

Weighted Betweenness and Algebraic Connectivity

JAVIER MARTÍN HERNÁNDEZ*

Faculty of Electrical Engineering, Mathematics, and Computer Science
Delft University of Technology, P.O. Box 5031, 2600 GA Delft

*Corresponding author: J.MartinHernandez@tudelft.nl

ZONGWEI LI

Faculty of Electrical Engineering, Mathematics, and Computer Science
Delft University of Technology, P.O. Box 5031, 2600 GA Delft

Zongwei.Li@tudelft.nl

AND

PIET VAN MIEGHEM

Faculty of Electrical Engineering, Mathematics, and Computer Science
Delft University of Technology, P.O. Box 5031, 2600 GA Delft

P.F.A.VanMieghem@tudelft.nl

[Received on 6 February 2014]

One of the better studied topology metrics of complex networks is the second smallest eigenvalue of the Laplacian matrix of a network's graph, referred to as the algebraic connectivity μ_{N-1} . This spectral metric plays a decisive role in synchronization of coupled oscillators, network robustness, consensus problems, belief propagation, graph partitioning, and distributed filtering in sensor networks. However, computing the graph spectra is computationally slow and its convergence greatly depends on the topology, thus a number of lower bounds have been proposed over the years in order to find good approximations. To date, the closest bound is the one proposed by Rad *et al.* [21] in 2009. The current paper proposes new approximations for the algebraic connectivity based on three variations of the betweenness centrality, a popular centrality score often used in social studies to characterize the importance of a node or link within a network. Based on numerical and a partly analytic analysis, we show that our approximations provide accurate lower bounds for the algebraic connectivity for a wide range of graphs, including random, power-law, small-world, and lattice graphs. In particular, we numerically show that the *average weighted Brandes* betweenness can be treated as a lower bound for large enough networks, which greatly improves state-of-the-art bounds.

Keywords: algebraic connectivity; betweenness; lower bound.

1. Introduction

The so called betweenness is one of the most frequently studied network metrics. Originally introduced by Freeman [11], the betweenness plays a vital role at identifying communities in social networks, as it quantifies the number of times an element (node or link in the network) acts as a relay along the shortest path between other pairs of nodes. In other words, assuming that all nodes in a network are exchanging information with each other, the betweenness of an element measures the fraction of pairwise conversations that cross such an element. Also in communication networks, betweenness is viewed

as a measure of network robustness, that tells us how many paths will stretch when such an element is removed from the network. For example, scale free networks display a surprisingly high degree of vulnerability against betweenness attacks, i.e. attacks targeting elements with high betweenness score [25]. The average path distance increases dramatically upon surgical node removal, which causes the network to rapidly break into isolated clusters. For instance, in a social network of injecting drug users, the removal of only 1.5% high betweenness drug dealers causes up to 50% network fragmentation [9]. Potential applications of betweenness include the study of epidemic behaviors in the world wide web and transportation networks [19] [28], and the characterization of cascading failures due to complex interactions among interdependent networks [6, 15, 29].

The graph spectrum or eigenvalue decomposition [27], which has already been studied for a century, offers a powerful way to represent a network. Probably due to the lack of a tangible meaning, the graph spectrum remains elusive, yet the properties of the largest eigenpairs have been thoroughly studied [12, 18, 24, 28]. The spectrum of a graph is an $\mathbb{R}^{N^2} \rightarrow \mathbb{R}^{N^2}$ algebraic transformation that offers an alternative representation of a graph with N nodes, such that the axis system is sorted by the principal directions, specified by the graph's eigenvectors and corresponding eigenvalues. Fiedler [10] first suggested in 1975 that the eigenvector associated to the second smallest eigenvalue (the so-called Fiedler eigenvector) could be used to find an approximate solution to the graph partitioning problem. Since then, a wealth of efficient heuristic spectral methods have been proposed to cope with the intrinsic difficulty of clustering a graph. Emerging partitioning algorithms provide partitions that account for the global graph structure, as opposed to the local vicinities (such as the nodal degree). In particular, the second smallest eigenvalue of the Laplacian matrix, (called by Fiedler the *algebraic connectivity*) proves to be strongly related to synchronization dynamics and graph robustness [14, 16, 19, 23]. Graphs with high algebraic connectivity are harder to disconnect, and synchronize quicker. In particular, if the algebraic connectivity of a graph reaches zero, the graph is split in more than one connected component [27], i.e. a dynamic process on the graph will never fully synchronize.

Albeit the algebraic connectivity and the betweenness seem to tell two different stories, they are closely related. From a data-communication point of view, link betweenness can be seen as a measure of congestion. For example, when a data link with high betweenness is subject to failure, a high percentage of routes are bound to be disrupted, and parts of the network may even become isolated. In the same way, networks with high algebraic connectivity prove to be harder to disconnect. At a first glance, the issues of congestion and disconnectivity may seem distant. However, one could argue that the betweenness and the algebraic connectivity are inversely proportional, because the more bottlenecks a network has, the more likely disconnectivity occurs. This intuition proves to be right: Comellas, and Rad *et al.* [7, 21] proved that the algebraic connectivity is lower bounded by the reciprocal of the link betweenness. This novel bound not only underlines the tight relation between these two metrics, but also proves to be the most accurate lower bound for the algebraic connectivity known to date. This paper further investigates this relation, in the search for more accurate answers, tighter bounds, and better interpretations for spectral metrics.

The remainder of the paper is structured as follows. Section 2 introduces the necessary notation and a bound for the algebraic connectivity as a function of the weighted link betweenness. Section 3 studies the relation between the two metrics, by means of theoretical results. Section 4 presents an extensive numerical analysis based on classic graph models. Conclusions are drawn in Section 5.

2. Definitions and Theory

2.1 Graph Notation

A graph $G(\mathcal{N}, \mathcal{L})$ consists of a set \mathcal{N} of N nodes and a set \mathcal{L} of L links. Connections among the set of nodes are characterized by an adjacency matrix A . The Laplacian matrix $Q = \Delta - A$, in which Δ is the degree diagonal matrix of graph G , has eigenvalues ranked as $\mu_N = 0 \leq \mu_{N-1} \leq \dots \leq \mu_1$.

The algebraic connectivity, denoted as μ_{N-1} , is the smallest non-zero eigenvalue of Q , which is greater than zero if and only if the graph G is connected. From a robustness standpoint, the magnitude of μ_{N-1} reflects how many nodes should be removed in order to disconnect a graph [8]. The larger the algebraic connectivity is, the harder a graph is to disconnect. To date, the problem of how the structural properties of a network influence the convergence and stability of its synchronized states has been extensively investigated and discussed, both numerically and theoretically [2, 3], with special attention given to networks of coupled oscillators [1, 14, 20]. In general, networks with high algebraic connectivity are desirable; these networks spread information faster and are more resilient to random failures.

