# Longer-term seeding effects on epidemic processes: a network approach

**Gergely Ódor[1], Domonkos Czifra[2], Júlia Komjáthy[3], László Lovász[2,4,*], Márton Karsai[2,5]**

[1]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[2]Alfréd Rényi Institute of Mathematics, Budapest, Hungary
[3]Eindhoven University of Technology, Eindhoven, The Netherlands
[4]Eötvös Loránd University, Budapest, Hungary
[5]Department of Network and Data Science, Central European University, Vienna, Austria
*Correspondence should be addressed to László Lovász (laszlo.lovasz@ttk.elte.hu)

Received: 11 December 2021; Accepted: 18 January 2022

## Summary

In this paper we touch upon three phenomena observed in real life as well as in simulations; in one case, we state mathematical results about the appearance of the phenomenon on arbitrary graphs (networks) under rather general conditions. We discuss a phenomenon of critical fluctuations, demonstrating that an epidemic can behave very differently even if it runs on the same network, with the same transmission probabilities and started from the same initial seeds. We explore a connection between the geographic distribution and intensity of the spreading epidemic. We argue that the speed of the spread of an epidemic depends not only on the number of current infections, but also on their geographic distribution over a country. Through the observations of these phenomena we suggest a dependence of the final epidemic size on the geometric position of initial seeds of an epidemic process.

**Keywords:** epidemic models, geometric networks, epidemic seeding, percolation, switchover phenomenon

# A korai fertőzések helyzetének hosszú távú hatásai a járványterjedésre: egy hálózati megközelítés

**Ódor Gergely[1], Czifra Domonkos[2], Komjáthy Júlia[3], Lovász László[2,4,*], Karsai Márton[2,5]**

[1]École Polytechnique Fédérale de Lausanne, Lausanne, Svájc
[2]Rényi Alfréd Matematikai Kutatóintézet, Budapest, Magyarország
[3]Eindhoven University of Technology, Eindhoven, Hollandia
[4]Eötvös Loránd Tudományegyetem, Budapest, Magyarország
[5]Central European University, Bécs, Ausztria
*Levelező szerző: Lovász László (laszlo.lovasz@ttk.elte.hu)

## Összefoglalás

A 2020. március és 2021. február közötti hazai COVID–19 járványügyi adatokat felhasználva a járványok geográfiai terjedését kutatjuk. Alapvető modellünk az, hogy a járvány azon emberek között terjed, akik mobilitási mintáik alapján egy városban tartózkodnak, így van esély arra, hogy találkozhatnak egymással. Ezt a hálózatot úgy közelítjük, hogy tekintjük az 1000 fő feletti települések hálózatát (gráfját), ahol a települések közötti élek súlya a közöttük becsült forgalomból, elsősorban ingázásból nyerhető. Az egyes településeken belül feltételezzük, hogy minden ember ugyanakkora valószínűséggel találkozhat. A településeken belül és között átlagolással (meanfield) kapjuk a terjedés paramétereit. Három kérdést tanulmányozunk a cikkben. Az első kettőben többé-kevésbé a várakozásnak megfelelőek az eredmények, a harmadik azonban meglepetéssel szolgált. Mennyire jelezhető előre a járvány lefutása? Szimulációink alapján az $R_0 = 1$ érték közelében a helyzet némileg a meteorológiai előrejelzésekhez hasonló, azonos hálózaton, azonos kezdeti fertőzésből, azonos paraméterekkel is nagyon eltérő járványgörbék keletkezhetnek. Mennyire befolyásolja a napi fertőzésszámot a járvány elterjedtsége, vagyis az, hogy mennyire oszlik meg a fertőzésszám a lakossággal arányosan az egyes településeken (járásokban, megyékben)? Szimulációink

egyértelműen mutatják, hogy a fertőzöttek azonos száma mellett, minél egyenletesebb a fertőzöttek eloszlása az országban, annál nagyobb a napi új fertőzések száma. Egy járvány első, korai esetei (pl. külföldről behozott fertőzés) hogyan befolyásolják a terjedés végkimenetelét, a teljes lefutás alatti megbetegedések számát? A kutatás egyik fontos célja az, hogy megtalálja a kezdeti fertőzöttek azon konfigurációját, amely a legnagyobb járványhoz vezet. Ezzel kapcsolatban egy új jelenséget figyeltünk meg. Az általánosan elfogadott kép szerint, ha a vizsgált betegség egy elszigetelt, alacsony népsűrűségű és nem túl sűrű tömegközlekedési kapcsolattal rendelkező településen jelenik meg először, akkor gyorsan kipusztulhat anélkül, hogy nagyobb járványt okozna. Egészen más lehet a dinamika, ha a betegség egy jól összekapcsolt, sűrűn lakott településről indul, ahol könnyebben túlélhet és terjedhet szét a lakosság többi részére. A kutatás során ezt a feltevést megkérdőjeleztük, és demonstráljuk, hogy az a járvány, ami a városok hálózatának legszorosabban összekapcsolt részéből indul, hosszú távon nem mindig vezet több fertőzötthöz. Ha a betegségnek nagy a fertőzőképessége, akkor egy járvány, ami véletlenszerűen kiválasztott városokból indul, akár nagyobb populációt is megfertőzhet. Eredményeinket magyar mobilitási és járványügyi adatok, szimulációk, illetve matematikai bizonyítások is alátámasztják. A tudományos megállapításaikon túlmenően eredményeink hozzájárulhatnak járvány-előrejelzések és az intervenciós stratégiák jobb megtervezéséhez egy adott országban egy éppen zajló világjárvány idején. Arra világítanak rá, hogy egy járvány korai szakaszában nem csak a terjedő betegség fertőzési eseteinek számát fontos követni, hanem a fertőzések geográfiai eloszlása is fontos információt hordoz. A most kidolgozott modell ilyen korai megfigyelésekből kiindulva, egy ország populációjának földrajzi szerkezetét, valamint a mobilitási hálózatának hatásait is figyelembe veszi és már a járvány kezdeti szakaszában segíthet hosszabb távú következtetések levonásában.

