Network connectivity during mergers and growth: Optimizing the addition of a module

Dane Taylor and Juan G. Restrepo

Phys. Rev. E **83**, 066112 — Published 20 June 2011

DOI: 10.1103/PhysRevE.83.066112

Network connectivity during mergers and growth: optimizing the addition of a module

Dane Taylor[1,*] and Juan G. Restrepo[1]

[1]*Department of Applied Mathematics, University of Colorado, Boulder, Colorado 80309, USA*

The principal eigenvalue λ of a network's adjacency matrix often determines dynamics on the network (e.g., in synchronization and spreading processes) and some of its structural properties (e.g., robustness against failure or attack) and is therefore a good indicator for how "strongly" a network is connected. We study how λ is modified by the addition of a module, or community, which has broad applications, ranging from those involving a single modification (e.g., introduction of a drug into a biological process) to those involving repeated additions (e.g., power-grid and transit development). We describe how to optimally connect the module to the network to either maximize or minimize the shift in λ, noting several applications of directing dynamics on networks.

I. INTRODUCTION

Spectral approaches for analysis of complex networks are becoming increasingly important due to their ability to describe the effect of network structure on dynamical processes. In particular, the principal eigenvalue λ of a network's weighted adjacency matrix A (A_{ij} is nonzero if there exists a link from node i to node j) is significant for dynamics on networks such as the synchronization of heterogeneous oscillators [1, 2], epidemic and information spreading [3], structural robustness (percolation) [4], the stability of equilibria for certain systems of network-coupled ordinary differential equations [5], the stability of gene expression in genetic networks [6], and criticality in network-coupled excitable systems [7].

Given the importance of λ in determining the outcome of so many dynamical processes on networks, there has been much interest in modifying λ through structural perturbations. In particular, the effect of removing node j can be quantified by its *dynamical importance*: $I_j = -\delta\lambda/\lambda \approx v_j u_j / v^T u$ [8], where u (v) is the right (left) eigenvector corresponding to the principal eigenvalue λ (i.e., $Au = \lambda u$, $v^T A = \lambda v^T$), and $\delta\lambda$ is the decrease in the principal eigenvalue that would result from the removal of node j. As an example application, a node removal strategy targeting nodes with large dynamical importance fragments a network more rapidly than targeting nodes with large degree (number of links) [8]. Ref. [9] extended these results by finding perturbative expressions for the change in eigenvalue $\delta\lambda$ due to the removal of groups of nodes as well as the addition or deletion of groups of links. Ref. [10] considered a perturbative approach to study the spectrum of networks with community structure.

In this study, we consider the effect on the largest eigenvalue of a network's adjacency matrix from the addition of a secondary network (referred to as the module or community). As opposed to previous work [8–10], we explicitly consider the effect of the module's topology on the resulting eigenvalue and use this information to discuss how one can make optimal connections to either maximize or minimize the effect on λ. There are many applications where smaller groups adhere to a larger network in social and economical networks [11] (e.g., the merging of corporations or markets) and biological networks (e.g., modifying a system of biochemical reactions with a drug [12, 13] or the merging of ecosystems [14]). For example, recent studies have shown that the effect on the largest eigenvalue of the Jacobian matrix describing interactions in an ecological network due to the addition of a species may be integral to the formation of ecological communities [15]. Moreover, our results offer new insight regarding the prevalence of subgraph motifs (recurrent subgraphs having a frequency higher than expected). While motifs have been cited as essential building blocks in biological networks [16], their role is not fully understood. For example, in contrast to several studies indicating that the global stability and robustness of a system is strongly influenced by the structure of motifs [17], our study suggests that "how" a motif is connected to the remaining network may be as significant as its structure (see Fig. 5).

This paper is organized as follows. In Sec. II we describe the problem and introduce variables. In Sec. II A we present perturbative approximations for $\delta\lambda$ in terms of spectral properties of A and the module to be added. In Sec. II B we test these approximations on several real networks. In Sec. III we discuss how our results can be used to optimize the connections between the original network and module. In Sec. IV we discuss our results, citing several applications of how they may be used to direct dynamics on networks. These results have application from cases in which just a single merger needs to be optimally designed to cases where a large number of small additions need to be optimized to quickly evolve λ to a desired value.

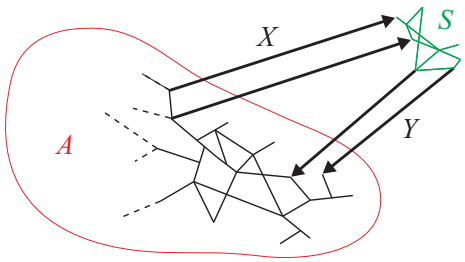

FIG. 1: (Color online) A module (described by matrix S) is connected to the original network (described by matrix A) using directed connections (described by the matrices X and Y).

