Epidemic risk assessment from geographic population density

Alessandro Celestini[1], Francesca Colaiori[2,3], Stefano Guarino[1*], Enrico Mastrostefano[1] and Lena Rebecca Zastrow[1]

*Correspondence:
s.guarino@iac.cnr.it

[1] CNR, Institute for Applied Mathematics "Mauro Picone", Rome, Italy
[2] CNR, Institute for Complex Systems, Rome, Italy
[3] Department of Physics, Sapienza University, Rome, Italy

Abstract

The geographic distribution of the population on a region is a significant ingredient in shaping the spatial and temporal evolution of an epidemic outbreak. Heterogeneity in the population density directly impacts the local relative risk: the chances that a specific area is reached by the contagion depend on its local density and connectedness to the rest of the region. We consider an SIR epidemic spreading in an urban territory subdivided into tiles (i.e., census blocks) of given population and demographic profile. We use the relative attack rate and the first infection time of a tile to quantify local severity and timing: how much and how fast the outbreak will impact any given area. Assuming that the contact rate of any two individuals depends on their household distance, we identify a suitably defined geographical centrality that measures the average connectedness of an area as an efficient indicator for local riskiness. We simulate the epidemic under different assumptions regarding the socio-demographic factors that influence interaction patterns, providing empirical evidence of the effectiveness and soundness of the proposed centrality measure.

Keywords: SIR, Epidemic, Risk assessment, Data driven, Urban system, Geographic spreading

Introduction and related work

In the presence of an epidemic outbreak, the availability of instruments to guide containment measures may have an enormous impact on public health, the economy, and society. Mathematical modeling and computer simulations provide powerful tools to understand the dynamics of epidemic outbreaks, design strategies to control and mitigate them, and project decisions and scenarios into potential alternative outcomes. The classical compartmental models of disease propagation (Kermack and McKendrick [1991](#); Hethcote [2000](#)) describe the system at the population level by grouping individuals in sub-populations according to their health states relevant for transmission and tracking changes in the sizes of the sub-populations, without specifying which individuals are involved. The movements in and out of the compartments are governed by constant transition rates in ordinary differential equations, corresponding to exponentially distributed waiting times in the compartments. This approach assumes a fully-mixed population, where an infectious individual is equally likely to spread the disease to any other

individual, and it is equivalent to a mean-field. The modeling of epidemic contagion has evolved from such simple compartmental schemes to sophisticated data-informed models using synthetic populations whose demographics are statistically indistinguishable from the census data. Socio-demographic and geographic factors, such as age and household distance, are essential in determining contact patterns and disease spreading, especially on an urban scale (Illenberger et al. 2013; Walsh and Pozdnoukhov 2011; Del Valle et al. 2007; Mossong et al. 2008).

In the last decades, lots of work has focused on incorporating heterogeneous contact patterns in the models. In individual-based models, the population-level behavior emerges from the microscopic interactions between agents that can carry a set of attributes such as age, spatial location, and social behavior. Socio-demographic attributes collected from census data and surveys or integrated with mobile, traffic, or wearable sensor data and online tools (Mossong et al. 2008; Cattuto et al. 2010; Mistry et al. 2020), allow the construction of realistic contact matrices describing the mixing patterns in typical social settings (e.g., households, schools, workplaces) (Mossong et al. 2008; Eubank et al. 2004; Merler and Ajelli 2009; Chang et al. 2021). Metapopulation models, built using data from inter-population mobility (Hufnagel et al. 2004; Colizza et al. 2006; Chang et al. 2021; Tizzoni et al. 2014), allowed remarkable progress in describing how the disease travels from one city/nation to the other and estimating epidemic paths predictability (Colizza et al. 2007). Compartmental models with stochastic and time-varying disease transmission rates allow for incorporating control measures and environmental variations due to unobserved processes (Cazelles et al. 2018; Gray et al. 2011; Bonaccorsi and Ottaviano 2016). By representing individuals in the population as vertices of a graph, network-based models aim to understand how the topological structure of the contact network affects the dynamics spreading upon them (Gupta et al. 1989; Newman 2002; Keeling and Eames 2005; Chakrabarti et al. 2008; Pastor-Satorras et al. 2015).

Ball and Neal (2002, 2008) incorporated some randomness in the classic stochastic Susceptible-Infected-Removed (SIR) model in a closed population, adding to the local contacts among neighbors on a social network, also global contacts among pairs of individuals chosen at random in the population. This model is also suited to mimic multiple transmission routes (Kiss et al. 2006). In Celestini et al. (2022), the authors address similar questions for epidemics at urban scale by simulating an agent-based SIR model in a synthetic urban population. As in Ball and Neal (2008), their model incorporates casual contacts on top of a structured set of social relations. In addition, they use a data-informed model to obtain an age-structured population connected by a geographically referenced social network and allow for global casual contacts to be dependent on distance and the age class of an agent. The model assumes that an individual has daily contact with her household members, frequent contact with a small network of acquaintances (e.g., friends or coworkers), and fortuitous contact with the rest of the population. As data about the population density are often available in a discretized form, the model used in Celestini et al. (2022) considers a regular tiling of the urban territory that induces a partition of the population based on their tile of residence and permits analysis of the geographic diffusion of the disease. According to this model, urban epidemic outbreaks are not predictable (as defined in Colizza et al. 2006) because of the many possible epidemic pathways. However, the authors highlight that the first infection

time and the final fraction of the population infected within a tile correlate to the logarithm of the tile population.

Building on these findings, in this paper, we investigate more in detail how the geographical distribution of the population influences the spatial and temporal evolution of an epidemic disease outbreak at an urban scale, where density patterns may impact the epidemic. We focus in particular on two local observables: the tile first infection time τ and relative attack rate α, defined as the attack rate within the tile divided by the total attack rate. Assuming that the frequency of contacts decays as an inverse power of the distance (Liben-Nowell et al. 2005; Illenberger et al. 2013; Kowald et al. 2013; Herrera-Yagüe et al. 2015; Büchel and Ehrlich 2020), we find that the expected value of τ and α can be described efficiently through a metric that we call *geographic harmonic centrality* (HC). We provide qualitative insights on the relations binding τ and α to HC, and support our findings through simulations. We find that these relations still hold when the model is generalized to include age-based mixing, vertex-intrinsic sociability fitness, and multiple levels of mixing with contacts that recur with different frequencies.