**Kulcsszavak:** járványmodellezés, geometriai hálózatok, korai járvány helyzete, perkoláció, átváltó jelenségek

## Introduction

Several aspects of a society can be represented as a network *(Newman 2018; Vega-Redondo 2007)*. Best examples are social networks, where single individuals can be identified as nodes connected by social interactions *(Wasserman–Faust 1994)*, or commuting networks, where nodes are places or populations (towns) connected by mobility links like local transportation or commuting *(Barthélemy 2011)*. Although these network representations describe a social system in rather different aggregation levels, they both serve as the structure underpinning the emergence of various global phenomena. In the case of epidemics, the transmission of a disease is usually assumed via direct social interactions, and the mixing of people between different populations is driven by local and global mobility. These are two aspects where the network approach has been shown to be useful *(Barrat et al. 2008)*. The COVID-19 epidemic has given new motivation to such studies as it highlighted several, earlier unseen or less understood phenomena challenging the actual mathematical description and modeling. A few examples are:

- **Stochastic fluctuations.** An epidemic process can behave very differently even if it runs on the same network, with the same transmission probabilities for contacts, and from the same initial seed *(Neri–Gammaitoni 2021)*. This holds in particular when the reproduction rate is close to 1, thus the spreading process is in its critical regime *(Aguiar et al. 2021)*.
- **Geographic distribution and intensity.** The spreading speed of an epidemic depends not only on the number of current infections, but also quite strongly on their geographic distribution in a given population (a country). Estimating how clustered or homogeneously distributed infected cases are in a country can be useful in predicting the intensity of the epidemic.
- **Seeding and epidemic size.** Related to the previous topic, but a more complicated phenomenon can be observed if we consider the dependence of the final size of an epidemic (the number of people infected during its whole duration) from the location where it started from *(Ódor et al. 2021)*. Comparing epidemic processes started from the central region of a country to similar cases started from its periphery leads to the observation of a new switchover phenomenon. It has been found that the spreading started from central populations are worse if the reproduction rate is small, but epidemic initiated from the periphery can reach more individuals if the reproduction rate is slightly over 1.

To formally study these aspects of epidemic spreading we need to model two things: The network itself, and the local rules of epidemic transition and recovery. Mathematical models of infection propagation is a well established area of research, taking into account the role of networks in epidemic spreading *(Barrat et al. 2008)*. Building on this ground, we study epidemic spreading on two levels of network aggregation *(Pastor-Satorras et al. 2015)*: on one hand, identifying nodes as individuals, who are connected via social interactions, possibly transmitting the disease directly between nodes; and on the other hand, identifying nodes as populations of individuals connected by mobility routes, possibly transmitting infected individuals between populations. Deeming them real, both of these structures can be considered to be embedded in geometric space. In turn, spatial and geometric constraints play important roles in the emergence of their structure leading to locally clustered, globally connected hierarchical network patterns *(Barthélemy 2011)*. In

this paper we summarize some results on epidemic models on geometric networks to highlight spreading phenomena where we found the geometric nature of the network structure to be relevant.

As follows, first in Section 1 we formally define the epidemic model to study and the underlying network structures built on data of the commuting network of Hungary. Subsequently, in Section 2 we elaborate on some observations about the fluctuations of simulated epidemic processes on simple geometric structures. In Section 3 we discuss our observations on how epidemic cases were geographically clustered during the course of the first two waves of the COVID-19 pandemic in Hungary. Finally, in Section 4, we summarize our results (earlier reported in *Ódor et al. (2021)*) on the observation and mathematical understanding of a new switchover phenomenon, suggesting that if an epidemic is more virulent, it may infect less people on the long run when seeded from large and well connected populations, as compared to cases when it strikes from randomly selected nodes of a network. We close our paper with a short summary and discussion of our results.

# 1 Models of epidemic spreading on metapopulation networks

## 1.1 The model of epidemic spreading

We use for simulation and mathematical analysis one of the most basic models of an epidemic spreading *(Barrat et al. 2008)*, an SIR process. In this model, each individual can be in one of three mutually exclusive states (S-susceptible, I-infected or R-recovered). At the start of a modeled epidemic everybody is susceptible, except for a small "seed" $S$ of infected people. At every step, every infected person infects every susceptible neighbor in the network with probability $\beta$, and recovers with probability $\mu$. So the expected duration of an infection is $1/\mu$. This simple SIR dynamic is often characterized by the *basic reproduction number*, as defined as the average number of people infected by one ill person in a fully susceptible population ($R_0 = \bar{d}\beta/\mu$), where people have $\bar{d}$ contacts on average.

