

NETWORK SCIENCE

Effective approach to epidemic containment using link equations in complex networks

Joan T. Matamalas*, Alex Arenas*, Sergio Gómez*

Epidemic containment is a major concern when confronting large-scale infections in complex networks. Many studies have been devoted to analytically understand how to restructure the network to minimize the impact of major outbreaks of infections at large scale. In many cases, the strategies are based on isolating certain nodes, while less attention has been paid to interventions on the links. In epidemic spreading, links inform about the probability of carrying the contagion of the disease from infected to susceptible individuals. Note that these states depend on the full structure of the network, and its determination is not straightforward from the knowledge of nodes' states. Here, we confront this challenge and propose a set of discrete-time governing equations that can be closed and analyzed, assessing the contribution of links to spreading processes in complex networks. Our approach allows a scheme for the containment of epidemics based on deactivating the most important links in transmitting the disease. The model is validated in synthetic and real networks, yielding an accurate determination of epidemic incidence and critical thresholds. Epidemic containment based on link deactivation promises to be an effective tool to maintain functionality of networks while controlling the spread of diseases, such as disease spread through air transportation networks.

INTRODUCTION

The problem of modeling the spread of a disease among individuals has been studied in depth over many years (1–4). The development of compartmental models, models that divide the individuals among a set of possible states, has given rise to a new collection of techniques that enable, for instance, analysis of the onset of epidemics (5–15) or study of the impact of a vaccination campaign (16–20). Previous studies have relied heavily on a mathematical approach to the study of epidemic spreading (21), and we follow a similar approach here in the same spirit.

The design of effective containment strategies constitutes a major challenge. Measures such as vaccination, improved hygiene, biosecurity, cattle culling, or education to prevent contagions operate on the biological aspects of the disease. On the other hand, isolation or mobility restrictions act on the physical routes or patterns of disease spread, which may transform a local event into a pandemic. Here, we concentrate on the role of the links of the spreading network. For example, if we identify the edges that are more involved in the propagation of a disease, then it is possible to design targeted countermeasures that affect only specific links instead of whole nodes while being more effective. This approach can be illustrated by a hypothetical pandemic disease propagated using the air transportation network: The isolation of one airport is a dramatic measure that is socially and politically difficult to accept and put into practice, but the suspension of only a few connections between selected airports could be more easily assumed and could also achieve a better containment of the disease.

Previous studies have directed their attention mostly toward schemes based on the actuation on single nodes, either randomly or according to node properties such as their degree, betweenness, PageRank, or eigenvector centrality (22–25). Following the same idea, some authors have introduced link removal using properties of the adjacent nodes (degrees or centralities) or of the link itself (edge betweenness) (22, 26, 27). A model of coevolution of epidemics with permanent and temporal link removals was proposed in (28), and methods from optimization and control have been applied to minimize the impact of the epidemics

(29–31). The core of what is currently considered to be the optimal approach is built upon finding the minimum set of edges whose removal leads to a maximum decrease in the spectral radius of the network, that is, the largest eigenvalue of the adjacency matrix (27, 32, 33). Because the epidemic threshold is, at first-order approximation of a susceptible-infected-susceptible (SIS) epidemic dynamics, inverse to the spectral radius, it seems the best and more mathematically grounded option. Unfortunately, it turns out to be an NP-complete problem; thus, only heuristics are available for large networks (27).

It must be emphasized that all the previous approaches make use only of the structural characteristics of the network to decide which nodes or edges have to be removed; the characteristics or parameters of the epidemic process are ignored. Even the spectral radius, which is closely related to the epidemic threshold, does not depend on the infection or recovery rates, the expected number of infected neighbors around a certain node, or any other local or global information of the spreading process.

Our proposal concentrates on the role of the links in the spreading of the epidemics, quantifying the importance of each link (34) and thus enabling containment strategies based on their removal. To this end, we first define the epidemic importance of a link as its capacity to infect other individuals once this link has been used to propagate the disease. The determination of this link epidemic importance requires the development of a mathematical model that is able to cope with the infection propagation at the level of links in complex networks. We will show that the proposed model facilitates the determination of the epidemics incidence and threshold with high accuracy. Moreover, the quantification of the epidemic importance of the edges leads to a link removal strategy that, in many cases, outperforms the previous approaches, even those based on minimizing the spectral radius, and also preserves most of the connectivity of the network.

The contributed model is built upon the relationships between the states of nodes connected by links, thus being related to pairwise approximations (35–41). However, our model is microscopic, at the level of individual links as in (40), thus allowing a clear identification of maximally infectious links, but with the additional advantage of being able to easily calculate not only the epidemic threshold but also the incidence of the epidemics and the importance of the links.

Copyright © 2018
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain.

*Corresponding author. Email: joantomas.matamalas@urv.cat (J.T.M.); alexandre.arenas@urv.cat (A.A.); sergio.gomez@urv.cat (S.G.)

RESULTS

Link epidemic importance

Let us consider a discrete-time SIS dynamics that runs on top of a complex network of N nodes and L edges, with adjacency matrix A , and where each node i can be in one of two different states σ_i , either susceptible (S) or infected (I), that is, $\sigma_i \in \{S, I\}$. We can say that a link (i, j) between nodes i and j is in state SI if $\sigma_i = S$ and $\sigma_j = I$. The parameters of the SIS dynamics are the infection and recovery probabilities, β and μ , respectively.

Our objective is to find an effective strategy to contain the SIS epidemic process through bond percolation. To determine which link should be removed first, we need a measure of the importance of each link in the spreading of the epidemics. A possible option would be to use the edge importance defined in (34), which accounts for the relative change of the spectral radius when the edge is removed. However, this constitutes an indirect way of containment, because we aim at lowering the incidence of the epidemics as much as possible, whereas the actuation on the spectral radius is directed to increase the epidemic threshold; both are different targets. In addition, the spectral radius only depends on the structure of the network, but not on the epidemic parameters of the process or the participation of the link in the spreading of the disease.

We assume that the system has reached the stationary state, which does not mean that the nodes remain in a certain fixed state, only that the average incidence of the epidemics is basically constant; thus, there is still margin for applying a containment strategy to minimize this incidence. In this regime, we can measure the probabilities of nodes and links in each of the epidemic states, for example, the probability $P(\sigma_i = I)$ of node i being infected or the joint probability $P(\sigma_i = I, \sigma_j = S)$ of link (i, j) being in state IS .

Consider we have a link in state SS or II . In both cases, the next step of the epidemic dynamics is not going to use this link because, in the former, there is no infected node to propagate the disease and, in the latter, both nodes are already infected. Thus, to propagate the epidemics, a link must be in an either SI or IS state. Let us suppose we have a link (i, j) in state IS . First, with probability β , node i can infect node j through this link, changing to state II . Next, infected node j may transmit the disease to some of its neighbors. Thus, if we had removed link (i, j) , then we would have cut this path of infections initiated at node i . This means that the larger the expected number of infected neighbors of node j , the largest the impact of removing link (i, j) for the spreading of the epidemics. Note that the degree of j , as well as the probability of its neighbors being susceptible when j is infected, is relevant, because you cannot infect nodes that are already infected. For example, if j is surrounded by many infected nodes, then cutting link (i, j) is not going to have too much effect on the overall incidence of the epidemics. The expected number of infected nodes produced in this way can be expressed as

$$\bar{n}_{ij} = \beta P(\sigma_j = S, \sigma_i = I) \sum_{r=1}^N A_{jr} \beta P(\sigma_r = S | \sigma_j = I) \quad (1)$$

where $P(\sigma_r = S | \sigma_j = I)$ is the conditional probability that node r is susceptible when its neighbor j is infected. Because this measure is asymmetric, and removing an edge affects the propagation of the disease in both directions, we define the link epidemic importance of a link, I_{ij} , as

$$I_{ij} = \bar{n}_{ij} + \bar{n}_{ji} \quad (2)$$

Now, the problem reduces to finding the joint and conditional probabilities for each link, and this is accomplished using our epidemic link equations (ELE). It can be shown that this definition of link epidemic importance has the property of trying to preserve the connectivity of the network (see section S1 and fig. S1), unlike other options such as edge betweenness, which quickly tend to produce a large number of disconnected components, thus hindering the functionality of the network.

Epidemic link equations

To simplify the notation, we first denote the previous joint probability as $\Phi_{ij} = P(\sigma_i = S, \sigma_j = I)$; the higher the Φ_{ij} , the larger the likelihood that the disease propagates from node j to node i . It is worth mentioning that this feature is generally asymmetrical, meaning that the propagation of the illness can be more probable from j to i than the other way around. In the same way, the epidemic is restrained by edges where the nodes are in the same state; thus, it is convenient to define the probabilities $\Theta_{ij}^S = P(\sigma_i = \sigma_j = S)$ and $\Theta_{ij}^I = P(\sigma_i = \sigma_j = I)$ for all pairs of neighboring nodes.

The evolution of the joint probability Φ_{ij} of one link depends on the probabilities Φ , Θ^I , and Θ^S to the rest of the neighboring links and on the infection rules of the SIS dynamics. Thus, we can write the following equation for each link

$$\begin{aligned} \Phi_{ij}(t+1) &= \Theta_{ij}^S(t) q_{ij}(t) (1 - q_{ji}(t)) + \Phi_{ij}(t) ((1 - \beta)q_{ij}(t))(1 - \mu) + \\ &\Phi_{ji}(t) \mu (1 - (1 - \beta)q_{ji}(t)) + \Theta_{ij}^I(t) \mu (1 - \mu) \end{aligned} \quad (3)$$

where $q_{ij}(t)$ stands for the probability that a susceptible node i is not infected by any of its neighbors (excluding node j). We have taken into account all the possible changes of state of the nodes i and j . The first term considers the probability that both nodes are in a susceptible state; then, node i remains susceptible, while node j is infected by any of its other neighbors. The second term accounts for both nodes remaining in the same state; node i is not infected by any of its neighbors, and node j is not recovered from the infection. Then, the third term represents the transition in which node i is infected and recovers, while node j is susceptible and is infected by any of its other neighbors. Last, in the fourth term, both nodes are infected, but node i recovers, while node j does not. The asymmetry of probability Φ_{ij} multiplies the number of equations by two, because, for each link between nodes i and j , we need an equation for $\Phi_{ij}(t+1)$ and another for $\Phi_{ji}(t+1)$.

Similarly, we can obtain an expression for probability Θ_{ij}^I

$$\begin{aligned} \Theta_{ij}^I(t+1) &= \Theta_{ij}^S(t) (1 - q_{ij}(t)) (1 - q_{ji}(t)) + \\ &\Phi_{ij}(t) (1 - (1 - \beta)q_{ij}(t)) (1 - \mu) + \\ &\Phi_{ji}(t) (1 - \mu) (1 - (1 - \beta)q_{ji}(t)) + \Theta_{ij}^I(t) (1 - \mu)^2 \end{aligned} \quad (4)$$

In this case, we have only L equations, one per link, because of its symmetry. There is no need for extra equations for probability Θ_{ij}^S , because the normalization leads to $\Theta_{ij}^S = 1 - \Phi_{ij} - \Phi_{ji} - \Theta_{ij}^I$.

$q_{ij}(t)$ in Eqs. 3 and 4 can be expressed as

$$q_{ij}(t) = \prod_{\substack{r=1 \\ r \neq j}}^N (1 - \beta A_{ri} h_{ir}) \quad (5)$$

where h_{ij} defines the hostility of j against i , that is, the probability that node j is infected when node i is susceptible, $h_{ij} = P(\sigma_j = I | \sigma_i = S)$. The hostility can be obtained in terms of Θ_{ij}^S and Φ_{ij} as

$$h_{ij} = \frac{\Phi_{ij}}{\Phi_{ij} + \Theta_{ij}^S} \quad (6)$$

Note that the denominator in Eq. 6 is a property of node i , given that $\Phi_{ij} + \Theta_{ij}^S = P(\sigma_i = S)$ for all neighboring nodes j of vertex i .

We call this system of $3L$ equations and unknowns our ELE model. It can be solved by iteration, starting from any meaningful initial condition, for example, $\Theta_{ij}^I(0) = \rho_0^2$ and $\Phi_{ij}(0) = \Theta_{ij}^I(0) = \rho_0(1 - \rho_0)$ (for any $0 < \rho_0 \leq 1$), until a fixed point is found. Apart from the solution where all nodes are susceptible, $\Theta_{ij}^S = 1$ for all the links, a nontrivial one appears when the system is above the critical value of the epidemic spreading (see Methods for the analytic derivation of the epidemic threshold from ELE model). Last, the incidence of the epidemic process, the average number of infected nodes in the whole system, can be computed as

$$\rho = \frac{1}{N} \sum_{i=1}^N \frac{1}{k_i} \sum_{j=1}^N A_{ji} (\Phi_{ji} + \Theta_{ji}^I) \quad (7)$$

where k_i is the degree of node i .

To test the agreement between our approach and empirical simulations, we have analyzed the incidence of the epidemics, ρ , in different synthetic and real network structures, covering the full range of infection probabilities, β (see Fig. 1). The results show a marked agreement between our ELE model and the Monte Carlo simulations and a good prediction of the epidemic threshold for all synthetic and real networks, pointing out the validity of our model to describe the global impact of the epidemics. Note that all networks, except the first one, have a large clustering coefficient, making the determination of the incidence difficult for standard mean field methods because of the effect of dynamical correlations.

Epidemic containment

Our approach for effective epidemic containment is based on removing the links with the largest link epidemic importance. This is possible once we have solved the ELE model, computing the I_{ij} for all the links in the network using Eq. 2, which can be expressed as

$$I_{ij} = \beta^2 \left(\Phi_{ji} \sum_{r=1}^N A_{jr} \frac{\Phi_{rj}}{\Phi_{rj} + \Theta_{rj}^I} + \Phi_{ij} \sum_{r=1}^N A_{ir} \frac{\Phi_{ri}}{\Phi_{ri} + \Theta_{ri}^I} \right) \quad (8)$$

Note that the value of β does not affect the ranking of the links, but we do not remove it from Eq. 8 to preserve the semantics of I_{ij} . Because the structure of the network changes after each link removal, it is convenient to recalculate the solution of the ELE model to ensure that we really remove the current link with the largest link epidemic importance.