SIR epidemics on a real territory: Model and methods

Let us consider a Susceptible-Infectious-Removed (SIR) epidemic model where the individuals of the population belong to one of three disjoint compartments: susceptible ($\mathcal{S}$), if they never caught the disease; infectious ($\mathcal{I}$), if they are currently infected and contagious; removed ($\mathcal{R}$), if they already recovered from or died due to the disease. Only two transitions are possible between the three compartments: a susceptible individual may become infected ($\mathcal{S} \to \mathcal{I}$) due to an interaction with an infectious individual; an infectious individual may spontaneously recover or die ($\mathcal{I} \to \mathcal{R}$). We consider a discrete-time *synchronous cellular automaton* in which the dynamic follows a *reactive process* (de Arruda et al. 2018). For each time step $t = 0, 1, \ldots$, we denote with $\mathcal{S}(t)$, $\mathcal{I}(t)$ and $\mathcal{R}(t)$ the sets of, respectively, susceptible, infected and recovered individuals at time t. The size of these compartments is, respectively, $S(t) = |\mathcal{S}(t)|$, $I(t) = |\mathcal{I}(t)|$ and $R(t) = |\mathcal{R}(t)|$. The interactions between the individuals of the population are modeled by a sequence of temporal contact networks $G^t = (V, E^t)$, where E^t is the set of contacts occurring at time t and $(u, v) \in E^t$ with probability $p_{u,v}$ independently of all other edges. If $v \in \mathcal{S}(t)$, for each $u \in \mathcal{I}(t)$ such that $(u, v) \in E^t$, the disease is transmitted from u to v with fixed probability β—i.e., the probability that $v \in \mathcal{I}(t+1)$ given $v \in \mathcal{S}(t)$ is $1 - \prod_{u \in \mathcal{I}(t)}(1 - p_{u,v}\beta)$. If $u \in \mathcal{I}(t)$, u recovers with a fixed rate μ—i.e., the probability that $u \in \mathcal{R}(t+1)$ given $u \in \mathcal{I}(t)$ is μ. We assume that a single individual, called *index case*, is infectious at time 0, i.e., $I(0) = 1$.

We simulate epidemic outbreaks in the Municipality of Viterbo, Italy, represented as a rectangular bounding box inhabited by a *synthetic population V* of $N \approx 60K$ agents. The territory of Viterbo is partitioned into a grid of n square tiles of fixed side $l = 500$ m and we disregard all tiles having less than 10 residents (see Fig. 1). Each agent v has the following three attributes: a tile of residence i_v, inferred from density estimates provided by the WorldPop Project (Bondarenko et al. 2020); an age-tag g_v —*child* (0–17), *young* (18–34), *adult* (35–64), or *elder* (65+)—drawn based on census data aggregated at the provincial level, provided by the Italian Institute of Statistics (ISTAT) (all details can be found in Guarino et al. 2021); a social fitness score f_u,
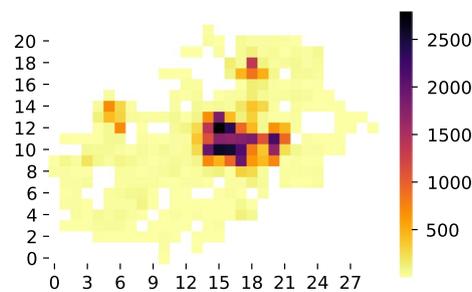
Fig. 1 The territory of the Municipality of Viterbo, tessellated into tiles. The color of each tile show the number of individuals living in the area, only tiles having 10 or more residents are considered

drawn from a Lognormal distribution. The sociability score f_u measures u's tendency to interact with other people; using a Lognormal distribution guarantees that the degree distribution of the interaction graphs has a heavy-tail and an adjustable mode (Guarino et al. 2021). In particular, we set the distribution of f_u to have "small" skewness and variance, so that most vertices have degree close to the average and the hubs are limited in both number and size (Kertész et al. 2021). Lognormally distributed data occur across several domains (Mitzenmacher 2004)—including social networks (Liben-Nowell et al. 2005; Illenberger et al. 2013)—and Lognormal distributions often fit the statistical properties of real-world network better than power laws (Broido and Clauset 2019). More details on the choice of the fitness distribution can be found in Guarino et al. (2021).

In the following, we will initially consider two cases: homogeneous mixing (HM), characterized by $p_{u,v} = p$ for constant $p > 0$; distance-based mixing (DM), where $p_{u,v} \propto d_{i,j}^{-1}$ for all $u \in V_i$ and $v \in V_j$ *only* depends on the distance $d_{i,j}$ between V_i and V_j. We will then assess, through extensive simulations, whether the predictions obtained for the DM model remain valid when $p_{u,v}$ also depends on the social fitness, on the age and/or on the underlying *social fabric*, encoded into a static urban social network (USN) that combines synthetic households and acquaintances. Intra-household (e.g., kinship) edges are defined by drawing a *clique* for each household, where the breakdown of the population into households is entirely inferred from the available data. Acquaintance edges are instead drawn based on the fundamental assumption that the probability of two individuals being "friends" is ultimately governed by their age group, their geographical distance, and their sociability. More details on our urban social network model can be found in Guarino et al. (2021) and Guarino et al. (2021).

In total, we consider five configurations, summarized in Table 1 and described more in details as needed. All configurations are fully dynamic—i.e., the temporal contact networks G^t are mutually independent—except for the *multi-layer mixing* (MLM), in which part of the interactions that occur at each time t are selected from the USN static social network of "strong ties".

