Chaos

AIP Publishing

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 Matemáticas i Enginyeria Informática, 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. Mckendrik and Kermack formulated the celebrated SIR (Suseptible-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 Richardson[3] 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 interactions[6]. 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 scales[7–9].

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 epidemics[10] or to design non-pharmaceutical interventions with minimal socioeconomic impact[11], 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 simulations[14] by incorporating more and more aspects of human behavior and mobility[15]. From the first works, including the complexity of human mobility networks[16–19] the focus has been put on the recurrent nature of human mobility, being it of special importance in urban and regional scales[20–23]. Recently[24], 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 destination[29], the heterogeneous of different contact patterns[30]. 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 countries[31–33]. Likewise, the MIR model has been used to optimize resource allocation to control epidemics[34] or to evaluate the role that individual awareness plays in hampering the spread of diseases[35].

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:

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

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

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

which can be finally simplified to:

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

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:

$$\rho(t+1) = (1-\mu)\rho(t) + P(t)\cdot(1-\rho(t)) \,. \tag{4}$$

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 scales[15]. To such aim, in[24] 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 Mucha[38], 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, in[24] 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 $\vec{\rho}^{\,T} = (\rho_1(t), ..., \rho_N(t))$ 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:

$$\rho_i(t+1) = (1-\mu)\rho_i(t) + (1-\rho_i(t))\Pi_i(t)\ , \tag{5}$$

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

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

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:

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

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

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:

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

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

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:

$$f_i = \frac{n_i^{eff}}{a_i} \,, \tag{11}$$

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 $\langle k^D \rangle$:

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

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 $\langle k^N \rangle$:

$$z^N = \frac{\sum_{i=1}^{N} n_i \langle k^N \rangle}{\sum_{i=1}^{N} n_i \sigma_i} \,. \tag{13}$$

As shown in[24], 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:

$$\lambda_c = \frac{\mu}{\Lambda_{max}(\boldsymbol{M})} \,, \tag{14}$$

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

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

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:

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

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

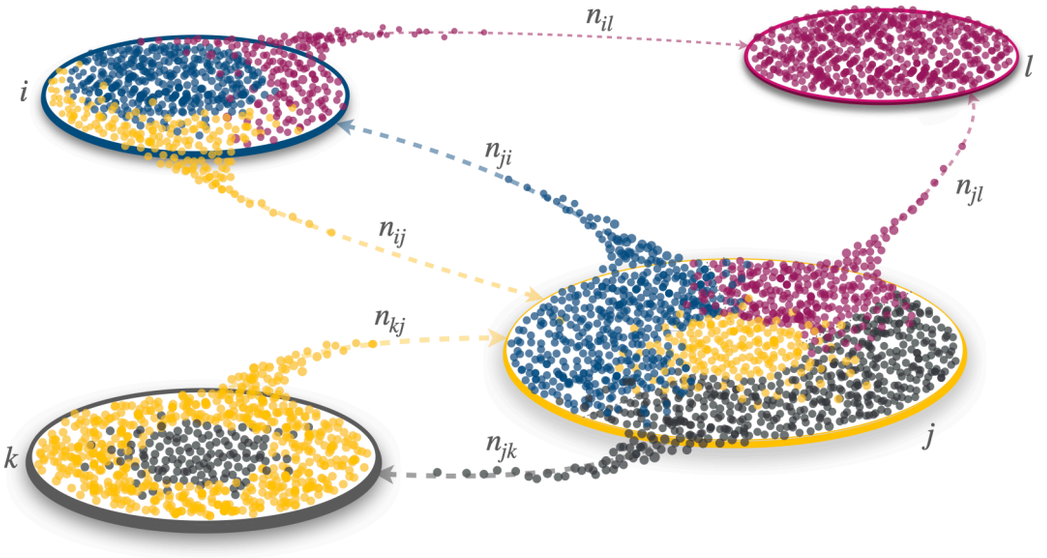

Figure 1. Schematic representation of a toy metapopulation network with distinguishable agents. The population is divided into 4 interconnected subpopulations denoted as i, j, k and l. In each subpopulation the agents are divided (colored) according to their destinations. This way the number of agents 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:

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

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:

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

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:

$$f_i = \frac{n_i^{eff}}{a_i} \, , \tag{19}$$

scaled by:

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

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:

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

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:

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

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:

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

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:

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

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 flows[39], the distribution of the population across patches[40] and the area of each patch[41] 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:

