

RESEARCH

Open Access



A framework for reconstructing transmission networks in infectious diseases

Sara Najem^{1,2*}, Stefano Monni^{1,3}, Rola Hatoum², Hawraa Sweidan⁴, Ghaleb Faour⁵, Chadi Abdallah⁵, Nada Ghosn⁴, Hamad Hassan⁶ and Jihad Touma^{1,2*}

*Correspondence:
sn62@aub.edu.lb; jt00@aub.edu.lb; sn62@aub.edu.lb

¹ Department of Physics,
American University of Beirut,
Beirut, Lebanon

² Center for Advanced
Mathematical Sciences,
American University of Beirut,
Beirut, Lebanon

³ Department of Mathematics,
American University of Beirut,
Beirut, Lebanon

⁴ Epidemiological Surveillance
Program, Ministry of Public
Health, Beirut, Lebanon

⁵ National Center for Remote
Sensing, National Council
for Scientific Research (CNRS),
Beirut, Lebanon

⁶ Faculty of Public Health,
Lebanese University, Beirut,
Lebanon

Abstract

In this paper, we propose a general framework for the reconstruction of the underlying cross-regional transmission network contributing to the spread of an infectious disease. We employ an autoregressive model that allows to decompose the mean number of infections into three components that describe: intra-locality infections, inter-locality infections, and infections from other sources such as travelers arriving to a country from abroad. This model is commonly used in the identification of spatiotemporal patterns in seasonal infectious diseases and thus in forecasting infection counts. However, our contribution lies in identifying the inter-locality term as a time-evolving network, and rather than using the model for forecasting, we focus on the network properties without any assumption on seasonality or recurrence of the disease. The topology of the network is then studied to get insight into the disease dynamics. Building on this, and particularly on the centrality of the nodes of the identified network, a strategy for intervention and disease control is devised.

Keywords: Network reconstruction, Betweenness centrality, Autoregressive model, COVID-19, Optimal control

Introduction

Modeling the evolution of infectious diseases with the goal of forecasting the numbers of infections answers a wealth of questions in a range of disciplines from medicine, genetics, pharmacology to social and economic sciences [1, 2], and is at the heart of subsequent investigations on practical aspects: from management of resources (assessment of the preparedness for disease containment and readiness of the healthcare system) to possible intervention measures (vaccination and testing strategies [6], government control measures) and their consequences [7]. Some of these studies have a regional focus, investigating a disease propagation in countries, or regions of a country, while others have considered its dynamics in larger geographical contexts [3–5].

The different types of modeling that have been applied to investigate the dynamics of infectious disease have a long history in epidemic modeling. Compartmental models (e.g., SIR, SIER) are a well known class of models. They study the interplay between susceptible, infected and recovered individuals within communities, with different degrees of spatial refinement. For instance, in the so called networked compartmental models, interactions

between communities are encoded in a network [8], often to identify the spatial and temporal origin of the disease [9]. In statistics, spatial/and/or temporal point processes are often employed to study the dynamics of the disease. Some models allow for the number of infections to be triggered by those at previous times, others can incorporate, as covariates, additional available information such as demographics, human mobility, and policy decisions. Some recent work follows this direction [10–12]. Epidemic models can also be recast in a standard regression framework, where the time series of infection counts are fitted by specifying a distribution for the counts and the associated conditional mean function [13, 14]. In most of these studies, the goal is to predict and capture the spatio-temporal patterns of disease spread. The major contribution of [13, 14] is in the study of diseases which exhibit a certain periodicity or recurrence, like influenza for example. Recently, this approach was implemented on COVID-19 data [3–5] to mainly break down the contributing factors to the disease spread. We observe that this model's richness is not fully explored as it has an underlying time-evolving transmission network, which was never fully identified and whose properties have never been explored.

In this paper, we propose a generic framework for the recovery of the transmission network in infectious diseases. Our method does not assume any periodicity in the dynamics or any underlying recurrence. It hinges on the seminal work of [13, 14], which is a statistical regression analysis of infection counts over time and aggregated over localities, in a country, with a mean function that takes account of the spatial proximity of these localities. The fitted model is then used to reconstruct a weighted network, which constitutes the second component of our framework. The salient point of our method of network recovery is that smoothness conditions on the temporal data are not required [15], and neither is the near steady-state dynamics that is instead necessary for the perturbation/response approaches to work [16–18]. Our reconstruction is similar to that in [19], while it differs from procedures that rely on the deterministic evolution of the disease [20, 21]. The third component of our framework is the study of the changes in the topology of the underlying recovered network and the computation of centrality measures (specifically the nodes' betweenness centrality), from which the recommendation of optimal control measures ensues. We apply the method in the case of COVID-19 in Lebanon.

Model definition

Our starting point of the analysis is a statistical model that captures the spatio-temporal dynamics of the infections, under the statistical framework discussed in [13]. Namely, we consider the number Y_{it} of infections recorded in a given locality i in a given day t , as independent, conditionally on the counts at previous times, random variables distributed according to a negative binomial distribution having a mean function decomposed into three terms as follows $\mu_{it} = E(Y_{it}|Y_{j,t-1}, e_i)$:

$$E(Y_{it}|Y_{j,t-1}, e_i) = \lambda_{it} Y_{i,t-1} + \phi_{it} \sum_{j \neq i} \omega_{ij} Y_{j,t-1} + v_{it} e_i. \quad (1)$$

The first two terms constitute the auto-regressive part of the model: one being the contribution to the mean infection μ_{it} in locality i at time t , due to the infections within i at the previous day, the other being the contribution to μ_{it} due to positive cases from other localities j also at the previous day. The final term accounts for all other contributions not

captured by the first two, such as infected people who entered the country under study from abroad. For simplicity we will refer to the last term $v_{it}e_i$ as the component due to travel and assume that it is proportional to the size of the population e_i of the locality. The log-transforms of non-negative coefficients λ_{it} and ϕ_{it} , which quantify the contribution of the past observations to future counts, and the log-transform of the parameter v_{it} are each modeled as a linear function of time, with a locality-specific slope to allow more flexibility across localities. Intercepts and slopes are estimated from the data. Finally, we model ω_{ji} as a power function of the geographical distance d_{ij} of the localities: $\omega_{ji} \propto d_{ij}^{-f}$. This is assumed because previous studies have shown that mobility flows are governed by power-law functions of inter-localities distances [22–25].

