The importance of superspreaders on the spread of the global COVID-19 pandemic

Jaan Kalda (✉ kalda@ioc.ee)

Tallinn University of Technology

Mart Ratas

University of Reading

Taavet Kalda

Tallinn University of Technology

Azer Ramazanli

Tallinn University of Technology

Heiko Herrmann

Tallinn University of Technology <https://orcid.org/0000-0001-9892-0670>

Juri Belikov

Tallinn University of Technology

Letter

Keywords:

Posted Date: December 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1167477/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

The importance of superspreaders on the spread of the global COVID-19 pandemic

Mart Ratas^{1,4}, Taavet Kalda¹, Azer Ramazanli^{2,3}, Heiko Herrmann¹, Juri Belikov²
and Jaan Kalda^{1*}

¹Department of Cybernetics, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia.

^{2*}Department of Software Science, Tallinn University of Technology, Akadeemia tee 15a, 12618 Tallinn, Estonia.

³Department of Computer Systems, Tallinn University of Technology, Akadeemia tee 15a, 12618 Tallinn, Estonia.

⁴Department of Meteorology, University of Reading, Meteorology Building, Whiteknights Road, Earley Gate, RG6 6ET, Reading, United Kingdom.

*Corresponding author(s). E-mail(s): kalda@ioc.ee;

Contributing authors: mart.ratas@gmail.com; taavet.kalda@gmail.com;
azer.ramazanli@taltech.ee; hh@cens.ioc.ee; juri.belikov@taltech.ee;

Abstract

The dynamics of pandemics is most often analyzed using a variation of the SIR (Susceptible-Infected-Recovered) model¹, the key parameter of which is the basic reproduction number R_0 . Some evidences suggest that the contagion-spreading networks are scale-free, with the biggest nodes corresponding to superspreaders^{2,3}. However, current understanding of the scale-free topology of these networks, and of the implications of such topology for the dynamics of pandemics is incomplete. Here we show that the world-wide spreading rate of COVID-19 gives an indirect evidence that the underlying virus-spreading network is scale-free, with the degree distribution exponent close to 2. Furthermore, our results show that the spreading rate of a virus is predominantly controlled by superspreaders who typically get infected and acquire immunity during the initial outbreak stage of the pandemic. Thereby the biggest nodes get immune and hence, removed from the network, resulting in a rapid decrease of the effective reproduction number. These findings are important for understanding the dynamics of pandemics, and for designing optimal virus control strategies. In particular, screening a population for the number of antibodies of a set of viruses can reveal potential superspreaders, the vaccination or isolation of whom can impede a pandemic at its early stage.

Main

The basic reproduction number R_0 has been serving virologists as the key parameter to describe the

virulence of pathogens for almost a century. Meanwhile, during the recent decades it has become clearer and clearer that the network structure of the virus spreading pathways has an important role to play, see the reviews^{4,5}. It was realized

that viruses can spread over complex networks the scale-free topology of which can affect dramatically the dynamics of the process^{6–10}, particularly in the case of COVID-19^{11–17}, SARS^{18–20}, MERS²¹, avian influenza³, and sexual diseases².

The scale-free nature of the virus-spreading networks has deep implications. To begin with, a small community without potent superspreaders is expected to have a smaller value of effective reproduction number R_E than a large community with potent superspreaders.

More importantly, the value of R_E may decrease rapidly in time due to what we shall be referring to as *self-targeting attacks*. This term is derived from the fact that scale-free networks are vulnerable to the so-called *targeted attacks*^{22,23}: the operation of such networks is disrupted if the biggest nodes are destroyed. Self-targeting attacks are acts of the system itself destroying its biggest nodes. The probability of a node getting infected and subsequently removed from a virus spreading network (assuming a recovery from the infection leads to becoming immune) is proportional to the number of links connecting it to the other, possibly infected nodes. Hence, the nodes with the largest number of links are the ones with the highest likelihood of becoming infected and getting removed from the virus spreading network at the earliest stage of the pandemics.