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

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

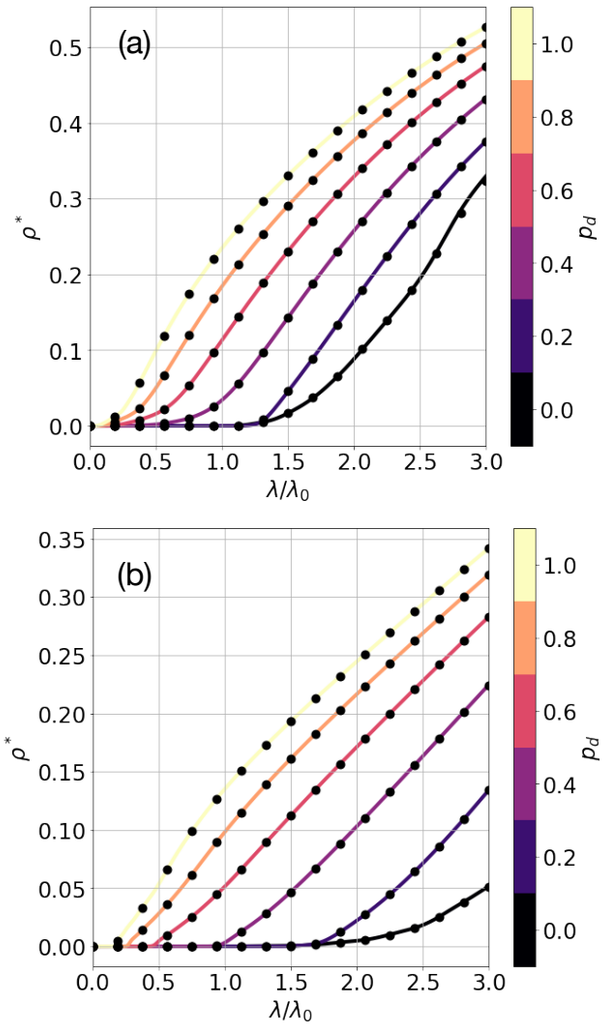

Figure 2. Validation of the markovian equations for the cities of (a) New York and (b) Boston. On the vertical axis we represent the fraction of infected ρ^* in the stationary state for the whole population, and on the horizontal axis the parameter λ normalized to the epidemic threshold for $p_d = 0$: $\lambda_0 = \lambda_c(p_d = 0)$. 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:

$$\mu \rho^*_{ij} = (1 - \rho^*_{ij}) \Pi_{ij}(\vec{\rho}^*) \, . \tag{26}$$

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: $\rho^*_{ij} \equiv \varepsilon_{ij} \ll 1\ \forall i, j$. 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 $\{P^D_i(\vec{\rho}^*)\}$ and $\{P^N_i(\vec{\rho}^*)\}$ and linearize each of them as:

$$P^D_i \simeq p_d \lambda \frac{z^D f_i}{n^{eff}_i} \sum_{j=1}^{N} n_{ji} \varepsilon_{ji} + \lambda (1 - p_d) \frac{z^D f_i}{n^{eff}_i} \sum_{j=1}^{N} n_{ij} \varepsilon_{ij} \, , \tag{27}$$

$$P^N_i \simeq \lambda \frac{z^N \sigma_i}{n_i} \sum_{j=1}^{N} n_{ij} \varepsilon_{ij} \, . \tag{28}$$

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

$$\begin{aligned} \Pi_{ij}(\vec{\varepsilon}) = \lambda \sum_{k,l=1}^{N} \Bigg[& (1 - p_d) p_d \frac{z^D f_k}{n^{eff}_k} n_{lk} \delta_{ik} \\ & + (1 - p_d)^2 \frac{z^D f_l}{n^{eff}_l} n_{lk} \delta_{il} \\ & + p_d^2 \frac{z^D f_k}{n^{eff}_k} n_{lk} \delta_{jk} \\ & + p_d (1 - p_d) \frac{z^D f_l}{n^{eff}_l} n_{lk} \delta_{jl} \\ & + \frac{z^N \sigma_l}{n_l} n_{lk} \delta_{il} \Bigg] \varepsilon_{lk} \end{aligned} \tag{29}$$

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 $\vec{\varepsilon}$:

$$\mu \varepsilon_{ij} \simeq \Pi_{ij}(\vec{\varepsilon}^*) = \lambda \sum_{k,l=1}^{N} M^{il}_{jk} \varepsilon_{lk} \tag{30}$$