The configurations are characterized by a fixed expected volume of daily contacts, i.e., they all have the same average interaction probability $\langle p_{u,v} \rangle$. In line with Biggerstaff et al. (2014) and Liu et al. (2018), all the experimental results presented in this paper are obtained calibrating the SIR model on a common Influenza Like Illness

Table 1 The interpersonal contact models used throughout the paper

Model		Contact probability $p_{u,v}$
Name	Acronym	
Homogeneous mixing	HM	constant
Distance–based mixing	DM	$\propto d_{u,v}^{-1}$
Distance and fitness–based mixing	DFM	$\propto d_{u,v}^{-1} f_u f_v$
Distance and age–based mixing	DAM	$\propto s_{g(u),g(v)} d_{u,v}^{-1}$
Multi–layer mixing	MLM	$= 1$ if $(u,v) \in E_H, = 0.5$ if $(u,v) \in E_A,$ $\propto s_{g(u),g(v)} d_{u,v}^{-1} f_u, f_v$ otherwise

$p_{u,v}$ is the probability that u and v are connected at each step of the simulation

f_u is the Lognormally distributed social fitness of u

s_{g_u,g_v} is the data–driven age–based social mixing for u and v's age groups g_u and g_v

E_H and E_A are the edge sets of two static networks, respectively used to represent households and acquaintances (see Sect. 4 and Guarino et al. (2021) for more details)

(ILI) by setting $\mu = 1/3$, so that the average time to recovery is 3 days, and $\langle p_{u,v} \rangle \beta$ such that the expected number of infections caused by a single case in a fully susceptible population is $R_0^{\text{index}} = 1.3$.

The tiling of the city's territory induces a partition of the population V into n blocks $\{V_i\}_{i=1}^n$, having size $|V_i| = N_i$. With a slight abuse of notation, V_i will be used to indicate both the ith tile and the population living therein. $\mathcal{I}_i(t)$ and $I_i(t)$ will denote, respectively, the set and the number of infected individuals in block i at time t. The partition of the population into tiles mirrors the fact that the spatial density of the population always comes at some given resolution. The tiling, however, has no meaningful interpretation: at urban scale, in fact, there are no clear boundaries and truly distinct means of transportation (e.g., air vs. road traffic) that permit to identify meaningful sub-populations. The probabilistic rule used to establish if u and v come into contact is the same regardless of whether u and v belong to the same or to different tiles. Albeit the tiling produces a discretization of distances—and, hence, a stratification of contact frequencies—our model significantly differs from typical meta-population models, where the local and global dynamics of contagion are well separated.

On large scale territories, the high number of possible spreading channels may be compensated by the presence of dominant connections in human traffic flow, making the geographic spread of the disease predictable (Colizza et al. 2006), in the sense that different outbreak realizations present similar patterns of spatial epidemic diffusion over time. As shown in Celestini et al. (2022), in our case the predictability of the geographic patterns of diffusion of the epidemic is instead limited—at least far from the epidemic peak—due to the absence of backbones capable of defining preferred epidemic pathways. Still, the stratification of contact frequencies may impact on the "riskiness" of different blocks with respect to epidemic outbreaks. As proxies of riskiness we will use the (expected) *relative attack rate* and local *first infection time*, defined as follows:

$$\alpha_i = \frac{\mathbb{E}[R_i]}{\mathbb{E}[R]} \frac{N}{N_i}, \tag{1}$$

$$\tau_i = \mathbb{E}[\min\{t : I_i(t) > 0\}]. \tag{2}$$

where $R = R(+\infty)$ and $R_i = R_i(+\infty)$ denote the final epidemic size, in V and V_i, respectively, and $\mathbb{E}[\cdot]$ is the expected value. For all $i = 1, \ldots, n$, α_i is the expected attack rate of V_i divided by the total expected attack rate, a measure of the chances for an individual resident in V_i to be infected with respect to the average population. τ_i is the expected time of the first infection in V_i, a measure of how quickly V_i must be isolated in order to protect its sub-population. The expected value in the definition of α_i and τ_i will be replaced by the empirical mean in the experimental analysis presented in Sect. 4. In all cases, the averages will be computed over 100 independent epidemic outbreaks, simulated with identical configuration parameters, that involve at least 25% of the population. We empirically verified that, with a single index case, the epidemic size follows a bimodal distribution with a minimum at around 25%N, so all simulations resulting in $R/N < 0.25$ have been considered cases of early extinction and, as such, discarded.

Geographic distributions of individuals and the spread of epidemics

In the simplest scenario, the geographic distribution of the population impacts the relative risk of an infection event occurring in a specific local area in two ways: (i) through the local population density, and (ii) through the connectedness of the local area to the rest of the territory. If we assume that the contact rate of any two individuals depends on their household distance, we can quantify the average connectedness of an area in terms of its geographical centrality. Thus, the chances for a given area to be reached by the contagion ultimately depend on its local density and centrality. To formalize this idea, we focus on the relative attack rate α and on the first infection time τ, defined in (1) and (2), starting from two simple models: homogeneous mixing (HM) and distance-based mixing (DM).

Homogeneous mixing

At this level of description, the geography of an outbreak of size R can be described by means of a vector of indices $O = (j_0, j_1, ..., j_{R-1})$ such that $j_h = i$ if and only if the hth infection occurs in V_i (simultaneous infection events are sorted in random order). If we assume that the population is homogeneously mixed, all individuals have, on average, the same chances of contracting the disease. The probability that $j_h = i$ is thus $S_i(t_h)/S(t_h)$, i.e., the probability that, at the time t_h when the hth infection occurs, a susceptible individual chosen uniformly at random belongs to V_i. Under homogeneous mixing, for fixed t_h, the expected value of the ratio $S_i(t_h)/S(t_h)$ can be approximated by the local population density $\rho_i = N_i/N$.

