Localization of nonbacktracking centrality on dense subgraphs of sparse networks

G. Timár ●,[*] S. N. Dorogovtsev ●, and J. F. F. Mendes ●

Departamento de Física da Universidade de Aveiro and I3N, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

The nonbacktracking matrix and the related nonbacktracking centrality (NBC) play a crucial role in models of percolation-type processes on networks, such as nonrecurrent epidemics. Here we study the localization of NBC in infinite sparse networks that contain an arbitrary finite subgraph. Assuming the local tree likeness of the enclosing network, and that branches emanating from the finite subgraph do not intersect at finite distances, we show that the largest eigenvalue of the nonbacktracking matrix of the composite network is equal to the highest of the two largest eigenvalues: that of the finite subgraph and of the enclosing network. In the localized state, when the largest eigenvalue of the subgraph is the highest of the two, we derive explicit expressions for the NBCs of nodes in the subgraph and other nodes in the network. In this state, nonbacktracking centrality is concentrated on the subgraph and its immediate neighborhood in the enclosing network. We obtain simple, exact formulas in the case where the enclosing network is uncorrelated. We find that the mean NBC decays exponentially around the finite subgraph, at a rate which is independent of the structure of the enclosing network, contrary to what was found for the localization of the principal eigenvector of the adjacency matrix. Numerical simulations confirm that our results provide good approximations even in moderately sized, loopy, real-world networks.

I. INTRODUCTION

Recent decades have witnessed a surge of scientific activity focused on understanding the behavior of dynamical models on complex network substrates. Much of this research effort has been oriented towards models of epidemic spreading on networks of social acquaintances, such as the paradigmatic susceptible-infected-susceptible (SIS) endemic model, and the susceptible-infected-removed (SIR) model of nonrecurrent epidemics [1,2]. Mean-field theories have been developed for both classes of models, making use of two characteristic matrices: the adjacency matrix and the nonbacktracking (or Hashimoto) matrix [3] of a given network, respectively. In the quenched mean field approximation of the SIS model on a static network, the epidemic threshold is given as the inverse of the largest eigenvalue (LEV) of the adjacency matrix [4–6]. Close to the threshold the probability of a given node being infected is proportional to the corresponding component of the principal eigenvector (PEV). This has the consequence that close to the epidemic threshold, disease may become localized on the largest hubs and their immediate neighborhoods [7]. These results are partly an artifact of the quenched mean field approximation, which is not exact even in infinite locally treelike networks due to neglecting dynamical correlations between the infection states of neighboring nodes. Nevertheless the predicted disease localization also does occur in real-life systems [8–11] and is a crucial phenomenon to understand in epidemic surveillance and control [12–14].

A similar approach for nonrecurrent epidemics, analogous to the quenched mean field in the case of the SIS model, is

the message-passing approximation where it is assumed that the contribution of a node j to the behavior of a neighboring node i is completely determined by the contribution of the neighbors of j, excluding i. This assumption holds exactly for infinite locally treelike networks in the case of percolation [15] and, e.g., the SIR epidemic model [16,17]. The relevant matrix in message-passing theories is the nonbacktracking (NB) matrix **H**, which is a $2L \times 2L$ nonsymmetric matrix (L being the number of links in the network) whose elements are indexed by directed links $i \leftarrow j$, instead of nodes. It is defined as $H_{i \leftarrow j, k \leftarrow l} = \delta_{j,k}(1 - \delta_{i,l})$, where δ is the Kronecker symbol.

Similarly to the quenched mean field in the case of the SIS model, message-passing predicts a percolation or SIR epidemic threshold that is the inverse of the LEV of the NB matrix [17]. Also, the probability, close to the threshold, of node i belonging to the giant connected component (or of being infected in an SIR epidemic) is proportional to the nonbacktracking centrality (NBC) of node i (see Ref. [18]), defined as $\mathrm{NBC}_i \equiv \sum_{j \in \partial_i} v_{i \leftarrow j}$, where ∂_i is the set of neighbors of node i and $v_{i \leftarrow j}$ is the component of the PEV of the NB matrix that corresponds to the directed link $i \leftarrow j$. In other words, NBC_i is the sum of incoming PEV components to node i. We will assume the NBC values to be normalized according to the condition $\sum_i^N \mathrm{NBC}_i = 1$, in a network consisting of N nodes. The NBC emerges in various applications related to nonrecurrent dynamical models, e.g., in designing optimal percolation and immunization strategies [19–21] or identifying influential spreaders [22,23]. Compared to the classical eigenvector centrality (the PEV components of the adjacency matrix), NBC suffers to a much lesser degree from localization, as discussed in Ref. [24]. In particular, isolated hubs cannot be centers of localization, that is, the NBC is able

────────
[*]gtimar@ua.pt

to avoid the artificial "self-inflating" phenomenon associated with high-degree nodes in the case of the adjacency matrix PEV. This circumstance has also made the NB matrix a useful tool in spectral community detection methods [25,26].

Although the PEV of the NB matrix cannot be localized on isolated hubs, localization may still occur on dense subgraphs, e.g., cliques [24], if the subgraph in question has a LEV higher than that of the surrounding network. A recent study of a large set of real-world networks showed that the presence of such dense subgraphs is quite common, in which case the LEV of the network is dominated by the LEV of the subgraph and nodes outside the subgraph have NBC close to zero [27]. In particular, the authors of Ref. [27] identified two specific subgraphs in real-world networks that were often found to be centers of localization of the NBC: the maximum *k*-core and a bipartite structure called *overlapping hubs*. These subgraphs may correspond to localized percolation clusters, and localized disease spreading in the SIR model. In such cases typically a double transition is found, where the first transition is determined by the structure of the dense subgraph and the second by the entire network [27]. The message-passing method is only able to identify the first transition, and treats it as a strict phase transition, which is an artifact of this approach [28]. As an extreme example, consider a small clique with high LEV embedded in a larger network with smaller LEV. Message passing will predict a single phase transition at the inverse of the LEV of the small clique. This is not a true phase transition, in the sense that susceptibility here is still small, of the order of 1. It is rather an indication that localized cooperative phenomena may already occur, namely, the clique and its neighborhood may experience a small epidemic outbreak, in the case of the SIR model. To be able to interpret message-passing results correctly it is vital to have a solid understanding of such localization phenomena.