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

$$\begin{aligned} M^{il}_{jk} = {} & (1 - p_d) p_d \frac{z^D f_k}{n^{eff}_k} n_{lk} \delta_{ik} + (1 - p_d)^2 \frac{z^D f_l}{n^{eff}_l} n_{lk} \delta_{il} \\ & + p_d^2 \frac{z^D f_k}{n^{eff}_k} n_{lk} \delta_{jk} + p_d (1 - p_d) \frac{z^D f_l}{n^{eff}_l} n_{lk} \delta_{jl} \\ & + \frac{z^N \sigma_l}{n_l} n_{lk} \delta_{il} \, . \end{aligned} \tag{31}$$

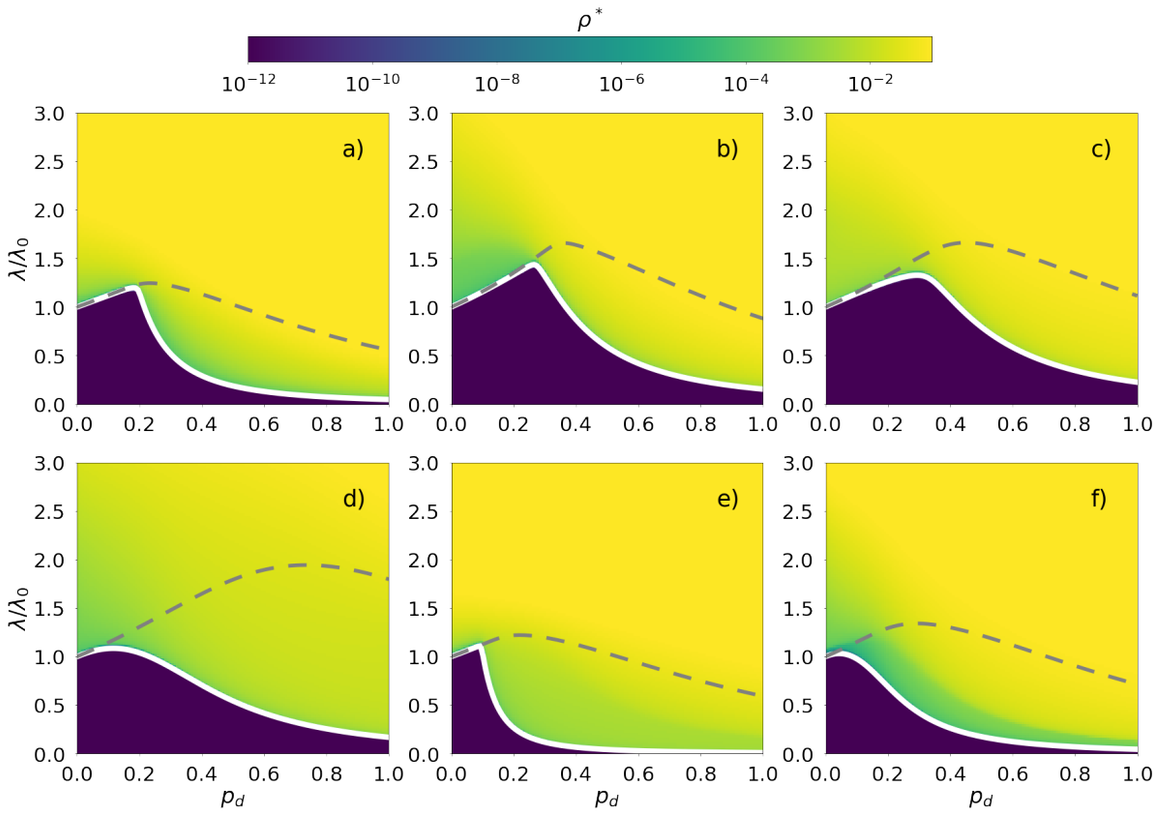

Figure 3. Epidemic diagrams $\rho^*(p_d, \lambda/\lambda_0)$ for the cities of (a) New York, (b) Boston, (c) Austin, (d) Miami, (e) Detroit and (f) Seattle. The continuous white line shows the epidemic threshold obtained by solving Eq. (34), whereas the dashed grey line accounts for the epidemic threshold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

$$\mathbf{M} = \begin{pmatrix} M_{11}^{11} & \cdots & M_{1N}^{11} & \cdots & M_{11}^{1N} & \cdots & M_{1N}^{1N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ M_{N1}^{11} & \cdots & M_{NN}^{11} & \cdots & M_{N1}^{1N} & \cdots & M_{NN}^{1N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ M_{11}^{N1} & \cdots & M_{1N}^{N1} & \cdots & M_{11}^{NN} & \cdots & M_{1N}^{NN} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ M_{N1}^{N1} & \cdots & M_{NN}^{N1} & \cdots & M_{N1}^{NN} & \cdots & M_{NN}^{NN} \end{pmatrix} . \tag{32}$$