Under these assumptions, V_i's epidemic size R_i follows a binomial distribution of parameters ρ_i and R, hence, its expected value is $\rho_i\mathbb{E}[R]$. This gives:

$$\alpha_i = \rho_i \frac{N}{N_i} = 1. \tag{3}$$

Let W_i be the number of infection events needed to observe the first infection in V_i, i.e. $W_i = h_i$ if $h_i = \min\{h : j_h = i\}$. W_i follows the geometric distribution $\rho_i(1 - \rho_i)^{h_i - 1}$, and the expected value of W_i is $1/\rho_i$. In the exponential phase of the epidemic the number of contagion events grows exponentially in time and we can assume that the hth contagion

Celestini *et al. Applied Network Science*   (2022) 7:39

Page 7 of 15

event occurs at time $t_h \propto \ln(h)$, yielding the following estimate for the first infection time:

$$\tau_i \simeq -\ln(\rho_i) \cdot \text{const.} = -\ln\left(\frac{N_i}{N}\right) \cdot \text{const.} \tag{4}$$

Connection flow to a sub-population

In the HM case, the local density $\rho_i = N_i/N$ is a good approximation of the probability of an infection reaching the ith tile, and it is therefore the key ingredient in determining both α_i and τ_i. In many cases, it is possible to partition the population into sub-populations $\{V_i\}_{i=1}^{n}$ and to assume that, at least with good approximation, $p_{u,v} \approx p_{i,j}$ for all $u \in V_i$ and $v \in V_j$, i.e., that $p_{u,v}$ mostly depends on the sub-populations of u and v. A simple example is the geographic partitioning into subpopulations achieved by dividing a territory with a tiling. More generally, one can partition a population based on a combination of geographic and demographic attributes, such as location and age.

In this case, the probability that at least one individual of V_i comes into contact with an infectious individual is intuitively proportional to the quantity

$$k_i = \frac{N_i \sum_j N_j p_{i,j}}{N} = \frac{N_i}{N} \sum_j N_j p_{i,j}. \tag{5}$$

k_i can be interpreted as a *per node connection flow* to the ith tile: a randomly chosen node $u \in V$ has on average k_i daily contacts with the nodes $v \in V_i$—i.e., the dynamic network G^t contains, on average, k_i edges incident to u and V_i. If we know that none of the individuals in V_i is infectious, a better estimate is provided by

$$\bar{k}_i = \frac{N_i}{N} \sum_{j \neq i} N_j p_{i,j}, \tag{6}$$

which only differs from k_i for the fact that the average does not involve the members of V_i itself.

The same rationale used for the HM case now brings us to the following general estimate for α_i in a partitioned population:

$$\alpha_i \simeq k_i \frac{N}{N_i} \cdot \text{const.} = \sum_j N_j p_{i,j} \cdot \text{const.} \tag{7}$$

Along the same line, if the infection reaches V_i during the phase of exponential growth of the epidemic, τ_i can be estimated as:

$$\tau_i \simeq -\ln(\bar{k}_i) \cdot \text{const.} = -\ln\left(\frac{N_i}{N} \sum_{j \neq i} N_j p_{i,j}\right) \cdot \text{const.} \tag{8}$$

Distance-based mixing

Let $d_{i,j}$ be the normalized geographic distance between the center of tiles V_i and V_j. The normalization is obtained through a division by $\frac{l}{2}$, where l is the tile side. Additionally, we impose $d_{i,i} = 1$, so that individuals in the same tile are at half the distance of individuals living in neighboring tiles. Finally, we set $d_{u,v} = d_{i,j}$ for all $u \in V_i$ and $v \in V_j$. This approximation is often inevitable, at least for small values of l, because population density data are generally discrete and we may only have access to the residence tile of a vertex.

In line with previous empirical findings (Liben-Nowell et al. 2005; Illenberger et al. 2013; Kowald et al. 2013; Herrera-Yagüe et al. 2015; Büchel and Ehrlich 2020), to model distance-based mixing (DM) we assume that the probability that $u \in V_i$ and $v \in V_j$ come into contact decays as a power of the distance: $p_{u,v} = p_{i,j} \propto d_{i,j}^{-\gamma}$ for $\gamma > 0$. It is straightforward that substituting $p_{i,j} \propto d_{i,j}^{-\gamma}$ in (5) and (6) gives

$$k_i \propto \frac{N_i}{N} \sum_j N_j d_{i,j}^{-\gamma}, \tag{9}$$

$$\bar{k}_i \propto \frac{N_i}{N} \sum_{j \neq i} N_j d_{i,j}^{-\gamma}. \tag{10}$$

The right hand sides of (9) and (10) are, actually, measures of geographic centrality. This leads us to introduce the following metrics:

$$\mathrm{HC}^\gamma(v) = \frac{1}{N} \sum_{u \in V} d_{u,v}^{-\gamma}, \tag{11}$$

$$\mathrm{HC}_i^\gamma = \mathrm{HC}^\gamma(V_i) = \frac{N_i}{N} \sum_{j=1}^n N_j d_{i,j}^{-\gamma}, \tag{12}$$

$$\bar{\mathrm{HC}}^\gamma(v) = \frac{1}{N} \sum_{u \in V \setminus V_i}^n d_{u,v}^{-\gamma} \quad \text{for } v \in V_i, \tag{13}$$

$$\bar{\mathrm{HC}}_i^\gamma = \bar{\mathrm{HC}}^\gamma(V_i) = \frac{N_i}{N} \sum_{\substack{j=1 \\ j \neq i}}^n N_j d_{i,j}^{-\gamma} = \mathrm{HC}_i^\gamma - \frac{N_i}{N} d_{i,i}^{-\gamma}. \tag{14}$$

$\mathrm{HC}^\gamma(v)$ and HC_i^γ are the *harmonic geographic centrality* of order γ of vertex v and of tile V_i, respectively. $\bar{\mathrm{HC}}^\gamma(v)$ and $\bar{\mathrm{HC}}_i^\gamma$ are the *outward* versions of $\mathrm{HC}^\gamma(v)$ and HC_i^γ that only measure the centrality of the individuals $v \in V_i$ with respect to the population that does not belong to V_i.

Substituting $k_i \propto \mathrm{HC}_i^\gamma$ and $\bar{k}_i \propto \bar{\mathrm{HC}}_i^\gamma$ in (7) and (8), respectively, we obtain the following estimates for the first infection time and the relative attack rate under the DM model:

$$\alpha_i \simeq \frac{N}{N_i} \mathrm{HC}_i^\gamma \cdot \mathrm{const.}, \tag{15}$$

$$\tau_i \simeq -\ln(\bar{\mathrm{HC}}_i^\gamma) \cdot \mathrm{const.} \tag{16}$$

Experimental analysis

The analysis presented so far suggests that, under the widely accepted hypothesis that the frequency of contacts decays with the distance, the riskiness of different areas of a city is tied to their local density and geographic centrality.