In the present work we explore the localization of the NBC in a more general setting than previously considered, deriving some exact results and useful approximations. The paper is organized as follows. In Sec. II we describe our network construction which allows for the derivation of exact results: an arbitrary finite subgraph embedded in an arbitrary infinite locally treelike network, where branches emanating from the finite subgraph do not intersect at finite distances. In Sec. III, using the concept of nonbacktracking expansion, we derive an exact expression for the NBCs of nodes in the finite subgraph (*child network*). We show that the LEV of the composite network is exactly given by the maximum of the two LEV values: that of the child network and that of the embedding infinite locally treelike network (*mother network*). Thus localization occurs when the composite LEV is equal to the LEV of the child. We obtain compact approximate formulas for the NBC values in the case where the embedding mother network is uncorrelated. Performing simulations we check that our predictions also work well with a real-world mother network with degree correlations and short loops. In Sec. IV we show that in the localized state the mean NBC decays exponentially with distance from the child network. Interestingly the rate of decay depends only on the LEV of the child network, and is completely independent of the mother network. We show that in finite mother networks the exponential form has a cutoff distance depending on network size and the difference between the LEVs of child and mother network. We give a summary and conclusions in Sec. V.

II. ARBITRARY GRAPH INSERTED IN AN INFINITE LOCALLY TREELIKE NETWORK

We aim to quantify the localization of the NBC on arbitrarily structured small, dense subgraphs embedded in large sparse networks. Our setup is somewhat idealized, enabling us to derive exact results and simple approximate expressions. Nonetheless, as we show in Secs. III and IV in various examples, the strict assumptions can be greatly relaxed and the theory still provides good predictions.

Throughout our derivations we will make extensive use of the concept of nonbacktracking expansion of graphs, introduced in [28], which corresponds to computational trees in computer science [29]. For a given graph $\mathcal{G}$, the nonbacktracking expansion provides an infinite tree constructed as follows. Starting from an arbitrary node i in $\mathcal{G}$ as the root of the tree, perform all possible nonbacktracking walks from this node. Each nonbacktracking walk corresponds to a branch of the infinite tree. Thus the nonbacktracking expansion "unfolds" the original graph into an infinite tree, generating an infinite number of replicas for each node in $\mathcal{G}$. It was shown in [28] that for an arbitrary graph, the LEV of the NB matrix corresponds to the asymptotic branching of the nonbacktracking expansion. (The branching of an infinite treelike network is defined as the limit $\lim_{\ell \to \infty} L_{\ell+1}/L_{\ell}$, where L_{ℓ} is the number of links connecting the nodes at distance ℓ from an arbitrary node to the nodes at distance $\ell + 1$; see Ref. [30].) Furthermore, the NBC of node i in graph $\mathcal{G}$ is equal to the relative frequency of replicas of node i on the surface of the tree at infinity, irrespective of the starting node (root). This useful property was applied recently in Ref. [31] to approximate NBC values of nodes in correlated networks.

Let us consider a system in which an arbitrary finite child network of n nodes is inserted into an infinite locally treelike network in the following way. Each node i, of $n_c \leqslant n$ selected nodes in the child network, is merged with a node of degree $b_1^{(i)}$ in the mother network. By merging we mean that the resulting node (of the composite network) retains the links of both the child node and the mother node. As an example, see Fig. 1(a), where child node i of degree 2 is merged with a mother node of degree 3, resulting in a node of degree 5 in the composite network. Let the sequence of branchings on the mother network branches attached to node i be $b_2^{(i)}, b_3^{(i)}, \ldots, b_m^{(i)}, \ldots$ [see Fig. 1(b)], where $b_{m \to \infty}^{(i)} \to b$. Here b is the branching of the mother network, i.e., the LEV of its NB matrix. For all nodes j of the child network that were not chosen for merging with mother network nodes, $b_1^{(j)}, b_2^{(j)}, b_3^{(j)}, \ldots = 0$. Importantly, the branches attached to different nodes in the child network do not intersect within any finite distance. Figure 1 demonstrates this construction. The noninteraction of branches emanating from child nodes is a strong assumption, which certainly does not hold in real-world networks that tend to have many short loops. As we will see, the theory also provides good approximations in these more realistic scenarios.

Now let us denote the LEV of the NB matrix of a child network (to be inserted into a mother network) by z. This is
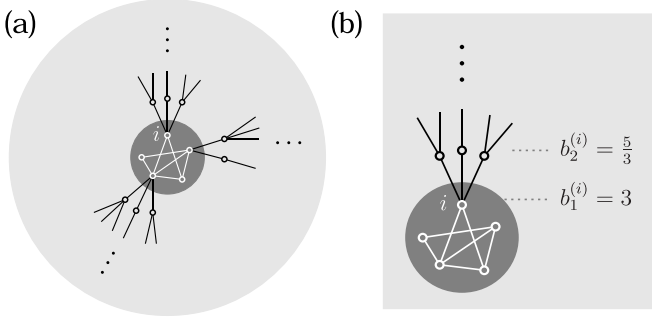
FIG. 1. (a) A finite child network (white nodes and links on dark grey shaded area) inserted into an infinite locally treelike mother network (black nodes and links on light gray shaded area). The branches emanating from nodes of the child network do not intersect within any finite distance. In this example the number of nodes selected for merging with mother network nodes is $n_c = 3$. (b) First two elements in the sequence of branching numbers on the branch connected to node i of the child network.

the asymptotic branching of the nonbacktracking expansion of this subgraph. This means that asymptotically the number of vertices on the surface of the expansion at distance m is

$$H_m \cong Cz^m, \tag{1}$$

where C is a constant. Let the NBC of node i in the child network be x_i, $i = 1, 2, \ldots, n$, normalized according to $\sum_{i=1}^{n} x_i = 1$.

III. NONBACKTRACKING CENTRALITIES IN THE CHILD NETWORK

A. Results for general networks

Let us perform the nonbacktracking expansion of the composite network, starting from one of the child nodes (see Ref. [28] for details of this procedure). The number of replica nodes of the child network on the surface of the nonbacktracking expansion at large distances m will still be Cz^m, but there will be additional vertices from the mother network. We want to find the fraction ρ of replica nodes of the child network on the surface of the expansion as $m \to \infty$. This is equal to the sum of NBCs of child nodes. (We denote the NBC of child node i by NBC_i. Note that this value—the NBC measured in the composite network—differs from x_i.) Let Q_m be the total number of nodes on the surface of the expansion at distance m. In the limit of infinite m,

$$\rho = \lim_{m \to \infty} \frac{Cz^m}{Q_m}. \tag{2}$$

For large m we have the following recursion relation:

$$Q_m = Cz^m + Cz^{m-1} \sum_{i=1}^{n} x_i b_1^{(i)} + Cz^{m-2} \sum_{i=1}^{n} x_i b_1^{(i)} b_2^{(i)}$$

$$+ Cz^{m-3} \sum_{i=1}^{n} x_i b_1^{(i)} b_2^{(i)} b_3^{(i)} + \cdots . \tag{3}$$