2.2 Betweenness and Weighted Betweenness

The betweenness of a link B_l (or node B_n) is defined as the total number of shortest paths that traverse such link l (or node n),

$$B_l = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} 1_{\{l \in \mathcal{P}(i,j)\}} \quad (2.1)$$

where $1_{\{x\}}$ is the indicator function and equal to 1 if the condition x is true, and where the shortest path $\mathcal{P}(i, j)$ between two nodes i and j is an ordered sequence of links, such that (a) two consecutive links are incident to the same node, and (b) no links are repeated. The node betweenness proposed by Freeman [11] has been heavily used in the past as a centrality measure in social networks.

Although the subset of links $\mathcal{P}(i, j)$ is generally defined as the *shortest* path between the node pair (i, j) , the shortest path is not necessarily unique, because there may exist multiple paths with equal hop count. Hence, the path $\mathcal{P}(i, j)$ in (2.1) may not be uniquely defined, causing B_l to be undetermined. We solve this indeterminacy by introducing the set of all shortest paths $S_{\mathcal{P}(i, j)}$

$$S_{\mathcal{P}(i, j)} = \{\mathcal{P}_1(i, j), \mathcal{P}_2(i, j), \dots, \mathcal{P}_{\varphi(i, j)}(i, j)\} \quad (2.2)$$

where $\varphi(i, j)$ is the number of different shortest paths between the node pair (i, j) . The set $S_{\mathcal{P}}$ allows us to conceive three variations of link betweenness, which we name *Full* link betweenness $B_l^{(F)}$, *Empty* link betweenness $B_l^{(E)}$, and *Brandes* link betweenness $B_l^{(B)}$. The latter is defined as

$$B_l^{(B)} = \frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \frac{1}{\varphi(i, j)} \sum_{Z \in S_{\mathcal{P}(i, j)}} 1_{\{l \in Z\}}$$

where $\frac{1}{\varphi(i, j)} \sum_{Z \in S_{\mathcal{P}(i, j)}} 1_{\{l \in Z\}}$ represents the fraction of shortest paths between nodes i and j that cross link l . The Brandes Betweenness $B_l^{(B)}$ is the *de facto* standard for betweenness in most scientific publications. It was first introduced by Freeman [11] and further studied extensively as a node centrality metric in the last decades [5]. The contribution to link l 's betweenness of a node pair (i, j) is the proportion of shortest paths that cross the link l over the total number of existing shortest paths $\varphi(i, j)$, as illustrated in Figure 1. This value is then added over all $\binom{N}{2}$ node pairs to find the betweenness $B_l^{(B)}$ of link l . The

Brandes betweenness $B_l^{(B)}$ can be interpreted as an automatic load balancing, where a communicating node pair (i, j) splits its bandwidth fairly over all shortest paths between i and j .

As opposed to the load balancing policy assumed in *Brandes* betweenness, any node may also choose to transmit its full load through a single path. Given a link l , a random path selection may lead to a worst case scenario where all node pairs $\binom{N}{2}$ transmit their full load over l . This maximum offered load is quantified by the *Full* betweenness of link l , or $B_l^{(F)}$, which represents the highest possible load that link l may offer (assuming all-to-all node transmission). On the contrary, one may contemplate the opposite scenario for link l , where every node pair has chosen an alternate route not including l , whenever possible. The *Empty* betweenness $B_l^{(E)}$ of link l represents the lowest possible load that link l may offer (again, assuming all-to-all node transmission). The *Full* and *Empty* betweenness are defined as,

$$\begin{aligned} B_l^{(F)} &= \frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \max_{Z \in S_{\mathcal{P}}(i,j)} \mathbf{1}_{\{l \in Z\}} \\ B_l^{(E)} &= \frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \min_{Z \in S_{\mathcal{P}}(i,j)} \mathbf{1}_{\{l \in Z\}} \end{aligned}$$

respectively. Where $\max_{Z \in S_{\mathcal{P}}(i,j)} \mathbf{1}_{\{l \in Z\}}$ equals 1 if there is at least one path Z in the set $S_{\mathcal{P}}(i, j)$ that crosses link l , and $\min_{Z \in S_{\mathcal{P}}(i,j)} \mathbf{1}_{\{l \in Z\}}$ equals 1 only if every path in the set $S_{\mathcal{P}}(i, j)$ crosses link l , as illustrated in Figure 1.

We additionally define weighted versions of the *Full* link betweenness $r_l^{(F)}$, the *Brandes* link betweenness $r_l^{(B)}$, and the *Empty* link betweenness $r_l^{(E)}$, as

$$\begin{aligned} r_l^{(F)} &= \frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} h(\mathcal{P}(i, j)) \max_{Z \in S_{\mathcal{P}}(i,j)} \mathbf{1}_{\{l \in Z\}} \\ r_l^{(B)} &= \frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} h(\mathcal{P}(i, j)) \frac{1}{\varphi(i,j)} \sum_{Z \in S_{\mathcal{P}}(i,j)} \mathbf{1}_{\{l \in Z\}} \\ r_l^{(E)} &= \frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} h(\mathcal{P}(i, j)) \min_{Z \in S_{\mathcal{P}}(i,j)} \mathbf{1}_{\{l \in Z\}} \end{aligned} \quad (2.3)$$

where $h(\mathcal{P}(i, j))$ is the hopcount of path $\mathcal{P}(i, j)$. The weighted link betweenness is more closely related to the algebraic connectivity than the raw link betweenness, as we will show in the next section. By definition, $r_l^{(F)} \geq r_l^{(B)} \geq r_l^{(E)}$, and they are lower bounded by 1, because the node pair (l^+, l^-) at both ends of link l always contributes to the *Empty* link betweenness of l .

2.3 Bounds for the Algebraic Connectivity

If x is the eigenvector of the Laplacian matrix Q belonging to μ_{N-1} , and $x(i)$ is the component of x corresponding to the i -th node, then [27, p. 81]

$$\mu_{N-1} = \frac{2N \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2}{\sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{N}} (x(u) - x(v))^2} \quad (2.4)$$

The bound $\mu_{N-1} \geq \frac{1}{\rho N}$ was derived in [27, p. 84], where ρ is the diameter or the maximum hopcount between any pair of nodes of a given graph. This bound can be improved by considering the Cauchy-Schwarz bound [27, p. 84]

$$(x(v) - x(u))^2 \leq h(\mathcal{P}(v, u)) \sum_{l \in \mathcal{P}(v, u)} (x(l^+) - x(l^-))^2 \quad (2.5)$$

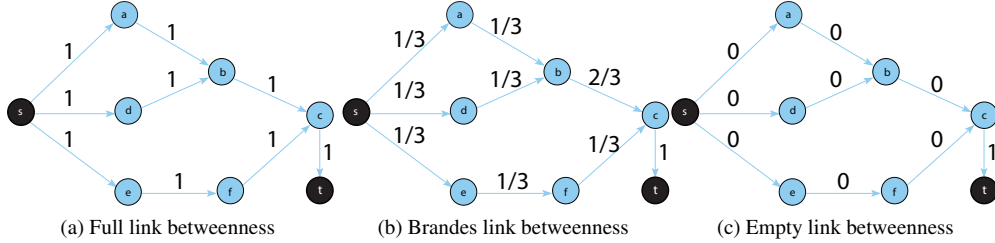


FIG. 1: Contribution of a single pair (s,t) to the Full betweenness (left), the Brandes betweenness (center), and the Empty betweenness (right) of all links. There exist a total of three shortest paths between node s and t : $\mathcal{S}_P(s,t) = \{\{a,b,c\}, \{d,b,c\}, \{e,f,c\}\}$. The Brandes betweenness equally weights each path (upper, middle and bottom paths) by $\frac{1}{3}$; the Full betweenness equally contributes to every link that lies in any shortest path; the Empty betweenness only contributes to the betweenness of link l if and only if a link appears in all shortest paths, link (c,t) in this example.

