction number $R_0 = 1$, meaning that, on average, an infected agent makes 1 new infection during its infectious period μ^{-1} . Obviously when $\lambda > \lambda_c$ the corresponding reproduction number is $R_0 > 1$ and corresponds to an epidemic regime.

The mean-field SIS model contains many simplifications and can be improved in many ways to become a more realistic framework. One of its main assumptions is that the population under study is completely isolated. However, as many recent real epidemics reveal, the main aspect behind the explosion of localized outbreaks into epidemic (or even pandemic) scenarios is the high mobility of individuals between different populations^{36,37}. To incorporate this important feature of real epidemics into any compartmental model as the SIS one draws on the metapopulation formalism. A basic metapopulation model describes the contacts between individuals at two different scales by dividing the total population into subpopulations of different sizes. In each subpopulation, individuals interact and have contact with each other, facilitating the contagion of the pathogen. In turn, agents are allowed to move and visit different subpopulations, thus favoring the transmission of the pathogen to disease-free regions.

When defining the metapopulation framework one has to set the microscopic contagion dynamics happening within the patches (here the SIS model) and the type of diffusion that better describes the mobility of agents. To capture the typical mobility patterns at urban or regional scales, one has to take into account the recurrent nature of human mobility at these scales¹⁵. To such aim, in²⁴ we introduced the MIR model in which each individual has an associated subpopulation (the residence) and comprises a three-stage process in every time step. The first stage comprises the diffusion step in which each individual decides, with probability p_d , to travel to a destination or to stay (with probability $1 - p_d$) in its residential patch. Second, the reaction process in which contacts occur between individuals being in the same subpopulation at time t . Finally, the return to the original subpopulations of those agents that decided move in the first stage of the MIR sequence. In this final stage, we also take into account, as introduced by Granell and Mucha³⁸, that agents make the second round of interactions at the household level. This way interactions can be split

into those made during the day (D) and those made at night (N) with their corresponding housemates.

The recurrent mobility of agents is typically provided by Origin-Destination (OD) matrices. Thus for a population divided into N subpopulations one has an $N \times N$ OD matrix whose (i, j) -entry contains the number n_{ij} of agents with residence in i that typically perform daily commutes to patch j . This matrix can be viewed as a directed (as commuting patterns between two patches are not symmetric) and weighted network connecting the collection of N patches.

Once equipped with the OD matrix characterizing the mobility flows of a metapopulation and the knowledge about the census of each subpopulation ($\{n_i\}$), one can tackle the formulation of the dynamical equations that rule the evolution of a metapopulation. To this aim, in²⁴ the authors assign to each node i a probability that an individual is infected $\rho_i(t)$, which means that we can define the state of the metapopulation via the vector $\vec{\rho}^T = (\rho_1(t), \dots, \rho_N(t))$ that can be used to compute the local prevalences at each patch i as $n_i \cdot \rho_i(t)$. The time evolution of the variables $\rho_i(t)$ defining the epidemic state of each patch can be written as the following time-discrete Markovian chain:

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

where $\Pi_i(t)$ is the probability that an individual with residence in patch i becomes infected at time t :

$$\Pi_i(t) = (1-p_d) [P_i^D(t) + (1-P_i^D(t))P_i^N(t)] + p_d \left\{ \sum_{j=1}^N R_{ij} [P_j^D(t) + (1-P_j^D(t))P_i^N(t)] \right\}. \quad (6)$$

The former expression contains in its *r.h.s.* two terms accounting for either the infection at its residence i and the infection at a patch j that, in general, is different from i . These two terms are weighted by $(1-p_d)$ and p_d respectively since, as introduced above, p_d is the probability that agents move in the metapopulation, *i.e.* a control parameter that allows us to tune the level of spatial confinement in the population. Besides, the second infection probability in Eq. (6) makes use of matrix $\mathbf{R}$ in which each entry R_{ij} is defined as the probability that a moving agent with residence in i chooses patch j as the destination. The matrix $\mathbf{R}$ is constructed directly from the OD matrix, $R_{ij} = n_{ij} / \sum_l n_{il}$, and satisfies the normalization condition of a row stochastic matrix: $\sum_{j=1}^N R_{ij} = 1 \forall i$.

Finally, both terms in the *r.h.s.* of Eq. (6) contain two sets of probabilities, $\{P_i^D(t)\}$ and $\{P_i^N(t)\}$, that account for the probabilities of getting infected being placed at a patch i during the day (D) and of contracting the disease at the household (N) placed in patch i respectively. In both cases, these probabilities adopt a similar form to Eq. (3). Namely:

$$P_i^D(t) = 1 - \left(1 - \lambda \frac{I_i^{eff}(t)}{n_i^{eff}} \right)^{z^D f_i}, \quad (7)$$

$$P_i^N(t) = 1 - (1 - \lambda \rho_i(t))^{z^N \sigma_i}. \quad (8)$$

Eq. (7) incorporates the overall population n_i^{eff} and the effective number of infected individuals I_i^{eff} located at patch i af-

ter the Movement stage. Under the assumptions of the model, both quantities read:

$$n_i^{eff} = \sum_{j=1}^N [(1-p_d)\delta_{ij} + p_d R_{ji}] n_j, \quad (9)$$

$$I_i^{eff} = \sum_{j=1}^N [(1-p_d)\delta_{ij} + p_d R_{ji}] n_j \rho_j(t). \quad (10)$$

In contrast, Eq. (8) only considers the fraction of infected individuals residing in patch i , as it governs the probability of contagion with the members from the household. Note that, in both expressions, we have denoted the number of contacts made in patch i during the D and N cycles by $z^D f_i$ and $z^N \sigma_i$ respectively. On one hand, the number of contacts during D is proportional to the patch population density:

$$f_i = \frac{n_i^{eff}}{a_i}, \quad (11)$$

where n_i^{eff} is the population that happens to be inside that patch at that time, and a_i is the area of that same patch. Likewise, z_D is a scaling factor that ensures that the average number of contacts during D across the entire metapopulation remains equal to $\langle k^D \rangle$:

$$z^D = \frac{\sum_{i=1}^N n_i^{eff} \langle k^D \rangle}{\sum_{i=1}^N n_i^{eff} f_i}. \quad (12)$$

On the other hand, for the N cycle, the average number of contacts is proportional to the average housemates at given patch i , σ_i and z_N ensures that the average number of contacts across the entire metapopulation is $\langle k^N \rangle$:

$$z^N = \frac{\sum_{i=1}^N n_i \langle k^N \rangle}{\sum_{i=1}^N n_i \sigma_i}. \quad (13)$$

As shown in²⁴, the former formulation nicely agrees with the results obtained in mechanistic simulations of general metapopulations, such as cities or regions, for which the local census and the daily commuting trips between neighborhoods or municipalities are known. In addition, the mathematical formulation allows deriving an analytical expression for the epidemic threshold:

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})}, \quad (14)$$

where $\Lambda_{\max}(\mathbf{M})$ is the maximum eigenvalue of the *mixing matrix* $\mathbf{M}$ defined as^{24,33}:

$$\begin{aligned} M_{ij} = & \left[\left((1-p_d)^2 \frac{z^D f_i}{n_i^{eff}} + \frac{z^N \sigma_i}{n_i} \right) \delta_{ij} + \right. \\ & p_d (1-p_d) \left(R_{ji} \frac{z^D f_i}{n_i^{eff}} + R_{ij} \frac{z^D f_j}{n_j^{eff}} \right) + \\ & \left. p_d^2 \sum_{l=1}^N R_{il} R_{jl} \frac{z^D f_l}{n_l^{eff}} \right] n_j. \end{aligned} \quad (15)$$

Each term, M_{ij} , of $\mathbf{M}$ contains the three elementary processes by which agents from patches i and j can interact. The main novelty of this formalism shows up when analyzing the dependence of the epidemic threshold with the degree of mobility p_d since, counter intuitively, it is shown that mobility can help to decrease the epidemic prevalence and hence increase the epidemic threshold. This result is the product of incorporating both the heterogeneous demographic distribution of real cities or regions and the commuting nature of human mobility.

B. Incorporating the distinguishable nature of individuals

The original MIR model and its subsequent refinements assume that all individuals belonging to the same subpopulation i are equivalent, *i.e.* they explore all the possible destinations that are connected to this patch i according to the data contained in the OD matrix and captured in the right stochastic matrix $\mathbf{R}$. However, this assumption neglects that each agent has its patterns of movement.

To overcome this limitation, here we consider that the agents with residence in a patch i are distinguishable according to their preferred destination. As a result, we cannot assume that there will be the same proportion $\rho_i(t)$ of infected people living in i among the subsets of individuals traveling to different destinations. Thus, we have to further divide the set of individuals having residence in i and consider a new set of variables $\{\rho_{ij}(t)\}$ that account for the probability that an individual with residence in patch i whose usual destination is node j is infected at time t . This way, the number of variables raises from N (indistinguishable case) to $L \leq N^2$ (indistinguishable case), being L the total number of non-zero entries in the OD matrix. In the most general scenario, the connectivity through commuting patterns between the N patches is characterized by an $N^2 \times N^2$ matrix $\mathbf{N}$, whose (i, j) -entry, n_{ij} , accounts for the number of commuters with residence in i whose usual destination is patch j . Let us note that for each patch i there is a subset of individuals whose usual commuting destination is node i , *i.e.* in general $n_{ii} \neq 0$. Naturally, these variables must satisfy that $\sum_{j=1}^N n_{ij} = n_i \forall i$. This distinguishable framework is illustrated in Fig. 1, where we show a schematic simple metapopulation of 4 patches in which the agents are distinguished (colored) according to their usual commuting destination.

To formulate the Markovian equations we consider again that at each time step susceptible agents can be infected during the two stages of the MIR sequence. First, during the day at Interaction stage, when agents interact either at their destination with probability p_d or at their residential patch with probability $(1-p_d)$, and second, at the household level at Return stage. The infection probability in the first case, when an agent is placed at patch i :

$$P_i^D = 1 - \left(1 - \lambda \frac{I_i^{eff}(t)}{n_i^{eff}} \right)^{z^D f_i}, \quad (16)$$

where n_i^{eff} is, as in Eq. (9), the effective number of individuals

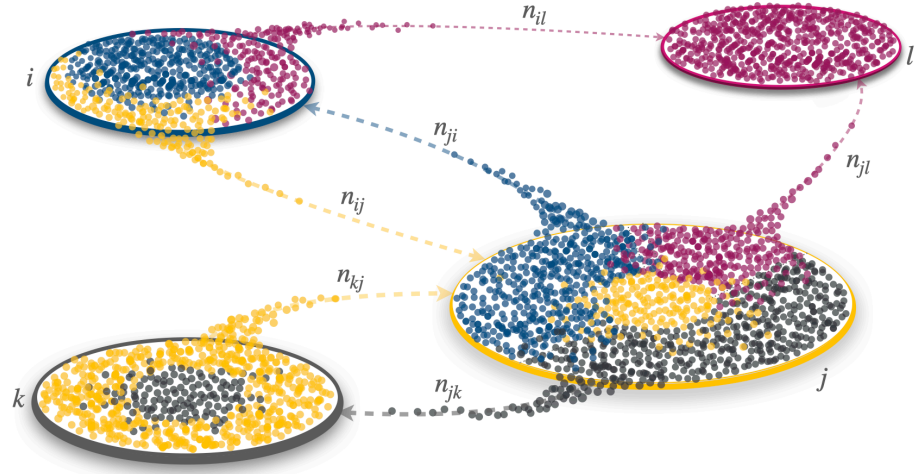


Figure 1. Schematic representation of a toy metapopulation network with distinguishable agents. The population is divided into 4 interconnected subpopulations denoted as i , j , k and l . In each subpopulation the agents are divided (colored) according to their destinations. This way the number of agents that travel from i to j correspond to a subset of size n_{ij} . Note that each patch also contains a subset of agents whose destination is the same as the residence and thus are colored according to the color of the corresponding patch.

that are placed in a patch i that, for the distinguishable case, reads:

$$n_i^{eff} = p_d \sum_{j=1}^N n_{ji} + (1 - p_d) \sum_{j=1}^N n_{ij}. \quad (17)$$

The r.h.s. of the former expression contains the number of people that decide to travel to i at the M stage from any patch j while the second contains accounts for all the commuters departing from patch i that decide not to travel. Following the same rationale is easy to write the effective number of infected individuals that visit a patch i once the movement stage has taken place:

$$I_i^{eff}(t) = p_d \sum_{j=1}^N n_{ji} \rho_{ji}(t) + (1 - p_d) \sum_{j=1}^N n_{ij} \rho_{ij}(t). \quad (18)$$

With the former two expressions, I_i^{eff} and n_i^{eff} , the probability that an agent at patch i is infectious, I_i^{eff}/n_i^{eff} can be constructed and used in Eq. (16). In close analogy with Eq. (7) for the indistinguishable case we consider that the number of contacts at patch i is proportional to the effective population density:

$$f_i = \frac{n_i^{eff}}{a_i}, \quad (19)$$

scaled by:

$$z^D = \frac{\sum_{i=1}^N n_i^{eff} \langle k^D \rangle}{\sum_{i=1}^N n_i^{eff} f_i}, \quad (20)$$

to ensure that the average number of contacts in the entire metapopulation at the Interaction stage (D period) is equal to $\langle k^D \rangle$.

For the probability of infection at the Return stage (N period) we have:

$$P_i^N = 1 - \left(1 - \lambda \frac{\sum_{j=1}^N n_{ij} \rho_{ij}(t)}{\sum_{j=1}^N n_{ij}} \right)^{z^N \sigma_i}, \quad (21)$$

where the fraction in the r.h.s has as numerator the expected number of infected individuals with residence at patch i while the denominator is the population of patch i . Thus, this fraction encodes the probability that a resident at patch i is infectious. As in Eq. (8) we consider that in the Return stage (period N) interactions are restricted to the household level so that they are given by $z_N \sigma_i$, where σ_i is the average housemate number of households at patch i and z^N reads:

$$z^N = \frac{\sum_{i=1}^N n_i \langle k^N \rangle}{\sum_{i=1}^N n_i \sigma_i}, \quad (22)$$

so that the average household contacts across the whole metapopulation during period N is $\langle k^N \rangle$.

Equipped with the infection probabilities during periods D and N, i.e. Eqs. (16) and (21), we can construct the infection probability for a resident of patch i having patch j as her usual commuting destination:

$$\Pi_{ij} = (1 - p_d) [P_i^D + (1 - P_i^D) P_i^N] + p_d [P_j^D + (1 - P_j^D) P_i^N]. \quad (23)$$

Name	Population	Patches	Links	Density ($sq.miles^{-1}$)
New York	12 423 494	498	148 001	5 084
Boston	4 146 213	232	42 064	1 704
Austin	1 775 659	98	7 281	306
Miami	5 590 269	186	31 790	1 912
Detroit	4 475 286	232	40 220	964
Seattle	3 502 087	176	22 375	565

Table I. Main characteristics of the 6 US cbsa studied as metapopulations. Namely, the total population, the number of patches and links that compose each metapopulation, and the average population density of the patches.

The former probability allows us to write the dynamical evolution for the fraction of infected agents with residence in i and destination j :

$$\rho_{ij}(t+1) = (1-\mu)\rho_{ij}(t) + (1-\rho_{ij}(t))\Pi_{ij}, \quad (24)$$

where the first term accounts for the fraction of infectious commuters between i and j at time t that do not recover whereas the second term adds the new infections of susceptible (i, j) -commuters that take place at time t both at the residence i and the destination j . The former expression takes part of a set of L equations that can be solved by iterating a given initial condition $\{\rho_{ij}(0)\}$.

C. Model Validation

To validate the set of Markovian equations given by Eq. (24) we now construct real metapopulations incorporating the demographic distribution and the commuting flows of the population of two different core-based statistical areas (cbsa) in the United States (see Table I for details); namely New York-Newark-Jersey City and Boston-Cambridge-Newton. For brevity, we will refer to them in what follows as New York and Boston respectively. For each cbsa, the patches represent the different zip codes, whereas the corresponding OD matrices have been obtained from surveys capturing the population moving daily from one area to another. Data about the commuting flows³⁹, the distribution of the population across patches⁴⁰ and the area of each patch⁴¹ are publicly available. A summary of the main attributes for every metapopulation studied throughout the paper is available in Table I.

Once the census and mobility data have been translated into a metapopulation we first perform mechanistic simulations by considering the n agents of each cbsa and simulating the microscopic dynamics corresponding to both individual displacements and also the pairwise interactions of susceptible and infectious agents that give rise to contagions of the former. For every time step t , the following operations are carried out:

- i) First, each agent decides with probability p_d whether to travel to its destination patch or to stay in its node of origin (with probability $1 - p_d$).

- ii) The number of contacts made on each patch i are calculated according to its effective population density $f_i = n_i^{eff}/a_i$. As $z^D f_i$ is a rational number, we round to the two nearest integers and assign a probability to each of them so that the average value is equal to $z^D f_i$. For instance, if $z^D f_i = 5.20$, there is a probability 0.2 that the agent make 6 contacts while this probability is 0.8 for 5 contacts.

- iii) Then, for each susceptible individual, we randomly choose another agent from the same patch. If the latter is infected, we use a random variable with probability λ to determine if the former gets infected. In that case, we move on while, otherwise, we randomly choose another agent. We repeat the process until either the susceptible individual makes the expected number of contacts or gets infected.

- iv) Once all the pairwise contacts have been simulated for all the susceptible agents in all patches, individuals come back to their residence. Then, contacts at the household level are simulated. To this aim, every agent repeat the previous steps ii) and iii) considering $z^N \sigma_i$ as the number of interactions.

- v) Finally, infected individuals at time t become susceptible at time $t+1$ with probability μ .

We carry these simulations by considering different values of the infectivity λ and the mobility parameter p_d while keeping the recovery probability $\mu = 0.2$. For each simulation, we let the system evolve for a reasonable time of T days (here the natural time unit imposed by the commuting data) and compute the stationary value for the fraction of infected individuals, ρ^* . Since mechanistic simulations are stochastic, for each value of λ and p_d , we consider the average value of ρ^* obtained from 100 realizations of the mechanistic simulation with different initial conditions for the infectious seeds.

The results obtained through mechanistic simulations are confronted with those obtained by iterating the Markovian equations by comparing the stationary value for the fraction infectious individuals. In the case of the Markovian equations, the steady fraction of infectious agents, *i.e.* the average prevalence of the diseases, is calculated as:

$$\rho^* = \frac{1}{n} \sum_{i,j=1}^N n_{ij} \rho_{ij}^* \quad (25)$$

where the values ρ_{ij}^* are the stationary values of the entries of vector ρ and n is the total population in the system. Note that, since Markovian dynamics is deterministic and there is a unique stable fixed point attracting any initial condition with a non-zero number of infected individuals, no realizations are needed, allowing a fast calculation of epidemic curves $\rho^*(\lambda)$.