Network identification and characterization

We wish to focus now on the inter-locality term and study it from a different perspective. To do so, observe that the second term in the mean equation (1) can be re-written as follows: $\sum_{j \neq i} A_{ij}(t)Y_{j,t-1}$, where $A_{ij}(t) = \phi_{it} \cdot \omega_{ij}$. It can be interpreted as the contribution to the cases at time t in locality i from cases from locality j at the previous day. We can suggestively think of $A = (A_{ij})$ as defining the weights of a network between localities: the transport network ω describes the traffic flow between localities, and thus predates the disease, while ϕ_{it} is the number of transported cases from i into neighboring localities. $A(t)$ explicitly depends on t since the coefficients of $\log \phi_{it}$, which are linear functions of t , and the power f in the definition of ω are estimated over each interval $[0, t]$.

This complex network A drives the cross-localities dynamics. We will now suggest employing some useful summary metrics for $A(t)$ and its time evolution in order to understand its properties, and accordingly prescribe adequate control measures.

One useful summary metric is the modularity, which is a measure of cluster formation in a network. More specifically, the modularity Q of a given network A is defined with respect to a given grouping of its nodes. We follow [27] where the grouping of the nodes is determined by a stochastic procedure that reveals densely connected subgraphs. An illustration of a grouping is given in Fig. 1.

Given this group membership, the modularity of A is then computed according to the formula:

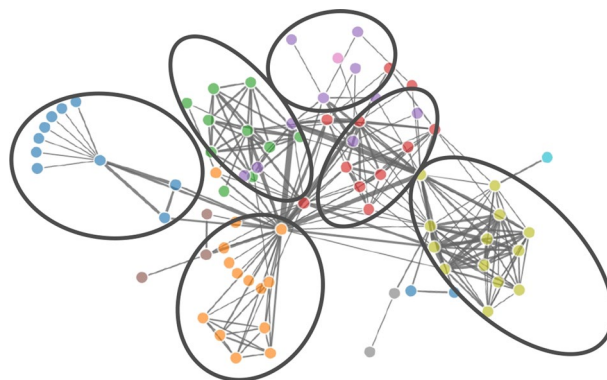


Fig. 1 An example of modular network along with its detected communities (encircled) is shown. The nodes are color-coded based on their memberships to these communities

$$Q(A) = \frac{1}{2h} \sum_{i,j} \delta(c_i, c_j) (A_{ij} - k_i k_j / (2h)),$$

where h denotes the total number of edges, k_i and k_j are the degree of nodes i and j respectively, c_i labels the group to which i belongs, and δ is the Kronecker delta.

We also suggest to analyse additional topological measures for the networks: mainly, the clustering coefficient, the average path length, and the strength distribution [29, 30]. These are generally used to classify networks into random, scale-free, or regular. The clustering coefficient C of a network is a measure of transitivity that counts the ratio of the number closed triplets to the number of all (closed and open) triplets. A triplet is closed if all the three connections between the three nodes exist and is open if one of the links is missing. The average path length l of a network is given by the mean distance over all pairs of vertices, where distance is the number of edges in the shortest path joining them. An illustration is shown in Fig. 2.

Finally, a node's strength is the sum of the weights of its edges. Namely, for the i -th node:

$$s_i = \sum_j A_{ij}$$

Small-world or scale-free networks (that is, networks with node degrees and strengths distributed according to a power-law) are characterized by high clustering coefficients and low average path lengths compared with those of regular/ordered graphs [29, 30]. Random graphs are, on the other hand, characterized by low average path lengths and low clustering coefficients compared to regular graphs. An illustration of the three different network types is shown in Fig. 3.

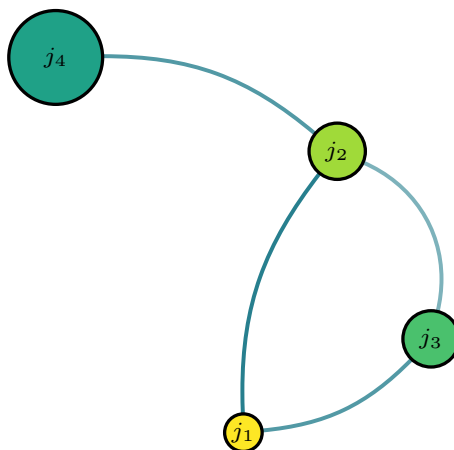


Fig. 2 Example of a network with $C = 1/3$ and $l = 8/6$ is shown. It has three triplets, one of which is closed (triangle). The lengths of the paths from j_2 to all others is 1, while that from j_1 to j_4 is 2