which holds for any randomly chosen shortest path $\mathcal{P}(u,v) \in \mathcal{S}_{\mathcal{P}}(u,v)$. However, if we were to sum (2.5) over all node pairs, there will exist a set of particular paths that minimizes (or maximizes) the right hand side of the equation. To avoid this indeterminacy, we generalize the previous expression by considering all $\varphi(v,u)$ shortest paths joining the node pair (u,v) by extending (2.5) into

$$(x(v) - x(u))^2 \leq \frac{1}{\varphi(u,v)} \sum_{Z \in \mathcal{S}_{\mathcal{P}}(u,v)} h(Z) \sum_{l \in Z} (x(l^+) - x(l^-))^2 \quad (2.6)$$

where each path Z in the set $\mathcal{S}_{\mathcal{P}}(u,v)$ has the same hopcount, thus $h(Z) = h(\mathcal{P}(u,v))$. Summing (2.6) over all node pairs yields

$$\begin{aligned} \sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{N}} (x(u) - x(v))^2 &\leq \sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{N}} \frac{h(\mathcal{P}(u,v))}{\varphi(u,v)} \sum_{Z \in \mathcal{S}_{\mathcal{P}}(u,v)} \sum_{l \in Z} (x(l^+) - x(l^-))^2 \\ &= \sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{N}} \frac{h(\mathcal{P}(u,v))}{\varphi(u,v)} \sum_{Z \in \mathcal{S}_{\mathcal{P}}(u,v)} \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 \mathbf{1}_{\{l \in Z\}} \\ &= \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 \sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{N}} \frac{h(\mathcal{P}(u,v))}{\varphi(u,v)} \sum_{Z \in \mathcal{S}_{\mathcal{P}}(u,v)} (\mathbf{1}_{\{l \in Z\}}) \end{aligned}$$

Using the definition (2.3) of the weighted Brandes betweenness $r_l^{(B)}$, we obtain

$$\sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{N}} (x(u) - x(v))^2 \leq 2 \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 r_l^{(B)} \quad (2.7)$$

Introducing the last equation into Fiedler's expression (2.4) for the algebraic connectivity yields the lower bound

$$\mu_{N-1} \geq \frac{N \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2}{\sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 r_l^{(B)}} \quad (2.8)$$

In general, the denominator of (2.8) can be bounded as

$$\min_{l \in \mathcal{L}} r_l^{(B)} \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 \leq \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 r_l^{(B)} \leq \max_{l \in \mathcal{L}} r_l^{(B)} \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 \quad (2.9)$$

which simplifies (2.8) down to

$$\mu_{N-1} \geq \frac{N}{\max_{l \in \mathcal{L}} r_l^{(B)}} \quad (2.10)$$

which is the tightest lower bound for the algebraic connectivity known to date, proposed by Rad *et al.* [21]. While the minimum of $\sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2 r_l^{(B)}$ would yield the largest lower bound for μ_{N-1} , we cannot guarantee that it equals $\min_{l \in \mathcal{L}} r_l \sum_{l \in \mathcal{L}} (x(l^+) - x(l^-))^2$, nor that $\frac{N}{\min_{l \in \mathcal{L}} r_l^{(B)}}$ will still lower bound μ_{N-1} . Intuitively though, we may expect that μ_{N-1} can be expressed as a function of the weighted betweenness, thus we are interested in studying the validity of the following nine inequalities:

$$\begin{array}{lll} \text{Empty Betweenness:} & \mu_{N-1} \stackrel{?}{\geq} \frac{N}{\min r_l^{(E)}} & \mu_{N-1} \stackrel{?}{\geq} \frac{N}{E[r_l^{(E)}]} & \mu_{N-1} \stackrel{?}{\geq} \frac{N}{\max r_l^{(E)}} \\ \text{Brandes Betweenness:} & \mu_{N-1} \stackrel{?}{\geq} \frac{N}{\min r_l^{(B)}} & \mu_{N-1} \stackrel{?}{\geq} \frac{N}{E[r_l^{(B)}]} & \boxed{\mu_{N-1} \geq \frac{N}{\max r_l^{(B)}}} \\ \text{Full Betweenness:} & \mu_{N-1} \stackrel{?}{\geq} \frac{N}{\min r_l^{(F)}} & \mu_{N-1} \stackrel{?}{\geq} \frac{N}{E[r_l^{(F)}]} & \boxed{\mu_{N-1} \geq \frac{N}{\max r_l^{(F)}}} \end{array} \quad (2.11)$$

where the *min* and *max* indicators iterate over all links in \mathcal{L} . We know that the two framed inequalities, i.e. the ones based on the maximum Brandes and the maximum Full betweenness, hold for any graph. The purpose of the present article is to study the validity of (2.11), and more generally, the relation between μ_{N-1} and various measures of the weighted betweenness, such as $\min_{l \in \mathcal{L}} r_l^{(F)}$, $\max_{l \in \mathcal{L}} r_l^{(E)}$, and the average $E[r_l^{(B)}] = \frac{1}{L} \sum_{l=1}^L r_l^{(B)}$.

The average Brandes betweenness can be rewritten as

$$E[r_l^{(B)}] = \frac{1}{L} \sum_{l=1}^L r_l^{(B)} = \frac{1}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} h(\mathcal{P}(i, j)) \frac{1}{\varphi(i, j)} \sum_{z \in \mathcal{S}_{\neq}(i, j)} \frac{1}{L} \sum_{l=1}^L 1_{\{l \in \mathcal{Z}\}}$$

and with $\sum_{l=1}^L 1_{\{l \in \mathcal{Z}\}} = h(\mathcal{P}(i, j))$,

$$E[r_l^{(B)}] = \frac{1}{2L} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} h^2(\mathcal{P}(i, j)) = \frac{N(N-1)}{2L} E[H^2] = (N-1) \frac{E[H^2]}{E[D]}$$

where H is the hopcount of an arbitrary path and D is the degree of an arbitrary node in the graph G .

The corresponding inequality $\mu_{N-1} \stackrel{?}{\geq} \frac{N}{E[r_l^{(B)}]}$, which is promising as shown below, becomes

$$\mu_{N-1} \stackrel{?}{\geq} \frac{NE[D]}{(N-1)E[H^2]} > \frac{E[D]}{E[H^2]}$$

where $E[H^2] = \text{Var}[H] + (E[H])^2$. For example, in the complete graph K_N , the variance $\text{Var}[H] = 0$, since the hopcount of any shortest path equals $H = 1$, the average degree $E[D] = N - 1$ and the algebraic connectivity $\mu_{N-1} = N$, so that equality is achieved in the first inequality, whereas $\frac{E[D]}{E[H^2]} = N - 1$.