We show the results of our approach for epidemic containment in Fig. 2. For comparison purposes, we also test four additional containment strategies. First, we consider two strategies that only make use of the structure of the network: removal based on maximum edge betweenness (22) and targeting the link with the highest eigenscore, that

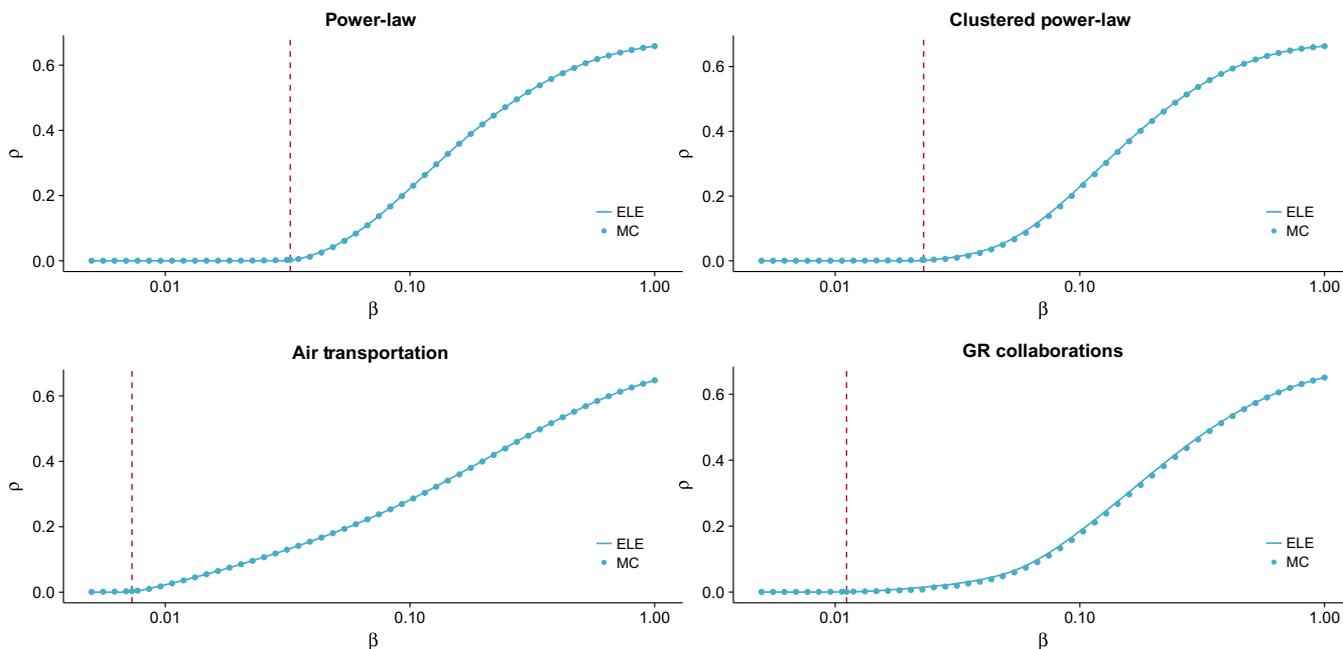


Fig. 1. Incidence of the epidemic process ρ as a function of the infection probability β . We show the incidence level for the ELE model (solid lines) and for Monte Carlo (MC) simulations (circles). The theoretical epidemic threshold calculated using Eq. 19 is marked with a vertical line. We have made use of two synthetic and two real networks: two scale-free networks (top) with an exponent of 3, one of them with high clustering coefficient; the world air transportation network; and the network of scientific collaborations in the field of general relativity (GR). We have set the recovery rate for all the networks to $\mu = 0.5$ (see Methods for the description of the networks and the details of the MC simulations).

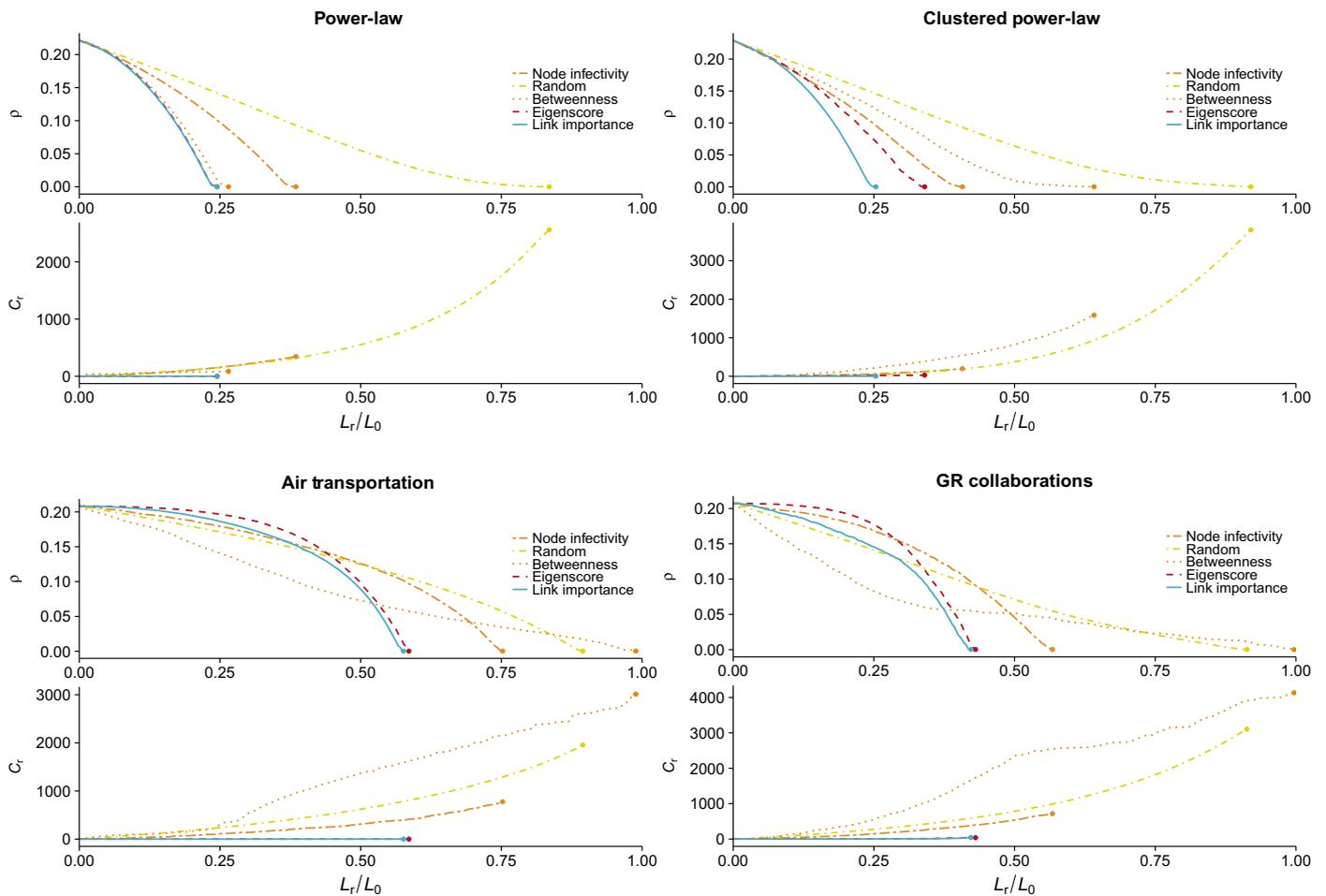


Fig. 2. Targeted bond percolation. We show the incidence of the epidemics, (ρ), and the number of connected components, (C_r), as functions of the occupation probability, (L_r/L_0), where L_0 is the number of links of the network and L_r is the number of removed edges in the bond percolation process. We compare five different epidemic containment strategies: removing the edges of the node with the highest probability of being infected, $P(\sigma_i = I)$ (orange dash-dash lines); a random edge removal (yellow dash-dot lines); removing the edge with the highest edge betweenness (light orange dotted lines); targeting the edge with the highest eigenscore (red dashed line); and, last, removing the edge that has the largest link epidemic importance (blue solid line). We apply these processes to the same networks as in Fig. 1 (see Methods). We have set the recovery rate to $\mu = 0.5$ and have chosen the infection probability β such that the stationary incidence of the epidemics is about $\rho_{ini} \approx 0.2$ for all the networks, that is, $\beta = 0.1$ for both power-law networks, $\beta = 0.06$ for air transportation network, and $\beta = 0.11$ for the collaboration network. The dots mark the achievement of total containment.

is, the product of the eigenvector centralities of the nodes connected by the link (27). Then, we consider a measure based on the epidemic process at the level of nodes, the removal of all the links of the node that has the maximum probability of being infected. Last, we carried out a simple random edge removal. As in the case of our strategy, we recalculate all the measures after each removal (see Methods for further details). We have also checked a promising approach based on communicability distances (42); however, the computational costs involved in computing communicability angles and distances for large networks preclude this approach to be used in large networks. For this reason, we have not included results on this one.

We observe in Fig. 2 that link epidemic importance leads to the fastest extinction of the epidemics for the four considered networks, and it is the only method that preserves their connectivity (thus, functionality). Note that the strategy based on node infectivity, though better than the random removal, performs poorly for all the networks despite having information about the epidemic process. This means that the use of information at the level of links is crucial to contain the epidemics.

For the power-law network, our approach using link epidemic importance yields the best performance, but the results are very similar to the those obtained using eigenscore and edge betweenness [equivalent results hold for Erdős-Rényi (ER) networks; see fig. S6]. However, when the transitivity of the network is increased, we can see the benefits of using link epidemic importance, both in epidemic containment and on preservation of the connectivity of the network (see figs. S2 and S3 for more details on the containment process for each method).

The effect of the clustering coefficient is also present when we look at the epidemic containment results for the two empirical networks in Fig. 2. Moreover, as in most real networks, the air transportation and the scientific collaboration networks have a significant modular structure. This plays an important role on the epidemic containment process. Here, we can see how the strategy based on edge betweenness apparently performs better when few links are removed because of the fact that links with higher edge betweenness are those connecting different modules (43). When the bond percolation process isolates modules, each module may sustain its own epidemic process, and thus, it may happen that some

of the modules are subcritical for the given infection probability β . This will lead to a decrease of the global prevalence of the epidemics at the expense of losing the connectivity of the network. Furthermore, if we look at the prevalence on the giant connected component, an important increase above the initial average number of infected individuals is revealed (see figs. S4 and S5). A consequence of this fragmentation process is the appearance of multiple isolated supercritical components, for which the removal of a link in one of them does not affect the incidence on the other components. As a result, the edge betweenness procedure needs to remove more links to arrive to the total epidemic extinction than any of

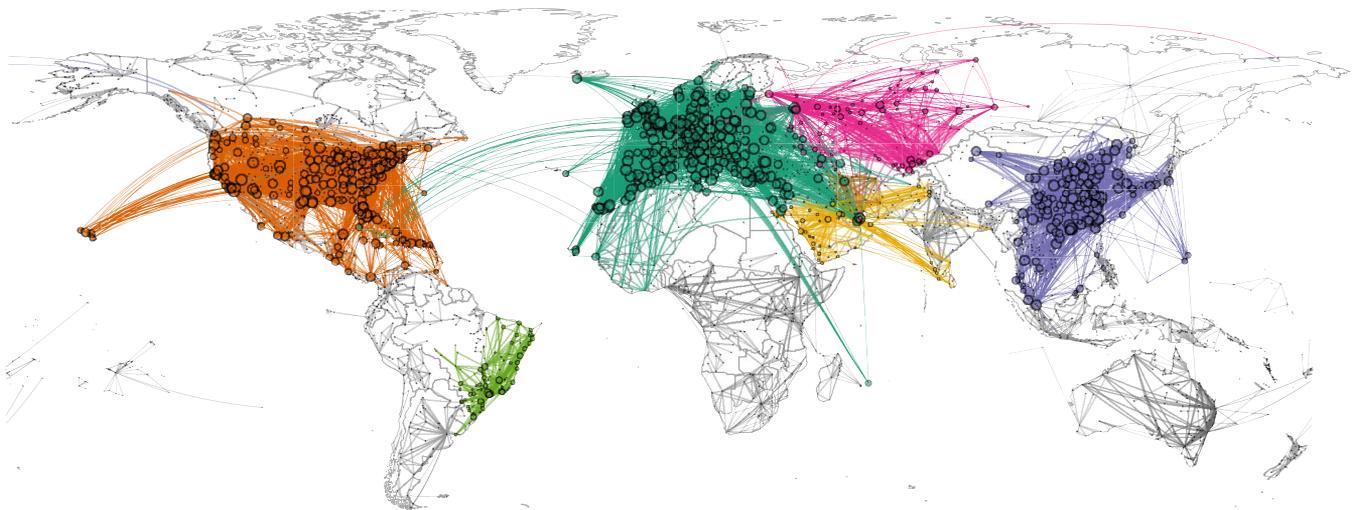
the other methods, even the random one. For the sake of completeness, we have analyzed two benchmark networks with community structure, obtaining similar results (see figs. S7 and S8).

In Fig. 3 (top), we illustrate the survival links in the air transportation network after 33.3% of the edges have been removed according to our epidemic containment strategy proposal (see fig. S9 for the original network before the containment process). As it is observed, the global connectivity, and thus functionality, of the worldwide connections is preserved (links of the same color are part of the same connected component). In Fig. 3 (bottom), we plot the network after deactivating

Link epidemic importance



Edge betweenness



$P(\sigma_i = l) \cdot 0.0 \circ 0.2 \circ 0.4 \circ 0.6$

Fig. 3. Epidemic containment on the air transportation network. We show the networks after 33.3% of the links have been removed using link epidemic importance (**top**) and edge betweenness (**bottom**). Nodes and edges with the same color belong to the same connected component, with subcritical components in gray scale and using darker gray for larger components. The area of the nodes is proportional to their probability of being infected $P(\sigma_i = l)$ from 0.0 to 0.6. We have set the epidemic probabilities to $\mu = 0.5$ and $\beta = 0.06$.

the same fraction of links (33.3%) using the recursive deactivation of links according to edge betweenness. The edge betweenness containment method, in contrast with our proposal, generates two kinds of components: small or sparsely connected subcritical modules such as the ones in Australia, Africa, or South America, where the epidemics vanishes, and large supercritical communities in Europe, North America, and East Asia, with a large prevalence of the epidemics. This means that, for instance, there is no path to go from London to New York, or from Tokyo to Los Angeles, thus disconnecting the world by air transportation.

For a better assessment of the performance of the different containment strategies, we show in Fig. 4 their comparison in terms of the required fraction of removed links to attain total containment, L_{TC}/L_0 , when applied to a large set of synthetic networks and epidemic parameters. The results point to a clear advantage of the link epidemic importance method over the node infectivity and random approaches, and better or equal results with respect to edge betweenness and eigenscore. Only eigenscore achieves results comparable to link epidemic importance, with a slight advantage for our method. When applied to a set

of 27 real networks (see Methods), the differences between link epidemic importance and eigenscore become more evident, showing again the effectiveness of our approach, but with some exceptions in which eigenscore performs better (see Fig. 5). In addition, we plot in figs. S10 and S11 the comparison of the number of connected components between the different containment strategies, which show the better performance of link epidemic importance to keep the number of components low, with only a few exceptions.