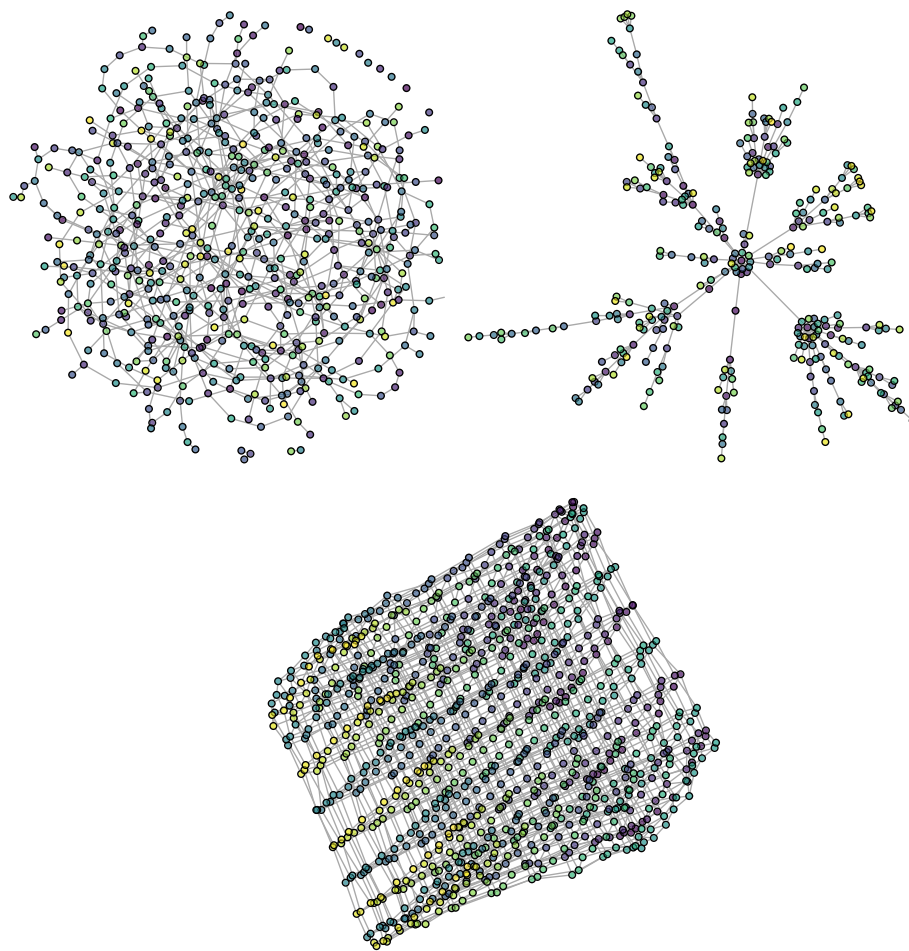


Fig. 3 Examples of random, scale-free, and ordered (lattice-like) networks are shown respectively

Data description and fitting of the regression model

As an application of the framework, the model (1) was applied to the COVID-19 data collected in Lebanon. On daily basis, the laboratories from the public and private sectors report the confirmed cases to the Epidemiological Surveillance Program of Lebanon's Ministry of Public Health (ESUMOH). Later, the cases are investigated in order to get additional demographic information and health condition. The data are then archived in a national platform. Specifically, the data we have considered consist of counts of COVID-19 recorded daily in each of the 1544 localities of Lebanon from February 21, 2020 to January 20, 2022. Such localities correspond to Lebanon's smallest statistical units called "circonscriptions foncières" or cadastral villages following the Central Administration for Statistics (CAS) nomenclature [26]. Recommendations on possible interventions and updates on the disease evolution were sought for by the Ministry of Public Health at 20 days intervals. Model (1) was fitted using the R package *surveillance* [14] over the intervals $[0, t]$, $t = 20 \cdot n$ for $n = 1, \dots, 41$. This allows us to follow the evolution in time of the model parameters until day 820, the last observation point.

Figure 4 displays the aggregated counts over all localities, $\sum_i Y_{i,t}$, and the fitted values over the complete time period of our study broken down into the three

Table 1 Comparison of the two models with the same mean function given in Eq. (1) with distribution of the counts being negative binomial and Poisson

	ses	AIC
Negative binomial	2.88	1,164,119
Poisson	3.65	1,333,453

AIC is Akaike's Information criterion. SES is the mean squared error: the mean squared difference over the localities of the observed and predicted counts at the final time point of the study

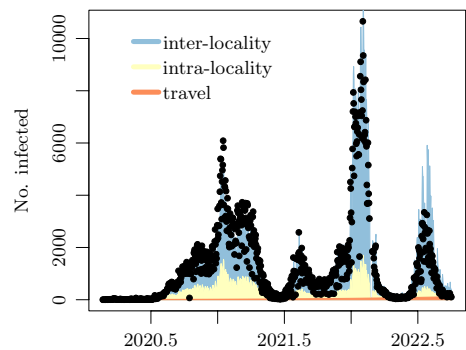


Fig. 4 Data and fit under the model of Eq. (1) with negative binomial distributed counts. The three colors show the decomposition of the fitted aggregate counts into travel, intra-locality, and inter-locality contributions to infections amounting to 3%, 10%, and 87% respectively

components of the mean function. The fit appears quite adequate. It is in fact a better fit to the data than the model with counts assumed to be Poisson-distributed, which is an indication of overdispersion in the data. A further comparison of these two models in terms of AIC value and prediction errors is provided in Table 1.

We further notice that the model of equation (1) considers a time-lag of one day; that is, the future counts depend on the counts recorded on the previous day. Changes in the time-lag from one to a few days did not result in any noticeable difference. Our analysis provides evidence that the inter-locality infection drives the overall transmission of the disease [22]. Then, for this reason we shift our focus to the network that governs the interaction between localities and observe that it is not purely a static spatially-dependent network but rather dynamic and time-evolving; in fact the product of time-dependent coefficients with the spatial proximity matrix. Figure 4 indicates that the inter-locality term has the most important contribution to the increase in the mean number of infections compared to the intra-locality and travel terms. This suggests that the inter-locality transmissions should be the main focus of analysis, and what one learns from their study would be useful for disease control.

The parameter estimates of model (1) for all 1554 localities and their errors can not be displayed in an uncluttered fashion but are available from the corresponding authors. A sample of the evolution of the inter-locality term ϕ_{it} is shown in Fig. 5.

An example of the reconstructed network is provided in Fig. 6 which is a graphical representation of $A^{(15)}(t = 300)$. The superscript 15 in the notation of A indicates that the latter was estimated on the counts data of 15 contiguous 20-day time intervals, that is the 300-day time span $[0, 20 \cdot 15]$ from February 21, 2020 to December 16th, 2020.

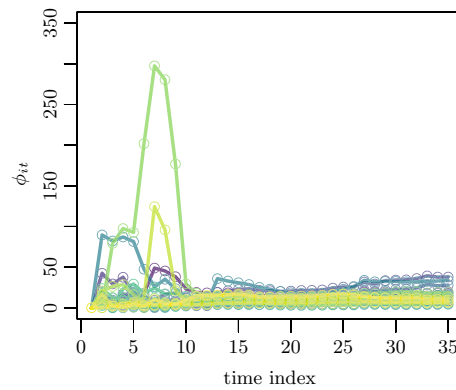


Fig. 5 Time evolution of the inter-locality parameter ϕ_{it} is shown for the regions with the highest centrality (see Sect. 3)