Real-world spreading rate of COVID-19. In order to see if the world-wide spreading rate of COVID-19 carries a fingerprint of self-targeted attacks, we utilized publicly accessible data of COVID-19 spread²⁴ covering the period from January 2020 until September 2021 by plotting the effective reproduction number R_E for a certain country against the percentage r of those who have suffered the virus earlier and have been thereby removed from the virus spreading network. The instantaneous value of R_E is calculated as the ratio of the number of new cases at this point in time divided by the number of new cases one week ago. One week is expected to approximately correspond to the average time interval between getting the virus, and infecting the next subjects. This includes the average incubation period of six days²⁵, and a time gap from the first symptoms to isolation. When calculating the percentage r , we have excluded those cases which happened before the all-time maximum value of R_E was reached. Fig. 1 shows the dependence of R_E on

Fig. 1: Calculated averaged reproduction number of COVID-19 cases for different regions and countries, plotted against the cumulative number of cases, and normalized to the total population in the initial stage of the pandemic. The dashed line depicts a dependence of $R_E \propto r^{-0.5}$.

the cumulative fraction of infected population. For single countries, the obtained $R_E(r)$ -curves fluctuate due to insufficient statistical basis. More stable results are obtained by performing this analysis for a set of countries/regions: for European countries, US counties, and for a set of biggest countries in the world (cf. the section “Data availability”). For each set of countries, we have binned the values of r , and averaged $R_E(r)$ over the countries in the set to find the mean value $\langle R_E \rangle$. A straight line in this log-log plot corresponds to a power law; for reference, a dependence $R_E \propto r^{-\alpha}$ with $\alpha = 1/2$ is depicted by a dashed line. The fact that already a very small percentage r of the population with acquired immunity results in a noticeable decrease of R_E provides a strong evidence of self-targeted attacks being in action.

Scale-free model. In order to verify whether the behaviour observed from the real-world data is really the consequence of the scale-free nature of the virus-spreading network, we have constructed the following mathematical model which, similarly to Ref. 11, combines the SIR model with a scale-free network and is kept as simple as possible.

- (i) The network links are (a) fixed in time, and (b) of equal strength: there is a fixed probability p for the virus to spread over each of the links connecting an infected node with a susceptible node. Regarding these simplifications, (a) this *model network* is to be considered as a snapshot of the real world: all those contacts which are made between people during a certain time period (of duration equal to the average infectious time) form the links of the model network; (b) while in the real world, there are strong and weak links, the strong links (e.g. between family members) can be represented via multiple weak links of the model.
- (ii) Nodes which have recovered from the infection remain immune until the end of the simulation run. For some simulation runs, a certain fraction of nodes have additionally been “immunized” through “vaccination”: random nodes were selected and marked as immune, prior to any infections.
- (iii) The incubation time is taken to be equal to one time step, i.e. iteration, and each infected node remains infectious during one time step. In other words, if a node got the virus at time t_n , it will be infectious at time t_{n+1} , and will neither be able to get the virus again nor transmit it to anyone for $t \geq t_{n+2}$. This means that in the real world, one time step correspond to the average time interval between getting the virus, and infecting the next subjects, i.e. slightly more than the average incubation period. Keeping in mind that the average incubation time is around six days, one time step corresponds to around one week in real world. Using a probability distribution of the incubation period can increase the accuracy of simulations, but is not expected to affect the scaling behavior. Indeed, scale-free systems which differ only in correlations at the smallest available scale belong usually to the same *universality class*, i.e. are characterized with the same scaling exponents, cf. 26.
- (iv) Initially, the whole population is susceptible to the virus, which is introduced by selecting and infecting a random node. The simulation run is considered as a successful one if at least 1% of the population got the virus; otherwise, the data is discarded. For successful runs,

only the data after the maximum reproduction number was reached was considered; the averaged $\langle R_E \rangle$ curves were calculated based on 200 successful simulation runs, mimicking the procedure employed by the analysis of the real-world data.

Our networks have been synthesized using the generator from the NetworkX package^{27,28}, and are characterized by the *cumulative degree distribution exponent* κ — the number of nodes n with degree exceeding k scales as $n \propto k^{-\kappa}$.

Technical analysis. Simulations were run for different scale-free networks with $\kappa = \{1.6, 2.0, 2.5, 3.4\}$ while the virus transmission probability was kept equal to $p = 0.1$. The middle plot in Fig. 2 shows $\langle R_E \rangle$ plotted against the cumulative number of infected nodes. Observe that for $\kappa \leq 2$, there is a fairly good power law over a wide range of values, $\langle R_E \rangle(r) \propto r^{-\alpha}$, corresponding to a straight line in Fig. 2. Such a behaviour can be explained as follows. The probability a node will get infected is proportional to its degree, hence the *expected reproduction number* R_{exp} is proportional to the weighted average degree (averaged over the network), with the weights being equal to the degree: $R_{\text{exp}} = p\langle k^2 \rangle / \langle k \rangle$, where p denotes the probability of the infection being transmitted over a link, and angular braces—averaging over all nodes of the network.

