Three faces of node importance in network epidemiology: Exact results for small graphs

Petter Holme[*]

Institute of Innovative Research, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku Yokohama, Kanagawa 226-8503, Japan

We investigate three aspects of the importance of nodes with respect to susceptible-infectious-removed (SIR) disease dynamics: influence maximization (the expected outbreak size given a set of seed nodes), the effect of vaccination (how much deleting nodes would reduce the expected outbreak size), and sentinel surveillance (how early an outbreak could be detected with sensors at a set of nodes). We calculate the exact expressions of these quantities, as functions of the SIR parameters, for all connected graphs of three to seven nodes. We obtain the smallest graphs where the optimal node sets are not overlapping. We find that (i) node separation is more important than centrality for more than one active node, (ii) vaccination and influence maximization are the most different aspects of importance, and (iii) the three aspects are more similar when the infection rate is low.

I. INTRODUCTION

One of the central questions in theoretical epidemiology [1–3] is how to identify individuals that are important for an infection to spread [4,5]. What "important" means depends on the particular scenario—what kind of disease spreads and what can be done about it. In the literature, three major aspects of importance have been discussed. First, *influence maximization* is aimed at identifying the nodes that, if they are sources of the outbreak, would maximize the expected outbreak size Ω (the number of nodes infected at least once) [6,7]. Second, *vaccination* is aimed at finding the nodes that, if vaccinated (or, in practice, deleted from the network), would reduce the expected outbreak size the most [5]. Third, *sentinel surveillance* is aimed at finding the nodes that are likely to get infected early [8,9]. These three notions of importance do not necessarily yield the same answer as to which node is most important. In this work, we investigate how the ranking of important nodes for these three aspects differs and why (see Fig. 1).

In this paper, we evaluate the three aspects of importance with respect to the susceptible-infectious-removed (SIR) disease-spreading model [1–3,10] on small connected graphs (all connected graphs from three up to seven nodes). The main reason we restrict ourselves to small graphs is that it allows us to use symbolic algebra, and thus exact calculations [11]. In this way we can discover, e.g., the smallest graph where the three aspects of importance disagree about which node is most important; cf. Ref. [12]. We argue that graphs of seven nodes are still large enough to illustrate the effects of distance.

Nevertheless, large networks are important to study. A possible future extension of this work will be to address the relationship between the three importance measures for larger networks. In the related Ref. [13], the difference between influence maximization and vaccination problems on (some rather large) empirical networks is studied. The authors compare the top results of heuristic algorithms to identify influential single nodes, whereas in this paper we will consider the influence of all nodes, and also all sets of two and three nodes. (The terminology of Ref. [13] is a bit different from ours—they call important nodes for vaccination "blockers" and important nodes for influence maximization "spreaders".)

We will proceed by discussing our setup in greater detail: our implementation of the SIR model, how to analyze the three aspects of importance, network centrality measures that we need for our analysis, and our results, including the smallest networks where different nodes come out as most important.

II. PRELIMINARIES

In this section, we provide the background to our analysis. The basis of our analysis is graphs $G(V,E)$ consisting of N nodes V and M links E.

A. Importance

As mentioned earlier, there are three ways to think of importance in theoretical infectious disease epidemiology. Influence maximization was first studied in computer science with viral marketing in mind [6,7]. As was mentioned, a node is important for influence maximization if it is a seed of an infection that could cause a large outbreak. For epidemiological applications, therefore, it might be interesting if one could immunize people against a disease before an outbreak happens. We will simply measure the expected outbreak size $\Omega(S)$ (the expected number of nodes to catch the disease), with S as the set of source nodes, and we will rank the set of nodes according to their Ω.

For vaccination, we will use the average outbreak size from one random seed node to estimate the importance of a node [1,2,14,15]. One could rephrase it as a cost problem [16]. We assume the vaccinees are deleted from the network before the outbreak starts. The node with the smallest Ω is the one that is most important for the vaccination problem.

Sentinel surveillance assumes a response after the outbreak already started (compared to influence maximization and vaccination, where the action affecting the nodes in question is assumed to take place before the outbreak happens). A node is important for sentinel surveillance if it gets infected early so that the health authorities can activate their countermeasures. This is usually determined by the lead time—the expected difference between the time a sentinel node gets infected, or the outbreaks dies out, and the infection time of any node in the graph [8]. We will instead measure the average discovery time time $\tau(i)$ from the beginning of the infection until a node i gets infected or the outbreak dies [9]. The node with the