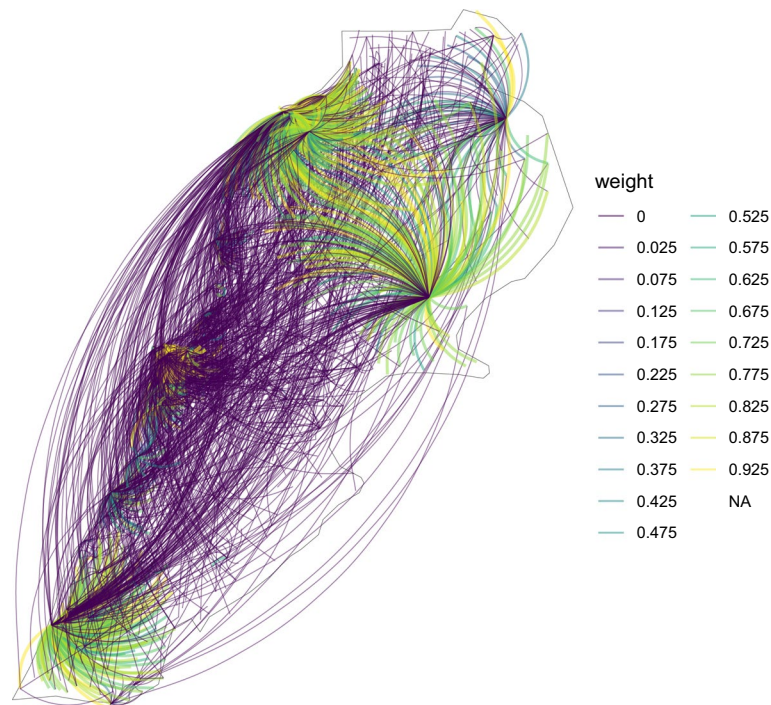


Fig. 6 The network $A^{(15)}(t = 300)$ is overlaid on the map of Lebanon to illustrate its complexity. It is evaluated at the 300-th day of the pandemic using all 15 time interval of our collected data, that is the time span $[0, 20 \cdot 15]$

Further, Fig. 7 shows the modularities of the 41 matrices $A^{(41)}(t)$ at days $t = 20 \cdot n$ for $n = 1, \dots, 41$. The superscript n in the notation $A^{(n)}(t)$, as mentioned above, indicates that A is estimated using the counts of the $20 \cdot n$ days of the study.

Figure 8 shows the clustering coefficients and the average paths lengths for the matrices $A^{(41)}(t)$. Similar behavior of both C and l was observed for all $A^{(n)}$, with $n \geq 10$.

One can see a jump in modularity on the tenth 20-day time interval, which we will denote by I_c . This behavior may signal the onset of an emerging power-law [28]. The evolution of both C and l gives additional evidence for a transition at a point I_c . The

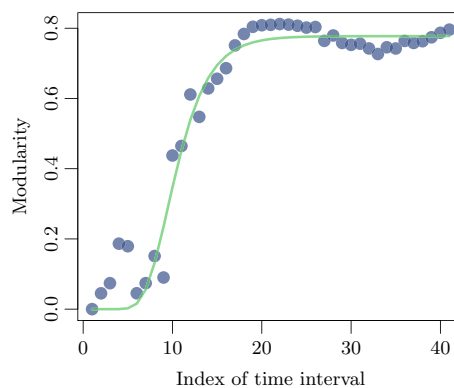


Fig. 7 The modularity of the network $A^{(n)}(t)$ is shown as a function of time $t = 20 \cdot n$, with $n = 1 \dots 41$. It is measure of the quality of the division of a graph into subgraphs

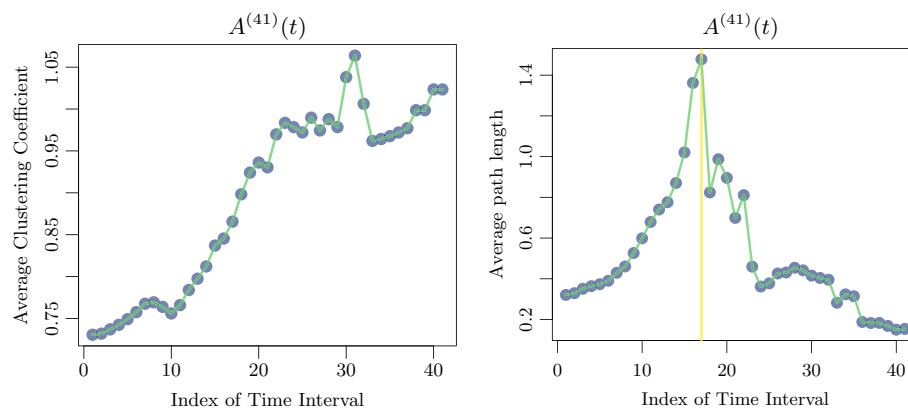


Fig. 8 The figure shows the clustering coefficient and average path length for $A^{(41)}(t)$, which are the matrices estimated using the data from the start of the pandemic to the 820-th day ($41 \cdot 20$) evaluated at $t = 20 \cdot n$, where n is the index of the time intervals

clustering coefficient starts suddenly to increase. At the 10-th interval there is an abrupt jump in the average path length as well at I_c (Fig. 8). This is an indication of scale-freeness of the network. This property expedites the spread of epidemics unlike what would occur in ordered networks, which are characterized by a slower spread because they possess a high C and an l that scales with system size [31, 32].

To characterize the transition to scale-freeness, we now analyse the distribution of the strengths of the nodes, as additional evidence for change in the network topology at the 10-th interval I_c . Figures 9 and 10 show the empirical and estimated distributions of the strengths (in fact, the survival function $P(S > s)$) of $A^{(n)}(t = 20 \cdot n)$, at the time intervals $n = 1, \dots, 41$ on a log-log scale. We note that a transition occurs at $t = 20 \cdot 10$, where the distribution becomes linear, which is indicative of a power-law (Pareto distribution): $P(X \leq x) = 1 - (\beta/x)^{\alpha-1}$, for $x \geq \beta$. The exponent α and the boundary value (scale) β are estimated by maximum likelihood following [33].

Figure 11 summarizes the estimates of the exponents for these networks and their standard errors (obtained by non-parametric bootstrap).

