On Random Eccentricity in Complex Networks

Gyan Ranjan[1,*], Zhi-Li Zhang

Dept. of Computer Science, University of Minnesota, Twin Cities, USA.

Abstract

We introduce random eccentricity for the nodes of a complex network as the detour overhead incurred when a random walk between any two nodes is forced through the node in question, averaged over all source-destination pairs. We provide an $O(n^3)$ method to compute random eccentricity for all nodes of the network by means of the Moore-Penrose pseudo-inverse of its combinatorial laplacian. This yields a geometric interpretation for random eccentricity in terms of the length of the position vectors for the nodes in a Euclidean space. Finally, we show how random eccentricity of a node captures its position in all the spanning trees of the associated graph or, equivalently, all possible cuts which divide the graph into exactly two connected components; thereby reflecting the vulnerability/immunity of a node to random edge failures in the network.

1. Introduction

Ever since Leonhard Euler's intellectual jaunt along the bridges of Königsberg in 1753, graph theory has become one of the pillars of discrete modeling and analysis. In its latest incarnation and rechristening as complex network theory, it is more widely used than ever before. Simply put, a network is a discrete structure comprising of entities and the binary relationships between them. Consequently, complex networks pervade through a variety of fields ranging from molecular biology to statistical physics, epidemiology to sociometry, *blogospheres* to the more general world wide web. Despite the varied application landscape, however, some of the problems at the heart of the subject have remained unchanged; one amongst them being the overall connectedness of a node in the network.

Perhaps the simplest of all measures which capture the connectedness of a node is its degree. Degree is simply the number of edges incident on, or emanating from, a node. It defines the one-hop neighborhood of a node. Despite its apparent local nature, degree has been shown to characterize the global connectedness of a node in *scale free* networks [3, 16], where the node degree distribution follows a power law. Scale free networks are often found in the real world [15] and are characterized by the so called *rich club connectivity* whereby nodes with high degrees connect with each other with high probability. Degree of a node in scale free networks is, therefore, believed to be highly correlated with its structural properties such as resilience against random failures and targeted attacks [2]. This, however, is not true in general and has recently been challenged in the domain of Internet's router infrastructure [30]. A way to characterize the overall connectedness of a node is in terms of its *betweenness*. Geodesic or shortest-path betweenness of a node i is defined as the fraction of shortest paths between all pairs of nodes that go through i. If, as is frequently the case, there are multiple shortest paths connecting two vertices, each such path is given an equal fractional weight, summing up to unity. Geodesic betweenness, therefore, measures a node's contribution towards communications between other nodes in the network [20, 21]. Geodesic betweenness and degrees of nodes invariably show

[*]Corresponding author.

Email addresses: `granjan@cs.umn.edu` (Gyan Ranjan), `zhzhang@cs.umn.edu` (Zhi-Li Zhang)

[1]This work was supported in part by the NSF grants CNS-0905037, CNS-1017647 and CNS-1017092 and the DTRA grant HDTRA1-09-1-0050.

high correlation in many real world networks [24, 32]. In most networks, however, communication is not confined to shortest or geodesic paths alone [22, 38]. News, rumour or communicable diseases do not take predetermined paths for spreading across the underlying network of human contacts. Their spread is random as has been confirmed by the famous small-world experiment [31] and its modern day equivalent [12]. Inspired by such experiments, several betweenness measures, which take into account geodesic as well as non-geodesic paths, have recently been proposed. Principal amongst these are flow betweenness [22], which measures the amount of flow that passes through a node i when a unit flow is inserted at all other nodes, and random-walk betweenness [33], which is defined as the number of times a random walker visits a node i in a random walk between a pair of nodes in the graph, summed over all possible source-destination pairs. For a detailed discussion of the advantages of choosing multiple/all-paths based measures over their shortest-paths based counterparts, we refer the reader to [33]. Every betweenness measure, in turn, can be used to compute a *closeness/ farness* index. Geodesic *closeness /farness* of a node is defined as the average length of geodesic distances between the node and the rest of the nodes in the network [20, 21]. Correspondingly, information [38] and random-walk centralities [35] have been proposed, which in effect measure the harmonic sums of all path lengths connecting the node to other nodes.

Recently, Estrada et al [14] have proposed subgraph centrality which aims at capturing the number of subgraphs a node participates in. This centrality index, in principle, has a clearer structural connotation. Alas, subgraph centrality is, for all practical purposes, computationally intractable. In practice, the method suggested in [14] is a two level approximation. The authors first approximate the proposed subgraph centrality of a node i by the number of cycles in which the node participates, a problem no less formidable than the original one. They then resort to computing the weighted sum of closed walks of all lengths, starting and ending at the node in question. The weight assigned to a walk is inversely proportional to the factorial of its length, a mathematical necessity to make the sum converge as much as the intended preference for shorter paths. This inevitably leads to subgraph centrality becoming biased towards local connectivity.

We aim at providing an alternative means to measure the overall connectedness of a node in the network. We propose random eccentricity which

a. Reflects the average overhead in a random detour taken through the node in question in a random walk between any two nodes.
b. Is based on the harmonic sum of all path-lengths ending at a node, particularly desirable because it automatically gives greater weight to shorter paths.
c. Is related to the voltage distribution in the equivalent electrical network (EEN).
d. Is computable for all nodes in the network in $O(n^3)$ time, n being the number of nodes in the network, through a simple matrix inversion.
e. Has a geometric interpretation in terms of a Euclidean embedding in an $n-$dimensional space.
f. Captures the structural connectedness of a node in terms of the number of *dense spanning forests* rooted at it [10] or, equivalently, the size of the vertex set from which it gets detached in the connected partitions of the network.

The rest of this paper is organized as follows. §2 presents the preliminaries for modeling a complex network as a simple graph and its equivalent electrical network (EEN). §3 outlines the idea of a two-phase random walk, the overhead involved and its relationship with the electrical properties of the EEN. §4 provides a method for computing random eccentricity using the Moore-Penrose pseudo-inverse of the combinatorial laplacian of the associated graph. §5 gives a Euclidean embedding of the network and a geometric interpretation of random eccentricity. §6 demonstrates the relationship between random eccentricity and *dense spanning forests* and *connected partitions* of the graph. §7 presents empirical analysis with the help of example networks and *scale free* graphs. Finally, section 8 concludes with the summary of the current results and future work.