*Electronic address: dane.taylor@colorado.edu

II. MODULE ADDITION

We consider the addition of a secondary network, or module, to an existing network, as shown schematically in Fig. 1. The original network of size n is described by an $n \times n$ weighted adjacency matrix A such that its entries A_{ij} satisfy $A_{ij} \neq 0$ if there exists a link from node i to node j and $A_{ij} = 0$ otherwise. Another network of size m (described by a $m \times m$ adjacency matrix S) is to be connected to the original network. We will refer to this secondary network as the *module*. In what follows, we will sometimes refer to both the original network and the module by their respective adjacency matrices, A and S.

Assuming that the original eigenvalue problems $Au = \lambda u$ and $v^T A = \lambda v^T$ have been solved, the modified eigenvalue problem after module addition may be formulated as

$$\begin{bmatrix} A & X \\ Y^T & S \end{bmatrix} \begin{bmatrix} u + \Delta^U \\ \Delta^L \end{bmatrix} = (\lambda + \delta\lambda) \begin{bmatrix} u + \Delta^U \\ \Delta^L \end{bmatrix}, \quad (1)$$

where we use the following definitions: (i) $\delta\lambda \geq 0$ denotes the shift in the largest eigenvalue; (ii) matrix X (Y) is size $n \times m$, has positive entries, and describes all directed links from A to S (S to A); (iii) Δ^U is a vector of length n which represents the shift in eigenvector u; and (iv) Δ^L is a vector of length m which represents the new eigenvector components. For the typical case in which no negative weights are allowed (i.e., $A_{ij} \geq 0$), the principal eigenvalues $\{\lambda, \lambda + \delta\lambda\}$ and all entries in $\{u, u + \Delta^U, \Delta^L\}$ are guaranteed to be nonnegative by the Perron-Frobenius theorem for nonnegative matrices [18]. Although in this paper we only consider matrices with positive entries so that the Perron-Frobenius theorem can be applied, in general our analysis only requires that the eigenvalue with largest magnitude λ is real and well separated from the remaining eigenvalues. While this is typical for networks with positive links [10], it is also observed for networks with negative links provided that they represent a small fraction of the number of links (e.g., see Fig. 6 in [2]).

A. Effect of module addition

We restrict our analysis to cases where the effect of the module addition is small, which will allow us to study its effect as a perturbation to the original eigenvalue problem. This restriction is applicable for describing heavy-sided mergers and applications for which a network is modified gradually, such as the expansion of infrastructure. Considering the upper and lower blocks of Eq. (1) independently and after left-multiplying the top block by the left principal eigenvector v^T (i.e., $v^T A = \lambda v^T$), we obtain

$$\begin{aligned} \delta\lambda &= \frac{v^T X \Delta^L}{v^T(u + \Delta^U)} \\ \Delta^L &= ((\lambda + \delta\lambda) I_m - S)^{-1} Y^T (u + \Delta^U), \end{aligned}$$

where I_m is the identity matrix of size m.

Assuming that the effect of the module addition is small, we have $\delta\lambda \ll \lambda$, $v^T \Delta^U \ll v^T u$, and $\delta\lambda \ll |\lambda - \lambda_S|$, where λ_S is the largest eigenvalue of the module. To first order, we find

$$\delta\lambda \approx \frac{1}{\lambda v^T u} v^T X K^S Y^T u \quad (2)$$

$$\Delta^L \approx \lambda^{-1} K^S Y^T u, \quad (3)$$

where we have defined $K^S \equiv (I_m - S/\lambda)^{-1}$. These expressions relate the change in the dominant eigenvalue $\delta\lambda$ to the topology of the added module S, the spectral properties of the original networks (u, v, and λ), and the way in which the module is coupled to the network by matrices X and Y. When the module contains few nodes, approximating $\delta\lambda$ by inverting an $m \times m$ matrix is significantly more efficient and, as we will see, offers more insight than solving the original $(m + n) \times (m + n)$ eigenvalue problem. Using $v = u$ and $X = Y$ for undirected networks, Eq. (2) simplifies to $\delta\lambda \approx \lambda^{-1}(X^T u)^T K^S (X^T u)$.

If the connections between the module and original network are made randomly, we can use Eq. (2) to estimate average values of $\delta\lambda$. Suppose that the entries of the matrix X are independent random variables such that $X_{ij} = 1$ with probability $x/(nm)$ and 0 otherwise, so that the expected number of links from the original network to the added module is x. Similarly, we assume that the entries of Y are independent random variables which are 1 with probability $y/(nm)$ and 0 otherwise. By averaging Eq. (2) and using the independence of X and Y, we find

$$\langle \delta\lambda \rangle = \frac{\bar{u}\bar{v}}{\lambda v^T u} \left(\frac{x}{m}\right) \left(\frac{y}{m}\right) \sum_{i,j} K^S_{ij}, \quad (4)$$

where $\bar{u} = n^{-1} \sum_{j=1}^n u_j$ and $\bar{v} = n^{-1} \sum_{j=1}^n v_j$. Thus, in addition to properties dependent on the original network, $\langle \delta\lambda \rangle$ is proportional to the product of the relative number of connections to and from the module (xy/m^2) and on the sum of elements in the matrix K^S. Moreover, for large λ/λ_S we have $\sum_{ij} K^S_{ij} \approx m$, the number of nodes in S. While this expression provides us with the average $\langle \delta\lambda \rangle$ when X and Y are chosen randomly, as discussed in Section III, strategically selecting connection matrices (X,Y) (e.g., to maximize $\delta\lambda$) can lead to significant variations in $\delta\lambda$ for a given module.