The results obtained by both methods are shown in Figure 2. In both panels, the fraction of infected individuals is represented as a function of the contagion probability $\rho(\lambda)$ for different values of p_d . The infectivity value is re-scaled by a

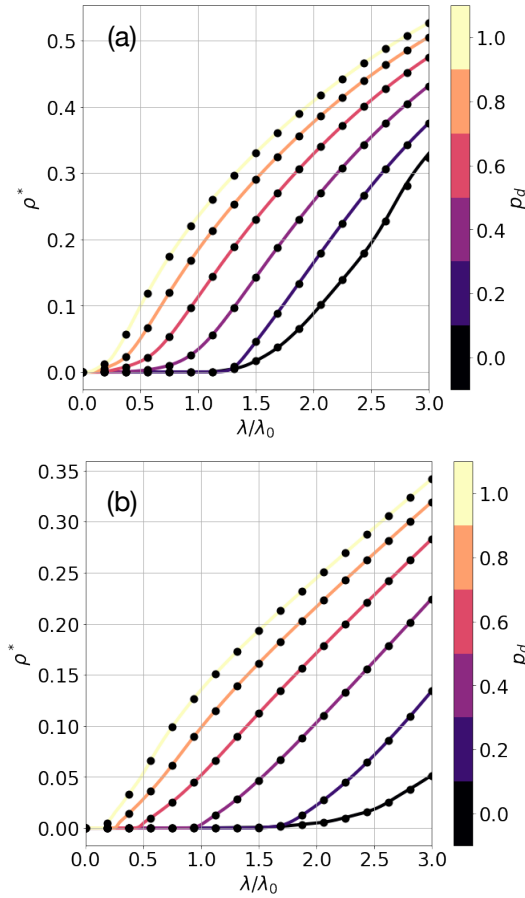


Figure 2. Validation of the Markovian equations for the cities of (a) New York and (b) Boston. On the vertical axis we represent the fraction of infected ρ^* in the stationary state for the whole population, and on the horizontal axis the parameter λ normalized to the epidemic threshold for $p_d = 0$: $\lambda_0 = \lambda_c(p_d = 0)$. Each panel shows different epidemic curves $\rho^*(\lambda/\lambda_0)$ for different values of p_d obtained by iterating the Markovian equations. In its turn, points represent the average of ρ^* obtained after 100 realizations of the mechanistic (Monte Carlo) simulations for each value of λ and p_d (note that error bars corresponding to the 95% confidence interval are smaller than the size of points). In all the cases the recovery probability is set to $\mu = 0.2$, the average contacts during the D cycle are $\langle k^D \rangle = 8$, and those that take place at the N one are set to $\langle k^D \rangle = 3$.

value λ_0 , which is the epidemic threshold λ_c when $p_d = 0$, i.e. $\lambda_0 = \lambda_c(p_d = 0)$. The value λ_0 is that of the SIS model in the most vulnerable patch, i.e. $\lambda_0 = \mu / \max(z^N \sigma_i + z^D f_i)$. From the plots it is clear that the agreement between the Markovian solution (curves) and the results from mechanistic simulations (points) is excellent, thus confirming the validity of the Markovian equations (24).

D. Epidemic threshold

We now tackle the analysis of the epidemic threshold λ_c . To this aim we focus on finding the stationary state of the SIS dynamics and thus assume that all the variables are time-independent: $\rho_{ij}(t+1) = \rho_{ij}(t) \equiv \rho_{ij}^* \forall i, j$. In this stationary regime Eq. (24) transforms into:

$$\mu \rho_{ij}^* = (1 - \rho_{ij}^*) \Pi_{ij}(\vec{\rho}^*). \quad (26)$$

Solving this equation by numerical means allows us to find the stationary value ρ^* used in Fig. (2). However, since our focus is those solutions close enough to λ_c , we are interested in a very particular situation that consists of arbitrarily small local prevalences: $\rho_{ij}^* \equiv \epsilon_{ij} \ll 1 \forall i, j$. This way we can linearize Eq. (26).

Since the function $\Pi_{ij}(\vec{\rho}^*)$ in Eq. (26) is highly nonlinear we start by considering the two sets of contagion probabilities $\{P_i^D(\vec{\rho}^*)\}$ and $\{P_i^N(\vec{\rho}^*)\}$ and linearize each of them as:

$$P_i^D \simeq p_d \lambda \frac{z^D f_i}{n_i^{eff}} \sum_{j=1}^N n_{ji} \epsilon_{ji} + \lambda (1 - p_d) \frac{z^D f_i}{n_i^{eff}} \sum_{j=1}^N n_{ij} \epsilon_{ij}, \quad (27)$$

$$P_i^N \simeq \lambda \frac{z^N \sigma_i}{n_i} \sum_{j=1}^N n_{ij} \epsilon_{ij}. \quad (28)$$

Next we insert the former expressions into Eq. (23) to obtain the linearized version of $\Pi_{ij}(\vec{\rho}^*)$ as:

$$\begin{aligned} \Pi_{ij}(\vec{\epsilon}) = \lambda \sum_{k,l=1}^N \left[(1 - p_d) p_d \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{ik} \right. \\ + (1 - p_d)^2 \frac{z^D f_l}{n_l^{eff}} n_{lk} \delta_{il} \\ + p_d^2 \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{jk} \\ + p_d (1 - p_d) \frac{z^D f_l}{n_l^{eff}} n_{lk} \delta_{jl} \\ \left. + \frac{z^N \sigma_l}{n_l} n_{lk} \delta_{il} \right] \epsilon_{lk} \end{aligned} \quad (29)$$

Finally, inserting this formula into Eq. (26) and neglecting nonlinear terms in ϵ_i , we obtain the following set of linear equations for the stationary prevalence $\vec{\epsilon}$:

$$\mu \epsilon_{ij} \simeq \Pi_{ij}(\vec{\epsilon}^*) = \lambda \sum_{k,l=1}^N M_{jk}^{il} \epsilon_{lk} \quad (30)$$

where, for convenience, we have written the expression in Eq. (29) as the product of the prevalence vector $\vec{\epsilon}$ by a matrix $\mathbf{M}$ whose terms are defined as:

$$\begin{aligned} M_{jk}^{il} = (1 - p_d) p_d \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{ik} + (1 - p_d)^2 \frac{z^D f_l}{n_l^{eff}} n_{lk} \delta_{il} \\ + p_d^2 \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{jk} + p_d (1 - p_d) \frac{z^D f_l}{n_l^{eff}} n_{lk} \delta_{jl} \\ + \frac{z^N \sigma_l}{n_l} n_{lk} \delta_{il}. \end{aligned} \quad (31)$$

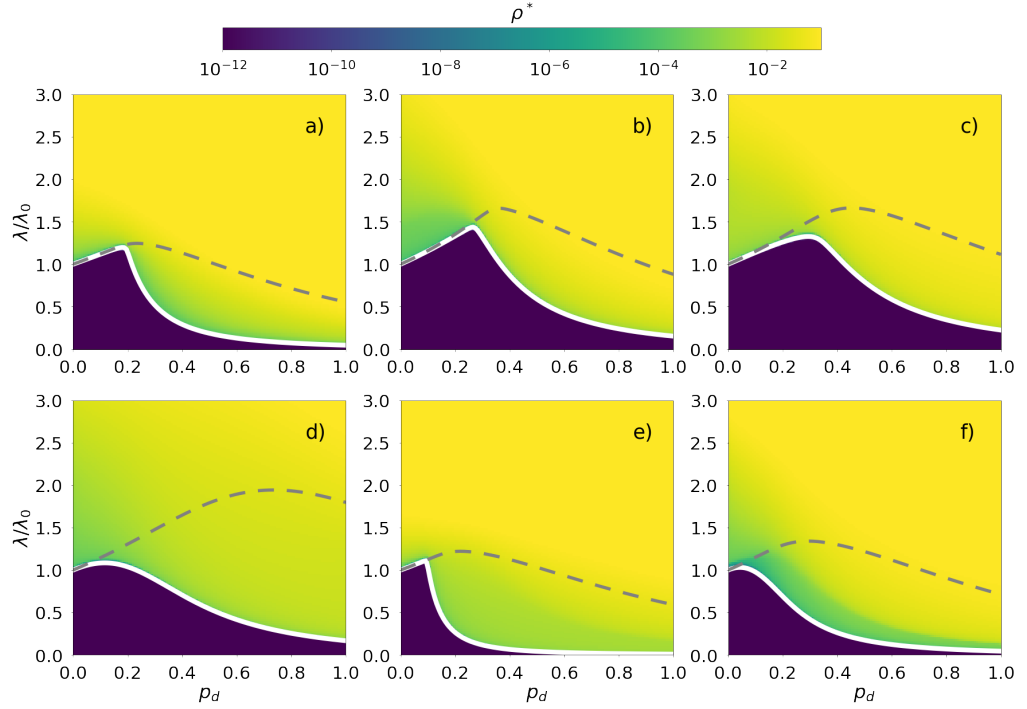


Figure 3. Epidemic diagrams $\rho^*(p_d, \lambda/\lambda_0)$ for the cities of (a) New York, (b) Boston, (c) Austin, (d) Miami, (e) Detroit and (f) Seattle. The continuous white line shows the epidemic threshold obtained by solving Eq. (34), whereas the dashed grey line accounts for the epidemic threshold of the indistinguishable case, *i.e.* obtained by solving Eq. (14).

The former matrix is the new *mixing matrix* for the distinguishable MIR model and, as its indistinguishable counterpart, Eq. (15), captures the different ways that residents in patch i traveling to j mix with agents from l traveling to k . Specifically, we can represent matrix $\mathbf{M}$ as follows:

$$\mathbf{M} = \begin{pmatrix} M_{11}^{11} & \dots & M_{1N}^{11} & \dots & | & M_{11}^{1N} & \dots & M_{1N}^{1N} \\ \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ M_{N1}^{11} & \dots & M_{NN}^{11} & \dots & | & M_{N1}^{1N} & \dots & M_{NN}^{1N} \\ \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ M_{11}^{N1} & \dots & M_{1N}^{N1} & \dots & | & M_{11}^{NN} & \dots & M_{1N}^{NN} \\ \dots & \dots & \dots & \dots & | & \dots & \dots & \dots \\ M_{N1}^{N1} & \dots & M_{NN}^{N1} & \dots & | & M_{N1}^{NN} & \dots & M_{NN}^{NN} \end{pmatrix}. \quad (32)$$

Note that with the former formulation each sub-matrix M^{il} of $\mathbf{M}$ contains the elements that relate those individuals with residence in patch i with those living in l .

Taking advantage of the definition of the mixing matrix $\mathbf{M}$ Eq. (30) can be written in a compact form as:

$$\frac{\mu}{\lambda} \vec{\epsilon} = \mathbf{M} \vec{\epsilon}, \quad (33)$$

i.e. an eigenvalue problem. From all the possible solutions (eigenvectors) of Eq. (30) we are interested in the one corresponding to the minimum value of λ (as it defines the epidemic threshold of the metapopulation), that corresponds to the maximum eigenvalue of $\mathbf{M}$. Thus, the epidemic threshold reads:

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})}. \quad (34)$$

Finally, we verify that Eq. (34) and the derivation of the mixing matrix (31) are correct by computing the epidemic diagram $\rho(p_d, \lambda)$ and comparing with the theoretical prediction for $\lambda_c(p_d)$. We have performed this analysis for 6 US cbsa (see Table I for details): New York, Boston, Austin, Miami, Detroit and Seattle. These results are shown in Fig. 3 where λ has been normalized to λ_0 so that at $p_d = 0$ the normalized epidemic threshold is 1. In each panel we overlay to each contour plot $\rho(p_d, \lambda/\lambda_0)$ the curve $\lambda_c(p_d)/\lambda_0$ as derived by calculating the spectral radius of the mixing matrix $\mathbf{M}$ of each city and applying Eq. (34). The agreement of the analytical formula is excellent and highlights the importance of considering the specific commuting patterns and the spatial distribution of the populations (the two key elements of the matrix $\mathbf{M}$) to assess the robustness of populations subjected to the spread of communicable diseases.

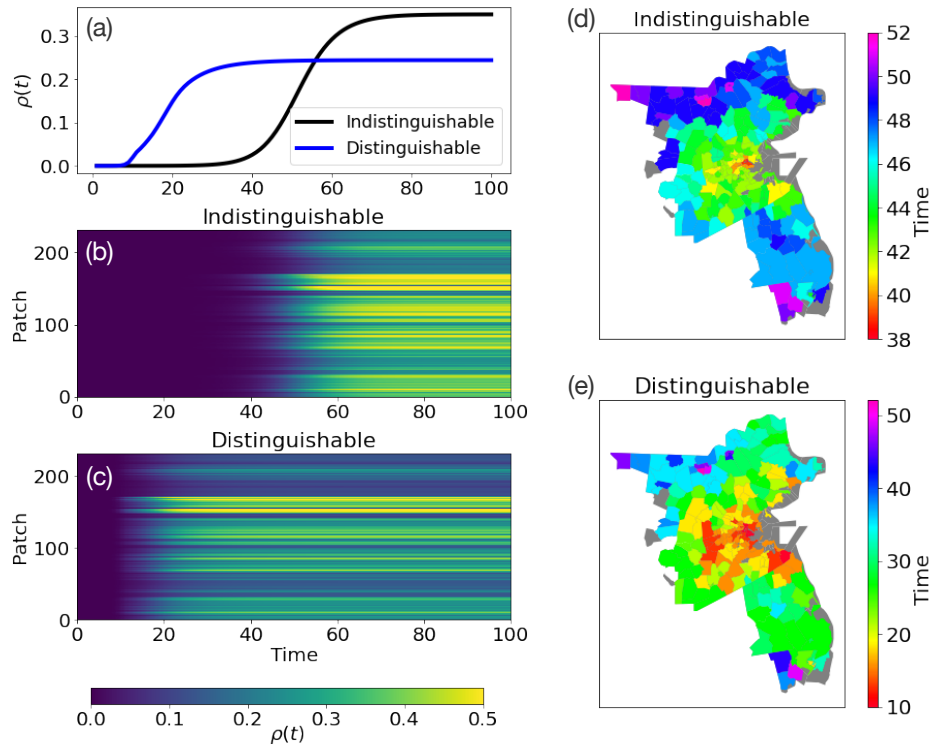


Figure 4. Spatio-temporal unfolding of an epidemic in the Boston area under the indistinguishable and distinguishable frameworks. In (a) we show the time evolution of the global prevalence ρ^* . Panels (b) and (c) show the time evolution of the local prevalence of each patch. Finally, panels (d) and (e) show the map of the Boston metropolitan area in which each subpopulation is colored according to the time of arrival of the first infections. This time is calculated as the time required for the instantaneous prevalence to reach 5% of the local population.

Remarkably, we observe the so-called epidemic detriment in all the explored cities, *i.e.* the increase of the epidemic threshold λ_c for values $p_d > 0$ compared to the case $p_d = 0$. Thus, the distinguishable MIR framework preserves the main result derived by the indistinguishable one²⁴, as it is rooted on the heterogeneous distribution of the population and the flows across patches. This counter-intuitive phenomenon can be easily explained considering the case $p_d = 0$ and $\lambda \gtrsim \lambda_0$. In this case, infectious agents are concentrated in the most vulnerable areas while the rest of the patches are disease-free subpopulations. By increasing p_d we spread the carriers of the disease to other areas with lower infection risk where, for the same value of $\lambda \gtrsim \lambda_0$, will not spread the pathogen. Besides, those individuals moving from the least to the most exposed areas are generally healthy. Thus, for $p_d \gtrsim 0$ and $\lambda \gtrsim \lambda_0$, the infected individuals will recover without making secondary infections in small size patches, while the localized outbreak in the largest population patch dies out.

E. Distinguishable vs. Indistinguishable behaviors

Although in terms of the epidemic detriment the results for indistinguishable and distinguishable agents are qualitatively similar, there are important quantitative differences between the two cases that we now analyze. First, we focus on the dependence of the epidemic threshold with mobility. In the panels of Fig. 3 we have included the function $\lambda_c(p_d)/\lambda_0$ (dashed grey lines) for the case of indistinguishable agents calculated as in Eq. (14), *i.e.* through the computation of the spectral radius of the mixing matrix for the indistinguishable case, whose elements are given by Eq (15). Let us note that the normalization factor in both curves, λ_0 , remains the same for the indistinguishable and distinguishable cases since it corresponds to the epidemic threshold when no mobility is at work, $\lambda_c(p_d = 0)$, which is identical for both scenarios.

From the panels in Fig. 3 it becomes clear that the epidemic threshold is always smaller in the distinguishable case than in the indistinguishable one. This difference is related to the underlying flow heterogeneity of the mobility networks. In the distinguishable case, individuals who visit the main focus of infection, and thus import contagions to their residences,

maintain this infectious flow between these two specific areas over time. However, in the indistinguishable case, an individual who has visited the focus may end up in a different location in the following time steps, diluting the effect of contagions between different subpopulations and thus avoiding outbreaks of secondary contagions in these subpopulations.

These differences become more pronounced as the value of p_d increases. However, when mobility is extremely low, $p_d \gtrsim 0$, the two formalisms coincide exactly and the curves $\lambda_c(p_d)/\lambda_0$ have the same slope. It is reasonable to think that, since the detriment is because the residents of the infectious focus begin to leave that node, at first order when travel is scarce the indistinguishable formalism behaves the same as the distinguishable one. In particular, for the indistinguishable case, if the interval between two consecutive trips is much longer than the duration of the infectious window of an infected individual, it will rarely visit different patches during its contagious cycle and will behave, for contagion purposes, as a distinguishable individual.

We now focus on studying their behavior in the supercritical regime. In particular, in this regime we are interested in monitoring the transient from the initial state in which a small infectious seed is placed in a single patch to the endemic regime after the subsequent infections spread through the entire metapopulation. In Fig. 4 we use as a test framework the city of Boston, and placing the same infectious seed for the two formalisms we analyze its expansion for $\lambda = 2\lambda_0$ and $p_d = 1$.

In Fig. 4.a we show the time evolution for the global fraction of infected individuals. It is clear that in the distinguishable case the pathogen spreads initially faster than in the indistinguishable framework. Moreover, the distinguishable prevalence reaches its stationary value when the indistinguishable prevalence is still negligible. However, once the epidemic unfolds in the indistinguishable scenario, it reaches a stationary prevalence much larger than in the distinguishable.

Apart from the different time scales involved in the transient process to the endemic equilibrium we can monitor the different spatio-temporal evolution from the initial outbreak to the stationary regime. The two epidemic trajectories are compared in Figs. 4.b-c. From these two plots it is clear that the epidemic unfolding under the two frameworks follows different paths although, given the different times scales involved, it is difficult to pinpoint those areas for which these differences become more relevant. To shed more light on the different spatio-temporal patterns of each propagation process, in Figs. 4.d-e we have drawn the map of the Boston area by assigning each patch a time value. This value accounts for the exact time at which the infected fraction of the patch reaches 5% of its population. It is clear from the two colored maps that the indistinguishable case follows a more explosive behavior, so that once the contagions spread out of the central Boston area, where the infectious seeds are located, the disease reaches the entire metropolitan population in a few time-steps. In contrast, for the distinguishable frame the process is highly sequential and takes more than 40 time steps to reach the entire population following, moreover, a radial pattern from the center to the periphery.

III. CONCLUSIONS

In this paper we have developed a Markovian framework based on the MIR model to analyze the spread of communicable diseases in networked metapopulations. Unlike its original formulation in which agents are labeled according to their residence, the new framework incorporates the distinguishable nature of agents according to both their residence and recurrent destinations. This new framework incorporates a further partition of subpopulations that can be easily gathered from commuting data and is certainly valuable for studying the spread of diseases in metropolitan areas since recurrent paths apply to most of the mobility flows.

After validating the Markovian equations by comparing with results obtained with mechanistic agent-based simulations, we have derived a new mixing matrix that captures the basic interaction mechanisms between the subpopulations of agents at different patches. By calculating the spectral radius of this mixing matrix we can estimate the epidemic threshold of the metapopulation, allowing us to estimate its vulnerability to the spread of pathogens.