2. The Network as a Graph

We model the complex network as a simple graph $G(V, E)$, without self loops. Let $n = |V(G)|$ be the number of vertices/nodes (order of the graph), and $m = |E(G)|$ be the number of links/edges in G. Let

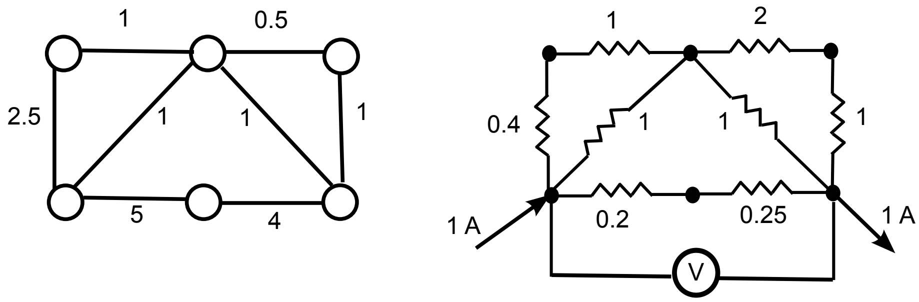


Figure 1: A simple graph G and its Equivalent Electrical Network.

$e_{ij} \in E(G)$ be an edge between nodes i and j with a weight $w_{ij} > 0$. Here, w_{ij} is a measure of affinity between the nodes i and j. If the graph is unweighted, all edges are assigned a unit weight i.e. $w_{ij} = 1 \ \forall e_{ij} \in E(G)$. Moreover, we denote by $\mathcal{N}_i$ the neighbor set of node i; i.e. $j \in \mathcal{N}_i \quad \Leftrightarrow \quad e_{ij} \in E(G) \quad \Leftrightarrow \quad i \in \mathcal{N}_j$. Let $d(i) = \sum_{j \in \mathcal{N}_i} w_{ij}$ the degree of the node i. Clearly, $d(i) = |\mathcal{N}_i|$ if G is unweighted. $Vol(G) = \sum_{i=1}^{n} d(i)$ is called the *volume* of the graph. Again, for an unweighted graph, $Vol(G) = 2|E(G)| = 2m$. $Vol(G)$, therefore, can be thought of as the aggregate affinity for the graph G.

As we make extensive use of the electrical properties of the network in the first half of this work, an introduction is warranted. The equivalent electrical network (EEN) for an undirected graph is formed by replacing an edge $e_{ij} \in E(G)$ with a conductance equal to w_{ij}, [13]; while the nodes of the graph G form the terminals. The dual of conductance of an edge e_{ij} is its resistance $r_{ij} = w_{ij}^{-1}$. Figure 1, shows a simple finite graph on the left and its EEN on the right. The unit of edge weights in the EEN is ohm. We denote by Ω_{ij} the effective resistance between nodes i and j, not necessarily neighbors in the graph, defined as the voltage developed across terminals i and j when a unit current is injected at i and is extracted from j, or vice versa. Effective resistance is a Euclidean distance metric which captures the harmonic sum of all path lengths between the two nodes in question [29]. In what follows, the graph is considered to be simple, finite, connected and unweighted; correspondingly all edges have unit resistance in the EEN.

3. Detour in a Random Walk and Random Eccentricity

Consider a simple random walk $(i \to j)$ [23, 25, 27], which starts at a node i, the source, visits other nodes in the graph G with varying and strictly positive probabilities until it finally reaches the destination j for the first time and stops. At each stage of this discrete stochastic process, the random walker has an equal probability of transitioning to the neighbors of the current node in one step. Now consider another process in which a random walker starts from a source node i, first goes to a node $k \neq i$, henceforth called a transit node, then starts again from k and tries to reach the destination $j \neq k$. This two-phase process $(i \to k \to j)$, in fact, comprises of two different random walks $(i \to k)$ and $(k \to j)$. The random walker may visit the node j several times during the first phase $(i \to k)$ but must terminate on reaching j for the first time in the second phase $(k \to j)$. We call this two-phase walk $(i \to k \to j)$ a random detour from source i to destination j through the transit k. We quantify the difference between the random detour $(i \to k \to j)$ and the simple random walk $(i \to j)$ in terms of the number of steps required to complete each of the two processes; first passage/ hitting time is a suitable measure for this purpose.

Definition 1. *First Passage/ Hitting Time (H_{ij}) : The expected number of steps in a random walk starting at node i before it reaches node j for the first time.*

Clearly, the expected number of steps required to complete the random detour $(i \to k \to j)$ is given by $H_{ik} + H_{kj}$ while the expected number of steps required in a simple random walk $(i \to j)$ is H_{ij}. Intuitively, it must take at least as many steps to complete the random detour ($i \rightarrow k \rightarrow j$) as it does to complete the simple random walk ($i \rightarrow j$). We, therefore, define the overhead for a random detour through transit k with respect to a source i and destination j as

$$\Delta H^{i \rightarrow k \rightarrow j} = H_{ik} + H_{kj} - H_{ij} \tag{1}$$

In general, $H_{ij} \neq H_{ji}$ and hitting time is not a distance measure in a strict Euclidean sense. Another measure, called commute times, has therefore been defined in literature to capture the number of steps in random walks.

Definition 2. *Commute Time (C_{ij}) : The expected number of steps in a random walk starting at node i which visits node j at least once before returning to node i again.*

In terms of hitting times, $C_{ij} = H_{ij} + H_{ji} = C_{ji}$ and, more importantly, $C_{ij} = Vol(G) \times \Omega_{ij} = C_{ji}$ [7]. Consequently, commute time is a metric. In the following theorem, we demonstrate the equivalence between the non-metric hitting times and the metric commute times in the form defined above in (1):

Theorem 1.

$$H_{ik} + H_{kj} - H_{ij} = \frac{C_{ik} + C_{kj} - C_{ij}}{2} \tag{2}$$

Proof : From proposition $9 - 58$ in [28] we have;

$$H_{ik} + H_{kj} - H_{ij} = \frac{U_i^{jk}}{\pi(i)} \tag{3}$$

Here, U_i^{jk} is the number of times a random walker visits the node i in a random walk from node j to k and $\pi(i)$ is the steady state stationary probability of node i. Given G is undirected and connected, from Perron-Frobenius theory [26];

$$\pi(i) = \frac{d(i)}{Vol(G)} \tag{4}$$

So in effect,

$$H_{ik} + H_{kj} - H_{ij} = \frac{Vol(G) \times U_i^{jk}}{d(i)} \tag{5}$$

From [39], we have the so called triangle inequality defined for the resistance distances $\forall (i, j, k) \in V(G) \times V(G) \times V(G)$;

$$\Omega_{ik} + \Omega_{kj} - \Omega_{ij} = \frac{U_i^{jk}}{d(i)} + \frac{U_j^{ik}}{d(j)} = \frac{2\ U_i^{jk}}{d(i)} = \frac{2\ U_j^{ik}}{d(j)} \tag{6}$$

From (5) and (6) we get,

$$H_{ik} + H_{kj} - H_{ij} = \frac{Vol(G) \times (\Omega_{ik} + \Omega_{kj} - \Omega_{ij})}{2} \tag{7}$$

Again, $\forall (i, j) \in V(G) \times V(G)$, $C_{ij} = Vol(G) \times \Omega_{ij}$, which yields

$$H_{ik} + H_{kj} - H_{ij} = \frac{C_{ik} + C_{kj} - C_{ij}}{2} \tag{8}$$

This completes the proof.