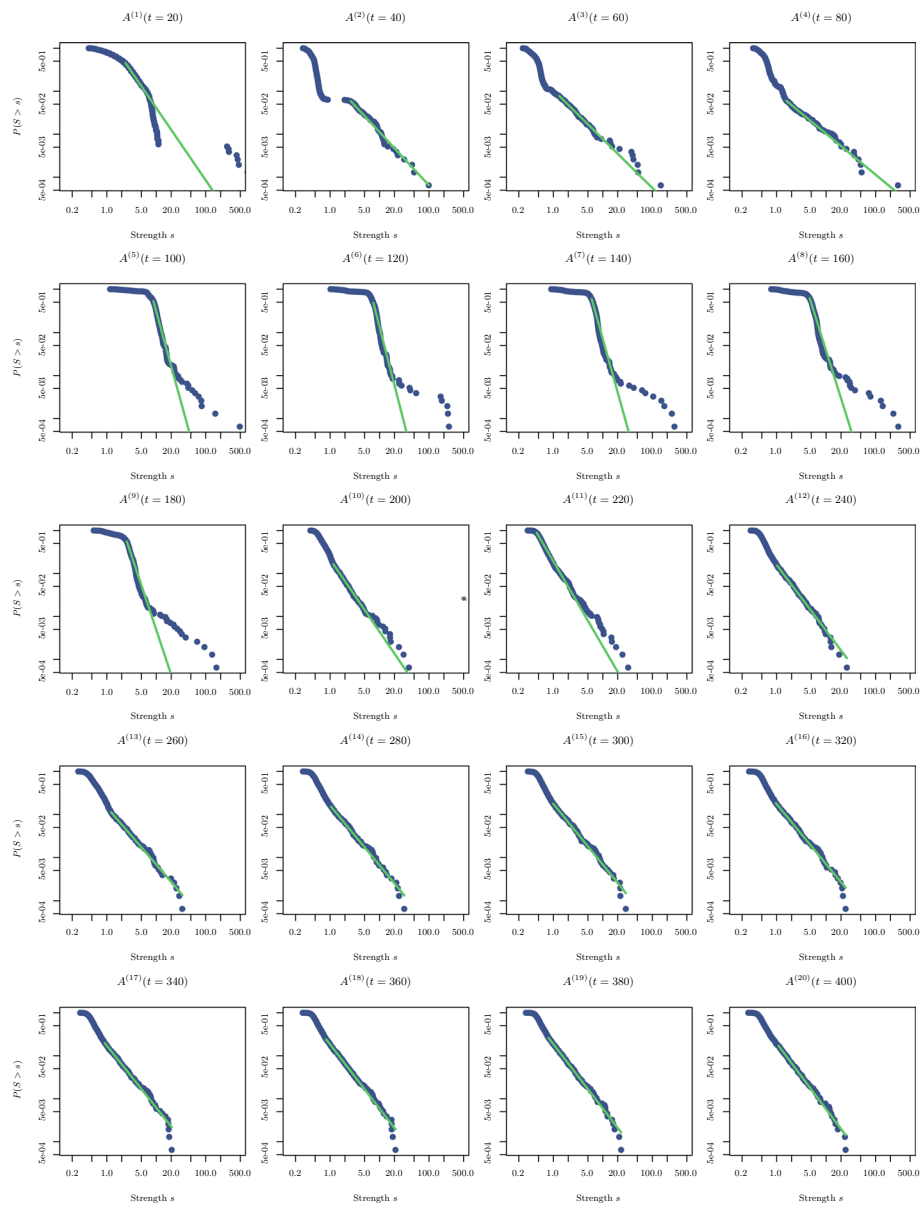


Fig. 9 Empirical and estimated distributions of the strength for the matrices $A^{(n)}(t = 20 \cdot n)$ with $n = 1, \dots, 20$ are shown

After the 180th day, that is for time intervals labeled by the index $n \geq 10$, most power-laws have very close exponents of about 2.5. This signals the stabilization of the network topology. Thus, I_c marks the onset of the emergence of the steady state network. We think that only above this point any prescription of control measures is likely to be efficient as the revealed network topology, relying on the daily counts, has stabilized. One can wonder if there is any explanation on why the stable phase has set in during this interval I_c , and not before or after it. I_c chronologically coincides with the period between August, 19, 2020, and September 7, 2020. Perhaps, the blast in Beirut which occurred on August 4th and in the following weeks of social protests, personal precaution measures (such as social distancing and wearing of masks) were compromised.