In our studies we consider $\mu = 1$, so every illness lasts for 1 time unit. The phenomena we are reporting already occur in this simplest model.

## 1.2 Metapopulation network of Hungary

For a more realistic modeling of epidemic phenomena, the network structure can represent the mobility patterns of individuals at various spatial scale (local transportation, commuting, international travels). The concept of *metapopulation networks (Colizza et al. 2007)* provides a broadly accepted framework to this approach. A metatpopulation network consists of $n$ nodes, which represent populations of individuals (which we also call towns or settlements from now on), connected by weighted and/or directed edges, encoding the number of people traveling between these populations.

To construct a metapopulation network for Hungary, we use the microcensus collected and released by the Hungarian Statistical Office in 2016 *(Hungarian Microcensus 2016)*. We relied on two important information from this dataset: (a) the population size of each 3,186 settlement in Hungary (a sample of them visualized in *Figure* 1(a); and (b) the number of people commuting to work or school on a daily basis between these settlements, 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 this 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

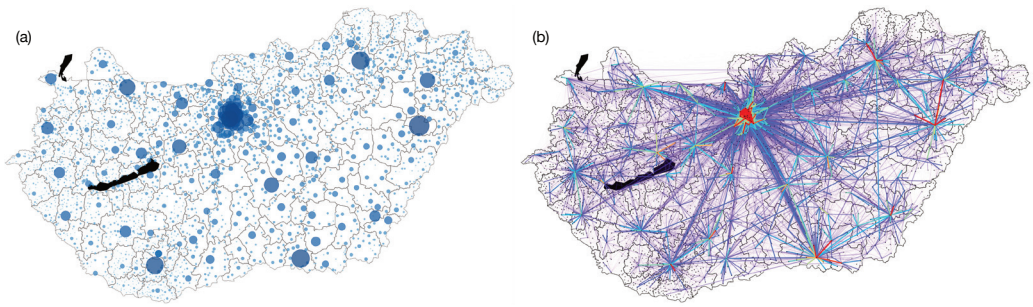


**Figure 1** Hungarian census data. (a) Population size of Hungarian settlements, indicated by blue circles with size proportional to the logarithm of the number of inhabitants. (b) Commuting network of Hungary, where two settlements are connected if there are at least 25 persons between them for school or work purposes. The colors of edges indicate the number of commuters. Source: authors

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 basis. We visualize this weighted spatial commuting network of Hungary in *Figure* 1b.

To model epidemic spreading (using an SIR or other processes) on a metapopulation network, we can employ a reaction-diffusion process. In this case the metapopulation network epidemic model evolves in two phases in each iteration. During the reaction phase, individuals inside each population mix homogeneously, and in case they are infected, they pass the infection to susceptible contacts with probability $\beta$. Meanwhile, an infected individual may recover spontaneously with probability $\mu$ after which, in an SIR scenario, it would never get infected again. 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 populations.

Specifically in our model implementation, we assign to each agent a home population $i$ where they were located in the beginning of the simulation. Each agent is assigned exactly one hometown, and the home assignments do not change for the rest of the simulation. We initialize the epidemic spreading by infecting some selected individuals at $t = 0$ according to some seed selection scenarios (discussed later) and proceed with the simulation for $t$ iteration steps. In the reaction phase, 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 $\beta$ is the infection rate. In the final recovery step, each infected agent recovers with rate $\mu$. In the diffusion phase, each agent who is at its hometown $i$ is selected to move to another town with probability $p_m$. The selected agents then chose a target town $j$ with probability $w_{ij}$, 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. For the exact implementation see *(Code)*. Note that in the reported study we concentrate on the conventional SIR model, however, our observations hold for more realistic models too, including the SEIR model with an addition compartment of exposed (E) state, better describing the reaction scheme of the SARS-Cov-2 disease.

## 2 Fluctuations of modeled epidemics

The intensity curve of an epidemic can be very different, even if it runs on the same network, starts from the same initial seed, and has the same transmission probability $\beta$. *Figure* 2 shows a few runs of simulation of an SIR process on two different networks. The first is a simple square grid; this shows that the phenomenon occurs even if the network is totally homogeneous. The second is a very simple geometric random graph, and its inhomogeneity further increases the fluctuations.

These fluctuations can be observed when the infection rate is close to the critical value $R_0 = 1$ or $\beta = |V|/(2|E|\mu)$. This may look like a serious restriction on the occurrence of fluctuations, but in reality it is a common goal to keep the infection rate near the critical value (if $R_0 > 1$, restrictions are imposed, if $R_0$ drops below 1, these restrictions are eased or lifted).

These examples indicate that epidemic predictions are at least as difficult as weather predictions. In fact, unpredictable circumstances (government intervention, new variants, poorly understood mass psychology) may make it even worse.