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:

$$\frac{\mu}{\lambda}\vec{\varepsilon} = \mathbf{M}\vec{\varepsilon} \, , \tag{33}$$

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:

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

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.

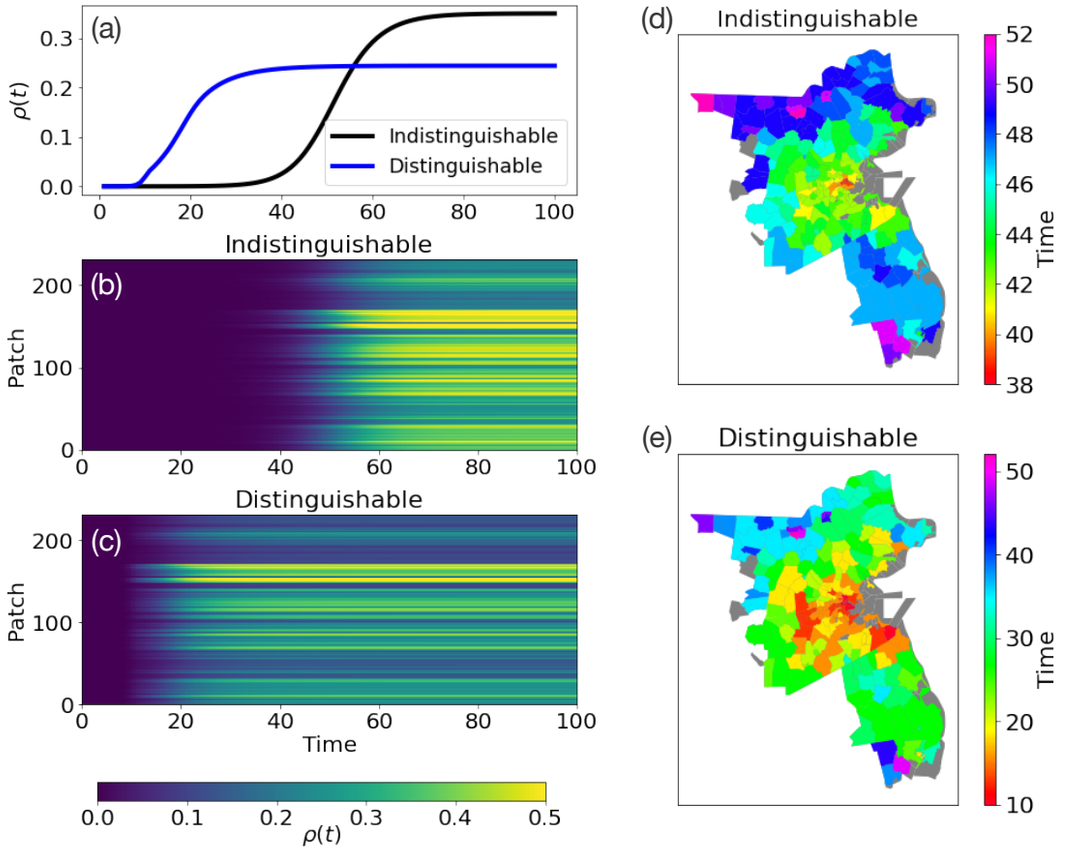

Figure 4. Spatio-temporal unfolding of an epidemic in the Boston area under the indistinguishable and distinguishable frameworks. In (a) we show the time evolution of the global prevalence ρ^*. Panels (b) and (c) show the time evolution of the local prevalence of each patch. Finally, panels (d) and (e) 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 one[24], 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 scenarios[29]. 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 matrices[45], the introduction of biases in the collection of mobility data[46,47] or the misuse of raw mobility data as OD matrices[48] 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.