Finally, we have compared the results obtained considering indistinguishable and distinguishable agents, taking advantage of the fact that, in the limit of null mobility, both frames are identical. Firstly, we have shown that indistinguishability overestimates the value of the epidemic threshold, demonstrating how distinguishability does not allow contagions produced in areas of high prevalence to be diluted between different patches, making it easier for them to cause secondary infections. This overestimation of the epidemic threshold, however, does not eliminate the phenomenon of the epidemic detriment that is still observed in the case of distinguishable agents. Likewise, we have observed that the spatio-temporal diffusion for the distinguishable case occurs more progressively than in the indistinguishable case, spreading spatially much more explosively.

Our manuscript fuels the discussion on the relevance of the nature of the mobility schemes introduced in the theoretical framework to provide a fair assessment of the evolution of diseases. In random-walker dynamics, Castioni et al.⁴² demonstrate that the outcome of control policies shaping mobility is strongly shaped by the ratio between the time scales involved in both movements and contagions. The latter ratio is also crucial when one accounts for recurrent mobility patterns and its variation changes the critical properties of the metapopulation, leading to a vanishing of the epidemic detriment here reported in some scenarios²⁹. On more general grounds, different theoretical works have shown that the inclusion of different mobility schemes^{43,44}, the information loss when translating higher-order flows in origin-destination matrices⁴⁵, the introduction of biases in the collection of mobility data^{46,47} or the misuse of raw mobility data as OD matrices⁴⁸ lead to substantial differences in the evolution of epidemics.

Despite its simplicity, the new distinguishable framework of the MIR model opens the door to the implementation of more accurate descriptions of real urban environments, a context where recurrent mobility flows predominate and precise identification of possible contagion pathways is much needed.

Nevertheless, it is worth stressing that the model here proposed does not capture entirely the weekly mobility rhythms of the population, as assuming a fixed destination over the weekends misrepresents the heterogeneous and variable nature of our usual mobility patterns in leisure time^{49,50}. The formulation of a framework taking into account the time-varying nature of our mobility patterns remains as future work. In epidemiological terms, we have focused here on a simplified version in which an epidemic SIS model is at work, but this formalism can be generalized to any other compartmental dynamics along with other refinements such as considering the age partitioning of the population or the inclusion of more complex interactions and mobility patterns.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS

P.V., D.S.P. and J.G.G. acknowledge financial support from the Departamento de Industria e Innovación del Gobierno de Aragón y Fondo Social Europeo (FENOL group grant E36-17R), from grant PID2020-113582GB-I00 funded by MCIN/AEI/10.13039/501100011033, and from Fundación Ibercaja and Universidad de Zaragoza (grant 224220). A.A. acknowledges financial support from Spanish MINECO (Grant No. PGC2018-094754-B-C21), Generalitat de Catalunya (Grant No. 2017SGR-896 and 2020PAND00098), Universitat Rovira i Virgili (Grant No. 2019PFR-URV-B2-41), Generalitat de Catalunya (PDAD14/20/00001), ICREA Academia, and the James S. McDonnell Foundation (Grant No. 220020325).

REFERENCES

- ¹A. M'Kendrick, "Applications of mathematics to medical problems," *Proceedings of the Edinburgh Mathematical Society* **44**, 98–130 (1925).
- ²W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* **115**, 700–721 (1927).
- ³L. F. Richardson, *Weather predictions by Numerical Process* (Cambridge University Press, 1922).
- ⁴R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, 1992).
- ⁵M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals* (Princeton university press, 2011).
- ⁶R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of modern physics* **87**, 925 (2015).
- ⁷S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature* **429**, 180–184 (2004).
- ⁸D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, and A. Vespignani, "Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model," *Journal of computational science* **1**, 132–145 (2010).
- ⁹M. Tizzoni, P. Bajardi, C. Poletto, J. J. Ramasco, D. Balcan, B. Gonçalves, N. Perra, V. Colizza, and A. Vespignani, "Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm," *BMC medicine* **10**, 1–31 (2012).
- ¹⁰C. C. Kerr, R. M. Stuart, D. Mistry, R. G. Abeysuriya, K. Rosenfeld, G. R. Hart, R. C. Núñez, J. A. Cohen, P. Selvaraj, B. Hagedorn, *et al.*, "Covasim: an agent-based model of covid-19 dynamics and interventions," *PLOS Computational Biology* **17**, e1009149 (2021).
- ¹¹F. Parino, L. Zino, M. Porfiri, and A. Rizzo, "Modelling and predicting the effect of social distancing and travel restrictions on covid-19 spreading," *Journal of the Royal Society Interface* **18**, 20200875.
- ¹²L. Sattenspiel and K. Dietz, "A structured epidemic model incorporating geographic mobility among regions," *Mathematical biosciences* **128**, 71–91 (1995).
- ¹³B. Grenfell and J. Harwood, "(meta) population dynamics of infectious diseases," *Trends in ecology & evolution* **12**, 395–399 (1997).
- ¹⁴F. Ball, T. Britton, T. House, V. Isham, D. Mollison, L. Pellis, and G. S. Tomba, "Seven challenges for metapopulation models of epidemics, including households models," *Epidemics* **10**, 63–67 (2015).
- ¹⁵H. 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* **734**, 1–74 (2018).
- ¹⁶V. Colizza, R. Pastor-Satorras, and A. Vespignani, "Reaction–diffusion processes and metapopulation models in heterogeneous networks," *Nature Physics* **3**, 276–282 (2007).
- ¹⁷V. Colizza and A. Vespignani, "Invasion threshold in heterogeneous metapopulation networks," *Physical review letters* **99**, 148701 (2007).
- ¹⁸V. Colizza and A. Vespignani, "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations," *Journal of theoretical biology* **251**, 450–467 (2008).
- ¹⁹D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences* **106**, 21484–21489 (2009).
- ²⁰D. Balcan and A. Vespignani, "Phase transitions in contagion processes mediated by recurrent mobility patterns," *Nature physics* **7**, 581–586 (2011).
- ²¹V. Belik, T. Geisel, and D. Brockmann, "Natural human mobility patterns and spatial spread of infectious diseases," *Physical Review X* **1**, 011001 (2011).
- ²²D. Balcan and A. Vespignani, "Invasion threshold in structured populations with recurrent mobility patterns," *Journal of theoretical biology* **293**, 87–100 (2012).
- ²³V. Belik, T. Geisel, and D. Brockmann, "Recurrent host mobility in spatial epidemics: beyond reaction-diffusion," *The European Physical Journal B* **84**, 579–587 (2011).
- ²⁴J. Gómez-Gardeñes, D. Soriano-Paños, and A. Arenas, "Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks," *Nature Physics* **14**, 391–395 (2018).
- ²⁵D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, "Spreading processes in multiplex metapopulations containing different mobility networks," *Physical Review X* **8**, 031039 (2018).
- ²⁶P. Bosetti, P. Poletti, M. Stella, B. Lepri, S. Merler, and M. De Domenico, "Heterogeneity in social and epidemiological factors determines the risk of measles outbreaks," *Proceedings of the National Academy of Sciences* **117**, 30118–30125 (2020).
- ²⁷D. 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," *Physical Review Research* **2**, 013312 (2020).
- ²⁸A. Anzo-Hernández, B. Bonilla-Capilla, J. Velázquez-Castro, M. Soto-Bajo, and A. Fraguela-Collar, "The risk matrix of vector-borne diseases in metapopulation networks and its relation with local and global R_0 ," *Communications in Nonlinear Science and Numerical Simulation* **68**, 1–14 (2019).
- ²⁹D. Soriano-Paños, G. Ghoshal, A. Arenas, and J. Gómez-Gardeñes, "Impact of temporal scales and recurrent mobility patterns on the unfolding of epidemics," *Journal of Statistical Mechanics: Theory and Experiment* **2020**, 024006 (2020).
- ³⁰W. Cota, D. Soriano, A. Arenas, S. Ferreira, and J. Gomez-Gardenes, "Infectious disease dynamics in metapopulations with heterogeneous transmission and recurrent mobility," *New Journal of Physics* (2021).

- ³¹A. Arenas, W. Cota, J. Gómez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Paños, and B. Steinegger, "Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions," *Physical Review X* **10**, 041055 (2020).
- ³²G. S. Costa, W. Cota, and S. C. Ferreira, "Outbreak diversity in epidemic waves propagating through distinct geographical scales," *Physical Review Research* **2**, 043306 (2020).
- ³³S. Hazaric, D. Soriano-Paños, A. Arenas, J. Gómez-Gardeñes, and G. Ghoshal, "Interplay between population density and mobility in determining the spread of epidemics in cities," *Communications Physics* **4**, 1–10 (2021).
- ³⁴X. Zhu, Y. Liu, S. Wang, R. Wang, X. Chen, and W. Wang, "Allocating resources for epidemic spreading on metapopulation networks," *Applied Mathematics and Computation* **411**, 126531 (2021).
- ³⁵B. Wang, M. Gou, and Y. Han, "Impacts of information propagation on epidemic spread over different migration routes," *Nonlinear Dynamics* **105**, 3835–3847 (2021).
- ³⁶D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *science* **342**, 1337–1342 (2013).
- ³⁷M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, O. C.-, D. W. Groupf, L. du Plessis, N. R. Faria, R. Li, *et al.*, "The effect of human mobility and control measures on the covid-19 epidemic in china," *Science* **368**, 493–497 (2020).
- ³⁸C. Granell and P. J. Mucha, "Epidemic spreading in localized environments with recurrent mobility patterns," *Physical Review E* **97**, 052302 (2018).
- ³⁹(), united States mobility data sources are available at: <https://1ehd.ces.census.gov/data/1odes/L0DES7/> (Accessed: 2021-09-30).
- ⁴⁰(), united States population distribution data sources are available at: <https://www2.census.gov/geo/tiger/TIGER2010BLKPOP0PHU/> (Accessed: 2021-09-30).
- ⁴¹United States data for the areas of patches are available at: <https://www2.census.gov/geo/tiger/TIGER2010/ZCTA5/2010/> (Accessed: 2021-09-30).
- ⁴²P. Castioni, R. Gallotti, and M. De Domenico, "Critical behavior in interdependent spatial spreading processes with distinct characteristic time scales," *Communications Physics* **4**, 1–10 (2021).
- ⁴³M. J. Keeling, L. Danon, M. C. Vernon, and T. A. House, "Individual identity and movement networks for disease metapopulations," *Proceedings of the National Academy of Sciences* **107**, 8866–8870 (2010).
- ⁴⁴L. Danon, T. House, and M. J. Keeling, "The role of routine versus random movements on the spread of disease in great britain," *Epidemics* **1**, 250–258 (2009).
- ⁴⁵J. T. Matamalas, M. De Domenico, and A. Arenas, "Assessing reliable human mobility patterns from higher order memory in mobile communications," *Journal of the Royal Society Interface* **13**, 20160203 (2016).
- ⁴⁶M. Tizzoni, P. Bajardi, A. Decuyper, G. Kon Kam 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* **10**, e1003716 (2014).
- ⁴⁷F. Schlosser, V. Sekara, D. Brockmann, and M. Garcia-Herranz, "Biases in human mobility data impact epidemic modeling," *arXiv preprint arXiv:2112.12521* (2021).
- ⁴⁸S. Gómez, A. Fernández, S. Meloni, and A. Arenas, "Impact of origin-destination information in epidemic spreading," *Scientific reports* **9**, 1–9 (2019).
- ⁴⁹L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann, and A. Baronchelli, "Evidence for a conserved quantity in human mobility," *Nature human behaviour* **2**, 485–491 (2018).
- ⁵⁰M. Schläpfer, L. Dong, K. O'Keeffe, P. Santi, M. Szell, H. Salat, S. Ankersaria, M. Vazifeh, C. Ratti, and G. B. West, "The universal visitation law of human mobility," *Nature* **593**, 522–527 (2021).

Capítulo 5

Control epidémico basado en el comportamiento humano en paisajes urbanos

<https://doi.org/10.1038/s44260-024-00021-z>

Human behavior-driven epidemic surveillance in urban landscapes

Check for updates

P. Valgañón^{1,2}, A. F. Useche^{3,4}, F. Montes⁴, A. Arenas⁵✉, D. Soriano-Paños^{2,5,6}✉ & J. Gómez-Gardeñes^{1,2,7}✉

We introduce a surveillance strategy specifically designed for urban areas to enhance preparedness and response to disease outbreaks by leveraging the unique characteristics of human behavior within urban contexts. By integrating data on individual residences and travel patterns, we construct a Mixing matrix that facilitates the identification of critical pathways that ease pathogen transmission across urban landscapes enabling targeted testing strategies. Our approach not only enhances public health systems' ability to provide early epidemiological alerts but also underscores the variability in strategy effectiveness based on urban layout. We prove the feasibility of our mobility-informed policies by mapping essential mobility links to major transit stations, showing that few resources focused on specific stations yields a more effective surveillance than non-targeted approaches. This study emphasizes the critical role of integrating human behavioral patterns into epidemic management strategies to improve the preparedness and resilience of major cities against future outbreaks.

Throughout history, the interplay between epidemics and human societies has been profound, each significantly influencing and shaping the course of the other^{1–3}. This interplay, already present during our early hunter-gatherer days, was notably boosted by the establishment of agrarian societies around 10,000 years ago, when the creation of communities provided fertile breeding grounds for diseases to thrive. It was during this transformative period that humanity first encountered diseases such as malaria, tuberculosis, leprosy, influenza, and smallpox, which have since significantly altered the course of human history⁴.

As human settlements transformed into nowadays vast urban centers, they also introduced new challenges for contemporary epidemiology^{5,6}. Urban environments, characterized by dense populations, complex social interactions, and socioeconomic disparities, create ideal conditions for the spread of communicable diseases^{7–9}. Moreover, the rapid movement of people and goods within and across cities sweeps out national and continental boundaries, thus facilitating the global dissemination of pathogens between major city centers¹⁰.

The advent of big data and advanced modeling techniques offers new avenues for understanding and managing this cocktail of epidemic boosters, emphasizing the importance of integrating mobility data and demographic insights into urban epidemic management strategies¹¹. In this line, data

analytics have significantly advanced epidemic modeling^{12,13}, enabling the mathematical formalization of complexities associated with demographic segregation, mobility patterns, and heterogeneous contacts-key factors in disease propagation¹⁴.

Utilizing metapopulation frameworks, epidemic models effectively manage the interaction among the above ingredients by coupling contact-driven transmission with mobility-related dispersal^{15–22}. In recent decades, such behaviorally informed metapopulation models have become indispensable for mechanistic forecasting, enabling precise predictions of epidemic trajectories across diverse scales—from local communities^{23,24} through national^{25–27} to global levels^{28–30}. Moreover, contemporary epidemic models extend beyond mapping the spatio-temporal spread of diseases, but also facilitate the development of data-informed containment strategies with maximal resource efficiency and minimal socioeconomic disruption^{31–34}.

These efforts highlight the critical role that incorporating social dynamics into epidemic models has on our capacity to respond to infectious disease threats, an aspect that has recently materialized in research agendas^{35–37} aimed at advancing this endeavor. These agendas, among other problems, specifically highlight the challenge of employing realistic human contact structures to explore localization behavior in *key subpopulations* for epidemic control policies. Addressing this precise challenge, here we

¹Department of Condensed Matter Physics, University of Zaragoza, 50009 Zaragoza, Spain. ²GOTHAM lab, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain. ³Department of Health Management and Policy, Dornsife School of Public Health, Drexel University, Philadelphia, PA, 19104, USA. ⁴Department of Industrial Engineering, School of Engineering, Universidad de Los Andes, 111711 Bogotá, Colombia.

⁵Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain. ⁶Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal. ⁷Center for Computational Social Science, University of Kobe, 657-8501 Kobe, Japan. ✉e-mail: alexandre.arenas@urv.cat;

sorianopanos@gmail.com; gardenes@unizar.es

introduce a data-informed approach tailored for large urban centers, leveraging their unique demographic and mobility patterns. Our framework allows identifying critical human flows and leads to actionable surveillance strategies to obtain valuable early warnings of incoming epidemic outbreaks.

Results

Mixing matrix $\mathbf{M}$

To address the former challenge, we first present a formalism aimed at capturing the complex spatiotemporal structure of human mobility, particularly within urban environments. Over the last decades, extensive research has been devoted to reveal the complex and rich spatiotemporal structure of human mobility patterns^{38,39}. When focusing on urban environments, public surveys reveal that human flows are mainly dominated by recurrent mobility patterns^{40,41} connecting the residence of individuals with the work place to which they usually commute. Consequently, recurrent human flows have been incorporated into theoretical frameworks^{42–46} as the backbone of metapopulation models tailored to track urban epidemics. This way, the analysis of origin-destination matrices (OD), either constructed from mobile phone devices or from census surveys, have been instrumental in shedding light into the vulnerability of cities during the initial stages of epidemics and the spatially uneven distribution of cases^{47–49}.

Regardless of the former advances, the feasibility and efficiency of surveillance campaigns and control strategies driven by commuting data remains to be explored. To tackle this challenge here we leverage the formalism proposed in ref. 50 for the study of contagion-diffusion processes with recurrent mobility patterns of distinguishable agents. This model represents a city as a collection of P interconnected patches (see Fig. 1a), each one representing a geographical area. The mobility data informs the matrix $\mathbf{n}$, whose elements, n_{ij} , identify subgroups of individuals living inside patch i and regularly traveling to patch j . Therefore, a generic patch, say i , is populated by n_i residents, divided into subgroups based on their travel destinations ($n_i = \sum_j n_{ij}$), as schematized in Fig. 1a.

Once constructed the matrix $\mathbf{n}$, we can simulate reactive processes, such as contagion dynamics, coupled with the diffusion of the population. To this aim, the model assumes a discrete-time approach, where each time step consists of four different stages. First, each individual decides to move to the usual destination with probability p_d or stay in the residence area. Second, (day) interactions take place with those agents sharing the same location. Third, those individuals that moved in the first stage return to their residence and, fourth, another interaction stage takes place when agents make local contacts inside their subpopulation of origin (night interactions). These contacts reflect interactions with members of their own household or neighborhood. In any of the two interaction stages occurring in each time step, contagions can occur. A more exhaustive description of the steps present in the simulation scheme is found in the Methods section.

The processes described above can be encapsulated into a Mixing matrix $\mathbf{M}$ which contains the expected number of interactions among the different subgroups of individuals found in an urban environment. We illustrate in Fig. 1b–e the elements of the mixing matrix associated to a specific subgroup of individuals with residence in patch i and destination in patch j . These elements capture all the contagion venues affecting these individuals as a result of their mobility across the city. Namely, these individuals can interact with other residents staying there (Fig. 1b) and with others coming from a different patch (Fig. 1c). In contrast, when moving to their usual destination, they interact with its residents (Fig. 1d) and with other visitors also moving there (Fig. 1e). The mathematical expression of the elements of the Mixing matrix $\mathbf{M}$ can be found in the Methods section whereas their derivation from the equations governing the epidemiological model used here is explained in the Supplementary Information.

Mobility-informed testing and quarantine policies

The leading eigenvector of Mixing matrices, derived from epidemic models on networked populations, plays a pivotal role in pinpointing key actors in epidemic outbreaks, whether they are individuals⁵¹ or specific geographical areas^{32,33}. Expanding upon these insights, we introduce a testing strategy that prioritizes screening of those agents participating in origin-destination trips with a large contribution to the leading eigenvector of the matrix $\mathbf{M}$. The operational framework of this policy is the Susceptible-Exposed-Infected-Recovered-Quarantine (SEIRQ) model, as depicted in Fig. 2a, so that those agents that test positive are isolated from the population as a control policy.