_____
[*]holme@cns.pi.titech.ac.jp

(a) ○ susceptible ● infectious ● recovered

(b) influence maximization

(c) vaccinization
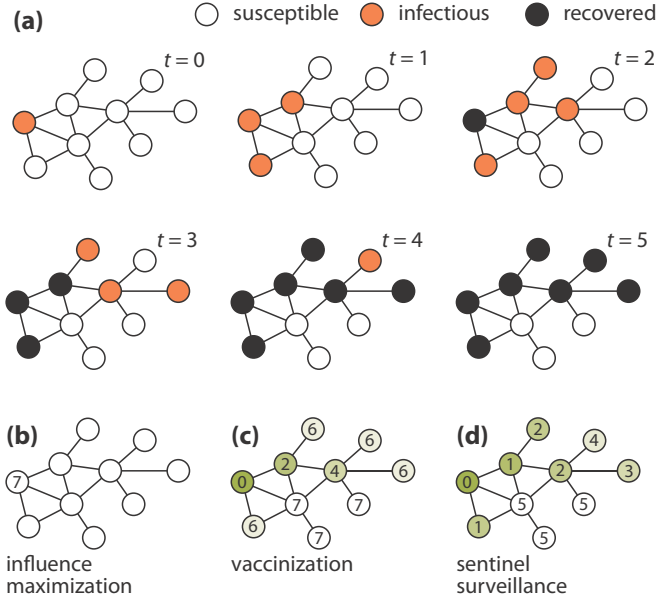
(d) sentinel surveillance

FIG. 1. Illustration of the three different notions of importance we explore in this work. Panel (a) shows an example of an SIR outbreak in a seven-node network. Panels (b)–(d) show how this outbreak influences maximization (a), vaccination (b), and sentinel surveillance, respectively. The idea of influence maximization (b) is that a node is important if the outbreak originating at it is expected to be large. The idea of vaccination (c) is that a node is important if removing it would reduce significantly the average outbreak size. The idea of sentinel surveillance (d) is that a node is important if a sensor on it would detect the outbreak early. The shades of the nodes in (c) and (d) are proportional to their contribution. In a stochastic simulation, one would average the values over many runs and, for (c) and (d), many seeds of the outbreak. In this work, however, rather than running simulations, we calculate the exact expectation values of these quantities.

smallest discovery time is then considered most important for sentinel surveillance. If the purpose of the surveillance is just to discover the outbreak—not to rid the population of the disease as early as possible—one could measure $\tau(i)$ conditioned on the outbreak reaching a sentinel before it dies out. We will briefly discuss such a conditioned τ and refer to it as τ'.

For all of the three problems mentioned above, one can consider sets of nodes rather than individuals. There can be more than one source (for influence maximization), vaccinee, or sentinel. We will, in general, call these sets *active nodes* and denote their number as n. We will try to find the optimal sets of active nodes (and call them *optimal nodes*). Note that this is not the same as ranking the nodes in order of importance and taking the n most important ones—such a "greedy" approach can in many cases fail [7,15].

Note that for vaccination and sentinel surveillance, we use one source node of the infection. This is the standard approach in infectious disease epidemiology simply because most outbreaks are thought to start with one person [3,17].

B. SIR model

We will use the constant infection and recovery-rate version of the SIR model [17]. In this formulation, if a susceptible node

i is connected to an infectious node j, then i becomes infected at a rate β. Infected nodes recover at a rate ν. Without loss of generality, we can set $\nu = 1$ (equivalently, this means we are measuring time in units of $1/\nu$). Let C be a configuration (i.e., a specification of the state—S, I, or R—of every node), M_{SI} is the number of links between S and I nodes, and N_{I} is the number of infected nodes. Then, the rate of events (either infections or recoveries) is $\beta M_{\mathrm{SI}} + N_{\mathrm{I}}$, which gives the expected duration of C as

$$\Delta_t = \frac{1}{\beta M_{\mathrm{SI}} + N_{\mathrm{I}}}. \tag{1}$$

Proceeding in the spirit of the Gillespie algorithm, the probability of the next event being an infection event is $\beta M_{\mathrm{SI}} \Delta_t$, and the probability of a recovery event is $N_{\mathrm{I}} \Delta_t$ [2,18].

C. Exact calculations of Ω and τ

Exactly calculating the outbreak size and time to discovery or extinction is, in principle, straightforward. Consider the change from configuration C into C' by an infection event (changing node i from susceptible to infectious). This can happen in m_i ways, where m_i is the number of links between i and an infectious node. Thus the probability for the transition from C to C' is $\beta m_i \Delta_t$. The probability that the next event will be a recovery event is simply Δ_t. To compute the probability of a chain of events, one simply multiplies these probabilities over all transitions. To compute the expected time for a chain of events, one sums the Δ_t for all configurations of the chain.

We will illustrate the description above with an example. See Fig. 2. The probability of the outbreak chain 7 is (multiply the probabilities of the transitions)

$$p_7 = \frac{\beta}{2\beta + 1} \times \frac{1}{2\beta + 2} \times \frac{\beta}{\beta + 1} = \frac{\beta^2}{2(\beta + 1)^2(2\beta + 1)}. \tag{2}$$

The expected duration of the infection chain is

$$\tau_7 = \frac{1}{2\beta + 1} + \frac{1}{2\beta + 2} + \frac{1}{\beta + 1} = \frac{8\beta + 5}{2(\beta + 1)(2\beta + 1)}, \tag{3}$$

giving a contribution

$$p_7 \tau_7 = \frac{8\beta^3 + 5\beta^2}{4(\beta + 1)^3(\beta + 2)^2} \tag{4}$$

of chain 7 to τ. Then these contributions need to be summed up for all chains, and averaged over all starting configurations. For the example in Fig. 2, this gives

$$\tau = \frac{3\beta^2 + 7\beta + 2}{6\beta^3 + 15\beta^2 + 12\beta + 3}. \tag{5}$$

The expressions of Ω and τ are fractions of polynomials. For the largest networks we study (seven nodes), these polynomials can be of order up to 43 with up to 54-digit integer coefficients.