2.4 Time Complexity

This section briefly exposes the time complexity of both spectral and betweenness computation state-of-the-art algorithms.

2.4.1 Spectral Complexity Computing the exact spectra of a given graph G is equivalent to calculating the eigenvalues and eigenvectors of its corresponding adjacency matrix [13]. The matrix multiplications of the eigenvalue decomposition lead to a time complexity of $O(N^3)$. Heuristic algorithms have been proposed to accelerate the product of sparse matrices, but this upper bound remains immutable.

Computing pseudo-spectra (i.e. approximations instead of exact values) has successfully been approached by a number of heuristics: first, the QR and later Arnoldi-Lanczos (know as IRAM, Iterative Restarted Arnoldi Method [22]), which have been implemented in popular libraries such as LAPACK and Matlab. The advantage of these methods over raw eigenvalue decomposition is two fold. First, they can compute only a portion of the spectra, and second they can approximate the spectra up to a relative error margin, thus speeding the computations by orders of magnitude. The disadvantage of pseudo-spectral algorithms resides in their iterative nature, i.e. the convergence time highly depends on the topology and the desired relative error margin. Using state of the art algorithms, eigenvalues of 200,000 node networks can be computed within one minute [30].

2.4.2 Betweenness Complexity The simplest variant of the link betweenness, i.e. the weighted Brandes betweenness, can be computed in order $O(NL)$. Proposed by Brandes [4], the fastest algorithm known takes advantage of the fact that dependency fractions are propagated at every bifurcation between a source-destination pair. Thus, one only requires a single iteration over all links to compute a single node's contribution to the Brandes betweenness of a node.

The Empty and Full weighted betweenness, as opposed to the Brandes weighted betweenness, (2.11) do not retain the propagation property. Path fractions cannot be merged at bifurcations, thus additional information is needed to know whether the crossing paths belong to the same source-destination pair or not. If not, the algorithm requires up to $O(N)$ additional floating point operations, i.e. the time complexity of the algorithm increases with the number of bifurcations along any shortest path. To date, there is no general expression for the average number of bifurcations, which means that we cannot estimate an average time complexity. The slowest, worst case scenario corresponds to the network where every single node bifurcates into two or more paths, which scales as $O(\log N)$. Hence, the computational complexity of Empty and Full betweenness of small-world networks is upper bounded by $O(NL \log N)$.

To conclude, exact spectral metrics are slower to compute than betweenness metrics in sparse networks (where $L = O(N)$). However, depending on the type of topology, the number of links in the network and the diversity factor of shortest paths, pseudo-spectral algorithms may outperform Full and Empty betweenness algorithms.

3. Analytic Results

In this section we provide exact results for the the Full link betweenness of the complete graph, the complete bipartite graph, the star graph, and the lattice graph. We exclusively focus our analysis on the

Full link betweenness because $r_l^{(F)}$ represents the worst case scenario, i.e. $r_l^{(F)}$ is always greater than both the Brandes $r_l^{(B)}$ and the Empty $r_l^{(E)}$ link betweenness.

3.1 Complete Graph

The complete graph K_N is a graph in which every node pair is connected with a link. In a complete graph, the algebraic connectivity is proven [27] to be $\mu_{N-1} = N$. By definition, each shortest path in a complete graph crosses only a single link, i.e. $S_{\mathcal{P}}(i, j) = \{(i, j)\}$,

$$\min_{l \in L} r_l^{(x)} = \max_{l \in L} r_l^{(x)} = 1$$

for $x \in \{E, B, F\}$. Hence, all inequalities (2.11) are equalities for the complete graph,

$$\mu_{N-1} = N = \frac{N}{\min_{l \in L} r_l^{(F)}}$$

3.2 Complete Bipartite Graph

A complete bipartite graph $K_{m,n}$ consists of two disjoint node sets \mathcal{S}_1 and \mathcal{S}_2 , containing m and n nodes respectively, in which every node in set \mathcal{S}_1 is connected to every node in set \mathcal{S}_2 . The star graph is a special case $K_{1,N-1}$ of the complete bipartite graph, where $m = 1, n = N - 1$, i.e. one set contains just one node and the other set contains $N - 1$ nodes. Because all links in the complete bipartite graph are equivalent, for any link l , it holds that

$$\begin{aligned} \min_{l \in L} r_l^{(F)} = \max_{l \in L} r_l^{(F)} &= \sum_{i \in \mathcal{S}_1} \sum_{j \in \mathcal{S}_2} h(\mathcal{P}(i, j)) 1_{\{l \in \mathcal{P}(i, j)\}} + \frac{1}{2} \sum_{i \in \mathcal{S}_1} \sum_{j \in \mathcal{S}_1} h(\mathcal{P}(i, j)) 1_{\{l \in \mathcal{P}(i, j)\}} \\ &+ \frac{1}{2} \sum_{i \in \mathcal{S}_2} \sum_{j \in \mathcal{S}_2} h(\mathcal{P}(i, j)) 1_{\{l \in \mathcal{P}(i, j)\}} \\ &= 1 + 2(m-1) + 2(n-1) = 2N - 3 \end{aligned}$$

where $m + n = N$. The first sum accounts for all the shortest paths with length one, i.e. the contribution of node pairs belonging to different sets. The second and third sums account for all the shortest paths with length two. We know [27] that the algebraic connectivity of a complete bipartite graph equals $\mu_{N-1} = \min\{m, n\} \geq 1$, thus, for $N > 2$,

$$\mu_{N-1} \geq \frac{N}{\min_{l \in L} r_l^{(F)}} = \frac{N}{2N - 3}$$

3.3 Square Lattice Graph

A square lattice $L_{n \times n}$ of size n is a bounded $n \times n$ two-dimensional grid containing $N = n^2$ nodes, where all nodes have four incident links to their four closest neighbours, except those around the four boundaries.