There is an important difference between the cases with $\kappa \leq 2$, and $\kappa > 2$, cf.⁹. For $\kappa \leq 2$, R_{exp} is dominated by the contributions from the largest nodes whose degree is on the order of the maximal degree k_{max} . Hence, R_{exp} can be expected to decrease proportionally to the maximal degree of the *effective network*: $R_{\text{exp}} \propto k_{\text{max}}$. Here, the term *effective network* refers to the network where the links to the ineffective nodes (recovered or vaccinated) are removed. The fact that the largest nodes dominate the contribution to R_{exp} means also that the largest nodes catch the infection at the earliest stage of the pandemics, and will be removed from the effective network henceforth. If the number of the biggest removed nodes is proportional to the overall number r of the infected and recovered nodes then $r \sim k_{\text{max}}^{-\kappa}$, and hence the maximal degree k_{max} of the remaining nodes scales as $k_{\text{max}} \propto r^{-1/\kappa}$, implying that $R_{\text{exp}} \propto r^{-\alpha}$ with $\alpha = 1/\kappa$, in an approximate agreement with the simulation results.

Fig. 2: Effective value of the reproduction number R_E versus the percentage of infected and recovered population for various infection probabilities for the same scale-free network (left), for different scale-free infection networks with $p = 0.1$ (middle), and for various vaccination rates for $p = 0.1$ using a fixed scale-free network (right). The plots depict the binned average of 200 different runs, with the error bars denoting one standard deviation within a bin.

The scenario described above will cease to work if $\kappa > 2$, when R_{exp} is no longer dominated by the contributions from the biggest nodes. Removal of a small number of the biggest nodes at the earliest stages of the pandemics can no longer dramatically affect the value of R_{exp} . This is confirmed by the simulation results: the curves corresponding to $\kappa > 2$ in the middle plot of Fig. 2 start with a nearly horizontal segment, and turn to a steeper descent only once a considerable number of nodes have been infected.

The real-world data are best matched with $\kappa \approx 2$. This fact can be considered as an indirect evidence that the virus spreading network is scale-free, with the degree distribution exponent being close to 2. This match is reasonably good, except that the real world curves follow the law $R_E \sim r^{-\alpha}$ only when $R_E > 1$, see Fig. 1. The reason is that while the simulation runs are stopped as soon as the first wave of pandemics dies out, the real world time series include also data from the subsequent waves.

The remaining plots of Fig. 2 show the simulation results for the same scale-free network with $\kappa = 2$. The left plot depicts $\langle R_E \rangle$ against cumulative number of infected nodes for different values of the virus propagation probabilities. While the curves with smaller p lie beneath the curves with larger p , all the curves exhibit almost the same

power law $\langle R_E \rangle \propto r^{-\alpha}$ with $\alpha \approx 0.5$. Only in the case $p = 0.04$, this behavior starts breaking down, the reason being that with so small p , the pandemics fluctuates at the verge of extinction and undergoes multiple minor “waves”: the infected population remains so small that contrary to the assumptions used to motivate the equality $\alpha = 1/\kappa$, sometimes, by chance, there are no infectious superspreaders.

The right plot shows the effect of initial vaccination of the population from 0% to 45%. The scaling law remains intact, vaccination will only affect the prefactor, resulting in a vertical shift of the graphs.

Simulation results show that if the virus spreading network is scale-free, with a big enough degree distribution exponent, $\kappa \geq 2$, the virus propagation rate is controlled by superspreaders, i.e. by subjects with the biggest number of contacts. The fact that the world-wide spreading rate of COVID-19 is consistent with the dynamics on a synthetic scale-free network with $\kappa \approx 2$ suggests that the same is valid for the real world. This triggers a question whether there is a way to identify the superspreaders so that by vaccinating or isolating them, it would be possible to avert the pandemic. Because the superspreaders are highly prone to catch the virus, they are expected to

Fig. 3: Simulation results with screening for a population of 100 000 individuals: the number of total infections is shown depending on the screening virus count and the initial vaccination rate p_t for a fixed value of the screening virus spreading probability $p_v = 0.07$ (left), and depending on the spreading probability p_v and the vaccination rate p_t if $N = 5$ screening viruses are used (right). The contours are calculated using cubic interpolation for the logarithm of the total infection count.

have caught almost all the similar viruses by a certain point in time, developing the corresponding antibodies. Therefore, the potential superspreaders could be found by carrying out a screening based on the number of antibodies for a selected collection of viruses.