For the optimization objectives explored later in this text, as well as situations in which computing K^S is inconvenient, it is useful to represent Eqs. (2) and (3) using a series expansion for K^S. For $\lambda_S < \lambda$, we have $K^S = (I_m - S/\lambda)^{-1} \approx \sum_{j=0}^k (S/\lambda)^j$. We thus define the k-th order approximations:

$$\delta\lambda_k = \frac{1}{\lambda v^T u} \sum_{j=0}^{k-1} \lambda^{-j} v^T X S^j Y^T u \quad (5)$$

$$\Delta^L_k = \sum_{j=0}^{k-1} \lambda^{-(j+1)} S^j Y^T u. \quad (6)$$

Because the matrices (X,Y) are often sparse and the module is often much smaller than the network, Eq. (5) is typically

Network and reference	N	$\langle d\rangle$	λ	λ_2
Neural network of C. elegans [19]	297	7.9	9.2	5.7
Network of political blogs [20]	1490	12.8	34.4	26.8
Yeast PPI network [21]	2361	5.6	12.1	9.4
Word association network [22]	5018	12.7	13.4	10.2

TABLE I: Test networks used and their characteristics: number of nodes N; mean degree $\langle d\rangle$; largest eigenvalue λ; and second largest eigenvalue λ_2.

very computationally efficient. We note that for large enough k, the error introduced by Eq. (2) dominates the error of series truncation in Eqs. (5) and (6). No gain was found by using $k > 4$ in the experiments that are to follow.

B. Numerical tests on real networks

We test our approximations by considering module additions to four networks: a neural network of C. elegans [19]; a network of political blogs [20]; a network of protein-protein interactions in the organism S. cerevisiae (i.e., brewers/bakers yeast) [21]; and a network of associations between words [22]. Their characteristics are summarized in Table I. We begin by examining the average effects for adding a module using random connections. First, matrices were constructed by randomly selecting 10 entries in X and Y to be 1 and the rest to be 0 (i.e., in the previous notation, $x = y = 10$). For each realization $\delta\lambda^{\text{act}}$, the actual eigenvalue shift [i.e. solving Eq. (1)], was compared to our approximations given by Eqs. (2) and (5). In the top panel of Fig. 2, Eq. (4) (stars) is shown to accurately predict the numerically-observed average $\langle\delta\lambda\rangle$ (solid line) for connecting the modules to the directed neural network of C. elegans using 10^4 realizations of (X,Y). Average values for the relative error $\epsilon = (\delta\lambda - \delta\lambda^{\text{act}})/\delta\lambda^{\text{act}}$ are plotted in the bottom panel for both the neural network (circles) and an network of political blogs (triangles) (see TableI) for all 13 non-isomorphic, directed modules of size 3. (Results for the other networks were found to be similar and are omitted for clarity.) Solid lines correspond to Eq. (2), while dotted (dashed) lines correspond to Eq. (5) with $k = 1$ ($k = 2$).

It can be observed in Fig. 2 that $\langle\delta\lambda\rangle$ changes substantially for the different three-node modules (for all networks, $\langle\delta\lambda\rangle$ typically increased $\sim 20\%$ from module 1 to module 13). Observe that the average error $\langle\epsilon\rangle$ of Eq. (5) when the module structure is not used [$k = 1$ (dotted lines in lower plot)] is strongly correlated with $\langle\delta\lambda\rangle$ (upper plot). This is to be expected as the error from neglecting module structure should be related to that structure's ability to modify λ. Note that for the political blog network (triangles), using $k = 2$ in Eq. (5) is nearly as accurate as directly using Eq. (2). As previously mentioned, for large enough k the dominant source of error comes from neglecting higher orders of $\delta\lambda/\lambda$ in the derivation of Eq. (2) [as opposed to series truncation in Eq. (5)].