DISCUSSION

We have presented a methodology for assessing epidemic spreading based on links instead of nodes. The model, named ELE, allows the determination of the epidemic importance of each link in transmitting the disease. The method accounts for the first-order correlations between links, although it could be extended to higher orders, assuming a larger analytical and computational cost. The results are used to develop an epidemic containment strategy built on deactivating recursively

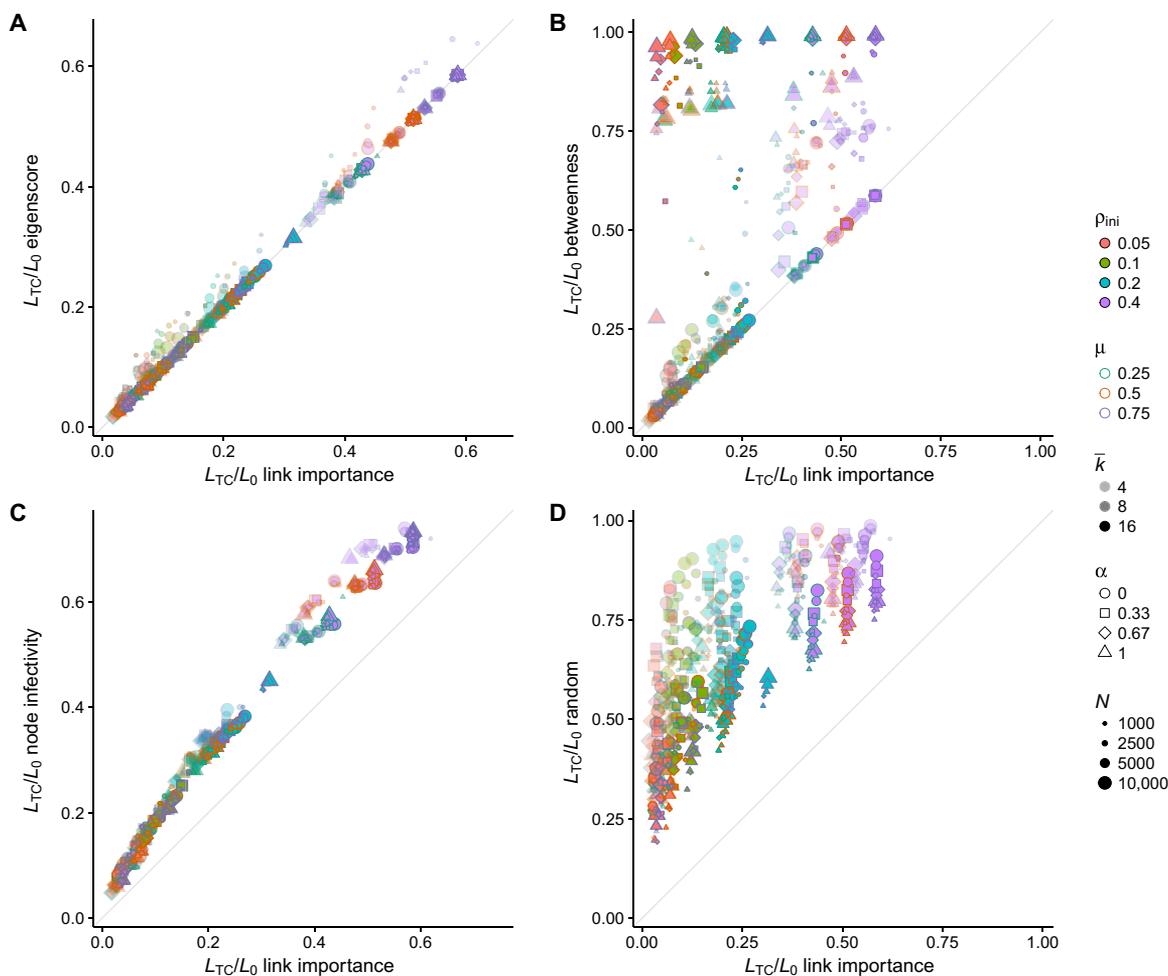


Fig. 4. Fraction of links removed for total epidemic containment on synthetic networks. We show the fraction of links that have to be removed to obtain total epidemic containment using link epidemic importance, compared with the fractions for the other four strategies: (A) eigenscore, (B) edge betweenness, (C) node infectivity, and (D) random removal. Each point represents a configuration consisting of a network and a set of epidemic parameters. The networks have been generated with the model in (48), which interpolates between ER ($\alpha = 1$) and BA networks ($\alpha = 0$). We use four values of the interpolating parameter: $\alpha = 0.0, 0.33, 0.67, \text{ and } 1.0$. These networks are generated in four sizes ($N = 1000, 2500, 5000, \text{ and } 10,000$) and three average degrees ($\langle k \rangle = 4, 8, \text{ and } 16$), thus amounting to 48 different networks. For each network, we apply the five containment strategies for three different values of the recovery probability ($\mu = 0.25, 0.50, \text{ and } 0.75$) and four values of the infection probability β selected such that, before removing links, the incidence of the epidemics at the stationary state is equal to $\rho_{ini} = 0.05, 0.1, 0.2, \text{ and } 0.4$. Therefore, each plot contains 576 different configurations.

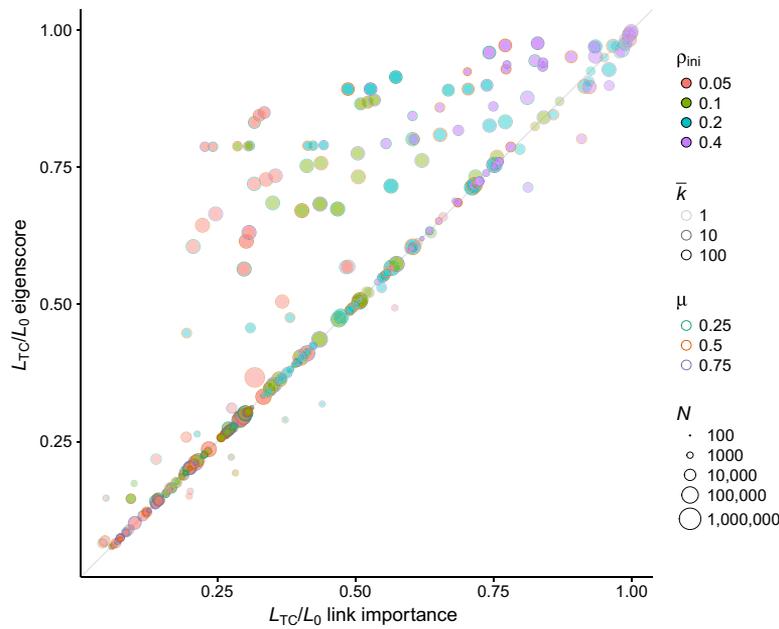


Fig. 5. Fraction of links removed for total epidemic containment on real networks. We compare the link epidemic importance and eigenscore methods on a set of 27 real networks selected from the Network Repository (49) (see Methods), with sizes ranging from 410 to 404,719 nodes. The epidemic parameters are the same as in Fig. 4, thus amounting to 324 different configurations.

the links with the largest link epidemic importance while preserving the connectivity of the full network, that is, avoiding fragmentation. We have validated our proposal in synthetic and empirical networks, comparing with other alternative containment strategies, which show its better performance, with few exceptions. In the empirical case of the worldwide air transportation network, we identify the most important connections between airports for the spreading of epidemics and evaluate the epidemic incidence after its deactivation, considering an SIS epidemic spreading dynamics. Our results open the door to new approaches in the analysis of dynamical diffusive-like models on complex networks at the level of links instead of nodes.

METHODS

Epidemic threshold

The determination of the epidemic threshold was performed by considering a state of the system in which the epidemic incidence is very small ($\Phi_{ij}, \Phi_{ji}, \Theta_{ij}^I \ll 1$, for all links); thus, the system of equations can be linearized (see section S2 for full details), resulting in

$$\Theta_{ij}^I = \beta(1 - \mu)\Phi_{ij} + \beta(1 - \mu)\Phi_{ji} + (1 - \mu)^2\Theta_{ij}^I \tag{9}$$

$$\Phi_{ij} = \beta \sum_r (A_{jr} - (1 - \mu)\delta_{ri})\Phi_{jr} + (1 - \beta)(1 - \mu)\Phi_{ij} + \mu(1 - \mu)\Theta_{ij}^I \tag{10}$$

Here, we removed the dependence on time to emphasize that we are considering the steady state. From Eq. 9, we can write

$$\Theta_{ij}^I = \frac{\beta(1 - \mu)}{\mu(2 - \mu)} (\Phi_{ij} + \Phi_{ji}) \tag{11}$$

Now, calling $\epsilon_i = \Phi_{ji} + \Theta_{ij}^I \ll 1$, which does not depend on node j because $P(\sigma_i = I, \sigma_j = S) + P(\sigma_i = I, \sigma_j = I) = P(\sigma_i = I)$, we made the following ansatz

$$\Theta_{ij}^I = \Upsilon(\epsilon_i + \epsilon_j) \tag{12}$$

$$\Phi_{ij} = X\epsilon_i + Z\epsilon_j \tag{13}$$

where Υ, X , and Z are constants independent of the link. These ansatz include the assumptions of symmetry of Θ_{ij}^I and asymmetry of Φ_{ij} , respectively. We can determine the constants by substitution in Eq. 11 and using the definition of ϵ_i , which leads to

$$\Upsilon = \frac{\beta(1 - \mu)}{\mu(2 - \mu) + 2\beta(1 - \mu)} \tag{14}$$

$$X = -\Upsilon \tag{15}$$

$$Z = 1 - \Upsilon \tag{16}$$

Last, we built equations for ϵ_i by substituting Eqs. 9 and 10 in $\epsilon_i = \Phi_{ji} + \Theta_{ij}^I$ and using the ansatz. The result is

$$\frac{\mu}{\beta} \epsilon_i = \sum_j B_{ji} \epsilon_j \tag{17}$$

where B is a matrix whose elements depend on the adjacency matrix of the network, on Υ , and on the degrees k_i of the nodes

$$B_{ij} = (1 - \Upsilon)A_{ij} - \Upsilon k_i \delta_{ij} \tag{18}$$

δ_{ij} stands for the Kronecker delta function, which is 1 if $i = j$ and 0 otherwise. If μ/β is an eigenvalue of matrix B , then Eq. 17 has nontrivial solutions. Hence, the onset of the epidemics β_c , the lowest value of β that yields nontrivial solutions of Eq. 17, is given by

$$\beta_c = \frac{\mu}{\Lambda_{\max}(B)} \quad (19)$$

where $\Lambda_{\max}(B)$ is the largest eigenvalue of matrix B . Note that matrix B depends on β and μ ; thus, Eq. 19 is implicit for β_c , which can be solved by iteration (see section S3 for a discussion on the determination of the epidemic threshold).

Estimation of the incidence of the epidemic from numerical simulations

The numerical incidence of the epidemics, ρ , is calculated using discrete-time and synchronous Monte Carlo simulations. We made use of the quasistationary (QS) approach (44, 45) to avoid the effect that a large number of realizations end up in the absorbing state with no infected individuals in the system. Basically, the QS method focuses the simulation on active configurations, that is, with one or more infected individuals. Every time the system reaches the absorbing state, this state is replaced by one of the previously stored active states of the system. We kept 50 active configurations with an update probability of 0.20. We gave the systems a transient time of 10^5 time steps and then calculated ρ as an average over a relaxation time of 2×10^4 time steps.

Networks

In this work, we evaluated our methodology on synthetic and empirical networks. We built a network with power-law degree distribution $P(k) \sim k^{-\gamma}$ with exponent $\gamma = 3$ and $\langle k \rangle = 6$ using the configuration model. To evaluate the impact of transitivity, we also built another network with the same characteristics of the previous one but with a clustering coefficient of 0.6 using the algorithm of Holme and Kim (46) with a parameter $p = 0.8$.

We considered also two empirical networks: the air transportation network and the network of scientific collaborations in the field of general relativity. The air transportation network was constructed using data from the website openflights.org, which has information about the traffic between airports updated to 2012. This network accounts for the largest connected component, with 3154 nodes and 18,592 edges (see data file S1). The network of scientific collaborations was obtained from (47); it is composed of 5242 nodes linked by 14,496 edges.

The synthetic networks in Fig. 4 were generated with the model in (48), which interpolates between ER and Barabási-Albert (BA) networks. In this way, we were able to evaluate the performance of the containment strategies on networks with degree distributions that range from Poisson (ER) to power-law (BA). By construction, these networks have no community structure and low transitivity.

The 27 real networks in Fig. 5 were obtained from the Network Repository (49), selecting only the largest connected component. They cover wide ranges of number of nodes (from 410 to 404,719), number of links (from 1043 to 713,319), average degree (between 2.04 and 84.82), average clustering coefficient (from 0.0023 to 0.1105), and assortativity (between -0.88 and 0.64) (see section S4).

Containment process

To perform the deactivation of links, we imposed an adiabatic process: After each removal step, we let the system converge to the meta-stable

equilibrium before removing any other link. For a fair comparison between different containment strategies, we removed on each deactivation step as many edges as we have removed using the node infectivity strategy. In the case of the real networks in Fig. 5, we set the maximum number of adiabatic processes to 1000 because of its large computational cost on the largest networks. This means that, if the network has 20,000 links, we removed 20 links at each deactivation step. We considered that we reached total containment when the incidence of the epidemics becomes lower than $1/N$. The computational cost of calculating the links to remove for each containment strategy is shown in fig. S13.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/12/eau4212/DC1>

Section S1. Link epidemic importance and connected components

Section S2. Linearization of the ELE model

Section S3. Epidemic threshold

Section S4. Data description

Fig. S1. Ratio between the link epidemic importance I^A of a link in a subnetwork A and the link epidemic importance I^{AB} of a link that acts as the only bridge between subnetworks A and B .

Fig. S2. Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, and average degree $\langle k \rangle = 6$.

Fig. S3. Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, high clustering coefficient, and average degree $\langle k \rangle = 6$.

Fig. S4. Epidemic containment for the air transportation network.

Fig. S5. Epidemic containment for the general relativity collaborations network.

Fig. S6. Epidemic containment for an ER network with 5000 nodes and average degree $\langle k \rangle = 6$.

Fig. S7. Epidemic containment for a network with 5000 nodes generated with a stochastic block model, with four blocks of 250 nodes, two blocks of 1000 nodes, and one block of 2000 nodes, average degree of 5, and mixing probability of 0.3.

Fig. S8. Epidemic containment for a network with 5000 nodes generated using the LFR algorithm, with average degree of 6, exponent of 3, and mixing probability of 0.1.

Fig. S9. Original air transportation network (top) and the results after a removal of 33.3% of the links using link epidemic importance (middle) and edge betweenness (bottom).

Fig. S10. Comparison of the number of connected components after total containment between the link epidemic importance strategy and the other four methods, calculated for the synthetic networks and parameters as in Fig. 4.

Fig. S11. Comparison of the number of connected components after total containment between the link epidemic importance and eigenscore strategies, calculated for the real networks and parameters as in Fig. 5.

Fig. S12. Graphical representation of the determination of the epidemic threshold.

Fig. S13. Computational time invested for each method to perform a single ranking and removal for BA networks ranging from 100 to 400,000 nodes, averaged over 36 repetitions.

Table S1. Structural characteristics of the 27 real networks obtained from the Network Repository (<http://networkrepository.com>) and used in Fig. 6 and fig. S11.

Data file S1. Air transportation network data.

REFERENCES AND NOTES

1. R. M. Anderson, R. M. May, B. Anderson, *Infectious Diseases of Humans: Dynamics and Control* (Wiley Online Library, 1992), vol. 28.
2. H. W. Hethcote, The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).
3. D. J. Daley, J. Gani, J. M. Gani, *Epidemic Modelling: An Introduction* (Cambridge Univ. Press, 2001), vol. 15.
4. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
5. R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
6. M. E. Newman, Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).
7. L. Hufnagel, D. Brockmann, T. Geisel, Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15124–15129 (2004).
8. Y. Wang, D. Chakrabarti, C. Wang, C. Faloutsos, *Proceedings of the 22nd International Symposium on Reliable Distributed Systems* (IEEE, 2003), pp. 25–34.
9. D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, C. Faloutsos, Epidemic thresholds in real networks. *ACM Trans. Inform. Syst. Security* **10**, 13 (2008).
10. S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, Y. Moreno, Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *Europhys. Lett.* **89**, 38009 (2010).