□

The result in theorem 1, not only justifies our choice of hitting times to measure the detour overhead despite it not being a metric, it also highlights two very interesting properties of the detour overhead stated in the following corollaries.

Corollary 1. *Non-Negativity:* $\forall (i,j,k) \in V(G) \times V(G) \times V(G)$

$$H_{ik} + H_{kj} - H_{ij} \geq 0 \tag{9}$$

Proof : Given that commute time is a metric, $\forall (i,j,k) \in V(G) \times V(G) \times V(G)$

$$C_{ik} + C_{kj} - C_{ij} \geq 0 \tag{10}$$

Therefore, from theorem 1, the proof follows.
□

Corollary 1 vindicates our assertion that it takes at least as many steps to complete the random detour $(i \to k \to j)$ as it does in the simple random walk $(i \to j)$. Equivalently, the detour overhead $\Delta H^{i \to k \to j}$ is strictly non-negative, a desired property.

Corollary 2. *Symmetry:* $\forall (i,j,k) \in V(G) \times V(G) \times V(G)$

$$H_{ik} + H_{kj} - H_{ij} = H_{jk} + H_{ki} - H_{ji} \tag{11}$$

Proof : Again, from theorem 1,

$$H_{ik} + H_{kj} - H_{ij} = \frac{C_{ik} + C_{kj} - C_{ij}}{2} \tag{12}$$

Swapping indices i and j in the above, we get

$$H_{jk} + H_{ki} - H_{ji} = \frac{C_{jk} + C_{ki} - C_{ji}}{2} \tag{13}$$

Observing, $C_{ij} = C_{ji}, \forall (i,j) \in V(G) \times V(G)$, we get the desired proof.
□

In other words, while $H_{ij} \neq H_{ji}$ in general, $\Delta H^{i \to k \to j} = \Delta H^{j \to k \to i}$. Therefore, random detours through a transit node k, between a pair of terminal nodes, incurs the same overhead in either direction.

In the rest of this section we exploit some well known results from electrical network theory to enhance our understanding of the detour overhead in terms of the underlying random processes. Let U_k^{ij} be the expected number of times a random walker visits a node k in the random walk $(i \to j)$. In the EEN, let V_k^{ij} be the voltage that develops at node k when a unit current is injected at i and a unit current is extracted from j. Then, from [39], we have the following relationship:

$$V_k^{ij} = \frac{U_k^{ij}}{d(k)} \tag{14}$$

Substituting $k = i$ in (14) we get, $U_i^{ij} = V_i^{ij}\ d(i)$; the expected number of times a random walker returns to the source in the random walk $(i \to j)$. If $U_i^{ij} > 0$, the random walk $(i \to j)$ is called recurrent. For a finite, connected and undirected graph G, the random walk $(i \to j)$ is recurrent for all pair of vertices $\{(i,j) \in V(G) \times V(G) : i \neq j\}$, [13]. By convention the destination is always grounded i.e. $V_j^{ij} = 0$. All voltages are measured relative to the destination. Therefore, for $k = j$, we obtain $U_j^{ij} = V_j^{ij}\ d(j) = 0$. We are now ready for the following lemma, which provides an insight into the possible cause for the overhead.

Lemma 1.

$$U_i^{jk} = U_i^{ik} + U_i^{kj} - U_i^{ij} \tag{15}$$

Proof: Given the relationship in (14) and dividing (15) by $d(i)$, the degree of node i, on both sides we get;

$$V_i^{jk} = V_i^{ik} + V_i^{kj} - V_i^{ij} \tag{16}$$

Proving (16) is equivalent to proving (15). We prefer the voltage form (16) in order to make use of some of the interesting properties of electrical current and its duality/reciprocity with voltage. For details the reader is referred to [39]. From the *superposition principle* of electrical current, we have

$$V_i^{ij} = V_k^{ij} + V_k^{ji} \tag{17}$$

Similarly,

$$V_i^{ik} = V_j^{ik} + V_j^{ki} \tag{18}$$

From, (17) and (18),

$$V_i^{ik} + V_i^{kj} - V_i^{ij} = (V_j^{ik} + V_j^{ki}) + V_i^{kj} - (V_k^{ij} + V_k^{ji}) \tag{19}$$

Rearranging the terms in the RHS,

$$V_i^{ik} + V_i^{kj} - V_i^{ij} = V_j^{ik} + (V_j^{ki} + V_i^{kj}) - (V_k^{ij} + V_k^{ji}) \tag{20}$$

From the *reciprocity principle* of voltage and electrical current,

$$V_k^{ij} = V_i^{kj} \tag{21}$$

Therefore;

$$V_i^{ik} + V_i^{kj} - V_i^{ij} = V_i^{jk} + (V_k^{ji} + V_k^{ij}) - (V_k^{ij} + V_k^{ji}) = V_i^{jk} \tag{22}$$

Or;

$$V_i^{jk} = V_i^{ik} + V_i^{kj} - V_i^{ij} \tag{23}$$

Multiplying (23) by $d(i)$, the degree of node i, on both sides we obtain;

$$U_i^{jk} = (U_i^{ik} + U_i^{kj}) - U_i^{ij} \tag{24}$$

□

The RHS of lemma 1 has a form similar to that of the detour overhead in (1). U_i^{ik} is the expected number of times a random walker returns to the source node i in the random walk $(i \rightarrow k)$ and U_i^{kj} is the expected number of times a random walker visits the node i in the random walk $(k \rightarrow j)$. Therefore, $(U_i^{ik} + U_i^{kj}) - U_i^{ij}$ can be interpreted as the expected extra number of times a random walker returns to the source i in the random detour $(i \rightarrow k \rightarrow j)$ as compared to the simple random walk $(i \rightarrow j)$. The following theorem relates lemma 1 to our detour overhead.

Theorem 2.

$$H_{ik} + H_{kj} - H_{ij} = \frac{Vol(G) \times (U_i^{ik} + U_i^{kj} - U_i^{ij})}{d(i)} \tag{25}$$

Proof: Recall from (5),

$$H_{ik} + H_{kj} - H_{ij} = \frac{Vol(G) \times U_i^{jk}}{d(i)} \tag{26}$$

Substituing for U_i^{jk} from lemma 1,

$$H_{ik} + H_{kj} - H_{ij} = \frac{Vol(G) \times (U_i^{ik} + U_i^{kj} - U_i^{ij})}{d(i)} \tag{27}$$

□

In words, the expected extra number of steps taken in a random detour $(i \rightarrow k \rightarrow j)$ as compared to the random walk $(i \rightarrow j)$ is proportional to the extra number of times a random walker returns to the source node i per edge incident on i, in the random detour $(i \rightarrow k \rightarrow j)$ than in the random walk $(i \rightarrow j)$. Each instance of the random process that returns to the source, must effectively start all over again thereby increasing the expected number of steps required to complete the process. Therefore, greater the detour overhead $H^{i\rightarrow k\rightarrow j} = H^{j\rightarrow k\rightarrow i}$ is, more eccentric the transit k is with respect to the source destination pair (i, j) which makes k a less preferred transit. We are now ready to define random eccentricity of a node k, as the additive detour overhead, averaged over all source-destination pairs $\forall(i, j) \in V(G) \times V(G)$;

$$\Delta H^{(k)} = \frac{1}{n^2\ Vol(G)} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta H^{i\rightarrow k\rightarrow j} \tag{28}$$

Note that the random eccentricity of node k, defined above in 28, has been normalized by a factor of $Vol(G)$ which is a constant for the graph G, an invariant over the set of nodes $V(G)$.