Calculating Ω for the influence maximization or vaccination problems follows the same path as the τ calculation above. The difference is that instead of multiplying by the expected time of a chain, one would multiply by the number of recovered

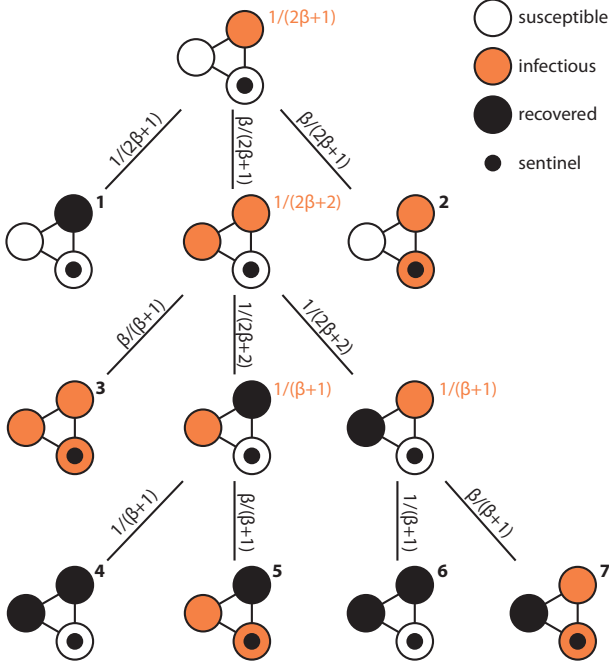FIG. 2. Illustration of how to calculate the expected time τ to discovery or extinction in a small network with one sentinel. The probabilities of transitions are marked on the links (with black text). The expected time to stay in a configuration is next to the configurations (in orange). The final states (where there are either no infectious nodes or a sentinel is infected) are enumerated by boldface numbers.

nodes in that branch. Furthermore, there are no sentinels to stop outbreaks, so trees (like Fig. 2) become larger.

In practice, our approach to analyzing network epidemiological models is time-consuming. The major bottleneck is the polynomial algebra (to be precise, calculating the greatest common divisor needed to reduce the fractions of polynomials to their canonical form). Because of this, we could not handle networks of more than seven nodes. The code was implemented in both Python (with the SymPy library [19]) and C with the FLINT library [20]. It also uses the subgraph isomorphism algorithm VF2 [21] as implemented in the igraph C library [22]. Our code is available at http://github.com/pholme/exact-importance, which also includes code to calculate τ' (mentioned above but not investigated in the paper).

D. Centrality measures

To better understand how the network structure determines what nodes are most important, we measure the average values of static importance predictors. In general, there are many ways to be the central means for a node—is it a node often passed by things traveling over the network, or is it a node for which short paths exist to other nodes? Different rationales give different measures. These are typically positively correlated, but they do not rank the nodes in the exact same way, and thus they can complement each other [23]. We focus on three measures: degree, closeness centrality, and vitality.

Degree centrality is simply the number of neighbors of a node. If a node has twice the neighbors of another, it has twice as many nodes to which to spread an infection. This makes it more important for influence maximization and vaccination. It also has twice as many nodes from which to get the infections, which contributes to its importance for vaccination and sentinel surveillance. On the other hand, degree is not a global quantity—it could happen that the neighbors of a high-degree node are so peripheral that a disease could easily die out there. The simplest way of modifying the degree to become a global measure is to operationalize the idea that a node is central if it is the neighbor of many central nodes. With the simplest possible assumptions, this reasoning leads to *eigenvector centrality*, i.e., the centrality of node i can be estimated as the ith entry of the leading eigenvector of the adjacency matrix [10]. For the small graphs that we consider, however, the eigenvector centrality is so strongly correlated with degree (intuitively so, because "everything is local" in a very small graph) that it makes little sense to include it in the analysis.

Many centrality measures are based on the shortest paths. Perhaps the simplest of these measures is *closeness centrality*—using the idea that a node is central if it is on average close to other nodes [10,23]. This leads to a measure of the centrality of i as the reciprocal distance to all other nodes in the network:

$$c(i) = \frac{N-1}{\sum_{j \neq i} d(i,j)}. \tag{6}$$

The main problem, in general, with closeness centrality may be that it is ill-defined on disconnected graphs. In our work, however, we consider only connected graphs.