The validity of our approximations for specific connections

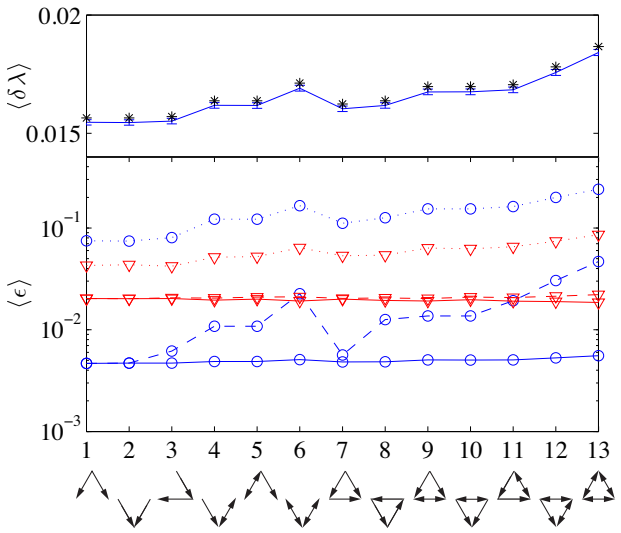

FIG. 2: (Color online) $\delta\lambda$ and approximation errors, ϵ, were averaged over 10^4 realizations of connecting a three-node module to the networks in Table I using 10 random links (see text). Eq. (4) (stars) is shown to be accurate in the upper panel for typical results for the neural network of C. elegans. The average relative error $\langle\epsilon\rangle$ for the neural network of C. elegans (circles) and a network of political blogs (triangles) are given in the lower plot, where Eq. (2) (sold lines) and Eq. (5) with $k = 1$ (dotted) and $k = 2$ (dashed) are shown.

is shown by considering the addition of two bidirectionally-linked nodes ($m = 2$) to an undirected protein-protein interaction (PPI) network and a directed network of word associations (see Table I). In order to illustrate the dependence of $\delta\lambda$ on the matrices X and Y, we will consider two connection strategies: connecting the module to nodes with either (A) increasing nodal degrees or (B) increasing eigenvector entries. For strategy A, the nodes in the original network are ordered so that the in-degrees monotonically increase: $d_1^{in} \le d_2^{in} \le \cdots \le d_N^{in}$. Then for $k \in \{1, 2, ..., N-20\}$, we establish a directed link from nodes $\{k, k+1, \ldots, k+20\}$ to both nodes in the module. The nodal out-degrees are then ordered such that $d_{i_1}^{out} \le d_{i_2}^{out} \le \cdots \le d_{i_N}^{out}$, and links are made to nodes $\{i_k, i_{k+1}, \ldots, i_{k+20}\}$ from both nodes in the module. The case $k = 0$ corresponds to connecting the network nodes with smallest d^{in} to both module nodes, which in turn connect to the nodes with smallest d^{out}; whereas the case $k = N - 20$ corresponds to connecting the nodes with largest d^{in} to both module nodes, both of which in turn connect to the nodes with largest d^{out} (shown schematically in the cartoon in Fig. 3a). For strategy B, we now order the nodes in the original network in order of increasing entries of the left eigenvector v so that $v_1 \le v_2 \le \cdots \le v_N$. As before, for $k \in \{1, 2, ..., N-20\}$, we connect nodes $\{k, k+1, \ldots, k+20\}$ in the network to both module nodes, both of which in turn connect to nodes $\{i_k, i_{k+1}, \ldots, i_{k+20}\}$, where the indices i_j now correspond to the ordering of the right eigenvector entries such that $u_{i_1} \le u_{i_2} \le \cdots \le u_{i_N}$. For both strategies, the indices simplify for undirected net-

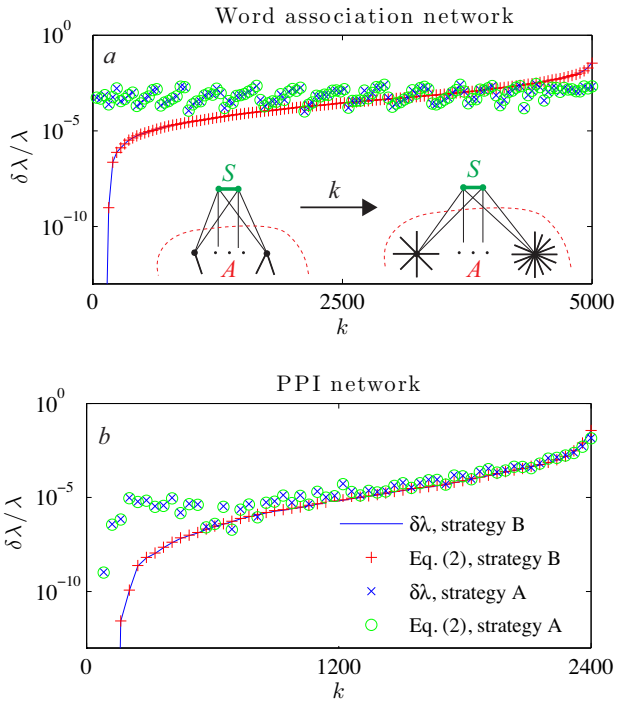

FIG. 3: (Color online) Eigenvalue shift $\delta\lambda$ for connecting a two-node module to (a) the word association network and (b) the PPI network. Eq. (2) (crosses) agrees well with actual values $\delta\lambda^{\text{act}}$ (solid line) for strategy B. The x's and circles show the same respective quantities for strategy A. As indicated by the cartoon, increasing k corresponds to connecting the module to nodes with increasing degrees (strategy A) or eigenvector entries (strategy B).

works, for which we have $u = v$, $d^{out} = d^{in}$, and $i_k = k$.