To support our thesis and to assess whether these findings remain valid under more realistic models, in this section we evaluate empirically the distribution of α_i and τ_i for five different models. We consider the HM model and the DM model (with $\gamma = 1$), and three other models that include additional factors influencing the patterns of interpersonal contacts:

- In the distance and fitness-based mixing (DFM) model, $p_{u,v} \propto d_{u,v}^{-1} f_u f_v$, where f_u is a measure of u's social fitness (Caldarelli et al. 2002), drawn from a Lognormal distribution.
- In the distance and age-based mixing (DAM) model, $p_{u,v} \propto s_{g_u,g_v} d_{u,v}^{-1}$, where $s_{i,j}$ is the age-based social mixing for age groups i and j, deduced from aggregated contact data (Mossong et al. 2008) through the SOCRATES Data Tool (Willem et al. 2020).
- Finally, in the multi-layer mixing (MLM) model, different sets of edges recur with different probabilities. We first construct two edge-disjoint static networks: the *household graph* $G_H = (V, E_H)$, composed of a *clique* for each synthetic household drawn based on census data (Guarino et al. 2021); the *acquaintance graph* $G_A = (V, E_A)$, where, for each $u, v \in V$, $(u, v) \in E_A$ independently of all others with probability $\psi_{u,v} = s_{g_u,g_v} d_{u,v}^{-1} f'_u f'_v$ ($f'_u \neq f_u$ is extracted independently of f_u, but from the same distribution). Household relations induce a static frame of *daily* contacts, i.e., $p_{u,v} = 1$ for all $(u, v) \in E_H$, while acquaintance relations induce *frequent* contacts, with $p_{u,v} = 0.5$ for all $(u, v) \in E_A$. All $(u, v) \notin E = E_H \cup E_A$ have $p_{u,v} \propto s_{g_u,g_v} d_{u,v}^{-1} f_u f_v$. Thus, in this model part of the connections—all $(u, v) \in E_H$—are static.

To allow for a fair comparison, in all configurations $p_{u,v}$ is normalized so that $\sum_{u<v} p_{u,v} = |E|$. This guarantees that the expected density of G^t and, thus, the expected number of potential contagion events $\sum_{u<v} p_{u,v} \beta$ are fixed. The choice of 1 as the distance dependence exponent, and of a Lognormal fitness distribution is justified on an empirical basis in Guarino et al. (2021) to best represent social contacts at the urban level. In the MLM model, a fixed social network instance is considered, so that all fluctuations are due to epidemic dynamics.

All the figures presented in the current section are scatterplots fitted by a linear regression model, with a dot for each tile of the territory of Viterbo. Since each epidemic simulation is inherently stochastic, each run reaches a possibly different subset of tiles. Grey dots in the plots represent tiles that are reached by only a fraction of the simulations and that are ignored by the linear regression due to their inherent bias. Configurations with

Relative attack rate

Figure 2 shows α_i as a function of $\rho_i = N_i/N$ for the HM model (panel a) and as a function of $\frac{N}{N_i}\mathrm{HC}_i^\gamma$ for the other four configurations described in the previous section (panels b–e). The homogeneous mixing case reported in panel (a) is shown only as a confirmation of the expected behavior. Panel (b) depicts the trend of α_i for the DM case, the good agreement with (15) is the first confirmation that HC_i^1, weighted by N/N_i, is indeed a good candidate to predict the epidemic spread when the frequency of contacts decays as d^{-1}. Adding fitness to the model (DFM configuration, panel (c)) yields more heterogeneity to the degree of nodes and apparently weakens the role of HC_i^γ as a predictor of the relative attack rate. For DAM and MLM (panels (d) and (e), respectively), however, we see again a good agreement with (15). Let us remind that in the MLM model both the static acquaintance graph and the fully dynamic part of the interaction graph are built relying on a social fitness score that guarantees a heavy-tailed degree distribution. Nonetheless, $\frac{N}{N_i}\mathrm{HC}_i^\gamma$ is a good predictor of α_i for the MLM configuration (Fig. 2e), contrary to the DFM (Fig. 2c). The insight is that the approximation works even for heterogeneous networks, provided that a significant volume of interactions tend to recur over time. Finally, in all panels (except (a)) we see that the relative attack rate (and thus the final epidemic size in each tile) grows linearly with HC_i^γ: the more central the tile is, the greater the fraction of its population that is infected.
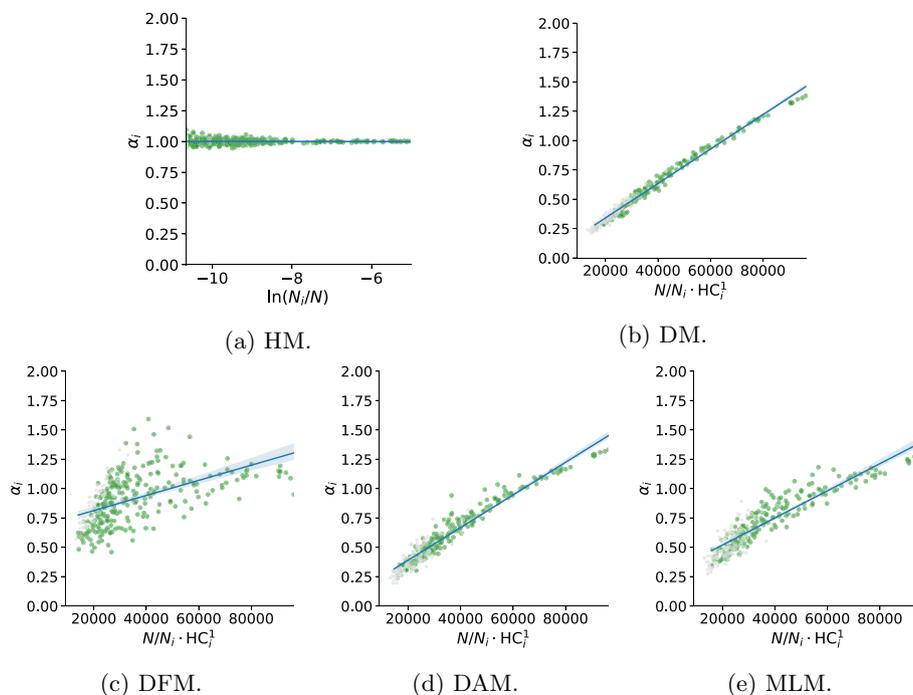

Fig. 2 The dependence of α_i upon N_i and HC_i^1 for the five considered models

First infection time

Figure 3 shows the plot of τ_i versus $\bar{\mathrm{HC}}_i^\gamma$. Despite the range of τ_i being highly dependent on the configuration—with more heterogeneous models corresponding to faster epidemics—all panels show a very good agreement between the average time of first infection in the simulations and the value predicted by (16). For all the configurations tested the time of first infection decreases linearly with the logarithm of $\bar{\mathrm{HC}}_i^\gamma$. In particular, as expectable, the most centrally located tiles are infected first.