We chose the third centrality measure—*vitality*—with the vaccination problem in mind. Vitality is, in general, a class of measures that estimate node centrality based on its impact on the network if it is deleted [23]. In our work, we let vitality denote the particular metric

$$v(i) = \frac{s(G) - 1}{s(G \setminus \{i\})}, \tag{7}$$

where $s(G)$ is the number of nodes in the largest component of G. This measure is thus in the interval $[1, N-1]$, and it increases with i's ability, if removed, to fragment the network. Since vaccination is, in practice, like removing nodes from the network, we expect v to identify important nodes for β close to 1. For large graphs, we expect v to be very close to 1, so we only recommend it for small graphs such as the ones we used here. Another popular centrality measure—*betweenness centrality* (roughly how many shortest paths there are in the network that passes a node) [10]—is very strongly correlated with vitality for our set of small graphs, and it is thus omitted from the analysis.

E. Small distinct graphs

In our work, we systematically evaluate small distinct (nonisomorphic) connected graphs. We use all such graphs with $3 \leqslant N \leqslant 7$. There are two such graphs with $N = 3$, six with $N = 4$, 20 with $N = 5$, 112 with $N = 6$, and 853 with $N = 7$. To generate these, we use the program GENG [24].
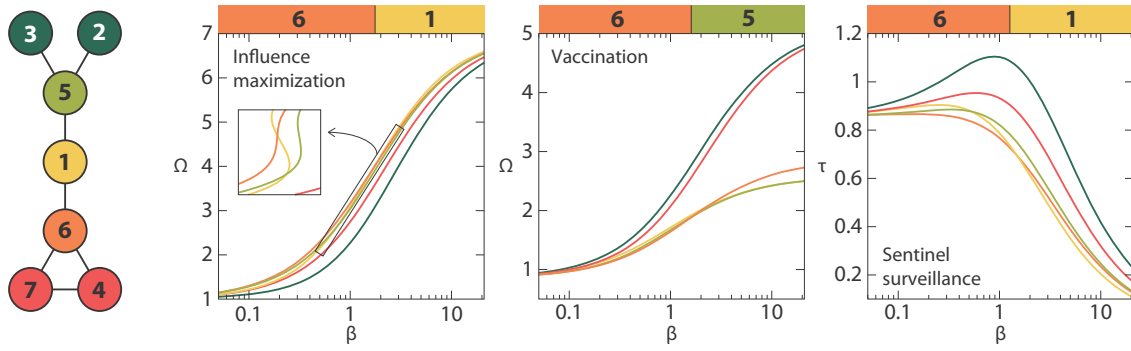
FIG. 3. A small graph with special properties. It is the smallest graph where the three aspects of importance all have distinct nodes as the most important. For β in the interval $[(1 + \sqrt{5})/2, (3 + \sqrt{17})/4] \approx (1.62, 1.78)$, 6 is the most important node for influence maximization, 5 for vaccination, and 1 for sentinel surveillance. The curves are color-coded according to the graph to the left (with one curve and color per automorphic equivalence class). The inset in panel (a) shows a zoomed and rotated view of some of the crossings of the curves (one of them being the point where node 1 takes over from node 6 as the most important).

III. RESULTS

In our analysis, we will focus on when and why the three cases of node importance rank nodes differently. We will start with some extreme examples, and continue with general properties of all small graphs.

A. Special cases

Inspired by Ref. [12], we will start with a special example (Fig. 3). This is the smallest graph where the most important single node ($n = 1$) is different for influence maximization, vaccination, and sentinel surveillance. For $[(1 + \sqrt{5})/2, (3 + \sqrt{17})/4] \approx (1.62, 1.78)$, node 6 is the most important node for influence maximization, 5 is most important for vaccination, and 1 is most important for sentinel surveillance. For small β values, 6 is most important for all three aspects of importance. In this region, the outbreaks die out easily. The fact that 5 and 6 have a larger degree than the others is, of course, helpful for an outbreak to take hold in the population. Node 6 is slightly more important as a seed node since the extra link in its neighborhood helps the outbreak to persist longer [there are the (6,7,4) and (6,4,7) infection paths that, although unlikely, do not exist for diseases starting at 5]. This reasoning also

explains why 6 is most important for vaccination. For sentinel surveillance and for low enough β, the outbreak would typically end by the outbreak becoming extinct rather than hitting a sentinel. Thus, for low β, when an outbreak has the highest chance of surviving if it starts at 6, then putting in a sentinel is good because an outbreak is either instantly discovered or will likely soon be extinct. With a conditional discovery time τ', the curves are strictly decreasing (since the early die-off is omitted), so 1 is the most important node for all β.

For larger β, node 1 becomes, relatively speaking, more important for influence maximization and sentinel surveillance. This is the most central node in aspects other than degree. For vaccination, however, node 5 is most important as it fragments the network most [the vitality is the same for both nodes $v(5) = v(1) = 2$, but the size of the second biggest component is larger if 1 is deleted]. So since 1 becomes more important than 6 at a larger β value for influence maximization compared with sentinel surveillance, there is an interval of beta where the network of Fig. 3 has three distinct most important nodes for the three aspects of importance that we investigate.