Let us define the *elementary square* of link l as the region between rows $k, k + 1$ and columns $m, m + 1$, where k and m are the row and column coordinates of the link l , respectively, and n is the

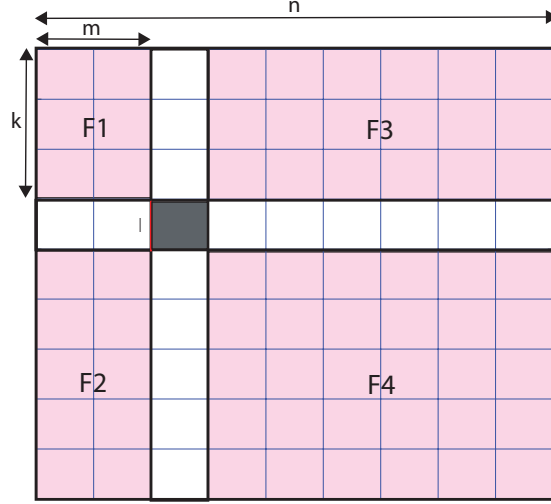


FIG. 2: This figure illustrates the computation of the Full link betweenness of a given link l in a square lattice. The shadowed region represents the elementary square of link l . The Full link betweenness of link l equals the number of covering fields that fully overlap the elementary square.

number of rows (and columns). Define the *covering field* between two nodes in the lattice as the region covered by a rectangle that contains the nodes in diagonally opposite corners. Then, by definition, the full weighted betweenness of link l equals the total number of *covering fields* that overlap the *elementary square*, over all node pairs. The computation of the Full betweenness of a link l is illustrated in Figure 2, which can be expressed as

$$\begin{aligned}
 r_l^{(F)} &= 2 \sum_{i \in F_1} \sum_{j \in F_4} ((y_j - y_i) + (x_j - x_i)) + \sum_{i=1}^m \sum_{j=m+1}^n (y_j - y_i) + 2 \sum_{i \in F_1} \sum_{j \in F_2} ((y_j - y_i) + (k + 1 - x_i)) \quad (3.1) \\
 &= \sum_{i=0}^{n-2} \sum_{j=0}^{n-1} (1 + i + j) + (m - 1) \sum_{i=0}^{n-k-2} \sum_{y_j=m+1}^n (y_j - y_i) + 2(k - 1) \sum_{i=0}^{n-2} \sum_{y_j=n+1}^n ((y_j - y_i) + (k + 1 - x_i)) \quad (3.2)
 \end{aligned}$$

where F_1, F_2, F_3, F_4 are 4 covering fields that lie in the four corners of the elementary square rooted at link l ; the indices i and j represent the start and end nodes of a given path, respectively; and y_i, x_i are the column and row of node i . The link that minimizes (3.2) leads to the expression

$$\min r_l^{(F)} = \sum_{i=0}^{n-2} \sum_{j=0}^{n-1} (1 + i + j) = \frac{N(N-1)}{2} + 2N(N-1)^2 \quad (3.3)$$

The algebraic connectivity of the square lattice $L_{n \times n}$ (see [26]) can be expressed as

$$\mu_{N-1}(L_{n \times n}) = 2 \left(1 - \cos \left(\frac{\pi}{N} \right) \right) \quad (3.4)$$

By comparing (3.4) to (3.3), we can establish that inequality $\mu_{N-1} \geq \frac{N}{\min r_l^{(F)}}$ only holds for values of N smaller than 324. In other words, the minimum Full Betweenness can be used as a lower bound for the algebraic connectivity only in square lattices with less than 18×18 nodes.

4. Simulation results

4.1 Small graphs

In order to test whether the Brandes, Full and Empty weighted betweenness are good approximations for the algebraic connectivity, we exhaustively computed μ_{N-1} , $r_l^{(F)}$, $r_l^{(B)}$, and $r_l^{(E)}$ for every existing connected undirected *small* graph, i.e. graphs up to 10 nodes. The full set of non-isomorphic connected graphs $C(N)$ with N nodes can be swiftly generated with McKay's algorithm [17]. We choose to exclude disconnected graphs because the algebraic connectivity of such graphs is zero ($\mu_{N-1} = 0$), which is an uninteresting case study. The results of our exhaustive exploration of the graph are summarized in Figure 3, Table 1, and Appendix A.

The **average** values of the Brandes, Full and Empty betweenness provide bounds that hold up to 43%, 95% and 19% of the graph space, respectively. Each of the three betweenness shows a particular behavior: first, the accuracy of the average Empty betweenness declines as N increases, reaching 0.0001% for $N = 10$. On the other hand, the accuracy of the average Full betweenness improves as N increases: it starts at 66% accuracy for all graphs with $N = 5$ nodes, and reaches 94% accuracy for $N = 10$. Lastly, Brandes betweenness shows a double regime as illustrated in Figure 3a: for $N < 7$, the accuracy declines for increasing N ; for $N \geq 8$ the accuracy rises for increasing N .

Similar to the average betweenness, the **minimum** betweenness shows varied behaviors. The accuracy of both the minimum Empty and Brandes betweenness steadily decline, whereas the minimum Full betweenness shows a double regime, i.e decreasing for $N < 8$, and increasing for $N \geq 8$. Surprisingly, the **maximum** Empty betweenness is outperformed by the average Full betweenness even for small networks. This means that even for small values of N , an average-based estimator outperforms a worst-case estimator.

To sum up, only two of the seven weighted betweenness metrics in (2.11), i.e. the average Full and the maximum Empty betweenness, are fair approximations for the algebraic connectivity of small graphs. These approximations provide us with two inequalities that hold for more than 58% of the graph space.

4.2 Large graphs

This section studies the relation between the maximum, average, and minimum of the Full Betweenness $r_l^{(F)}$, the Brandes Betweenness $r_l^{(B)}$, the Empty Betweenness $r_l^{(E)}$ and the algebraic connectivity for large undirected graphs, i.e. graphs with $N \gg 10$.

Given that obtaining the betweenness of every non-isomorphic connected graph with more than $N = 10$ nodes is computationally unfeasible, we resorted to a random sample of four statistically significant graph models. The four proposed models are the Erdős-Rényi random graphs, the Barabási-Albert *scale-free* graphs, the Watts-Strogatz, and the Lattice graph. These models mimic common features of real-world like graphs, such as the *small-world* and the *robust-yet-fragile* properties. The input parameters for our graph models were tuned such that the resulting Erdős-Rényi random graph and Barabási-Albert Free-scale graphs all had an average degree of $E[D] = 8$, and Watts-Strogatz and Lattice have an average degree of $E[D] = 4$. This multiple choice arises from the fact that most Erdős-Rényi graphs with average

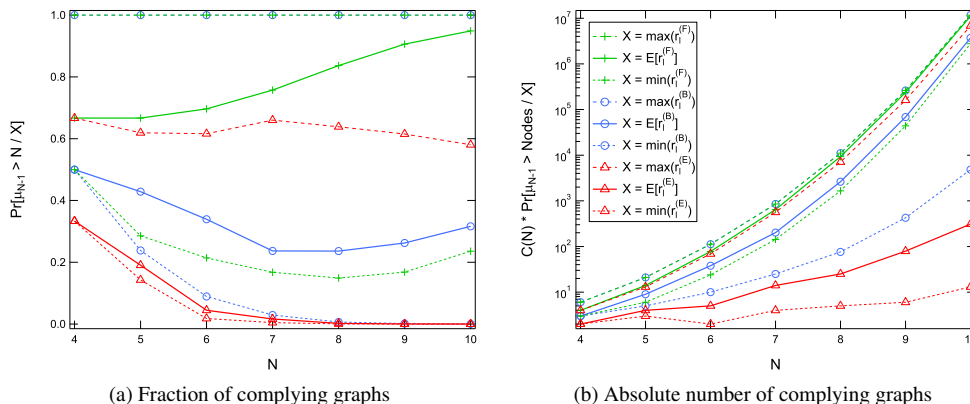


FIG. 3: Number of graphs that obey the indicated inequalities for networks sizes in the range [4,10], both as a percentage and as an absolute number (left image and right image, respectively). The top dashed curve represents $C(N)$, i.e. the total number of non-isomorphic connected graphs with N nodes. The legend applies to both images.

degree 4 are disconnected.

