Analyzing network reliability using structural motifs

Yasamin Khorramzadeh, Mina Youssef, Stephen Eubank, and Shahir Mowlaei

Analyzing Network Reliability Using Structural Motifs

Yasamin Khorramzadeh*
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia 24061, USA and
Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA

Mina Youssef†
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia 24061, USA

Stephen Eubank‡
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia 24061, USA
Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA and
Department of Population Health Sciences, Virginia Tech, Blacksburg, Virginia 24061, USA

Shahir Mowlaei§
Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA and
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia 24061, USA

This paper uses the reliability polynomial, introduced by Moore and Shannon in 1956, to analyze the effect of network structure on diffusive dynamics such as the spread of infectious disease. We exhibit a representation for the reliability polynomial in terms of what we call *structural motifs* that is well suited for reasoning about the effect of a network's structural properties on diffusion across the network. We illustrate by deriving several general results relating graph structure to dynamical phenomena.

I. INTRODUCTION

Characterizing networks in a way that is directly relevant to diffusion phenomena on the network is important, but difficult. We argue that the *Network Reliability Polynomial* introduced by Shannon and Moore [1] is a characterization that folds together static measures like degree, modularity, and measures of centrality into precisely the combinations that are most relevant to the dynamics. [2] Conversely, knowledge of reliability can be used to infer structure, in the sense of network tomography. [3] Furthermore, reliability is a useful concept for reasoning more generally about the consequences of structural changes. Colbourn [4] gives a comprehensive introduction to network reliability, while Youssef [5] provides a brief derivation of the form of the reliability polynomial.

A network's *reliability* is the probability that it continues to function after sustaining damage to its component edges and/or vertices. Reliability depends on a parameterized *damage model* $D(\vec{x})$, which specifies the probability of damaging a particular set of components, and a parameterized property $\mathcal{P}(\vec{\alpha})$, which specifies what it means

for the network to function. Canonical problems in diffusion over fixed networks can be cast in terms of reliability. For example, a damage model $D(x)$ under which edges (respectively, vertices) fail independently with probability $1 - x$ represents bond (respectively, site) percolation. The choice of $\mathcal{P}$ then selects the dynamical phenomenon of interest. For example, the property "contains a connected component including at least a fraction α of the vertices" is appropriate for studying the existence of a giant component. For ease of notation, we often express $\mathcal{P}$ in terms of a corresponding *reliability rule* $r_{\mathcal{P}}$, a binary function indicating whether property $\mathcal{P}$ holds for a given graph. We say that the rule *accepts* a graph g if and only if $r(g) = 1$, that is g has the property $\mathcal{P}$.

Here we introduce four different reliability rules. The first three are the most commonly used rules, followed by the last one that could be of use to study percolation problems. To make it easier to understand, we refer to real world problems like designing reliable communication networks and epidemiology which benefit from each of these rules:

1. *Two-Terminal:* a graph is accepted if it contains at least one directed path from a distinguished vertex S (the *source*) to another distinguished vertex T (the *terminus*). Reliability under this rule is the probability that the specified source can send a message to the specified terminus in a damaged network. [6]

———————
* yasi@vbi.vt.edu
† myoussef@vbi.vt.edu
‡ seubank@vbi.vt.edu
§ shahir@vbi.vt.edu

2. *K-Terminal:* a graph is accepted if every vertex is in a connected component that includes at least one of K specified vertices. For example, consider a set of K nodes as service centers. Then reliability under this rule is the probability that every operational site in a damaged network is connected to at least one service center. [7]

3. *All-Terminal:* a graph is accepted if it consists of a single connected component. The reliability under this rule, also known as *system reliability* is the probability that every pair of nodes in a damaged network can communicate with each other, or alternatively, that any vertex in a damaged network can broadcast to all the other vertices. [8]

4. *EAR-α:* To understand this rule we discuss the application of bond percolation for the study of the spread of infectious diseases on networks. Such a bond percolation disease model was explained by [9] . The probability that an edge does not fail represents the *transmissibility*, i.e., the conditional probability of transmitting infection from one person to another, conditioned on the source being infectious and the destination being susceptible. One of the most important properties of disease dynamics is the *attack rate*, defined as the fraction of the population infected in an outbreak. Most models of infectious disease exhibit a sharp transition in the attack rate at a critical value of transmissibility. Indeed, this is a percolation phase transition. Using reliability we can find the relationship between critical transmissibility and all structural information of a network contained in its edge list. In this case the reliability is the probability that the expected attack rate for an outbreak seeded in a single person chosen uniformly at random from the population is at least α. Hence, we call this rule *EAR-α*.

These rules are all *coherent*. That is, any graph formed by adding an edge to an accepted graph is also accepted. In a companion to this paper [5], we have shown how the concept of network reliability together with an efficient, scalable estimation scheme can shed light on complicated dynamical trade-offs between local structural properties such as assortativity-by-degree and the number of triangles. Here we introduce a different representation of the reliability polynomial that highlights the role certain network structures play in dynamical phenomena. We show how coefficients of the reliability polynomial can be interpreted in terms of topological motifs in the network and their overlaps. Conversely, we illustrate how knowledge of these motifs and their overlaps can be used to infer important constraints on the dynamics of diffusion processes on the network. The representation in [5] is well-suited for computational analysis of networks with up to 10^8 edges, but is analytically tractable only for small networks; the representation presented in the current work is analytically tractable, but computationally

feasible only for small networks because of its combinatorial complexity. Thus the results of this paper exactly complement those of the previous paper.

II. RELIABILITY POLYNOMIALS

We use the common notation of $G(V, E)$ for a graph with V vertices and E edges. The graph may be directed or undirected, and it is possible to have multiple edges between two vertices.

The vertices and edges may be labeled. The general case of directed edges and labeled vertices and edges is powerful enough to represent extremely complex networks such as interdependent infrastructure networks. Here, without loss of generality, we restrict ourselves to homogeneous networks represented as undirected, unlabeled graphs.

