

Switchover phenomenon induced by epidemic seeding on geometric networks

Gergely Ódor^a, Domonkos Czifra^{b,c}, Júlia Komjáthy^d, László Lovász^{b,e,1}, and Márton Karsai^{b,f}

^aInformation and Network Dynamics Lab, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; ^bDepartment of Combinatorics and its Applications, Alfréd Rényi Institute of Mathematics, H-1053 Budapest, Hungary; ^cArtificial Intelligence Research Group, Alfréd Rényi Institute of Mathematics, H-1053 Budapest, Hungary; ^dDepartment of Mathematics and Computer Science, Eindhoven University of Technology, NL-5612AZ Eindhoven, The Netherlands; ^eInstitute of Mathematics, Eötvös Loránt University, H-1053 Budapest, Hungary; and ^fDepartment of Network and Data Science, Central European University, A-1100 Vienna, Austria

Contributed by László Lovász, August 27, 2021 (sent for review July 8, 2021; reviewed by Marc Barthelemy and Alan Frieze)

It is a fundamental question in disease modeling how the initial seeding of an epidemic, spreading over a network, determines its final outcome. One important goal has been to find the seed configuration, which infects the most individuals. Although the identified optimal configurations give insight into how the initial state affects the outcome of an epidemic, they are unlikely to occur in real life. In this paper we identify two important seeding scenarios, both motivated by historical data, that reveal a complex phenomenon. In one scenario, the seeds are concentrated on the central nodes of a network, while in the second one, they are spread uniformly in the population. Comparing the final size of the epidemic started from these two initial conditions through data-driven and synthetic simulations on real and modeled geometric metapopulation networks, we find evidence for a switchover phenomenon: When the basic reproduction number R_0 is close to its critical value, more individuals become infected in the first seeding scenario, but for larger values of R_0 , the second scenario is more dangerous. We find that the switchover phenomenon is amplified by the geometric nature of the underlying network and confirm our results via mathematically rigorous proofs, by mapping the network epidemic processes to bond percolation. Our results expand on the previous finding that, in the case of a single seed, the first scenario is always more dangerous and further our understanding of why the sizes of consecutive waves of a pandemic can differ even if their epidemic characters are similar.

epidemics propagation | networks | random graphs | switchover phenomenon

Whether a local epidemic becomes a global pandemic depends on several conditions. Biological (1), environmental (2), and behavioral (3) factors are important but the final outcome of the epidemic is also strongly determined by the size and location of the seed population from which it originates (4–9). If the epidemic strikes first at an isolated place with low population density and few local transportation connections, it may become rapidly extinct without causing a major breakout. The dynamics can be entirely different if the epidemic starts from a well-connected, more populated place where it can survive and spread to the rest of the population more easily. Although this is the broadly accepted picture, we challenge this intuition and show that seeding an epidemic from the most tightly connected core of a network does not always lead to a larger epidemic in the long run, in terms of the number of final infected people: If the disease transmits easily, seeding the spreading from nodes selected uniformly at random from the network could reach a larger population.

Among many factors, similar processes could act in the background during the early phase of the COVID-19 pandemic: Even though the circulating severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic variants had similar transmission profiles, the number of infections differed significantly in subsequent waves of the pandemic in several countries (10–12).

This was especially true for Hungary with an order of magnitude more daily number of detected cases observed at the peak of the second wave compared to the first outbreak (Fig. 1A). Reasons behind this variation could be the effect of several factors. This includes seasonal effects as people may have spent more time outside during the first wave (13); regulations were followed less strictly during the second wave that may have potentially induced a larger number of contacts per person transmitting the disease (14); the testing capacities also developed considerably since the beginning of the pandemic, allowing for more observations during the second wave; and further, while the first wave of the epidemic was boosted by institutional outbreaks (e.g., in hospitals and care homes) that were easier to identify and contain (15), the second wave circulated freely in the population without effective control (16).

The global and local mobilities of people are among the most important driving factors behind the spatial spread of most diseases (17–19). How people commute locally or travel between cities and countries can be well represented by mobility networks. Concentrating on Hungary, we consider a spatial mobility network (Fig. 1C) describing the average number of daily commuters, who travel to work and school between 1,398 settlements with populations larger than 1,000 inhabitants according to the 2016 Hungarian microcensus (20). From epidemic data

Significance

Initial seeding of an epidemic from the best-connected nodes of a network would intuitively lead to the largest outbreak. We challenge this picture and explore a switchover phenomenon: Epidemics started from the central part of a geometric metapopulation network can reach more individuals only if the basic reproduction number is small, but if the epidemic is more infectious, it reaches a larger population when seeded from uniformly selected nodes. We identify spatial geometry to amplify this effect in both data-driven and synthetic epidemic models, and we give rigorous proofs that this phenomenon appears in various random networks. Our results help us understand why real epidemics started from seemingly similar conditions may have significantly different outcomes.

Author contributions: G.Ó., D.C., J.K., L.L., and M.K. designed research; G.Ó., D.C., J.K., L.L., and M.K. performed research; G.Ó., J.K., and L.L. contributed new analytic tools; G.Ó., D.C., and M.K. analyzed data; G.Ó., D.C., J.K., L.L., and M.K. wrote the paper; and G.Ó., D.C., and M.K. developed computational tools.

Reviewers: M.B., French Alternative Energies and Atomic Energy Commission–Paris Saclay; and A.F., Colorado Mesa University.

The authors declare no competing interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹ To whom correspondence may be addressed. Email: laszlo.lovasz@ttk.elte.hu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2112607118/-/DCSupplemental>.

Published October 7, 2021.

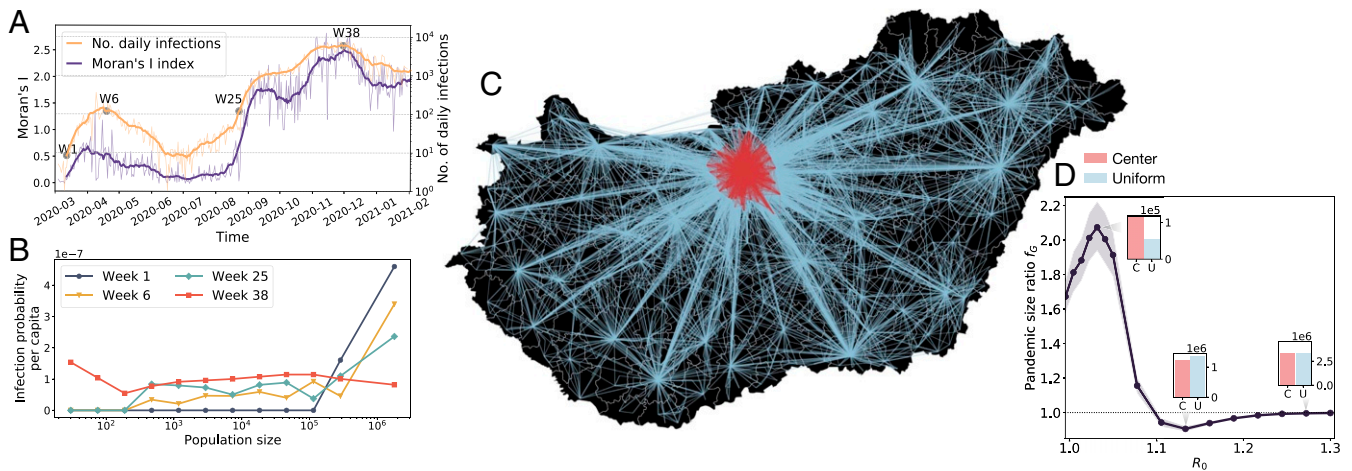


Fig. 1. Data-driven observations of the switchover phenomenon. (A) Dynamics of the number of daily infections (orange) and Moran's I index (purple) for Hungary. Indicated time points match the observation weeks in B. (B) Distribution of per capita infection probabilities in settlements of different sizes at different observation times (in weeks). (C) Commuting network map of Hungary with settlements larger than 1,000 inhabitants and commuting links with more than 25 travelers depicted. Central Hungary (called Center) is highlighted with red. (D) Pandemic size ratios $f_G(R_0, s)$ measured between the endemic sizes of simulated SIR epidemic processes seeded from s towns selected from the center or uniformly at random from the whole metapopulation network. Epidemic seeded from the center may lead to larger outbreaks for small R_0 basic reproduction numbers (left bar plot), while uniform seeding results in more infections for larger R_0 (center bar plot). For very large R_0 values, differences due to different seeding strategies disappear (right bar plot).

we can follow the daily number of new COVID-19 infection cases in each of these settlements to explore their spatiotemporal distribution in this geometric network. The analysis of the epidemic on this structure sheds light on a so-far neglected effect associated to the different initial seeding conditions of the virus, which may have contributed to the emerging large differences between the first and the second waves.