For each model, we computed two performance metrics: (a) the average μ_{N-1} and betweenness values as illustrated in Figure 4; and (b) the number of graphs that obey each of the nine inequalities, as depicted in Table 1. Simulations were iterated 10^5 times per sample point, for network sizes N in the range [25, 1014]. The following subsections provide a quantitative analysis of each of the four graph types under study.

4.2.1 Square Lattice Graphs In agreement with our analytic results (3.1) and (??), the algebraic connectivity and the minimum Full betweenness curves intersect at $N = 324$, as illustrated in Figure 4a. For values of N slightly below 324, the minimum Full betweenness provides the tightest lower bound; for values of N greater than 324, the average Brandes betweenness becomes the closest lower bound.

In contrast to other graph models, none of the Empty link betweenness inequalities provide a reliable lower bound. This is due to the high path diversity observed in grid structures, in which there is an absence of bottlenecks. By the same token, the Full link betweenness tends to be a better bound in the Square Lattice than in the other three models.

4.2.2 Erdős-Rényi random graphs Numerical results for Erdős-Rényi are illustrated in Figure 4b, which shows the three closest weighted betweenness estimators to be the average Brandes $E[r_l^{(B)}]$, the minimum Full $r_{l\min}^{(F)}$ and the maximum Empty betweenness $r_{l\max}^{(E)}$, i.e. the three curves immediately below μ_{N-1} . However, the values depicted in Figure 4b represent only averaged values, i.e. there exist outlier graphs for which $\mu_{N-1} < \frac{N}{E[r_l^{(B)}]}$, i.e. inequality $\mu_{N-1} \geq \frac{N}{E[r_l^{(B)}]}$ does not comply. The number of complying cases is depicted in the first three columns of Table 1. We can see that less than 90% of average Brandes samples provide a valid lower bound for the networks smaller than ≈ 729 nodes. On

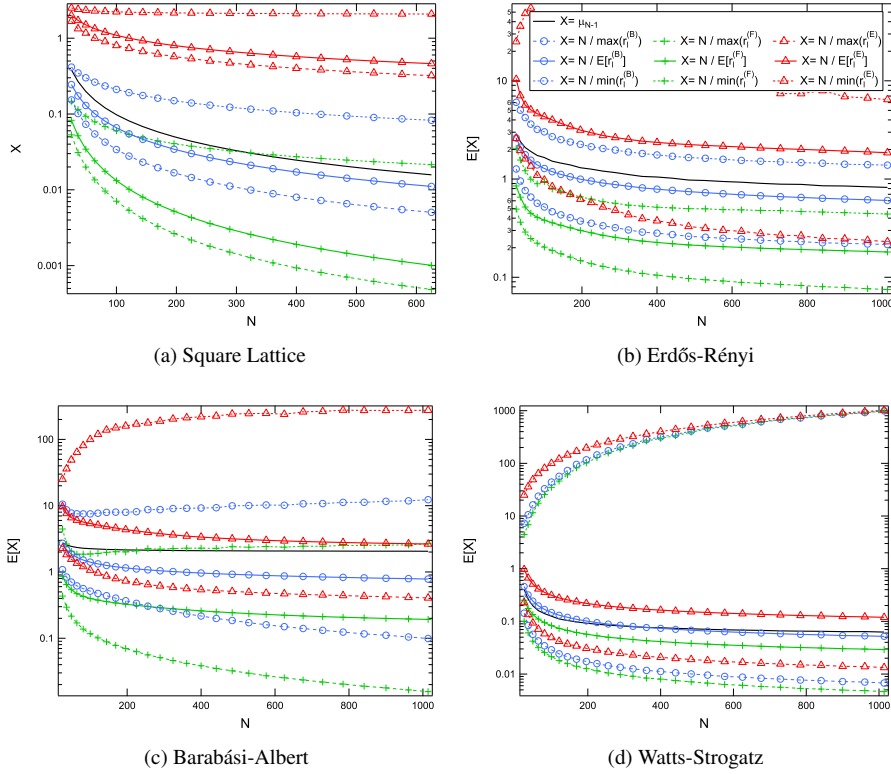


FIG. 4: (Color online) Average values for the algebraic connectivity (black curve), Full, Brandes, and Empty betweenness (green, blue, and red curves, respectively) of 10^4 sampled networks. The simulated graphs are: Square Lattice (top left), Erdős-Rényi (top right) and Barabási-Albert (bottom left), and Watts-Strogatz (bottom right). The curves lying the closest to the black line (i.e. the algebraic connectivity) provide the better bounds. The legend applies to the four images.

the other hand, the accuracy of $r_{l\min}^{(F)}$ and $r_{l\max}^{(E)}$ -based bounds is always greater than 90% irrespectively of the network size.

The rest of the curves, besides the mentioned three, pose uninteresting case studies. These curves show either 0% or 100% compliance, because they either never hold, or lie below well-known lower bounds, respectively. Among the average Brandes $E[r_i^{(B)}]$, the minimum Full $r_{l\min}^{(F)}$ and the maximum Empty betweenness $r_{l\max}^{(E)}$, the latter two provide lower bounds with the highest compliance for Erdős-Rényi random graphs. Nonetheless, as $N \rightarrow \infty$, the average Brandes betweenness seems to approach 100% compliance, while still being the closest lower bound.

4.2.3 Barabási-Albert Free-scale graphs Simulation results are illustrated in Figure 4c. We have simulated 10^4 iterations of Barabási-Albert Free-scale graphs from 25 to 1014 nodes.

Table 1: The top table displays the percentage of small graphs that obey the nine inequalities, for all graphs with small $N \leq 10$. The bottom table displays, for large $N \gg 10$, the percentage of Erdős-Rényi, Barabási-Albert and Watts-Strogatz graphs that obey three handpicked inequalities: maximum Empty, average Brandes, and minimum Full betweenness. The six inequalities we left out are either $\approx 100\%$ accurate, or $\approx 0\%$ accurate.