Let us explain this equation. The first term on the right-hand side is the number of nonbacktracking walks of length m that stay inside the child network throughout the m steps. The

second term is the number of nonbacktracking walks of length m that spend $m-1$ steps inside the child network and exit it only in the last step. In general, the ith term gives the number of nonbacktracking walks of length m that spend $m - i + 1$ steps inside the child network and exit it in the next step, never returning due to the nonintersecting tree nature of all branches emanating from the child network. Equations (2) and (3) lead to the following relation:

$$\frac{1}{\rho} = 1 + \sum_{i=1}^{n} x_i \frac{b_1^{(i)}}{z} + \sum_{i=1}^{n} x_i \frac{b_1^{(i)} b_2^{(i)}}{z^2} + \sum_{i=1}^{n} x_i \frac{b_1^{(i)} b_2^{(i)} b_3^{(i)}}{z^3} + \cdots . \tag{4}$$

Hence we have

$$\rho = \left[\sum_{i=1}^{n} x_i \left(1 + \frac{b_1^{(i)}}{z} + \frac{b_1^{(i)} b_2^{(i)}}{z^2} + \frac{b_1^{(i)} b_2^{(i)} b_3^{(i)}}{z^3} + \cdots \right) \right]^{-1}. \tag{5}$$

We immediately see from Eq. (5) that $\rho = 0$ if $b \geqslant z$, which implies that, in this case, the LEV of the composite network is $\lambda_{\mathrm{cmp}} = b$. On the other hand, in the localized state, when $z > b$, the sum in parentheses converges, resulting in $\rho > 0$. In this case, as we show in Sec. IV, almost all nodes outside the child network have negligible NBC: only the nodes of the child network and nodes close to it have NBC noticeably different from zero. Since the branches emanating from the child network do not intersect, no finite neighborhood of the child network contains additional loops (compared to those already present in the child network), and so the subgraph comprising the nodes with nonnegligible NBCs has the same LEV as the child network, and hence, $\lambda_{\mathrm{cmp}} = z$. Therefore, we have in general that

$$\lambda_{\mathrm{cmp}} = \max(z, b). \tag{6}$$

Making use of the Collatz-Wielandt formula, a corollary of the Perron-Frobenius theorem, one can easily show that the LEV of any network is greater than or equal to the LEV of an arbitrary subgraph of the network (see, e.g., Ref. [24]). According to Eq. (6) this inequality is actually a strict equality in the case of an arbitrary child network in an infinite locally treelike mother network. Equation (6) is confirmed in Fig. 2, where, as an example, the small *Karate club* network [32] is inserted into a large Erdős-Rényi (ER) network of varying mean degree. (An ER mother network was chosen for convenience, as it is locally treelike, and its LEV is closely approximated by its mean degree, for large sizes.) The insertion is made in a way that a number n_c of nodes of the child network are merged with n_c randomly chosen nodes in the mother network. The LEV of the composite network was found to follow Eq. (6) very closely for small n_c, where the condition of nonintersecting branches is a decent approximation. For larger n_c there is less "space" in the mother network for the branches to be independent, making the theory less accurate, and resulting in a smoother crossover of the composite LEV. This is a finite size effect, and would disappear for $N \to \infty$ [see Fig. 2(a)]. A rough criterion for the validity of the results derived for $N \to \infty$ is given in Sec. III D.

We can write Eq. (5) in simpler form. Since $b_{m\to\infty}^{(i)} \to b$, we have, asymptotically,

$$\prod_{k=1}^{m} b_k^{(i)} \cong c_i b^m, \qquad (7)$$

where c_i is a constant,

$$c_i \equiv \lim_{m\to\infty} \frac{1}{b^m} \prod_{k=1}^{m} b_k^{(i)}. \qquad (8)$$

The function ρ^{-1} contains a singularity at $z = b$, hidden in Eqs. (4) and (5). Extracting this singularity we obtain

$$\rho = \left[\sum_{i=1}^{n} x_i \left(c_i + 1 - c_i + \frac{c_i b + b_1^{(i)} - c_i b}{z} + \frac{c_i b^2 + b_1^{(i)} b_2^{(i)} - c_i b^2}{z^2} + \frac{c_i b^3 + b_1^{(i)} b_2^{(i)} b_3^{(i)} - c_i b^3}{z^3} + \cdots \right) \right]^{-1}$$

$$= \left[\sum_{i=1}^{n} x_i \left(\frac{c_i}{1 - b/z} + 1 - c_i + \frac{b_1^{(i)} - c_i b}{z} + \frac{b_1^{(i)} b_2^{(i)} - c_i b^2}{z^2} + \frac{b_1^{(i)} b_2^{(i)} b_3^{(i)} - c_i b^3}{z^3} + \cdots \right) \right]^{-1}, \qquad (9)$$

or, in more compact form,

$$\rho = \left\{ \frac{\sum_{i=1}^{n} x_i c_i}{z - b} z + \sum_{i=1}^{n} x_i \left[1 - c_i + \sum_{k=1}^{\infty} \frac{1}{z^k} \left(\prod_{j=1}^{k} b_j^{(i)} - c_i b^k \right) \right] \right\}^{-1}. \qquad (10)$$

Note that the only assumption for the mother network was local tree likeness. The mother network may possess any correlations, as long as tree likeness is fulfilled; it may be random or deterministic.

The NBC of a given node i of the child network is simply

$$\text{NBC}_i = x_i \rho. \qquad (11)$$

This is easily confirmed by noticing that the fraction of replicas of node i, among the replicas of all child nodes, on the surface of the nonbacktracking expansion at infinity is x_i regardless of the structure of the surrounding mother network.

Throughout the rest of the paper, unless explicitly stated otherwise, we will always consider the localized state, i.e., when $z > b$.

B. Approximation for finite distances

Equation (10) is an exact result, given the somewhat strict assumptions of an infinite locally treelike mother network, and nonintersecting branches emanating from child nodes. To turn this result into a useful expression in real networks, we consider distances only up to a finite m value. One can measure the branching numbers $b_1^{(i)}$, $b_2^{(i)}$, ... up to $b_m^{(i)}$, for all child nodes i, and assume that the branching numbers are equal to b thereafter. The constants c_i are in this case defined as

$$c_i \equiv \frac{1}{b^m} \prod_{k=1}^{m} b_k^{(i)}, \qquad (12)$$

and the formula for ρ changes to

$$\rho \approx \left\{ \frac{\sum_{i=1}^{n} x_i c_i}{z - b} z \right.$$
$$\left. + \sum_{i=1}^{n} x_i \left[1 - c_i + \sum_{k=1}^{m-1} \frac{1}{z^k} \left(\prod_{j=1}^{k} b_j^{(i)} - c_i b^k \right) \right] \right\}^{-1}. \qquad (13)$$

We can derive a first order approximation by considering just one step away from the child network, i.e., $m = 1$,

$$\rho \approx \frac{z - b}{\sum_{i=1}^{n} x_i (z - b + k_i)}, \qquad (14)$$

where $k_i = b_1^{(i)}$ is the degree of the mother network node with which the child node i was merged.