In Fig. 3, $\delta\lambda$ is plotted for strategies A and B as a function of the parameter k for both (a) the directed word-association network and (b) the undirected PPI network. For strategy B, the crosses show the approximation given by Eq. (2) and the solid line shows the numerically-calculated value from directly solving the eigenvalue problem Eq. (1). The x's and circles respectively show the same quantities for strategy A. The first observation is that the approximation for $\delta\lambda$ works well, with only a small deviation as the perturbation becomes large (not shown). One can observe that strategy B is superior for yielding either large or small $\delta\lambda$ for both networks. However, the two strategies are similar for producing large $\delta\lambda$ for the PPI network in Fig. 3b. This is expected when the first-order approximations to the eigenvectors ($u_i \propto d_i^{out}$ and $v_i \propto d_i^{in}$ [8]) are valid. The results of this experiment suggest that it may be useful to devise connection strategies to systematically maximize (or minimize) $\delta\lambda$.

III. OPTIMIZING CONNECTIONS

The issue of efficiently decreasing λ by removing nodes or links from a network has been recently addressed [8], where it was found that when removing a single node, λ is most decreased by removing the node with largest dynamical importance. We consider a closely related issue: given a module S to be added to a network A with given constraints (such as a fixed number of connections), how should the links between the network and module be chosen to either maximize or minimize $\delta\lambda$? Given some set of constraints and staying within our previous assumptions, we will look for matrices (X,Y) that maximize (or minimize) $\delta\lambda$ in Eq. (2). In the examples that follow, it is helpful to assume that the node indices are now ordered such that the left eigenvector entries are in decreasing order: $v_1 \geq v_2 \geq \cdots \geq v_n \geq 0$. In addition, the entries of the right eigenvector are ordered using indices $\{l_i\}$ so that $u_{l_1} \geq u_{l_2} \geq \cdots \geq u_{l_n} \geq 0$. (If A is symmetric, $u = v$ and $l_i = i$.) We will present our optimization methodology for two examples, yet the techniques presented are general and have potential application beyond these particular constraints.

A. Example I: multiple links per module node

In the first example we assume that the number of connections from the original network to the module, x, and the number of connections from the module to the original network, y, are fixed and less than n, the number of nodes in the original network. It is also assumed that all links have strength one (i.e., $X_{ij}, Y_{ij} \in \{0, 1\}$) and multiple links per module node are allowed.

The right hand side of Eq. (2), which approximates the quantity to be maximized, is proportional to $\sum_{i,j}(X^T v)_i^T K_{ij}^S (Y^T u)_j$. This sum can be maximized by (i) finding indices a and b such that $K_{ab} = \max_{ij}\{K_{ij}\}$ and (ii) choosing X and Y to make $(X^T v)_a^T$ and $(Y^T u)_b$ as large as possible. The scalar $(X^T v)_a^T$ is maximized by placing the x ones in the a-th column of X and in positions $1, 2, \ldots, x$ corresponding to the largest values of v, while $(Y^T u)_b^T$ is maximized by placing the y ones in the b-column of Y and in positions $l_1, l_2, \ldots, l_y$ corresponding to the largest values of u. In this way, $(X^T v)_i^T = \delta_{ia}\sum_{j=1}^{x} v_j$ and $(Y^T u)_i^T = \delta_{ib}\sum_{j=1}^{y} u_{l_j}$, where δ_{ij} is Kronecker's delta. The maximum of Eq. (2) is then

$$\delta\lambda^{max} \approx \frac{K_{ab}^s}{\lambda v^T u}\sum_{i=1}^{y} v_i \sum_{j=1}^{x} u_{l_j} \qquad (7)$$

This result implies $\delta\lambda$ may be maximized for the constraints of example (I) by connecting the x nodes with the largest left eigenvector entries v_i in the original network to a single node in the module (having index a), and by also originating all links from the module to the original network from a single module node (having index b) to the y nodes in the original network with the largest entries of the right eigenvector u. For large values of λ/λ_s, the maximum entry of matrix K^s is typically in its diagonal, yielding $a = b$ as shown in Fig.4.

For a heuristic interpretation of this result, let us assume that $A_{ij} \in \{0, 1\}$ and denote $L_i^{o,p} = \sum_j (A^p)_{ij}$ and $L_i^{t,p} = \sum_j (A^p)_{ji}$ as the number of paths of length p originating

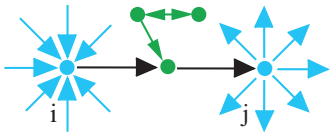

FIG. 4: (Color online) A typical optimal connection for example (I): node i (a *point of contraction* with large left eigenvector entry v_i) points to a node in the module, which in turn points to node j (a *point of expansion* with large right eigenvector entry u_j).

from and terminating at node i, respectively. Thus $L_p = \sum_{ij}(A^p)_{ij}$ is the total number of paths of length p. These quantities satisfy $||L_i^{o,p}||_2^{-1} L_i^{o,p} \to u_i$, $||L_i^{t,p}||_2^{-1} L_i^{t,p} \to v_i$, and $L_{p+1}/L_p \to \lambda$ as $p \to \infty$ [23]. Therefore connecting nodes with large v_i (which receive many paths) to nodes with large u_i (which distribute many paths) will have the largest impact in how L_p grows with p, which determines λ. We therefore define a node i with large v_i as a *point of contraction* and a node j with large u_j as a *point of expansion*. Therefore our result for example (I) is that the effect of the whole module is to act as a bridge from points of contraction to points of expansion in the original network.