Characterizing epidemic front waves

Based on (16), the epidemic front waves, i.e., the sets of tiles having equal τ, can be characterized by the condition

$$\ln\left(\bar{\mathrm{HC}}_i^\gamma\right) = H \tag{17}$$

for some constant H. Given the spatial distribution of the population, we would like to predict the shape of these contours on the city map, but the geographic interpretation of (17) is not straightforward. In this section, we study these epidemic front waves to see if they can be described using some simple features of V_i, ideally, its population and its location in the territory. To this end, we observe that

$$\bar{\mathrm{HC}}_i^\gamma = N_i \langle \bar{\mathrm{HC}}^\gamma(v) \rangle_i,$$

i.e., $\bar{\mathrm{HC}}_i^\gamma$ is N_i times the average of $\bar{\mathrm{HC}}^\gamma(v)$ over the elements of V_i, which disentangles the dependance of τ_i on the population and the position of V_i. Condition (17) can now be reformulated as
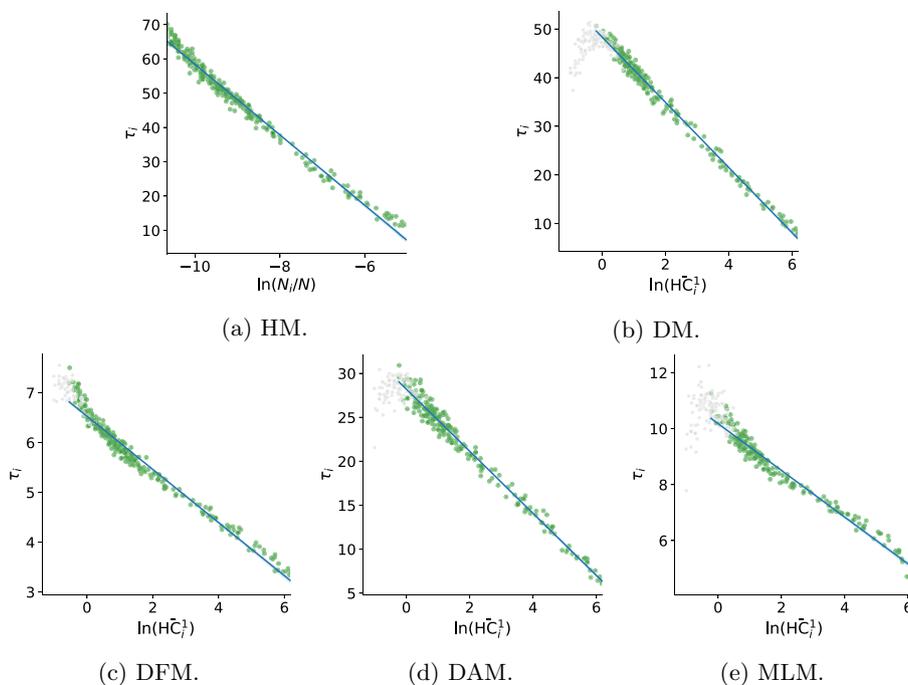

(a) HM.　　　　　　　　　(b) DM.

(c) DFM.　　　　　(d) DAM.　　　　　(e) MLM.

Fig. 3 The dependence of τ_i upon $\bar{\mathrm{HC}}_i^1$ for the five considered configurations

$$\ln(N_i) = -\ln\left(\langle \bar{H}C^{\gamma}(v)\rangle_i\right) + H \tag{18}$$

Arguably, in most real world cities $\langle \bar{H}C^{\gamma}(v)\rangle_i$ is correlated, to some extent, with the distance $d_{C,i}$ of V_i from the center of the city. The actual relation between $\langle \bar{H}C^{\gamma}(v)\rangle_i$ and $d_{C,i}$ depends on the spatial density of the considered population, but the intuition that these two quantities are correlated is validated empirically in Fig. 4, where a relation of the type $\langle \bar{H}C^{\gamma}(v)\rangle_i \propto e^{-d_{C,i}}$ seems to emerge, regardless of whether we consider the real, uniform or shuffled population density for Viterbo—here, we identified C as the tile having max HC^1.

As a consequence of the estimated relation $\langle \bar{H}C^{\gamma}(v)\rangle_i \propto e^{-d_{C,i}}$, the epidemic front waves can be characterized by a simpler condition of the type

$$\ln(N_i) = a d_{C,i} + b \tag{19}$$

for $a, b > 0$. This relation can be clearly seen in Fig. 5, where we grouped the tiles having similar τ and we included a linear regression for each class. While in the HM model τ only depends on the population, as per equation (4), in the DM model (Fig. 5a, c) the epidemic front waves lie along parallel non-horizontal lines in the $d_{C,i}, \ln(N_i)$ plane—albeit with variable slope according to the configuration parameters. The different shapes that these epidemic front waves take on the map show the role that the (generally, positive) correlation between the centrality of a tile and its population have in shaping the riskiness map of urban epidemics.