C. Random mother network

Informative expressions can be found for average quantities when the child network is inserted into a random uncorrelated network with an arbitrary degree distribution. In this case we can obtain $\langle \rho^{-1} \rangle$ exactly, averaged over the
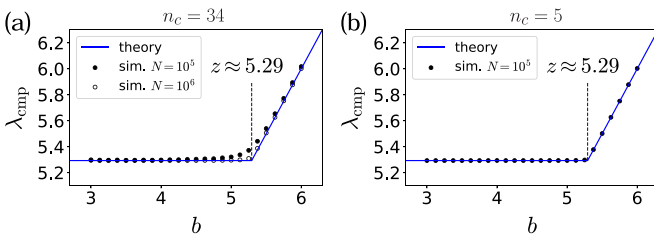

FIG. 2. LEV of the composite network consisting of the Karate club network ($n = 34$, $z \approx 5.29$) inserted into an ER network of varying mean degree $\langle k \rangle \approx b$. Solid black dots correspond to simulation results for a mother network size of $N = 10^5$, and empty black dots correspond to a mother network size of $N = 10^6$. The solid blue line corresponds to the result for $N \to \infty$, Eq. (6). (a) The Karate club network was inserted into the ER mother network by merging all $n_c = n = 34$ nodes with random nodes of the mother network. (b) Only $n_c = 5$ random nodes of the Karate club network were merged with random nodes of the mother network. Simulation results were averaged over 100 realizations for each point in both panels.

members of the configuration model ensemble, using Eq. (10). In this expression only c_i and $b_j^{(i)}$ are random variables, and they appear independently, so we can perform the average:

$$\langle \rho^{-1} \rangle = \frac{\sum_{i=1}^{n} x_i \langle c_i \rangle}{z - b} z$$

$$+ \sum_{i=1}^{n} x_i \left[1 - \langle c_i \rangle + \sum_{k=1}^{\infty} \frac{1}{z^k} \left(\left\langle \prod_{j=1}^{k} b_j^{(i)} \right\rangle - \langle c_i \rangle b^k \right) \right]. \tag{15}$$

Here, for an uncorrelated network, the branching $b = \langle k^2 \rangle / \langle k \rangle - 1$, where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moments of the mother network's degree distribution. (We assume a finite second moment.) Note that in uncorrelated networks the sequence $b_1^{(i)}, b_2^{(i)}, b_3^{(i)}, \dots$ almost surely converges to b for any i. Also, $b_1^{(i)}, b_2^{(i)}, b_3^{(i)}, \dots, b_k^{(i)}$ are independent random variables with mean values $\langle k \rangle, b, b, \dots, b$. Therefore, $\langle \prod_{j=1}^{k} b_j^{(i)} \rangle = \langle k \rangle b^{k-1}$, for any $k \geqslant 1$, and hence, using Eq. (8), we have that $\langle c_i \rangle = \langle k \rangle / b$. Using these observations Eq. (15) can be written as

$$\langle \rho^{-1} \rangle = \frac{\langle k \rangle + z - b}{z - b}. \tag{16}$$

Note the similarity between Eq. (16) and the first-order approximation for general mother networks, Eq. (14). Assuming that the variance of ρ is small, we can estimate

$$\langle \rho \rangle \approx \frac{z - b}{\langle k \rangle + z - b}. \tag{17}$$

By Jensen's inequality we know that

$$\langle \rho \rangle \geqslant \langle \rho^{-1} \rangle^{-1}, \tag{18}$$

so the expression in Eq. (17) is a lower bound on the exact value. Equality holds only if the variance of ρ is zero. Simulations show that this approximation is fairly accurate.

For an ER mother network, $\langle k \rangle = b$, so we have the simple relationship

$$\langle \rho^{-1} \rangle = \frac{z}{z - b}, \tag{19}$$

which is readily observed in Fig. 3. Equations (16) and (19) are completely general exact results, for an arbitrary child network, given that all nodes of the child network are merged with random mother network nodes. If only a random fraction f of nodes in the child network is merged with random mother network nodes, then

$$\langle \rho^{-1} \rangle = \frac{f \langle k \rangle + z - b}{z - b}, \tag{20}$$

and hence, again assuming a small variance of ρ, we have

$$\langle \rho \rangle \approx \frac{z - b}{f \langle k \rangle + z - b}. \tag{21}$$

As an example, Fig. 3 shows results for the sum of NBCs of nodes in the Karate club network inserted into an ER network of size $N = 10^5$. The exact formula, Eq. (20), fits the simulation results perfectly when the number of connection nodes n_c is small. For larger n_c deviations from the theory are observed close to the localization transition due to the intersection of
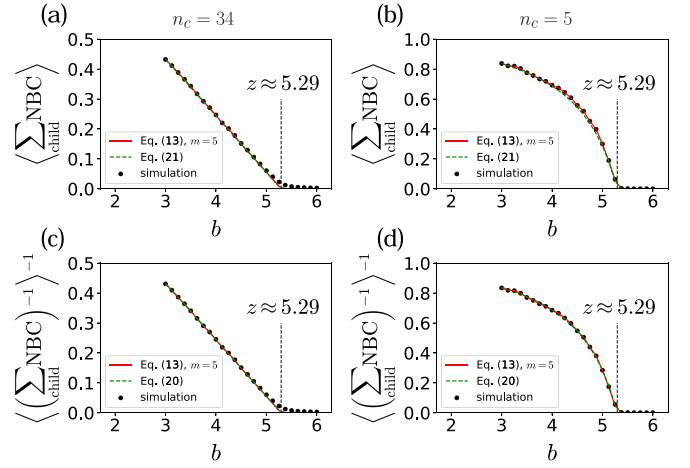

FIG. 3. The Karate club network ($n = 34$, $z \approx 5.29$) inserted into an ER network of $N = 10^5$ with varying mean degree $\langle k \rangle \approx b$. (a), (b) Average sum of NBCs of child nodes and (c), (d) inverse of average inverse sum of NBCs of child nodes, as functions of the LEV of the mother network. The solid red line corresponds to the result Eq. (13), evaluated up to a distance $m = 5$. The dashed green line corresponds to the results for random mother networks, Eqs. (20) and (21). (a), (c) The Karate club network was inserted into the ER mother network by merging all $n_c = n = 34$ nodes with random nodes of the mother network. (b), (d) Only $n_c = 5$ random nodes of the Karate club network were merged with random nodes of the mother network. Simulation results were averaged over 100 realizations for each point in all panels.

branches already at short distances. For decreasing mean degree of the mother network the formula works progressively better, as expected. Equation (21) also works remarkably well, despite being strictly only a lower bound. This supports our assumption of a small variance for the distribution of ρ.