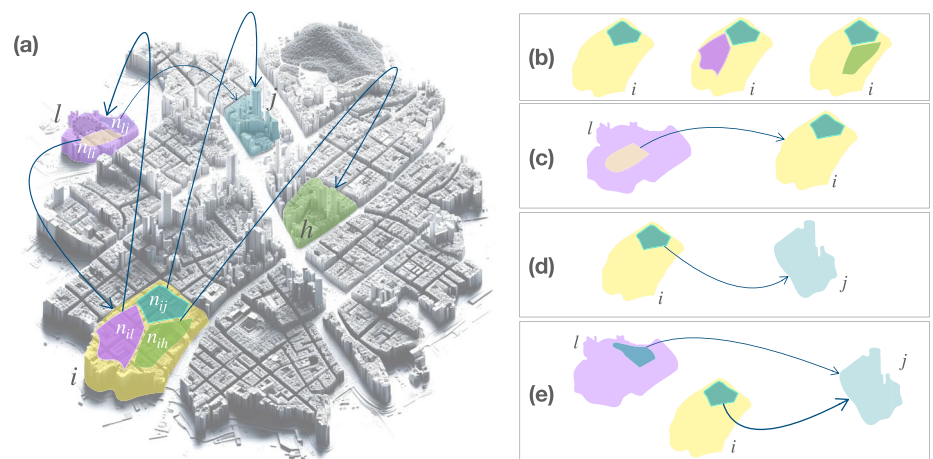
As illustrated in Fig. 2b, our testing policy hinges on two critical parameters: the total number of tests n_{tests} at our disposal, and the links in the mobility network L targeted by these tests. By adjusting the parameter L , our strategy can transition from a focused approach, targeting individuals taking part in those trips with the largest contributions to the leading eigenvector of $\mathbf{M}$, to a more distributed allocation of tests across the urban landscape. A more detailed exposition of the strategies for the distribution of testing resources is provided in the Methods section.

We apply our mobility-informed testing strategy to two real scenarios, the cities of Bogotá and Miami, whose mixing matrices are constructed from demographic and mobility data drawn from public surveys, as detailed in the Methods section. For this purpose, we simulate several outbreaks caused by a pathogen with basic reproduction number $\mathcal{R}_0 = 4$ and for which we have $n_{\text{tests}} = 2 \times 10^5$ daily tests to counteract its spread. This way, varying the number L allows us to evaluate the impact of different testing distributions.

In Fig. 2c we plot the time evolution of the number of infected for different strategies. Starting from an scenario when resources are distributed quite evenly across the city ($L = 5 \times 10^4$), decreasing L yields the usual flattening of the epidemic curves, reducing and delaying the peak of

Fig. 1 | Mobility model and mixing of the population.

a Illustration of the mobility model here considered. We assume a metapopulation from daily back-and-forth movements recorded in public surveys. Each geographical patch i in the metapopulation is partitioned into different subgroups according to their usual destination, being n_{ij} the number of residents inside i daily commuting to patch j . **b–e** Mixing patterns of the subpopulation of n_{ij} residents in i with typical destination in j . When staying in their residential patch i , these residents might interact with others staying also there (**b**) and also with residents from other patches (**c**) that have as usual destination patch i . In turn, when moving, residents in i will interact with those residents from other patches: (**d**) either those that belong to the usual destination j and that decide not to move or (**e**) with residents of a patch l who also have as usual destination patch j .



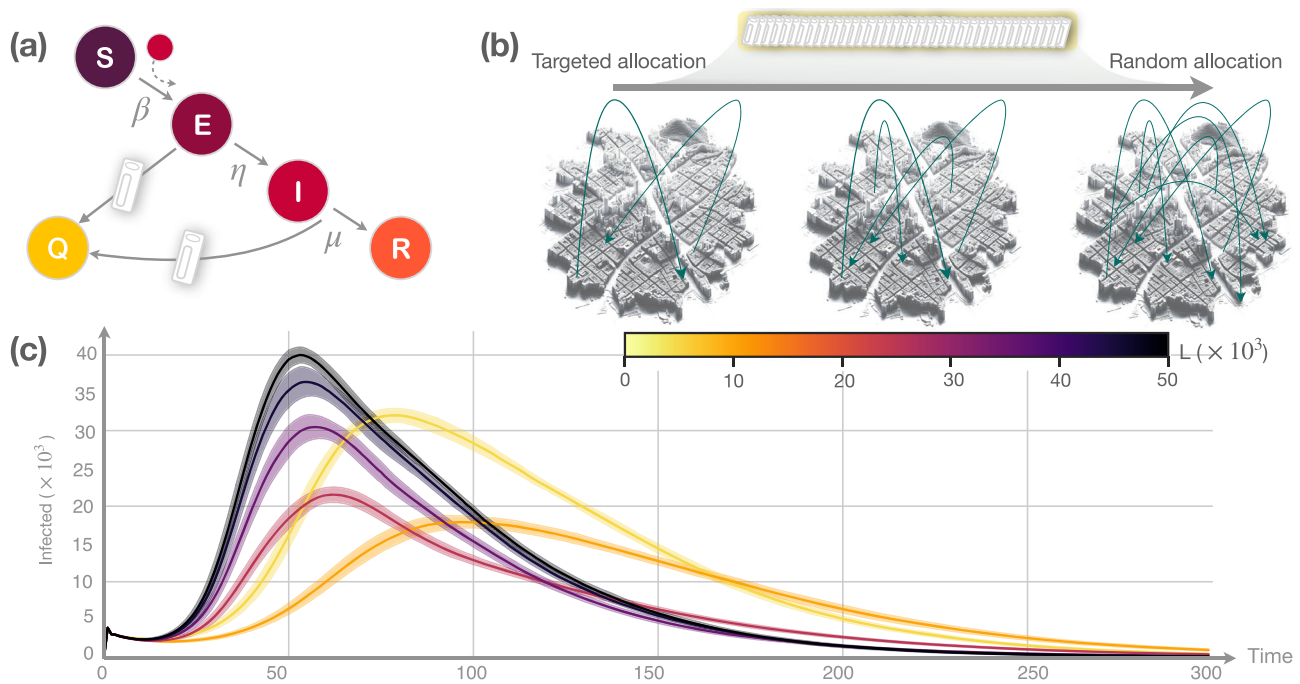


Fig. 2 | Mobility-informed testing and quarantine policies. **a** Susceptible-Exposed-Infected-Recovered-Quarantined (SEIRQ) compartmental model used in the epidemic simulations. In addition to the usual flows governed by the pathogen infectiousness λ , the incubation period η^{-1} and the recovery probability μ , we assume that Infected and Exposed individuals are quarantined after being tested. **b** Illustration of the testing strategy implemented in the model. Such strategy depends on the amount of daily tests available to screen the population n_{tests} and on how these tests are allocated across the city, governed by L . Small values of L capture

targeted resource allocation on the most vulnerable subgroups defined from the Mixing matrix $\mathbf{M}$ whereas large L values correspond to their random allocation. **c** Time evolution of the number of infected individuals in Bogotá when distributing $n_{\text{tests}} = 2 \times 10^5$ tests every day on the L most critical links (color code). Solid lines represent the average time evolution and the shadowed region corresponds to 95% confidence interval over 25 simulations for each L value. The basic reproduction number of the pathogen responsible for the outbreak is $\mathcal{R}_0 = 4$.

contagions. However, when tests are concentrated in the individuals of a few critical links ($L = 2 \times 10^3$), spreading them out by increasing L also favors epidemic control. Overall, our analysis for the city of Bogotá reveals the existence of an optimal number of links L_{opt}^* on which testing and isolation policies should be prioritized to efficiently control an outbreak. Remarkably, Supplementary Fig. 1 shows that such L_{opt}^* value does not exist for Miami, pinpointing that in this city the concentration of tests on a few subgroups of individuals, i.e. reducing L , seems to be the best control strategy.

Optimal control policies depend on urban landscapes

To further showcase the effects of targeted testing and quarantine policies, we now focus on the impact of the mobility-informed policies on the attack rate, defined as the total number of individuals contracting the disease during an epidemic outbreak. Fig. 3a–b show how the latter indicator depends on both the basic reproduction number of the disease $\mathcal{R}_0$ and the spatial allocation of the n_{tests} across L mobility links in Bogotá and Miami respectively. Figure 3a reaffirms the existence of an optimal value $L_{\text{opt}}^*(\mathcal{R}_0)$ of targeted links reducing the long-term impact of epidemics on Bogotá. Interestingly, such value is not independent of the epidemic process but strongly varies with the basic reproduction number $\mathcal{R}_0$, yielding a complex interplay between the extent of an outbreak and the optimal spatial allocation of resources to reduce its burden. Conversely, Fig. 3b shows that, in Miami, concentrating testing efforts on a few critical links allows mitigating an outbreak even for relatively large $\mathcal{R}_0$ values. Similar findings are presented in Supplementary Fig. 2 for other cities retrieving the same behaviors observed in Bogotá or Miami. The optimal L_{opt}^* value is found in the cities of New York and Boston, while the city of Austin returns the same outcomes as Miami. The results are further reinforced in Supplementary Figs. 3 and 4 where we perform a sensitivity analysis on other parameters, namely the variables η (the latency probability) and the mobility p , proving the qualitative results remain consistent.

Our findings underscore that mobility-informed policies should be adapted to the unique characteristics of each urban environment^{52,53}. Specifically, in cities with complex and segregated socio-economic activities, concentrating resources on the most critical links does not always ensure community-wide protection and thus proves ineffective. Conversely, this strategy yields significant benefits in cities where the most relevant links are primarily confluent towards a few and close patches. These two contrasting patterns correspond to the two cities under study and can be further elucidated by examining the spatial distribution of cases for a resource allocation characterized by $L \gtrless L_{\text{opt}}^*(\mathcal{R}_0)$.

In Fig. 3c, we consider that tests are allocated across the $L = 2.8 \times 10^4$ most critical links in Bogotá and depict the attack rate in each patch R_i^∞ , defined as the total number of recovered agents living in patch i at the end of the epidemic outbreak. This reveals a widespread penetration of the disease with varying degrees of impact across different areas. We also represent in Fig. 3d the spatial allocation of resources, that is, the proportion of tests conducted on individuals moving to each patch in this scenario. This spatial distribution indicates that the mobility-informed policies prioritize acting over human flows spanning multiple neighborhoods. An in-depth exploration of the dependence of both cases and test distribution on L is presented in Supplementary Fig. 5. There, it becomes evident that the optimal L_{opt}^* arises from the trade-off between the spatial allocation of resources required to address different contagion sources and the minimum number of tests needed for local outbreak control in each patch. In fact, when $L < L_{\text{opt}}^*(\mathcal{R}_0)$, the epidemic is contained in very specific locations but emerges elsewhere. Conversely, beyond the optimal value, i.e. when $L > L_{\text{opt}}^*(\mathcal{R}_0)$, the dispersion of tests across the urban landscape is such that preventing the spread in the main outbreak becomes totally unfeasible.

The epidemic scenario observed in Bogotá starkly contrasts with that of Miami for $L = 6 \times 10^3$, where the majority of both cases (Fig. 3e) and the allocated tests (Fig. 3f) concentrate around a single patch and its immediate

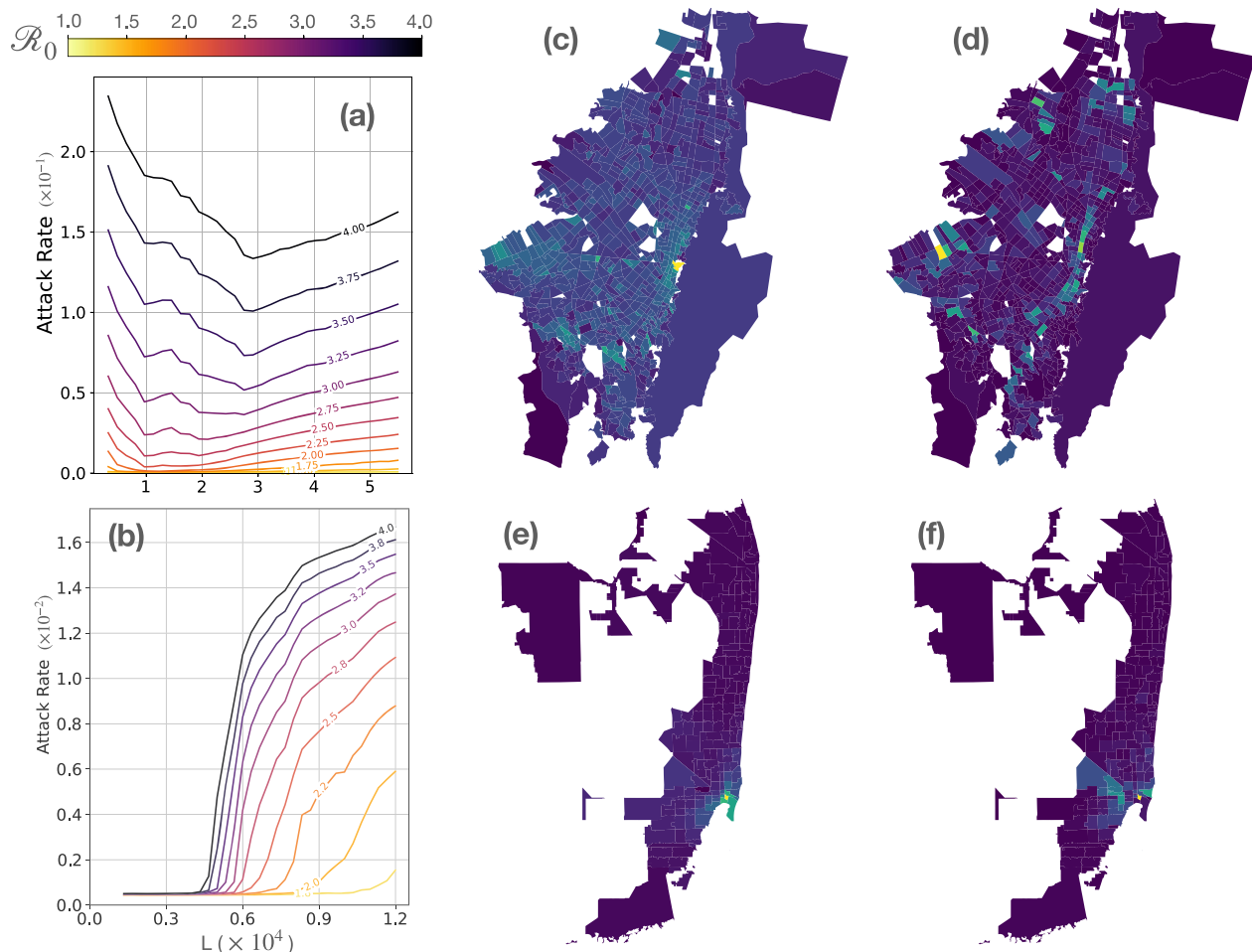


Fig. 3 | Urban landscapes and optimal distributions of tests across the mobility network. Panels **a** and **b** show the attack rate of an epidemic wave as a function of the basic reproduction number R_0 of the pathogen (color code) and the number L of mobility links used to distribute the $n_{\text{tests}} = 2 \times 10^5$ daily tests in Bogotá (**a**) and Miami (**b**). The central maps show the distribution of cases across Bogotá (**c**) and Miami (**e**) by plotting the attack rate in each patch after implementing a control strategy that

sets n_{tests} tests across the most important mobility links ($L = 2.8 \times 10^4$ for Bogotá and $L = 6 \times 10^3$ for Miami). Finally, the maps in the right show how the n_{tests} were distributed according to the destination of the aforementioned most critical links in Bogotá (**d**) and Miami (**f**). In all the maps, brighter colors correspond with higher values of the represented quantities.

vicinity. In this context, the aforementioned trade-off leans towards maximizing local outbreak control, thereby favoring strategies that adopt lower L values. Specifically, Supplementary Fig. 5 reveals that as tests become more evenly distributed (i.e., as L increases), the surge in attack rate shown in Fig. 3b is attributable to a significant outbreak in the most vulnerable area, which cannot be contained with the limited resources available there.

Mobility-informed surveillance policies

The results shown above have illustrated that the Mixing matrix $\mathbf{M}$ can guide the optimal allocation of tests to reduce the long-term extent of epidemic outbreaks. Regardless of the potential adequacy of the different control policies here proposed, their timing of implementation also represents a crucial factor shaping their ultimate benefits to reduce the epidemic burden on the population^{54–56}. Such timing is closely related to the design of efficient surveillance strategies, which should provide the authorities with an up-to-date picture of the evolution of the number of cases for the early diagnosis of ongoing epidemic crises.

We now explore whether the Mixing matrix $\mathbf{M}$ can inform surveillance campaigns for an efficient screening of the infected population in an epidemic outbreak. In what follows, we restrict our analysis to Bogotá and consider an epidemic outbreak with $R_0 = 4$. Our surveillance strategy distributes $n_{\text{tests}} = 2 \times 10^4$ tests across the L most critical mobility links identified in the matrix $\mathbf{M}$. Unlike the previous analysis, we assume an

uncontrolled scenario, meaning that infected individuals are no longer quarantined upon detection but, instead, their number is monitored for surveillance purposes as shown in the compartmental diagram in Fig. 4a. Thus, our baseline scenario (see Fig. 4b) consists in an uncontrolled epidemic wave in which contagions naturally propagate through the urban environment. This enables us to quantify the temporal evolution of positive cases as the epidemic unfolds without case detection having any effect on its evolution.

In Fig. 4c, we present the temporal evolution of the total number of detected cases for different allocation strategies of testing resources, regulated by L . These results indicate that concentrating testing resources on a limited number of key mobility links (indicated by a small L) enables surveillance systems to detect an impending epidemic more swiftly. In addition, the cumulative number of cases detected 30 days after the beginning of the outbreak decreases as tests are more evenly distributed across the city, for different values of R_0 and amounts of testing resources, n_{tests} (see Supplementary Fig. 6).