A. Definition and a common representation

The *reliability* $R(G, \mathcal{P}(\vec{\alpha}), D(\vec{x}))$ of a network G with respect to the property $\mathcal{P}$ under damage model D is the probability that a subgraph of G chosen with probability given by D has property $\mathcal{P}$, the binary rule $r_\mathcal{P}$ examines whether subgraph has property $\mathcal{P}$ or not, if it has then:

$$r_{\mathcal{P}(\vec{\alpha})}(g) = 1,$$

otherwise

$$r_{\mathcal{P}(\vec{\alpha})}(g) = 0.$$

i.e. , reliability can be interpreted as the expected value of the reliability rule operator over different subgraphs of G. We will explicitly include the dependence on the network G and the property $\mathcal{P}$ in notation such as $R(G, \mathcal{P}(\vec{\alpha}), D(\vec{x}))$ only when we wish to distinguish the reliability of two different graphs or two different properties. Moreover, we will not include the damage model itself, but only the values of its parameters $\vec{x}$. Finally, for a homogeneous network in which all edges (or all vertices) fail with the same probability, $\vec{x}$ is a scalar, x. Thus we can write the reliability simply as $R(x)$.

$$R(x) \equiv \sum_{g \subseteq G} r_{\mathcal{P}(\vec{\alpha})}(g) p_{D(\vec{x})}(g) \tag{1}$$

For the independent-edge damage model, in which the probability of selecting a subgraph $g \subseteq G$ depends only on the number of its edges, $|g| = k$, and is $x^k (1-x)^{E-k}$, we have:

$$R(x) = \sum_{g \subseteq G} r_{\mathcal{P}(\vec{\alpha})}(g) x^k (1-x)^{E-k} \tag{2}$$

We can re-write 2 in terms of sum over subgraphs of different sizes, introduced by Alon et al.[10] as *motifs*:

$$R(x) = \sum_{k=0}^{E} R_k x^k (1-x)^{E-k}. \tag{3}$$

R_k is the number of subgraphs of G with exactly k edges that are accepted by the rule. For computational convenience, we often prefer to work with normalized coefficients

$$P_k \equiv R_k / \binom{E}{k}. \tag{4}$$

P_k is the *fraction* of subgraphs of kG with exactly k edges that are accepted by the rule. $P_k \le P_{k+1}$ for a coherent rule. P_k can be estimated efficiently via monte Carlo simulation. [5]

Substituting R_k coefficients in Equation 2 with P_ks from 4, we can see the resemblance to binomial distribution, since $P_k \le 1$ it is clear that

$$R(x) = \sum_{k=0}^{E} \binom{E}{k} P_k x^k (1-x)^{E-k}$$
$$\le \sum_{k=0}^{E} \binom{E}{k} x^k (1-x)^{E-k} \le 1 \tag{5}$$

Therefore, $R(x)\colon [0,1] \to [0,1]$ is a continuous polynomial with only a finite, but possibly large, number of coefficients $R_k, k \in \{0, \ldots, E\}$. That is, the reliability can be thought of as a vector in an $E+1$-dimensional vector space, and the R_k's as the components of the vector in the basis $x^k(1-x)^{E-k}$. There are, of course, many other bases we could choose for this space. An orthogonal basis, such as the first $E+1$ Legendre polynomials, might have useful estimation properties. Here we use another non-orthogonal basis – the functions x^k i.e. the Taylor series expansion – because of its simplicity and its attractive interpretation. There is a unique mapping from coefficients in one basis to those in the other, which can be derived by expanding the factor $(1-x)^{E-k}$ in Equation 3:

$$R(x) = \sum_{k=0}^{E} R_k x^k \sum_{m=0}^{E-k} \binom{E-k}{m}(-1)^m x^m$$
$$= \sum_{k=0}^{E} R_k \sum_{l=k}^{E} \binom{E-k}{l-k}(-1)^{l-k} x^l$$
$$= \sum_{l=0}^{E} (-1)^l x^l \sum_{k=0}^{l} (-1)^k R_k \binom{E-k}{l-k}$$
$$= \sum_{l=0}^{E} N_l x^l \tag{6}$$

where

$$N_l \equiv (-1)^l \sum_{k=0}^{l} (-1)^k \binom{E-k}{l-k} R_k. \tag{7}$$

The N_l coefficients are signed integers. In section II C we will explain how we can interpret these coefficients.

B. Structural motifs

We can express the reliability polynomial in terms of overlaps among certain distinguished subgraphs. These subgraphs are the $\mathcal{P}$-minimal subgraphs of G. A graph g is $\mathcal{P}$-minimal if and only if:

1. g has property $\mathcal{P}$; and

2. there is no proper subgraph $g' \subset g$ that has property $\mathcal{P}$.

Obviously, whether a graph is $\mathcal{P}$-minimal or not depends on the property $\mathcal{P}$, or equivalently here, the reliability rule. For example, $\mathcal{P}$-minimal graphs under a Two-Terminal rule are paths from S to T with no extraneous edges, i.e., no loops or dead ends; under the All-Terminal rule, they are spanning trees. In general, the rule selects a distinctive topological pattern (e.g. path, spanning tree) that may occur many times in a given graph, i.e., if we consider all subgraphs of a distinctive pattern to be the *motif*s introduced by Alon et al.[10] then reliability rule selects a subset of these motifs that has the property $\mathcal{P}$. We refer to subsets generated by a particular rule as *structural* motifs because, as we will demonstrate, they are the structural elements of the network that completely determine the occurrence of dynamical phenomena of interest, as specified by $\mathcal{P}$. One advantage of using this representation is that the contribution of the structural motifs to the reliability is known exactly for all $\mathcal{P}$ and all D, as is shown below.

To describe this, we apply two terminal reliability rule with *source* and *terminus* nodes as its parameters on a small toy network, depicted in Figure 1. For this rule structural motifs are simple paths (including but not limited to shortest paths) connecting these two nodes. The structural motifs and their overlap can be seen in Figure 1.

In another example, we consider the two-terminal reliability rule on a two dimensional grid with 4 nodes in each dimension. For a given source node we look at two different terminus nodes to illustrate the dependence of structural motifs on parameters of the reliability rule. Several structural motifs of different sizes for each set of selected sources and termini are shown in Figure 2.