Small graphs ($N \leq 10$)									
N	Empty			Brandes			Full		
	min	avg	max	min	avg	max	min	avg	max
5	14.2%	19.0%	61.9%	23.8%	42.8%	100%	28.5%	66.6%	100%
6	1.7%	4.4%	61.6%	8.9%	33.9%	100%	21.4%	69.6%	100%
7	0.4%	1.6%	66.0%	2.9%	23.6%	100%	16.7%	75.7%	100%
8	0.04%	0.2%	63.8%	0.6%	23.6%	100%	14.8%	83.6%	100%
9	0.002%	0.03%	61.5%	0.1%	26.2%	100%	16.8%	90.5%	100%
10	0.0001%	0.002%	58.0%	0.04%	31.6%	100%	23.6%	94.8%	100%

Large graphs ($N \gg 10$)									
N	Erdős-Rényi			Barabási-Albert			Watts-Strogatz		
	$r_{l\max}^{(E)}$	$E[r_l^{(B)}]$	$r_{l\min}^{(F)}$	$r_{l\max}^{(E)}$	$E[r_l^{(B)}]$	$r_{l\min}^{(F)}$	$r_{l\max}^{(E)}$	$E[r_l^{(B)}]$	$E[r_l^{(F)}]$
49	94.6%	74.7%	91.1%	99.8%	100%	77.5%	99.8%	8.1%	97%
144	100%	70.6%	95.3%	100%	100%	74.4%	100%	17.3%	97.3%
289	100%	63.8%	98.6%	100%	100%	58.8%	100%	41.4%	98.6%
484	100%	88.3%	99.0%	100%	100%	46.7%	100%	68.9%	99.0%
729	100%	94.7%	98.6%	100%	100%	35.6%	100%	78.4%	98.8%
1014	100%	96.4%	99.4%	100%	100%	28.4%	100%	83%	99.1%

Similarly to the Erdős-Rényi scenario, the three closest weighted betweenness estimators are the average Brandes $E[r_l^{(B)}]$, the minimum Full $r_{l\min}^{(F)}$ and the maximum Empty betweenness $r_{l\max}^{(E)}$. However, their behaviors differ considerably from those in the Erdős-Rényi case, as seen in the three middle columns of Table 1. For extremely small values of $N < 100$, both the maximum Empty and average Brandes betweenness show a compliance greater than 99%. However, as N increases, the accuracy of the $r_{l\max}^{(E)}$ -based bound decreases all the way down to 28.4% for $N = 1014$; where average Brandes remains at 100% compliance.

The rest of the curves display 100% compliance, whereas all curves above the algebraic connectivity display near 0% compliance. Given that Barabási-Albert graphs with less than $N \approx 500$ nodes do not possess *scale-free* features, we can conclude that for scale-free graphs, both the maximum Empty and the average Brandes betweenness provide the better bounds for the algebraic connectivity.

4.2.4 Watts-Strogatz Graphs Simulation results related to Watts-Strogatz are illustrated in Figure 4d and the three rightmost columns of Table 1. The three closest weighted betweenness estimators are: the average Brandes $E[r_l^{(B)}]$, the average Full $r_{l\min}^{(F)}$ and the maximum Empty betweenness $r_{l\max}^{(E)}$. The main difference between the Watts-Strogatz model and the three previous graph models lies in the Full betweenness. For Erdős-Rényi and Barabási-Albert graphs, the maximum Full betweenness proves to be a reasonably good lower bound for small values of N . However, the algebraic connectivity of Watts-Strogatz is far removed from the maximum Full betweenness, always lying at 0% compliance.

Similarly to the random graph scenario, the average Brandes betweenness is a poor lower bound for $N < 700$, but it shows a steady increase in compliance as N increases, as shown in Table 1. We expect that, if this tendency remains as N grows larger, the average Brandes will have near 100% compliance while providing the most accurate bound.

5. Conclusions

In this paper, we found a new estimate for the algebraic connectivity based on the averaged weighted Brandes betweenness over all network links,

$$\mu_{N-1} \gtrsim \frac{N}{E[r_i^{(B)}]} \approx \frac{E[D]}{E[H^2]} \quad (5.1)$$

For large enough N , simulations show that our approximation (5.1) is exceedingly sharp and can be treated almost as a lower bound, which is nearly an order of magnitude better than the closest lower bound (2.10) known to date. Unfortunately, strict inequality in (5.1) does not hold and even for large N , we are unable to prove asymptotic inequality, although asymptotic relations for the hopcount H exist. We showed the validity of our lower bound (2.10) by means of numerical simulations, where it holds for 96% of Erdős-Rényi graphs, 100% of Barabási-Albert graphs, 83% of Watts-Strogatz graphs, and 100% of Square Lattice graphs with more than 1014 nodes. Our simulations suggest that the bound (2.10) approaches 100% compliance as $N \rightarrow \infty$. In a case by case scenario, alternative definitions of betweenness (based on the Empty betweenness, and the Full betweenness) may provide even higher quality algebraic connectivity estimators. For instance, in Erdős-Rényi graphs, inequality $\mu_{N-1} \leq \frac{N}{\max r_i^E}$ holds for 100% of the sampled graphs, while still being tighter than the state-of-the-art analytic bounds.

The present study brings us one step closer to understanding the graph spectrum as an alternative representation of a graph. It is well known that the algebraic connectivity μ_{N-1} plays an important role in, among others, synchronization of coupled oscillators and network robustness to disconnectivity. In addition to these applications, one may understand the algebraic connectivity to be the average length-scaled distance between all node pairs. Thus, the effect of a given path on synchronization and robustness processes is directly proportional to the length of such path, i.e. short paths have less of an influence on synchronization than longer paths.

Computationally, our work increases the amount of graphs for which computing the algebraic connectivity is timely feasible. A substantial advantage of hopcount algorithms over spectral approximations is that hopcount-based processes can be parallelized, thus a raid of computers is likely to compute the Brandes betweenness multiple times faster than the algebraic connectivity. In addition, computing any variation of the weighted betweenness also yields the all-to-all node distances matrix, which proves useful in the optimization of routing protocols.

REFERENCES

1. Acebrón, J. A., Bonilla, L. L., Pérez-Vicente, C. J., Ritort, F. & Spigler, R. (2005) The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.*, **77**, 137–185.
2. Atay, F. M., Biyikoglu, T. & Juergen, J. (2006) Synchronization of networks with prescribed degree distributions. *IEEE Transactions on Circuits and Systems-I*, **53** (1), 92–98.
3. Barahona, M. & Pecora, L. M. (2002) Synchronization in Small-World Systems. *Phys. Rev. Lett.*, **89**, 054101.
4. Brandes, U. (2001) A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*, **25**, 163–177.