The screening viruses in this collection should have neither too high nor too low transmission probability which is henceforth denoted as p_v . Viruses with too small p_v fail in infecting large enough fraction of the population, and viruses with too large p_v infect too many non-superspreaders. We have tested the feasibility of this approach by running another set of simulations. First, a collection was built using N different “viruses”, all having the same transmission probability p_v . For each virus from the collection, a pandemic was initiated by infecting random nodes and letting it evolve until it naturally dies out. Then, the nodes were ranked according to the number of “antibodies”, i.e. by the number of times they were infected previously. Then, the target for the vaccination coverage p_t was set, and nodes were “vaccinated” (i.e. marked as immune) according to the ranking list, starting with the node with the largest number of “antibodies”, and moving downwards until the vaccination coverage target was reached. Finally, the main pandemic was initiated by infecting a random node of the network, and the total number of nodes M which got the virus during the main pandemics was

determined. The last step was repeated 300 times to find the average value $\langle M \rangle$.

The simulation results for a population of 100 000 individuals are presented in Fig. 3. The left plot shows $\langle M \rangle$ as a function of N and p_t by $p_v = 0.07$, and the right plot— $\langle M \rangle$ as a function of p_v and p_t by $N = 5$. Observe that the number of viruses in the collection does not need to be too big, even just with five different viruses in the collection, vaccination of merely 6% of the nodes can prevent the virus from spreading. In the real world, the results will be somewhat inferior as the contact links between people will vary over time and are not frozen in time as in the case of our simulations. For instance, big conferences temporarily create many superspreading nodes of the contact network. On the other hand, temporary superspreaders create a noise for the screening procedure. Because of that, somewhat bigger virus collection is probably needed for a reasonably reliable screening procedure. Notice also that the suggested screening procedure can neutralize only persistent superspreading nodes, and the temporary ones need to be handled via administrative measures.

References

- [1] Kermack, W. O., McKendrick, A. G. & Walker, G. T. A contribution to the mathematical theory of epidemics. *Proceedings of*

- the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **115** (772), 700–721 (1927).
- [2] Liljeros, F., Edling, C. R., Amaral, L. A., Stanley, H. E. & Aberg, Y. The web of human sexual contacts. *Nature* **411** (6840), 907–908 (2001).
 - [3] Small, M., Walker, D. M. & Tse, C. K. Scale-free distribution of avian influenza outbreaks. *Phys. Rev. Lett.* **99**, 188702 (2007).
 - [4] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
 - [5] Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. Critical phenomena in complex networks. *Rev. Mod. Phys.* **80**, 1275–1335 (2008).
 - [6] Dezső, Z. & Barabási, A.-L. Halting viruses in scale-free networks. *Phys. Rev. E* **65**, 055103 (2002).
 - [7] Volchenkov, D., Volchenkova, L. & Blanchard, P. Epidemic spreading in a variety of scale free networks. *Phys. Rev. E* **66**, 046137 (2002).
 - [8] Barthélemy, M., Barrat, A., Pastor-Satorras, R. & Vespignani, A. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* **92**, 178701 (2004).
 - [9] Vazquez, A. Polynomial growth in branching processes with diverging reproductive number. *Phys. Rev. Lett.* **96**, 038702 (2006).
 - [10] Lv, W., Ke, Q. & Li, K. Dynamical analysis and control strategies of an SIVS epidemic model with imperfect vaccination on scale-free networks. *Nonlinear dynamics* **99** (2), 1507–1523 (2019).
 - [11] Herrmann, H. A. & Schwartz, J.-M. Why COVID-19 models should incorporate the network of social interactions. *Physical Biology* **17** (6), 065008 (2020).
 - [12] Singer, H. M. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *Physical Biology* **17** (5), 055001 (2020).
 - [13] Aleta, A. *et al.* Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav* **4** (9), 964–971 (2020).
 - [14] Manchein, C., Brugnago, E. L., da Silva, R. M., Mendes, C. F. O. & Beims, M. W. Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30** (4), 041102 (2020).
 - [15] Cheng, X., Wang, Y. & Huang, G. Global dynamics of a network-based SIQS epidemic model with nonmonotone incidence rate. *Chaos, Solitons & Fractals* **153**, 111502 (2021).
 - [16] Thurner, S., Klimek, P. & Hanel, R. A network-based explanation of why most COVID-19 infection curves are linear. *Proceedings of the National Academy of Sciences* **117** (37), 22684–22689 (2020).
 - [17] Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in china. *Science* **368** (6492), 742–746 (2020).
 - [18] Small, M. & Tse, C. K. Small world and scale free model of transmission of SARS. *International Journal of Bifurcation and Chaos* **15** (05), 1745–1755 (2005).
 - [19] Small, M. & Tse, C. Clustering model for transmission of the SARS virus: application to epidemic control and risk assessment. *Physica A: Statistical Mechanics and its Applications* **351** (2-4), 499–511 (2005).
 - [20] Small, M., Tse, C. & Walker, D. M. Superspreaders and the rate of transmission of the SARS virus. *Physica D: Nonlinear Phenomena* **215** (2), 146–158 (2006).