C. Contribution of structural motifs to $R(x)$

The sizes of each motif and their unions completely determine a network's reliability. We demonstrate this using an Inclusion-Exclusion argument motivated by a series of straightforward examples. A detailed proof and more examples are provided in Appendices A and B.
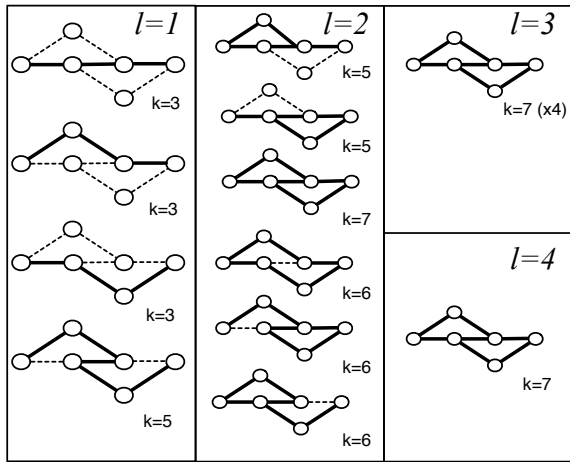
FIG. 1. Example calculation of two-terminal reliability in the toy network in the left panel. The second panel from the left shows the four motifs; the third shows all unions of two motifs. There are $\binom{4}{3} = 4$ unions of 3 motifs, all of which give the entire graph as shown on the top of the right panel; there is one union of all motifs, which is also the entire graph. The number next to each union of motifs gives its size.
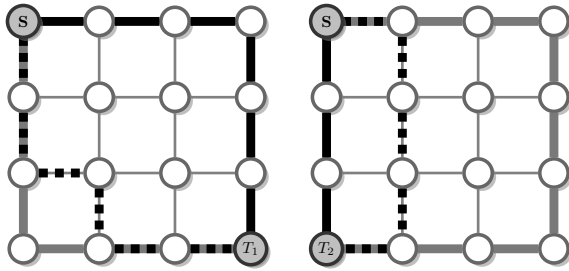

FIG. 2. Structural motif for the two-terminal reliability rule are shown for same source node S and two different target nodes T_1 and T_2. We can see that structural motifs of the same network varies with the parameters of property $\mathcal{P}$. In the left grid three structural motifs of sizes 6 are shown: note that black dashed path and gray line have all but two edges in common. For the grid on the right, black normal, black dashed and gray paths represent motifs of sizes 3, 5, 9 respectively. The last two have two edges in common.

1. Example 1: A single structural motif

Suppose the network G contains only one structural motif and that it is a set of k_0 edges. For example, for Two-Terminal reliability, suppose that there is exactly one path between S and T, and that it has length k_0. Then the motif will occur exactly once among all subgraphs of size k_0. For $k > k_0$, we must "use up" k_0 edges to build the structural motif. This leaves $E - k_0$ other edges from which, because the rule is assumed to be coherent, any set of $k - k_0$ produces an acceptable subgraph

of size k. Hence for this case

$$R_k = \begin{cases} 0 & k < k_0 \\ \binom{E-k_0}{k-k_0} & k \geq k_0 \end{cases} \tag{8}$$

2. Example 4: N disjoint structural motifs

Suppose the graph has exactly N structural motifs, that all have k_0 edges, and that the N edge sets are disjoint. Arguing as above, with the convention that $\binom{a}{b} = 0$ $\forall b < 0$, gives:

$$R_k = \sum_{i=1}^{N} (-1)^{i+1} \binom{N}{i} \binom{E - ik_0}{k - ik_0} \tag{9}$$

3. Example 5: Two overlapping structural motifs

Suppose the graph has exactly two structural motifs, that both have k_0 edges, and that the number of edges in the union of the two is $k_0 + \Delta$. Arguing as in Example 1, we get a similar result, with $2k_0$ replaced by $k_0 + \Delta$:

$$R_k = \begin{cases} 0 & k < k_0 \\ 2\binom{E-k_0}{k-k_0} & k_0 \leq k < k_0 + \Delta \\ 2\binom{E-k_0}{k-k_0} - \binom{E-k_0-\Delta}{k-k_0-\Delta} & k_0 + \Delta \leq k \end{cases} \tag{10}$$

4. The general case

Suppose the graph has exactly N structural motifs. As above, its reliability polynomial will be determined by the size of each structural motif and the overlaps among them. Define $N_k^{(l)}$ as the number of combinations of l structural motifs whose union contains exactly k edges. Also, define

$$N_k \equiv \sum_{l=1}^{N} (-1)^{l+1} N_k^{(l)}. \tag{11}$$

Then arguing as above gives

$$R_k = \sum_{k'=0}^{k} N_{k'} \binom{E - k'}{k - k'}. \tag{12}$$

In Appendix C, we present constraints on N_k. Using Equation 11 and 12 we can determine the reliability coefficients for the two-terminal reliability rule, by analyzing the four motifs as shown in Figure 1. Results of this computation is shown in Table I.

For two dimensional grid, we study only the left case of two cases in Figure 2. For two terminal reliability from S to T there exist 184 structural motifs of sizes 6 to 14, thus it is not trivial to draw all structural motifs and their overlap. We provide computation for N_k and R_k for $k \leq 10$ in Table II.

TABLE I. By inspection, we have the values for $N_k^{(l)}$ given in the table on the right, and the values of N_k and R_k as given by Equations 11 and 12.

l	k	$N_k^{(l)}$	N_k	R_k
1	3	3	3	3
	4		0	12
1	5	1		
2	5	2	-1	17
2	6	3	-3	7
2	7	1		
3	7	4		
4	7	1	2	1

TABLE II. Number of structural motifs and their overlap for the two terminal reliability rule on the left grid in Figure 2.

k	N_k^1	N_k^2	N_k^3	N_k	R_k
5	0	0	0	0	0
6	20	0	0	20	20
7	0	0	0	0	360
8	36	-30	0	6	3066
9	0	-84	0	-84	16332
10	48	-146	144	10	60670

D. Reliability in terms of structural motifs

Given the rather complicated relationship between R_k and N_k in Equation 12, it is somewhat surprising that $R(x)$ can be expressed very simply in terms of N_k. Consider the contribution of a single structural motif of size k_0 to $R(x)$. Using Equation 12, $R_k = \binom{E-k_0}{k-k_0}$. This set of coefficients determines $R(x)$

$$
\begin{aligned}
R(x) &\equiv \sum_{k=0}^{E} R_k x^k (1-x)^{E-k} \\
&= \sum_{k=0}^{E} \binom{E-k_0}{k-k_0} x^k (1-x)^{E-k} \\
&= x^{k_0} \sum_{k=k_0}^{E} \binom{E-k_0}{k-k_0} x^{k-k_0} (1-x)^{E-k_0-(k-k_0)} \\
&= x^{k_0} \sum_{k'=0}^{E-k_0} \binom{E-k_0}{k'} x^{k'} (1-x)^{E-k_0-k'} \\
&= x^{k_0}
\end{aligned}
\tag{13}
$$

Since the effect of each structural motif, and each motif overlap, is additive on R_k, we can reduce the general case to sums like the above, so we immediately find:

$$
R(x) = \sum_{k=0}^{E} N_k x^k
\tag{14}
$$

Thus the N_k defined in Equation 11 are indeed the same coefficients as those introduced in Equation 7.