4. Computing Random Eccentricity

In this section, we provide a linear algebra based tool to compute random eccentricity for the nodes of a network. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be the adjacency matrix of $G(V, E)$, with elements $a_{ij} = a_{ji} = w_{ij}$ if $e_{ij} \in E(G)$ and $a_{ij} = a_{ji} = 0$ otherwise. By definition of G as a simple graph, $a_{ii} = 0$. Let $\mathbf{D}$ be a diagonal matrix with elements $d_{ii} = \sum_{j=1}^{n} a_{ij}$ and $d_{ij} = 0$ if $i \neq j$. The combinatorial Laplacian $\mathbf{L}$ is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$. $\mathbf{L}$ is known to be singular and is therefore not invertible. However, the Moore-Penrose pseudo-inverse [4, 26] of $\mathbf{L}$, denoted henceforth by $\mathbf{L}^+$, is of interest to us. Traditionally, these matrices have been used as computational tools in the study of electrical circuits, hence the name admittance and Kirchhoff matrix for $\mathbf{L}$ in literature. Klein et al. [29] first expressed the effective resistance between two vertices i and j in terms of the elements of $\mathbf{L}^+$ as;

$$\Omega_{ij} = l^+_{ii} + l^+_{jj} - 2l^+_{ij} = l^+_{ii} + l^+_{jj} - 2l^+_{ji} \tag{29}$$

We can, therefore, express random eccentricity for the nodes of a graph in terms of the elements of $\mathbf{L}^+$ as well. From (28):

$$\begin{aligned}
\Delta H^{(k)} &= \frac{1}{n^2\ Vol(G)} \sum_{i=1}^{n} \sum_{j=1}^{n} H_{ik} + H_{kj} - H_{ij} \\
&= \frac{1}{2n^2\ Vol(G)} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ik} + C_{kj} - C_{ij} \\
&= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Omega_{ik} + \Omega_{kj} - \Omega_{ij} \\
&= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (l^+_{ii} + l^+_{kk} - 2l^+_{ik}) + (l^+_{kk} + l^+_{jj} - 2l^+_{kj}) - (l^+_{ii} + l^+_{jj} - 2l^+_{ij}) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (l^+_{kk} - l^+_{ik} - l^+_{kj} - l^+_{ij})
\end{aligned}$$

Using the fact that $\mathbf{L}^+$ is doubly centered [19], we have:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} l^+_{ik} = \sum_{i=1}^{n} \sum_{j=1}^{n} l^+_{kj} = \sum_{i=1}^{n} \sum_{j=1}^{n} l^+_{ij} = 0$$

Hence,

$$\Delta H^{(k)} = l^+_{kk} \tag{30}$$

Therefore, random eccentricity of a transit node k in the graph G is simply the corresponding diagonal entry in the $\mathbf{L}^+$ matrix, a result upon which we shall expand in the following sections in an attempt to understand what it truly means to be *randomly eccentric*?

5. A Geometric Embedding of the Network

We now look at an $n-$dimensional Euclidean space for embedding the nodes of the network $G(V, E)$ in order to obtain a geometric interpretation for random eccentricity. Following is a well known result from metric multidimensional scaling [6, 11];

$$\mathbf{B} = -\frac{1}{2}(\mathbf{I} - \frac{\mathbf{11'}}{N})\ \mathbb{D}^{(2)}\ (\mathbf{I} - \frac{\mathbf{11'}}{N}) \tag{31}$$

Here, $\mathbb{D}^{(2)}$ is the matrix of pairwise squared distances between n points, $\mathbf{I}$ is the identity matrix of dimension $n \times n$ and $\mathbf{1}$ is a vector of all $1's$. It is known that the matrix $\mathbf{B}$ is symmetric, doubly centered and positive semi-definite. The eigen decomposition of $\mathbf{B} = \Phi\Lambda\Phi'$, on Cholesky factorization $\mathbf{B} = (\Lambda^{1/2}\Phi')'(\Lambda^{1/2}\Phi') = \mathbf{X}'\mathbf{X}$, yields an n-dimensional coordinate system in which the k^{th} column of the matrix $\mathbf{X} = \Lambda^{1/2}\Phi'$ represent the coordinates for node k. With this in view, we establish the following relationship between $\mathbf{L}^+$ and $\mathbf{\Omega}$, the matrix of pairwise effective resistances between the vertices of a graph.

Lemma 2.

$$\mathbf{L}^+ = -\frac{1}{2}(\mathbf{I} - \frac{\mathbf{11'}}{n})\ \mathbf{\Omega}\ (\mathbf{I} - \frac{\mathbf{11'}}{N}) \tag{32}$$

Proof: Expanding the RHS of (32) and substituting $\Omega_{ij} = l^+_{ii} + l^+_{jj} - l^+_{ij} - l^+_{ji}$, we obtain the required proof.
□

Surely, $\mathbf{L}^+$ is symmetric, doubly centered and positive semi-definite [18, 19]. Therefore, $\mathbf{L}^+ = \Phi\Lambda\Phi'$ yields the n-dimensional co-ordinate system spanned by the columns of $\mathbf{X} = \Lambda^{1/2}\Phi'$. The properties of this n-dimensional space have been studied in detail in [17, 37] and applied to a collaborative recommendation task in [19]. In particular, it is known that the squared distance between two points in the n-dimensional space represented by the columns of the matrix $\mathbf{X}$ is exactly equal to the effective resistance between the corresponding nodes of the EEN. But, what about the length of the position vectors?

Let $\mathbf{x}_k$ be the k^{th} column of the matrix $\mathbf{X}$. As the columns of $\mathbf{X}$ are centered at the origin [19], the squared length of the position vector for the point representing node k in this Euclidean space is $||\mathbf{x}_k||_2^2 = \mathbf{x}'_k\mathbf{x}_k$. We, therefore, have the following theorem:

Theorem 3.

$$\Delta H^{(k)} = ||\mathbf{x}_k||_2^2 \tag{33}$$

Proof: Given that $\mathbf{x}_k$ is the k^{th} column of $\mathbf{X} = \Lambda^{1/2}\Phi'$,

$$\mathbf{x}'_k\mathbf{x}_k = \sum_{i=1}^{N} \Lambda_{ii}\Phi^2_{ik} = (\Phi\Lambda\Phi^T)_{kk} = l^+_{kk} \tag{34}$$

But from (30), we know that $\Delta H^{(k)} = l^+_{kk}$. This completes the proof.
□

Therefore, farther the point corresponding to the node k is from the origin, in the space spanned by the columns of $\mathbf{X} = \Lambda^{1/2}\Phi'$, more randomly eccentric it is.