To more precisely measure this effect, we introduce the concept of the *Early Warning Time* (EWT), defined as the time needed to declare an epidemic scenario after detecting n_{alarm} infected individuals (indicated by the dashed line in Fig. 4c). The increasing trend of the EWT with the number of links, L , across which tests are allocated is evident in Fig. 4d–e. This observation aligns with the premise that mobility-informed

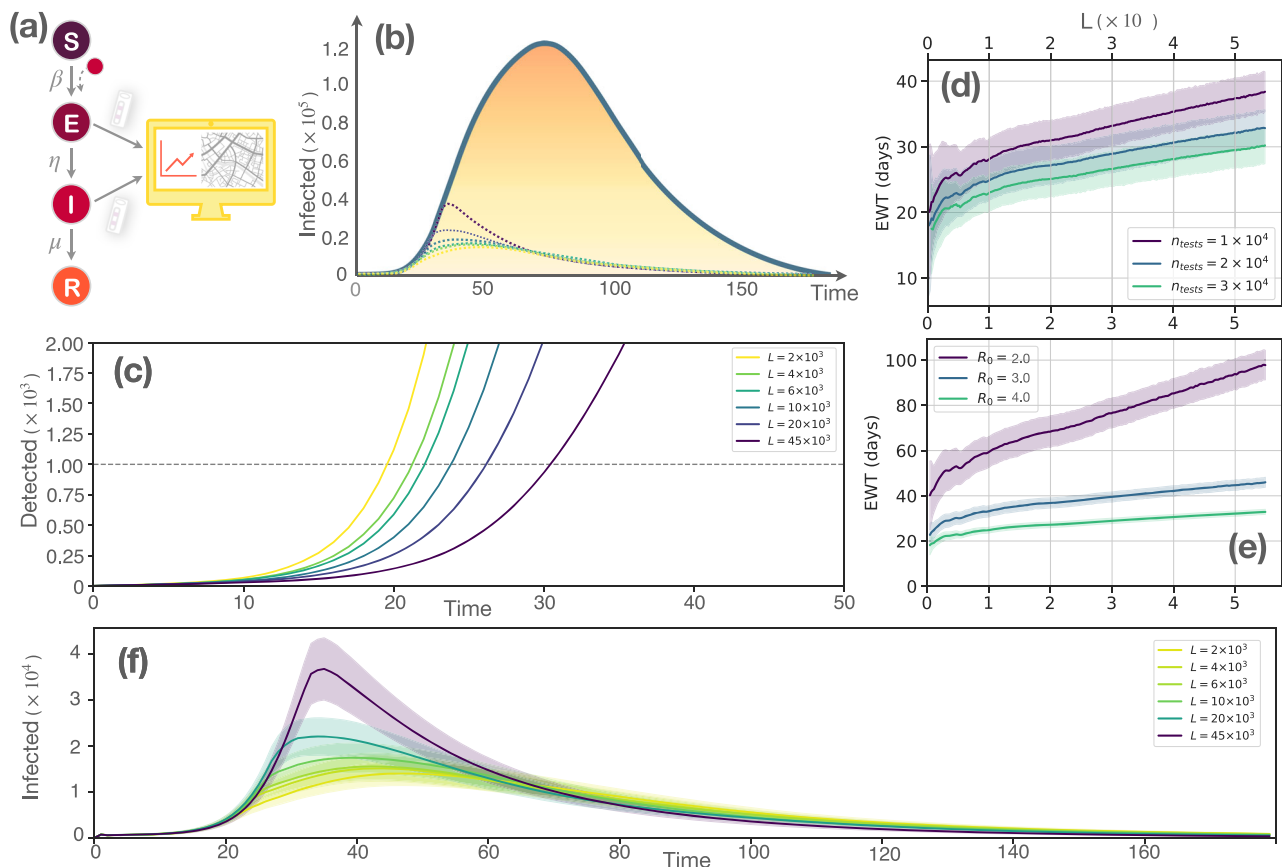


Fig. 4 | Enhancing epidemic response with targeted surveillance. **a** The SEIR model is augmented with a surveillance system that records positive tests from exposed and infectious individuals. **b** The solid line represents the natural (without interventions) epidemic curve. The dotted lines correspond to the epidemic curves shown in **f** (commented below). In panel **c** we show the time evolution of detected cases for different concentrations of testing resources on critical links. The dashed horizontal line signifies the threshold, n_{alarm} , which triggers containment actions. The Early Warning Time (EWT) - the interval spanned from the initial infections to

the time when n_{alarm} detections occur - is plotted against the degree of test concentration L for **d** various numbers of available tests, n_{tests} , and for **e** distinct basic reproduction numbers, R_0 . Panel **f** shows the mitigated epidemic curves upon the imposition of containment measures that reduce R_0 by 50% at time $t = \text{EWT}$ (these curves are also represented with dashed lines in panel **b** for illustrating the mitigation effect with respect to the uncontrolled scenario). The shaded areas in panels **d–f** represent the 95% confidence interval in the results obtained from 100 distinct simulations.

strategies enhance the preemptive detection of an epidemic by testing individuals in high-risk transmission flows. Moreover, these two panels corroborate the persistent benefit of such surveillance campaigns focused on the most critical human mobility links, irrespective of the volume of testing resources, n_{tests} (Fig. 4d) and the basic reproduction number, R_0 , of the spreading pathogen (Fig. 4e). This also holds for the other cities included in the study (Miami, New York, Boston and Austin) as shown in Supplementary Fig. 7.

We further delineate the implications of timely control strategies following the epidemic alarm raised by mobility-informed surveillance campaigns. To do so, we consider that, when n_{alarm} cases are detected, containment strategies are deployed, halving the reproduction number R_0 . Figure 4f shows the trajectories of the epidemic after the activation of containment measures at the EWT ascertained through the surveillance strategies shown in Fig. 4c. Significantly, in all instances, a marked decline in case numbers is observed, as indicated by the dashed lines in Fig. 4b (that replicate the evolution of the average values shown in Fig. 4f). Focusing on the differences among the scenarios presented in Fig. 4c, we note that an anticipation of roughly 10 days in the EWT not only postpones the epidemic peak but also substantially curtails peak incidence. A more granular analysis of the effects of the containment measures implemented at the EWT is presented in Supplementary Fig. 8, which showcases a significant reduction in the size of the epidemic peak and the attack rate when focusing the tests on the most vulnerable mobility links, as well as a temporal shift of the epidemic peak.

Mobility-informed surveillance at Transit Stations

So far, mobility-informed strategies have demonstrated effectiveness in early epidemic detection and in palliating their long-term burden on the population. However, the granular application of testing based on the anticipated knowledge of citizen's residences and workplaces is impractical during an actual epidemic due to privacy and logistical constraints. Consequently, other control strategies leveraging critical mobility data obtained from Mixing matrix **M**, while still actionable, are necessary. To this aim, here we seek for alternative targeted strategies based on interventions at the level of transportation systems.

In Bogotá, the *Transmilenio* bus rapid transit system accounted for 38% of urban travel in 2022. Its extensive network, shown in Fig. 5a, vertebrates the daily flows through 12 routes that feature 144 stations with varied geographic distribution. The system's electronic card check-ins and check-outs offer a digital trace of users' travel patterns. This feature, along with the massive use of the Transmilenio mobility network, facilitates the implementation of mobility-informed policies by allocating tests at strategic transport hubs.

To identify those stations where resources should be prioritized, we need to map the critical links identified in the mobility network, connecting two geographical patches, with critical journeys between two stations in the transportation network. To do so, we assume that the *Transmilenio* network is used by those individuals whose residence and workplace are located within a radius of 800 m of distance from their respective nearest stations, which are considered to be the origin and destination of the trip in the

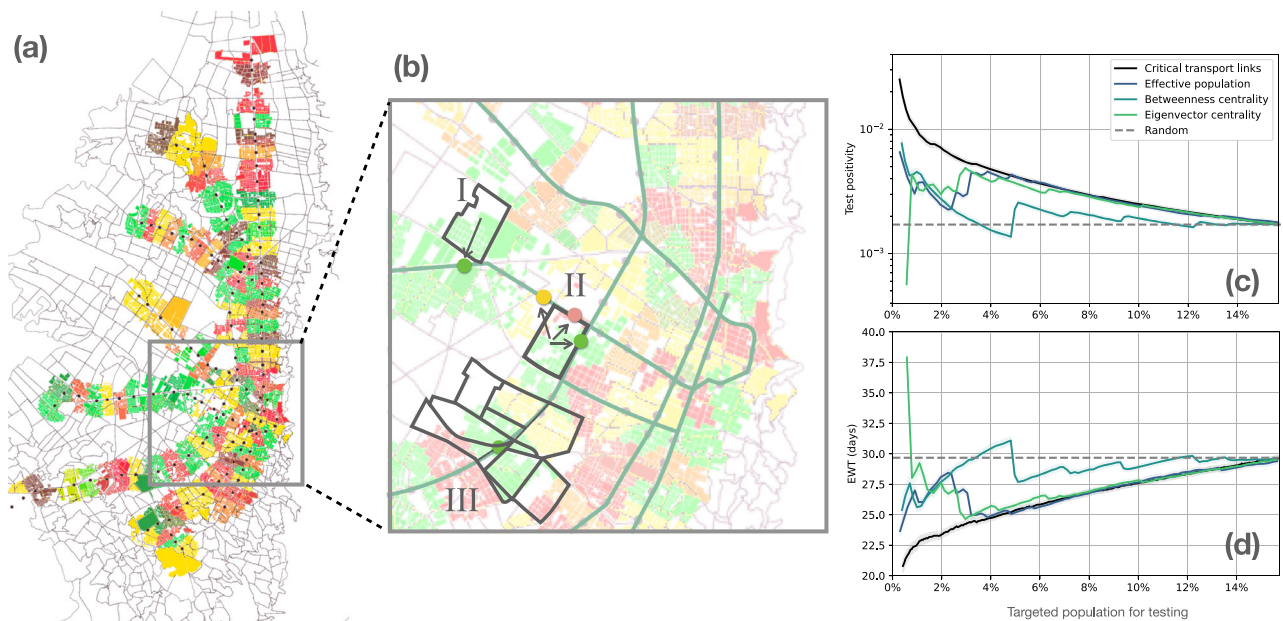


Fig. 5 | Transferring mobility-informed policies into massive transportation systems. **a** The spatial distribution of the Transmilenio stations in Bogotá (Colombia): colored areas correspond to the different patches whereas dots correspond to stations. The alignment of raw mobility flows with Transmilenio journeys is made by assigning (see Methods) the fraction of population of a patch making use of a particular station. Different situations of the assignment processes are highlighted in **b**: (I) all the population of a patch is assigned to the same station, (II) three stations serve a single patch, and (III) a single station serves the population of a collection of patches. Test positivity (**c**) and EWT (**d**) as a function on relative size of

the pool of public transport users targeted for testing (relative to the city's population) following 5 different surveillance strategies: prioritizing critical mobility links projected on the transportation network (black), the amount of daily users of each transport station (dark blue), the betweenness centrality of the stations (light blue), their eigenvector centralities (green) and randomly across all transportation users (gray dashed line). In these panels, solid lines represent the average quantities and the shadowed region constitute the 95% confidence interval for the results obtained for each scenario across 100 simulations.

transportation network. Overall, the aggregation process, sketched in Fig. 5b and further explained in the Methods section, preserves 15% of the human movements recorded in the mobility network.

Figure 5c–d present the comparison between different strategies for allocating tests among users of the public transportation network. To this aim we show how detected cases (Fig. 5c) and EWT (Fig. 5d) depend on the pool of the public transport users chosen for testing according to different criteria. The results, complemented with those presented in Supplementary Fig. 10, show the same qualitative behavior as before, where concentrating the tests on a few critical links leads to a much higher positivity rate (or detections) and lower epidemic warning time. Therefore, even when restrained to the lower resolution of the transportation network, targeted surveillance campaigns on vulnerable subgroups of individuals provide a better early assessment of an epidemic crisis. Note, however, that the loss of accuracy when projecting the mobility flows onto the transport network gives suboptimal results compared to the distribution of tests directly using the mobility network, as shown in Supplementary Fig. 11. This is expected, as the coarse-graining of the flows into subgroups of diverse exposure levels results in focusing tests on many individuals that are not at high risk. The information needed to move from number of links to percentage of population tested is represented in Supplementary Fig. 12.

To round off our analysis, we compare the performance of the control policies implemented on the links of the transportation networks with other simpler, more straightforward strategies focusing on transportation stations. These strategies encompass focusing the resources on the transport stations with the highest number of users, highest betweenness centrality and highest eigenvector centrality, as detailed in the Materials and Methods section. Such strategies, yet more effective than a complete random selection, do not achieve the same performance as the policies relying on the critical transportation links, thus showing the usefulness of the mixing matrix **M** for surveillance and testing strategies.

Discussion

In this work, we have explored the intersection of urban mobility and epidemic surveillance, harnessing the capacity of metapopulation frameworks to enrich epidemic models with a, often overlooked, layer of behavioral complexity. Our findings contribute to the emerging field of digital epidemiology³⁷, addressing the challenge of incorporating human behavioral data to construct realistic human mixing models that improve our preparedness and response to epidemic scenarios³⁵.

We have shown that urban mobility flows and demographic data can be combined into a Mixing matrix that captures the interactions of residents in urban environments driven by their daily recurrent mobility patterns. Our results underscore that the dominant eigenvector of this matrix has the potential to streamline testing and quarantine efforts by identifying critical pathways conducive to pathogen transmission. Thus, equipped with the particular mixing matrices of real cities, we have leveraged the knowledge of critical mobility links, identifying vulnerable subgroups of population with a given origin and destination, to design targeted mitigation measures (based on test and quarantine policies) and surveillance campaigns. In both scenarios, we have assumed limited testing resources, a common situation when facing sanitary crises in large urban environments, and studied how the allocation of testing resources to the most critical mobility links affects the outcomes of both processes.

For containment purposes, we have demonstrated that an optimal concentration of testing resources can significantly reduce the impact of an epidemic outbreak. However, the contrasting scenarios observed in cities like Bogotá and Miami illustrate that the optimal distribution of tests is contingent upon the specific socio-economic urban landscape. This insight is critical for urban planners and public health officials seeking to enhance the resilience of cities against future outbreaks.

In the realm of surveillance, we have shown that the concentration of testing resources in critical mobility links results in a considerable anticipation of the epidemic wave, allowing for prompt action even with limited resources. Specifically, the concept of Early Warning Time (EWT) emerges

as a key metric for assessing the timeliness of epidemic alerts, reinforcing that targeted testing of individuals within high-risk transmission flows can lead to more expedient interventions. We have also demonstrated that the proposed human behavior-informed surveillance campaigns are significant and actionable by showing that strategic testing at public transport stations - a more practical approach than widespread testing - can be an effective method of epidemic control.

Our findings should be considered in light of several limitations of the model here introduced. First, our model assumes that the population fully complies with the implemented policies. While this assumption greatly simplifies the simulations and the data requirements, it neglects the socio-economic factors leading to heterogeneous adherences to public health interventions^{57–59} or the pandemic fatigue effect^{60,61}, reducing the overall compliance of the population as a result of a prolonged implementation of control policies. Consequently, our model overestimates the expected impact of mobility-informed surveillance strategies on the public health of the population. In addition to this, our model does not account for the economic constraints or the logistic challenges^{62,63} associated with the spatial deployment of limited resources to monitor the advance of epidemic outbreaks. Additional economic metrics such as the cost-effectiveness of mobility-informed surveillance strategies would be thus needed to obtain a more informed assessment on their suitability to improve our preparedness and response to future epidemic outbreaks.

In conclusion, our research highlights the importance of strategically leveraging human behavior data to design efficient control policies when facing an epidemic crisis⁶⁴. While challenges in the implementation of such strategies persist, particularly regarding data privacy and ethical considerations, the advantages of enhancing epidemic intelligence through astute data analysis are clear. As globalization advances and urban populations continue to expand⁶⁵, the necessity for informed and multidisciplinary approaches such as the one presented here becomes increasingly vital, providing a shift from merely reacting to epidemic crisis to an integrative program that includes, among others, epidemiology and social sciences^{35,36}.

Methods

Construction of real metapopulations

Our framework represents cities as metapopulations whose nodes correspond to residence areas and their links encodes the mobility patterns of the population. The construction of a metapopulation then requires demographic data, capturing how residents are distributed across the city, and mobility data encoding their commuting patterns. For the cities here analyzed, we can construct their associated metapopulation by using publicly available data.

Demographic data for the city of Bogotá, Colombia, come from the 2018 National Census of Population and Housing⁶⁶. The public database provides household, residence, and individual data at the spatial resolution of *city blocks* (*manzanas*). Moreover, origin-destination data **T** for commuters is sourced from the 2018 Bogotá mobility survey⁶⁷ with the resolution of the *Transport Analysis Zones* (*Zonas de Análisis de Transporte* or *ZATs*). These zones are considerably larger in area than the *city blocks*. To match both data sources, we use available geometric spatial data to aggregate the populations of the *city blocks* within each *ZAT* and use the latter as the patches of our metapopulation model.

We also analyze the mixing matrix corresponding to different metropolitan statistical areas (also known as core based statistical areas or CBSAs) in the United States: Miami-Fort Lauderdale-West Palm Beach (referred to as Miami), Austin-Round Rock-San Marcos (referred to as Austin), Boston-Worcester-Providence (referred to as Boston) and New York-Newark-Jersey City (referred to as New York). We extract the population distribution in each city at the *census block* level from the 2010 US census⁶⁸. The census also provides mobility data on the Longitudinal Employer-Household Dynamics⁶⁹ (LEHD) database. Specifically, daily mobility patterns of the population can be extracted from LEHD Origin-Destination Employment Statistics⁷⁰ (LODES), which provides commuting data over the entire country, also at the level of *census blocks*. As census

blocks represent very small geographical areas, we aggregate both data sets to ZIP Code Tabulation Areas (ZCTAs) resolution, thus being able to provide a more coarse-grained description of the locations on which resources should be prioritized.

For both cities, the constructed metapopulations are fully characterized by the number of residents in each patch i , n_i and the mobility flows n_{ij} across patches, which can be constructed from the elements of the origin-destination matrix **T** as:

$$n_{ij} = n_i \frac{T_{ij}}{\sum_k T_{ik}}, \quad (1)$$

where T_{ij} is the number of survey respondents commuting from i to j . The main characteristic of the constructed metapopulation for each city under study can be found in Supplementary Table 1.

Agent-based simulations

All the epidemiological curves shown in the manuscript are computed through agent-based simulations combining mobility data n_{ij} and the epidemiological processes driving the evolution of epidemic outbreaks. These simulations allow tracking the epidemiological state of each individual of the population according to a SEIRQD (Susceptible-Exposed-Infected-Recovered-Quarantined-Detected) dynamics.

We use a discrete-time approach considering that each time step in the simulations corresponds to a day. For each time step, different processes are simulated:

- First, we distribute the population according to the mobility flows n_{ij} recorded in the metapopulation, giving rise to a new (temporary) spatial distribution in which each patch i is effectively populated by $n_i^{\text{eff}} = \sum_j n_{ji}$ agents.
- Then, day contacts are simulated at each patch. These contacts aim at capturing interactions occurring at workplaces, schools etc. We assume that all individuals who are concurrently in the same subpopulation make the same number of contacts. These contacts are proportional to the effective population density of the patch in which they are located. Therefore, an individual located at patch i makes $z_D f_i$ interactions with randomly chosen individuals within their temporary patch, where $f_i = n_i^{\text{eff}} / a_i$ captures the effective population density at patch i , being a_i its area. Likewise, z_D is a scaling factor to ensure that the average number of day contacts is a fixed value $\langle k_D \rangle$, i.e. $z_D = \langle k_D \rangle / \sum_i n_i^{\text{eff}} f_i$. This means that, while the number of interactions each agent makes is different depending on their origin and destination, their average values in the total population is fixed. Throughout the manuscript, we assume $\langle k_D \rangle = 8$ for all epidemic scenarios. To simulate contagions, we assume that a susceptible (S) individual becomes exposed (E) with a probability β for each contact with an infected (I) person.
- Then, we simulate local contagion processes within their neighborhood or household members. Those processes add a total number of $\langle k_H \rangle$ interactions to every susceptible individual, which are established with randomly chosen individuals within their residence patch. Throughout the manuscript, we assume $\langle k_H \rangle = 3$ for all epidemic scenarios. Unlike the previous case, this number of contacts does not vary across patches.
- The rest of compartments get updated according to their respective dynamics. Every exposed (E) individual has a probability η of turning infectious (I), and every infected (I) agent has a probability μ of becoming recovered (R).
- Finally, depending on the control strategy in question, a specific number of daily tests n_{tests} are randomly distributed every time step over agents belonging to the selected links in the mobility network, n_{ij} , or in the transportation network $n_{\alpha\gamma}$, where i and j are patches in the mobility network and α and γ represent stations in the transport network. When enforcing quarantine, every positive test will set the state of the individual to a different compartment, quarantined (Q), isolating them from the rest of the population.

The parameters used for all the simulations can be found in Supplementary Table 2 whereas the deterministic equations that describe the dynamics of the systems are contained in the Supplementary Information.