E. Alternative damage models

The reasoning above is all done in the context of the usual edge damage model introduced by Moore and Shannon. This damage model is appropriate for studying bond percolation. An entirely analogous set of arguments applies to a vertex damage model, in which a set of k *vertices* is chosen uniformly at random, producing a unique subgraph containing all the edges whose endpoints are both in the selected set of vertices. This damage model is appropriate for studying site percolation. Coefficients analogous to P_k and N_k can be derived (substituting the number of vertices V for the number of edges E wherever it appears) and structural motifs can be defined in terms of vertex removal instead of edge removal. The physical interpretation of N_k in terms of these structural motifs is the same. It is likely that there are many other damage models with these properties. Here, we consider only the edge damage model, because it serves to illustrate the role of structural motifs and its analysis is simpler.

III. ESTIMATES AND BOUNDS ON RELIABILITY FOR SPECIAL CASES

In this section, we illustrate how the study of structural motifs and their overlaps helps understand network reliability under several different rules.

A. Exact expressions

If the network contains m structural motifs that are disjoint, and they all have the same size k_0, the only non-zero coefficients are $N_{lk_0}^{(l)} = (-1)^{l+1}\binom{m}{l}$, yielding

$$
R(x) = 1 - (1 - x^{k_0})^m.
\tag{15}
$$

If the network contains m structural motifs, every pair overlaps in all but one edge, and they all have the same size k_0, the only non-zero coefficients are $N_{k_0+l-1}^{(l)} = (-1)^{l+1}\binom{m}{l}$, yielding

$$
R(x) = x^{k_0-1}[1 - (1-x)^m].
\tag{16}
$$

We can use the identities given in Eq. C4 and C5 to evaluate N_k for the following case: there are f structural motifs; they all have the same size k_0; and all unions of structural motifs have one of only one or two other sizes. Although this case is somewhat artificial, note that the first two conditions are satisfied for any network under the AR-α rule. It seems likely that the last restriction can be relaxed if additional combinatorial identities are

brought to bear on the problem. First suppose that the only nonzero coefficients N_k are for k_0 and k_1. Then we must have the following:

$$N_{k_0} = m;$$
$$m + N_{k_1} = 1; \qquad (17)$$
$$m + |N_{k_1}| = 2^m - 1.$$

These simultaneous equations admit a solution only for $m = 2$, for which $N_{k_1} = 1 - m$ and hence

$$R(x) = x^{k_0} \left[2 - x^{k_1 - k_0} \right]. \qquad (18)$$

Note that k_1 is not determined by this argument; however, it is easy to see that $k_0 + 1 \leq k_1 \leq mk_0$. Now consider the above case, but with three nonzero coefficients instead of two. We have:

$$N_{k_0} = m;$$
$$m + N_{k_1} + N_{k_2} = 1; \qquad (19)$$
$$m + |N_{k_1}| + |N_{k_2}| = 2^m - 1.$$

If we look for solutions with $N_{k_2} \geq 1$, we must have $m \geq 3$. Then the solution is

$$N_{k_1} = 1 - 2^{m-1}; \qquad N_{k_2} = 2^{m-1} - m. \qquad (20)$$

This gives

$$R(x) = m(x^{k_0} - x^{k_2}) + x^{k_1} + 2^{m-1}(x^{k_2} - x^{k_1}). \qquad (21)$$

B. Perturbative estimates of reliability

Since $R(x)$ is defined for x in the interval $[0, 1]$, it is tempting to think that the lowest-order term in x that appears in the reliability polynomial, i.e., $N_{k_{min}} x^{k_{min}}$, is a good estimate of its value. Note that $N_{k_{min}} = N_{k_{min}}^{(1)}$, since any union of 2 or more structural motifs must contain more than k_{min} edges. Moreover $N_{k_{min}} = R_{k_{min}}$. Unfortunately, because the coefficients N_k may grow combinatorially and may be either positive or negative, the leading order coefficient may not be sufficient to determine behavior of the reliability polynomial far from 0. For example, if there are very few different structural motifs with k_{min} edges (specifically, if $N_{k_{min}} \ll N_{1+k_{min}} x$), the contribution of $N_{k_{min}}$ may be overwhelmed by larger structural motifs. Nevertheless, evaluating the lowest-order term provides insight into the relationship between graph structure and reliability.

• **All-Terminal reliability**: Recall that the structural motifs for the All-Terminal rule are spanning trees. Each such tree has exactly $V - 1$ edges. N_{V-1} is thus the number of spanning trees, so the lowest-order term in the reliability polynomial is $N_{V-1} x^{V-1}$. The (Kirchhoff) Matrix Tree Theorem [11] gives N_{V-1} in terms of a cofactor of the graph Laplacian matrix.

• *AR-α* **reliability**: The structural motifs for the *AR-α* reliability rule are trees that contain at least αV vertices. Letting t be the number of such trees, the leading order term in $R(x)$ is $t x^{\alpha V - 1}$. Higher-order terms depend on how the trees overlap. We can use this to establish a tight lower bound on $R(x)$ for one particular choice of α.

The lower bound is generated by graphs that minimize the coefficient of the next higher order term $x^{\alpha V}$. This in turn requires that as many as possible of the motifs overlap in all but one edge. For example, beginning with a single tree, we can change one edge to any other edge that is not already in the tree and does not create a loop in the tree. There are at most $E - (\alpha V - 1)$ ways to do this, depending on the graph. Thus there is a graph with t trees, each of which contains α vertices, each of which differs from any other by exactly 2 edges, if and only if $t \leq E + 2 - \alpha V$. In this case, $R(x) = x^{\alpha V - 2}(1 - x)^t$. As far as we know, this particular tree structure occurs only for $\alpha V = E - 1$. The graph in which it occurs has a central vertex of degree t connected to t linear chains of length E/t (thus t must divide E evenly). The trees contain every edge except the last edge on one of the chains.