6. The Topological Connection

Random eccentricity is closely related to the set of spanning acyclic subgraphs in a network $G(V, E)$ which are structural derivatives of the set of spanning trees in G. Henceforth, we focus exclusively on the structural aspect of random eccentricity in order to establish it as a true measure of a node's overall connectedness.

6.1. Dense Spanning Forests of a Graph

In a series of articles, Chebotarev et al. have investigated the *rooted spanning forests* of a graph [8, 9, 10], defined as:

Definition 3. *Rooted Spanning Forest : An acyclic subgraph of G that has the same node set as G and one marked node (a root) in each component.*

Of particular interest is a result in [10] regarding *dense rooted spanning forests* of an undirected graph which is related to the random eccentricity measure.

Definition 4. *Dense Rooted Spanning Forest : A rooted spanning forest which has greater than or equal to $n-2$ edges.*

Let $\mathcal{F}_x$ be the set of all rooted spanning forests in $G(V, E)$, with x edges and $\mathcal{F}_x^{ij} \subset \mathcal{F}_x$ be the set of all spanning forests in G, rooted at i in which vertex i and j are in the same tree. From Theorem 3 in [10], we have the following topological interpretation for the elements of $\mathbf{L}^+$;

$$l_{ij}^+ = \frac{\varepsilon(\mathcal{F}_{n-2}^{ij}) - \frac{1}{n}\varepsilon(\mathcal{F}_{n-2})}{\varepsilon(\mathcal{F}_{n-1})} \tag{35}$$

A few notational clarifications are in order. If G_i be a subgraph of G, then

$$\varepsilon(G_i) = \prod_{e_{ij} \in E(G_i)} w_{ij}$$

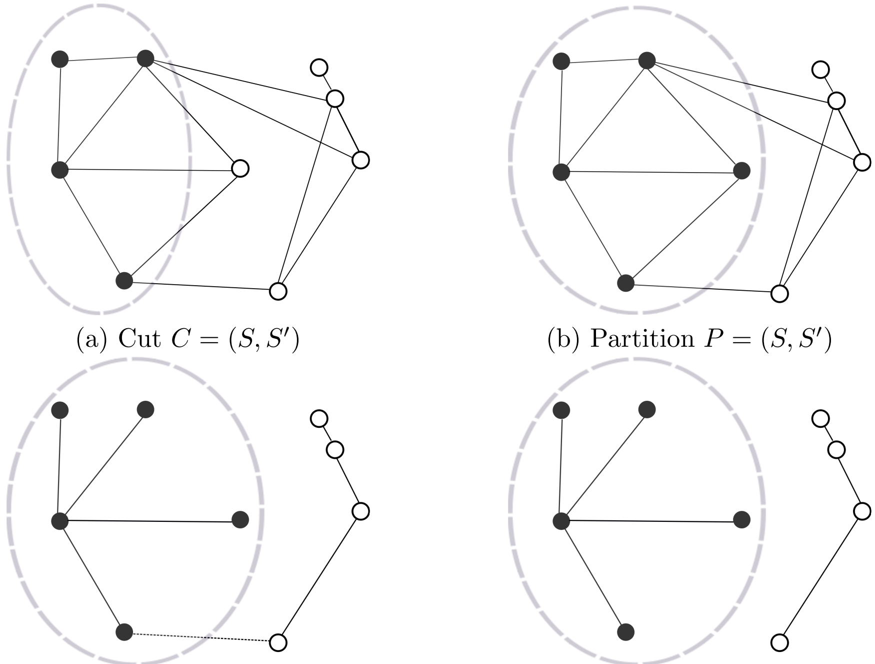
is the product of the weights of edges in G_i. If $\mathcal{G} = \{G_1, G_2, ..., G_s\}$ be a set of s subgraphs of G, then $\varepsilon(\mathcal{G}) = \sum_{i=1}^{s} \varepsilon(G_i)$. For a simple, connected and unweighted graph G, the product of edge weights in any subgraph is unity which means that $\varepsilon(G_i) = 1, \forall G_i \in \mathcal{G}$. Consequently, $\varepsilon(\mathcal{G}) = |\mathcal{G}| = s$ i.e. $\varepsilon(\mathcal{G})$ is the number of subgraphs in the set $\mathcal{G}$. Therefore, $\varepsilon(\mathcal{F}_x)$ is the number of rooted spanning forests with x edges and $\varepsilon(\mathcal{F}_x^{ij})$ is the number of spanning forests, rooted at i in which i and j are in the same tree. Substituting, $i = j = k$ in (35), we obtain:

$$l_{kk}^+ = \frac{\varepsilon(\mathcal{F}_{n-2}^{kk}) - \frac{1}{n}\varepsilon(\mathcal{F}_{n-2})}{\varepsilon(\mathcal{F}_{n-1})} \tag{36}$$

Here, $\varepsilon(\mathcal{F}_{n-2}^{kk})$ is the number of spanning forests with $n-2$ edges in which k is the root of the tree in which it belongs. Note that $\varepsilon(\mathcal{F}_{n-2})$ and $\varepsilon(\mathcal{F}_{n-1})$ are invariants over the set of nodes $V(G)$ for a given graph G. Therefore,

$$l_{kk}^+ \propto \varepsilon(\mathcal{F}_{n-2}^{kk}) \tag{37}$$

Clearly, from (37) we have a direct correspondence between the random eccentricity of a node and the set of dense spanning forests with $n-2$ edges rooted at it. Henceforth, we focus on the set $\mathcal{F}_{n-2}$, to put in place the last pieces of this puzzle.



(a) Cut $C = (S, S')$ (b) Partition $P = (S, S')$ (c) Spanning tree across $P = (S, S')$ (d) Spanning forest in $P = (S, S')$.

Figure 2: Capturing dense spanning forests in terms of the partitions of a graph.

6.2. Random Eccentricity and Partitions of a Graph

We now generalize the topological interpretation presented in section 6.1 in a way which highlights the effectiveness of random eccentricity in capturing the structural connectivity of a node. But first we introduce some notation here.

Let $C = (S, S')$ be a cut for the graph $G(V, E)$. $C = (S, S')$ is characterized by the mutually exclusive and exhaustive vertex sets $V(S)$, $V(S')$ and the edge sets $E(S)$, $E(S')$ and $E(S \times S')$ where $E(S \times S') = \{e_{ij} : i \in V(S) \Leftrightarrow j \in V(S') \text{ and } i \in V(S') \Leftrightarrow j \in V(S)\}$, is the set of edges that violate the cut.