Construction of the mixing matrix $\mathbf{M}$

As explained in the main text, the Mixing matrix $\mathbf{M}$ simplifies the multitude of mobility-driven contagion processes into a single, mathematically manageable entity, whose dimension corresponds to the number of different subgroups in the metapopulation under study. In particular, we focus on the Mixing matrix $\mathbf{M}$ when $p_d = 1$, for it captures the essence of behavioral-driven surveillance in urban landscapes subjected to the baseline mobility scenario. Note that $p_d = 1$ does not imply that all the individuals move across the metapopulation, as the mobility network contains self-loops n_{ii} , capturing individuals who daily commute to a workplace located at their residential patch i .

To construct this matrix, we must take into account the two different types of contagion processes occurring at each time step in our model. First, we assume that individuals interact in their destination, being their number of contacts proportional to the effective population density inside that patch. In addition, we assume that individuals make $\langle k_H \rangle$ contacts with individuals from their household or neighborhood.

From these assumptions and as shown in the Supplementary Information, the element M_{ijkl} , encoding all the interactions of one individual whose residence (destination) is located in i (j) with the subgroup of residents in l traveling to k , is given by:

$$M_{ijkl} = (1 - p_d) z_D^2 \frac{z_{f_l}^D}{n_i^{\text{eff}}} n_{lk} \delta_{il} + \frac{\langle k_H \rangle}{n_i} n_{lk} \delta_{il} + (1 - p_d) p_d \left(\frac{z_{f_k}^D}{n_k^{\text{eff}}} n_{lk} \delta_{ik} + \frac{z_{f_l}^D}{n_i^{\text{eff}}} n_{lk} \delta_{jl} \right) + p_d^2 \frac{z_{f_k}^D}{n_k^{\text{eff}}} n_{lk} \delta_{jk} \quad (2)$$

where δ_{il} denotes the Kronecker delta and z^D ensures that an average number of $\langle k_D \rangle$ daily contacts are observed across all destinations, i.e. $z_D = \langle k_D \rangle / \sum_i f_i n_i^{\text{eff}}$.

Mobility-informed testing strategies and the Mixing matrix $\mathbf{M}$

The mixing matrix governs the evolution of the spatial distribution of infected individuals throughout the city at the initial stages of epidemic outbreaks, as proven in the Supplementary Information. For a Susceptible-Infected-Recovered (SIR) dynamics, defining e_{ij} as the fraction of population with residence in i and destination in j in the infected state, the former time evolution can be approximated as:

$$\frac{\mu}{\beta} e_{ij}(t) = \sum_{lk} M_{ijkl} e_{lk}(t-1), \quad (3)$$

where β and μ denote the transmission and recovery probabilities. Expressing the former equation in matrix form, the evolution of the epidemic state of the population is given by the following linear equation:

$$\frac{\mu}{\beta} \vec{e}(t) = \mathbf{M} \vec{e}(t-1). \quad (4)$$

The former expression neglects all nonlinear terms involved in contagions across the metapopulation, as it assumes a finite but negligible fraction of infected individuals across the metapopulation at early stages of the outbreak, i.e. $e_{ij}(t) \ll 1 \forall i, j$. This assumption is no longer valid when the epidemic prevalence is higher across the metapopulation; however, it allows us to envisage testing policies relying on the spectral policies of the Mixing matrix $\mathbf{M}$. In particular, its leading eigenvector $\vec{e}$ is expected to capture the spatial distribution of infected agents at the onset of an outbreak, thus providing an early estimate of the most vulnerable groups of population. As this eigenvector corresponds to individuals with different origins and

destinations, we will refer to its elements as links L (origin-destination pairs) of the metapopulation.

Spatial allocation of tests

Mobility-informed policies relying on the Mixing matrix $\mathbf{M}$ are shaped by two parameters: n_{tests} , determining the amount of resources available for testing the population, and L , dictating how many subgroups of individuals (links) are prone to be tested. Specifically, for each time step (day) in the agent-based simulations, n_{tests} are distributed randomly among the pool of population composing the most critical L links identified by analyzing the Mixing matrix $\mathbf{M}$. Specifically, following our analysis in the previous section, we assume that these links correspond to the L largest entries of the leading eigenvector of matrix $\mathbf{M}$. This methodology allows us to evaluate whether concentrating resources on the most critical links or spreading them out across the city constitutes an optimal strategy to face an emerging epidemic.

Critical mobility links on the public transport network

While quite useful for surveillance purposes and the long-term control of epidemic outbreaks, mobility-informed policies that rely on the Mixing matrix $\mathbf{M}$ are not feasible due to practical difficulties in identifying specific subgroups of the population. As explained in the main text, we can take advantage of the information provided by the Mixing matrix $\mathbf{M}$ and implement a more plausible testing policy by allocating strategically test in the transport stations, based on the passengers' entrance and exit. For this purpose, we need to project the information obtained from $\mathbf{M}$ to estimate the likelihood of a specific subgroup of users in the transportation network being infected. We focus our analysis on the city of Bogotá, for which we have extremely fine-grained demographic information from census data, the daily mobility patterns of the population from public surveys and the spatial distribution of the Transmilenio network, the Bus Rapid Transit network of the city.

To aggregate the mobility flows into the transport network, we use geographical location data of every Transmilenio station and assume people often use the closest stations to their origin and destination patches, in case they are located within a radius of 800 m from their associated city blocks. Therefore, we assign a set of weights $W_{i\alpha}$ to every patch i , representing the fraction of residents there using a nearby station α to travel to their usual destination. Note that these weights are indistinctively used to map stations with patches for both the origin and destination of each trip.

Oftentimes, every person from a patch i has the same nearest transport station α , i.e. $W_{i\alpha} = 1$. Nonetheless, there are also scenarios, as the one visualized in Fig. 5b, for which patches are surrounded by different nearby stations, which are chosen disparately by the residents across the city blocks of those areas. In that case, weights are assigned as the proportion of residents who choose either of the stations.

Assuming that there is no correlation between the origin and destination stations, the number of individuals traveling from node i to j using the transportation line from station α to γ is:

$$n_{i\alpha}^{j\gamma} = n_{ij} W_{i\alpha} W_{j\gamma}, \quad (5)$$

which works for the number of infected agents in the same way:

$$I_{i\alpha}^{j\gamma} = I_{ij} W_{i\alpha} W_{j\gamma}. \quad (6)$$

Adding the populations of all the nodes that use a particular pair of stations gives us the projection of the mobility network onto the transport network. Therefore, we can compute the fraction of the population infected at the onset of the outbreak and using the transportation route connecting stations α and γ , $\epsilon_{\alpha\gamma}$ as

$$\epsilon_{\alpha\gamma} = \frac{I_{\alpha\gamma}}{n_{\alpha\gamma}} = \frac{\sum_{ij} W_{i\alpha} W_{j\gamma} n_{ij} \epsilon_{ij}}{\sum_{ij} W_{i\alpha} W_{j\gamma} n_{ij}}, \quad (7)$$

where $n_{\alpha\gamma}$ ($I_{\alpha\gamma}$) represents the total number of (infectious) users with origin at station α and destination at γ . Note that the proposed projection loses some accuracy in the localization of critical mobility flows because it can group people from different links in the same transport line. Nonetheless, as argued above, it represents a realistic approach that is very useful in diagnosing and tackling the progression of an emerging epidemic wave. The results shown in the main text using this criterion were obtained in the same way as previously described: ranking the transport lines according to their risk $\varepsilon_{\alpha\gamma}$ and targeting the most vulnerable ones when distributing the tests.

Control strategies on transport stations

To highlight the advantages of the control policies devised from the mixing matrix, other public transport-based strategies are also considered and compared to the main selection criteria from the study. These use information from the transport network to assign a certain value of centrality using different metrics in order to rank the transport lines and distribute the tests accordingly.

The most straightforward metric is focusing on the transit of transport stations, n_{γ}^{eff} , which represents the total number of public transport users arriving in a transport station γ each day. Namely,

$$n_{\gamma}^{\text{eff}} \equiv \sum_{\alpha} n_{\alpha\gamma}. \quad (8)$$

This quantity is assumed to be equivalent for all users arriving in that station, i.e. $n_{\alpha\gamma}^{\text{eff}} \equiv n_{\gamma}^{\text{eff}}$.

Another strategy here considered is leveraging the betweenness centrality of the transport stations. The betweenness centrality of a station (node) γ can be calculated with the number of shortest paths, between any pair of nodes (α, δ) in the network) that go through said node ($\sigma_{\alpha\delta}(\gamma)$) in relation to the number of shortest paths ($\sigma_{\alpha\delta}$) that connect them. This can be expressed as

$$b_{\gamma} = \sum_{\alpha \neq \delta} \frac{\sigma_{\alpha\delta}(\gamma)}{\sigma_{\alpha\delta}}, \quad (9)$$

where the distance network used to estimate the shortest path between nodes is defined as $d_{\alpha\gamma} = 1/n_{\alpha\gamma}$. This allows us to rank the transportation lines by the betweenness centrality of their destination as $b_{\alpha\gamma} \equiv b_{\gamma}$.

Lastly, we also consider the eigenvector centrality p_{γ} of the different transportation stations γ in the transportation network, as it measures the transitive influence of nodes in a network and is used as an estimation of importance in many epidemic scenarios. Note that this eigenvector centrality is computed over the network whose weights correspond to $n_{\alpha\gamma}$, i.e. the number of transportation users with origin at station α and destination at γ . We can rank the different transportation lines according to the eigenvector centrality of their destination as $p_{\alpha\gamma} \equiv p_{\gamma}$.

Data availability

Data used to construct the metapopulation of Bogotá have been obtained from the national census of population and housing⁶⁶ and mobility surveys⁶⁷ whereas data for the US cities are extracted from the 2010 US census^{68–70}.

Code availability

The underlying code for this study is not publicly available but may be made available on reasonable request from the corresponding author.

Received: 1 July 2024; Accepted: 18 October 2024;

Published online: 04 December 2024

References

1. Diamond, J. *Guns, Germs, and Steel: The Fates of Human Societies* (W.W. Norton, 1997).
2. Snowden, F. *Epidemics and Society: From the Black Death to the Present* (Yale University Press, 2019).
3. McNeill, W. *Plagues and Peoples* (Anchor, 1976).
4. Domínguez-Andrés, J. et al. Evolution of cytokine production capacity in ancient and modern european populations. *eLife* **10**, e64971 (2021).
5. Alirol, E. et al. Urbanisation and infectious diseases in a globalised world. *Lancet Infect. Dis.* **11**, 131 (2011).
6. Lee, V. et al. Epidemic preparedness in urban settings: new challenges and opportunities. *Lancet Infect. Dis.* **20**, 527 (2020).
7. Brizuela, N., García-Chan, N., Gutiérrez-Pulido, H. & Chowell, G. Understanding the role of urban design in disease spreading. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **477**, 20200524 (2021).
8. Bilal, U. et al. Scaling of mortality in 742 metropolitan areas of the americas. *Sci. Adv.* **7**, eabl6325 (2021).
9. Kache, P. A. et al. Bridging landscape ecology and urban science to respond to the rising threat of mosquito-borne diseases. *Nat. Ecol. Evol.* **6**, 1601 (2022).
10. Baker, R., Mahmud, A. & Miller, I. et al. Infectious disease in an era of global change. *Nat. Rev. Microbiol.* **20**, 193 (2022).
11. Buckee, C., Noor, A. & Sattenspiel, L. Thinking clearly about social aspects of infectious disease transmission. *Nature* **595**, 205 (2021).
12. Althouse, B. M. et al. Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Sci.* **4**, 17 (2015).
13. Scarpino, S. V., Allard, A. & Hébert-Dufresne, L. The effect of a prudent adaptive behaviour on disease transmission. *Nat. Phys.* **12**, 1042 (2016).
14. Manfredi, P. & D'Onofrio, A. *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases* (Springer, 2015).
15. Watts, D., Muhamad, R., Medina, D. & Dodds, P. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl Acad. Sci. USA* **102**, 11157 (2005).
16. Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276 (2007).
17. Colizza, V. & Vespignani, A. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl Acad. Sci.* **104**, 12487 (2007).
18. Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *J. Theor. Biol.* **251**, 450 (2008).
19. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21484 (2009).
20. Belik, V., Geisel, T. & Brockmann, D. Epidemic spreading in metapopulation networks with heterogeneous connectivity patterns. *Chaos Solitons Fractals* **44**, 404 (2011).
21. Meloni, S. et al. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci. Rep.* **1**, 62 (2011).
22. Castioni, P., Gallotti, R. & De Domenico, M. Critical behavior in interdependent spatial spreading processes with distinct characteristic time scales. *Commun. Phys.* **4**, 131 (2021).
23. Soriano-Paños, D., Lotero, L., Arenas, A. & Gómez-Gardeñes, J. Spreading processes in multiplex metapopulations containing different mobility networks. *Phys. Rev. X* **8**, 031039 (2018).
24. Soriano-Paños, D. et al. Vector-borne epidemics driven by human mobility. *Phys. Rev. Res.* **2**, 013312 (2020).
25. Gatto, M. et al. Spread and dynamics of the covid-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Natl Acad. Sci. USA* **117**, 10484 (2020).
26. Bertuzzo, E. et al. The geography of covid-19 spread in Italy and implications for the relaxation of confinement measures. *Nat. Commun.* **11**, 1 (2020).
27. Arenas, A. et al. Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions. *Phys. Rev. X* **10**, 041055 (2020).
28. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl Acad. Sci.* **103**, 2015 (2006).

29. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337 (2013).
30. Zhang, Q. et al. Spread of zika virus in the americas. *Proc. Natl Acad. Sci. USA* **114**, E4334 (2017).
31. Bosetti, P. et al. Heterogeneity in social and epidemiological factors determines the risk of measles outbreaks. *PNAS* **117**, 30118 (2020).
32. Zhu, X. et al. Allocating resources for epidemic spreading on metapopulation networks. *Appl. Math. Comput.* **411**, 126531 (2021).
33. Reyna-Lara, A., Soriano-Paños, D., Arias-Castro, J., Martínez, H. & Gómez-Gardeñes, J. A metapopulation approach to identify targets for wolbachia-based dengue control. *Chaos* **32**, 041105 (2022).
34. Mazzoli, M., Gallotti, R., Privitera, F., Colet, P. & Ramasco, J. Spatial immunization to abate disease spreading in transportation hubs. *Nat. Commun.* **14**, 1448 (2023).
35. Bedson, J. et al. A review and agenda for integrated disease models including social and behavioural factors. *Nat. Hum. Behav.* **5**, 834 (2021).
36. Bedford, J. et al. A new twenty-first century science for effective epidemic response. *Nature* **575**, 130 (2019).
37. Salathé, M. Digital epidemiology: what is it, and where is it going? *Life Sci. Soc. policy* **14**, 1 (2018).
38. González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779 (2008).
39. Barbosa, H. et al. Human mobility: Models and applications. *Phys. Rep.* **734**, 1 (2018).
40. Jiang, S. et al. The timegeo modeling framework for urban motility without travel surveys. *Proc. Natl Acad. Sci. USA* **113**, E5370 (2016).
41. Bokányi, E., Juhász, S., Karsai, M. & Lengyel, B. Universal patterns of long-distance commuting and social assortativity in cities. *Sci. Rep.* **11**, 20829 (2021).
42. Balcan, D. & Vespignani, A. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat. Phys.* **7**, 581 (2011).
43. Belik, V., Geisel, T. & Brockmann, D. Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1**, 011001 (2011).
44. Apolloni, A., Poletto, C., Ramasco, J. J., Jensen, P. & Colizza, V. Commuting in metapopulation epidemic modeling. *Sci. Rep.* **4**, 4857 (2014).
45. Charaudeau, S., Pakdaman, K. & Boëlle, P.-Y. Commuter mobility and the spread of infectious diseases: Application to influenza in france. *PLoS One* **9**, e83002 (2014).
46. Gómez-Gardeñes, J., Soriano-Paños, D. & Arenas, A. Critical regimes driven by recurrent mobility patterns of reaction-diffusion processes in networks. *Nat. Phys.* **14**, 391 (2018).
47. Rader, B. et al. Crowding and the shape of covid-19 epidemics. *Nat. Med.* **26**, 1829 (2020).
48. Hazarie, S., Soriano-Paños, D., Arenas, A., Gómez-Gardeñes, J. & Ghoshal, G. Interplay between population density and mobility in determining the spread of epidemics in cities. *Commun. Phys.* **4**, 191 (2021).
49. Aguilar, J. et al. Impact of urban structure on infectious disease spreading. *Sci. Rep.* **12**, 3816 (2022).
50. Valgañón, P., Soriano-Paños, D., Arenas, A. & Gómez-Gardeñes, J. Contagion-diffusion processes with recurrent mobility patterns of distinguishable agents. *Chaos* **32**, 043102 (2022).
51. Torres, L., Chan, K. S., Tong, H. & Eliassi-Rad, T. Nonbacktracking eigenvalues under node removal: X-centrality and targeted immunization. *SIAM J. Math. Data Sci.* **3**, 656 (2021).
52. Roth, C., Kang, S., Batty, M. & Barthélemy, M. Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLOS ONE* **6**, e15923 (2011).
53. Bassolas, A. et al. Hierarchical organization of urban mobility and its connection with city livability. *Nat. Commun.* **10**, 4817 (2019).
54. Oraby, T. et al. Modeling the effect of lockdown timing as a covid-19 control measure in countries with differing social contacts. *Sci. Rep.* **11**, 3354 (2021).
55. Steinegger, B. et al. Joint analysis of the epidemic evolution and human mobility during the first wave of covid-19 in spain: Retrospective study. *JMIR Public Health Surveill.* **9**, e40514 (2023).
56. Morris, D. H., Rossine, F. W., Plotkin, J. B. & Levin, S. A. Optimal, near-optimal, and robust epidemic control. *Commun. Phys.* **4**, 78 (2021).
57. Valgañón, P., Useche, A. F., Soriano-Paños, D., Ghoshal, G. & Gómez-Gardeñes, J. Quantifying the heterogeneous impact of lockdown policies on different socioeconomic classes during the first covid-19 wave in colombia. *Sci. Rep.* **13**, 16481 (2023).
58. Mamelund, S.-E., Dimka, J. & Bakkeli, N. Z. Social disparities in adopting non-pharmaceutical interventions during covid-19 in norway. *J. Dev. Societies* **37**, 302 (2021).
59. Napoli, L., Sekara, V., García-Herranz, M. & Karsai, M. Socioeconomic reorganization of communication and mobility networks in response to external shocks. *Proc. Natl Acad. Sci.* **120**, e2305285120 (2023).
60. Delussu, F., Tizzoni, M. & Gauvin, L. Evidence of pandemic fatigue associated with stricter tiered covid-19 restrictions. *PLOS Digital Health* **1**, e0000035 (2022).
61. Lilleholt, L., Zettler, I., Betsch, C. & Böhm, R. Development and validation of the pandemic fatigue scale. *Nat. Commun.* **14**, 6352 (2023).
62. Dasaklis, T. K., Pappis, C. P. & Rachaniotis, N. P. Epidemics control and logistics operations: A review. *Int. J. Prod. Econ.* **139**, 393 (2012).
63. Mandyata, C. B., Olowski, L. K. & Mutale, W. Challenges of implementing the integrated disease surveillance and response strategy in zambia: a health worker perspective. *BMC Public Health* **17**, 1 (2017).
64. Bubar, K. M. et al. Model-informed covid-19 vaccine prioritization strategies by age and serostatus. *Science* **371**, 916 (2021).
65. Baker, R. E. et al. Infectious disease in an era of global change. *Nat. Rev. Microbiol.* **20**, 193 (2022).
66. Colombia's 2018 National Census of Population and Housing, <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018> accessed: 20 September 2023 (2018).
67. Encuesta de Movilidad de Bogotá 2018, https://datosabiertos.bogota.gov.co/dataset/encuesta_movilidad_bogota accessed: 4 August 2020 (2018).
68. TIGER/Line Shapefiles and TIGER/Line Files, <https://www2.census.gov/geo/tiger/TIGER2010BLKPOPHU/> accessed: 30 September 2021 (2010).
69. United States Longitudinal Employer-Household Dynamics (LEHD) Data, <https://lehd.ces.census.gov/data/> accessed: 30 September 2021 (2021).
70. United States mobility data sources are available at: <https://lehd.ces.census.gov/data/lodes/LODES7/> (Accessed: 2021-09-30).