It is an interesting open problem to determine, or at least estimate, this instability of an epidemic on a given network, given starting infected set, and given transmission rate. How large is the variance of the epidemic size? How large is the variance of the number of new infections at a given day? The answer will depend on the structure of the underlying network, but perhaps those properties of the network that influence the instability can be identified.

## 3 Geographic distribution and intensity

During the unfolding of an epidemic in a country the infected individuals may not always be homogeneously located across the whole population. First infected cases are commonly reported from larger cities, where case importation from abroad is more likely due to international mobility. In addition, an early phase epidemic survives easier in a densely populated area. In case of an outbreak, the epidemic reaches individuals in smaller settlements and spreads out more homogeneously in the whole population. This was the scenario actually during the first two waves of the COVID-19 pandemic in Hungary. The first wave started in March 2020, and as in many countries, the disease arrived via international travels, landing the epidemic first in larger cities *(Fauver et al. 2020; Kang et al. 2020; Karsai et al. 2020; Röst et al. 2020)*. These initial seeds resulted in outbreaks clumped around highly populated areas *(Ódor et al. 2021)*. This is evident from *Figure* 3a, where we measure how a Pearson's chi-squared test *(Greenwood–Nikulin 1996)*, measured on the distribution of newly infected cases between different settlements, changes over time. It shows that in the beginning and during the first wave of the pandemic the daily new infected cases are not distributed homogeneously according to the population size of settlements. By looking at the per-capita infection probability at the beginning of the first wave (week 1 in *Figure* 3b), it indicates that infection cases were concentrated in cities with the largest populations.

In contrast, the second wave in Hungary arrived after the summer season from a significantly different initial condition. It was likely to be induced by people coming back from

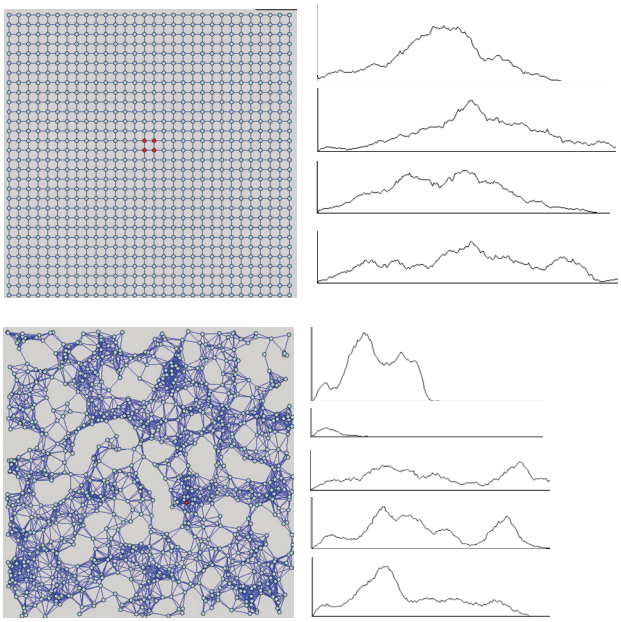

**Figure 2** Left: Different behavior of the same epidemics on a 30 × 30 grid, a very simple and regular network. The same transmission probabilities, and the same initial seed are used. Right: Instability is even more prominent on a more realistic network, where 900 random points in the unit square represent the towns, and two nodes are connected by an edge if their distance is less than .066.
Source: authors