For two active nodes ($n = 2$), the smallest network with no overlap between the optimal node sets is actually smaller than for $n = 1$. This network, displayed in Fig. 4, has six nodes and eight links. Note that $N = 6$ is the smallest number
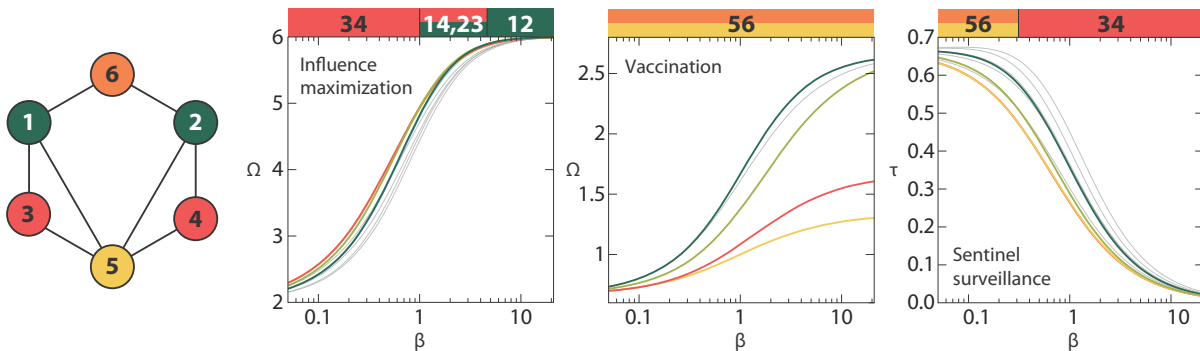

FIG. 4. A special graph for two active nodes. To the left, the colors highlight the automorphic equivalence classes. For $\beta > 4.62$, {1,2} are the most important source nodes (influence maximization), {5,6} are the most important for vaccination, and finally {3,4} are the most important for sentinel surveillance. The color bars indicate the most important node-pairs as a function of β, and the curves are showing the β dependence of the objective quantities Ω and τ. The four node combinations that are optimal at some β value for some scenario are highlighted; the others are gray.
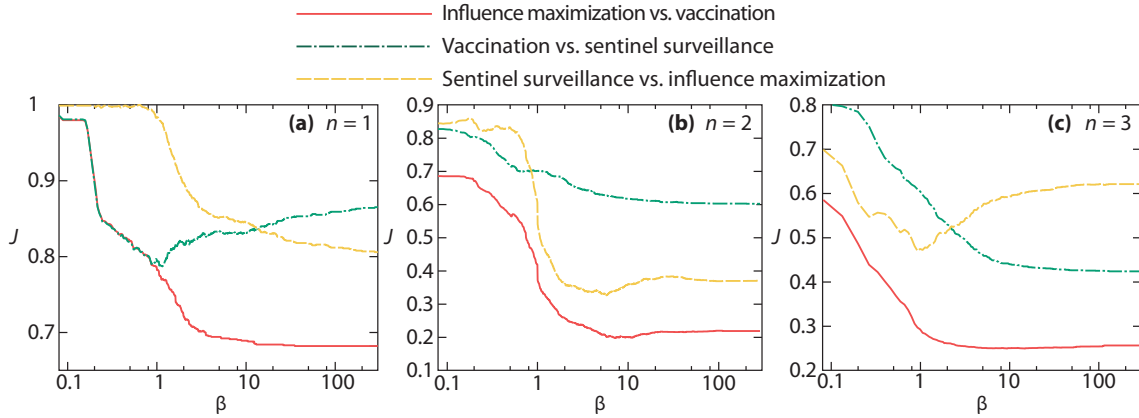
FIG. 5. The average pairwise Jaccard overlap between the sets of the most influential nodes for the three different aspects of importance. Panels (a), (b), and (c) show the overlap for one, two, and three active nodes, respectively. For β larger than shown, the curves are constant.

of nodes to make three distinct sets of two nodes, so in that sense the $n = 2$ example seems more extreme than the $n = 1$ counterpart, Fig. 3.

For large β values, 1 and 2 are the most important nodes for influence maximization, 5 and 6 are most important for vaccination, and 3 and 4 are most important for sentinel surveillance. 5 and 6 are the nodes that, if deleted, break the network into the smallest components, which explains why they are most important for vaccination (at least for large β). In addition to 5 and 6, 1 and 2 are the only pair of nodes whose neighborhoods contain all other nodes. Nodes 1 and 2 both have degree 3, as opposed to 5 and 6, which have degrees 4 and 2, respectively. It is not clear whether that makes 1 and 2 better than 5 and 6 for influence maximization, or why. Similarly, it is hard to understand why 3 and 4 are the best nodes for sentinel surveillance. The neighborhoods of these nodes do not even contain the entire graph.

We can see that the optimal sets of nodes in Fig. 4 do not have links within themselves. This seems natural for most networks and all three notions of importance. This means that as n grows, the distance between the optimal nodes will be larger than 1. This is an observation we will make more quantitatively in the next section. Another such observation is that for small β, the optimal nodes for the three importance aspects are overlapping. In this parameter region, most outbreaks die out before they reach a sentinel. If the outbreak starts at a high-degree node in a highly connected neighborhood, there is a larger chance for it to survive. For all three importance aspects, it is important to have active nodes where an outbreak would be likely to survive. Still, as evident from Fig. 4, there are examples where the optimal nodes are not overlapping.