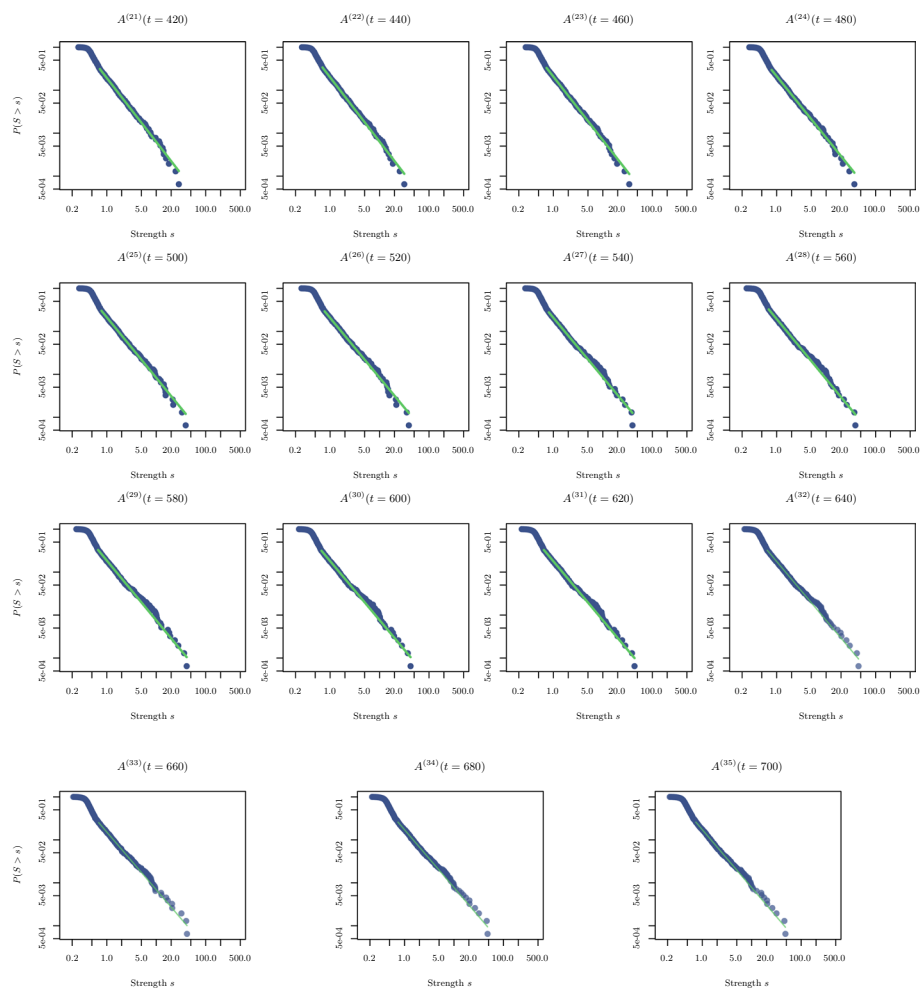


Fig. 10 Empirical and estimated distributions of the strength for the matrices $A^{(n)}(t = 20 \cdot n)$ with $n = 21, \dots, 35$ are shown

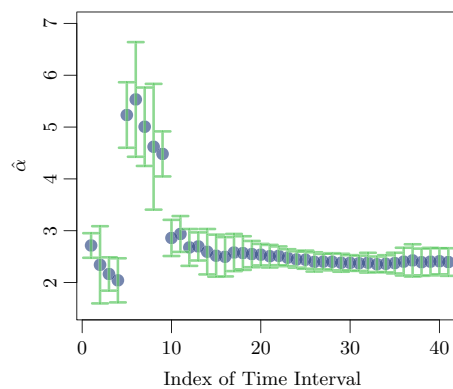


Fig. 11 Estimated values $\hat{\alpha}$ of the exponent of the power-law distributions of the strengths of the nodes of the 41 matrices $A^{(n)}(t = 20 \cdot n)$, $n = 1, \dots, 41$, which are represented in Fig. 9. Vertical bars indicate $\pm 2\delta_{\hat{\alpha}}$

Either of these occurrences may have contributed to the detected change in the network type. See the Appendix for the chronology.

Putting the analysis into action: control measures

Having fully characterized the network and identified the steady-state, we now turn to a possible use of this analysis to guide an optimal strategy for disease control. The strategy will identify some localities as candidates for being isolated or for having their connections to other localities curtailed. The measure on which the identification is based is that of centrality of a node. The betweenness centrality of a node v is defined as [30]:

$$g(v) = \sum_{i \neq v \neq j} \frac{\sigma_{ij}(v)}{\sigma_{ij}}$$

where σ_{ij} is the total number of shortest paths from node i to node j passing through $\sigma_{ij}(v)$. Therefore, the more central the node is, the more its removal has an effect on the network's connectivity, since its removal would yield a network with more disconnected subgraphs. The control strategy we propose involves an iterative procedure, where at each step the centralities of the nodes are computed, the node with the resulting highest centrality is removed, and the matrix A is updated, as illustrated in Fig. 12.

Other removal schemes of nodes in network exist, but the one we have just described has been suggested to incur the highest loss of connectivity for scale-free networks [34–39]. In practice, candidate targets for intervention the localities corresponding to nodes with higher centralities. We notice that at the policy level this strategy based on our analysis was indeed adopted. The localities we have identified through this strategy were given priority in the national vaccination campaign. On the other hand, the recommendations we put forward based on this analysis were only partially adopted in targeting the high centrality localities for lockdown and intervention measures, as the decision making process involved other ministries and stakeholders. However, we conclude by considering theoretically the would-be repercussions of such implementation. Clearly, the loss of connectivity would impede the evolution of the disease since the localities which are contributing the most to the infection would be isolated. For example, removing around 20% of the most connected localities on the basis of their betweenness centrality would lead to 80% loss of connectivity as shown in Fig. 13.

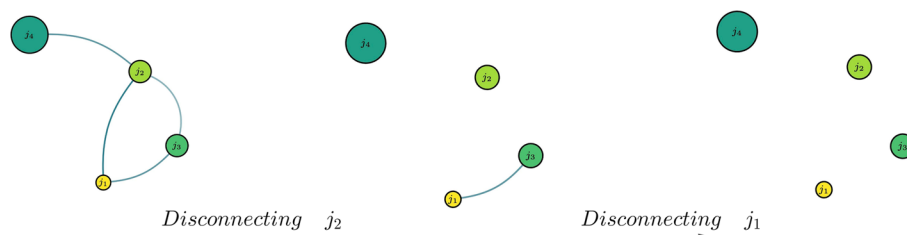


Fig. 12 The figure illustrates the iterative scheme. First, the node with the highest centrality j_2 is disconnected by removing all its links, as it is the node with the highest number of shortest paths. The resulting network has j_1 and j_3 with equal centralities and either one can be disconnected. This leads to a total loss of connectivity in the network at the end of the process

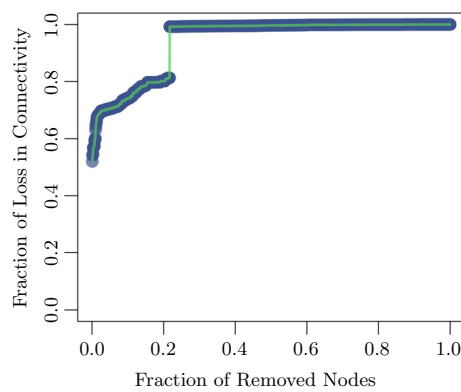


Fig. 13 The loss of connectivity versus the fraction of removed nodes for the cascading and non-cascading strategies