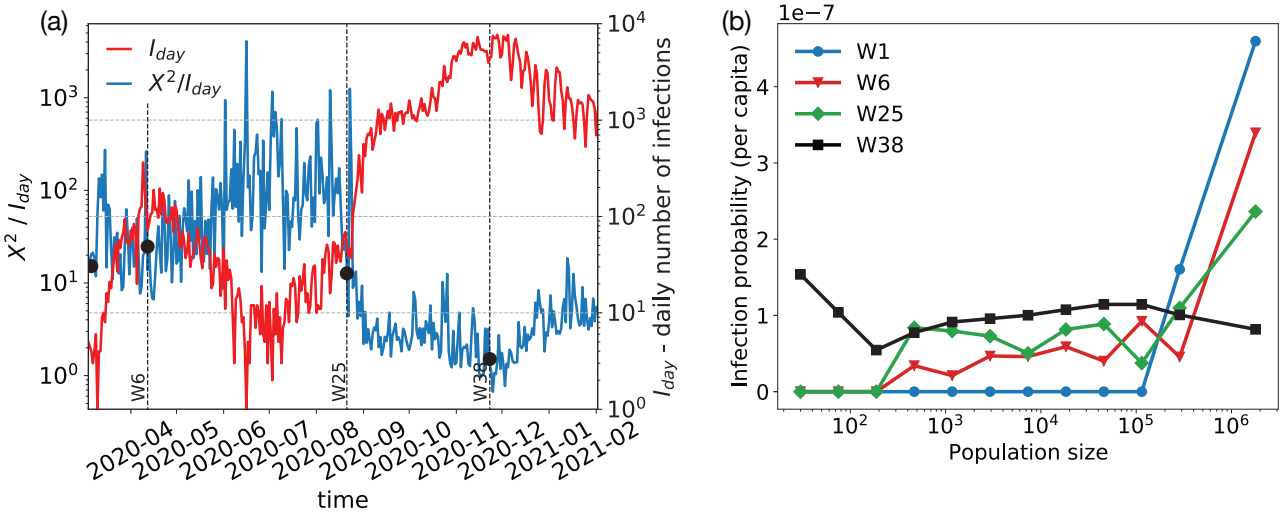


**Figure 3** Geographic distribution of COVID-19 infection cases in Hungary as the function of time. (a) The time evolution of Pearson's chi-squared test (blue line) computed for the distribution of actual newly infected cases in different settlements, shown together with the number of actual newly infected cases (red line). (b) The probability of infections in cities of different size at the first (W1) , sixth (W6), twenty-fifth (W25), and thirty-sixth (W36) of the pandemic.
Source: authors

holidays, thus bringing back the virus to their local communities. Indeed, at the beginning of the second wave (end of August 2020) new infected cases were distributed more homogeneously all around the country. On the one hand, this is

evident from *Figure* 3a, where the corresponding chi-squared values (week 25) are lower, as compared to week 1. On the other hand, the same conclusion can be drawn from *Figure* 3b (week 25), where infected cases appear to be more homogeneously distributed among settlements of different population sizes, although the infection numbers are comparable to the peak of the first wave (week 6). 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 of the second wave (week 38 in *Figure* 3a and b). Surprisingly, the first wave that started from the most tightly connected, central, and largest populations led to significantly smaller number of infections as 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.

# 4 The switchover phenomenon

From the results above it is evident that central and densely populated settlements of the country played a distinct role in the beginning of the pandemic. To identify them in the spatial commuting network of Hungary, we define a *central node set* $\mathcal{C}$, containing Budapest and all its districts and the towns in the suburb of the capital, which all together represent about 30% of the total population of the country *("Central Hungary", 2020)*. Note that we could find more general definitions for $\mathcal{C}$, that are based solely on the network structure. This could be a prescribed number of nodes with the highest degrees, or the core of the network, which is obtained by repeatedly deleting all nodes with the lowest degrees as long as only nodes with prescribed degrees remain.

Using the selected $\mathcal{C}$, we define 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 experiment, we select $s$ (being $< |\mathcal{C}| \ll n$) number of towns chosen randomly from $\mathcal{C}$. In the other, we choose $s$ towns uniformly at random from the whole set of nodes in the network. When starting the spreading, we initially infect a small fraction $i_0$ 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), seeded towns were infected on average with the same number of agents, as the $i_0/s$ fraction of the total population. To quantify the relative effects of the two seeding scenarios, we introduced the *experimental pandemic size ratio* $f_G(R_0, s)$, given by the ratio of the average final infection sizes of epidemic processes seeded from central or uniformly randomly selected nodes. Interestingly, as shown in *Figure* 4, we observe that $f_G(R_0, s) > 1$ for small $R_0 \simeq 1$, indicating that the epidemic started from the $\mathcal{C}$ central set leads to larger outbreaks. However, as we increase $R_0$, the fraction $f_G(R_0, s)$ falls under 1, suggesting that in this case epidemic seeded from uniformly random selected towns over the whole country induces a larger outbreak. This *switchover phenomenon* appears in the regime slightly above critical point of the SIR process, 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 two seeding scenarios lead to different infected cluster sizes, inducing the observed switchover phenomenon in this regime. On the other hand, when $R_0$ grows larger, the difference between the the two different seedings vanishes as the epidemic reaches essentially the whole population in each case. This observed switchover phenomenon challenges the commonly accepted intuition that the size of the epidemic is always the largest if seeded from the best connected sub-graph, or from the largest degree nodes of a network. It provides a good example on cases of real epidemics, which start from seemingly similar conditions but appear with significantly different outcomes.

## *4.1 Epidemic and percolation*

On a (simple, not metapopulation) network $G$, the distribution of the epidemic (in particular, the final number of recovered individuals) of a SIR model with deterministic, unit recovery time has an elegant and useful connection with the commonly used simple mathematical framework of *bond percolation*. In this model, every edge of the network $G$ has a chance to transmit the disease at most once, when one endpoint is infected but the other is still susceptible. Therefore one may decide about every edge *in advance*, independently with probability $\beta$, whether it will transmit the disease (if it gets a chance for this mischief at all). We can think about this as deleting those edges from the graph which are deemed non-transmitting, and so retaining any edge with probability $\beta$. The retained edges form a random subgraph $G^\beta$ of $G$; this procedure is familiar from random graph theory, where it is called a *percolation* of $G$. If a set $S$ of nodes is selected as infection seeds in the network, then the epidemic will spread exactly over the connected components (also called clusters) of $G^\beta$ that contain at least one node of $S$.