B. β dependence

We will now move to a more statistical evaluation of all graphs with 3–7 nodes. We will present average quantities over all these graphs as functions of β. Other summary statistics, including grouping the graphs according to size, give the same conclusions.

Let $u_i^{a,b}$ be the optimal sets for a given network, β, and importance classes a and b. The first quantity we look at is the

pairwise overlap of sets of optimal nodes as measured by the Jaccard overlap,

$$J(a,b) = \frac{|U_a \cap U_b|}{|U_a \cup U_b|}, \tag{8}$$

where

$$U_{a,b} = \bigcup_i u_i^{a,b}. \tag{9}$$

For example, in Fig. 4 at $\beta = 2$, we have

$$u_1^a = \{1,4\}, \tag{10a}$$

$$u_2^a = \{2,3\}, \tag{10b}$$

$$u_1^b = \{3,4\}, \tag{10c}$$

where a is influence maximization and b represents sentinel surveillance, giving

$$J(a,b) = \frac{|\{1,2,3,4\} \cap \{3,4\}|}{|\{1,2,3,4\} \cup \{3,4\}|} = \frac{1}{2}. \tag{11}$$

As seen in Fig. 5, for $n = 1$, the overlap between the optimal nodes for vaccination and sentinel surveillance has a minimum as a function of β. The same is true for sentinel surveillance versus influence maximization when $n = 3$. It is hard to say why, beyond that, for individual graphs the $J(a,b,\beta)$ curves can of course be nonmonotonous as different aspects of the graph structure determine the role of the nodes. We note that (for a different spreading model and much larger networks), Ref. [13] finds the Jaccard similarity between influence maximization and vaccination to have a minimum as a function of β.

C. Structural predictors

Next, we investigate the structural properties of the most influential nodes and how they depend on β. In Fig. 6, we plot the degree, closeness centrality, and vitality as a function of β for all aspects of importance and $n \in \{1,2,3\}$. We start by examining the case $n = 1$; see Figs. 6(a), 6(b), and 6(c). The first thing to notice is the general impression that centralities of the optimal nodes decrease with β. The only case with an opposite trend is vitality [Fig. 6(c)], where the curves are
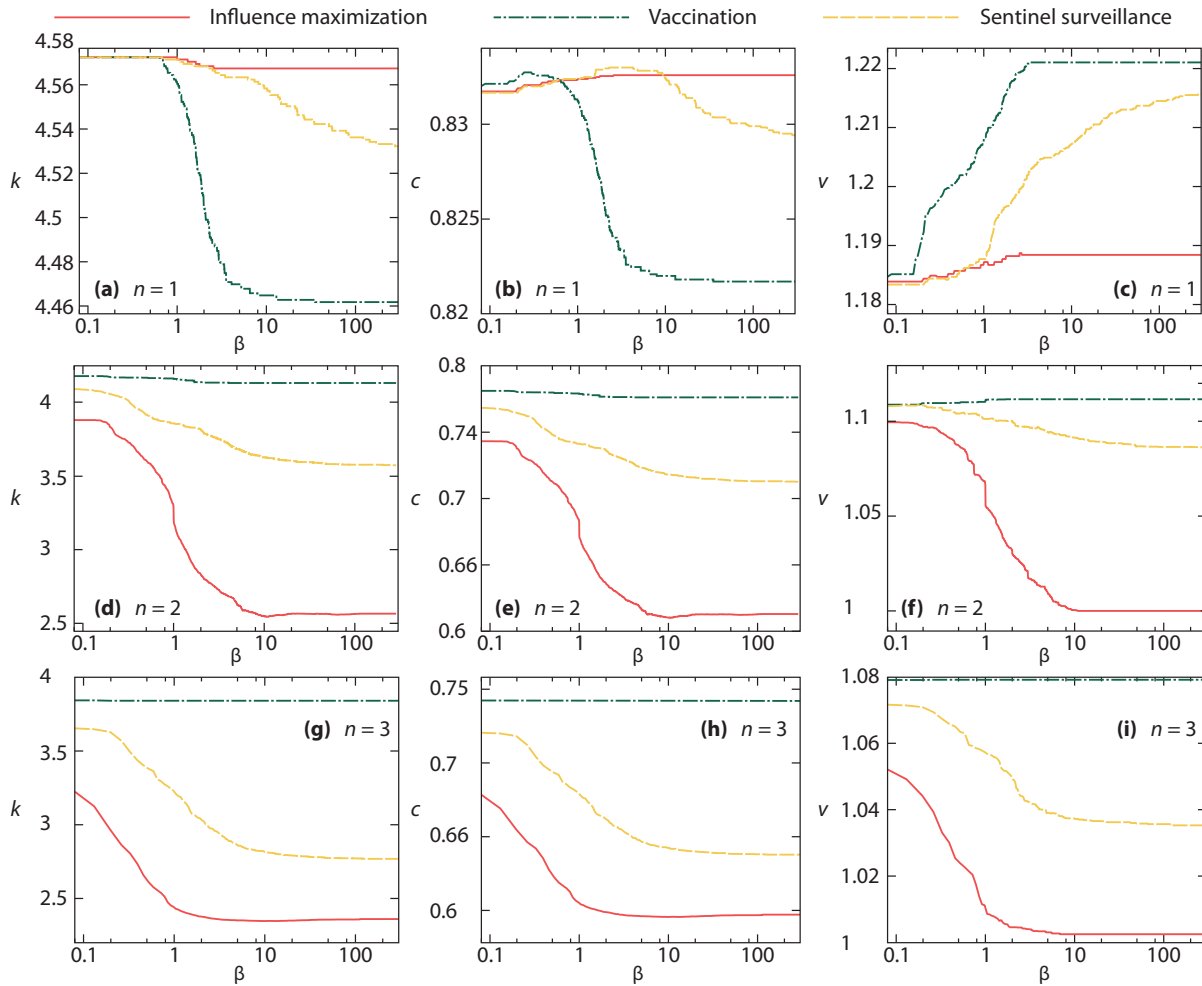
FIG. 6. Average centrality values for the different aspects of centrality and the different numbers of active nodes. Panels (a), (d), and (g) show the average degree; panels (b), (e), and (h) show closeness centrality; and panels (c), (f) and (i) show curves for vitality. Panels (a), (b), and (c) show results for $n = 1$; (d), (e), and (f) for $n = 2$; and (g), (h), and (i) for $n = 3$.