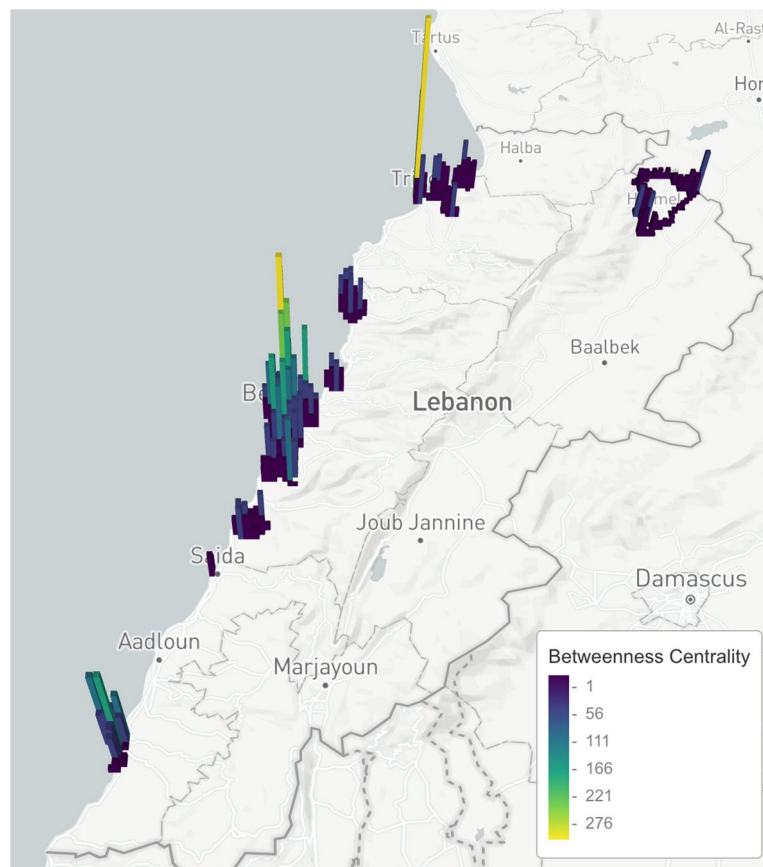


Fig. 14 The map shows the localities with the highest centrality whose removals lead to 80% loss of connectivity

Specifically, the localities causing 80% loss of connectivity are shown in Fig. 14, while the fitted model of the top sixteen localities is shown in Fig. 15. An animated map of the control strategy is available on this <https://www.dropbox.com/s/guhamz3p7op6b3y/animated.gif?dl=0>.

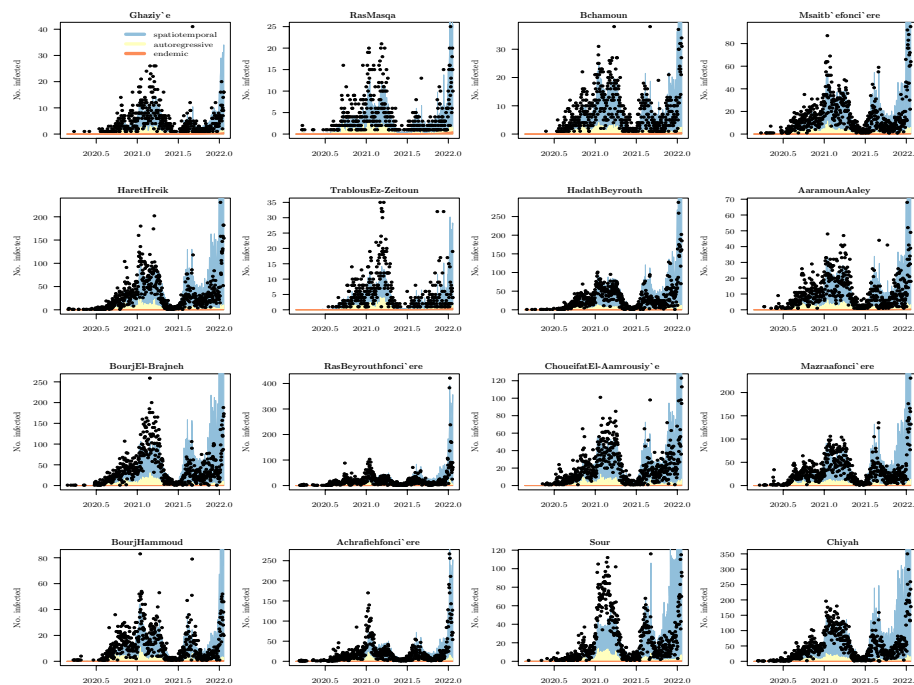


Fig. 15 The figure shows the counts, along with the fitted model, for twelve localities ordered by decreasing betweenness centrality

Conclusion

In this paper, we have proposed a framework that can be used to inform control measures for epidemics in a country for which infections counts aggregated over local regions are available over time. In particular, and as an example, we have followed the evolution of the counts of COVID-19 cases in Lebanon at the level of local administrative units at a daily resolution. The framework entails fitting an auto-regressive model to the data; recovering an underlying network over which the disease propagates; analyzing such time-evolving network to identify topological measures of node centrality that suggest an optimal control of the spread of the disease. Specifically, for the data about COVID-19 in Lebanon the analysis of the topological metrics of the network has given us a hint into a transition to a steady state structure that governs interactions between localities. After identifying this steady state network, and characterizing it as a scale-free, we have proposed control measures based on betweenness centrality of its nodes. The findings were taken into consideration in the national vaccination campaign for COVID-19, with the identified localities given priority for vaccination.

Appendix: Chronology of COVID-19 pandemic in Lebanon

In this appendix we summarize the chronology of the COVID-19 pandemic in Lebanon from the first recorded case to February 2022. We divide this time interval into 4 periods, and highlight the main governmental interventions taken to control the spread of the disease.

Period 1 (February 2020 to June 2020). The first cases are documented. Early lockdown measures are implemented with airport closure. Testing is carried out for suspected cases, close contacts, and travelers. Cases are mainly within clusters. Aggressive contact tracing is adopted.

Period 2 (July 2020 to December 2020): The airport reopens in July 2020. The daily number of cases increases progressively and community transmission sets in. On August 4th, the Beirut blast occurs.

Period 3 (January 2021 to June 2021): The alpha variant is introduced. The case counts increase. Lockdown measures are implemented resulting in a decrease of the recorded cases. However, after lockdown release, an increase of the number of infections is observed until mid-March, with a progressive and sustained decrease up to June.

Period 4 (July 2021 to December 2021). Introduction of the delta variant, which progressively replaces the alpha variant. Two waves of delta are observed: July–September and November–December.

Period 5 (January 2022 to February 2022). Introduction of the omicron variant. High transmissibility of the new variant leads to high daily case counts reaching 10,000 on 1st Feb 2022.