D. Range of validity of the theory

We can give a rough criterion for the validity of the results derived in the limit $N \to \infty$, considering the normalization condition for the NBCs, $\sum_{i=1}^{N} \text{NBC}_i = 1$. As we show in Sec. IV, in the localized state the NBC of nodes in the mother network decays exponentially with distance from the child network. This means that nodes outside the child network, on average, have NBC much smaller than ρ/n, if the network is large. Together with the normalization condition, for large networks, this means that the network size N must satisfy

$$\frac{\rho}{n} N \gg 1. \tag{22}$$

Using Eqs. (22) and (20), and assuming $z \approx b$, we arrive at a rough criterion for the validity of the theory,

$$\frac{N}{n_c} \gg \frac{\langle k \rangle}{z - b}. \tag{23}$$

Equation (23) shows that, as z approaches b, N must increase approximately as $\sim (z - b)^{-1}$ for the theory to remain valid. Conversely, for a fixed network size, the theory becomes less accurate close to the point $z = b$, which is confirmed by Figs. 2 and 3.
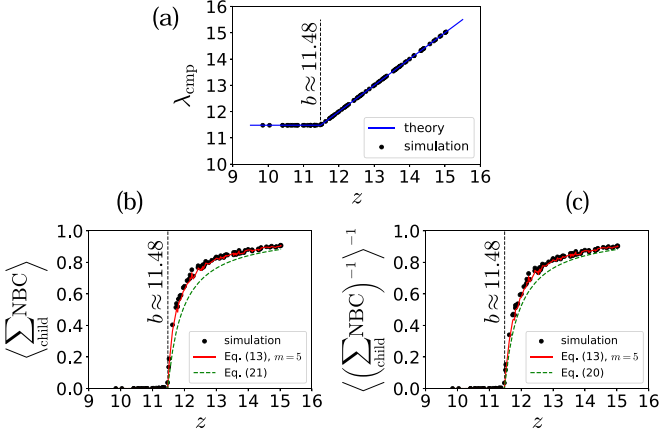
FIG. 4. An ER network of $n = 50$ and varying mean degree, inserted into the Gnutella p2p network ($N = 62\,561$, $b \approx 11.48$) at $n_c = 5$ random nodes. (a) LEV of the composite network as a function of the LEV of the child network. (b) Average sum of NBCs of child nodes. (c) Inverse of average inverse sum of NBCs of child nodes. Simulation results were averaged over 100 realizations for each point in all panels.

E. Real-world mother network

Many real-world networks, particularly social networks, tend to have a multitude of short loops, violating our assumption of nonintersecting branches emanating from the child network. Simulation results indicate that Eq. (13) still works well, and even the results derived for uncorrelated random networks, Eqs. (20) and (21), provide good approximations (see Fig. 4). As an example, a small ($n = 50$) ER network is inserted into the *Gnutella p2p* network [32] by merging $n_c = 5$ random nodes of the child network with random mother network nodes, and the sum of NBCs of the child network is compared with our predictions. (An ER network was, again, chosen only for convenience.)

IV. EXPONENTIAL DECAY OF NONBACKTRACKING CENTRALITY AROUND THE CHILD NETWORK

A. Results for general networks

Let us find how nonbacktracking centrality is distributed in the composite network. Returning to the nonbacktracking expansion of the composite network, let $\zeta_m^{(i)}$ be the fraction of replicas—on the surface of the expansion—of nodes at distance $m \geqslant 1$ from child node i (on the branch attached to node i). This is the sum of the NBCs of these nodes. For $z > b$ Eqs. (3) and (4) immediately lead to the following expression:

$$\zeta_m^{(i)} = \frac{\rho}{z^m} x_i \prod_{k=1}^{m} b_k^{(i)} \cong \rho c_i x_i \left(\frac{b}{z}\right)^m. \qquad (24)$$

According to Eq. (5),

$$\rho + \sum_{i=1}^{n} \sum_{m=1}^{\infty} \zeta_m^{(i)} = 1, \qquad (25)$$

as is natural. Furthermore, since the number of nodes in the branches attached to node i at distance m from this node grows as $\sim c_i b^m$, the NBC for an "average node" at distance m decays

as

$$\frac{\zeta_m^{(i)}}{c_i b^m} \cong \rho x_i z^{-m}, \qquad (26)$$

which is an exponential decay with rate z. Notably the branching of the mother network doesn't play a role here. Using Eq. (24) the sum of NBCs of all nodes at a distance $m \geqslant 1$ from the child network can be written as

$$S_m = \frac{\rho}{z^m} \sum_{i=1}^{n} x_i \prod_{k=1}^{m} b_k^{(i)} \cong \rho \left(\frac{b}{z}\right)^m \sum_{i=1}^{n} c_i x_i, \qquad (27)$$

while the number of nodes at distance $m \geqslant 1$ from the child network,

$$N_m = \sum_{i=1}^{n} \prod_{k=1}^{m} b_k^{(i)} \cong b^m \sum_{i=1}^{n} c_i. \qquad (28)$$

So the mean NBC of nodes at a distance $m \geqslant 1$ from the child network is

$$\langle \text{NBC} \rangle_m = \frac{S_m}{N_m} \cong \rho z^{-m} \frac{\sum_{i=1}^{n} c_i x_i}{\sum_{i=1}^{n} c_i}. \qquad (29)$$

If the child network is a regular graph, then $x_i = 1/n$, and

$$\langle \text{NBC} \rangle_m = \frac{S_m}{N_m} \cong \frac{\rho z^{-m}}{n}. \qquad (30)$$

These results are similar to the case of the adjacency matrix PEV in a localized state, studied in Ref. [7]. The authors also found an exponential decay of eigenvector centrality as a function of the distance from the localization center: the node of highest degree. The authors of Ref. [7] considered a less general situation, a hub of degree q inserted into a Bethe lattice of branching B, and found that the adjacency matrix PEV components decay with distance m as $f_m \propto a^{-m}$, with $a = (q - B)^{1/2}$, and the critical value for localization $q_{\text{loc}} = B^2 + B$. In other words the rate of decay depends on the branching of the surrounding network, whereas, interestingly, in the case of the NBC the decay rate depends only on the LEV of the child network. On the other hand, as the degree of the hub q approaches q_{loc} from above, a tends to B. Thus, at the localization transition both the eigenvector centrality and NBC decay at a rate given by the branching of the mother network.