increasing monotonically. If we first focus on the case with one active node, this could be understood as the ability of nodes to (if removed) fragment the network. This ability is captured by vitality and becomes more important as β increases.

Continuing the analysis for $n = 1$, when β is low the most important thing is for the outbreak to persist in the population. If an active node has a high degree, it is likely to be the source of a large outbreak, meaning it is important for influence maximization (which was also concluded by Ref. [13]). If a high-degree node is deleted, it would remove many links that could spread the disease and thus be important for vaccination [25]. It would also be important not to put a sentinel on a low-degree node for sentinel surveillance and low β as diseases reaching low-degree nodes would likely die out. So panels Figs. 6(a) and 6(c) can be understood as a shift from nodes of high degree to nodes of high vitality. Closeness centrality—seen in Fig. 6(b)—is harder to explain. Values of c increase with β for influence maximization but decrease for vaccination. One way of understanding this is from the observation that vitality is most important for vaccination [as evident from Fig. 6(c)], and degree is most important for influence maximization [as seen in Fig. 6(a)]. The results of

Fig. 6(b) then suggest that the high vitality nodes optimizing the solution of the vaccination problem have a lower closeness centrality. Indeed, for many of the graphs we study, the highest vitality node has many degree-1 neighbors—cf. node 5 in Fig. 3—which does not necessarily contribute to the closeness centrality. For influence maximization, it seems that the optimal nodes are central in the closeness sense—the closer to average the seed node is to the rest of the network, the higher is the chance for the outbreak to reach the entire network.

For $n = 2$ and 3, the picture is somewhat different than for $n = 1$. In these cases, all centrality measures are decreasing monotonically. The order of importance measures are all the same, with vaccination having the largest values, and influence maximization the smallest. It is no longer the case for vaccination that the optimizing nodes have high vitality and low closeness centrality (as it was for $n = 1$). Indeed, for the vaccination case, the optimal nodes are usually independent of β, which is why the curves for vaccination in Figs. 3(d)–3(i) are almost straight. Naively, one would think that some centrality measure needs to increase with β. However, as we will argue further below, the optimal nodes would usually not be close to each other. One could think of each node being
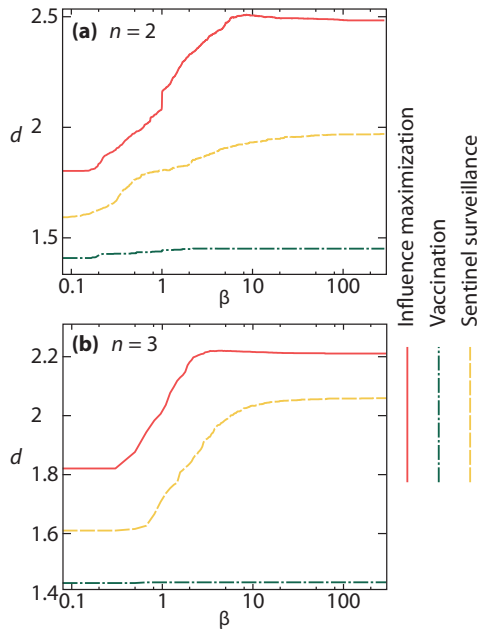
FIG. 7. The average distance between optimal nodes when the number of active nodes is $n = 2$ (a) and 3 (b).

responsible for (and centrally situated within) a region of the network, and that that tendency is so strong that it overrides all simple centrality measures. On the other hand, there are group centrality measures that could perhaps increase with β [26] (that could be a theme for another paper).