- [21] Kim, Y., Ryu, H. & Lee, S. Effectiveness of intervention strategies on MERS-CoV transmission dynamics in south korea, 2015: Simulations on the network based on the real-world contact data. *International Journal of Environmental Research and Public Health* **18** (7), 3530 (2021).
- [22] Crucitti, P., Latora, V., Marchiori, M. & Rapisarda, A. Efficiency of scale-free networks: error and attack tolerance. *Physica A: Statistical Mechanics and its Applications* **320**, 622–642 (2003).
- [23] Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406** (6794), 378–382 (2000).
- [24] Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* **20** (5), 533–534 (2020).
- [25] Cheng, C. *et al.* The incubation period of COVID-19: a global meta-analysis of 53 studies and a chinese observation study of 11 545 patients. *Infectious Diseases of Poverty* **10** (2021).
- [26] Stanley, H. E. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev. Mod. Phys.* **71**, S358–S366 (1999).
- [27] Hagberg, A. A., Schult, D. A. & Swart, P. J. Varoquaux, G., Vaught, T. & Millman, J. (eds) *Exploring network structure, dynamics, and function using NetworkX*. (eds Varoquaux, G., Vaught, T. & Millman, J.) *Proceedings of the 7th Python in Science Conference*, 11 – 15 (Pasadena, CA USA, 2008).
- [28] Bollobás, B., Borgs, C., Chayes, J. T. & Riordan, O. N/A (ed.) *Directed scale-free graphs*. (ed.N/A) *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 12-14, 2003, Baltimore, Maryland, USA*, 132–139 (ACM/SIAM, 2003).
- [29] United Nations Department of Economic and Social Affairs. Total population - both sexes (2020). URL <https://population.un.org/wpp/Download/Standard/Population/>.
- [30] U.S. Census Bureau. Population, population change, and estimated components of population change (2019). URL <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html>.

Methods

Real world data

For the purposes of aggregating and showcasing the spread of COVID-19 from multiple countries, two sets of countries were formed. The first set included all the countries from Europe, and were as follows:

- Albania, Andorra, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Moldova, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, and United Kingdom.

The second set included the most populous countries in the world. Some of the countries were removed due to lack of reported data, or the data being erroneous, such as the reported number of cases on a day being negative. The included countries were as follows:

- Argentina, Brazil, Colombia, France, Germany, India, Indonesia, Iran, Italy, Japan, Mexico, Peru, Poland, Russia, South Africa, Spain, Turkey, Ukraine, United Kingdom, and United States.

For United States counties, the data from 1809 counties were aggregated. Only counties that had reliable COVID-19 and population census data were considered.

Scale-free network

The underlying graph of the scale-free networks used in our model was generated using the `scale_free_graph` function from the NetworkX package, which provides a directed scale-free graph. The function takes as its argument three parameters, α , β , and γ where $\alpha + \beta + \gamma = 1$ ²⁸. Changing these parameters changes the cumulative degree distribution exponent of the ingoing and outgoing connections. The directed graph is then turned into an undirected

graph and double connections between nodes as well as self-connections are removed.

As an example, in order to get $\kappa = 2.0$, values of $\alpha = 0.1$, $\beta = 0.8$, $\gamma = 0.1$ can be used. See Extended Data Table 1 for other κ and parameter value pairs. Examples of the cumulative degree distribution for several different resulting graphs with $N = 500\,000$ nodes are shown on Extended Data Fig. 1.