Localization in a given vector $\vec{h} = (h_1, h_2, \ldots, h_N)$ is often studied using the inverse participation ratio, $Y_4(\vec{h}) = (\sum_i^N h_i^4)/(\sum_i^N h_i^2)^2$, observing how this quantity scales with system size N. For the case of NBCs in our study, when $z < b$, we have $Y_4(\text{NBC}) \sim N^{-1}$, indicating a delocalized state. For $z > b$, considering that the decay of NBC around the child network is exponential, we obtain $Y_4(\text{NBC}) \sim \text{const}$, indicating that localization occurs on a finite subgraph, as expected.

B. Random mother network

If the mother network is a random uncorrelated network, we can calculate some averages approximately. In this case let us define an average NBC at distance m as

$$\overline{\text{NBC}}_m \equiv \frac{\langle S_m \rangle}{\langle N_m \rangle}, \qquad (31)$$
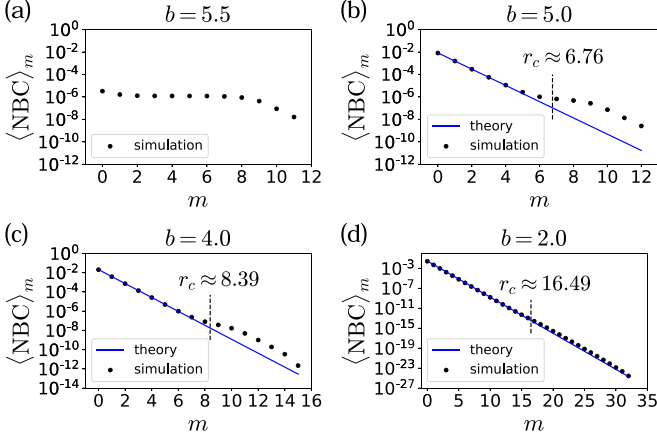
FIG. 5. Exponential decay of NBC around the child network: the Karate club network ($n = 34$, $z \approx 5.29$) inserted into an ER network of $N = 10^6$ with varying mean degree $\langle k \rangle \approx b$. The insertion was made at $n_c = 5$ random points. Black dots show mean NBC as a function of distance from the child network, for different LEV values of the ER mother network. Solid blue lines correspond to the result, Eq. (34). The range of exponential decay, r_c, given by Eq. (43) is marked with dashed black lines. In panel (a) localization of the NBC is absent, and the mean NBC is almost uniform. Simulation results were averaged over 10 realizations for each point in all panels.

where $\langle S_m \rangle$ and $\langle N_m \rangle$ are the averages over the network ensemble. We have

$$\langle S_m \rangle \approx f \langle \rho \rangle \langle k \rangle \frac{b^{m-1}}{z^m}, \tag{32}$$

$$\langle N_m \rangle = f n \langle k \rangle b^{m-1}, \tag{33}$$

where f is the fraction of child nodes that are merged with mother network nodes. [In Eq. (32) we made the approximation $\langle \rho \prod_{k=1}^{m} b_k^{(i)} \rangle \approx \langle \rho \rangle \langle \prod_{k=1}^{m} b_k^{(i)} \rangle$, again assuming a small variance for ρ.] We obtain

$$\overline{\mathrm{NBC}}_m = \frac{\langle S_m \rangle}{\langle N_m \rangle} \approx \frac{\langle \rho \rangle z^{-m}}{n}. \tag{34}$$

As an example, Fig. 5 shows results for the decay of the NBC around the child network in a composite network consisting of the Karate club network inserted, at $n_c = 5$ random points, into an ER network of $N = 10^6$ and varying mean degree. The LEV of the child network is ≈ 5.29, therefore in Fig. 5(a) localization is absent due to the higher LEV of the mother network. In the cases where localization is present [Figs. 5(b)–5(d)] the theoretical prediction, Eq. (34) provides a very good fit up to a cutoff distance which depends on the LEV (i.e., the branching, which in this case is closely approximated by the mean degree) of the mother network. As an example of a real-world mother network Fig. 6 shows results for a small ER network inserted into the Gnutella p2p network. For small distances away from the child network, the theory works remarkably well.

In both Figs. 5 and 6 the mean NBC noticeably deviates, above a certain distance, from the exponential decay predicted by Eq. (34). This is a finite-size effect related to the normalization of the NBCs and can be understood as follows. Due to
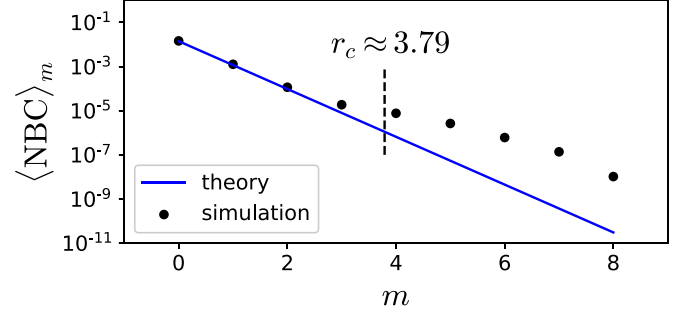

FIG. 6. Mean NBC as a function of distance from the child network: an ER network of $n = 50$ and $z \approx 12.11$ inserted into the Gnutella p2p ($N = 62\,561$, $b \approx 11.48$) network by merging $n_c = 5$ random nodes in both networks. The range of exponential decay, r_c, given by Eq. (43) is marked with a dashed black line. Simulation results were averaged over 100 realizations.

the normalization of the NBCs we must have

$$1 \approx \langle \rho \rangle \left[1 + \frac{f \langle k \rangle}{b} \sum_{m=1}^{\infty} \left(\frac{b}{z} \right)^m \right], \tag{35}$$

which results in

$$\langle \rho \rangle \approx \frac{z - b}{f \langle k \rangle + z - b}, \tag{36}$$

in accordance with Eq. (21). Equation (34) is approximately valid in an infinite locally treelike mother network for any m value. The rate of decay of the exponential in Eq. (34) is such that when summing the NBC values of all distances, $m \to \infty$, the normalization condition, Eq. (35), holds. For a finite network, assuming the same ρ value and the same exponential decay rate but a finite maximum distance $m_{\max}$, the normalization condition is necessarily violated,

$$1 > \langle \rho \rangle \left[1 + \frac{f \langle k \rangle}{b} \sum_{m=1}^{m_{\max}} \left(\frac{b}{z} \right)^m \right]. \tag{37}$$

Therefore, to restore correct normalization, for finite mother networks there must exist a cutoff distance above which the NBC decays more slowly. Note that the finite network considered here is still large enough so that ρ is given by the theory derived for infinite mother networks, i.e., we assume that the criterion in Eq. (23) is satisfied.