11. S. Gómez, J. Gómez-Gardenes, Y. Moreno, A. Arenas, Nonperturbative heterogeneous mean-field approach to epidemic spreading in complex networks. *Phys. Rev. E* **84**, 036105 (2011).
12. D. Brockmann, D. Helbing, The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
13. C.-R. Cai, Z.-X. Wu, J.-Y. Guan, Effective degree Markov-chain approach for discrete-time epidemic processes on uncorrelated networks. *Phys. Rev. E* **90**, 052803 (2014).
14. C.-R. Cai, Z.-X. Wu, M. Z. Q. Chen, P. Holme, J.-Y. Guan, Solving the dynamic correlation problem of the susceptible-infected-susceptible model on networks. *Phys. Rev. Lett.* **116**, 258301 (2016).
15. P. E. Paré, B. E. Kirwan, J. Liu, T. Başar, C. L. Beck, 2018 Annual American Control Conference (ACC) (IEEE, 2018), pp. 404–409.
16. D. J. D. Earn, P. Rohani, B. M. Bolker, B. T. Grenfell, A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670 (2000).
17. R. Pastor-Satorras, A. Vespignani, Immunization of complex networks. *Phys. Rev. E* **65**, 036104 (2002).
18. N. Madar, T. Kalisky, R. Cohen, D. ben-Avraham, S. Havlin, Immunization and epidemic dynamics in complex networks. *Eur. Phys. J. B* **38**, 269–276 (2004).
19. J. Gómez-Gardeñes, P. Echenique, Y. Moreno, Immunization of real complex communication networks. *Eur. Phys. J. B* **49**, 259–264 (2006).
20. L. Hébert-Dufresne, A. Allard, J.-G. Young, L. J. Dubé, Global efficiency of local immunization on complex networks. *Sci. Rep.* **3**, 2171 (2013).
21. E. T. Lofgren, M. E. Halloran, C. M. Rivers, J. M. Drake, T. C. Porco, B. Lewis, W. Yang, A. Vespignani, J. Shaman, J. N. S. Eisenberg, M. C. Eisenberg, M. Marathe, S. V. Scarpino, K. A. Alexander, R. Meza, M. J. Ferrari, J. M. Hyman, L. A. Meyers, S. Eubank, Opinion: Mathematical models: A key tool for outbreak response. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18095–18096 (2014).
22. P. Holme, B. J. Kim, C. N. Yoon, S. K. Han, Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056109 (2002).
23. R. Cohen, S. Havlin, D. ben-Avraham, Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91**, 247901 (2003).
24. F. Chung, P. Horn, A. Tsiatas, Distributing antidote using PageRank vectors. *Internet Math.* **6**, 237–254 (2009).
25. B. A. Prakash, L. A. Adamic, T. Iwashyna, H. Tong, C. Faloutsos, *Proceedings of the 2013 SIAM International Conference on Data Mining* (SIAM, 2013), pp. 659–667.
26. C. M. Schneider, T. Mihaljev, S. Havlin, H. J. Herrmann, Suppressing epidemics with a limited amount of immunization units. *Phys. Rev. E* **84**, 061911 (2011).
27. P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. van de Bovenkamp, D. Liu, H. Wang, Decreasing the spectral radius of a graph by link removals. *Phys. Rev. E* **84**, 016101 (2011).
28. D. H. Zanette, S. Risau-Gusmán, Infection spreading in a population with evolving contacts. *J. Biol. Phys.* **34**, 135–148 (2008).
29. V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, G. J. Pappas, Optimal resource allocation for network protection against spreading processes. *IEEE Trans. Control Netw. Syst.* **1**, 99–108 (2014).
30. M. Zargham, V. Preciado, 2014 48th Annual Conference on Information Sciences and Systems (CISS) (IEEE, 2014), pp. 1–6.
31. C. Nowzari, V. M. Preciado, G. J. Pappas, Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Syst. Mag.* **36**, 26–46 (2016).
32. A. N. Bishop, I. Shames, Link operations for slowing the spread of disease in complex networks. *Europhys. Lett.* **95**, 18005 (2011).
33. S. Saha, A. Adiga, B. A. Prakash, A. K. S. Vullikanti, *Proceedings of the 2015 SIAM International Conference on Data Mining* (Society for Industrial and Applied Mathematics, 2015), pp. 568–576.
34. J. G. Restrepo, E. Ott, B. R. Hunt, Characterizing the dynamical importance of network nodes and links. *Phys. Rev. Lett.* **97**, 094102 (2006).
35. P. Van Mieghem, J. Omic, R. Kooij, Virus spread in networks. *IEEE ACM Trans. Netw.* **17**, 1–14 (2009).
36. M. Taylor, P. L. Simon, D. M. Green, T. House, I. Z. Kiss, From Markovian to pairwise epidemic models and the performance of moment closure approximations. *J. Math. Biol.* **64**, 1021–1042 (2012).
37. E. A. Cator, P. Van Mieghem, Second-order mean-field susceptible-infected-susceptible epidemic threshold. *Phys. Rev. E* **85**, 056111 (2012).
38. I. Z. Kiss, L. Berthouze, T. J. Taylor, P. L. Simon, Modelling approaches for simple dynamic networks and applications to disease transmission models. *Proc. R. Soc. A* **468**, 1332–1355 (2012).
39. J. P. Gleeson, Binary-state dynamics on complex networks: Pair approximation and beyond. *Phys. Rev. X* **3**, 021004 (2013).
40. A. S. Mata, S. C. Ferreira, Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *Europhys. Lett.* **103**, 48003 (2013).
41. A. Szabó-Solticzky, L. Berthouze, I. Z. Kiss, P. L. Simon, Oscillating epidemics in a dynamic network model: Stochastic and mean-field analysis. *J. Math. Biol.* **72**, 1153–1176 (2016).
42. E. Estrada, E. Vargas-Estrada, H. Ando, Communicability angles reveal critical edges for network consensus dynamics. *Phys. Rev. E* **92**, 052809 (2015).
43. M. E. Newman, M. Girvan, Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
44. S. C. Ferreira, C. Castellano, R. Pastor-Satorras, Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Phys. Rev. E* **86**, 041125 (2012).
45. K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt, S. McKay, Monte Carlo simulation in statistical physics. *Comput. Phys.* **7**, 156 (1993).
46. P. Holme, B. J. Kim, Growing scale-free networks with tunable clustering. *Phys. Rev. E* **65**, 026107 (2002).
47. J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data* **1**, 2 (2007).
48. J. Gómez-Gardeñes, Y. Moreno, From scale-free to Erdős-Rényi networks. *Phys. Rev. E* **73**, 056124 (2006).
49. R. A. Rossi, N. K. Ahmed, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (AAAI, 2015), vol. 15, pp. 4292–4293.

Acknowledgments: We acknowledge comments on the manuscript by C. Granell, M. Barahona, and P. Hövel. **Funding:** J.T.M., A.A., and S.G. were supported by the Generalitat de Catalunya project 2017-SGR-896, Spanish MINECO project FIS2015-71582-C2-1, and Universitat Rovira i Virgili project 2017PFR-URV-B2-41. A.A. acknowledges financial support from the ICREA Academia and the James S. McDonnell Foundation (220020325). J.T.M. acknowledges an FI-2015 grant from the Generalitat de Catalunya AGAUR. **Author contributions:** All authors have contributed equally to this manuscript. **Competing interests:** All authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 8 June 2018

Accepted 7 November 2018

Published 5 December 2018

10.1126/sciadv.aau4212

Citation: J. T. Matamalas, A. Arenas, S. Gómez, Effective approach to epidemic containment using link equations in complex networks. *Sci. Adv.* **4**, eaau4212 (2018).

Supplementary Materials for

Effective approach to epidemic containment using link equations in complex networks

Joan T. Matamalas*, Alex Arenas*, Sergio Gómez*

*Corresponding author. Email: joantomas.matamalas@urv.cat (J.T.M.); alexandre.arenas@urv.cat (A.A.); sergio.gomez@urv.cat (S.G.)

Published 5 December 2018, *Sci. Adv.* **4**, eaau4212 (2018)
DOI: 10.1126/sciadv.aau4212

The PDF file includes:

Section S1. Link epidemic importance and connected components

Section S2. Linearization of the ELE model

Section S3. Epidemic threshold

Section S4. Data description

Fig. S1. Ratio between the link epidemic importance I^A of a link in a subnetwork A and the link epidemic importance I^{AB} of a link that acts as the only bridge between subnetworks A and B .

Fig. S2. Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, and average degree $\langle k \rangle = 6$.

Fig. S3. Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, high clustering coefficient, and average degree $\langle k \rangle = 6$.

Fig. S4. Epidemic containment for the air transportation network.

Fig. S5. Epidemic containment for the general relativity collaborations network.

Fig. S6. Epidemic containment for an ER network with 5000 nodes and average degree $\langle k \rangle = 6$.

Fig. S7. Epidemic containment for a network with 5000 nodes generated with a stochastic block model, with four blocks of 250 nodes, two blocks of 1000 nodes, and one block of 2000 nodes, average degree of 5, and mixing probability of 0.3.

Fig. S8. Epidemic containment for a network with 5000 nodes generated using the LFR algorithm, with average degree of 6, exponent of 3, and mixing probability of 0.1.

Fig. S9. Original air transportation network (top) and the results after a removal of 33.3% of the links using link epidemic importance (middle) and edge betweenness (bottom).

Fig. S10. Comparison of the number of connected components after total containment between the link epidemic importance strategy and the other four methods, calculated for the synthetic networks and parameters as in Fig. 4.

Fig. S11. Comparison of the number of connected components after total containment between the link epidemic importance and eigenscore strategies, calculated for the real networks and parameters as in Fig. 5.

Fig. S12. Graphical representation of the determination of the epidemic threshold.

Fig. S13. Computational time invested for each method to perform a single ranking and removal for BA networks ranging from 100 to 400,000 nodes, averaged over 36 repetitions.

Table S1. Structural characteristics of the 27 real networks obtained from the Network Repository (<http://networkrepository.com>) and used in Fig. 6 and fig. S11.

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/4/12/eaau4212/DC1)

Data file S1 (.graphml format). Air transportation network data.

Section S1. Link epidemic importance and connected components

One important property of our proposed definition of *link epidemic importance* is that it tends to maintain the connectivity of the network when the selected link is removed. Here we show that, when the network is formed by two subnetworks A and B , which are connected by just one link (i_A, j_B) , the link epidemic importance of that link is lower than that of another link internal to A or B . This means our containment strategy of removing the edge with largest link epidemic importance will not break the network in two disconnected components A and B , unlike the betweenness approach, for which the link (i_A, j_B) plays the role of a bridge and thus it has maximal edge betweenness.

Let $\langle k \rangle_A$ and $\langle k \rangle_B$ be the average degrees of subnetworks A and B , respectively, with $\langle k \rangle_A > \langle k \rangle_B$. Let us also call ρ^A and ρ^B their respective incidence of the epidemics. The link epidemic importance of a link between nodes i and j has been defined as

$$I_{ij} = \bar{n}_{ij} + \bar{n}_{ji}, \quad (\text{S.1})$$

where

$$\bar{n}_{ij} = \beta P(\sigma_j = S, \sigma_i = I) \sum_{r=1}^N A_{jr} \beta P(\sigma_r = S | \sigma_j = I). \quad (\text{S.2})$$

Supposing independence of the states of the nodes, \bar{n}_{ij} can be approximated as

$$\bar{n}_{ij} \approx \beta P(\sigma_j = S) P(\sigma_i = I) \sum_{r=1}^N A_{jr} \beta P(\sigma_r = S). \quad (\text{S.3})$$

In a homogeneous mean field approximation we may substitute $k_j = \langle k \rangle$, $P(\sigma_i = I) \approx \rho$, and $P(\sigma_j = S) \approx 1 - \rho$, which lead to the following expressions for the importance of a link:

$$I^A \approx 2\beta^2 \rho^A (1 - \rho^A)^2 \langle k \rangle_A, \quad (\text{S.4})$$

$$I^B \approx 2\beta^2 \rho^B (1 - \rho^B)^2 \langle k \rangle_B, \quad (\text{S.5})$$

$$I^{AB} \approx \beta^2 [\rho^A (1 - \rho^B)^2 \langle k \rangle_B + \rho^B (1 - \rho^A)^2 \langle k \rangle_A]. \quad (\text{S.6})$$

Here, I^A and I^B denote the link epidemic importance of links inside A and B , respectively, and I^{AB} the link epidemic importance of the link connecting subnetworks A and B .

We need an expression relating the average degree $\langle k \rangle$ and the incidence of the epidemics ρ to be able to calculate the approximate values of the link epidemic importances in Eqs. (S.4) to (S.6). It can be obtained using the nonperturbative heterogeneous mean field (npHMF) equations in [Gómez et al., Phys. Rev. E, 84 (2011) 036105]. In particular, the npHMF equations for the SIS model without one-step reinfections (WOR) read as

$$0 = -\mu\rho_k + (1 - \rho_k)(1 - q_k), \quad (\text{S.7})$$

$$q_k = \prod_{k'} (1 - \beta\rho_{k'})^{C_{kk'}}, \quad (\text{S.8})$$

where ρ_k represents the fraction of infected nodes of degree k , q_k the probability that nodes of degree k are not infected by nodes of any other degree k' , and $C_{kk'} = kP(k'|k)$ the expected number of links from a node of degree k to nodes of degree k' . In the nonperturbative homogeneous mean field (npHoMF) approximation, this reduces to

$$0 = -\mu\rho + (1 - \rho)(1 - q), \quad (\text{S.9})$$

$$q = (1 - \beta\rho)^{\langle k \rangle}. \quad (\text{S.10})$$

Thus, after some algebra we get

$$\langle k \rangle = \frac{\log\left(1 - \mu\frac{\rho}{1 - \rho}\right)}{\log(1 - \beta\rho)}. \quad (\text{S.11})$$

An immediate consequence of Eq. (S.11) is that $0 \leq \rho \leq 1/(1 + \mu)$. We can see a plot of the npHoMF relationship between ρ and $\langle k \rangle$ in the inset of Fig. S1. Note that ρ is an increasing function of $\langle k \rangle$, thus the larger the average degree, the greater the incidence of the epidemics.

Now, we can substitute Eq. (S.11) for $\langle k \rangle_A$ and $\langle k \rangle_B$ into Eqs. (S.4) to (S.6) to obtain approximations of the three different link epidemic importances. The results are presented in Fig. S1. Fixing a certain average degree $\langle k \rangle_B$ for subnetwork B , we consider subnetworks A with $\langle k \rangle_A \geq \langle k \rangle_B$. Since ρ increases with $\langle k \rangle$, this is equivalent to fixing ρ^B and consider subnetworks A with $\rho^A \geq \rho^B$. We observe that, in all cases, the link epidemic importance I^A of links in subnetwork A is larger than I^{AB} of the bridge link between subnetworks A and B ($I^A \geq I^{AB}$), thus confirming that our epidemic containment strategy driven by link epidemic importance does not disconnect the network. This result has been obtained under independence and homogeneous mean field approximations, and for a specific structural configuration of the network. In practice, we observe that the epidemic containment approach based on link epidemic importance is the one considered which better preserves the connectivity of the networks, as shown in Figs. S10 and S11.

Section S2. Linearization of the ELE model

The determination of the epidemic threshold from the Epidemic Link Equations (ELE) requires the consideration of states in which the probabilities of having infected nodes are very small, i.e. $\Phi_{ij}, \Phi_{ji}, \Theta_{ij}^I \ll 1$. Therefore, we may suppose that $\Phi_{ij}, \Phi_{ji}, \Theta_{ij}^I \sim O(\varepsilon)$, with $\varepsilon \ll 1$, and in consequence $\Theta_{ij}^S \sim 1 - O(\varepsilon)$. Using these approximations the epidemic link equations become linear in these $O(\varepsilon)$ probabilities, since $O(\varepsilon^2)$ terms should be neglected.