• *EAR-α* **reliability**: Satisfying the *EAR-α* rule demands that the sum of squared component sizes equals or exceeds αV^2. What are the structural motifs for this rule? Consider a partition Π of V, i.e., a set of positive integers π_i whose sum is V. The number of elements in Π varies from one partition to another. Then π_i could represent the number of vertices in the ith connected component. Furthermore, if each component is a tree, the number of edges in the ith component is just $\pi_i - 1$, hence the number of edges in the entire subgraph is $\sum_i (\pi_i - 1) = V - C$, where $C = |\Pi|$ is the number of components. There are many ways to assign vertices to components, even for a single Π. Each will generate a different structural motif, as long as the reliability condition $\sum_i \pi_i^2 \geq \alpha V^2$ is satisfied. The smallest number of edges results from a subgraph with the largest number of components. The result is that k_{min}, the size of the smallest structural motif, is the size of a subgraph with all isolated vertices except for one large tree with v vertices. k_{min} can be determined by the constraint

$$\sum_i \pi_i^2 = v^2 + (V - v) \geq \alpha V^2, \qquad (22)$$

or

$$v > \sqrt{\alpha} \left[1 - \frac{1}{\alpha V} \left(1 - \frac{1}{4V} \right) \right]^{-1/2} V. \qquad (23)$$

Thus $k_{min} = V - (V - v + 1) \approx \sqrt{\alpha} V - 1$, and $N_{k_{min}}$ is the number of different trees that can be made with k_{min} edges.

IV. STRUCTURAL MOTIFS TO FIND EDGE IMPORTANCE

In section II B we demonstrated how structural motifs for 2 dimensional grid depend on parameters of the two terminal reliability rule i.e. S and T. Here we explain how this fact can effect edge importance based on reliability rules. We computed reliability for two-terminal reliability rule for two case on two dimensional grids in Figure 2. It is clear that grid is more reliable for the reliability rule parameters in the right i.e. it is more probable to have a path from S to T_2 than to T_1. This can be seen in Figure 3. Next we remove two out of three edges on the shortest path connecting S and T_2 and we compute the reliability for both cases again. We see that reliability of the grid decreases more for the left case as expected. This result suggests to employ structural motifs for finding most important edges in a way that reflects the choices of parameters for the reliability rule.
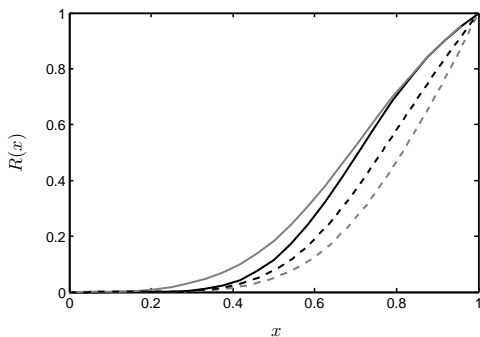

FIG. 3. Reliability polynomial for two terminal reliability black curve represents reliability for $S - T_1$, gray curve for $S - T_2$ and dashed curves represent the reliability polynomial after removing two edges.
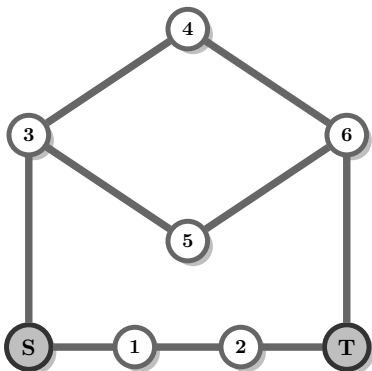

FIG. 4. Toy graph illustrating the possible x-dependence of an edge's importance.

The reliability polynomials of different graphs may intersect multiple times. [12] This means that, for different values of the damage model's parameters $\vec{x}$, the relative reliability of the two graphs switches signs. If both graphs are subgraphs of the network of interest with the same number of elements removed, then clearly the Birnbaum importance ranking of the elements is different for different parameter values. This is perhaps surprising, but it is an important feature of this approach compared to, for example, a graph statistic that is independent of $\vec{x}$.

For example, consider the graph in Figure 4 under $S - T$ reliability for the indicated S and T. First, we write down the reliability for the graph by inspection from its structural motifs. There are three motifs, $A \equiv S12T$, $B \equiv S354T$, and $C \equiv S364T$. The first has size 3; the second and third, size 4. The second and third overlap in two edges, but are disjoint from the first. There are no edges that do not appear in any structural motif. Taken together, this gives:

$$R(x) = x^3 + 2x^4 - x^6 - 2x^7 + x^9. \qquad (24)$$

(Note that $R(x)$ satisfies the constraints $\sum N_k = 1$ and $\sum |N_k| = 2^m - 1$, where $m = 3$ is the number of structural motifs.) By symmetry, we expect the three edges $S1$, 12, and $2T$ to be equally important, and also the pair $S3$ and $4T$, and finally the four edges 35, 36, 54, and 64. Which edge is most important? A moment's thought shows that any edge from the last four is less important than any other edge. The real choice is between $S1$, which is part of a single structural motif of size 3, and $S3$, which is part of *two* structural motifs of size 4. We consider the reliability $R_1(x)$ of the graph after removing edge $S1$, leaving motifs B and C, and the reliability $R_2(x)$ after removing edge $S3$, leaving only motif A. Again by inspection, these are:

$$R_1(x) = 2x^4 - x^6 \qquad (25)$$
$$R_2(x) = x^3. \qquad (26)$$

By definition, the importance of the edges is $I_{S1}(x) \equiv R(x) - R_1(x)$ and $I_{S2}(x) \equiv R(x) - R_2(x)$. Hence, the rank of the edges switches if the polynomial $I_{S1}(x) - I_{S2}(x) = R_2(x) - R_1(x)$ changes sign. In fact, this polynomial has a zero in the interval $[0, 1]$. That is,

$$R_2(x) - R_1(x) = x^3(1 - 2x + x^3) \begin{cases} > 0 \text{ for } x < 0.618... \\ < 0 \text{ for } 0.618... < x \end{cases} \qquad (27)$$

V. APPLYING RELIABILITY CONCEPTS TO OTHER NETWORK ANALYSIS PROBLEMS

The representation of the reliability polynomial in terms of structural motifs provides a convenient organizing principle for thinking about general network analysis problems. As one example, consider the tradeoffs between two systems: one with only a few completely redundant reliable subsystems and another with more, but only partially redundant, ones. To study this we consider

two extreme cases of overlap. One contains r_1 structural motifs of size k_1, any two of which differ by only two edges. They are thus built using a total of $2r_1 + k_1 - 2$ edges. The reliability of this combination can be written as:

$$R_1(x) = \sum_{i=1}^{r_1} (-1)^{i+1} \binom{r_1}{i} x^{k_1 + r_1(i-1)}. \qquad (28)$$

Using the same number of edges we can construct $r_2 = \frac{2r_1 + k_1 - 2}{k_2}$ motifs of size k_2 that are completely disjoint. The reliability of this combination of motifs is:

$$R_2(x) = \sum_{i=1}^{r_2} (-1)^{i+1} \binom{r_2}{i} x^{ik_2}. \qquad (29)$$

Knowing the reliability for these two cases, we are able to compare the reliability of networks with different configurations of structural motifs of different sizes. As an example we compared the reliability of a network composed of 20 motifs with 18 edges that are different from one another only in two edges with a network of 4 completely disjoint motifs of size 6. Figure 5 shows the reliability curves for these two networks and their difference as a function of x. The analysis shows that the network of disjoint motifs is more reliable for smaller values of x while the opposite is true for larger x values.