The first wave started in March 2020 in Hungary (W1 in Fig. 1A). As in many countries, the disease arrived in the country via international air travel and first landed in larger cities (21–24), resulting in outbreaks clumped around highly populated areas. This is evident from Fig. 1B, where the per-capita infection probability at the beginning of the first wave (week 1) indicates that infection cases were concentrated in cities with the largest populations. To further demonstrate how much of the infection spreading can be attributed to everyday mobility (as opposed to atypical mobility patterns, such as going on a vacation), we computed Moran's I index on this network (for definition see *Materials and Methods*). This is a spatial autocorrelation function, which has been previously used to measure the spatial association of the COVID-19 infections by ref. 24. Looking at the time dependency of Moran's I index (Fig. 1A), during the beginning of the first wave (W1) the index indicates low spatial correlation, meaning that infected cases were concentrated only in a few places during this initial stage of the epidemic. In contrast, the second wave in Hungary (and Europe) emerged after the summer season and was potentially induced by people coming back from holidays bringing back the virus to their local community, and thus restarting the pandemic from a significantly different initial condition. Indeed, at the beginning of the second wave (at the end of August 2020 in Hungary) (Fig. 1A), new infected cases were distributed more homogeneously all around the country. On the one hand, this is evident from Fig. 1B where the corresponding probability distribution (week 25) is more stretched toward smaller population, compared to week 1. On the other hand, the same conclusion can be drawn from Fig. 1A (W25) where Moran's I index starts to grow rapidly from a state where infections were even more homogeneously distributed than at the peak of the first wave (W6), although the infection numbers were comparable. This homogenization of infected cases continued during the unfolding of the second wave, leading to a fully

uniform distribution—corresponding to population densities—at the peak (W38 in Fig. 1A and B). Surprisingly, the first wave that started from the most tightly connected, central, and largest populations led to a significantly smaller number of infections compared to the second wave that reached an order of magnitude more people, even though it was initiated from more uniformly distributed populations of the network.

To better understand this phenomenon, we build a metapopulation network (25) using the spatial commuting network of Hungary (20). In such a network we consider n nodes, which represent populations of individuals (which we also call towns or settlements from now on), connected by weighted edges, encoding the number of people traveling between them. On this network we simulate one of the most basic models for an epidemic spreading (26), a susceptible-infected-recovered (SIR) process, where each individual can be in one of three mutually exclusive states (S, susceptible; I, infected; or R, recovered). In an SIR model on a metapopulation network, the epidemic evolves in two phases in each iteration. During the reaction phase, individuals inside each town mix homogeneously and, if infected, pass the disease to susceptible others with probability β . Meanwhile, an infected individual may recover with probability μ , after which the individual would never get infected again. This simple SIR dynamic is often characterized by the basic reproduction number, defined as the average number of people infected by one ill person in a fully susceptible town ($R_0 = \beta/\mu$). Subsequently, during the diffusion phase, individuals (possibly infected) are selected with probability p_m to move to neighboring nodes in the metapopulation network, this way migrating the epidemic to other towns (for a more formal definition see *Materials and Methods*). Note that we concentrate on the conventional SIR model to demonstrate a phenomenon. However, our observations hold for more realistic models too, including the susceptible-exposed-infected-recovered (SEIR) model with an additional compartment of exposed (E) state, better describing the reaction scheme of the SARS-Cov-2 disease.

To capture the observed structural distinction of the central towns in the case of the spatial commuting network of Hungary, we identify a central node set \mathcal{C} , containing the districts of Budapest and its suburbs (red nodes in Fig. 1C), which represent about 30% of the total population of the country (27).

While this definition of \mathcal{C} relies on the specific urban structure of Hungary, we could find more general definitions for \mathcal{C} that are based solely on the network structure. The simplest formal definition would be to take a prescribed number of nodes with the highest degrees. We could also use the core of the network for this purpose (for definition see *Materials and Methods*), which is obtained by repeatedly deleting all nodes with the lowest degrees as long as only nodes with prescribed degrees remain. Later we exploit all these definitions to identify \mathcal{C} when studying different types of model network structures.

Once we have selected \mathcal{C} , we consider two initial conditions to seed the SIR process in the metapopulation network, starting the spreading from the same number of towns and individuals in both cases. In one case, we choose s ($\ll |\mathcal{C}| \ll n$) number of towns selected randomly from the \mathcal{C} central set, while in the other case we choose s towns uniformly at random from the whole network. To initiate the spreading, we infect a small i_0 fraction of the total population selected uniformly at random from the chosen s towns, irrespective of their size. This way, for both seeding strategies (centralized or uniform), each seeded town is infected on average with the same number of agents (i_0/s fraction of the total population). To observe the relative effects of the two seeding scenarios, we look at the experimental pandemic size ratio $f_G(R_0, s)$ that we define as the ratio of average final infection sizes of epidemic processes seeded from central or uniformly randomly selected towns (for a related but more formal definition see *Theoretical Results*). Interestingly, as shown in Fig. 1D, we find $f_G(R_0, s) > 1$ for small $R_0 \simeq 1$, which means that the epidemics seeded from the \mathcal{C} central set lead to larger outbreaks. However, as we increase R_0 , the fraction $f_G(R_0, s)$ falls under 1, and thus seeding from uniformly random selected towns over the whole country induces a larger outbreak. This switchover phenomenon appears in the slightly supercritical regime, where R_0 is not too large and where the epidemic never reaches the total population. Instead, due to network effects, it stays clustered around the seeded towns until it dies out. The differences in the infected cluster sizes induced by the two seeding scenarios lead to the observed switchover phenomenon in this regime. On the other hand, if R_0 grows larger, the difference between these seeding scenarios vanishes as the epidemic reaches essentially the whole population in each case.

The switchover phenomenon challenges the commonly accepted intuition that the size of the epidemic is always the largest if seeded from the best-connected subgraph or from the largest degree nodes of a network. In the remaining sections of this paper, we show that the switchover phenomenon is ubiquitous in various network models and argue that the geometric nature of the underlying network plays an important role in amplifying the size of the phenomenon. We perform data-driven and synthetic simulations of spreading processes on real, geometric, and random metapopulation networks and provide a rigorous proof of the phenomenon after mapping it to a bond percolation problem. Our results open a research direction toward understanding why real epidemics that started from seemingly similar conditions may have significantly different outcomes.

Results

Simulation Results. In the Introduction, we demonstrated the existence of the switchover phenomenon on a metapopulation model parameterized by the Hungarian commuting network, which is a spatially embedded geometric network (see Fig. 1C for Hungary) featuring various structural heterogeneities (for a detailed data description see *Materials and Methods*). Geometric constraints inducing commuting connections at various distances, link weights coding the daily commuting frequencies between towns, the number of commuting connections of each settlement (also called the node degree in the network), and the size of the different towns are all network characteristics taking

values ranging over orders of magnitude. These properties may all contribute to the emergence of the observed switchover phenomenon of simulated spreading processes (an SIR model in our case), with central vs. random seeding in the metanetwork.