We start with the linearization of the hostility:

$$\begin{aligned}
 h_{ij} &= \frac{\Phi_{ij}}{\Phi_{ij} + \Theta_{ij}^S} \\
 &= \frac{\Phi_{ij}}{1 - (\Phi_{ji} + \Theta_{ij}^I)} \\
 &= \Phi_{ij} (1 + \Phi_{ji} + \Theta_{ij}^I + O(\varepsilon^2)) \\
 &= \Phi_{ij} + O(\varepsilon^2),
 \end{aligned} \tag{S.12}$$

where we have used the normalization $\Phi_{ij} + \Phi_{ji} + \Theta_{ij}^I + \Theta_{ij}^S = 1$, and we realize that terms $\Phi_{ij}\Phi_{ji}$ and $\Phi_{ij}\Theta_{ij}^I$ are both $O(\varepsilon^2)$. Substituting hostility in the expression for q_{ij} we get:

$$\begin{aligned}
 q_{ij} &= \prod_{\substack{r=1 \\ r \neq j}}^N (1 - \beta A_{ri} h_{ir}) \\
 &= \prod_{\substack{r=1 \\ r \neq j}}^N (1 - \beta A_{ri} \Phi_{ir} + O(\varepsilon^2)) \\
 &= 1 - \beta \sum_{\substack{r=1 \\ r \neq j}}^N A_{ri} \Phi_{ir} + O(\varepsilon^2) \\
 &= 1 - \beta \sum_{r=1}^N A_{ri} (1 - \delta_{rj}) \Phi_{ir} + O(\varepsilon^2),
 \end{aligned} \tag{S.13}$$

where the Kronecker δ_{rj} has been introduced to make zero the j th term of the sum. Now we are in condition to find the linear approximations of the main ELE model equations. First, the equation for Θ_{ij}^I becomes:

$$\begin{aligned}
 \Theta_{ij}^I &= \Theta_{ij}^S (1 - q_{ij}) (1 - q_{ji}) \\
 &\quad + \Phi_{ij} (1 - (1 - \beta)q_{ij}) (1 - \mu) \\
 &\quad + \Phi_{ji} (1 - \mu) (1 - (1 - \beta)q_{ji}) \\
 &\quad + \Theta_{ij}^I (1 - \mu)^2 \\
 &= O(\varepsilon^2) \\
 &\quad + \Phi_{ij} \beta (1 - \mu) + O(\varepsilon^2) \\
 &\quad + \Phi_{ji} (1 - \mu) \beta + O(\varepsilon^2) \\
 &\quad + \Theta_{ij}^I (1 - \mu)^2 \\
 &= \beta (1 - \mu) \Phi_{ij} + \beta (1 - \mu) \Phi_{ji} + (1 - \mu)^2 \Theta_{ij}^I + O(\varepsilon^2).
 \end{aligned} \tag{S.14}$$

Note that, for the terms with a factor Φ_{ij}, Φ_{ji} or Θ_{ij}^I , which are $O(\varepsilon)$, we just need to keep $O(1)$ contributions in the rest of the term, thus we may use the approximations $q_{ij} = q_{ji} = 1 + O(\varepsilon)$.

The equation for Φ_{ij} reads:

$$\begin{aligned}
\Phi_{ij} &= \Theta_{ij}^S q_{ij}(1 - q_{ji}) \\
&\quad + \Phi_{ij} ((1 - \beta)q_{ij}) (1 - \mu) \\
&\quad + \Phi_{ji} \mu (1 - (1 - \beta)q_{ji}) \\
&\quad + \Theta_{ij}^I \mu(1 - \mu) \\
&= \beta \sum_{r=1}^N A_{rj}(1 - \delta_{ri}) \Phi_{jr} + O(\varepsilon^2) \\
&\quad + \Phi_{ij} (1 - \beta)(1 - \mu) + O(\varepsilon^2) \\
&\quad + \Phi_{ji} \mu\beta + O(\varepsilon^2) \\
&\quad + \Theta_{ij}^I \mu(1 - \mu) \\
&= \beta \sum_{r=1}^N (A_{rj}(1 - \delta_{ri}) + \mu\delta_{ri}) \Phi_{jr} \\
&\quad + \Phi_{ij} (1 - \beta)(1 - \mu) + \Theta_{ij}^I \mu(1 - \mu) + O(\varepsilon^2) \\
&= \beta \sum_{r=1}^N (A_{rj} - (1 - \mu)\delta_{ri}) \Phi_{jr} + (1 - \beta)(1 - \mu) \Phi_{ij} + \mu(1 - \mu) \Theta_{ij}^I + O(\varepsilon^2). \quad (\text{S.15})
\end{aligned}$$

For the last step we have made use of $A_{rj}\delta_{ri} = A_{ij}\delta_{ri} = \delta_{ri}$, since these equations correspond to a link between nodes i and j , thus we are implicitly assuming that $A_{ij} = 1$.

Summarizing, the linearized equations of the ELE model can be expressed as:

$$\Theta_{ij}^I = \beta(1 - \mu) \Phi_{ij} + \beta(1 - \mu) \Phi_{ji} + (1 - \mu)^2 \Theta_{ij}^I, \quad (\text{S.16})$$

$$\Phi_{ij} = \beta \sum_{r=1}^N (A_{rj} - (1 - \mu)\delta_{ri}) \Phi_{jr} + (1 - \beta)(1 - \mu) \Phi_{ij} + \mu(1 - \mu) \Theta_{ij}^I. \quad (\text{S.17})$$

Section S3. Epidemic threshold

We have shown in *Methods* that the epidemic threshold is obtained by finding the non-trivial solutions of the system of equations

$$\frac{\mu}{\beta}\epsilon_i = \sum_j B_{ji}\epsilon_j, \quad (\text{S.18})$$

where the components of matrix B read

$$B_{ij} = (1 - \Upsilon)A_{ij} - \Upsilon k_i \delta_{ij}, \quad (\text{S.19})$$

with the constant Υ being

$$\Upsilon = \frac{\beta(1 - \mu)}{\mu(2 - \mu) + 2\beta(1 - \mu)}. \quad (\text{S.20})$$

The non-trivial solutions of Eq. (S.18) require $\frac{\mu}{\beta}$ to be an eigenvalue $\Lambda(B)$ of matrix B :

$$\Lambda(B) = \frac{\mu}{\beta} \quad (\text{S.21})$$

Unfortunately, for any given fixed value of the recovery rate μ , matrix B also depends on the infection rate β through Υ , thus Eq. (S.21) becomes an implicit equation for β . Moreover, since we are interested in the onset of the epidemics, we must find the lowest value of the infection probability, β_c , which satisfies Eq. (S.21). At first sight, one would say that we must choose the maximum eigenvalue of B to obtain the lowest value of β . However, the dependence of B on β may rise the question whether a different eigenvalue (e.g., the second largest eigenvalue) could solve the equation at a lower value of β , since each eigenvalue has a different functional form.

We show in Fig. S12 that, no matter the functional form of the eigenvalues of B on β , if there exists a solution to Eq. (S.21), the curve $\frac{\mu}{\beta}$ always crosses the largest eigenvalue line most-to-the-left than for any other of the eigenvalues, due to the decreasing behavior of $\frac{\mu}{\beta}$. Thus, we can safely say express that

$$\beta_c = \frac{\mu}{\Lambda_{\max}(B)}, \quad (\text{S.22})$$

which is the final implicit equation for the epidemic threshold β_c . In Fig. S12 we have made use of an Erdős-Rényi network with 100 nodes, average degree $\langle k \rangle = 6$, and setting $\mu = 0.5$, but the previous result is general no matter the network or the parameters.

Section S4. Data description

Description of the 27 real networks used in Figs. 6 and S11, sorted by increasing number of nodes. They have been obtained from the Network Repository (<http://networkrepository.com>). Table S1 provides their main structural characteristics.

ia-infect-dublin Human contact network where nodes represent humans and edges represent proximity (i.e., contacts in the physical world), during the Infectious SocioPatterns event that took place at the Science Gallery in Dublin, Ireland.

soc-wiki-Vote Wikipedia voting data from the inception of Wikipedia till January 2008. Nodes represent Wikipedia users and a directed edge from node i to node j represents that user i voted on user j .

ca-CSphd Genealogy network of PhD's in computer science.

ia-fb-messages The Facebook-like Social Network originate from an online community for students at University of California, Irvine. The dataset includes the users that sent or received at least one message.

soc-hamsterster Network of the friendship and family links between users of Hamsterster social network.

socfb-USFCA72 A social friendship network extracted from Facebook consisting of people (nodes) with edges representing friendship ties.

socfb-nips-ego A social friendship network extracted from Facebook consisting of people (nodes) with edges representing friendship ties.

socfb-Santa74 A social friendship network extracted from Facebook consisting of people (nodes) with edges representing friendship ties.

ca-GrQc Collaboration network of arXiv General Relativity. Nodes represent scientists, and links coauthorship.

web-spam Web Spam Challenge 2008 network.

power-US-Grid US Power grid graph.

ca-Erdos992 Erdős collaboration network. Nodes represent scientists, and links coauthorship.

soc-advogato Advogato is a social community platform where users can explicitly express weighted trust relationships among themselves. The dataset contains a list of all of the user-to-user links.

p2p-Gnutella08 Gnutella peer to peer network from August 8 2002.

ia-reality Reality mining network data consists of human mobile phone call events between a small set of core users at the MIT whom actually were assigned mobile phones for which all calls were collected. A node represents a person; an edge indicates a phone call or voicemail between two users.

ca-HepTh Collaboration network of arXiv High Energy Physics Theory. Nodes represent scientists, and links coauthorship.

soc-anybeat Anybeat is an online community, a public gathering place where you can interact with people from around your neighborhood or across the world.

ca-AstroPh Collaboration network of arXiv Astrophysics. Nodes represent scientists, and links coauthorship.

ca-CondMat Collaboration network of arXiv Condensed Matter. Nodes represent scientists, and links coauthorship.

soc-gplus Google+ social network.

tech-as-caida2007 Internet network at the level of autonomous systems as of 2017. Nodes represent autonomous systems, and there exists a link between them if they have a business agreement for the routing of packets.

ia-email-EU The network was generated using email data from a large European research institution. Nodes are users and edges represent email exchanges between two users in both directions.

ia-enron-large Network of emails exchanged between senior managers of Enron Corporation, during the period which lead to its bankruptcy.

soc-brightkite Brightkite is a location-based social networking service provider where users shared their locations by checking-in. The dataset contains all links among users.

soc-epinions Who-trust-whom online social network of the general consumer review site Epinions.com.

soc-slashdot A technology-related news website known for its specific user community. The dataset contains friend/foe tags between the users of slashdot.

soc-twitter-follows Twitter follower network.

The prefix in the name of each network indicates the category it belongs to, namely: (ia) interaction networks; (soc) social networks; (ca) collaboration networks; (socfb) Facebook networks; (web) we graphs; (power) power networks; (p2p) peer to peer networks; (tech) technological networks.

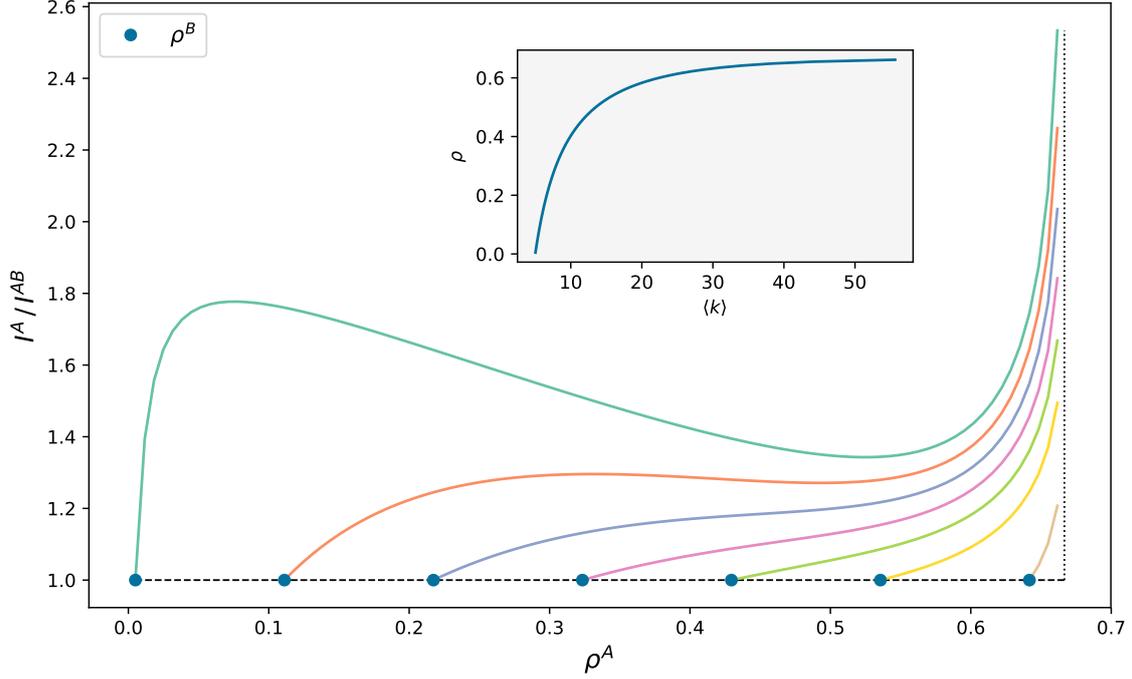


Fig. S1. Ratio between the link epidemic importance I^A of a link in a subnetwork A and the link epidemic importance I^{AB} of a link that acts as the only bridge between subnetworks A and B . First, we fix the average degree $\langle k \rangle_B$ of subnetwork B (or equivalently, we fix its incidence ρ^B , the red circles), and then we consider subnetworks A with average degree (and epidemic incidence) larger than that of B , i.e. $\langle k \rangle_A \geq \langle k \rangle_B$ (thus, $\rho^A \geq \rho^B$). We can see that, in all cases, $I^A \geq I^{AB}$, meaning that the ranking by link epidemic importance will not be led by the bridges. The vertical dotted line highlights the asymptote at $\rho = 1/(1 + \mu)$. The inset shows the relationship between the incidence and the average degree. We have set the epidemic parameters to $\mu = 0.5$ and $\beta = 0.1$, and the calculations rely on a nonperturbative homogeneous mean field approximation (npHoMF). See section S1 for further details.