Acknowledgements

P.V. and J.G.G. acknowledge financial support from the Departamento de Industria e Innovación del Gobierno de Aragón y Fondo Social Europeo (FENOL group grant E36-23R), and from Ministerio de Ciencia e Innovación through projects PID2023-147734NB-I00 and PID2020-113582GB-I00/AEI/10.13039/501100011033. A.F.U acknowledges funding support from the Department of Industrial Engineering, at Universidad de los Andes, Colombia through the doctoral training program. A.A. and D.S.P acknowledge support from Spanish Ministerio de Ciencia e Innovación (PID2021-128005NB-C21), Generalitat de Catalunya (2021SGR-00633) and Universitat Rovira i Virgili (2023PFR-URV-00633), the European Union's Horizon Europe Programme under the CREXDATA project. D.S.P. acknowledges the financial support of the Calouste Gulbenkian Foundation through the PONTE program and from Ministerio de Ciencia e Innovación through the Juan de La Cierva program through grant JDC2022-048339-I. A.A. acknowledges the Joint Appointment Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830, grant agreement no. 101092749, ICREA Academia, and the James S. McDonnell Foundation (Grant N.

220020325). FM acknowledges the financial support of the School of Engineering at Universidad de los Andes.

Author contributions

P.V., A.A., D.S.-P. and J.G.-G. conceived the study. A.F.U. and F.M. provided data. P.V. developed the methods and performed the simulations. All the authors contributed to the analysis of the results and the writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44260-024-00021-z>.

Correspondence and requests for materials should be addressed to A. Arenas, D. Soriano-Paños or J. Gómez-Gardeñes.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

Capítulo 6

Conclusiones

En este último capítulo recordamos las razones principales que han motivado la tesis, cómo encaja lo que hemos aprendido en el marco de la modelización epidémica junto con la elaboración de políticas de control y, finalmente, se detalla el estado actual y futuro de esta investigación.

Este trabajo y su evolución, al igual que el de cualquier otro, no se encuentran aislados en un vacío. Por el contrario, surgen y se transforman dentro del contexto de los eventos y avances que los precedieron. No parece una coincidencia que, durante la década de 2010, la epidemiología matemática experimentara un notable auge. Las recurrentes apariciones de enfermedades contagiosas, que provocaban brotes epidémicos y, en muchos casos, se mantenían como endémicas en ciertas poblaciones, actuaron como señales tempranas de la vulnerabilidad de un sistema vulnerable e insuficientemente preparado para enfrentar grandes crisis sanitarias como la acaecida poco tiempo después con la llegada del SARS-CoV-2, causante de la pandemia de COVID-19. A pesar de los esfuerzos de científicos, divulgadores e instituciones de atraer la atención a los riesgos que conlleva la carencia de infraestructura y de inversión en mejorar las condiciones higiénicas y sanitarias de poblaciones vulnerables [119–121], este coronavirus tomó desprevenidos gran parte del mundo. Al fin y al cabo, un porcentaje sustancial de estas epidemias se limitan a países en desarrollo donde no existen los recursos para frenarlas y, por supuesto, donde las consecuencias de dichos desastres no afectan a los más privilegiados. En este contexto, y dadas las secuelas dejadas por la pandemia de COVID-19, podemos encontrar suficiente motivación para afrontar los objetivos que hemos explicado con anterioridad.

La gran cantidad de información recolectada durante el primer periodo de confinamiento y el altruismo emergente en torno a la actividad investigadora, fomentó la publicación de grandes bases de datos por parte de entidades públicas y pri-

vadas. Esta actitud de colaboración y apertura informativa ha permitido abordar el primer objetivo de esta tesis: analizar las medidas de contención y su impacto tanto en los patrones de movilidad de las personas como en la contención del brote, considerando factores socioeconómicos. En particular, en el capítulo 2 se ha puesto a prueba un modelo compartimental simple, ya que utiliza una aproximación de campo medio, que tiene en cuenta los diferentes estados que atraviesan la mayor parte de pacientes de COVID-19. La cualidad más relevante del modelo es, sin embargo, su forma de integrar la respuesta de la población al confinamiento y su efecto en la propagación. Por un lado, se asume que un porcentaje equivalente de personas, equivalente a la reducción de movilidad observada en los datos empíricos, lleva a cabo la cuarentena y puede interactuar únicamente con otros miembros de su vivienda. Al mismo tiempo, se tiene en cuenta que cierta proporción de esas personas no es capaz de seguir el confinamiento estrictamente, por lo que a pesar de reducir sus interacciones, no se aíslan por completo. A este grupo se le atribuye la variable de permeabilidad social, gracias a la cual hemos podido replicar la curva de fallecidos de cuarenta y dos países. Además de permitirnos extraer información útil estudiando diferentes parámetros obtenidos en la calibración de todos los países, los resultados proveen para cada población un valor de dicha permeabilidad, permitiendo cuantificar lo inefectivas que han sido las restricciones en el mismo. El análisis de esta nueva métrica nos confirma que la riqueza de cada país está fuertemente correlacionada con su capacidad de frenar el crecimiento del brote, incluso para reducciones menores en la movilidad general de la población. La razón de estas diferencias puede asociarse a diversas causas, tales como la necesidad de infraestructura adecuada, falta de confianza en las autoridades políticas o, especialmente, la necesidad de muchas personas de mantener su actividad económica para continuar sus fuentes de ingresos. De cualquier forma, dado que se observa esta diferencia a nivel de países, es natural preguntarse hasta qué punto esto ocurre también a nivel de individuos.

En el capítulo 3 se aumenta la resolución hasta el nivel urbano (haciendo uso de datos obtenidos en las ciudades colombianas de Bogotá, Medellín y Cali), analizando los patrones de movilidad de individuos según su clase socioeconómica y cómo ésta cambia radicalmente en el momento en el que se lleva a cabo el confinamiento. A pesar de contar con datos mucho más precisos que en el capítulo anterior, llegamos a la misma conclusión: las personas de estratos económicos más altos tienen una mayor capacidad para reducir su actividad, confirmando, a nivel individual, las conclusiones previamente alcanzadas. La disponibilidad de datos origen-destino permitió, además, un análisis más detallado de cómo los patrones de viaje de los miembros de cada estrato se modificaron. El estudio revela que los

individuos de estratos más altos, aunque reducen significativamente la cantidad de viajes, mantienen en gran medida sus destinos habituales. En contraste, los de estratos más bajos experimentan mayores alteraciones en sus patrones, optando por destinos más cercanos y asociados a estratos económicos similares o inferiores al propio. Estos hallazgos evidencian el riesgo, ya señalado, de que ciertas políticas de confinamiento generalizado aumenten la segregación y perjudiquen desproporcionadamente a sectores específicos de la población.

Tras lo expuesto previamente, el segundo y principal objetivo de la tesis ha sido introducir una mejora significativa en los modelos epidémicos metapoblacionales para elaborar estrategias de control efectivas que sean capaces de alertar tempranamente de brotes o frenarlos con el mínimo impacto posible en la población. En el capítulo 4, se aborda uno de los principales desafíos de los modelos de movilidad utilizados para estudiar y predecir dinámicas epidémicas en entornos urbanos y con escalas temporales reducidas. Al igual que los modelos que reproducen los viajes diarios (*commuting*) consideran la recurrencia en los orígenes de los trayectos, estos suelen ignorar la recurrencia en los destinos, contradiciendo las fuentes de encuestas y censos que proporcionan datos individuales de origen-destino. Los resultados han demostrado que este fenómeno es responsable de una notable disminución del umbral epidémico en numerosos casos, lo que facilita el crecimiento exponencial de enfermedades que, en otras condiciones, no serían lo suficientemente infecciosas como para desencadenar una epidemia. Concluimos que la capacidad del sistema para entrar en un estado epidémico depende, en gran medida, de los flujos de movilidad hacia el foco principal de la metapoblación y, asimismo, del tiempo que los individuos infectados permanecen en ese foco antes de recuperarse.

Los hallazgos descrito en el capítulo 4 abrieron un nuevo camino a la hora de abordar el contagio en paisajes urbanos, donde la recurrencia de los viajes juega un papel fundamental. Por ello, en el capítulo 5, hemos hecho uso de la información disponible sobre los viajes de las personas para determinar qué flujos son críticos para la existencia de una fase epidémica y se han puesto a prueba diferentes medidas de contención que los prioricen sobre el resto. En este trabajo, a partir de los datos demográficos y de movilidad, se calcula para cada ciudad su matriz de mezcla, que está asociada a los contactos que ocurren entre todas las parejas de individuos a lo largo del día. Esta matriz provee información esencial con respecto a los comportamientos que conllevan más probabilidad de resultar en contagios, por lo que se obtiene una clasificación de los flujos más críticos. En la primera parte del artículo, se ha puesto a prueba esta métrica distribuyendo una cantidad diaria arbitraria de tests en los flujos más importantes e imponiendo el confina-

miento de aquellos individuos que resultan ser positivos. Se puede concluir que, efectivamente, la priorización de los individuos más vulnerables resulta en una prevalencia menor, por lo que estas medidas son beneficiosas. Sin embargo, también se ha observado que no todas las ciudades se comportan de forma idéntica y que el diseño de la localización de recursos de vigilancia, como pruebas de diagnóstico rápido, depende la fisonomía de los flujos humanos y la distribución demográfica específica. Así, para ciertas estructuras urbanas, la excesiva discriminación en el reparto de tests puede tener un efecto negativo, existiendo un punto óptimo que se sitúa entre la extrema localización de los recursos y su distribución aleatoria. Adicionalmente, e ignorando la posibilidad de confinar a las personas positivas, se comprobó que, en todos los casos estudiados, la positividad de los tests y el número de individuos detectados aumentan notablemente, permitiendo detectar brotes epidémicos con semanas de antelación. Por último, se ha abordado la dificultad práctica de clasificar a cada viajero según su lugar exacto de residencia o destino. Para superar este desafío, se proyectó la red de movilidad sobre la de transporte urbano. Los resultados muestran que, aunque se pierde cierta cantidad de información en el proceso, las estaciones de transporte pueden ser herramientas muy útiles para distribuir los recursos de forma óptima. Este enfoque, sin embargo, deja un vacío evidente: aunque se ha utilizado la red de transporte público para distribuir las pruebas, no se han considerado los contactos e infecciones que pueden ocurrir dentro de la misma. Este vacío ha abierto un nuevo camino de investigación, actualmente en desarrollo, que ya está produciendo resultados prometedores al integrar los contagios durante los viajes. Esto permitirá el desarrollo de modelos mucho más realistas en las ciudades donde se disponga de datos de transporte.

Por último, un proyecto que no ha sido incluido en la tesis por no encontrarse publicado al tiempo de su escritura examina una cuestión fundamental de los modelos desarrollados en esta tesis. En particular, en [122] nos planteamos cuáles son las consecuencias de elegir formalismos de movilidad aleatoria frente a la recurrente a la hora de simular epidemias. A partir del formalismo MIR hemos construido una descripción nueva que es capaz de fusionar los patrones de movilidad euleriana y lagrangiana en una misma dinámica. Al introducir la variable de la probabilidad de retorno r , se simula la propagación de epidemias en agentes que llevan a cabo una caminata aleatoria con reinicio, de forma que entre los extremos con r igual a cero (caminatas aleatorias) y uno (viajes recurrentes), se puede estudiar todo un rango de comportamientos que tienen un gran impacto en fenómenos como el umbral epidémico y el de invasión. Los resultados corroboran que, como se ha sugerido a lo largo de la tesis, los principales determinantes de

los umbrales son las escalas temporales. En concreto, el umbral epidémico se eleva cuando los agentes infecciosos no pueden invertir mucho tiempo en los parches más densos, algo que ocurre frecuentemente cuando domina la movilidad aleatoria. En casos de movilidad muy reducida, sin embargo, la probabilidad de retorno juega el papel opuesto, ya que impide que los agentes invasores permanezcan en su destino por mucho tiempo. El trabajo detalla de esta forma las implicaciones que tiene la recurrencia en los viajes, acelerando la propagación de la epidemia en ciertos regímenes críticos pero frenándola en casos de movilidad muy reducida.

Finalmente, es importante señalar algunos aspectos que no se han priorizado como objetivos en este trabajo y que, por tanto, han sido tratados de forma limitada. Los modelos matemáticos presentados no buscan ser reproducciones perfectas de la dinámica de propagación epidémica en poblaciones reales, ya que esto requeriría considerar numerosos factores adicionales y, seguramente, imposibilitaría su análisis matemático. Entre estos factores, a nivel individual se incluyen las distribuciones de edad y sexo, con sus respectivos parámetros sobre infectividad, asintomaticidad y fatalidad. También sería relevante poder incorporar políticas de mitigación como el uso de medidas profilácticas (mascarillas, vacunas), las redes de contactos sociales, y las interacciones entre diferentes enfermedades o cepas. Además, los modelos asumen ciertas simplificaciones que no son completamente realistas. Por ejemplo, se considera que los agentes que no viajan mantienen los mismos patrones de contacto en sus áreas de residencia, lo cual podría no reflejar con precisión la realidad.

Estas simplificaciones, no obstante, son deliberadas, ya que el objetivo principal del trabajo no era alcanzar el máximo nivel de realismo, sino avanzar en la comprensión del impacto de la movilidad humana recurrente en la propagación epidémica. Se han excluido niveles adicionales de complejidad para evitar que interfieran con el análisis exclusivo de las redes de movilidad. Por ejemplo, al introducir el formalismo de agentes distinguibles en el capítulo 4, se asume que la reducción de movilidad no implica cuarentena o aislamiento completo, sino una modificación en los patrones de contacto: los individuos interactúan en sus nodos de origen en lugar de los de destino. Esto asegura que los fenómenos observados y las conclusiones obtenidas reflejen exclusivamente el efecto de la complejidad de las redes de movilidad. Finalmente, esta descripción distinguible de los agentes puede integrarse en la mayoría de los modelos epidémicos que operan en redes de movilidad multiescala, no limitándose únicamente al formalismo MIR.

Con este capítulo concluye el cuerpo principal de esta tesis, en la que creemos haber logrado una caracterización significativa de la influencia de la movilidad humana en la difusión y el control de epidemias. A través del uso de datos empíricos en la primera parte y del desarrollo de marcos teóricos metapoblacionales en la segunda, hemos abordado el problema desde perspectivas complementarias, pero con un objetivo común. Sin embargo, replicar fenómenos sociales sigue siendo un desafío considerable, especialmente debido a las dificultades para contrastar empíricamente los resultados. En el ámbito de la epidemiología teórica, donde el propósito es informar y apoyar a las autoridades, resulta esencial que las variables y ajustes de los modelos estén sólidamente fundamentados en observaciones. Así como la corrección propuesta en los modelos metapoblacionales encuentra respaldo en los patrones de movilidad humana observados, las dinámicas derivadas de estos modelos deben ser validadas utilizando todas las fuentes de datos disponibles. La rápida digitalización de nuestro entorno, junto con el esfuerzo de entidades públicas y privadas por publicar información durante brotes recientes de enfermedades como el coronavirus o la gripe, abre la puerta a un futuro donde los modelos complejos orientados a los datos sean cada vez más viables. No obstante, este avance también plantea nuevos retos: al aumentar la capacidad computacional y la complejidad de los modelos, es crucial garantizar su capacidad de generalización, evitando el riesgo de sobreajustar sin un soporte empírico sólido. Queda, por tanto, mucho trabajo por hacer. En el futuro, esperamos incorporar la distinguibilidad de los agentes, los contagios en el transporte urbano y otras mejoras a formalismos que ya han sido probados ampliamente. Estas líneas de investigación no solo contribuirán a mejorar la precisión de los modelos, sino que también ofrecerán herramientas más robustas para afrontar las complejidades de las epidemias en un mundo cada vez más interconectado.

Apéndice

A

En los cuatro artículos incluidos en la tesis por compendio, el doctorando es primer autor, y ha sido responsable del tratamiento de los datos, desarrollo de los modelos, computación y representación de los resultados, y junto con otros coautores ha sido parte de la concepción y escritura de los mismos .

Revista	Factor de Impacto (JIF)	Área	Cuartil
Chaos: An Interdisciplinary Journal of Nonlinear Science	2.9	Physics, Mathematical	Q1
Frontiers in Public Health	3.0	Public, Environmental & Occupational Health	Q2
Scientific Reports	3.8	Multidisciplinary Sciences	Q1
npj Complexity	(aún no designado)	Multidisciplinary Sciences	

Bibliografia

- [1] P. C. Grammaticos, A. Diamantis, *et al.*, “Useful known and unknown views of the father of modern medicine, hippocrates and his teacher democritus,” *Hell J Nucl Med*, vol. 11, no. 1, pp. 2–4, 2008.
- [2] G. Kayali, “The forgotten history of pre-modern epidemiology: contribution of ibn an-nafis in the islamic golden era,” *Eastern Mediterranean Health Journal*, vol. 23, no. 12, pp. 854–857, 2017.
- [3] U. Kutschera, “Antonie van leeuwenhoek (1632–1723): Master of fleas and father of microbiology,” *Microorganisms*, vol. 11, no. 8, 2023.
- [4] N. Lane, “The unseen world: reflections on leeuwenhoek (1677) ‘concerning little animals’,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1666, p. 20140344, 2015.
- [5] J. Snow, *On the Mode of Communication of Cholera*. 1855.
- [6] T. H. Tulchinsky, “Chapter 5 - john snow, cholera, the broad street pump; waterborne diseases then and now,” pp. 77–99, 2018.
- [7] D. Cameron and I. G. Jones, “John snow, the broad street pump and modern epidemiology,” *International Journal of Epidemiology*, vol. 12, pp. 393–396, 01 1983.
- [8] L. Pasteur, *Mémoire sur les corpuscles organisés qui existent dans l’atmosphère, examen de la doctrine des générations spontanées*. Masson, 1861.
- [9] K. A. Smith, “Louis pasteur, the father of immunology?,” *Frontiers in Immunology*, vol. 3, 2012.
- [10] E. Metchnikoff, R. Koch, J. Lister, L. Pasteur, and D. Berger, *The Founders of Modern Medicine: Pasteur, Koch, Lister*. Walden publications, 1939.
- [11] N. Bacaër, *Ross and malaria (1911)*, pp. 65–69. London: Springer London, 2011.