Definition 5. *A cut $C = (S, S')$ of the graph G is called a partition of the graph if and only if the graph induced by the set $E(G) - E(S \times S')$ has exactly two connected components.*

Figure 2 illustrates the difference between a cut and a partition. By abuse of notation, we use S and S' to denote the connected subgraphs represented by the respective cut vertex sets of the partition $P = (S, S')$, as well. Let $\mathcal{P} = \{P : \text{A partition of the graph } G\}$. For a detailed treatment please see [**?**].

For a given partition $P = (S, S') \in \mathcal{P}$, let $\mathcal{T}_S$ and $\mathcal{T}_{S'}$ be the sets of spanning trees defined over the nodes of the subgraphs S and S' respectively. It is easy to see that, for any $T_1 \in \mathcal{T}_S$, $T_2 \in \mathcal{T}_{S'}$ and $e_{ij} \in E(S \times S')$, the subgraph induced by $E(T_1) \cup \{e_{ij}\} \cup E(T_2)$ is a spanning tree $T \in \mathcal{T}_G$. Therefore, a partition $P = (S, S')$ cuts across $(|\mathcal{T}_S| \times |E(S \times S')| \times |\mathcal{T}_{S'}|)$ number of spanning trees of the graph. In what follows, we select a particular node k and determine its random eccentricity in terms of the structural properties of the partitions of the graph. Without loss of generality, we assume $k \in V(S)$.

Lemma 3. *Let $\mathcal{F}^{kk}_{n-2|P}$ be the set of spanning forests with $n - 2$ edges rooted at node k in a partition $P = (S, S') \in \mathcal{P}$ and $k \in V(S)$. Then,*

$$\varepsilon(\mathcal{F}^{kk}_{n-2|P}) = |\mathcal{T}_S| \times |\mathcal{T}_{S'}| \times |V(S')| \tag{38}$$

Proof : Given a partition $P = (S, S')$, let $T_1 \in \mathcal{T}_S$ and $T_2 \in \mathcal{T}_{S'}$ be spanning trees respectively in the two parts of the partition $P = (S, S')$. Clearly, $|E(T_1)| = |V(S)| - 1$ and $|E(T_2)| = |V(S')| - 1$. Given that $V(S) \cup V(S') = V(G)$,

$$\begin{aligned} |E(T_1) \cup E(T_2)| &= |E(T_1)| + |E(T_2)| \\ &= |V(S)| - 1 + |V(S')| - 1 \\ &= |V(S)| + |V(S')| - 2 \\ &= n - 2 \end{aligned}$$

Each such pair (T_1, T_2) gives a spanning forest of $n-2$ edges. As we have already chosen k as the root of T_1 in S, we can choose $|V(S')|$ number of different roots for T_2 in S'. So for a given pair (T_1, T_2) there are $|V(S')|$ number of spanning forests rooted at k. Clearly, there are $|\mathcal{T}_S| \times |\mathcal{T}_{S'}|$ such pairs of (T_1, T_2). Therefore, for a given partition $P = (S, S')$,

$$\varepsilon(\mathcal{F}^{kk}_{n-2|P}) = |\mathcal{T}_S| \times |\mathcal{T}_{S'}| \times |V(S')|$$

□

Consider then, a node $i \in V(S) : i \neq k$. As i and k belong to the same subgraph S in $P = (S, S')$, the following result holds:

$$\varepsilon(\mathcal{F}^{kk}_{n-2|P}) = \varepsilon(\mathcal{F}^{ii}_{n-2|P})$$

However, for a node $j \in V(S')$,

$$\varepsilon(\mathcal{F}^{jj}_{n-2|P}) = |\mathcal{T}_S| \times |\mathcal{T}_{S'}| \times |V(S)|$$

We therefore see that for a given partition, the count of spanning forests rooted at a node depends upon which of the two subgraphs S or S', the node in question belongs to. In particular, for a given $P = (S, S')$, $i \in V(S)$ and $j \in V(S')$,

$$\varepsilon(\mathcal{F}^{ii}_{n-2|P}) - \varepsilon(\mathcal{F}^{jj}_{n-2|P}) = |\mathcal{T}_S| \times |\mathcal{T}_{S'}| \times (|V(S')| - |V(S)|)$$

Each partition $P \in \mathcal{P}$, therefore, divides the node set of G into two *equivalence* classes in terms of the counts of rooted spanning forests represented in P. We now need a result for the overall count of spanning forests rooted at a node in terms of all the partitions.

Lemma 4.

$$\varepsilon(\mathcal{F}^{kk}_{n-2}) = \sum_{P \in \mathcal{P}} \varepsilon(\mathcal{F}^{kk}_{n-2|P}) \tag{39}$$

Proof: By definition, if $P_1 = (S_1, S'_1)$ and $P_2 = (S_2, S'_2)$ are two distinct partitions then $V(S_1) \neq V(S_2)$ and $V(S'_1) \neq V(S'_2)$. Clearly, this implies that $\forall F \in \mathcal{F}_{n-2}$, F can only belong to one of the partitions of G. This in turn means:

$$\mathcal{F}^{kk}_{n-2} = \coprod_{P \in \mathcal{P}} \varepsilon(\mathcal{F}^{kk}_{n-2|P})$$

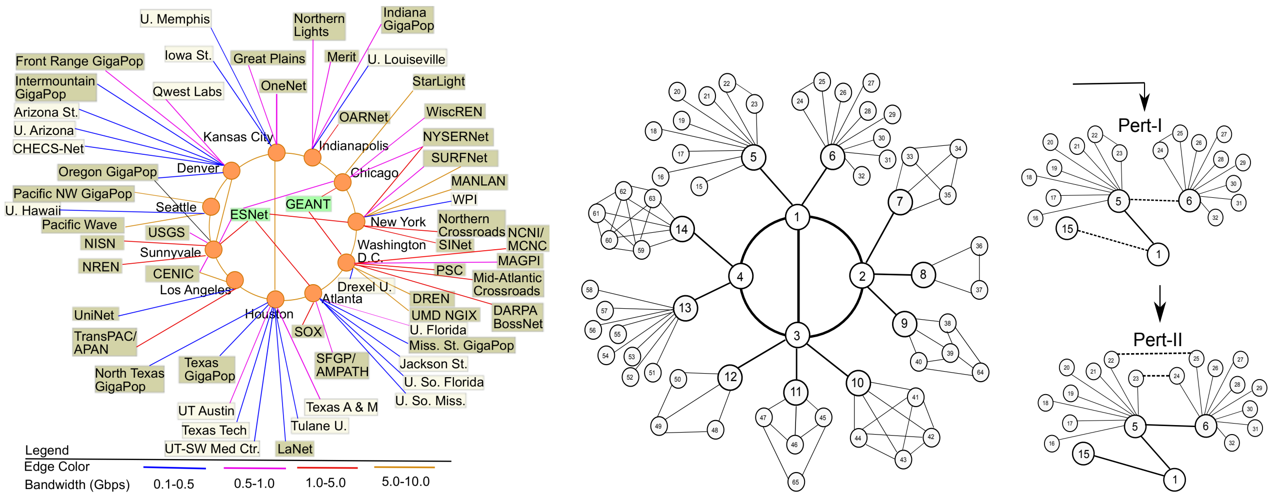
As the RHS above is a disjoint union, counting the members of the sets on both sides we get the desired proof.