B. Example II: one link per node

In the second example we require that, in addition to a fixed number of links x and y with unity strength, no more than one link can be added to a particular node in the module or original network. Because undirected links are equivalent to two links and violate our constraint, it is reasonable (although not necessary) to assume the network and module are directed. To treat this case, we maximize successive terms in Eq. (5). The first term, $v^T X Y^T u/\lambda$, vanishes since any entry of XY^T is nonzero only if there is a module node that has links both to and from the network, a situation which is not allowed by our constraint. Therefore, we maximize the next term, $(X^T v)^T S (Y^T u)/\lambda^2$. As shown in Fig. 5a, let us denote the set of nodes in the original network that point to the module as NO (network outgoing), the set of nodes in the module that are pointed to by the original network as SI (module incoming), the set of nodes in the module that point to the original network as SO (module outgoing), and the set of nodes in the original network that are pointed to by the module as NI (network incoming). Because no node can have more than one new link, there is a one-to-one correspondence between nodes in NO and nodes in SI. The index of nodes in SI will be represented as i_j, where node j in NO points to node i_j in SI. We have $(X^T v)^T_{i_j} = v_j$ if $j \in NO$ and $i_j \in SI$, and 0 otherwise. With a similar notation for SO and NI, we have $(Y^T u)_{m_k} = u_k$ if $m_k \in SO$ and $k \in NI$, and 0 otherwise. It follows that Eq. (5) yields

$$\delta\lambda_2 = \frac{1}{\lambda^2 v^T u} \sum_{j\in NO} v_j \sum_{k\in NI} S_{i_j m_k} u_k.$$

This expression is maximized if S contains a directed complete bipartite graph for disjoint subsets SI and SO such that every node in SI points to every node in SO (see Fig. 5a). Assuming that one can be found, we may set $S_{i_j m_k} = 1$ and look for sets NO and NI that solve

$$\delta\lambda_2^{max} = \left(\frac{1}{\lambda^2 v^T u}\right) \max_{NO\cap NI=\emptyset} \left(\sum_{j\in NO} v_j \sum_{k\in NI} u_k \right). \quad (8)$$

Let $Q = \{i\}_{i=1}^x \cap \{l_i\}_{i=1}^y$. If $Q = \emptyset$ then Eq. (8) is solved by letting $NO = \{i\}_{i=1}^x$ and $NI = \{l_i\}_{i=1}^y$, which yields

$$\delta\lambda^{max} \approx \frac{1}{\lambda^2 v^T u} \sum_{j=1}^{x} v_j \sum_{i=1}^{y} u_{l_i}. \quad (9)$$

As indicated in Fig. 5a, this corresponds to selecting nodes of contraction for NO and nodes of expansion for NI.

The significance of link choices for maximizing $\delta\lambda$ is shown in Fig. 5b, where the module in Fig. 5a was added to

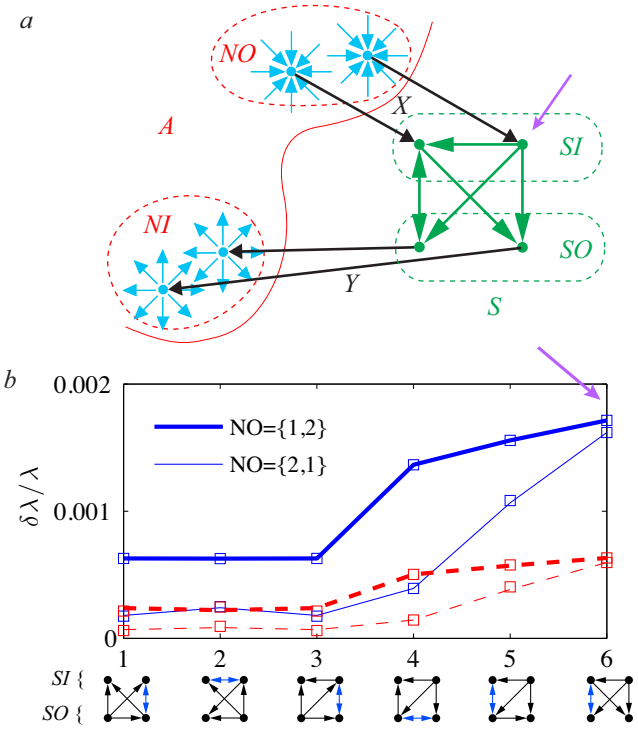

FIG. 5: (Color online) (a) Typical optimal link selections for example (II) for $x = y = 2$. Two points of contraction (NO) link to two module nodes (SI) and the remaining two module nodes (SO) link to two points of expansion (NI). The module also contains a directed complete bipartite graph pointing from SI to SO. (b) Under the restrictions of example (II), the module in Fig. 5a was connected to the neural network for C. elegans using various orientations. Solid lines indicate letting $NI = \{l_1, l_2\}$ and either $NO = \{1, 2\}$ (thick) or $NO = \{2, 1\}$ (thin). Symbols show Eq. (2). Approximating points of contraction (expansion) by nodes with large d^{in} (d^{out}) also offers a decent strategy (dashed).