To identify which underlying network characteristics are the most important to play a role, we use random reference network models (28). We homogenize the network in different ways to remove certain structural heterogeneities and compare the outcome of the experimental pandemic size ratio of simulated spreading processes on the randomized structures to our observations on the empirical network (see curve with blue dots in Fig. 3A). First, to reduce the effects of weight heterogeneities, we reset edge weights to the mean weight of all outgoing edges of each node (see curve with green diamonds in Fig. 3A). Although this way of homogenization changes somewhat the pandemic size ratio function, it does not have dramatic effects on the observed phenomena. Second, to remove the effects of heterogeneous town sizes and the varying number of commuting individuals from different settlements, we set each town's population to the system average ($\bar{N} = 6,581$) and choose the fraction of commuters to be the same (i.e., to $p_m = 0.001$) for each town. Interestingly, this way of homogenization makes the switchover phenomenon even stronger (see curve with red squares in Fig. 3A). Finally, we reshuffle the ends of network links using the configuration network model (28). This removes any structural correlations from the network beyond degree heterogeneity, including geometric effects such as long-distance connections, the central-periphery structure, structural hierarchy, and locally dense subgraphs. Due to this shuffling process the switchover phenomenon disappears or becomes too small to be observed (see curve with yellow triangles in Fig. 3A), indicating that geometric correlations play a central role behind its emergence.

Geometric Inhomogeneous Random Graphs. The specific effect of an underlying geometry can be studied by using geometric network models, opening directions for an analytical description of the phenomenon. Geometric inhomogeneous random graph (GIRG) models (29) provide a good framework to generate structurally heterogeneous synthetic metapopulation structures embedded in geometric space (for detailed definition see *Materials and Methods*). GIRGs have two robust parameters that control the qualitative features of the emerging network. The parameter τ determines the variability of the number of neighbors of individual nodes (smaller values of τ correspond to more variability, while keeping the average degree the same). This is apparent when comparing Fig. 2A–C where all parameters of the simulated network structures are identical, and only τ is increased gradually, leading to the disappearance of hubs, i.e., nodes with a large number of neighbors. The other robust parameter of GIRG, α , controls the number of long-range connections in the network coding the possible travels between far-apart nodes. If $\alpha \simeq 1$, many long-range edges appear, resembling an ageometric (or mean-field) structure (Fig. 2B), but when α is increased, the number of long-range contacts is reduced, and the network exhibits a more apparent underlying geometry as demonstrated in Fig. 2C. The values of (τ, α) determine different universality classes of GIRGs with respect to average distance in the network (see more detailed definition and explanation in *Materials and Methods*).

To distinguish the central set \mathcal{C} from the rest of the network, we adopt here the concept of core decomposition (for definition see *Materials and Methods*). Formally, this procedure provides the highest core as a subgraph with nodes having at least k neighbors inside the core, for the largest possible k (for definition see *Materials and Methods*). Similar to the data-driven simulations, we start the spreading process from two seeding conditions: by initially selecting s towns within the highest core of the

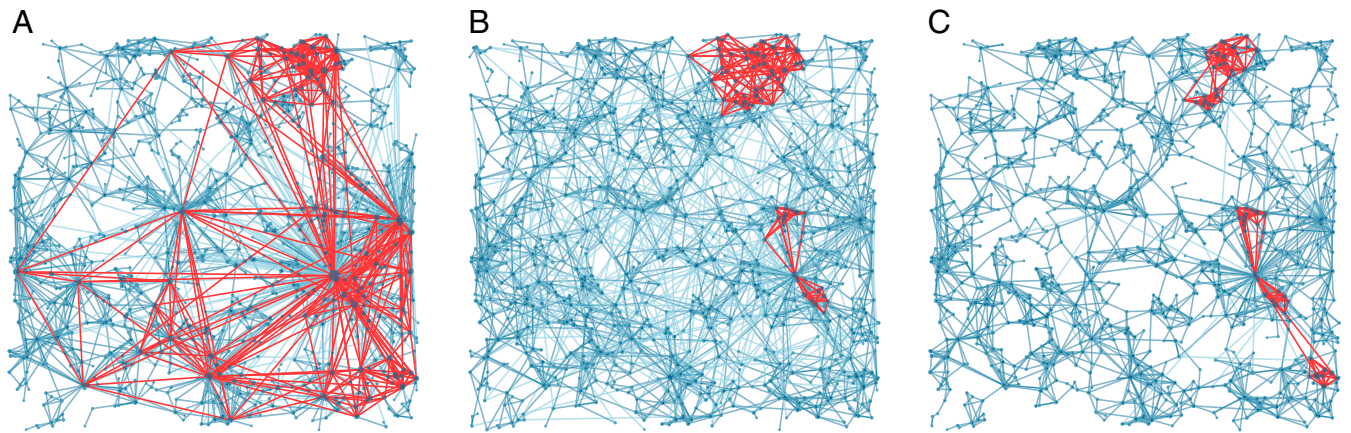


Fig. 2. GIRD models. (A) Model networks with connection parameter $\tau = 2.5$ and geometry parameter $\alpha = 2.3$. In case $\tau < 3$ the network appears with high degree variability and dominant hubs connected via (light blue) long-range edges. (B) By increasing $\tau > 3$ (here $\tau = 3.5$) and decreasing $\alpha = 1.3$, hubs' sizes reduce and long-range interactions become more random. (C) With parameter $\alpha = 2.3$ (and $\tau = 3.5$) the network is strongly geometric with (dark blue) short-range interactions and no long-range links. Networks were generated over the same $N = 1,000$ nodes randomly distributed in a unit square and the highest k-cores of each graph are colored in red. Red clusters highlight the largest k-cores in each network.

metapopulation network (corresponding to the central set C) or by selecting the same number of towns uniformly at random from the whole structure and infecting on average the same number of agents (i_0/s fraction of the total population) in each selected town in both scenarios.

We observe a similar but stronger switchover phenomenon of the pandemic size ratio f_G in GIRDs compared to the data-driven simulations. As seen in Fig. 3B, the “shape” of $f_G(R_0, s)$ as the function of R_0 strongly depends on the network properties controlled by the parameters of the model. If the network parameter $\tau \geq 3$, the modeled epidemic processes, which were initiated from uniform random seeds, reach larger populations, reflected by f_G (blue curve in Fig. 3B) falling well below 1 for a broad range of R_0 . This is because the hubs in the network have relatively smaller degrees compared to the case when $\tau < 3$. They are too far away from each other to form direct connections, and thus the highest cores are localized around some of them as demonstrated in Fig. 2B and C. Although high-degree seed nodes in these cores should have an advantage to effec-

tively induce a larger outbreak, this effect is not strong enough to compensate for the disadvantage of starting the infection from a localized setup. Beyond localized cores, long-range interactions also have important effects on the network structure. Rare long-range connections (induced by higher α values) reduce the number of edges leaving the localized cores, which leads to networks with dominant local geometric structures (as shown in Fig. 2C). This makes it even harder for the infection to spread from a localized setup. Thus, for $\tau \geq 3$, increasing α enhances the danger of the random seeding scenario, as evident from Fig. 3B where the (red) curve with $\tau = 3.5$ and $\alpha = 2.3$ dives below 1 more than a similar curve with $\alpha = 1.3$. Finally, when $\tau \in (2, 3)$ (green curve in Fig. 3B), the pandemic size ratio f_G goes well above 1 for $R_0 \simeq 1$ values and goes barely below 1 for larger R_0 . In this case the highest-degree nodes are so dominant that they connect to each other even when they are spatially remote, and this way they induce a delocalized core (Fig. 2A). Simulations on these networks with delocalized cores resemble the phenomenon that is closest to our data-driven simulations (Fig. 3A) where the