Conclusions

Thanks to the current availability of data regarding human mobility flows, the large-scale spreading of a disease is, to some extent, predictable (Colizza et al. 2006). To the best of our knowledge, it has not been ascertained whether major backbones of disease propagation also exist at the urban scale. In our previous work (Celestini et al. 2022), we explored the consequences of the widely-acknowledged assumption that the frequency of contacts decays as an inverse power of the distance (Liben-Nowell et al. 2005; Illenberger et al. 2013; Kowald et al. 2013; Herrera-Yagüe et al. 2015; Büchel and Ehrlich 2020). We found that this distance-based stratification of contacts is not enough to make the geographic spreading of urban epidemics predictable. Yet, we also showed that heterogeneity in population distribution reflects onto a local variation of epidemic timing and severity. Building on Celestini et al. (2022), in this paper we formalize the relative riskiness of living in different areas of a city by means of a suitable metric of geographic centrality.
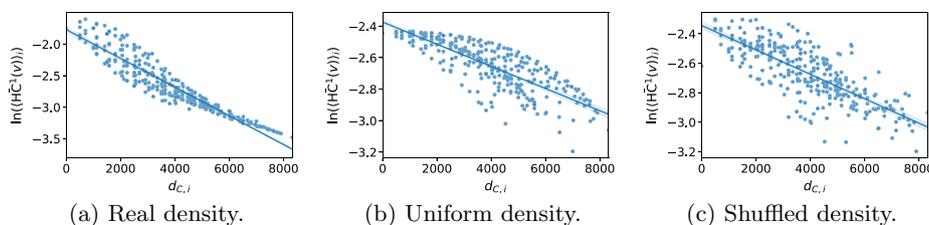

(a) Real density.   (b) Uniform density.   (c) Shuffled density.

Fig. 4 $\ln\left(\langle \bar{H}C^1(v)\rangle_i\right)$ versus $d_{C,i}$, for different population densities

(a) Epidemic front waves for DM.

(b) Epidemic front waves for DM w/ uniform density.

(c) Epidemic front waves for DM w/ shuffled density.

(d) Epidemic front waves for DM on the map.

(e) Epidemic front waves for DM w/ uniform density on the map.

(f) Epidemic front waves for DM w/ shuffled density on the map.

(g) Real density.

(h) Uniform density.

(i) Shuffled density.
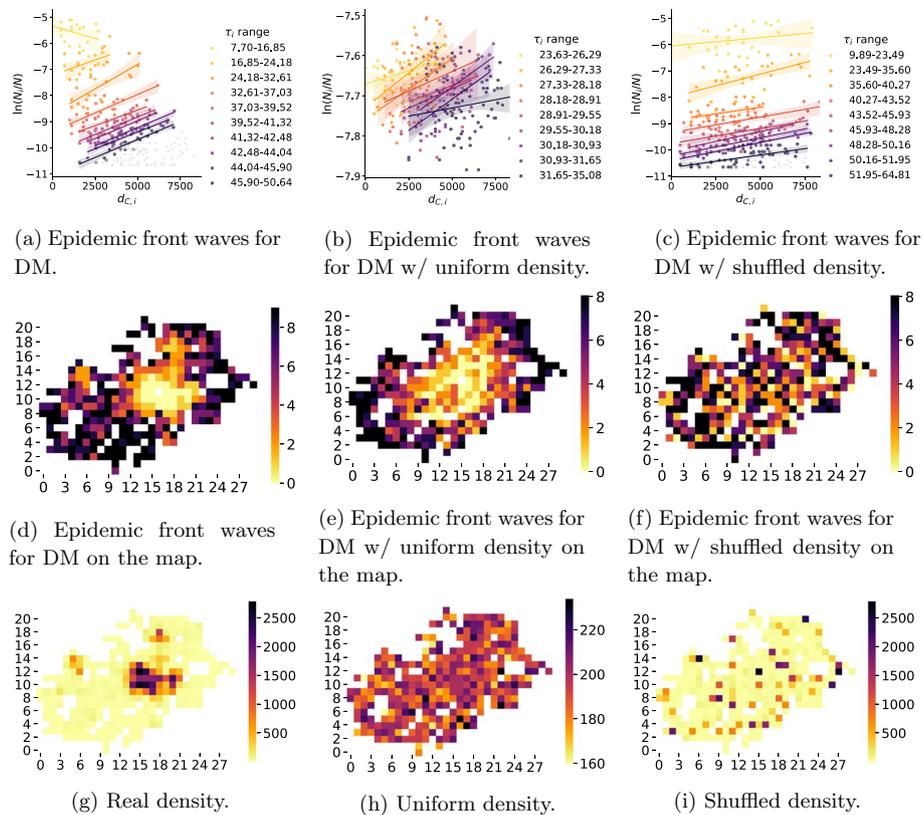
Fig. 5 Epidemic front waves, characterized in terms of their population and centrality (**a**–**c**), and shown on the map of Viterbo (**d**–**f**), together with the corresponding population density (**g**–**i**)

To measure the riskiness of a specific portion of the considered territory, we consider the relative attack rate and the first infection time. We find two closed-form expressions that approximate these two quantities when the rate of interpersonal contacts *only* depends on household distances. Both expressions can be formulated in terms of a new metric of geographic harmonic centrality, that quantifies the average connectedness of a local area with the rest of the territory. By means of simulations, we provide empirical evidence that our results can be generalized to more heterogeneous contact patterns.

The risk evaluation obtained in this way is independent of the actual evolution of the epidemic and can be done from geographic and demographic information only. In addition, under reasonable assumptions, the epidemic front waves can be characterized by a simple condition that depends on the local population density and on the distance from the city center. Our results can thus potentially inform containment strategies at a local scale.

It would be interesting to extend this analysis to the regional scale and compare our predictions and simulation results with actual epidemic data, where available. Another possibility we intend to explore is to replace the distance with the traveling time between different blocks, to include urban environments where physical barriers, such as a river or a railway constrain mobility within the city.

Author contributions
All authors designed the study. AC and EM acquired the data. AC, SG and EM created the software used to simulate the epidemic and perform the data analysis. All authors interpreted the results and wrote the paper. All authors revised the work and read and approved the final manuscript.