Chaos

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 2020PANDE00098), 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

[1] A. M'Kendrick, "Applications of mathematics to medical problems," Proceedings of the Edinburgh Mathematical Society **44**, 98–130 (1925).
[2] 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 **115**, 700–721 (1927).
[3] L. F. Richardson, *Weather predictions by Numerical Process* (Cambridge University Press, 1922).
[4] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Presss, 1992).
[5] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals* (Princeton university press, 2011).
[6] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," Reviews of modern physics **87**, 925 (2015).
[7] S. 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).
[8] D. 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).
[9] M. 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).
[10] C. 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).
[11] F. 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.
[12] L. Sattenspiel and K. Dietz, "A structured epidemic model incorporating geographic mobility among regions," Mathematical biosciences **128**, 71–91 (1995).
[13] B. Grenfell and J. Harwood, "(meta) population dynamics of infectious diseases," Trends in ecology & evolution **12**, 395–399 (1997).
[14] F. 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).
[15] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, "Human mobility: Models and applications," Physics Reports **734**, 1–74 (2018).
[16] V. Colizza, R. Pastor-Satorras, and A. Vespignani, "Reaction–diffusion processes and metapopulation models in heterogeneous networks," Nature Physics **3**, 276–282 (2007).
[17] V. Colizza and A. Vespignani, "Invasion threshold in heterogeneous metapopulation networks," Physical review letters **99**, 148701 (2007).
[18] V. Colizza and A. Vespignani, "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations," Journal of theoretical biology **251**, 450–467 (2008).
[19] D. 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).
[20] D. Balcan and A. Vespignani, "Phase transitions in contagion processes mediated by recurrent mobility patterns," Nature physics **7**, 581–586 (2011).
[21] V. Belik, T. Geisel, and D. Brockmann, "Natural human mobility patterns and spatial spread of infectious diseases," Physical Review X **1**, 011001 (2011).
[22] D. Balcan and A. Vespignani, "Invasion threshold in structured populations with recurrent mobility patterns," Journal of theoretical biology **293**, 87–100 (2012).
[23] V. Belik, T. Geisel, and D. Brockmann, "Recurrent host mobility in spatial epidemics: beyond reaction-diffusion," The European Physical Journal B **84**, 579–587 (2011).
[24] J. Gómez-Gardeñes, D. Soriano-Panos, and A. Arenas, "Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks," Nature Physics **14**, 391–395 (2018).
[25] D. 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).
[26] P. 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).
[27] D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, H. J. Martínez, S. Meloni, and J. Gómez-Gardeñes, "Vector-borne epidemics driven by human mobility," Physical Review Research **2**, 013312 (2020).
[28] A. 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 r0," Communications in Nonlinear Science and Numerical Simulation **68**, 1–14 (2019).
[29] D. 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).
[30] W. 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).

[31]A. 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).

[32]G. S. Costa, W. Cota, and S. C. Ferreira, "Outbreak diversity in epidemic waves propagating through distinct geographical scales," Physical Review Research **2**, 043306 (2020).

[33]S. Hazarie, 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).

[34]X. 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).

[35]B. Wang, M. Gou, and Y. Han, "Impacts of information propagation on epidemic spread over different migration routes," Nonlinear Dynamics **105**, 3835–3847 (2021).

[36]D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," science **342**, 1337–1342 (2013).

[37]M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, O. C.-. D. W. Group†, 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).

[38]C. 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://lehd.ces.census.gov/data/lodes/LODES7/` (Accessed: 2021-09-30).

[40](), united States population distribution data sources are available at: `https://www2.census.gov/geo/tiger/TIGER2010BLKPOPHU/` (Accessed: 2021-09-30).

[41]United States data for the areas of patches are available at: `https://www2.census.gov/geo/tiger/TIGER2010/ZCTA5/2010/` (Accessed: 2021-09-30).

[42]P. 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).

[43]M. 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).

[44]L. 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).

[45]J. 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).

[46]M. 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).

[47]F. Schlosser, V. Sekara, D. Brockmann, and M. Garcia-Herranz, "Biases in human mobility data impact epidemic modeling," arXiv preprint arXiv:2112.12521 (2021).

[48]S. Gómez, A. Fernández, S. Meloni, and A. Arenas, "Impact of origin-destination information in epidemic spreading," Scientific reports **9**, 1–9 (2019).

[49]L. 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).

[50]M. Schläpfer, L. Dong, K. O'Keeffe, P. Santi, M. Szell, H. Salat, S. Anklesaria, M. Vazifeh, C. Ratti, and G. B. West, "The universal visitation law of human mobility," Nature **593**, 522–527 (2021).