This approach could also be used to estimate the number of spanning trees in a graph. A spanning tree is a subgraph of the network that includes all vertices [13–15]; the number of spanning trees can be estimated by evaluating the All-Terminal reliability. Another problem that can be addressed using this method is to identify chordless loops of various sizes in a network. A chordless loop is a sequence of vertices with more than three vertices if for all $i = 1, \cdots, k$ there is exactly one link from vertex v_i to v_{i+1} and there is no other link between any two of the vertices in this sequence [16]. Recent studies on ecological networks have discovered the existence of many chordless cycles in these networks [17], therefore enumeration of all chordless cycles can make a significant impact on understanding the structure of these networks. An appropriately-designed reliability rule can be used to count the number of chordless cycles of different sizes.

VI. CONCLUSION AND FUTURE WORK

In this paper we focused on the representation of the reliability polynomial in terms of structural motifs. We have shown that network reliability is simply related to the number of edges in unions of structural motifs N_k (14). Whereas the coefficients P_k of $x^k (1 - x)^{E-k}$ are easy to estimate numerically but hard to work with analytically, the coefficients N_k of x^k are hard to estimate numerically but easy to work with analytically. To demonstrate this, we have derived closed-form expressions for N_k for several types of graphs. The resulting expressions were confirmed by numerical estimation. We anticipate that this approach can lead us to a measure of
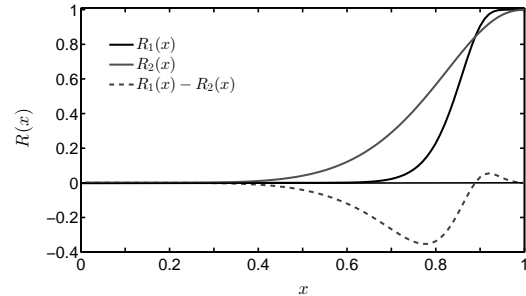

FIG. 5. Comparing the reliability of a network with many overlapping structural motifs with that of a network with a few disjoint motifs.

edge centrality that relates the importance of an edge to the frequency of its appearance in different structural motifs [18]. While we can use numerical simulation to study specific large, realistic networks – including epidemiology on social networks [19–21] – we can use the notion of structural motifs to understand the differences between networks that are discovered in simulation. We expect this approach to be particularly useful in studying the stability and robustness of interconnected networks [22–27].

ACKNOWLEDGMENTS

Appendix A: Inclusion-Exclusion proof of reliability polynomial

We begin by rewriting the expansion

$$R(x) = \sum_{k=0}^{E} R_k \, x^k \, (1 - x)^{E-k}, \qquad (A1)$$

as the following,

$$R(x) = \sum_{A \subseteq E} r(A)\, x^A\, (1-x)^{E-A}\, ; \qquad \text{(A2)}$$

here x^A is short for $x^{|A|}$ where $|A|$ refers to size of the subset $A \subseteq E$ that characterizes the edge-induced subgraph A. $(1-x)^{E-A}$, then, has a similar interpretation. We note that (A1) is equivalent to (A2) simply because R_k is counting the number of edge-induced subgraphs $A \subseteq E$ of size k that satisfy the reliability rule r; for these subgraphs $r(A) = 1$ and the equivalence follows –for those that do not, $r(A) = 0$.

If the reliability rule is coherent, there is a unique family of minimal subgraphs $A \subseteq E$ such that for every subgraph A' that is a *proper* subgraph of either one of them, $r(A') = 0$. This means that every subgraph that is accepted by the reliability rule r, contains at least one of these minimal subgraphs (entirely). Therefore, we can see that every reliability rule is in 1-1 correspondence with a certain family of (edge-induced) subgraphs A of E. As a result, we can now define the reliability rule in terms of the very family of *minimal* subgraphs that were obtained from this rule,

$$r(B) = \begin{cases} 1 & ; \quad \exists\, A \in \mathcal{A} \, : \, A \subseteq B \\ 0 & ; \qquad\qquad \text{otherwise} \end{cases} ; \qquad \text{(A3)}$$

here $\mathcal{A}$ represents this family.

We now show that $R(x)$ stands for the probability that at least one of these minimal subgraphs is *operating* in the sense that all of its corresponding edges function. Put differently, we show that

$$R(x) \stackrel{?}{=} \mathbf{Pr}\left(\bigvee_i \{A_i \text{ operates}\} \right)$$

$$= \mathbf{Pr}\left(\bigvee_i A_i \right), \qquad \text{(A4)}$$

where A_i's are different members of the above family of minimal subgraphs indexed by i; the second line is understood as a shorthand notation for the first.

Let $\langle A_1 \vee A_2 \rangle$ stand for the family of all subgraphs that contain all of the edges in at least one of the subgraphs A_1 or A_2. Also let $\langle \overline{A_1 \vee A_2} \rangle$ denote the family of all subgraphs that lack at least one edge from each of the subgraphs A_1 and A_2. Next, we define $\langle A_1 \wedge A_2 \rangle$ as the family of all subgraphs that contain all of the edges in both A_1 and A_2. And finally, let $\langle A_1 \wedge \overline{A_2} \rangle$ represent the family of all subgraphs that contain all of the edges in subgraph A_1 *and* lack at least one edge from A_2. We also utilize unambiguous generalizations of these notations at a later inductive step.