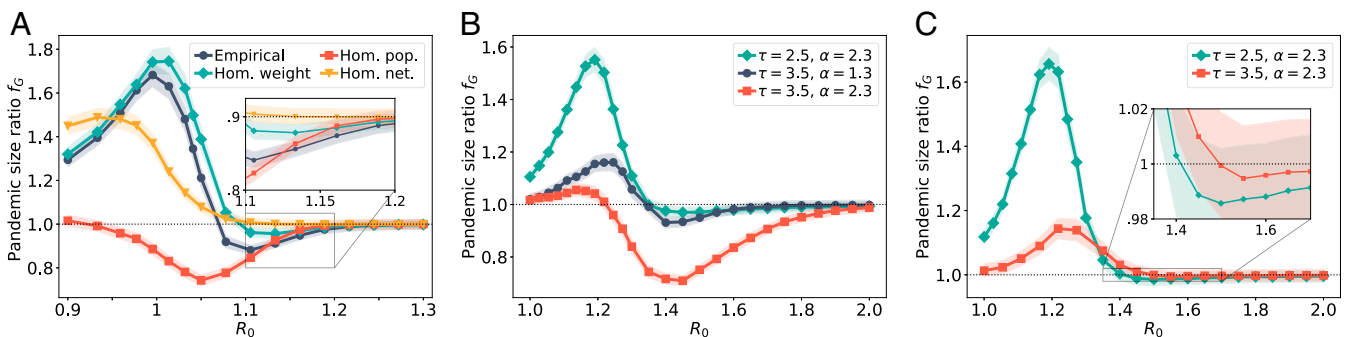


Fig. 3. The pandemic size ratio f_G as a function of R_0 . (A) Simulation results on the real commuting network of Hungary and its three homogenized versions as explained in the main text. Each data point is an average computed from 150 independent simulations, shown with 81% confidence interval. For each of them, initially $s = 97$ settlements are selected according to one of the seeding scenarios (central or uniform). Then, we infect $i_0 = 0.0005$ fraction of the 10^7 agents in the total population, and we distribute these agents in the s settlements uniformly at random, irrespective of the size of the settlements. (B) Pandemic size ratio computed on three geometric inhomogeneous random graph models corresponding to the three main universality classes with respect to graph distance. When $(\tau, \alpha) = (2.5, 2.3)$, the core is delocalized, and the switchover is weak. When $(\tau, \alpha) = (3.5, 1.3)$ or $(3.5, 2.3)$, the underlying geometry is more apparent, and the core is localized, so the parameter range for R_0 where random seeding is more dangerous ($f_G < 1$) is more spread out. In C, we see f_G on configuration model networks, where the switchover phenomenon appears weaker. For B and C the sizes of the metapopulation networks are $n = 1,000$ and each town is set with $n = 2,000$ individuals. Each pandemic size ratio data point is computed on 25 networks, with 35 simulations on each network, with $i_0 = 0.0005$ and $s = 30$. In all simulations we set $p_m = 0.001$, which means that 0.1% of the population moves in each iteration.

effects of the geometry are somewhat reduced due to the interconnectedness of larger cities all over a country. For $\tau \in (2, 3)$, the parameter α does not have a significant effect on the network structure.

For comparison, we also study the phenomenon on metanetworks sampled from the configuration model, a uniform distribution over networks with a given power-law degree sequence. The configuration model has no underlying geometry and features heterogeneity only in its degree distribution, parameterized by the exponent of the power-law distribution τ (details in *Materials and Methods*). For a fair comparison with the results obtained on GIRGs, we take $\tau = 2.5$ to obtain a configuration model with plenty of hubs and $\tau = 3.5$ for a model with reduced degree heterogeneity. These cases correspond to two different universality classes (in both GIRGs and configuration models) with respect to average distance (*Materials and Methods*). To keep the average degree and the number of nodes the same for the configuration model networks as in GIRGs, we obtain them by swapping randomly the links of the GIRG structures, while keeping the total number of connections for each node the same. Interestingly, when larger hubs are present in the structure (the case of $\tau = 2.5$ in Fig. 3C), the switchover phenomenon is recovered, even though the structure is fully uncorrelated. However, the switchover appears weaker, similar to the case in GIRGs where the effect of the geometry is suppressed due to the high interconnectedness of the network. We provide a heuristic explanation of this observation during the derivation of our theoretical results below. In summary, our simulation results demonstrate that while the emergence of the switchover phenomenon requires only degree heterogeneity in the network, it is certainly amplified by geometric correlations of the underlying structure.

Theoretical Results. To explain rigorously the switchover phenomena, we developed a mathematical framework relying on percolation theory.

Epidemics and percolation on metapopulation networks. The pandemic size (i.e., the final number of recovered individuals) of an SIR model with deterministic, unit recovery time (e.g., 1 d) on a (non-meta)network G has a useful connection with the commonly used simple mathematical framework of bond percolation. In such an SIR model, every edge of the network G transmits the disease at most once, when one endpoint is infected but the other is still susceptible. Equivalently, one may decide about every edge in advance, independently with probability p , whether it will do so. This is called retaining the edge, and p is then the retention probability of the model. The retained edges form the percolated random subgraph G^p of G . If a set S of nodes is selected as infected seeds in the network, then the epidemic will spread exactly over the connected components (also called clusters) of G^p that contain at least one node of S .

Metapopulation models are more difficult to treat mathematically, but a fundamental result in refs. 30 and 31 connects the behavior of SIR on metapopulation models to bond percolation. Following the arguments in refs. 30 and 31, once a large outbreak occurs in a town A , the proportion of infected people within the town concentrates around some $r_\infty \in (0, 1)$ (called local outbreak ratio). Infected people during the local pandemic carry the infection to a neighboring town B and cause a large outbreak there with a certain—computable—probability

$$p_{AB} = 1 - \exp\left(-\frac{Np_m w_{AB} r_\infty \left(1 - \frac{1}{R_0}\right)}{\mu}\right), \quad [1]$$

where N is the size of each population. Note that the dependence on A and B can be neglected if we assume an unweighted network. Since herd immunity is reached in each town after the

first large local epidemic outbreak of size $r_\infty N$, later infections in a town are no longer able to cause macroscopically visible outbreaks. Therefore, after time rescaling, the towns themselves go through an $S \rightarrow I \rightarrow R$ progression with unit recovery times and infection probability p . Consequently, the metapopulation model can be approximated by a simple SIR model on the network of towns and in turn with a bond percolation process with retention probability p .

The connection between metapopulation models and bond percolation allows us to understand the switchover phenomenon of the pandemic size ratio using a theoretical analysis of percolation cluster sizes, which have been extensively studied both in the mathematics and the physics literature for various network models, because they show a remarkable phase transition in the edge retention probability p . At a critical value p_c two phases are separated, where for $p < p_c$ all clusters are small, while for $p > p_c$ a single giant cluster emerges that contains a positive proportion of all nodes, while all other clusters are small. The critical parameter p_c depends only on the structure of the network G . For some networks, p_c can only be measured using numerical simulations. However, for the configuration model, the critical p_c can be explicitly computed, given the degree distribution of the network, as $p_c = \mathbb{E}[\deg(v)] / \mathbb{E}[\deg(v)(\deg(v) - 1)]$, which is asymptotically nonzero when $\tau > 3$. Using Eq. 1, the critical parameter p_c translates back to a critical basic reproduction number $R_c^{\text{glob}} > 1$ for the infection process. For $R_0 < 1$ the epidemic is subcritical already within a single town, while for $1 < R_0 < R_c^{\text{glob}}$ the epidemic is supercritical within towns but subcritical globally in the metanetwork (hence outbreaks containing only a few towns are possible). Finally, for $R_0 > R_c^{\text{glob}}$ the epidemic is supercritical in the entire network.

Beyond percolation cluster sizes, we also need to understand how the different seedings (central or uniform) interact with the clusters to explain the switchover phenomenon of the pandemic size ratio. Slightly deviating from the experimental setup, where central seeding corresponded to the highest core, here we define the central seeding set $\mathcal{CI}_0(s)$ as the s highest-degree nodes. This can be done as the two definitions are strongly correlated in the network models we focus on in this section (32–34). For the uniform seeding, just as earlier, we choose the seed set $\mathcal{UI}_0(s)$ as s nodes sampled uniformly at random in the whole network. In both setups we look at $\mathbb{E}_p[\mathbf{CI}(\mathcal{CI}_0(s))]$ and $\mathbb{E}_p[\mathbf{CI}(\mathcal{UI}_0(s))]$, the average percolation cluster sizes of the initially infected nodes, when edges are retained with probability p . This corresponds to the average number of populations that experience local large outbreaks in the two seeding scenarios. The percolation pandemic size ratio function is then defined as the ratio of these two averages,

$$f_G(p, s) = \mathbb{E}_p[\mathbf{CI}(\mathcal{CI}_0(s))] / \mathbb{E}_p[\mathbf{CI}(\mathcal{UI}_0(s))], \quad [2]$$

similar to the earlier defined experimental function.