the neural network of C. elegans with constant node selections for NI and NO but using several module orientations (defined as a particular choice for the disjoint sets SO and SI in the module, and shown in the horizontal axis of Fig. 5b). The solid lines show $\delta\lambda/\lambda$ found numerically using $NI = \{l_1, l_2\}$ and either $N0 = \{1, 2\}$ (thick) or $N0 = \{2, 1\}$ (thin) (see next paragraph for discussion). Symbols indicate $\delta\lambda/\lambda$ found using Eq. (2). One can observe that our maximization strategy for example (II) (Fig. 5a) does in fact maximize $\delta\lambda$ (see orientation 6). An important practical issue is that the eigenvectors may be unknown and require estimation using local information. One can observe that attempting to maximize $\delta\lambda$ using the first-order approximations $v_i \propto d_i^{in}$ and $u_i \propto d_i^{out}$ [8] may also be a good strategy (dashed lines). If necessary, a more refined approximation for the eigenvectors may be sought (e.g., using second-order neighbors [9]).

It is important to note that we have so far neglected higher order terms of Eq. (5) in addressing example (II), which are responsible for the difference in $\delta\lambda$ for the permutation $NO = \{1, 2\}$ or $NO = \{2, 1\}$. Attempting to maximize the third term of the series in Eq. (5) (which is proportional to $v^T X S^2 Y^T u$) while using the nodes of contraction, 1 and 2 (with $v_1 \approx 0.58$ and $v_2 \approx 0.23$), we see that the more-dominant point of contraction (node 1) should link to the module node indicated by the violet arrow in Fig. 5a. (Note that there is a path of length 2 stemming from this node to each node in SO, whereas there are none for the other node in SI.) Unlike permuting nodes in SI, permuting nodes in SO had little effect for this network since $u_{l_1} \approx u_{l_2} \approx 0.23$.

Up to this point we have assumed $Q = \emptyset$, where Q is defined just after Eq. (8); however this may not always be the case. For example, as more links are made (i.e., for increasing x, y), one would expect some nodes to have large values for both v_i and u_i. This may also occur for networks with correlations between d^{in} and d^{out} and, in fact, always occurs for undirected networks where $l_i = i \; \forall \; i$. For these situations, nodes in Q must be allocated to either NO or NI and additional nodes must be selected. Considering the limiting case of an undirected network under the constraints of example (II), maximization of the second-order term in Eq. (5) indicates that we should choose $NO, NI \subset \{i\}_{i=1}^{x+y}$. (Recall that the first-order term is zero by our constraints.) The allocation of these indices should then correspond to successively maximizing the third-, fourth-, ..., kth-order terms until all degrees of freedom have been exhausted. While this strategy of successive maximization does not guarantee the optimal connections (which would require considering all possible links between S and A), it is computationally efficient and ensures a near-optimal solution.

IV. DISCUSSION

While we have presented an efficient strategy for maximizing $\delta\lambda$ for the addition of a module under two examples of constraints, our methodology is general and is thus applicable for many constraints not discussed here. For example, the problem of minimizing $\delta\lambda$ under the constraints of example (II) may be solved by minimizing successive terms of Eq. (5). Heuristically, this corresponds to connecting nodes in A with small values of v_n to the module, and then from the module to nodes in A with small values of u_n. The module should also be oriented so as few links as possible point from SI to SO. We now discuss several applications of using module addition(s) to direct dynamics on networks.

Increasing λ has many real-world applications. For example, because λ relates to the ability of network-coupled oscillatory systems to synchronize [1, 2], one or several module additions to increase λ may be useful to promote synchronization in, for example, a biological process or power grid. Moreover, epidemic thresholds of spreading processes on networks are often dependent on λ^{-1} [3]. Increasing λ can increase the connectivity of a network, improving flow and reducing the epidemic threshold. This may be useful, for example, if one wants to improve communication over a social network or routing-system. The related problem of percolation on networks (where nodes and/or links are randomly removed) is also related to λ^{-1} [4]. Increasing λ can increase a network's robustness against network degradation under failure, blackout, jamming, or attack.

For other dynamical systems, it is beneficial to have a small value for λ. For example, the instability of equilibria for a system of network-coupled ODE's (e.g., interactions in a metabolic network) is related to the largest eigenvalue of a weighted adjacency matrix defined in terms of the system's Jacobian [5] (i.e., if $\lambda < 1$, then the equilibria are stable). When the eigenvalue of the Jacobian matrix with largest magnitude is real and well separated from the bulk of the spectrum, our method is applicable. For example, besides choosing appropriate link weights to keep $\delta\lambda$ small, choosing optimal connections and module orientation (as shown in Sect. III) may also aid in preserving the stability of equilibria for a system undergoing modification. Such analysis may be relevant, for example, in understanding the formation of ecological communities for which the largest eigenvalue of the system's Jacobian has already been suggested to guide the network's evolution under species additions and subtractions [15].