Now suppose that the reliability rule r is such that only one minimal subgraph, namely A_1, is accepted and, as a result, $R(x)$ becomes

$$R(x) = \sum_{A_1 \subseteq A \subseteq E} x^A\, (1-x)^{E-A}$$

$$= \sum_{A \in \langle A_1 \rangle} x^A\, (1-x)^{E-A}$$

$$= x^{A_1} \qquad \text{(A5)}$$

$$= \mathbf{Pr}\left(A_1 \text{ operates} \right); \qquad \text{(A6)}$$

where equation (A5) can be easily obtained by induction on the edges in subgraph $E - A_1$. We can take (A6) as the base case for an inductive proof of the equivalence of expressions (A2) and (A4) and proceed to the induction step. However, to break in the notation we introduced above, we also prove the case of two minimal subgraphs, and then proceed to the inductive step. Therefore, suppose that we have two minimal subgraphs A_1 and A_2 and we want to prove the following,

$$\mathbf{Pr}\left(\bigvee_{i \leq 2} A_i \right) \stackrel{?}{=} \sum_{A \in \langle A_1 \vee A_2 \rangle} x^A\, (1-x)^{E-A}. \qquad \text{(A7)}$$

The right hand side of this euqality is precisely the expansion of $R(x)$ for a reliability rule that is in 1-1 correspondence with the family of two minimal subgraphs A_1 and A_2; therefore our claim is proved for this case if the above equality holds. We start from the left hand side and note that

$$\mathbf{Pr}\left(\bigvee_{i \leq 2} A_i \right) = \mathbf{Pr}\,(A_1) + \mathbf{Pr}\,(A_2) - \mathbf{Pr}\,(A_1 \wedge A_2). \qquad \text{(A8)}$$

Now, we make the following observations,

$$\mathbf{Pr}\,(A_1) = \sum_{A \in \langle A_1 \rangle} x^A\, (1-x)^{E-A}$$

$$= \left(\sum_{A \in \langle A_1 \wedge \overline{A_2} \rangle} + \sum_{A \in \langle A_1 \wedge A_2 \rangle} \right) x^A\, (1-x)^{E-A}, \qquad \text{(A9)}$$

$$\mathbf{Pr}\,(A_1 \wedge A_2) = \sum_{A \in \langle A_1 \wedge A_2 \rangle} x^A\, (1-x)^{E-A}. \qquad \text{(A10)}$$

From these expansions, it immediately follows for (A8) that

$$\mathbf{Pr}\left(\bigvee_{i \leq 2} A_i \right)$$

$$= \left(\sum_{A \in \langle A_1 \wedge \overline{A_2} \rangle} + \sum_{A \in \langle \overline{A_1} \wedge A_2 \rangle} + \sum_{A \in \langle A_1 \wedge A_2 \rangle} \right) x^A\, (1-x)^{E-A}$$

$$= \sum_{A \in \langle A_1 \vee A_2 \rangle} x^A\, (1-x)^{E-A}. \qquad \text{(A11)}$$

This proves the base case of a family with only two minimal subgraphs A_1 and A_2. The inductive step is quite similar,

$$\mathbf{Pr}\left(\bigvee_{i \leq n} A_i\right)$$

$$= \mathbf{Pr}\left(\bigvee_{i<n} A_i\right) + \mathbf{Pr}\left(A_n\right) - \mathbf{Pr}\left(\left(\bigvee_{i<n} A_i\right) \wedge A_n\right)$$

$$= \left(\sum_{A \in \langle \bigvee_{i<n} A_i \wedge \overline{A_n}\rangle} + \sum_{A \in \langle \overline{\bigvee_{i<n} A_i} \wedge A_n\rangle}\right) \tag{A12}$$

$$+ \sum_{A \in \langle \bigvee_{i<n} A_i \wedge A_n\rangle}\right) x^A (1-x)^{E-A} \tag{A13}$$

$$= \sum_{A \in \langle \bigvee_{i<n} A_i\rangle} x^A (1-x)^{E-A} \tag{A14}$$

$$= \sum_{A \subseteq E} r(A)\, x^A (1-x)^{E-A},$$

where r is the reliability rule that corresponds to the family of minimal subgraphs $\{A_1, \cdots, A_n\}$. In (A13), $\langle \bigvee_{i<n} A_i \wedge \overline{A_n}\rangle$ represents the family of all subgraphs that contain all of the edges in at least one of the subgraphs A_1 through A_{n-1} *and* lack at least one edge from the subgraph A_n. Next, $\langle \overline{\bigvee_{i<n} A_i} \wedge A_n\rangle$ stands for the family of all subgraphs that contain all of the edges in subgraph A_n *and* lack at least one edge from each and every subgraph A_1 through A_{n-1}. Lastly, $\langle \bigvee_{i<n} A_i \wedge A_n\rangle$ refers to the family of all subgraphs that contain all of the edges in both A_n and at least on of the subgraphs A_1 through A_{n-1}. Equation (A14), then, follows from the fact that the combination of these three possibilities is precisely what $\langle \bigvee_{i<n} A_i\rangle$ stands for.

So far, we have established that

$$R(x) = \mathbf{Pr}\left(\bigvee_i A_i\right), \tag{A15}$$

and now we proceed to prove a final equivalent expansion, namely

$$R(x) \stackrel{?}{=} \sum_{k=0}^{E} N_k\, x^k \tag{A16}$$

$$N_k \equiv \sum_{l=1}^{E} (-1)^{l+1} N_k^{(l)}, \tag{A17}$$

where $N_k^{(l)}$ denotes the number of combinations of l minimal subgraphs whose union contains exactly k edges. To show that the above holds, we appeal to equation (A15) which can be now, by account of inclusion-exclusion principle, expanded as

$$\mathbf{Pr}\left(\bigvee_{1 \leq i \leq n} A_i\right)$$

$$= \sum_{i=1}^{n} (-1)^{i+1} \sum_{1 \leq j_1 < \cdots < j_i \leq n} \mathbf{Pr}\left(A_{j_1} \wedge \cdots \wedge A_{j_i}\right)$$

$$= \sum_{i=1}^{n} (-1)^{i+1} \sum_{1 \leq j_1 < \cdots < j_i \leq n} x^{A_{j_1} \cup \cdots \cup A_{j_i}}; \tag{A18}$$

this clearly coincides with the expansion (A16).