□

To summarize the results of the last two lemmas, in the overall context of random eccentricity of a node k, we see that a high value of $\mathcal{F}^{kk}_{n-2}$, and consequently of l^+_{kk}, shows a tendency of node k to be detached from a larger portion of the graph on the deletion of an arbitrary edge in $E(S \times S')$. Moreover, as a comparative measure for an arbitrary pair of nodes i and j;

$$l^+_{ii} - l^+_{jj} \propto \varepsilon(\mathcal{F}^{ii}_{n-2}) - \varepsilon(\mathcal{F}^{jj}_{n-2}) \tag{40}$$

$$= \sum_{P \in \mathcal{P}} \varepsilon(\mathcal{F}^{ii}_{n-2|P}) - \varepsilon(\mathcal{F}^{jj}_{n-2|P}) \tag{41}$$



(a) Abilene Topology (b) Simulated topology (c) Perturbations

Figure 3: Abilene Network and a simulated topology.

Notice that a term corresponding to a particular $P \in \mathcal{P}$ of the summand in (41), is zero if and only if

a. Nodes i and j belong to the same connected component of the partition P, or
b. The two connected components of the partition P have the same number of nodes, i.e. $|V(S)| = |V(S')|$.

Otherwise, the partition in question acts as a differentiator between the two nodes, contributing differently to the count of spanning rooted forests to each. For a node $i \in V(S)$, the contribution made by P is $|\mathcal{T}_S| \times |\mathcal{T}_{S'}| \times |V(S')|$ and for a node $j \in V(S')$, the contribution made is $|\mathcal{T}_S| \times |\mathcal{T}_{S'}| \times |V(S)|$.

Once again, a high value of random eccentricity for a node indicates lower structural significance and vice versa. Therefore, it is justifiably a dual of the centrality measures in its class.

7. Empirical Evaluations

We now empirically study the properties of structural centrality ($\mathcal{C}^*(i)$) and Kirchoff index (we use $\mathcal{K}^* = \mathcal{K}^{-1}$ henceforth to maintain *higher is better*). We first show in §7.1, how structural centrality can capture the structural roles played by nodes in the network and then in §7.2 demonstrate how it, along with Kirchoff index, is appropriately sensitivity to rewiring and local perturbations in the network.

7.1. Identifying Structural Roles of Nodes

Consider the router level topology of the Abilene network (Fig. 3(a)) [1]. At the core of this topology, is a ring of 11 POP's, spread across mainland US, through which several networks interconnect. Clearly, the connectedness of such a network is dependent heavily on the low degree nodes on the ring. For illustration, we mimic the Abilene topology, with a simulated network (Fig. 3(b)) which has a 4-node core $\{v_1, ..., v_4\}$ that connects 10 networks through gateway nodes $\{v_5, ..., v_{14}\}$ (Fig.3(b)). Fig. 4 shows the (max-normalized) values of GC, SC and $\mathcal{C}^*$ for the core $\{v_1, ..., v_4\}$, gateway $\{v_5, ..., v_{14}\}$ and other nodes $\{v_{15}, ..., v_{65}\}$ in topology (Fig.3(b)). Notice that v_5 and v_6 are the highest degree nodes ($d(v_5) = d(v_6) = 10$) in the network while v_{14} has the highest subgraph centrality (SC). In contrast, $\mathcal{C}^*$ ranks the core nodes higher than the gateway nodes with v_1 at the top. The relative peripherality of v_5, v_6 and v_{14} as compared to the core nodes requires no elaboration. As far as geodesic centrality (GC) is concerned, it ranks all the nodes in the subnetwork abstracted by v_5, namely $v_{15} - v_{23}$, as equally well connected even though v_{22} and v_{23} have

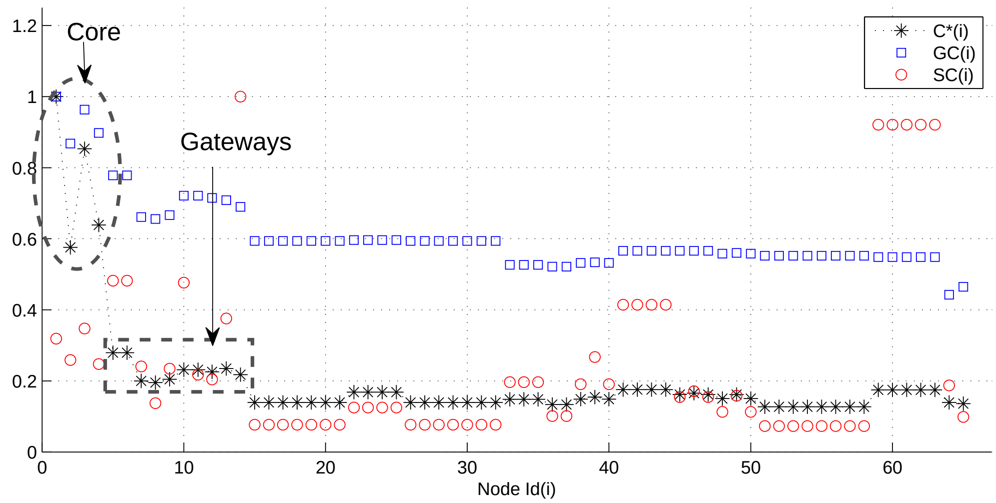


Figure 4: Max-normalized centralities for simulated topology.

redundant connectivity to the network through each other and are, ever so slightly, better connected than the others.

We see similar characterization of structural roles of nodes in two real world networks: western states power-grid network [40] and a social network of co-authorships [34], as shown through a color scheme based on $\mathcal{C}^*(i)$ values in Fig. 5. Core-nodes connecting different sub-communities of nodes in both these real world networks are recognized effectively by structural centrality as being more central (*Red* end of the spectrum) than several higher degree peripheral nodes.