For metapopulation models, the connection with graph percolation is more complicated, but a fundamental result of *Barthélemy et al. (2010)* and *Colizza–Vespignani (2008)* extends this connection in an approximate form. These authors argue that once a large outbreak occurs in a town $A$, the proportion of infected people within the town is concentrated around some value $r_\infty \in (0, 1)$ (called the "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:

$$\beta'_{AB} = 1 - \exp\left(-\frac{N_A p_m w_{AB} r_\infty \left(1 - \frac{1}{R_0}\right)}{\mu}\right), \tag{4.1}$$

where $N_A$ is the size of population $A$. (The dependence on $A$ and $B$ can be ignored if we work on an unweighted network.) Herd immunity is reached in each town $A$ after the

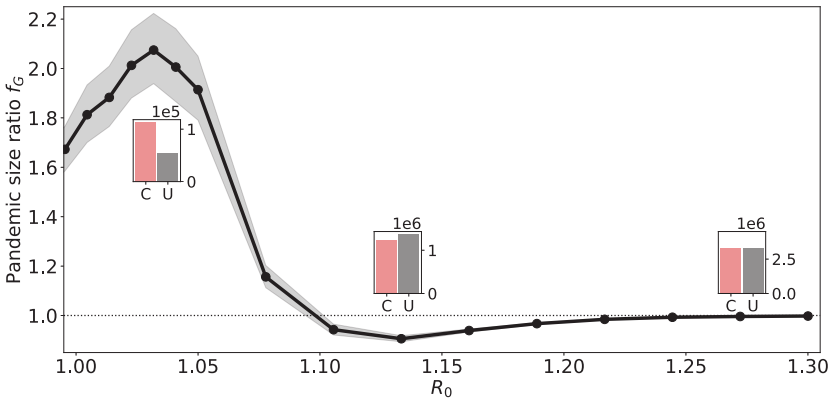


**Figure 4** The switchover phenomenon. The pandemic size ratio $f_G$ shown as the function of the basic reproduction number. When $f_G > 1$, it indicates that infection started from the central set $\mathcal{C}$ of the Hungarian commuting network would infect more people, while when $f_G < 1$, the epidemic seeded from random populations would be more dangerous.
Source: authors

first large local epidemic outbreak of size $r_\infty N_A$, later infections to a town are no longer able to cause large-scale outbreaks. Therefore, after rescaling the time, the towns themselves go through a progression $S \to I \to R$ with unit recovery times and infection probability $\beta'$ computed by (4.1). Consequently, the metapopulation model can be approximated by a simple SIR model on the network of towns, which in turn can be described by a bond percolation process with retention probability $\beta'$ (in the rest of this paper, we drop the prime from $\beta'$ and proceed simply with $\beta$ for simplicity).

Once we mapped our problem to bond percolation, it is important to note that for many graphs $G$ a *critical value* $\beta_c$ separates two phases: for $\beta < \beta_c$ all clusters of $G^\beta$ are small, while for $\beta > \beta_c$ a single *giant cluster* emerges that contains a positive proportion of all nodes, while all other clusters are small. If the central region is sufficiently more connected, then this giant cluster will contain most of $C$, but a smaller fraction of $V \setminus C$. Both seeds sets $S_1$ and $S_2$ infect the giant component, but if the giant component contains most of $C$, then uniform seeding over the whole country may pick up more of the small components.

To make this heuristic argument precise is nontrivial, and we discuss this in two cases: on the configuration model and on general graphs under suitable sparsity/density conditions.

## 4.2 Results about the configuration model

Our first mathematical result will be about the configuration model, a uniform distribution over networks with a given power-law degree sequence. We define the central seeding set $\mathcal{CI}_0(s)$ as the set of $s$ nodes with highest degrees. (Considering the core as would lead to similar results, since the two definitions are strongly correlated in these configuration models; see *Fernholz–Ramachandran (2003)*, *Janson–Luczak (2007)*, *Luczak (1991).*) For the uniform seeding, just as earlier, we choose the seed set $\mathcal{UI}_0(s)$ of $s$ nodes sampled uniformly at random in the whole network.

**Theorem 4.1.** *The sequence of random graphs sampled from the configuration model with exponent $\tau \in (2,4)$ and $n \to \infty$ exhibit the switchover phenomenon. Specifically, with infection probability $\beta = \beta_c + \theta$, under the assumptions $1 \ll s \ll n$ and $n^{-|\tau-3|/(\tau-1)} \ll \theta \ll 1$,*

(1) *if $\theta \ll \max((1/s)^{|\tau-3|}, (n/s)^{-\frac{|\tau-3|}{\tau-1}})$, then the central seeding is more dangerous,*
(2) *if $\theta \gg \max((1/s)^{|\tau-3|}, (n/s)^{-\frac{|\tau-3|}{\tau-1}})$, then the uniform seeding is more dangerous,*

*with high probability as $n \to \infty$.*

Theorem 4.1 formalizes the switchover phenomenon observed in *Figure* 4 (in terms of $R_0$ instead of $\theta$). See *Figure* 5(a) for a phase diagram of Theorem 4.1.

## 4.3 Results about general networks

Theorem 4.1 was a probabilistic statement about a specific distribution of configuration model networks. In this section we aim to generalize our results to all graphs that satisfy certain properties that are sufficient for the switchover phenomenon to appear. It is not hard to see that when $\beta$ is very small, and the epidemic dies out quickly, then essentially only the degrees of the seeds affect the outcome of the epidemic, and the central region is more dangerous. In the remainder of the section we show that under the right conditions, for larger values of $\beta$, uniform seeding can be more dangerous too.