Appendix B: Additional examples of contribution of structural motifs to $R(x)$

a. Example 2: Two disjoint structural motifs

Suppose the network G contains exactly two structural motifs, that both have k_0 edges, and that no edge is in both. Arguing as in Example 1, for $k < 2k_0$, R_k is simply twice what it is for the case of a single structural motif. But when $k = 2k_0$, the subgraph that consists of the union of the two structural motifs will have been counted twice instead of once. Similarly, for $k > 2k_0$, the number of graphs overcounted is given by assigning $2k_0$ of the edges and choosing the remaining $k - 2k_0$ in the subgraph from among the remaining $E - 2k_0$ in the graph G. Hence:

$$R_k = \begin{cases} 0 & k < k_0 \\ 2\binom{E-k_0}{k-k_0} & k_0 \leq k < 2k_0 \\ 2\binom{E-k_0}{k-k_0} - \binom{E-2k_0}{k-2k_0} & 2k_0 \leq k \end{cases} \tag{B1}$$

b. Example 3: Three disjoint structural motifs

Suppose the network contains exactly three structural motifs, that all three have k_0 edges, and that the three edge sets are disjoint. Again, when $k_0 \leq k < 2k_0$, each motif generates $\binom{E-k_0}{k-k_0}$ different reliable subgraphs, and for $2k_0 = k$, three of these subgraphs are counted twice. But in this case, when k reaches $3k_0$, the subgraph consisting of all three motifs is first included three times (once for each motif), then excluded three times (once for each pair of motifs) with the net result that it must be included again:

$$R_k = \begin{cases} 0 & k < k_0 \\ 3\binom{E-k_0}{k-k_0} & k_0 \leq k < 2k_0 \\ 3\binom{E-k_0}{k-k_0} - 3\binom{E-2k_0}{k-2k_0} & 2k_0 \leq k < 3k_0 \\ 3\binom{E-k_0}{k-k_0} - 3\binom{E-2k_0}{k-2k_0} + \binom{E-3k_0}{k-3k_0} & 3k_0 \leq k \end{cases} \tag{B2}$$

Appendix C: Constraints on coefficients

Several constraints apply to N_k. A union of l motifs can have size k only if all possible unions of $l-1$ of the same motifs have size less than k. This leads to a set of constraints of the form

$$N_k^{(2)} \leq \binom{\sum_{k'=0}^{k} N_{k'}^{(1)}}{2}. \tag{C1}$$

In addition, a union of l motifs can have size k only if all possible unions of $l-1$ and $l-2$ of the same motifs have size less than k. For instance, $N_k^{(3)}$ has the following upper bound

$$N_k^{(3)} \leq \left(\sum_{k'=0}^{k'=k} N_{k'}^{(1)}\right)\left(\sum_{k'=0}^{k'=k} N_{k'}^{(2)}\right) + \binom{\sum_{k'=0}^{k'=k} N_{k'}^{(1)}}{3}. \tag{C2}$$

Overall, since all unions of l structural motifs must be included in G, we have

$$\sum_{k=0}^{E} N_k^{(l)} = \binom{f}{l}, \tag{C3}$$

where f is the total number of structural motifs. Finally, the facts that $\sum_{l=0}^{f} \binom{f}{l} = 2^f$ and $\sum_{l=0}^{f}(-1)^l\binom{f}{l} = 0$ imply that

$$\sum_{k=0}^{E} N_k = \sum_{k=0}^{E}\sum_{l=1}^{f}(-1)^{l+1}N_k^{(l)}$$
$$= \sum_{l=1}^{f}(-1)^{l+1}\binom{f}{l} = 1; \tag{C4}$$

and

$$\sum_{k=0}^{E} |N_k| = \sum_{k=0}^{E}\sum_{l=1}^{f} N_k^{(l)}$$
$$= \sum_{l=1}^{f}\binom{f}{l} = 2^f - 1; \tag{C5}$$

[1] E. Moore and C. Shannon, Journal of the Franklin Institute **262**, 191 (1956).
[2] M. Newman, A. Barabsi, and D. Watts, Princeton University Press (2006).
[3] A. Ganesh, L. Massouli, and D. Towsley, Proceedings of the IEEE INFOCOM Miami, FL, 05 (2005).
[4] C. J. Colbourn, *The Combinatorics of Network Reliability* (Oxford University Press, 1987).
[5] M. Youssef, Y. Khorramzadeh, and S. Eubank, Physical Review E **88** (2013).
[6] C.-C. Jane, W.-H. Shenb, and Y.-W. Laihc, European Journal of Operational Research **195**, 427 (2009).
[7] E. S. Elmallah, Networks **22**, 369 (1992).
[8] J. Rong-Hong, Computers and Operations Research **20**, 25 (1993).
[9] S. LM, W. CP, S. IM, and K. J. Simon C, Mathematical Biosciences 180 (2002).
[10] U. Alon, Nature Reviews Genetics **8**, 450,461 (2007).
[11] G. Kirchhoff, Ann. Phys. Chem. **72**, 497 (1847).
[12] J. Brown, Y. Koc, and R. Kooij, in *Proceedings of the 3rd International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2011* (Budapest, Hungary, October 5-7, 2011, 2011).
[13] S. Chaiken, SIAM J. ALG. DISC. METH **3** (1982).
[14] D. M. Cvetkovi, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Applications* (Wiley, 1998).
[15] F. Buekenhout and M. Parker, Disc. Math **186**, 69 (1998).
[16] N. Sokhn, R. Baltensperger, L.-F. Bersier, J. Hennebert, and U. Ultes-Nitschey, *Identification of Chordless Cycles in Ecological Networks* (Springer, 2013) pp. 316–324.
[17] M. Huxham, S. Beaney, and D. Raffaelli, **76**, 284 (1996).
[18] S. Eubank, M. Youssef, and Y. Khorramzadeh, in *Proceedings of the second workshop on complex networks and their applications, SITIS 2013* (Tokyo, Japan, 2013).
[19] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature **429(6988)**, 180 (2004).
[20] N. Parikh, M. Youssef, S. Swarup, and S. Eubank, Scientific Reports **3** (2013).
[21] S. Y. D. Valle, J. M. Hyman, H. W. Hethcote, and S. G. Eubank, Social Network **29**, 539,554 (2007).
[22] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, Nat. Phys. **8**, 40,48 (2012).
[23] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Nature **464**, 1025,1028 (2009).
[24] S. Trajanovski, J. Martn-Hernndez, W. Winterbach, and P. Van Mieghem, Journal of Complex Networks **1**, 44 (2013).
[25] P. Van Mieghem, Computer Communications **35**, 1494 (2012).
[26] M. Youssef, R. Kooij, and C. Scoglio, Journal of Computational Science **2**, 286 (2011).
[27] W. Ellens and R. Kooij, arXiv:1311.5064 [cs.DM] (2013).