We define two approaches to the switchover phenomenon on a metanetwork of n cities. In the weak switchover phenomenon, we require that there exists a seed count $s \leq n$ and link-retention probabilities $0 < p_1, p_2 < 1$ with

$$f_G(p_1, s) > 1 + c, \quad \text{and} \quad f_G(p_2, s) < 1 - c, \quad [3]$$

for some constant c that might depend on the network size. Meanwhile, in the strong switchover phenomenon, we require that the constant c does not depend on the network size n and thus holds across a whole model class (e.g., GIRG or configuration model with fixed-degree heterogeneity). When the switchover occurs for a seed count s , we say that the switch happens at retention probability p_{switch} if $f_G(p, s) > 1$ for $p < p_{\text{switch}}$, while $f_G(p, s) < 1$ for $p > p_{\text{switch}}$.

While the switchover phenomenon in GIRGs is hard to study analytically due to the lack of percolation theory developed for this model, we borrow concepts from a simpler conventional network model, called the stochastic block model (SBM), to observe the strong switchover phenomenon. The SBM is able to mimic the central and rural areas of a town network, since it contains a “hidden geometry”: We group towns into two sets of central or rural areas. Within areas we assume geometric random networks; i.e., each pair of nodes is connected with the same probability, while the edge density between the two areas is lower.

Theorem 1. *In the stochastic block model with appropriately scaled parameters and $s_n = \Theta(n)$ the strong switchover phenomenon happens. (For a proof see SI Appendix.)*

In the case of the geometric configuration model, we are able to prove the weak switchover, already observed experimentally in Fig. 3C. Further, we are able to give quantitative bounds on c in Eq. 3 as a function of the size n of the population network G , the parameter τ expressing the prevalence of hubs, and the initial seed number $s = s_n$ that may also depend on the network size.

Theorem 2. *In the configuration model with exponent $\tau \in (2, 4)$ and $1 \ll s_n \ll n$ the weak switchover phenomenon appears with p_{switch} slightly above the critical percolation parameter p_c . (For a proof see SI Appendix.)*

While Theorem 2 is valid for $\tau \in (2, 4)$, the two regimes $\tau \in (2, 3)$ and $\tau \in (3, 4)$ quantitatively differ. In the former case, also called the scale-free regime, p_c tends to zero as the network size grows and the region of the parameter space where seeding from nodes selected uniformly randomly is more dangerous is described by different linear equations compared to the $\tau \in (3, 4)$ case. The switchover phenomenon disappears when $\tau > 4$ as hubs become too small and separated from each other to produce the desired effect.

An interpretation of our rigorous derivations yields the following heuristic explanation of the switchover phenomenon:

- 1) Below the percolation threshold, as demonstrated in Fig. 4A, the connected components of the central area seed nodes (indicated as red nodes in Fig. 4A) will be much larger than the components of the uniformly randomly selected seed nodes; however, they do not yet form a giant component. Nodes selected uniformly at random (blue nodes in Fig. 4A) are not likely to be in these large components; hence the union of the connected components of seeds selected uniformly at random from the network will be smaller than the pandemic that started from the central area. For very small values of p , seeding the highest-degree nodes is the most dangerous on every graph.
- 2) Slightly above the percolation threshold, there will be a single “giant component” of nodes experiencing a local pandemic, containing most of the central nodes. Thus, when seeding starts from the central nodes, their components will contain this giant component and a small portion of smaller components. In the uniform seeding scenario, it is likely that a few of the random seed nodes will also belong to the giant component, while the other seeds, spread out randomly over the rest of the network, will be contained in lots of additional smaller components. Hence the union of the components of the uniformly chosen nodes will be larger, as shown in Fig. 4B.
- 3) Well above the percolation threshold (Fig. 4C), there is essentially only one connected component; thus each node gets infected regardless of the position of the seed nodes in the network.

The phenomenon in explanation 2 is stronger when there are relatively few edges leaving the central area, which can be due to lack of long-range interactions amplifying local geometric effects (as observed in GIRGs) (Fig. 3). However, in Theorem 1 the geometry induced by the two blocks is already enough to cause a strong switchover. On the contrary, there is nothing to limit the number of edges leaving the central area in the configuration model (the degree-degree correlation coefficient is close to 0) (35–37); hence, the switchover phenomenon appears weak.

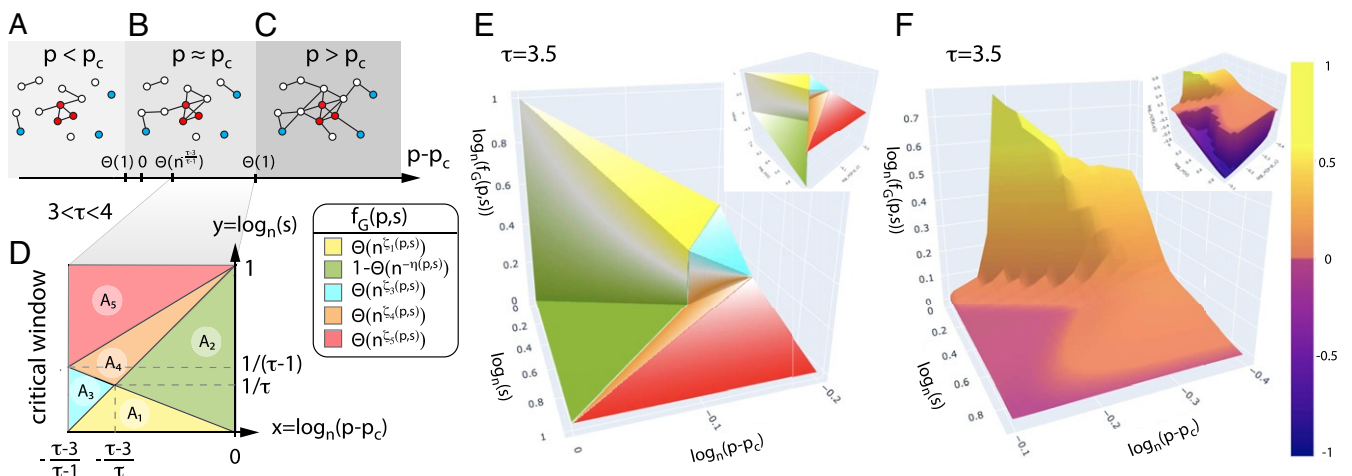


Fig. 4. A–C show the heuristic explanation of the switchover of the pandemic size ratio function f_G . D shows the phase diagram of the function f_G for $3 < \tau < 4$, for values of p slightly above the percolation threshold and for various values of s . The asymptotic of f_G is different in parameter regions with different colors. The precise values of C_i in the key in D are included in *Materials and Methods*, and the phase diagram for $2 < \tau < 3$ is included in *SI Appendix*. E shows the three-dimensional plot the limit function $\log_n(f_G)$ for $\tau = 3.5$, as the number of nodes in network G tends to infinity, and F shows the corresponding simulation results on configuration model networks with $n = 10^7$ nodes (each datapoint is an average of 1,000 independent percolation instances on 10 independent random networks, after outlier removal is performed as explained in *SI Appendix*). The coloring in E follows the coloring on the phase diagram in D. Since in the configuration model we have only weak switchover, the green part of the surface $\log_n(f_G)$ (which corresponds to $f_G < 1$) converges to 0. For a visualization of the precise deviation of $\log_n(f_G)$ below 0, in E and F, *Insets*, we plot the function $\log_n(f_G)$ for $f_G > 1$ and $-\log_n(1 - f_G) - 1$ for $f_G < 1$.

Quantitative results for the configuration model. For geometric networks with various node degree distributions, critical exponents have been already proposed earlier (38, 39), with some of them proved rigorously for the configuration model (possibly with power-law degree distribution) (40–43), as well as for rank-1 inhomogeneous random graphs (44–46) and for Erdős–Rényi graphs (47). Based on these results, we can prove that, after appropriate scaling, the pandemic size ratio f_G of the configuration model (simulated in Fig. 3C for fixed s) converges to a two-dimensional limit function, which can be precisely determined. To state our result, let us reparameterize $f_G(p, s)$ as a function of x, y , where $p = p_c + n^x$ for $x \in (-(\tau - 3)/(\tau - 1), 0)$ and $s = s_n = n^y$ for $y \in (0, 1)$; i.e., we consider $f_G(x, y) := f_G(p_c + n^x, n^y)$. For $\tau \in (3, 4)$, we divide the parameter space into five triangular regions A_1 to A_5 illustrated in Fig. 4D (defined precisely in *SI Appendix*). For $\tau \in (2, 3)$, the picture is similar, except there is an extra triangular region A_6 .