Abbreviations

ESUMOH Epidemiological Surveillance Program of Lebanon's Ministry of Public Health
CAS Central Administration for Statistics

Acknowledgements

The authors acknowledge the support of Prof. Rima Habib from the AUB's Faculty of Public Health for her unfailing support and invaluable advice to physicists and mathematicians as they worked with public data.

Author contributions

J.T contributed to the problem formulation and led the project, and together with S.N and S.M developed the model and numerical analysis involved in this manuscript. G.F and C.A worked on the geographical and the demographic data, R.H and H.S worked on the data processing of infection counts, N.G and H.H on the epidemiological analysis and the significance of the results and their implementation in the vaccination strategy. All authors read and approved the final manuscript.

Funding

The authors received no funding.

Availability of data and material

The data is available upon request from the Ministry of Public Health's Surveillance Program.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 6 September 2022 Accepted: 12 December 2022

Published online: 19 December 2022

References

1. Mohamadou Y, Halidou A, Kapen PT (2020) A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Appl Intell* 50(11):3913–3925
2. Guan J, Wei Y, Zhao Y, Chen F (2020) Modeling the transmission dynamics of COVID-19 epidemic: a systematic review. *J Biomed Res* 34(6):422
3. Celani A, Giudici P (2021) Endemic-epidemic models to understand COVID-19 spatio-temporal evolution. *Spat Stat* 49:100528
4. Ssentongo P, Fronterre C, Geronimo A, Greybush SJ, Mbabazi PK, Muvawala J, Nahalamba SB, Omadi PO, Opar BT, Sinnar SA et al (2021) Pan-African evolution of within-and between-country COVID-19 dynamics. *Proc Natl Acad Sci* 118(28):e2026664118

5. Dickson MM, Espa G, Giuliani D, Santi F, Savadori L (2020) Assessing the effect of containment measures on the spatio-temporal dynamic of COVID-19 in Italy. *Nonlinear Dyn* 101(3):1833–1846
6. Gozzi N, Bajardi P, Perra N (2021) The importance of non-pharmaceutical interventions during the COVID-19 vaccine rollout. *medRxiv*
7. Perra N (2021) Non-pharmaceutical interventions during the COVID-19 pandemic: a review. *Phys Rep* 913:1–52
8. Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science* 342(6164):1337–1342
9. Schlosser F, Brockmann D (2021) Finding disease outbreak locations from human mobility data. *EPJ Data Sci* 10(1):52
10. Zhu S, Bukharin A, Xie L, Santillana M, Yang S, Xie Y (2021) High-resolution spatio-temporal model for county-level COVID-19 activity in the US. *ACM Trans Manag Inf Syst (TMIS)* 12(4):1–20
11. Chiang W-H, Liu X, Mohler G (2021) Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *Int J Forecast* 38:505–520
12. Giudici P, Pagnottoni P, Spelta A (2021) Network self-exciting point processes to measure health impacts of COVID-19. Available at SSRN 3892998
13. Held L, Höhle M, Hofmann M (2005) A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Stat Model* 5:187–199
14. Meyer S, Held L, Höhle M (2014) Spatio-temporal analysis of epidemic phenomena using the R package *surveillance*. *arXiv preprint [arXiv:1411.0416](https://arxiv.org/abs/1411.0416)*
15. Shandilya SG, Timme M (2011) Inferring network topology from complex dynamics. *New J Phys* 13(1):013004
16. Prabakaran S, Gunawardena J, Sontag E (2014) Paradoxical results in perturbation-based signaling network reconstruction. *Biophys J* 106(12):2720–2728
17. Yu D (2010) Estimating the topology of complex dynamical networks by steady state control: generality and limitation. *Automatica* 46(12):2035–2040
18. Yu D, Parlitz U (2010) Inferring local dynamics and connectivity of spatially extended systems with long-range links based on steady-state stabilization. *Phys Rev E* 82(2):026108
19. Wan X, Liu J, Cheung WK, Tong T (2014) Inferring epidemic network topology from surveillance data. *PLoS ONE* 9(6):100661
20. Pajevic S, Plenz D (2009) Efficient network reconstruction from dynamical cascades identifies small-world topology of neuronal avalanches. *PLoS Comput Biol* 5(1):1000271
21. Braunstein A, Ingrosso A, Muntoni AP (2019) Network reconstruction from infection cascades. *J R Soc Interface* 16(151):20180844
22. Meyer S, Held L (2014) Power-law models for infectious disease spread. *Ann Appl Stat* 8(3):1612–1639
23. Zipf GK (1946) The p 1 p 2/d hypothesis: on the intercity movement of persons. *Am Sociol Rev* 11(6):677–686
24. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100
25. Barthélemy M (2011) Spatial networks. *Phys Rep* 499(1–3):1–101
26. Verdeil E, Faour G, Velut S, Hamzé M, Mermier F (2007) *Atlas du LIBAN*. Presses de l'Ifpo
27. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3–5):75–174
28. Grindrod P, Higham DJ (2018) High modularity creates scaling laws. *Sci Rep* 8(1):1–9
29. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74(1):47
30. Newman ME (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
31. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. *Rev Mod Phys* 87(3):925
32. Newman ME (2002) Spread of epidemic disease on networks. *Phys Rev E* 66(1):016128
33. Clauset A, Shalizi CR, Newman ME (2009) Power-law distributions in empirical data. *SIAM Rev* 51(4):661–703
34. Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance of complex networks. *Nature* 406(6794):378–382
35. Dong G, Gao J, Du R, Tian L, Stanley HE, Havlin S (2013) Robustness of network of networks under targeted attack. *Phys Rev E* 87(5):052804
36. Valdez LD, Shekhtman L, La Rocca CE, Zhang X, Buldyrev SV, Trunfio PA, Braunstein LA, Havlin S (2020) Cascading failures in complex networks. *J Complex Netw* 8(2):013
37. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464(7291):1025–1028
38. Albert R, Albert I, Nakarado GL (2004) Structural vulnerability of the North American power grid. *Phys Rev E* 69(2):025103
39. Edsberg Møllgaard P, Lehmann S, Alessandretti L (2021) Understanding components of mobility during the COVID-19 pandemic. *Phil Trans R Soc A* 380(2214):20210118

Publisher's Note

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