Simulation description

The stochastic SIR model used in the simulations is already described in some detail in the main text. At each time step, the effective reproduction number R_E is found as the ratio of the number of new infections I_t at the current time step t to the number I_{t-1} at the previous time step, and its value is recorded together with the cumulative fraction $r_t = N^{-1} \sum_{\tau=0}^{t-1} I_\tau$ of all the infections during the whole simulation run. Only successful runs that infected at least 1 % of the population were considered.

The parameters used for running the simulations that were used to generate figures 2 and 3 are displayed in Extended Data Table 2.

Simulations involving screening

The simulations used for figure 3 were carried out at 14 different values of initial vaccination percentage p_t , 10 different values of the screening virus spreading probability p_v , and 14 different values of screening virus count n_v . The values used can be found in Extended Data Table 2. Notably, a screening virus count of $n_v = 0$ was used in the simulations, corresponding to the case where random initial vaccination was used.

For each set of parameters (p_v, p_t, n_v) , 300 simulations were carried out and averaged with a virus with spreading probability $p = 0.1$ and a scale-free network with $\kappa = 2.0$ (refer to Table 2).

Data availability

The real world COVID-19 data is gathered from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University²⁴, available at <https://github.com/CSSEGISandData/COVID-19>.

The population census data for different world countries is gathered from United Nations Department of Economic and Social Affairs survey²⁹, available at <https://population.un.org/wpp/Download/Standard/Population/>. US county population census is gathered from a survey conducted by United Nations Department of Economic and Social Affairs³⁰, available at <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html>.

All simulation results discussed in the article are available from the corresponding author on request.

Code availability

The code used for the simulations discussed within this paper can be accessed at TalTech Data Repository <https://doi.org/10.48726/37qsj-66n08>.

Acknowledgment

The work was partly supported by the Estonian Research Council grant COVSG22. The simulations were carried out in the High Performance Computing Centre of TalTech. We would like to express thanks to Dr. Aleksei Tepljakov who managed part of the research project leading to the results presented in this paper.

Author contributions

MR and TK wrote most parts of the simulation code, and performed the data analysis; HH contributed also to the code. TK collected and analyzed the epidemic data. AR and HH provided a literature meta analysis. JB guided the data visualization. JK conceived the model and the framework of the data analysis, as well as supervised the research. All authors participated in discussions, review and editing, and approved the final draft of the article.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Jaan Kalda.

Extended Data

Extended Data Fig. 1: Cumulative frequency distribution plots of the four scale-free networks which were used for the simulations, with different scaling exponents κ , each with $N = 500\,000$ nodes. The cumulative frequency at a degree k is represented as the number of nodes with degree higher or equal to k .

Extended Data Table 1: Parameters used to generate the scale-free graph.

NetworkX parameters			κ
α	β	γ	
0.35	0.45	0.20	3.45
0.14	0.72	0.14	2.53
0.10	0.80	0.10	2.00
0.05	0.90	0.05	1.58

Extended Data Table 2: Parameters used for simulations and the number of simulations run for each set of parameters

Fixed parameters	Varied parameters	Number of simulations (per parameter set)	Total simulations	Figure
$N = 10^5$ $iv = 0$ $\kappa = 2.0$	$p = 0.04, p = 0.10$ $p = 0.20, p = 0.50$	200	800	2
$N = 10^5$ $p = 0.10$ $iv = 0$	$\kappa = 3.45, \kappa = 2.53$ $\kappa = 2.00, \kappa = 1.58$	200	800	2
$N = 10^5$ $p = 0.1$ $\kappa = 2.0$	$iv = 0.00, iv = 0.10$ $iv = 0.30, iv = 0.45$	200	800	2
$N = 5 \times 10^5$ $\kappa = 2.0$	$p = p_v \in \{0.025, 0.050, 0.0625, 0.075, 0.0875, 0.010, 0.012, 0.015, 0.018, 0.02\}$	25	250	3
$N = 5 \times 10^5$ $p = 0.1$ $\kappa = 2.0$	$p_v \in \{0.025, 0.050, 0.0625, 0.075, 0.0875, 0.010, 0.012, 0.015, 0.018, 0.02\}$ $s_{\%} \in \{0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10\}$ $n_v \in \{0, 1, 2, 3, 4, 5, 10, 15, 20\}$	300	378000	3
$N = 5 \times 10^5$ $p = 0.1$ $\kappa = 2.0$	$iv \in \{0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10\}$	300	4200	3