Theorem 3. *In the configuration model with exponent $\tau \in (2, 4)$, $\log_n(f_G(x, y))$ converges to a function $\zeta(x, y)$. In each triangular region A_i , $\zeta(x, y)$ can be expressed as a different linear function $\zeta_i(x, y)$.*

Theorem 3 implies that the percolation pandemic size ratio $f_G(p_c + n^x, n^y) = \Theta(n^{\zeta_i(x, y)})$ on A_i for each $i = 1$ to 5. We give the formula for each ζ_i in *Materials and Methods* and the proof of Theorem 3 in *SI Appendix*. A three-dimensional illustration of ζ can be seen in Fig. 4E. The limiting function ζ is discontinuous at the boundary line between A_1 and A_2 and between A_3 and A_4 , respectively. These discontinuities correspond to a discontinuous phase transition of the system's behavior at those boundaries. Curves in Fig. 3 correspond to horizontal cross-sections of the two-dimensional f_G function for fixed s values. We experience this phase transition on the curves of Fig. 3 dropping steeply from above 1 to below 1 when slightly increasing R_0 . Our result implies that the curves will look steeper and steeper as n , the size of the network, increases.

We also show that $f_G(p_c + n^x, n^y) = 1 - \Theta(n^{\eta(x, y)})$ in region A_2 . Hence, the scaling $\log_n(f_G)$ in Theorem 3 is not appropriate in region A_2 , which is reflected by the fact that the limiting function ζ_2 is identically 0 in this regime. To be able to compute how much f_G (Eq. 2) goes below 1 in A_2 , i.e., how much more dangerous uniform seeding can be compared to central seeding, we extract the limiting exponent $\eta(x, y)$ using a different normalization for f_G and give the formula in *Materials and Methods*. This different normalization is used on the area A_2 (green) in Fig. 4E, *Inset*, which in turn demonstrates that f_G falls below 1 in this regime. Finally, we validate our theoretical results for the configuration model by simulations in Fig. 4F (48). Despite the finite size of the simulations ($n = 5 \times 10^6$), the resemblance to the theoretical predictions is already apparent.

Discussion and Conclusions

Different seeding of an epidemic can lead to significantly different outcomes, depending on the actual value of the basic reproduction number R_0 . While R_0 is defined by the biological parameters of the spreading disease, it is only one factor determining the effective reproduction rate R_t , which characterizes the actual speed of reproduction during an ongoing epidemic. In the case of an influenza-like disease, R_t depends on many other factors, including the actual number of interactions of people, the actual interventions, the self-protection measures (e.g., masks, sanitizing, etc.), or even the seasonal variance of temperature and humidity. Considering the distribution of the initial seeds and the actual R_t values, our theory may suggest counterintuitive effects during the consecutive waves of a pandemic. This could be the case in Hungary, where the first wave of the COVID-19 pandemic was initiated from large, well-connected

towns, while social distancing was very effective at the time, causing a smaller actual R_t value during this period. Thus, the clumped initial seeds and the somewhat low R_t could set relatively favorable conditions for the epidemic to reach a larger population, compared to a uniformly seeded situation. Meanwhile, at the beginning of the second wave, seeded towns were more distributed all around the country, while social distancing was not followed rigorously. This induced larger R_t values, which yet again set relatively easier conditions for the epidemic to reach a larger population, now seeded from a uniform initial state.

In this paper we studied the effects of epidemic seeding on geometric metapopulation networks. We were interested in the long-term behavior of spreading processes and showed that the relative danger of infecting a larger population when starting the process from the core or uniformly at random in a network has a nonmonotonous dependency on R_0 . We explored a switchover phenomenon and demonstrated it on real and synthetic networks via numerical simulations. We provided a rigorous proof for the existence of this phenomenon on a large set of random graphs, while we are confident that our theory can be extended to a more general set of graphs, which resembles certain structural constraints. Importantly, we identified the spatial geometry of the underlying structure as an important amplifying factor of the switchover phenomenon.

We build our theory on some results (38, 39), which are broadly accepted by the network science community, yet it has not been proved rigorously for all network structure (for exceptions see *SI Appendix*). This implies certain limitations for our results, although assuming these results to hold, our proposed theory has been derived rigorously. In addition, we made some assumptions for the simplicity of our presentation but their generalization is possible. We demonstrated experimentally the switchover phenomenon on directed networks, while we assume undirected structures in our theory, which can be extended for directed structures easily. Moreover, we conjecture that the observed phenomenon occurs in most networks where a “central” region can be meaningfully distinguished in the structure. While metapopulation networks are generally used to model spreading phenomena at global spatial scales where towns are well separable, they provide a useful tool to study epidemics at shorter spatial distances too (49). In our case, we applied metapopulation networks on the country level but we carefully separated the scales of flows inside and between towns (by setting $p_m = 0.001$), this way evidently separating towns from each other. We concentrated on the conventional SIR model for the demonstration of the switchover phenomenon, but this observation holds for more realistic models, including the SEIR model with an additional compartment of E state, better capturing the reaction scheme of the SARS-Cov-2 disease. Our goal in this paper was to identify, verify, and mathematically prove the existence of the observed switchover phenomenon, not to provide predictions. Accordingly, we worked with the simplest models and ignored the effects of possible interventions, seasonal weather conditions, superinfection events, permanent residence changes, etc. It indicates the fundamental nature of our observations, that surprisingly they occurred even under these simplified circumstances.

Beyond scientific merit, our results may contribute to the better designs of epidemic forecasts and intervention strategies in a country during an ongoing pandemic. We highlight the importance to follow not only the rate but also the spatial distribution of new infection cases of a spreading disease or its variants during the early phase of an epidemic. This could lead to inventive testing strategies, which disclose the spatial distribution of the epidemic during its initial phase, as this was the case in some countries (like Denmark) (50) from the beginning of the COVID-19 pandemic. Based on these early-time observations our theory provides understanding about the long-term

consequences of an epidemic by considering the commonly overlooked convoluted effects of epidemic seeding and the geometric structure of human populations and mobility.

Materials and Methods

Data Description.

Settlement-level daily COVID-19 infection data for Hungary. For the analysis presented in Fig. 1 we used a dataset recording the daily number of newly infected cases in 3,118 Hungarian settlements. These data match the officially reported total number of daily cases (51, 52); however, just as in the official data, they suffer from some observational bias due to the limited capacity of testing in the country during certain periods of the pandemic. For the analysis presented in Fig. 1 we considered all settlements and obtained their population sizes from data shared by the Hungarian Statistical Office (53). A version of these data aggregated on the county level is openly available (52).

Daily commuting network of Hungary. For the data-driven simulations of the Hungarian epidemic we use a microcensus collected and released by the Hungarian Statistical Office in 2016 (20). The data contain the number of people commuting to work or school on a daily basis between the 3,186 settlements in Hungary, with the districts of the capital considered as separate towns. In our analysis we concentrated only on settlements with populations larger than 1,000 inhabitants and kept commuting links with at least 25 daily commuters. From these data we constructed an undirected metapopulation commuting network with 1,398 settlements as nodes (of which 97 were from the capital and its suburbs) and 8,322 commuting edges with weights computed as the average number of commuters between pairs of towns. The total population size of the network contained 95% (9,285,286 individuals) of the Hungarian population. Despite the sparsity of the network (0.85% of the possible edges are present), 19% of individuals commute between settlements on a daily base.