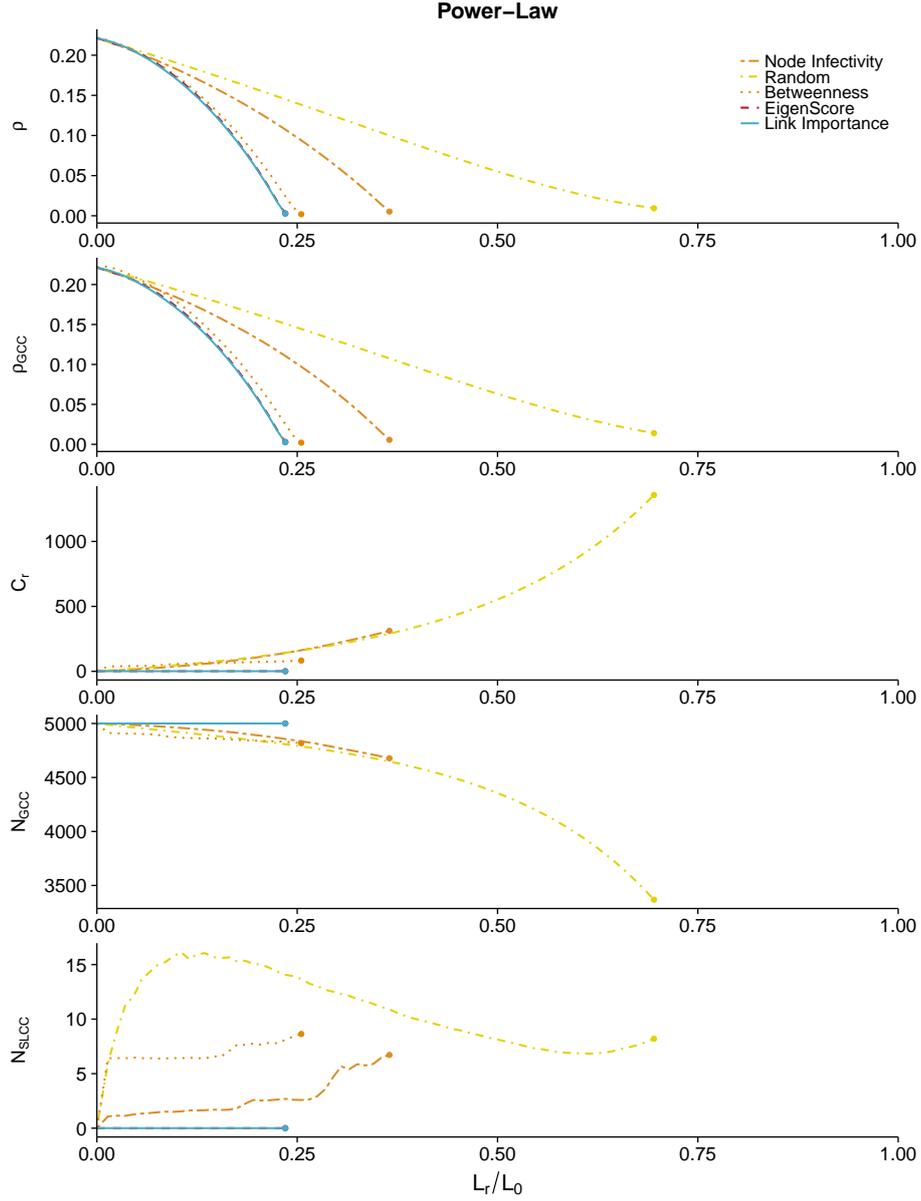


Fig. S2. Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, and average degree $\langle k \rangle = 6$. Five containment strategies are compared: maximum probability of being infected (Node Infectivity), random link selection (Random), maximum edge betweenness (Betweenness), maximum eigenscore (EigenScore), and maximum link epidemic importance (Link Importance). In the horizontal axis we represent the fraction of removed links during the containment process (L_r/L). We show, from top to bottom: ρ , incidence of the epidemics on the network; ρ_{GCC} , incidence of the epidemics on the Giant Connected Component; C_r , number of connected components; N_{GCC} , size of the GCC; N_{SLCC} , size of the Second Largest Connected Component. The dots mark the achievement of total containment.

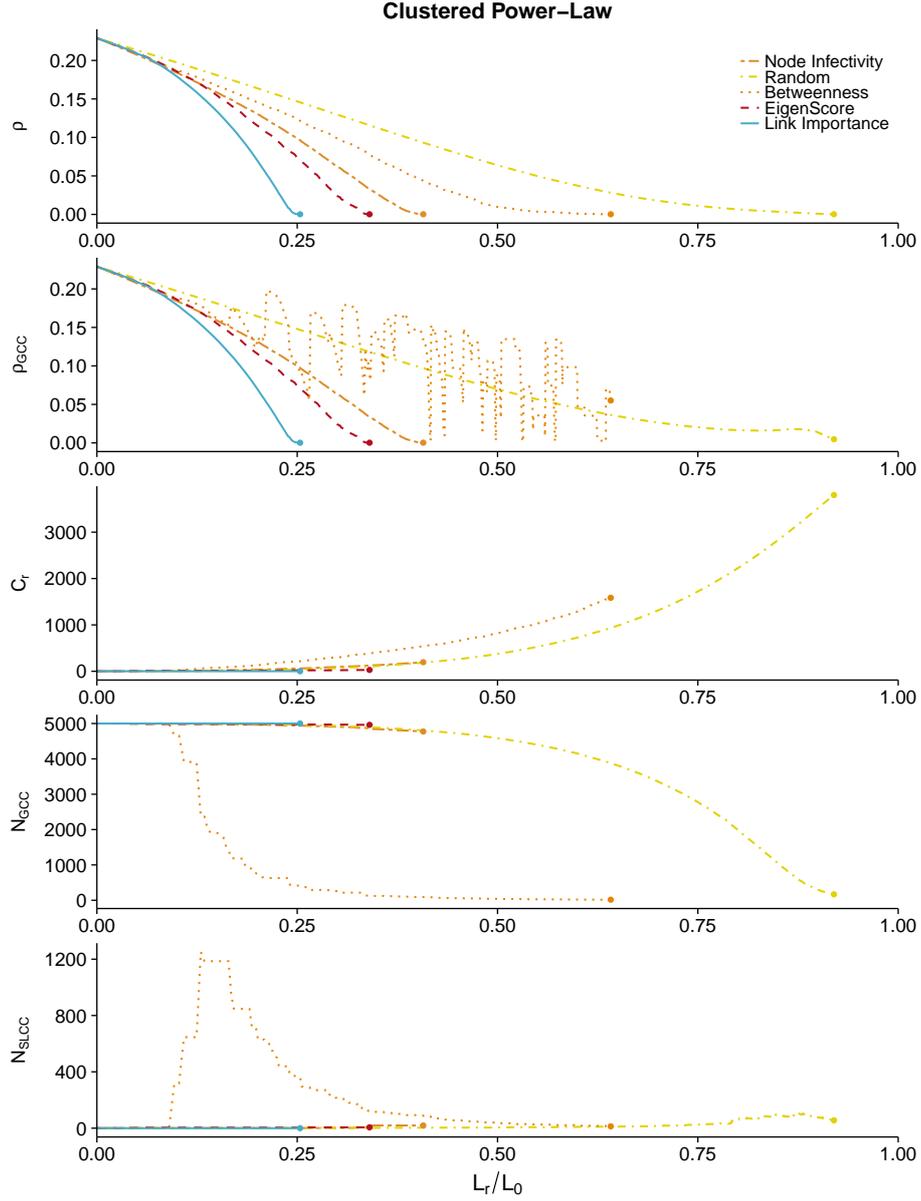


Fig. S3. Epidemic containment for a network with 5000 nodes, power-law degree distribution of exponent 3, high clustering coefficient, and average degree $\langle k \rangle = 6$ (see Methods for more details). Five containment strategies are compared: maximum probability of being infected (Node Infectivity), random link selection (Random), maximum edge betweenness (Betweenness), maximum eigenscore (EigenScore), and maximum link epidemic importance (Link Importance). In the horizontal axis we represent the fraction of removed links during the containment process (L_r/L). We show, from top to bottom: ρ , incidence of the epidemics on the network; ρ_{GCC} , incidence of the epidemics on the Giant Connected Component; C_r , number of connected components; N_{GCC} , size of the GCC; N_{SLCC} , size of the Second Largest Connected Component. The dots mark the achievement of total containment.

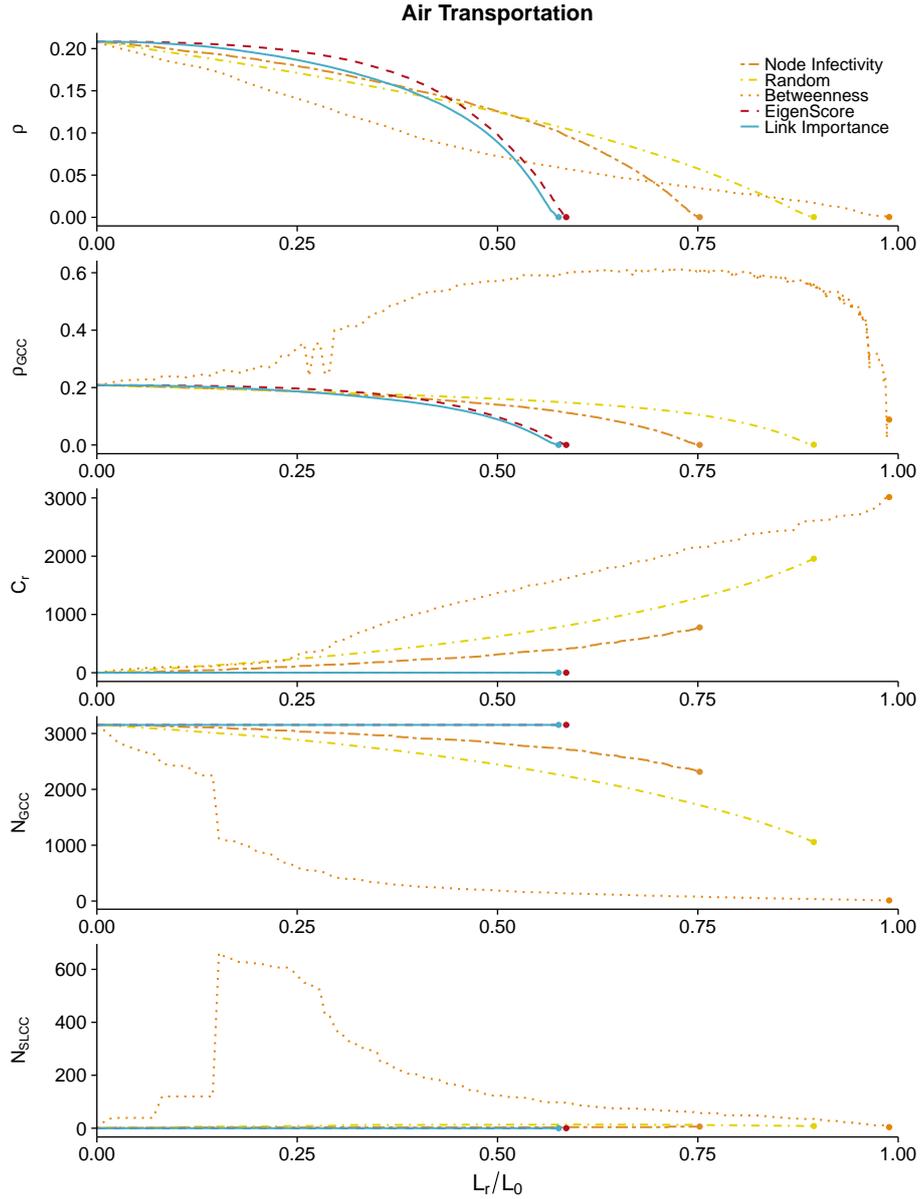


Fig. 4. Epidemic containment for the air transportation network (see Methods for more details). Five containment strategies are compared: maximum probability of being infected (Node Infectivity), random link selection (Random), maximum edge betweenness (Betweenness), maximum eigenscore (EigenScore), and maximum link epidemic importance (Link Importance). In the horizontal axis we represent the fraction of removed links during the containment process (L_r/L). We show, from top to bottom: ρ , incidence of the epidemics on the network; ρ_{GCC} , incidence of the epidemics on the Giant Connected Component; C_r , number of connected components; N_{GCC} , size of the GCC; N_{SLCC} , size of the Second Largest Connected Component. The dots mark the achievement of total containment.

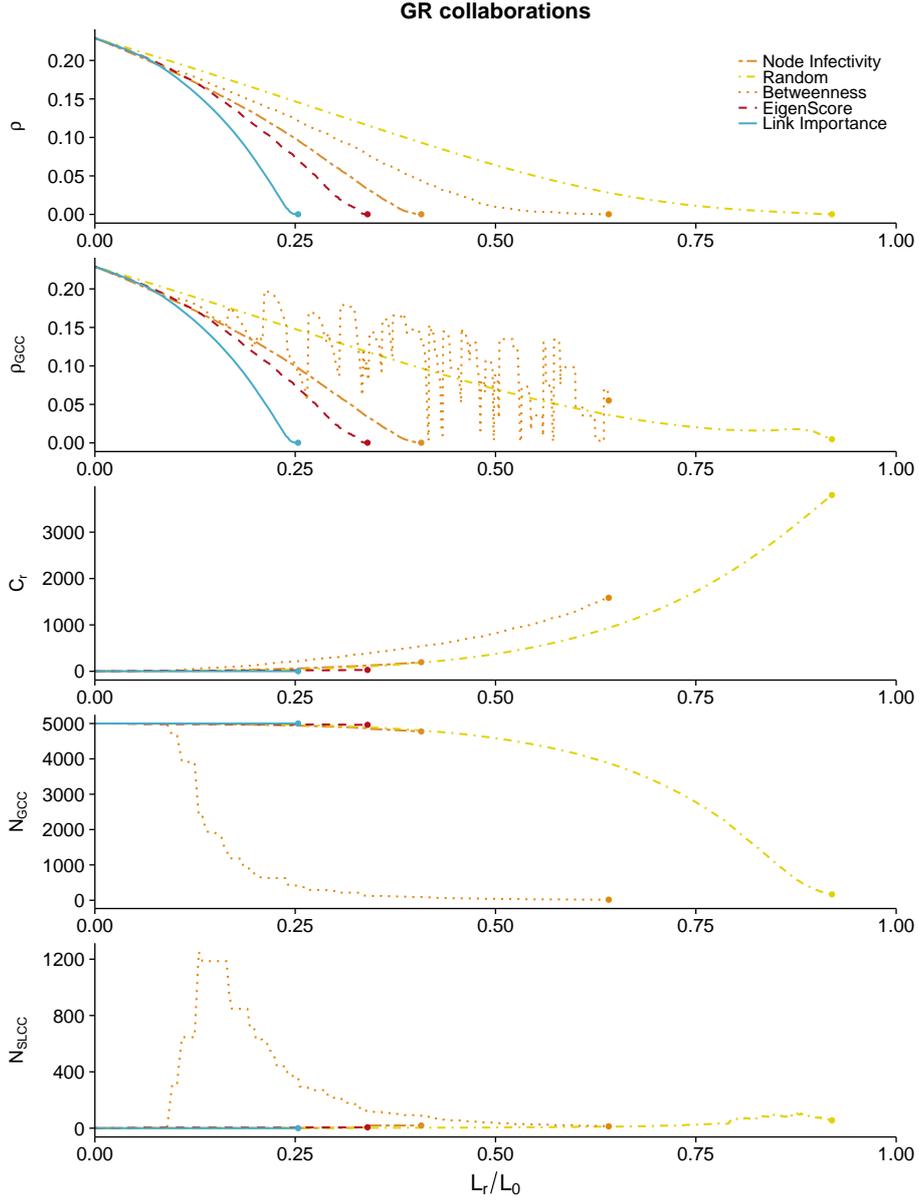


Fig. 5. Epidemic containment for the general relativity collaborations network (see Methods for more details). Five containment strategies are compared: maximum probability of being infected (Node Infectivity), random link selection (Random), maximum edge betweenness (Betweenness), maximum eigenscore (EigenScore), and maximum link epidemic importance (Link Importance). In the horizontal axis we represent the fraction of removed links during the containment process (L_r/L). We show, from top to bottom: ρ , incidence of the epidemics on the network; ρ_{GCC} , incidence of the epidemics on the Giant Connected Component; C_r , number of connected components; N_{GCC} , size of the GCC; N_{SLCC} , size of the Second Largest Connected Component. The dots mark the achievement of total containment.