**Definition 4.2.** We say that a graph $G = (V, E)$ has edge-expansion $a > 0$, if for every set $X \subset V$, $|X| \leq n/2$, the number of edges between $X$ and $V \setminus X$ is at least $a|X|$.

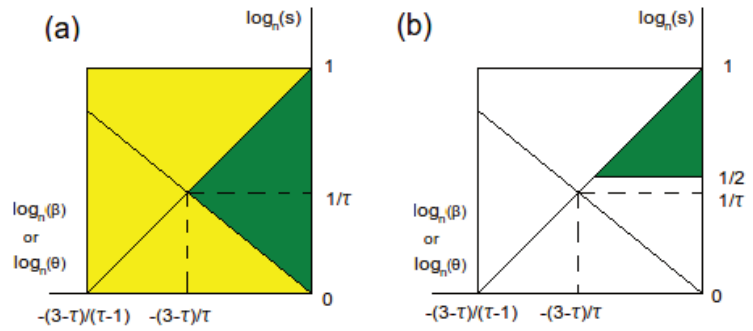


**Figure 5** Phase diagrams for the switchover phenomenon on the configuration model with $\tau \in (2,3)$. (a) With yellow we show the region where the central area is more dangerous by Theorem 4.1.1, and with green we show the region where the central area is more dangerous by Theorem 4.1.2. (b) With green we show the region where the central area is more dangerous by Theorem 4.4. Source: authors

**Definition 4.3.** Let $G = (V, E)$ be a connected graph on $n$ nodes, and let $C \subseteq V$, $|C| = c \leq n/2$. Let $G_1 = G[C]$, and $G_2 = G \setminus E(G_1)$. Let $b$ denote the average degree in $G_2$, and let $a$ be the edge expansion of $G_1$. Let $\beta < 1$ be the infection probability. For a seed set $S$, with $|S| = s$, we denote by $G^\beta(S)$ the (random) set of nodes that get infected during the epidemic when $S$ is the seed (the set of nodes infected at the start).

The following theorem is a version of our results that can be stated without too many technicalities.

**Theorem 4.4.** *Consider the setting and notation of Definition 4.3 with $a > 2b$. Let $S_1$ be a random $s$-subset of $C$ and $S_2$, a random $s$-subset of $V$. Suppose that the following three assumptions hold on the model parameters:*

$$\text{(i)} \quad \beta \geq 4/a,$$

$$\text{(ii)} \quad 11ne^{-\beta(a-2b)} \leq s \leq \frac{c}{2} \leq \frac{n}{4},$$

$$\text{(iii)} \quad sc(1-\beta)^{2b} \gg n\log(n).$$

*Then $\mathbb{E}(|G^\beta(S_2)|) > \mathbb{E}(|G^\beta(S_1)|)$, i.e., the uniform seeding is more dangerous. On the other hand, for all sufficiently small $\beta < 1$, central seeding is more dangerous. Hence, the switchover phenomenon occurs on $G$.*

Under slightly stronger assumptions on the parameters, the epidemic size for seed set $S_2$ can be larger than for seed set $S_1$ by a constant proportion of the population. The proof of Theorem 4.4, along with related results, will be published in a separate paper. While Theorem 4.4 talks about general graphs, it can partially reproduce the more involved and precise results about the configuration model. This is surprising because Theorem 4.4 only lists a few general conditions on the graph, and does not use its structure in much detail, while the proof of Theorem 4.1 does.

**Corollary 4.5** (Corollary to Theorem 4.4). *Theorem 4.4 applied to the configuration model with exponent $\tau \in (2,3)$ implies that the uniform seeding is more dangerous when the edge-retention probability $\beta$ and the size of the seed-set $s$ satisfies*

$$\beta \gg (n/s)^{-\frac{3-\tau}{\tau-1}} \log(n/s),$$
$$s \gg \sqrt{n\log(n)}.$$

See *Figure* 5 for an illustration of the applicability of Theorem 4.4 to the configuration model.

## 5 Conclusions

The COVID-19 pandemic highlighted several new phenomena, which challenged the state-of-the-art understanding and mathematical modeling of network epidemic processes. In this paper we briefly described three interesting phenomena that we studied lately using geographically detailed epidemic and mobility data from Hungary. We discussed that

(1) The unpredictability of the epidemic curve near the critical rate $R_0 = 1$ is not a shortcoming of the methods, but an inherent property of the process, somewhat similarly to the weather forecast. On the other hand, when there are only a few infection cases, thus the system's $R_0$ is far from its critical value 1, the predictions are much more reliable.

(2) The geographic distribution of a disease on a given day during a real pandemic is strongly correlated with the growth rate (daily number of new infections). In the early phase of the process, infections are more concentrated in large and well connected cities, while in later phases, infections are distributed more homogeneously between settlements. To improve predictions in the future, collection of data on a finer spatial scale would be necessary. This may follow administrative spatial structures like settlements or counties; perhaps data could be considered on the level of hospitals or family physician practices, which

are potentially easier to collect. This is particularly important in the phases of the epidemic when infection numbers are low. In these situations, the current infected set can be considered as the seed for the future dynamics, and could lead to significantly different epidemic outcomes on the long run, as we have demonstrated.