C. Decay of nonbacktracking centrality in finite graphs

Here we give an approximation to the cutoff distance above which the NBC must decay more slowly than the exponential predicted by Eq. (34). For simplicity let us assume that the mean NBC decays exponentially up to a cutoff distance r_c, and is uniform above that value up to the effective radius

$$R \equiv \frac{\ln N + \ln(b-1) - \ln\langle k \rangle - \ln n - \ln f}{\ln b} \approx \frac{\ln N}{\ln b}. \tag{38}$$

(Here the effective radius R is the average maximum distance from the child network in a random mother network of size N, with mean branching b.) First, let us write the mean

number of nodes up to distance r in a random mother network,

$$\sum_{m=0}^{r} \langle N_m \rangle \approx \frac{fn\langle k \rangle b^r}{b-1}, \tag{39}$$

and the mean sum of NBCs up to distance r, assuming the exponential decay of Eq. (32),

$$\sum_{m=0}^{r} \langle S_m \rangle \approx \langle \rho \rangle + \frac{f\langle \rho \rangle \langle k \rangle \left[\left(\frac{b}{z} \right)^r - 1 \right]}{z \left(\frac{b}{z} - 1 \right)}. \tag{40}$$

We can write an equation for normalization:

$$1 \approx \langle \rho \rangle + \frac{f\langle \rho \rangle \langle k \rangle \left[\left(\frac{b}{z} \right)^{r_c} - 1 \right]}{z \left(\frac{b}{z} - 1 \right)}$$
$$+ \frac{\langle \rho \rangle}{n} z^{-r_c} \left(N - fn\langle k \rangle \frac{b^{r_c}}{b-1} \right), \tag{41}$$

where the first two terms on the right-hand side give the sum of NBCs up to the cutoff distance r_c from the child network; see Eq. (40). The third term on the right-hand side gives the sum of NBCs of all remaining nodes, that is, nodes at distances larger than r_c, considering that NBC from distance r_c onwards is assumed to be uniform, and using Eq. (39). Accounting for Eq. (21), we have

$$\frac{f\langle k \rangle + z - b}{z - b} \approx 1 + \frac{f\langle k \rangle \left[\left(\frac{b}{z} \right)^{r_c} - 1 \right]}{z \left(\frac{b}{z} - 1 \right)}$$
$$+ z^{-r_c} \left(\frac{N}{n} - f\langle k \rangle \frac{b^{r_c}}{b-1} \right). \tag{42}$$

This results in the cutoff distance

$$r_c \approx \frac{\ln N + \ln(z - b)}{\ln b} + C, \tag{43}$$

with

$$C = \frac{\ln(b-1) - \ln(z-1) - \ln\langle k \rangle - \ln(fn)}{\ln b}. \tag{44}$$

Expressed using the effective radius R,

$$r_c \approx R - \frac{\ln(z-1) - \ln(z-b)}{\ln b} < R. \tag{45}$$

We see that $r_c \to R$ as $b \to 1$, i.e., approaching the percolation threshold in the mother network. Equation (43) slightly overestimates the cutoff distance, as seen in Figs. 5–7. This is due to the fact that the mean NBC displays only a short plateau at the cutoff distance and then continues to decay, contrary to our assumption. The prediction of the cutoff distance, relative to the actual observable value, progressively improves with increasing network size, as shown in Fig. 7, where the Karate club network was inserted into ER mother networks of size ranging from $N = 10^3$ to 10^6.

One can easily show that r_c in Eq. (43) is positive within the range of validity of our theory, Eq. (23), as it must be.

V. DISCUSSION AND CONCLUSIONS

In this paper we have considered the problem of localization of the NBC in networks where the LEV of the
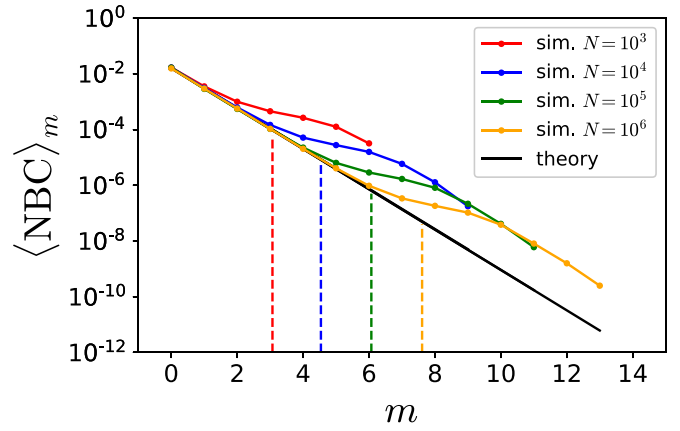

FIG. 7. The Karate club network ($n = 34$, $z \approx 5.29$) inserted into ER networks of mean degree $\langle k \rangle = 4.5 \approx b$ and sizes $N = 10^3$, 10^4, 10^5, and 10^6. The insertion was made at $n_c = 5$ random points. Solid lines with dots show mean NBC as a function of distance from the child network. The solid black line corresponds to the result, Eq. (34). The range of exponential decay, r_c, given by Eq. (43) is marked with dashed lines. Increasing the size of the mother network by a factor of 10 corresponds to a shift of the range r_c by a value $\ln 10 / \ln 4.5 \approx 1.53$, as predicted by Eq. (43).

nonbacktracking matrix of a small subgraph is larger than that of the surrounding network. Our results were obtained for a composite network where a completely arbitrary child network is inserted into an infinite locally treelike mother network. The mother network is allowed to have any correlations, but must be such that the branches emanating from child nodes do not intersect at finite distances. This condition allowed us to utilize the concept of nonbacktracking expansion to derive an exact expression for the sum of NBCs of child nodes. For uncorrelated random mother networks a simple expression was found. For large mother networks the NBCs of child nodes are small (zero in the limit of infinite mother network size) when the LEV of the child network is less than or equal to the LEV of the mother network. Localization on the child network occurs when the child LEV is larger than the mother LEV, and in this case the child node NBCs are strictly positive in the infinite size limit. We show that, in our construction, the LEV of the composite network coincides with the maximum of the two LEVs: that of the child and that of the mother network. This result is a useful addition to what is already known due to the Collatz-Wielandt formula: that the LEV of any network is greater than or equal to the LEV of an arbitrary subgraph of the network. Strict equality applies only in the case of an infinite locally treelike mother network with nonintersecting branches, however, as we show, the result is accurate also in moderately sized random networks and even in real-world sparse, loopy networks. Our expressions for the NBC of child nodes proved to be good approximations in all these situations.