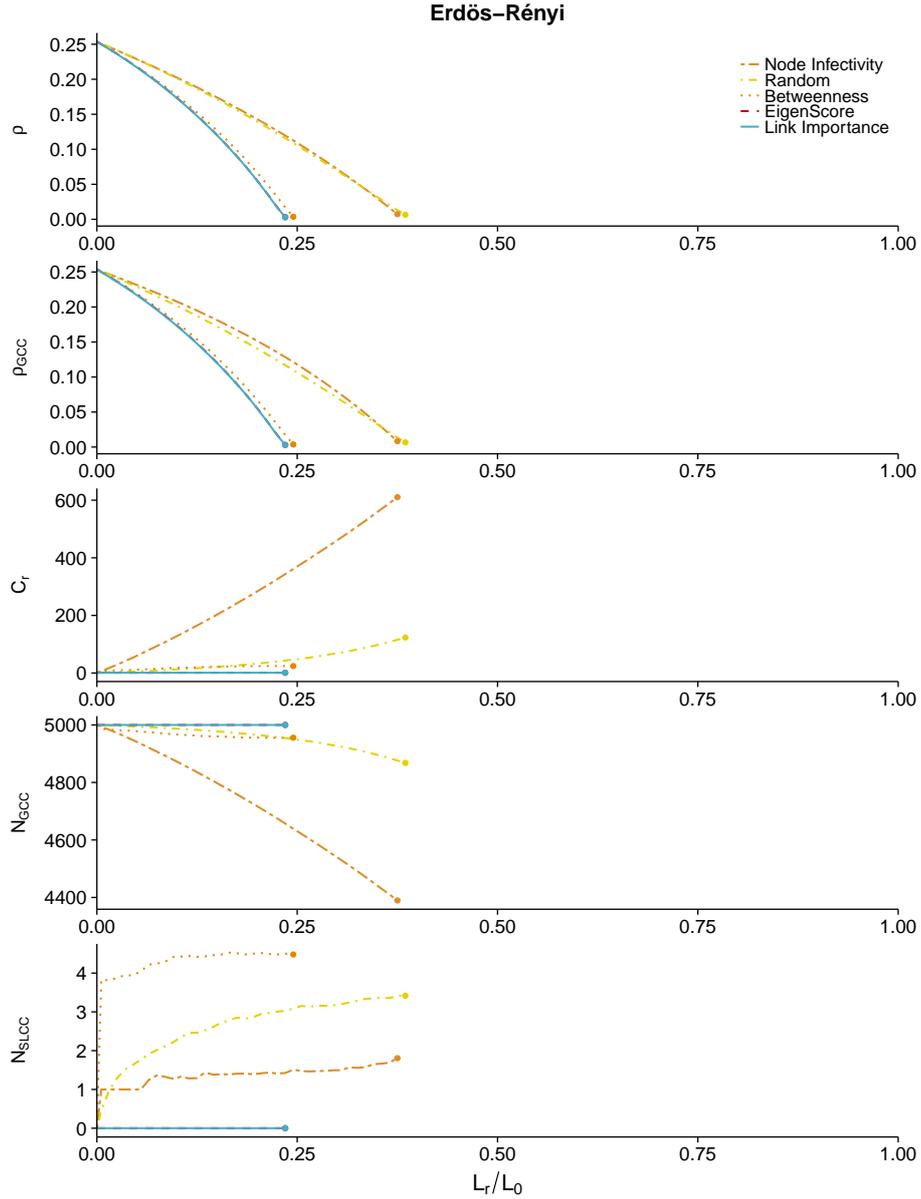


Fig. S6. Epidemic containment for an ER network with 5000 nodes and average degree $\langle k \rangle = 6$. Five containment strategies are compared: maximum probability of being infected (Node Infectivity), random link selection (Random), maximum edge betweenness (Betweenness), maximum eigenscore (EigenScore), and maximum link epidemic importance (Link Importance). In the horizontal axis we represent the fraction of removed links during the containment process (L_r/L). We show, from top to bottom: ρ , incidence of the epidemics on the network; ρ_{GCC} , incidence of the epidemics on the Giant Connected Component; C_r , number of connected components; N_{GCC} , size of the GCC; N_{SLCC} , size of the Second Largest Connected Component. The dots mark the achievement of total containment. We observe that the containment is similar to the one in the power-law network in Fig. S2 due to the absence of transitivity and modular structure.

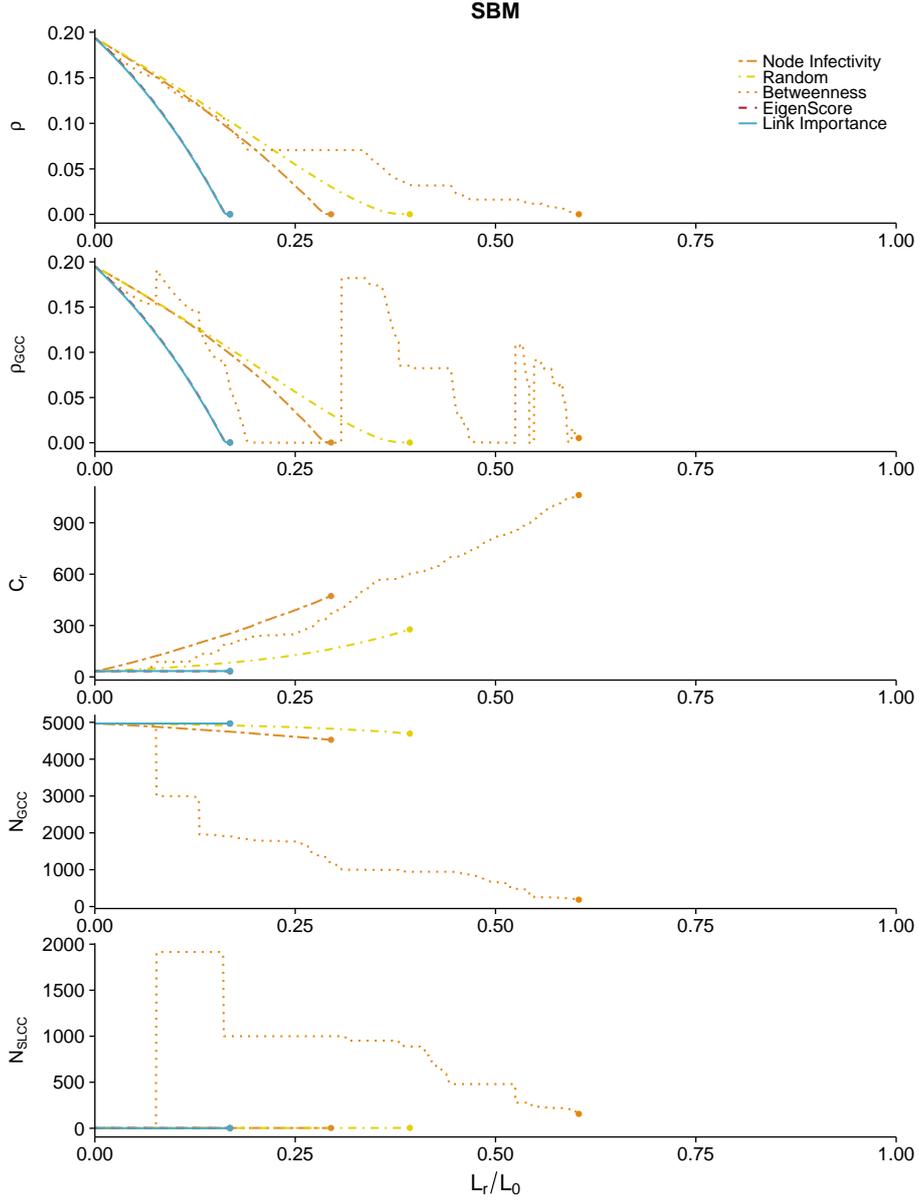


Fig. S7. Epidemic containment for a network with 5000 nodes generated with a stochastic block model, with four blocks of 250 nodes, two blocks of 1000 nodes, and one block of 2000 nodes, average degree 5, and mixing probability of 0.3. Five containment strategies are compared: maximum probability of being infected (Node Infectivity), random link selection (Random), maximum edge betweenness (Betweenness), maximum eigenscore (EigenScore), and maximum link epidemic importance (Link Importance). In the horizontal axis we represent the fraction of removed links during the containment process (L_r/L). We show, from top to bottom: ρ , incidence of the epidemics on the network; ρ_{GCC} , incidence of the epidemics on the Giant Connected Component; C_r , number of connected components; N_{GCC} , size of the GCC; N_{SLCC} , size of the Second Largest Connected Component. The dots mark the achievement of total containment. We observe that our containment strategy based on link epidemic importance outperforms all other methods except eigenscore, with similar results, and the large fragmentation induced by the other strategies.

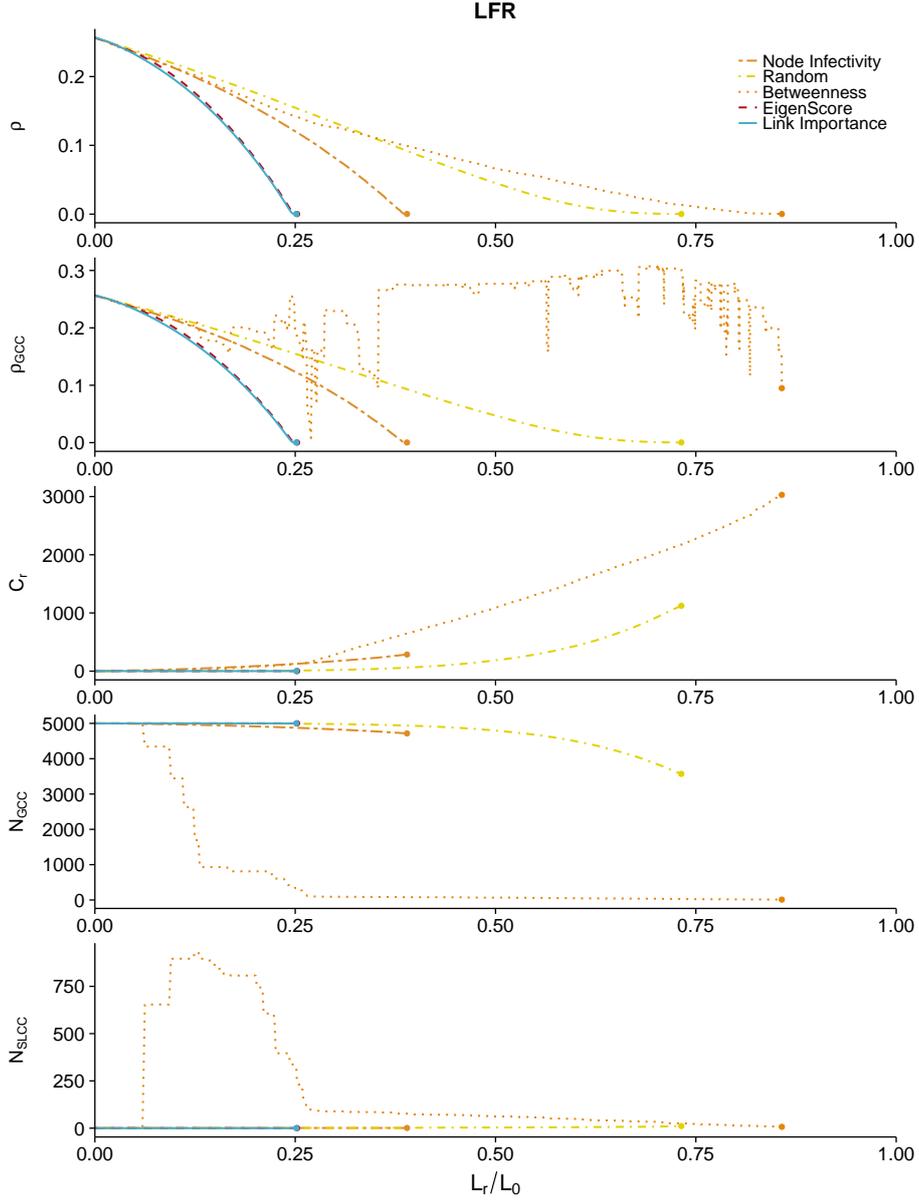


Fig. S8. Epidemic containment for a network with 5000 nodes generated using the LFR algorithm [Lancichinetti et al., Phys. Rev. E, 78 (2008) 046110], with average degree of 6, exponent of 3, and mixing probability 0.1. Five containment strategies are compared: maximum probability of being infected (Node Infectivity), random link selection (Random), maximum edge betweenness (Betweenness), maximum eigenscore (EigenScore), and maximum link epidemic importance (Link Importance). In the horizontal axis we represent the fraction of removed links during the containment process (L_r/L). We show, from top to bottom: ρ , incidence of the epidemics on the network; ρ_{GCC} , incidence of the epidemics on the Giant Connected Component; C_r , number of connected components; N_{GCC} , size of the GCC; N_{SLCC} , size of the Second Largest Connected Component. The dots mark the achievement of total containment. We observe that our containment strategy based on link epidemic importance outperforms all other methods except eigenscore, with similar results, and the large fragmentation induced by the other approaches.

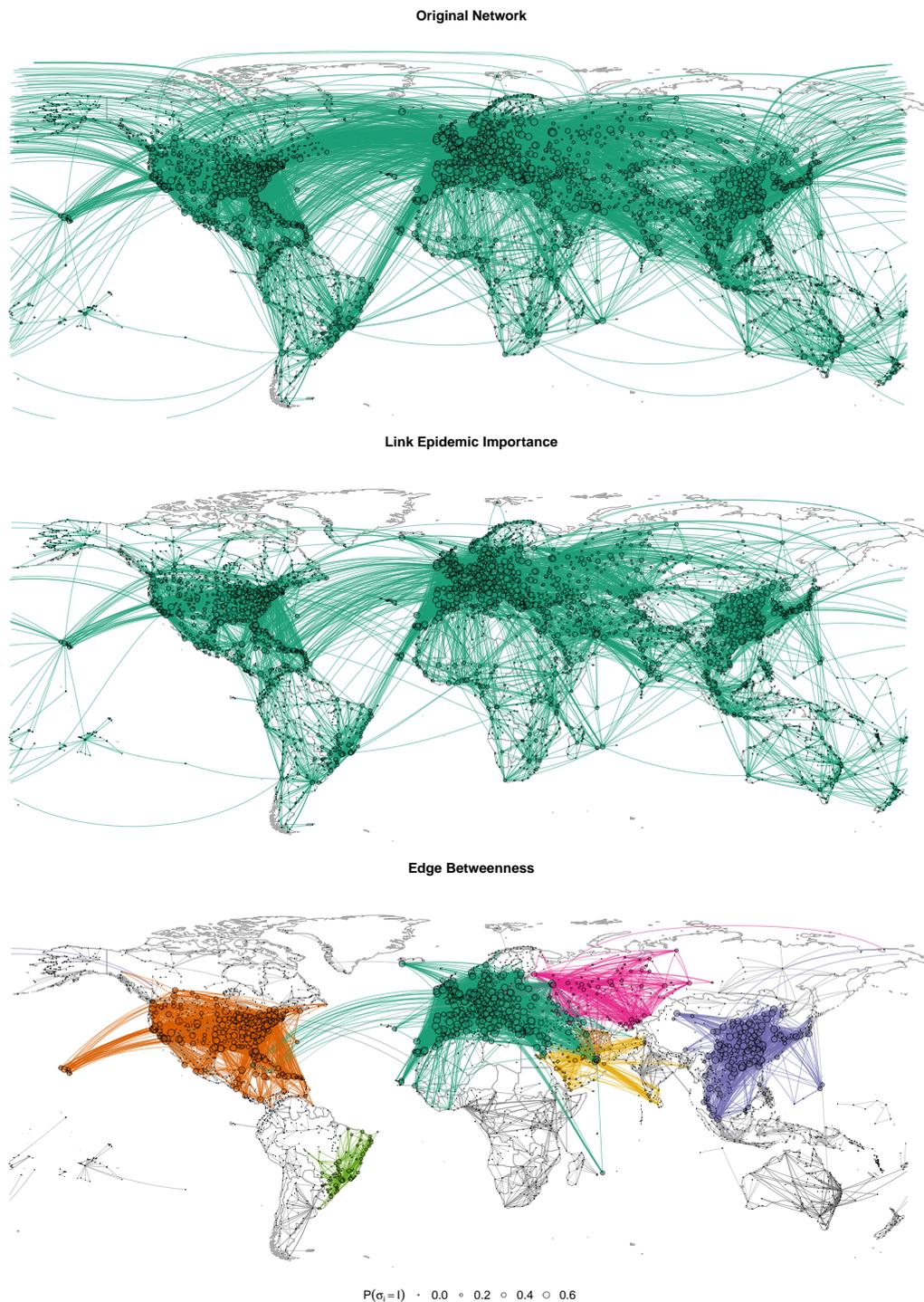


Fig. S9. Original air transportation network (top) and the results after a removal of 33.3% of the links using link epidemic importance (middle) and edge betweenness (bottom). Nodes and edges with the same color belong to the same connected component, with subcritical components in gray scale and using darker gray for larger components. The area of the nodes is proportional to their probability of being infected. We have set the epidemic parameters to $\mu = 0.5$ and $\beta = 0.06$.