(3) From the point of view of network science, an important long-term task is to develop theoretical foundations for network epidemic models. Currently, conventional network properties are considered, as the degree heterogeneity, clustering, multi-layer organization or the temporal evolution of the network, and their effects on the outcome of ongoing modeled epidemic processes. Meanwhile, there are several concepts borrowed from graph theory that could contribute to our deeper understanding of the interaction of structure and ongoing spreading phenomena.

Although most of our results contribute to the fundamental understanding and theoretical modeling of epidemic spreading, yet they have important policy implications. They highlight that it is important to follow not only the number of infected cases during an unfolding pandemic, but also the geographic distribution of the epidemics, which could provide critical insights, as it may determine the long-term outcome of the epidemic. Depending on the phase of a pandemic and the actual reproduction rate of the disease, different intervention efforts have to be concentrated in large and central, and small and peripheral towns for a more effective epidemic control.

## References


Aguiar, M., Van-Dierdonck, J. B., Mar, J., Cusimano, N., Knopoff, D., Anam, V., & Stollenwerk, N. (2021) Critical fluctuations in epidemic models explain covid-19 post-lockdown dynamics. Scientific Reports, Vol. 11. No. 1. pp. 1–12.

Barrat, A., Barthelemy, M., & Vespignani, A. (2008) Dynamical processes on complex networks. Cambridge University Press.

Barthélemy, M. (2011) Spatial networks. Physics Reports, Vol. 499. No. 1-3. pp. 1–101.

Barthélemy, M., Godreche, C., & Luck, J.-M. (2010) Fluctuation effects in metapopulation models: percolation and pandemic threshold. Journal of Theoretical Biology, Vol. 267. No. 4. pp. 554–564.

Colizza, V., Pastor-Satorras, R., & Vespignani, A. (2007) Reaction–diffusion processes and metapopulation models in heterogeneous networks. Nature Physics, Vol. 3. No. 4. pp. 276–282.

Colizza, V., & Vespignani, A. (2008) Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. Journal of Theoretical Biology, Vol. 251. No. 3. pp. 450–467.

Fauver, J. R., Petrone, M. E., Hodcroft, E. B., Shioda, K., Ehrlich, H. Y., Watts, A. G., … & Grubaugh, N. D. (2020) Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. Cell, Vol. 181. No. 5. pp. 990–996.

Fernholz, D., & Ramachandran, V. (2003) The giant k-core of a random graph with a specified degree sequence. Unpublished.

Greenwood, P. E., & Nikulin, M. S. (1996) A guide to chi-squared testing, John Wiley & Sons.

Janson, S., & Luczak, M. J. (2007) A simple solution to the k-core problem. Random Structures & Algorithms, Vol. 30. No. 1-2. pp. 50–62.

Kang, D., Choi, H., Kim, J.-H., & Choi, J. (2020) Spatial epidemic dynamics of the COVID-19 outbreak in China. International Journal of Infectious Diseases, Vol. 94. pp. 96–102.

Karsai, M., Koltai, J., Vásárhelyi, O., & Röst, G. (2020) Hungary in masks/"MASZK" in Hungary. Corvinus Journal of Sociology and Social Policy, Vol. 11. No. 2. pp. 139–146.

Łuczak, T. (1991) Size and connectivity of the k-core of a random graph. Discrete Mathematics, Vol. 91. No. 1. pp. 61–68.

Neri, I., & Gammaitoni, L. (2021) Role of fluctuations in epidemic resurgence after a lockdown. Scientific Reports, Vol. 11. No. 1. pp. 1–6.

Newman, M. (2018) Networks. Oxford University Press.

Ódor, G., Czifra, D., Komjáthy, J., Lovász, L., & Karsai, M. (2021) Switchover phenomenon induced by epidemic seeding on geometric networks. Proceedings of the National Academy of Sciences, Vol. 118. No. 41. e2112607118.

Pastor-Satorras, R., Castellano, C., Mieghem, P. V., & Vespignani, A. (2015) Epidemic processes in complex networks. Reviews of modern physics, Vol. 87. No. 3. p. 925.

Röst, G., Bartha, F. A., Bogya, N., Boldog, P., Dénes, A., Ferenci, T., … & Oroszi, B. (2020) Early phase of the COVID-19 outbreak in Hungary and post-lockdown scenarios. Viruses, Vol. 12. No. 7. p. 708.

Vega-Redondo, F. (2007) Complex social networks. Econometric Society Monographs Vol. 44. Cambridge University Press.

Wasserman, S., & Faust, K. (1994) Social network analysis: Methods and applications. Cambridge University Press.


## Links


Central Hungary (3 July 2020) in Wikipedia. https://en.wikipedia.org/w/index.php?title=Central_Hungary&oldid=965825126 (Accessed: 17.05.2021)

Hungarian Microcensus (2016): https://www.ksh.hu/mikrocenzus2016/?lang=en

Implementation of epidemic modelling with metapopulation and percolation models: https://github.com/dczifra/epidemic_seeding