- [12] R. Ross, *The Prevention of malaria*. J. Murray, 1910.
- [13] R. Ross, “Some quantitative studies in epidemiology,” *Nature*, vol. 87, no. 2188, pp. 466–467, 1911.
- [14] D. L. Smith, K. E. Battle, S. I. Hay, C. M. Barker, T. W. Scott, and F. E. McKenzie, “Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens,” *PLOS Pathogens*, vol. 8, pp. 1–13, 04 2012.
- [15] H. H. Weiss, “The sir model and the foundations of public health,” *Materials matematics*, pp. 0001–17, 2013.
- [16] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, 1927.
- [17] W. O. Kermack and A. G. McKendrick, “Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity,” *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, vol. 138, no. 834, pp. 55–83, 1932.
- [18] W. O. Kermack and A. G. McKendrick, “Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity,” *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, vol. 141, no. 843, pp. 94–122, 1933.
- [19] W. Kermack and A. McKendrick, “Contributions to the mathematical theory of epidemics iv. analysis of experimental epidemics of the virus disease mouse ectromelia,” *Epidemiology & Infection*, vol. 37, no. 2, pp. 172–187, 1937.
- [20] W. Kermack and A. McKendrick, “Contributions to the mathematical theory of epidemics: V. analysis of experimental epidemics of mouse-typhoid; a bacterial disease conferring incomplete immunity,” *Epidemiology & Infection*, vol. 39, no. 3, pp. 271–288, 1939.
- [21] R. M. Anderson, “Discussion: The kermack-mckendrick epidemic threshold theorem,” *Bulletin of Mathematical Biology*, vol. 53, no. 1, pp. 3–32, 1991.
- [22] H. Abbey, “An examination of the reed-frost theory of epidemics,” *Human Biology*, vol. 24, no. 3, pp. 201–233, 1952.

- [23] D. G. Kendall, “Mathematical models of the spread of infection,” *Mathematics and computer science in biology and medicine*, pp. 213–225, 1965.
- [24] D. Mollison, *The Rate of Spatial Propagation of Simple Epidemic*, pp. 579–614. Berkeley: University of California Press, 1972.
- [25] Y. Hosono and B. Ilyas, “Traveling waves for a simple diffusive epidemic model,” *Mathematical Models and Methods in Applied Sciences*, vol. 05, no. 07, pp. 935–966, 1995.
- [26] A. Winn, A. Konkol, and E. Katifori, “From localized to well mixed: How commuter interactions shape disease spread,” *Physical Review E*, vol. 108, no. 4, p. 044305, 2023.
- [27] S. Riley, “Large-scale spatial-transmission models of infectious disease,” *Science*, vol. 316, no. 5829, pp. 1298–1301, 2007.
- [28] L. Rass and J. Radcliffe, *Spatial deterministic epidemics*. American Mathematical Soc., 2003.
- [29] H. W. Hethcote, “The mathematics of infectious diseases,” *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.
- [30] C. W. Reynolds, “Flocks, herds and schools: A distributed behavioral model,” in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 25–34, 1987.
- [31] T. C. Schelling, “Dynamic models of segregation,” *The Journal of Mathematical Sociology*, vol. 1, no. 2, pp. 143–186, 1971.
- [32] C. M. Glen, M. L. Kemp, and E. O. Voit, “Agent-based modeling of morphogenetic systems: Advantages and challenges,” *PLoS computational biology*, vol. 15, no. 3, p. e1006577, 2019.
- [33] R. Beckman, C. Barrett, K. Berkgigler, K. Burris, B. Bush, S. Hull, J. Hurford, P. Medvick, D. Kubicek, M. Marathe, J. Morgeson, K. Nagel, D. Roberts, L. Smith, M. Stein, P. Stretz, S. Sydoriak, K. Cervenka, and R. Donnelly, “Transportation analysis simulation system (transims) - the dallas-ft. worth case study,” 12 2002.
- [34] R. J. Beckman, K. A. Baggerly, and M. D. McKay, “Creating synthetic baseline populations,” *Transportation Research Part A: Policy and Practice*, vol. 30, no. 6, pp. 415–429, 1996.

- [35] P. W. Anderson, “More is different,” *Science*, vol. 177, no. 4047, pp. 393–396, 1972.
- [36] H. H. Heng, “The conflict between complex systems and reductionism,” *Jama*, vol. 300, no. 13, pp. 1580–1581, 2008.
- [37] A. Ma’ayan, “Complex systems biology,” *Journal of The Royal Society Interface*, vol. 14, no. 134, p. 20170391, 2017.
- [38] A. Fornito, A. Zalesky, and M. Breakspear, “The connectomics of brain disorders,” *Nature Reviews Neuroscience*, vol. 16, no. 3, pp. 159–172, 2015.
- [39] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, “Quantifying echo chamber effects in information spreading over political communication networks,” *EPJ Data Science*, vol. 8, no. 1, pp. 1–13, 2019.
- [40] S. P. Fraiberger, R. Sinatra, M. Resch, C. Riedl, and A.-L. Barabási, “Quantifying reputation and success in art,” *Science*, vol. 362, no. 6416, pp. 825–829, 2018.
- [41] S. Milgram, “The small world problem,” *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.
- [42] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [43] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [44] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Rev. Mod. Phys.*, vol. 74, pp. 47–97, Jan 2002.
- [45] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Phys. Rev. Lett.*, vol. 86, pp. 3200–3203, Apr 2001.
- [46] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.
- [47] P. Sah, M. Otterstatter, S. T. Leu, S. Leviyang, and S. Bansal, “Revealing mechanisms of infectious disease spread through empirical contact networks,” *PLOS Computational Biology*, vol. 17, pp. 1–20, 12 2021.
- [48] M. E. J. Newman, “Spread of epidemic disease on networks,” *Phys. Rev. E*, vol. 66, p. 016128, Jul 2002.

- [49] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, “A high-resolution human contact network for infectious disease transmission,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 51, pp. 22020–22025, 2010.
- [50] C. Castellano and R. Pastor-Satorras, “Thresholds for epidemic spreading in networks,” *Phys. Rev. Lett.*, vol. 105, p. 218701, Nov 2010.
- [51] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of modern physics*, vol. 87, no. 3, p. 925, 2015.
- [52] K. J. Sharkey, “Deterministic epidemiological models at the individual level,” *Journal of Mathematical Biology*, vol. 57, pp. 311–331, 2008.
- [53] M. Youssef and C. Scoglio, “An individual-based approach to sir epidemics in contact networks,” *Journal of Theoretical Biology*, vol. 283, no. 1, pp. 136–144, 2011.
- [54] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, “Discrete-time markov chain approach to contact-based disease spreading in complex networks,” *Europhysics Letters*, vol. 89, no. 3, p. 38009, 2010.
- [55] A. S. Mata and S. C. Ferreira, “Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks,” *Europhysics Letters*, vol. 103, no. 4, p. 48003, 2013.
- [56] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, “Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results,” *Phys. Rev. E*, vol. 86, p. 041125, Oct 2012.
- [57] R. Levins, “Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control1,” *Bulletin of the Entomological Society of America*, vol. 15, pp. 237–240, 09 1969.
- [58] J. V. Wells and M. E. Richmond, “Populations, metapopulations, and species populations: what are they and who should care?,” *Wildlife Society Bulletin*, pp. 458–462, 1995.
- [59] P. Opdam, “Metapopulation theory and habitat fragmentation: a review of holarctic breeding bird studies,” *Landscape ecology*, vol. 5, pp. 93–106, 1991.

- [60] R. Gallotti, A. Bazzani, M. D. Esposti, and S. Rambaldi, “Entropic measures of individual mobility patterns,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, p. P10022, oct 2013.
- [61] L. A. Rvachev and I. M. Longini, “A mathematical model for the global spread of influenza,” *Mathematical Biosciences*, vol. 75, no. 1, pp. 3–22, 1985.
- [62] S. Charaudeau, K. Pakdaman, and P.-Y. Boëlle, “Commuter mobility and the spread of infectious diseases: Application to influenza in france,” *PLOS ONE*, vol. 9, pp. 1–9, 01 2014.
- [63] G. Le Treut, G. Huber, M. Kamb, K. Kawagoe, A. McGeever, J. Miller, R. Pnini, B. Veytsman, and D. Yllanes, “A high-resolution flux-matrix model describes the spread of diseases in a spatial network and the effect of mitigation strategies,” *Scientific Reports*, vol. 12, no. 1, p. 15946, 2022.
- [64] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis, “Population flow drives spatio-temporal distribution of covid-19 in china,” *Nature*, vol. 582, no. 7812, pp. 389–394, 2020.
- [65] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo, “Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 19, pp. 10484–10491, 2020.
- [66] L. A. Rvachev and I. M. Longini Jr, “A mathematical model for the global spread of influenza,” *Mathematical biosciences*, vol. 75, no. 1, pp. 3–22, 1985.
- [67] D. Brockmann and D. Helbing, “The hidden geometry of complex, network-driven contagion phenomena,” *science*, vol. 342, no. 6164, pp. 1337–1342, 2013.
- [68] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, “The role of the airline transportation network in the prediction and predictability of global epidemics,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 7, pp. 2015–2020, 2006.
- [69] J. Gómez-Gardeñes, D. Soriano-Panos, 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.
- [70] F. M. Snowden, *Epidemics and Society: From the Black Death to the Present*, pp. 30–39. Yale University Press, 2019.

- [71] F. M. Snowden, *Epidemics and Society: From the Black Death to the Present*, pp. 101–103. Yale University Press, 2019.
- [72] I. Kautner, M. J. Robinson, and U. Kuhnle, “Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis, and prevention,” *The Journal of pediatrics*, vol. 131, no. 4, pp. 516–524, 1997.
- [73] S. B. Halstead, “Dengue,” *The lancet*, vol. 370, no. 9599, pp. 1644–1652, 2007.
- [74] M. Andraud, N. Hens, C. Marais, and P. Beutels, “Dynamic epidemiological models for dengue transmission: A systematic review of structural approaches,” *PLOS ONE*, vol. 7, pp. 1–14, 11 2012.
- [75] S. Mandal, R. R. Sarkar, and S. Sinha, “Mathematical models of malaria-a review,” *Malaria journal*, vol. 10, pp. 1–19, 2011.
- [76] P. L. Alonso, G. Brown, M. Arevalo-Herrera, F. Binka, C. Chitnis, F. Collins, O. K. Doumbo, B. Greenwood, B. F. Hall, M. M. Levine, *et al.*, “A research agenda to underpin malaria eradication,” *PLoS medicine*, vol. 8, no. 1, p. e1000406, 2011.
- [77] A. A. Ansari, “Clinical features and pathobiology of ebolavirus infection,” *Journal of Autoimmunity*, vol. 55, pp. 1–9, 2014.
- [78] J.-P. Chretien, S. Riley, and D. B. George, “Mathematical modeling of the west africa ebola epidemic,” *eLife*, vol. 4, p. e09186, dec 2015.
- [79] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, *et al.*, “A novel coronavirus from patients with pneumonia in china, 2019,” *New England journal of medicine*, vol. 382, no. 8, pp. 727–733, 2020.
- [80] Y. Shi, G. Wang, X.-p. Cai, J.-w. Deng, L. Zheng, H.-h. Zhu, M. Zheng, B. Yang, and Z. Chen, “An overview of covid-19,” *Journal of Zhejiang University. Science. B*, vol. 21, no. 5, p. 343, 2020.
- [81] M. B. Thomas, “Epidemics on the move: Climate change and infectious disease,” *PLOS Biology*, vol. 18, pp. 1–2, 11 2020.
- [82] W. Martens, L. W. Niessen, J. Rotmans, T. H. Jetten, and A. J. McMichael, “Potential impact of global climate change on malaria risk.,” *Environmental health perspectives*, vol. 103, no. 5, pp. 458–464, 1995.

- [83] S. I. Hay, J. Cox, D. J. Rogers, S. E. Randolph, D. I. Stern, G. D. Shanks, M. F. Myers, and R. W. Snow, “Climate change and the resurgence of malaria in the east african highlands,” *Nature*, vol. 415, no. 6874, pp. 905–909, 2002.
- [84] P. R. Epstein, H. F. Diaz, S. Elias, G. Grabherr, N. E. Graham, W. J. Martens, E. Mosley-Thompson, and J. Susskind, “Biological and physical signs of climate change: focus on mosquito-borne diseases,” *Bulletin of the American Meteorological Society*, vol. 79, no. 3, pp. 409–418, 1998.
- [85] N. Perra, “Non-pharmaceutical interventions during the covid-19 pandemic: A review,” *Physics Reports*, vol. 913, pp. 1–52, 2021.
- [86] E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, B. Lake, C. Cattuto, and M. Tizzoni, “Covid-19 outbreak response, a dataset to assess mobility changes in italy following national lockdown,” *Scientific data*, vol. 7, no. 1, p. 230, 2020.
- [87] H. Fang, L. Wang, and Y. Yang, “Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china,” *Journal of Public Economics*, vol. 191, p. 104272, 2020.
- [88] A. Pan, L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, *et al.*, “Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, china,” *Jama*, vol. 323, no. 19, pp. 1915–1923, 2020.
- [89] N. Haug, L. Geyrhofer, A. Londei, E. Dervic, A. Desvars-Larrive, V. Loreto, B. Pinior, S. Thurner, and P. Klimek, “Ranking the effectiveness of worldwide covid-19 government interventions,” *Nature human behaviour*, vol. 4, no. 12, pp. 1303–1312, 2020.
- [90] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, *et al.*, “The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak,” *Science*, vol. 368, no. 6489, pp. 395–400, 2020.
- [91] N. Ferguson, “Capturing human behaviour,” *Nature*, vol. 446, no. 7137, pp. 733–733, 2007.
- [92] S. Collinson, K. Khan, and J. M. Heffernan, “The effects of media reports on disease spread and important public health measurements,” *PLOS ONE*, vol. 10, pp. 1–21, 11 2015.

- [93] M. Jung, L. Lin, and K. Viswanath, “Associations between health communication behaviors, neighborhood social capital, vaccine knowledge, and parents’ h1n1 vaccination of their children,” *Vaccine*, vol. 31, no. 42, pp. 4860–4866, 2013.
- [94] C. I. Jarvis, K. Van Zandvoort, A. Gimma, K. Prem, P. Klepac, G. J. Rubin, and W. J. Edmunds, “Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk,” *BMC medicine*, vol. 18, pp. 1–10, 2020.
- [95] A. Reznik, V. Gritsenko, V. Konstantinov, N. Khamenka, and R. Isralowitz, “Covid-19 fear in eastern europe: validation of the fear of covid-19 scale,” *International journal of mental health and addiction*, vol. 19, pp. 1903–1908, 2021.
- [96] J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, *et al.*, “Changes in contact patterns shape the dynamics of the covid-19 outbreak in china,” *Science*, vol. 368, no. 6498, pp. 1481–1486, 2020.
- [97] F. Fatoye, T. Gebrye, O. Arije, C. T. Fatoye, O. Onigbinde, and C. E. Mbadada, “Economic impact of covid-19 lockdown on households,” *Pan African Medical Journal*, vol. 40, no. 1, 2021.
- [98] R. Shretta, “The economic impact of covid-19,” *University of Oxford*, 2020.
- [99] L.-P. Béland, A. Brodeur, and T. Wright, “The short-term economic consequences of covid-19: exposure to disease, remote work and government response,” *Plos one*, vol. 18, no. 3, p. e0270341, 2023.
- [100] L. Ferretti, A. Ledda, C. Wymant, L. Zhao, V. Ledda, L. Abeler-Dörner, M. Kendall, A. Nurtay, H.-Y. Cheng, T.-C. Ng, *et al.*, “The timing of covid-19 transmission,” *MedRxiv*, pp. 2020–09, 2020.
- [101] X. R. Hoffmann and M. Boguñá, “Memory-induced complex contagion in epidemic spreading,” *New Journal of Physics*, vol. 21, no. 3, p. 033034, 2019.
- [102] R. Sender, Y. Bar-On, S. W. Park, E. Noor, J. Dushoff, and R. Milo, “The unmitigated profile of covid-19 infectiousness,” *eLife*, vol. 11, p. e79134, aug 2022.
- [103] K. Prem, K. v. Zandvoort, P. Klepac, R. M. Eggo, N. G. Davies, C. for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, A. R. Cook, and M. Jit, “Projecting contact matrices in 177 geographical

- regions: An update and comparison with empirical data for the covid-19 era,” *PLOS Computational Biology*, vol. 17, pp. 1–19, 07 2021.
- [104] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, “On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations,” *Journal of mathematical biology*, vol. 28, pp. 365–382, 1990.
- [105] O. Diekmann, J. Heesterbeek, and M. G. Roberts, “The construction of next-generation matrices for compartmental epidemic models,” *Journal of the Royal Society Interface*, vol. 7, no. 47, pp. 873–885, 2010.
- [106] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006.
- [107] M. Serafino, G. Cimini, A. Maritan, A. Rinaldo, S. Suweis, J. R. Banavar, and G. Caldarelli, “True scale-free networks hidden by finite size effects,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 2, p. e2013825118, 2021.
- [108] I. Voitalov, P. Van Der Hoorn, R. Van Der Hofstad, and D. Krioukov, “Scale-free networks well done,” *Physical Review Research*, vol. 1, no. 3, p. 033034, 2019.
- [109] A. D. Broido and A. Clauset, “Scale-free networks are rare,” *Nature communications*, vol. 10, no. 1, p. 1017, 2019.
- [110] P. Erdos and A. Renyi, “On the evolution of random graphs,” *Publ. Math. Inst. Hungary. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
- [111] K. J. Sharkey, “Deterministic epidemic models on contact networks: Correlations and unbiological terms,” *Theoretical Population Biology*, vol. 79, no. 4, pp. 115–129, 2011.
- [112] O. Givan, N. Schwartz, A. Cygelberg, and L. Stone, “Predicting epidemic thresholds on complex networks: Limitations of mean-field approaches,” *Journal of theoretical biology*, vol. 288, pp. 21–28, 2011.
- [113] M. KRIVELEVICH and B. SUDAKOV, “The largest eigenvalue of sparse random graphs,” *Combinatorics, Probability and Computing*, vol. 12, no. 1, p. 61–72, 2003.

- [114] E. Cator and P. Van Mieghem, “Second-order mean-field susceptible-infected-susceptible epidemic threshold,” *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 5, p. 056111, 2012.
- [115] L. Sattenspiel and K. Dietz, “A structured epidemic model incorporating geographic mobility among regions,” *Mathematical Biosciences*, vol. 128, no. 1, pp. 71–91, 1995.
- [116] C. Cosner, J. C. Beier, R. S. Cantrell, D. Impoinvil, L. Kapitanski, M. D. Potts, A. Troyo, and S. Ruan, “The effects of human movement on the persistence of vector-borne diseases,” *Journal of theoretical biology*, vol. 258, no. 4, pp. 550–560, 2009.
- [117] D. T. Citron, C. A. Guerra, A. J. Dolgert, S. L. Wu, J. M. Henry, H. M. Sánchez C., and D. L. Smith, “Comparing metapopulation dynamics of infectious diseases under different models of human movement,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 18, 2021.
- [118] P. Castioni, R. Gallotti, and M. De Domenico, “Critical behavior in interdependent spatial spreading processes with distinct characteristic time scales,” *Communications Physics*, vol. 4, no. 1, p. 131, 2021.
- [119] N. Mahroum, I. Seida, S. N. Esirgün, and N. L. Bragazzi, “The covid-19 pandemic – how many times were we warned before?,” *European Journal of Internal Medicine*, vol. 105, pp. 8–14, 2022.
- [120] W. H. Organization *et al.*, “Severe acute respiratory syndrome (sars): Status of the outbreak and lessons for the immediate future,” *World Health Organization*, 2003.
- [121] Y. Fan, K. Zhao, Z.-L. Shi, and P. Zhou, “Bat coronaviruses in china,” *Viruses*, vol. 11, no. 3, p. 210, 2019.
- [122] P. Valgañón, A. Brotons, D. Soriano-Paños, and J. Gómez-Gardeñes, “Balancing mobility behaviors to avoid global epidemics from local outbreaks,” *arXiv preprint arXiv:2412.07656*, 2024.