5. Brandes, U. (2008) On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, **30**(2), 136 – 145.
6. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. (2010) Catastrophic cascade of failures in interdependent networks. *Nature*, **464**(7291), 1025–1028.
7. Comellas, F. & Gago, S. (2007) Spectral bounds for the betweenness of a graph. *Linear Algebra and its Applications*, **423**(1), 74 – 80.
8. Donetti, L., Neri, F. & Munoz, M. A. (2006) Optimal network topologies: Expanders, Cages, Ramanujan graphs, Entangled networks and all that. *Journal of Statistical Mechanics: Theory and Experiment*, **8**.
9. Estrada, E. (2006) Network robustness to targeted attacks. The interplay of expansibility and degree distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, **52**(4), 563–574.
10. Fiedler, M. (1973) Algebraic connectivity of graphs.. *Czechoslovak Mathematical Journal* **23.2**.
11. Freeman, L. C. (1977) A Set of Measures of Centrality Based on Betweenness. *Sociometry*, **40**(1), 35–41.
12. Freeman, R., Yang, P. & Lynch, K. (2006) Distributed estimation and control of swarm formation statistics. In *American Control Conference, 2006*, page 7.
13. Golub, G. H. & van Van Loan, C. F. (1996) *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)* . The Johns Hopkins University Press, 3rd edition.
14. Jadbabaie, A., Motee, N. & Barahona, M. (2004) On the Stability of the Kuramoto Model of Coupled Non-linear Oscillators. In *In Proceedings of the American Control Conference*, pages 4296–4301.
15. jun Wu, J., you Gao, Z. & jun Sun, H. (2006) Cascade and breakdown in scale-free networks with community structure. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, **74**(6), 066111–5.
16. Martin Hernandez, J., Wang, H., Van Mieghem, P. & D'Agostino, G. (2013) Algebraic Connectivity of Interdependent Networks. *Physica A*, **tba**.
17. McKay, B. D. (1981) Practical Graph Isomorphism. *Congressus Numerantium*, **30**, 45–87.
18. Olfati-Saber, R. & Dartmouth Coll., H. (2007) Algebraic Connectivity Ratio of Ramanujan Graphs. In *American Control Conference*.
19. Pastor-Satorras, R. & Vespignani, A. (2001) Epidemic Spreading in Scale-Free Networks. *Phys Rev Lett*, **86**(14), 3200–3203.
20. Pecora, L. M. & Carroll, T. L. (1998) Master Stability Functions for Synchronized Coupled Systems. *Phys. Rev. Lett.*, **80**, 2109–2112.
21. Rad, A. A., Jalili, M. & Hasler, M. (2011) A Lower Bound for Algebraic Connectivity based on Connection Graph Stability Method. *Linear Algebra and its Applications*, **435**, 186–192.
22. Sorensen, D. C. (1992) Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Matrix Anal. Appl.*, **13**(1), 357–385.
23. Strogatz, S. (2000) From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators.. *Physica D*, **143**, 1–20.
24. Strogatz, S. H. (2001) Exploring complex networks. *Nature*, **410**(6825), 268–276.
25. Trajanovski, S., Martin Hernandez, J., Winterbach, W. & Van Mieghem, P. (2013) Robustness Envelopes of Networks. *OUP Journal of Complex Networks*, **1**.
26. Trajanovski, S. & Van Mieghem, P. (2010) Fiedler's Clustering on m-dimensional Lattice Graph. In *Proceedings of the 3rd International Workshop on Optimal Network Topologies (IWONT)*, Spain.
27. Van Mieghem, P. (2011a) *Graph Spectra for Complex Networks*. Cambridge University Press.
28. Van Mieghem, P. (2011b) The N-intertwined SIS epidemic network model. *Computing*, **93**(2-4), 147–169.
29. Wang, H., Li, Q., D'Agostino, G., Havlin, S., Stanley, H. E. & Van Mieghem, P. (2013) Effect of the Interconnected Network Structure on the Epidemic Threshold. *Physical Review E*, **88**(2), 022801.
30. Wright, T. G., Lloyd & Trefethen, L. N. (2001) Large-Scale Computation Of Pseudospectra Using ARPACK And Eigs. *SIAM J. Sci. Comput.*, **23**, 591–605.

A. Scatter plots for $N = 8$.

For all non-isomorphic connected graphs with $N = 8$ nodes, we display nine scatter plots for the algebraic connectivity vs. each of the nine weighted betweenness. There exist a total of 11,117 connected graphs with 8 nodes.

None of the point clouds seems to consistently stay below the cyan line (which illustrates the equality $\mu_{N-1} = N/r_l$), except the maximum Brandes and maximum Full betweenness. Even though the average Brandes betweenness seems to be the closest candidate to approximate μ_{N-1} (middle row, middle column), the number of complying graphs is lower than 40%, as illustrated in Section 4.1, Figure3.

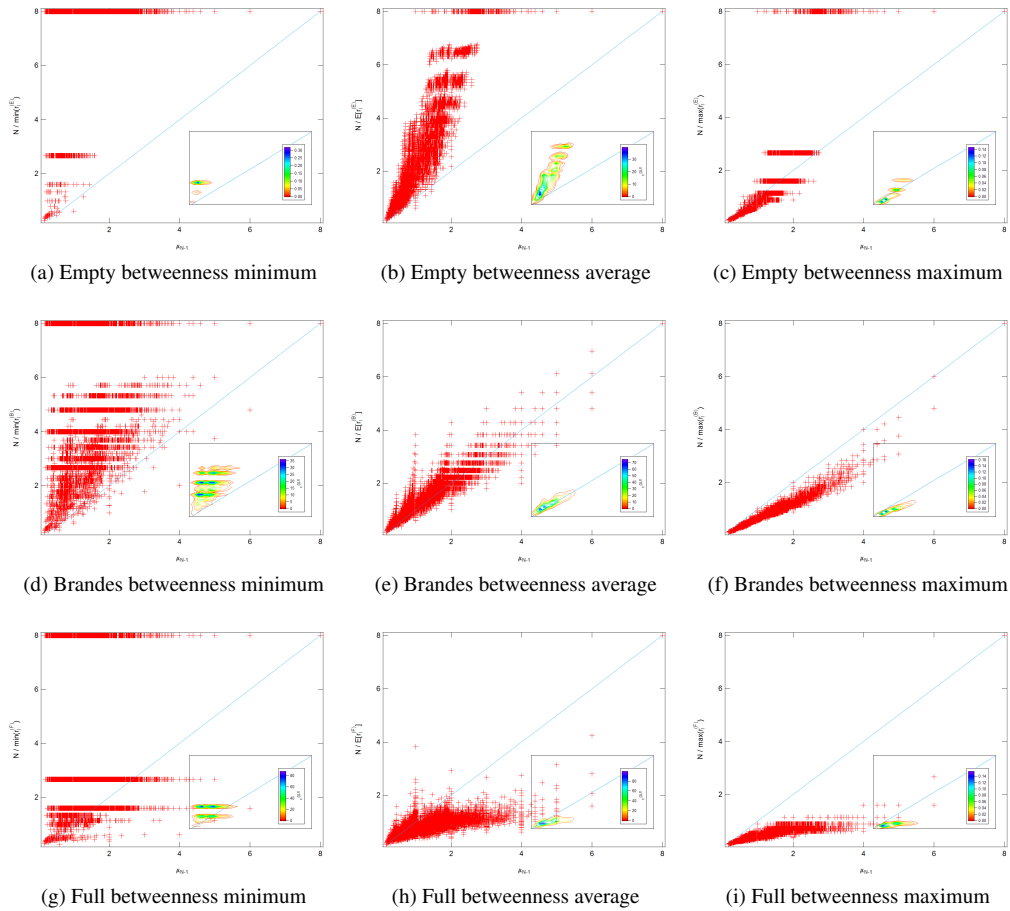


FIG. A.5: $N = 8$ nodes. The cyan line illustrates the equality $\mu_{N-1} = N/r_l$, thus graphs lying below the cyan line obey the respective inequality, points lying above do not. The inset graph displays a contour plot of the data: blue contours high density, red contours low density.