Moran's I Statistic. We compute Moran's I statistic at time t as

$$I(t) = \frac{n \sum_{ij} w_{ij} (y_i(t) - \bar{y}(t))(y_j(t) - \bar{y}(t))}{\sum_{ij} w_{ij} \sum_i (y_i(t) - \bar{y}(t))^2}, \quad [4]$$

where n is the number of nodes, w_{ij} is the edge weight between the nodes i and j , $y_i(t)$ is the number of new infected cases at node i at time t , and $\bar{y}(t) = (1/n) \cdot \sum_i y_i(t)$.

Generating Geometric Inhomogeneous Random Graphs. GIRG(τ, α) networks were generated by the following process: The locations of n nodes are sampled uniformly at random from the square $[0, 1]^2$, and each node u is assigned a "fitness" value (w_u) sampled from a power-law distribution with exponent τ . Each pair of nodes is connected by an edge with a probability

$$P(u, v) = p \min \left\{ \left(\frac{C w_u w_v}{n \|x_u - x_v\|^2} \right)^\alpha, 1 \right\}, \quad [5]$$

which after only the largest connected component of the network is kept. To generate models with different parameters comparable to each other, we fix the number of edges to 5,000, by selecting the constant C and p accordingly, since these two parameters are responsible for the edge density. For the exact implementation see ref. 54. When the fitness distribution w_u is set to be a power law, node degrees also satisfy a power law. The abundance of long-range connections is tuned by α in Eq. 5: The smaller α is, the more likely are long-range connections. The power-law exponent τ and the long-range parameter α tune the average graph distance in the network $\overline{\text{Dist}}(n)$ (55–58):

$$\overline{\text{Dist}}(n) = \begin{cases} \Theta(\log \log n) & \text{when } \tau \in (2, 3), \alpha > 1 \\ \Theta(\text{polylog } n) & \text{when } \tau > 3, \alpha \in (1, 2) \\ \Theta(\sqrt{n}) & \text{when } \tau > 3, \alpha > 2. \end{cases}$$

Comparing this to the average distance in the configuration model, where only the first two regimes are possible [$\Theta(\log \log n)$ when $\tau \in (2, 3)$, $\Theta(\log n)$ when $\tau > 3$], and to distances in lattice models (where $\overline{\text{Dist}}_n$ is polynomial in n), we observe that the underlying geometry of GIRGs with the long-range connections play a role when $\tau > 3$, and the model interpolates between the small-world configuration model and the lattice.

Generating Random Networks from the Configuration Model. We generate a uniform sample from the set of graphs with power-law degree distribution with degree exponent τ by first generating a GIRG with given parameter τ (and $\alpha = 2.3$), and we swap the endpoints of randomly selected

pairs of edges (59) to remove all geometric and structural correlations from the structure, while conserving the degree of each node. We perform $10 \times \text{number_of_edges}$ swaps, which mixes the edges enough so that the resulting network becomes close to a uniform sample from the set of networks that have exactly the same degree sequence as the original GIRG network (60).

Core Decomposition and Seed Selection. In metapopulation network models, we use k-shell decomposition to identify the largest k-core of the network (61, 62) and to select seeds in the central area. This algorithm computes the k-shell by recursively removing each node of the network that has degree less than k , until no more nodes can be removed. We take the largest k for which at least s nodes remain, and we select s nodes from this k-shell uniformly at random as our seed set in the central area. For the uniform seeding scenario, we select s nodes of the network uniformly at random. Finally, we infect $f_0 = 0.0005$ fraction of the agents in the total population, and we distribute these agents in the s settlements uniformly at random, irrespective of the size of the settlements.

In the theoretic computations and simulations, the s highest-degree nodes are selected for the central area, and s uniformly random nodes are selected for the uniform seeding scenario. Node degrees and core number of nodes in configuration network models are strongly correlated, allowing us to make this approximation.

SIR Model on Metapopulation Networks. To make our simulations somewhat realistic, we set the hometown of each of N_i agents to the agent's initial settlement i . Each agent is assigned exactly one hometown, and the home assignments do not change for the rest of the simulation. We initialize the infection according to one of the seed selection scenarios and proceed with the simulation in each iteration t in three steps. In the diffusion step, each agent who is at its hometown i is selected to move to another town with probability p_m . The selected agents then choose a target town j with probability proportional to the weight w_{ij} of the link connecting town i to j and move there. We set $p_m = 0.001$ in all simulations, which means that 0.1% of the total population moves in each iteration.

Agents that are not at their hometown simply move back to their home settlement. In the reaction step, each susceptible agent in town i becomes infected with probability $1 - (1 - \beta/N_i)^{I_i}$, where I_i is the number of infected agents in town i at iteration step t and β is the infection rate. In the final recovery step, each infected agent recovers with rate μ . For the exact implementation see ref. 48.

The Limiting Function of the Percolation Pandemic Size Ratio. In *Theorem 3*, we identified the scaling of $f_G(p, s) = f_G(p_c + n^x, n^y) = \Theta(n^\zeta(x, y))$, where $\zeta(x, y)$ is a piecewise linear function. On each region A_1 to A_6 , ζ is given as follows:

$$\zeta(x, y) = \begin{cases} \zeta_1(x, y) = 1 + \left(\frac{1}{|\tau-3|} + \mathbb{1}_{\tau \in (3,4)} \right) x - y & \text{on } A_1 \\ \zeta_2(x, y) = 0 & \text{on } A_2 \\ \zeta_3(x, y) = \mathbb{1}_{\tau \in (3,4)} x + \left(1 - \frac{1}{\tau-1} \right) (1-y) & \text{on } A_3 \cup A_6 \\ \zeta_4(x, y) = -\frac{1}{|\tau-3|} x - \frac{1}{\tau-1} (1-y) & \text{on } A_4 \\ \zeta_5(x, y) = \frac{1}{(\tau-1)(\tau-2)} (1-y) & \text{on } A_5. \end{cases}$$

Finally, on region A_2 , $f_G(p_c + n^x, n^y) = 1 - \Theta(n^{-\eta(x,y)})$, where

$$\eta(x, y) = \left(\frac{1}{|\tau-3|} + \mathbb{1}_{\tau \in (3,4)} \right) x - y.$$

Data Availability Previously published data were used for this work (50–52).

ACKNOWLEDGMENTS. We are grateful for the insightful comments by Miklós Abert, Géza Ódor, and Tim Hulshof and for the shared datasets by János Köllő, Péter Pollner, Miklós Szócska, Beatrix Oroszi, Gergely Röst, and the Hungarian COVID-19 Mathematical Modelling and Epidemiological Analysis Task Force. This work has been supported by the Dynasnet European Research Council Synergy project (ERC-2018-SYG 810115). G.Ó. acknowledges support from the Swiss National Science Foundation (200021-182407); D.C. received support from Hungarian National Artificial Intelligence Laboratory Grant 2018-1.2.1-NKP-00008; and M.K. was supported by the DataRedux French National Research Agency project (ANR-19-CE46-0008), the SoBigData++ H2020 project (H2020-871042), and the Hungarian National Development, Research, and Innovation Fund (NKFH 2020-2.1.1-ED-2020-00003).

1. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
2. L. Stone, R. Olinky, A. Huppert, Seasonal dynamics of recurrent epidemics. *Nature* **446**, 533–536 (2007).
3. S. Funk, E. Gilad, C. Watkins, V. A. Jansen, The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6872–6877 (2009).
4. Z. A. Memish, N. Aljerian, S. H. Ebrahim, Tale of three seeding patterns of SARS-CoV-2 in Saudi Arabia. *Lancet Infect. Dis.* **21**, 26–27 (2021).
5. V. Colizza, A. Vespignani, Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.* **99**, 148701 (2007).
6. A. Apolloni, C. Poletto, J. J. Ramasco, P. Jansen, V. Colizza, Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theor. Biol. Med. Model.* **11**, 3 (2014).
7. M. Kitsak *et al.*, Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
8. P. Crépey, F. P. Alvarez, M. Barthélemy, Epidemic variability in complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **73**, 046131 (2006).
9. A. Gautreau, A. Barrat, M. Barthélemy, Arrival time statistics in global disease spread. *J. Stat. Mech.* **2007**, L09001 (2007).
10. J. M. Brauner *et al.*, Inferring the effectiveness of government interventions against COVID-19. *Science* **371**, eabd9338 (2021).
11. C. Dye, R. C. H. Cheng, J. S. Dagpunar, B. G. Williams, The scale and dynamics of COVID-19 epidemics across Europe. *R. Soc. Open Sci.* **7**, 201726 (2020).
12. E. R. White, L. Hébert-Dufresne, State-level variation of initial COVID-19 dynamics in the United States. *PLoS One* **15**, e0240648 (2020).
13. C. Merow, M. C. Urban, Seasonality and uncertainty in global COVID-19 growth rates. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 27456–27464 (2020).
14. S. Reicher, J. Drury, Pandemic fatigue? How adherence to covid-19 regulations has been misrepresented and why it matters. *BMJ* **372**, n137 (2021).
15. J. K. Burton *et al.*, Evolution and effects of COVID-19 outbreaks in care homes: A population analysis in 189 care homes in one geographical region of the UK. *Lancet Healthy Longev.* **1**, e21–e31 (2020).
16. A. Aleta *et al.*, Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
17. M. U. G. Kraemer *et al.*, Open COVID-19 Data Working Group, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
18. P. Bajardi *et al.*, Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS One* **6**, e16591 (2011).
19. M. Szocska *et al.*, Countrywide population movement monitoring using mobile devices generated (big) data during the COVID-19 crisis. *Sci. Rep.* **11**, 5943 (2021).
20. Hungarian Central Statistical Office, Microcensus 2016: Population enumeration. <https://www.ksh.hu/mikrocenzus2016/?lang=en>. Accessed 2 January 2021.
21. G. Röst *et al.*, Early phase of the covid-19 outbreak in Hungary and post-lockdown scenarios. *Viruses* **12**, 708 (2020).
22. M. Karsai, J. Koltai, O. Vásárhelyi, G. Röst, Hungary in mask/maszk in Hungary. *Corvinus J. Sociol. Soc. Policy* **11**, 139–146 (2020).
23. J. R. Fauver *et al.*, Coast-to-coast spread of SARS-Cov-2 during the early epidemic in the United States. *Cell* **181**, 990–996.e5 (2020).
24. D. Kang, H. Choi, J. H. Kim, J. Choi, Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int. J. Infect. Dis.* **94**, 96–102 (2020).
25. V. Colizza, R. Pastor-Satorras, A. Vespignani, Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276–282 (2007).
26. A. Barrat, M. Barthélemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
27. Wikipedia, Central Hungary. https://en.wikipedia.org/wiki/Central_Hungary. Accessed 17 May 2021.
28. M. Newman, *Networks* (Oxford University Press, 2018).
29. K. Bringmann, R. Keusch, J. Lengler, Geometric inhomogeneous random graphs. *Theor. Comput. Sci.* **760**, 35–54 (2019).
30. M. Barthélemy, C. Godrèche, J. M. Luck, Fluctuation effects in metapopulation models: Percolation and pandemic threshold. *J. Theor. Biol.* **267**, 554–564 (2010).
31. V. Colizza, A. Vespignani, Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *J. Theor. Biol.* **251**, 450–467 (2008).
32. D. Fernholz, V. Ramachandran, The giant k-core of a random graph with a specified degree sequence. CiteSeerX [Preprint] (2003). <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.642.2499> (Accessed 29 September 2021).
33. S. Janson, M. J. Luczak, A simple solution to the k-core problem. *Random Struct. Algorithms* **30**, 50–62 (2007).
34. T. Łuczak, Size and connectivity of the k-core of a random graph. *Discrete Math.* **91**, 61–68 (1991).
35. N. Litvak, R. van der Hofstad, Uncovering disassortativity in large scale-free networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **87**, 022801 (2013).
36. R. Van Der Hofstad, N. Litvak, Degree-degree dependencies in random graphs with heavy-tailed degrees. *Internet Math.* **10**, 287–334 (2014).
37. W. L. F. van der Hoorn, N. Litvak, Degree-degree dependencies in directed networks with heavy-tailed degrees. *Internet Math.* **11**, 155–179 (2015).
38. M. E. Newman, S. H. Strogatz, D. J. Watts, Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **64**, 026118 (2001).
39. R. Cohen, D. Ben-Avraham, S. Havlin, Percolation critical exponents in scale-free networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **66**, 036113 (2002).
40. S. Dhara, R. van der Hofstad, J. S. van Leeuwen, Critical percolation on scale-free random graphs: New universality class for the configuration model. *Commun. Math. Phys.* **1**, 1–49 (2021).
41. S. Dhara, R. van der Hofstad, J. S. van Leeuwen, S. Sen, Heavy-tailed configuration models at criticality. *Ann. I. H. P. Probab. Stat.* **56**, 1515–1558 (2020).
42. S. Dhara, R. van der Hofstad, J. van Leeuwen, S. Sen, Critical window for the configuration model: Finite third moment degrees. *Electron. J. Probab.* **22**, 1–33 (2017).
43. R. van der Hofstad, S. Janson, M. Łuczak, Component structure of the configuration model: Barely supercritical case. *Random Struct. Algorithms* **55**, 3–55 (2019).
44. S. Bhamidi, R. van der Hofstad, J. S. van Leeuwen, Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.* **15**, 1682–1702 (2010).
45. S. Bhamidi *et al.*, Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.* **40**, 2299–2361 (2012).
46. R. van der Hofstad, S. Kliem, J. S. H. van Leeuwen, Cluster tails for critical power-law inhomogeneous random graphs. *J. Stat. Phys.* **171**, 38–95 (2018).
47. J. Ding, J. H. Kim, E. Lubetzky, Y. Peres, Anatomy of a young giant component in the random graph. *Random Structures Algorithms* **39**, 139–178 (2011).
48. D. Czirfa, Data from “Epidemic modelling with metapopulation and percolation models, based on the <https://arxiv.org/abs/2106.16070> paper.” GitHub. https://github.com/dczirfa/epidemic_seeding. Accessed 8 August 2021.
49. N. Gozzi *et al.*, Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nat. Commun.* **12**, 2429 (2021).
50. T. L. Holmager, E. Lynge, C. E. Kann, G. St-Martin, Geography of COVID-19 in Denmark. *Scand. J. Public Health* **49**, 88–95 (2020).
51. Daily new covid-19 infections in Hungary. <https://koronavirus.gov.hu>. Accessed 17 May 2021.
52. Wikipedia, Covid-19 corona virus pandemic in Hungary. https://hu.wikipedia.org/wiki/Covid19-koronavirus-járvány_Magyarországon. Accessed 17 May 2021.
53. Hungarian Statistical Office, Settlement and population registry. https://www.ksh.hu/apps/hnr.egyeb?p_lang=EN&p_sablon=LETOLTES. Accessed 1 April 2020.
54. J. Jorritsma, Data from “Random graph visualization.” GitHub. <https://github.com/joostjor/random-graphs/blob/master/girg.py>. Accessed 4 January 2021.
55. M. Deijfen, R. van der Hofstad, G. Hooghiemstra, Scale-free percolation. *Ann. I. H. P. Probab. Stat.* **49**, 817–838 (2013).
56. K. Bringmann, R. Keusch, J. Lengler, Average distance in a general class of scale-free networks with underlying geometry. arXiv [Preprint] (2016). <https://arxiv.org/abs/1602.05712> (Accessed 29 September 2021).
57. M. Biskup *et al.*, On the scaling of the chemical distance in long-range percolation models. *Ann. Probab.* **32**, 2938–2977 (2004).
58. P. Deprez, R. S. Hazra, M. V. Wüthrich, Inhomogeneous long-range percolation for real-life network modeling. *Risks* **3**, 1–23 (2015).
59. Networkx, Connected double edge swap function. https://networkx.org/documentation/networkx-1.9/reference/generated/networkx.algorithms.swap.connected_double_edge_swap.html. Accessed 13 October 2020.
60. C. Gkantsidis, M. Mihail, E. W. Zegura, “The Markov chain simulation method for generating connected power law random graphs” in *Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments* (SIAM, 2003), pp. 16–25.
61. S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavit, E. Shir, A model of Internet topology using k-shell decomposition. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11150–11154 (2007).
62. Networkx, K-shell decomposition. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.core.k_shell.html. Accessed 13 October 2020.