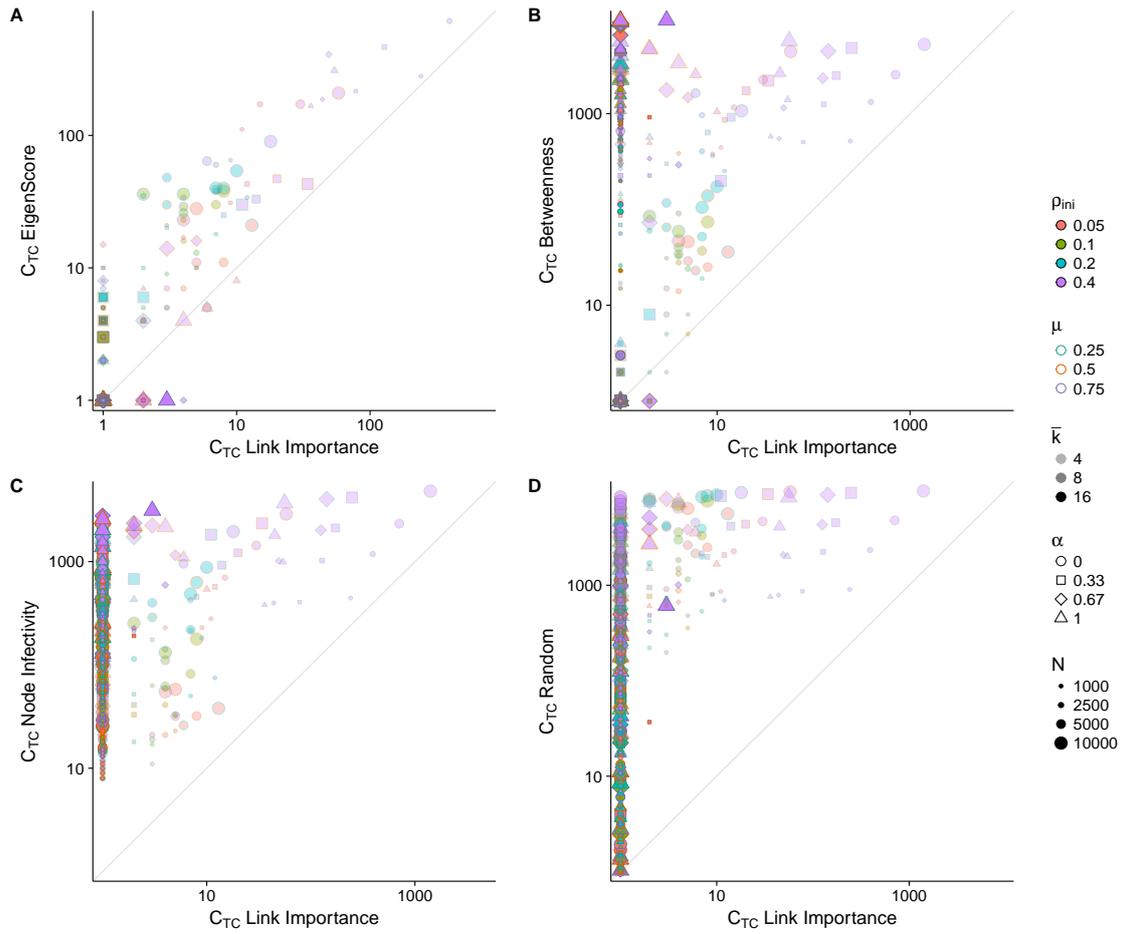


Fig. S10. Comparison of the number of connected components after total containment between the link epidemic importance strategy and the other four methods, calculated for the synthetic networks and parameters as in Fig. 4.

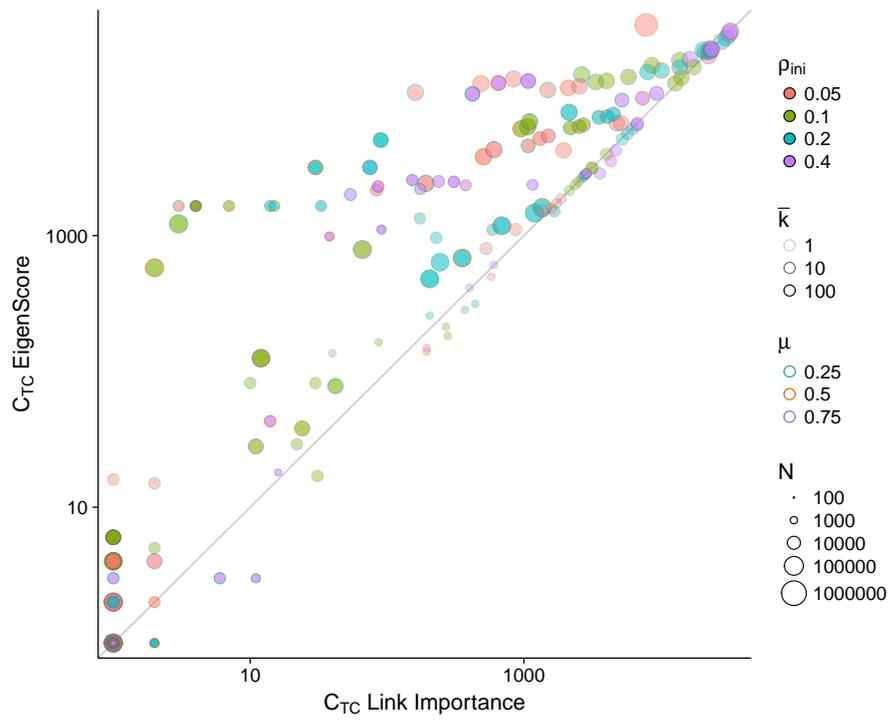


Fig. S11. Comparison of the number of connected components after total containment between the link epidemic importance and eigenscore strategies, calculated for the real networks and parameters as in Fig. 5.

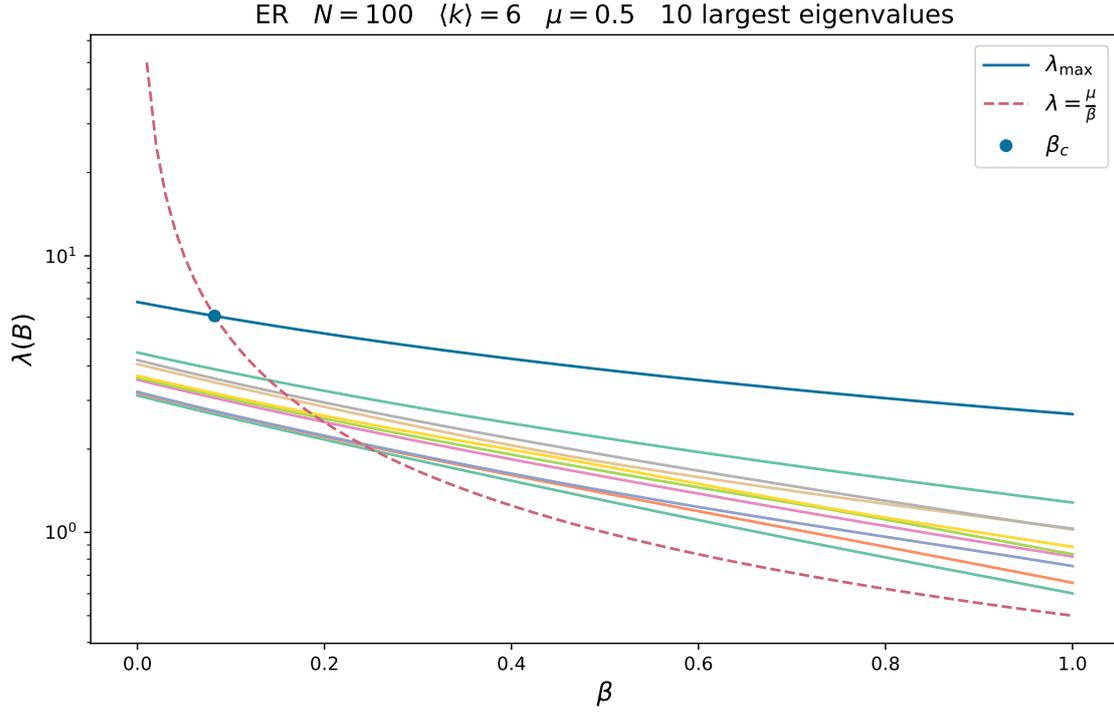


Fig. S12. Graphical representation of the determination of the epidemic threshold. The solid lines show the 10 largest eigenvalues of matrix B , which depends on the infection probability β and the recovery probability μ , for an Erdős-Rényi network of 100 nodes, average degree $\langle k \rangle = 6$, and setting $\mu = 0.5$. For the epidemic threshold, the eigenvalue must be equal to $\frac{\mu}{\beta_c}$, thus it must lay at the intersection between the eigenvalues curves and the dashed line $\frac{\mu}{\beta}$. From those intersections, the one providing the smallest value of β is given by the largest eigenvalue of B , thus showing that $\beta_c = \frac{\mu}{\lambda_{\max}(B)}$, which is an implicit equation for the epidemic threshold.

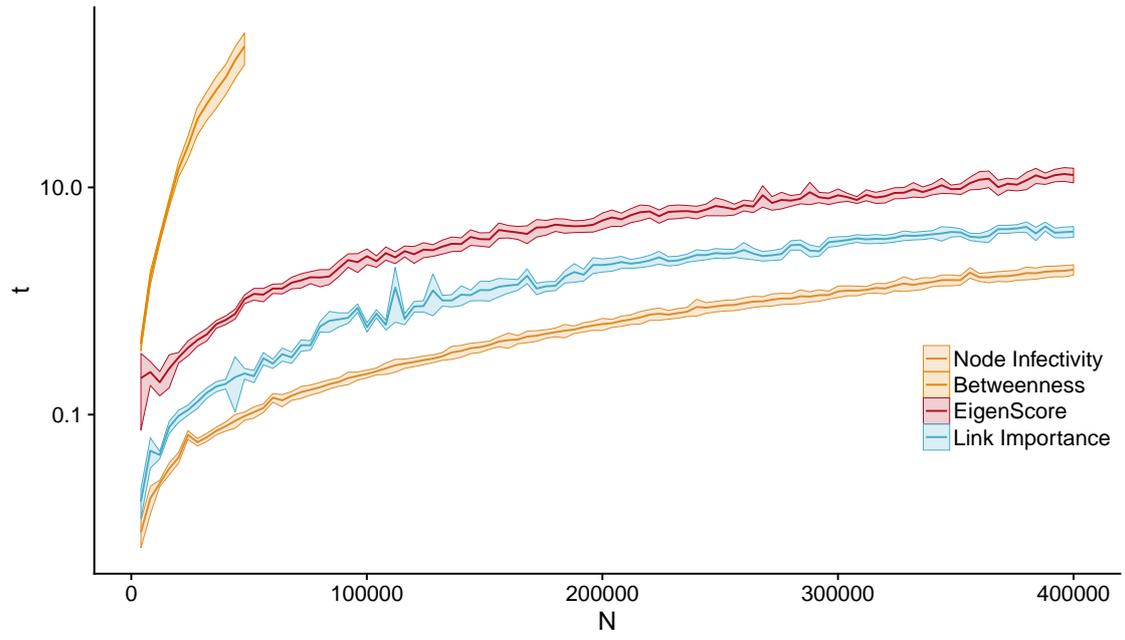


Fig. S13. Computational time invested for each method to perform a single ranking and removal for BA networks ranging from 100 to 400,000 nodes, averaged over 36 repetitions. The standard deviations are included as ribbons on the plot.

Table S1. Structural characteristics of the 27 real networks obtained from the Network Repository (<http://networkrepository.com>) and used in Fig. 6 and fig. S11. They correspond to the largest connected component of the networks with the same name in the repository. The structural descriptors shown are the number of nodes N , the number of links L , the average degree $\langle k \rangle$, the clustering coefficient c , and the assortativity r .

Name	N	L	$\langle k \rangle$	c	r
ia-infect-dublin	410	2765	13.4878	0.1452	0.0000
soc-wiki-Vote	889	2914	6.5557	0.0901	-0.0556
ca-CSphd	1025	1043	2.0351	0.0023	-0.2532
ia-fb-messages	1266	6451	10.1912	0.6289	0.6392
soc-hamsterster	2000	16097	16.0970	0.0268	-0.3816
socfb-USFCA72	2672	65244	48.8353	0.0037	-0.3886
socfb-nips-ego	2888	2981	2.0644	0.3178	0.2013
socfb-Santa74	3578	151747	84.8222	0.2618	0.1253
ca-GrQc	4158	13422	6.4560	0.2294	0.0227
web-spam	4767	37375	15.6807	0.0006	-0.8764
power-US-Grid	4941	6594	2.6691	0.0925	-0.0952
ca-Erdos992	4991	7428	2.9766	0.0439	-0.0844
soc-advogato	5054	39374	15.5813	0.0207	0.0355
p2p-Gnutella08	6299	20776	6.5966	0.0420	-0.4531
ia-reality	6809	7680	2.2558	0.1032	0.0035
ca-HepTh	8638	24806	5.7435	0.1907	0.0917
soc-anybeat	12645	49132	7.7710	0.0024	-0.6753
ca-AstroPh	17903	196972	22.0044	0.4357	0.2258
ca-CondMat	21363	91286	8.5462	0.1273	-0.0288
soc-gplus	23576	39145	3.3207	0.0073	-0.1946
tech-as-caida2007	26475	53381	4.0326	0.0260	-0.0651
ia-email-EU	32430	54397	3.3547	0.0004	-0.6682
ia-enron-large	33696	180811	10.7319	0.2022	0.0706
soc-brightkite	56739	212945	7.5061	0.2811	0.2389
soc-opinions	61355	494372	16.1151	0.0217	-0.1234
soc-slashdot	70068	358647	10.2371	0.0851	-0.1165
soc-twitter-follows	404719	713319	3.5250	0.1105	0.0096