Similarly to what was seen in the localization of eigenvector centrality (the PEV components of the adjacency matrix), we found an exponential decay of the mean NBC around the child network in the localized state. The rate of the decay, however, is given only by the LEV of the child network, i.e., is independent of the mother network. In infinite systems the

exponential decay means that almost all nodes outside the child network have zero NBC. In the language of epidemic spreading in the SIR model, for example, this would mean that disease is only able to spread locally, in the child network or very close to it. In finite networks the exponential decay is only valid up to a certain cutoff distance above which the mean NBC decays more slowly. Our estimate for this cutoff distance agrees well with the simulation results. Interestingly, in simulations of random networks, the mean NBC appears to return to an exponential decay after a short plateau at the predicted cutoff distance. This behavior, involving peripheral nodes, is not explained by our theory and will require further work to understand.

Our findings contribute to a better understanding of nonrecurrent dynamical models, such as the SIR model of epidemic spreading, on certain real-world network structures, where localization may occur due to small, dense subgraphs. Our results were derived for an arbitrary child network, albeit inserted into an infinite locally treelike mother network. We suggest that the presented ideas will provide a basis for future work on more realistic mother network structures containing finite cycles.

[1] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[2] M. E. J. Newman, Spread of epidemic disease on networks, Phys. Rev. E **66**, 016128 (2002).

[3] K. Hashimoto and Y. Namikawa, *Automorphic Forms and Geometry of Arithmetic Varieties* (Elsevier, Amsterdam, 1989).

[4] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, Epidemic spreading in real networks: An eigenvalue viewpoint, in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems, 2003* (IEEE, Piscataway, NJ, 2003), p. 25.

[5] C. Castellano and R. Pastor-Satorras, Thresholds for Epidemic Spreading in Networks, Phys. Rev. Lett. **105**, 218701 (2010).

[6] S. C. Ferreira, C. Castellano, and 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).

[7] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, Localization and Spreading of Diseases in Complex Networks, Phys. Rev. Lett. **109**, 128702 (2012).

[8] S. N. Dorogovtsev and J. F. F. Mendes, *The Nature of Complex Networks* (Oxford University Press, Oxford, 2022).

[9] M. Kulldorff and N. Nagarwalla, Spatial disease clusters: Detection and inference, Stat. Med. **14**, 799 (1995).

[10] F. Tanser, T. Bärnighausen, G. S. Cooke, and M.-L. Newell, Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic, Int. J. Epidemiol. **38**, 1008 (2009).

[11] R. S. Ferreira, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, Metastable localization of diseases in complex networks, Phys. Rev. E **94**, 062305 (2016).

[12] J. L. Herrera, R. Srinivasan, J. S. Brownstein, A. P. Galvani, and L. A. Meyers, Disease surveillance on complex social networks, PLoS Comput. Biol. **12**, e1004928 (2016).

[13] L. D. Valdez, P. A. Macri, and L. A. Braunstein, Intermittent social distancing strategy for epidemic control, Phys. Rev. E **85**, 036108 (2012).

[14] Z. Zhang, H. Wang, C. Wang, and H. Fang, Cluster-based epidemic control through smartphone-based body area networks, IEEE Trans. Parallel Distrib. Syst. **26**, 681 (2014).

[15] B. Karrer, M. E. J. Newman, and L. Zdeborová, Percolation on Sparse Networks, Phys. Rev. Lett. **113**, 208702 (2014).

[16] B. Karrer and M. E. J. Newman, Message passing approach for general epidemic models, Phys. Rev. E **82**, 016101 (2010).

[17] M. Shrestha, S. V. Scarpino, and C. Moore, Message-passing approach for recurrent-state epidemic models on networks, Phys. Rev. E **92**, 022821 (2015).

[18] R. Kühn and T. Rogers, Heterogeneous micro-structure of percolation in sparse networks, Europhys. Lett. **118**, 68003 (2017).

[19] F. Morone and H. A. Makse, Influence maximization in complex networks through optimal percolation, Nature (London) **524**, 65 (2015).

[20] F. Morone, B. Min, L. Bo, R. Mari, and H. A. Makse, Collective influence algorithm to find influencers via optimal percolation in massively large social media, Sci. Rep. **6**, 30062 (2016).

[21] L. Torres, K. S. Chan, H. Tong, and T. Eliassi-Rad, Nonbacktracking eigenvalues under node removal: X-centrality and targeted immunization, SIAM J. Math. Data Sci. **3**, 656 (2021).

[22] F. Radicchi and C. Castellano, Leveraging percolation theory to single out influential spreaders in networks, Phys. Rev. E **93**, 062314 (2016).

[23] B. Min, Identifying an influential spreader from a single seed in complex networks via a message-passing approach, Eur. Phys. J. B **91**, 18 (2018).

[24] T. Martin, X. Zhang, and M. E. J. Newman, Localization and centrality in networks, Phys. Rev. E **90**, 052808 (2014).

[25] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Spectral redemption in clustering sparse networks, Proc. Natl. Acad. Sci. USA **110**, 20935 (2013).

[26] C. Bordenave, M. Lelarge, and L. Massoulié, Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs, in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science* (IEEE, Piscataway, NJ, 2015), p. 1347.

[27] R. Pastor-Satorras and C. Castellano, The localization of non-backtracking centrality in networks and its physical consequences, Sci. Rep. **10**, 21639 (2020).

[28] G. Timár, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, Nonbacktracking expansion of finite graphs, Phys. Rev. E **95**, 042322 (2017).

[29] Y. Weiss and W. T. Freeman, On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs, IEEE Trans. Inf. Theory **47**, 736 (2001).

[30] R. Lyons, The Ising model and percolation on trees and tree-like graphs, Commun. Math. Phys. **125**, 337 (1989).

[31] G. Timár, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, Approximating nonbacktracking centrality and localization phenomena in large networks, Phys. Rev. E **104**, 054306 (2021).

[32] The Karate club and Gnutella p2p networks can be downloaded from http://icon.colorado.edu.