Future applications of our results are also not limited to network dynamics for which the dependency on λ is currently well-established. For example, minimizing $\delta\lambda$ for a module addition may present an effective strategy for minimizing global effects during a network modification. Possible applications may include aiding the in development of systems-level drug design by indicating candidate drug targets that are less invasive (e.g., nodes with middle-valued degrees are typical [12]). The importance of developing mathematical approaches for this promising field are often mentioned [13]. Another open question is the implications of our results on the prevalence of subgraph motifs, which have been proposed to be the basic building blocks of biological networks [16]. In contrast to several studies showing that global dynamics of a system can depend on the structure of subgraph motifs [17], our results suggest that "how" motifs are connected in the network may be as important as their structure.

The work of D. T. and J. G. R. was supported by NSF Grant No. DMS-0908221.

[1] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. E **71**, 036151 (2005); Physica D **224**, 114 (2006); Phys. Rev. Lett. **96**, 254103 (2006).

[2] J. G. Restrepo, E. Ott, and B. R. Hunt, Chaos **16**, 015107 (2005).

[3] Y. Wang *et. al.*, In SRDS, 25-34 (2003); D. Chakrabarti *et. al.*, ACM Trans. on Infor. and Syst. Sec. **4**, 13 (2008); P. Van Mieghem, J. S. Omic, and R. E. Kooij, IEEE ACM Trans. on Net. **17**(1), 1 (2009).

[4] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. Lett. **100**, 058701 (2008); B. Bollobas *et. al.*, The Annals of Prob. **38**, 150 (2010); D. Taylor and J. G. Restrepo, in preparation.

[5] R. May, Nature London **238**, 413 (1972); S. Sinha, Physica A **346**, 147 (2005); J. Feng, V. K. Jirsa, and M. Ding, Chaos **16**, 015109 (2006); M. Brede and S. Sinha, e-print arXiv:cond-mat/0507710. Our method is applicable to systems where the eigenvalue of the Jacobian matrix with largest magnitude is real and well separated from the bulk of the spectrum.

[6] A. Pomerance *et. al.*, PNAS **106**, 8209 (2009).

[7] D. B. Larremore, W. L. Shew, J. G. Restrepo, Phys. Rev. Lett. **106**, 058101 (2011).

[8] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. Lett. **97**, 094102 (2006).

[9] A. Milanese, J. Sun, and T. Nishikawa, Phys. Rev. E **81**, 046112 (2010).

[10] S. Chauhan, M. Girvan, and E. Ott, Phys. Rev. E **80**, 056114 (2009).

[11] A. K. Naimzada, S. Stefani, and A. Torriero, *Networks, Topology and Dynamics: Theory and Applications to Economics and Social Systems*, Springer-Verlag (2009).

[12] T. Hase *et. al.*, PloS Comp. Biol. **5**(10), e1000550 (2009).

[13] H. Kitano, Nature Rev. Drug Disc. **6**, 202 (2007); G. R. Zimmermann, J. Lehár, and C. T. Keith, Drug Disc. Today **12**, 34 (2007); H. Ma and I. Goryanin, Drug Disc. Today **13**, 402 (2008).

[14] S. Sinha and S. Sinha, Phys. Rev. E **74**, 066117 (2006).

[15] J. A. Drake, J. Theor. Biol. **147**, 213 (1990); C. C. Wilmers, S. Sinha, and M. Brede, Oikos **99**, 363 (2002).

[16] R. Milo *et. al.*, Science **303**, 1538 (2004); U. Alon, Nature Rev. Gene. **8**, 450 (2007).

[17] V. P. Zhigulin, PRL **92**, 238701 (2004); R. J. Prill *et. al.*, PLoS Biol. **11**, e343 (2005); P. Kaluza *et. al.*, Phys. Rev. E **75**, 015101 (2007); Chaos **18**, 026113 (2008);

[18] R. B. Bapat and T. E. S. Ranghavan, *Nonnegative Matrices and Applications*, Cambridge Univ. Press (1997); P. Van Mieghem, *Graph Spectra for Complex Networks*, Cambridge Univ. Press (2011).

[19] J. G. White *et. al.*, Phil. Trans. R. Soc. London **314**, 1 (1986).

[20] L. A. Adamic and N. Glance, in *WWW-2005 Workshop on the Weblogging Ecosystem*, May 10, 2005, Chiba, Japan.

[21] D. Bu *et. al.*, Nucleic Acids Research, **31**, 2443 (2003).

[22] D. L. Nelson, C. L. McEnvoy, and T. A. Schreiber (1998). Obtained from http://www.usf.edu/FreeAssociation/.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations, 3rd Ed.*, Johns Hopkins University Press, (1999).