Availability of data and materials
The code and data used in this paper are available at https://gitlab.com/cranic-group/usn.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

Ball F, Neal P (2002) A general model for stochastic sir epidemics with two levels of mixing. Math Biosci 180(1–2):73–102

Ball F, Neal P (2008) Network epidemic models with two levels of mixing. Math Biosci 212(1):69–87

Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L (2014) Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC Infect Dis 14(1):1–20

Bonaccorsi S, Ottaviano S (2016) Epidemics on networks with heterogeneous population and stochastic infection rates. Math Biosci 279:43–52

Bondarenko M, Kerr D, Sorichetta A, Tatem A (2020) Census/projection-disaggregated gridded population datasets for 189 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs. University of Southampton, Southampton

Broido AD, Clauset A (2019) Scale-free networks are rare. Nat Commun 10(1):1–10

Büchel K, Ehrlich MV (2020) Cities and the structure of social interactions: evidence from mobile phone data. J Urban Econ 119:103276

Caldarelli G, Capocci A, De Los Rios P, Muñoz MA (2002) Scale-free networks from varying vertex intrinsic fitness. Phys Rev Lett 89:258702

Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton J-F, Vespignani A (2010) Dynamics of person-to-person interactions from distributed RFID sensor networks. PLoS ONE 5(7):11596

Cazelles B, Champagne C, Dureau J (2018) Accounting for non-stationarity in epidemiology by embedding time-varying parameters in stochastic models. PLoS Comput Biol 14(8):1006211

Celestini A, Colaiori F, Guarino S, Mastrostefano E, Zastrow LR (2022) Epidemics in a synthetic urban population with multiple levels of mixing. In: Benito RM, Cherifi C, Cherifi H, Moro E, Rocha LM, Sales-Pardo M (eds) Complex networks and their applications X. Springer, Cham, pp 315–326

Chakrabarti D, Wang Y, Wang C, Leskovec J, Faloutsos C (2008) Epidemic thresholds in real networks. ACM Trans Inf Syst Secur (TISSEC) 10(4):1–26

Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J (2021) Mobility network models of covid-19 explain inequities and inform reopening. Nature 589(7840):82–87

Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci 103(7):2015–2020

Colizza V, Barrat A, Barthélemy M, Vespignani A (2007) Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. BMC Med 5(1):1–13

de Arruda GF, Rodrigues FA, Moreno Y (2018) Fundamentals of spreading processes in single and multilayer complex networks. Phys Rep 756:1–59

Del Valle SY, Hyman JM, Hethcote HW, Eubank SG (2007) Mixing patterns between age groups in social networks. Soc Netw 29(4):539–554. https://doi.org/10.1016/j.socnet.2007.04.005

Eubank S, Guclu H, Kumar VA, Marathe MV, Srinivasan A, Toroczkai Z, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. Nature 429(6988):180–184

Gray A, Greenhalgh D, Hu L, Mao X, Pan J (2011) A stochastic differential equation sis epidemic model. SIAM J Appl Math 71(3):876–902

Guarino S, Mastrostefano E, Bernaschi M, Celestini A, Cianfriglia M, Torre D, Zastrow LR (2021) Inferring urban social networks from publicly available data. Future Internet 13(5):108

Guarino S, Mastrostefano E, Celestini A, Bernaschi M, Cianfriglia M, Torre D, Zastrow LR (2021) A model for urban social networks. In: International conference on computational science. Springer, pp 281–294

Gupta S, Anderson RM, May RM (1989) Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS (London, England) 3(12):807–817

Herrera-Yagüe C, Schneider CM, Couronne T, Smoreda Z, Benito RM, Zufiria PJ, González MC (2015) The anatomy of urban social networks and its implications in the searchability problem. Sci Rep 5(1):1–13

Hethcote HW (2000) The mathematics of infectious diseases. SIAM Rev 42(4):599–653

Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci 101(42):15124–15129

Illenberger J, Nagel K, Flötteröd G (2013) The role of spatial interaction in social networks. Netw Spat Econ 13(3):255–282

Keeling MJ, Eames KT (2005) Networks and epidemic models. J R Soc Interface 2(4):295–307

Kermack WO, McKendrick AG (1991) Contributions to the mathematical theory of epidemics-I. Bull Math Biol 53(1–2):33–55

Kertész J, Török J, Murase Y, Jo H-H, Kaski K (2021) Modeling the complex network of social interactions. In: Pathways between social science and computational social science. Springer, Berlin, pp 3–19

Kiss IZ, Green DM, Kao RR (2006) The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. Math Biosci 203(1):124–136

Kowald M, van den Berg P, Frei A, Carrasco J-A, Arentze T, Axhausen K, Mok D, Timmermans H, Wellman B (2013) Distance patterns of personal networks in four countries: a comparative study. J Transp Geogr 31:236–248

Liben-Nowell D, Novak J, Kumar R, Raghavan P, Tomkins A (2005) Geographic routing in social networks. Proc Natl Acad Sci 102(33):11623–11628

Liu Q-H, Ajelli M, Aleta A, Merler S, Moreno Y, Vespignani A (2018) Measurability of the epidemic reproduction number in data-driven contact networks. Proc Natl Acad Sci 115(50):12680–12685

Merler S, Ajelli M (2009) The role of population heterogeneity and human mobility in the spread of pandemic influenza. Proc R Soc B Biol Sci 277(1681):557–565

Mistry D, Litvinova M, Chinazzi M, Fumanelli L, Gomes MF, Haque SA, Liu Q-H, Mu K, Xiong X, Halloran ME et al (2020) Inferring high-resolution human mixing patterns for disease modeling. arXiv preprint. arXiv:2003.01214

Mitzenmacher M (2004) A brief history of generative models for power law and lognormal distributions. Internet Math 1(2):226–251. https://doi.org/10.1080/15427951.2004.10129088

Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 5(3):1–1. https://doi.org/10.1371/journal.pmed.0050074

Newman ME (2002) Spread of epidemic disease on networks. Phys Rev E 66(1):016128

Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. Rev Mod Phys 87(3):925

Tizzoni M, Bajardi P, Decuyper A, Kon Kam King G, Schneider CM, Blondel V, Smoreda Z, González MC, Colizza V (2014) On the use of human mobility proxies for modeling epidemics. PLoS Comput Biol 10(7):1003716

Walsh F, Pozdnoukhov A (2011) Spatial structure and dynamics of urban communities

Willem L, Hoang TV, Funk S, Coletti P, Beutels P, Hens N (2020) SOCRATES: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for COVID-19. BMC Res Notes 13(1):1–8

Publisher's Note

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