D. Distance between optimal nodes

The fact that all the curves of Figs. 6(d)–6(i) are nonincreasing could be explained by the fact that the separation of the optimal nodes increases with β. In Fig. 7, we try to make this argument more quantitative by measuring the average (shortest path) distance d between the optimal nodes. In the limit of small β, these values come rather close to its minimum of 1, but as β increases, so does d. Essentially, the pattern from Fig. 7 is the reverse of Figs. 6(d)–6(i)—the vaccination curve is almost constant, sentinel surveillance increases moderately, but influence maximization increases much more.

A larger separation gives the sentinels the ability on average to be closer to outbreaks anywhere in the network, while for influence maximization a larger separation means that there are more susceptible-infectious links (fewer infectious-infectious links) in the incipient outbreak. For vaccination there is no such positive effect of a larger separation that we can think

of, which is a part of the explanation as to why the optimal sets are relatively independent of β for $n > 1$. The rest of the explanation, i.e., why the trends for $n = 1$ are so much weaker when $n > 1$, is not clear to us, and it is something we will investigate further in the future.

IV. DISCUSSION

We investigated the average properties of the optimal nodes for all our graphs. We found that the overlap between the optimal nodes of the different importance aspects are largest for small β. In the small-β region, a high degree seems most important for all importance aspects. For larger β nodes, it becomes more important for them to be positioned such that they would fragment the network if they were removed, particularly for the vaccination problem (slightly less for the sentinel surveillance problem, and much less for influence maximization). On the other hand, when the number of active nodes increases, it becomes important for the nodes to be spread out—the average distance between them increases. This effect is large for influence maximization, intermediate for sentinel surveillance, and very small for vaccination. The small effect for vaccination can be understood since all that matters is to fragment the network, and for that purpose the vaccine does not necessarily have to be distant.

Most of the behavior discussed above seems quite natural. For small β, the dominant aspect of the dynamics is how fast an outbreak will die out. For large β, the outbreak will almost certainly reach all nodes. For vaccination and sentinel surveillance, this leads to a question of deleting nodes that would break the network into the smallest components. (In the former case, this is trivial since the size of the outbreak is almost surely the size of the connected component to which the seed node belongs. In the latter, we conclude this from the monotonically increasing vitality).

As an extension, it would be interesting to confirm this work in larger networks using stochastic simulations. This would not allow for the discoveries of special graphs such as those in Figs. 3 and 4, but it could reinforce the connection between the different notions of centrality. We believe that many of our conclusions hold for larger networks, an indication being that our results are consistent with the results of Ref. [13] (comparing the vaccination and influence maximization for $n = 1$ in large empirical networks).

[1] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of Epidemics on Networks* (Springer, Heidelberg, 2017).

[2] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).

[3] H. W. Hethcote, SIAM Rev. **42**, 599 (2000).

[4] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, Phys. Rep. **650**, 1 (2016).

[5] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhao, Phys. Rep. **664**, 1 (2016).

[6] J. Sun and J. Tang, *A Survey of Models and Algorithms for Social Influence Analysis* (Springer, Boston, 2011), pp. 177–214.

[7] D. Kempe, J. Kleinberg, and É. Tardos, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining (ACM, Washington, DC, 2003), pp. 137–146.

[8] N. A. Christakis and J. H. Fowler, PLoS One **5**, e12948 (2010).

[9] P. Bajardi, A. Barrat, L. Savini, and V. Colizza, J. R. Soc. Interface **9**, 2814 (2012).

[10] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, UK, 2010).

[11] N. Masuda, J. Theor. Biol. **258**, 323 (2009).

[12] U. Brandes and J. Hildenbrand, Network Sci. **2**, 416 (2014).

[13] F. Radicchi and C. Castellano, Phys. Rev. E **95**, 012318 (2017).

[14] T. Britton, S. Janson, and A. Martin-Löf, Adv. Appl. Probab. **39**, 922 (2007).

[15] J. Gu, S. Lee, J. Saramäki, and P. Holme, Europhys. Lett. **118**, 68002 (2017).

[16] P. Holme and N. Litvak, PLOS Comp. Biol. **13**, e1005696 (2017).

[17] P. Holme, J. Logist. Eng. Univ. **30**, 1 (2014).

[18] R. Huerta and L. S. Tsimring, Phys. Rev. E **66**, 056115 (2002).

[19] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh *et al.*, PeerJ Comput. Sci. **3**, e103 (2017).

[20] W. B. Hart, in *Proceedings of the Third International Congress on Mathematical Software* (Springer-Verlag, Berlin, 2010), ICMS'10, pp. 88–91.

[21] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento, in *3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition* (Springer, Berlin, 2001), pp. 149–159.

[22] G. Csárdi and T. Nepusz, InterJournal Complex Systems 1695 (2006).

[23] D. Koschützki, K. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski, in *Network Analysis: Methodological Foundations* (Springer, Berlin, 2005), pp. 16–61.

[24] B. D. McKay and A. Piperno, J. Symbol. Computat. **60**, 94 (2014).

[25] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002).

[26] M. G. Everett and S. P. Borgatti, J. Math. Sociol. **23**, 181 (1999).