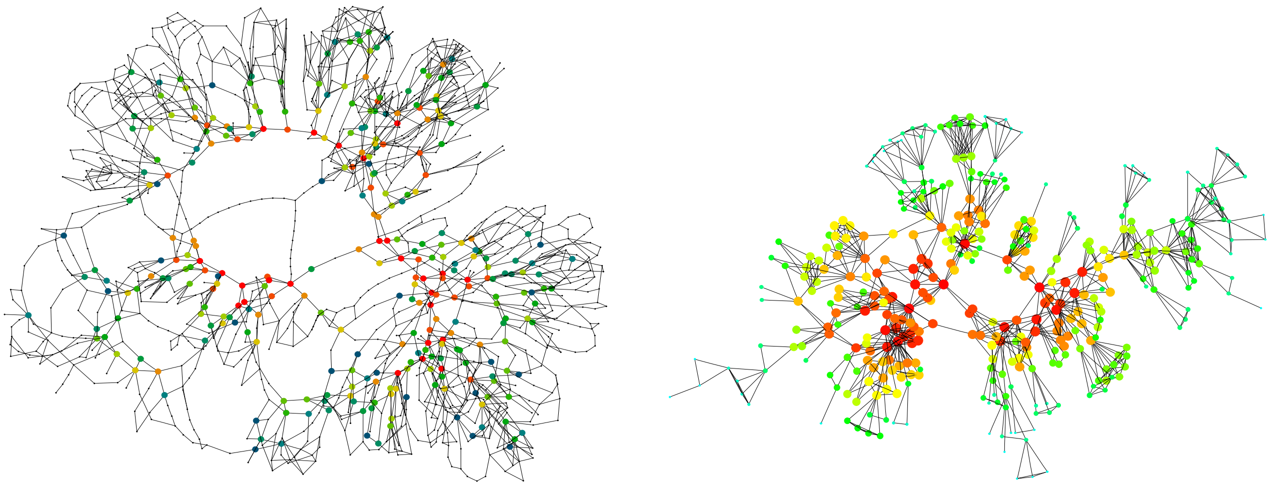
7.2. Sensitivity to Local Perturbations

An important property of centrality measures is their sensitivity to perturbations in network structure. Traditionally, structural properties in real world networks have been equated to average statistical properties like power-law/scale-free degree distributions and rich club connectivity [2, 15, 16]. However, the same degree sequence $D = \{d(1) \geq d(2) \geq ... \geq d(n)\}$, can result in graphs of significantly varying topologies. Let $\mathcal{G}(D)$ be the set of all connected graphs with scaling sequence D. The generalized Randic index $R_1(G)$ [5, 36]:

$$R_1(G) = \sum_{e_{ij} \in E(G)} d(i)d(j) \tag{42}$$

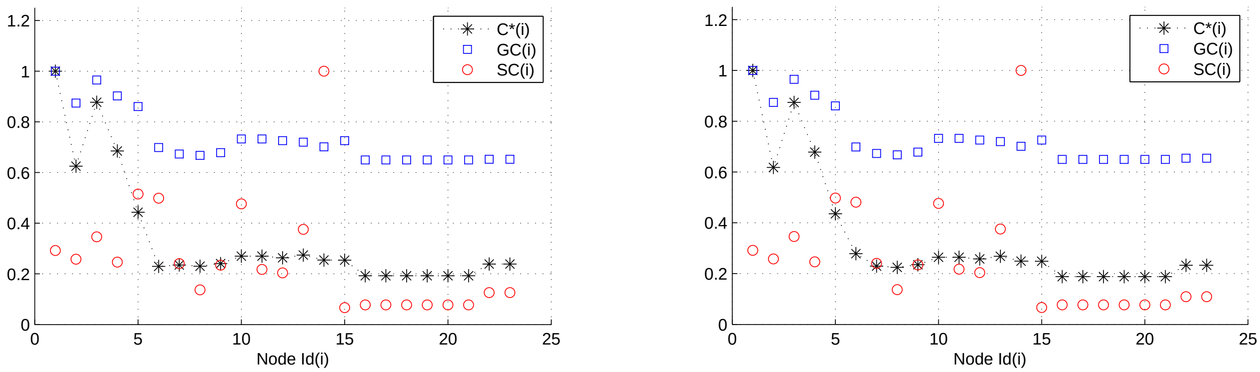
where $G \in \mathcal{G}(D)$, is considered to be a measure of overall connectedness of G as higher $R_1(G)$ suggests *rich club connectivity* (RCC) in G [30]. Also, the average of each centrality index (GC, SC, GB, RB averaged over the set of nodes), is in itself a global structural descriptor for the graph G [14]. We now examine the sensitivity of each index with respect to local perturbations in the subnetwork abstracted by the core node v_1 and its two gateway neighbors v_5 and v_6.

First, we rewire edges $e_{15,5}$ and $e_{6,1}$ to $e_{15,1}$ and $e_{6,5}$ respectively (PERT-I Fig. 3(c)). PERT-I is a degree preserving rewiring which only alters local connectivities. Fig. 6(a) and Fig. 7 respectively show the altered values of centralities ($\mathcal{C}^*, GC, SC$) and betweennesses, geodesic and random-walk i.e. (GB, RB), after PERT-I. Note, after PERT-I, v_{15} is directly connected to v_1 which makes $\mathcal{C}^*(v_{15})$ comparable to other gateway nodes while $SC(v_{15}), GB(v_{15}), RB(v_{15})$ seem to be entirely unaffected. Moreover, PERT-I also results in v_6 losing its direct link to the core, reflected in the decrease in $\mathcal{C}^*(v_6)$ and a corresponding increase in $\mathcal{C}^*(v_5)$. $\mathcal{C}^*$, however, still ranks the core nodes higher than v_5 (whereas SC, GB, RB do not) because PERT-I being a



(a) The western-states power grid network [40] (b) A network of co-authorships in network sciences [34]

Figure 5: Real world networks: *Red* $\rightarrow$ *Turqoise* reducing order of $\mathcal{C}^*(i)$.



(a) After PERT-I (b) After PERT-II

Figure 6: Max-normalized values of structural centrality, geodesic closeness and subgraph centrality for core, gateway and some other nodes.

local perturbation should not affect nodes outside the sub-network — v_1 continues to abstract the same sub-networks from the rest of the topology. We, therefore, observe that $\mathcal{C}^*$ is appropriately sensitive to the changes in connectedness of nodes in the event of local perturbations. But what about the network on a whole?

Let G and G_1 be the topologies before and after PERT-I. G_1 is less well connected overall than G as the failure of $e_{5,1}$ in G_1 disconnects 19 nodes from the rest of the network as compared to 10 nodes in G. However,

$$\Delta R_1(G \rightarrow G_1) = \frac{R_1(G_1) - R_1(G)}{R_1(G)} = 0.029$$

as the two highest degree nodes (v_5 and v_6) are directly connected in G_1 (see TABLE 1 for the sensitivity of other centrality based global structural descriptors). In contrast, $\Delta\mathcal{K}^*(G \rightarrow G_1) = -0.045$, which rightly reflects the depreciation in overall connectedness after PERT-I.

A subsequent degree preserving perturbation PERT-II of G_1, rewiring $e_{22,23}$ and $e_{24,25}$ to $e_{22,25}$ and $e_{23,24}$, to obtain G_2, creates two cycles in G_2 that safeguard against the failure of edge $e_{5,6}$. This significantly

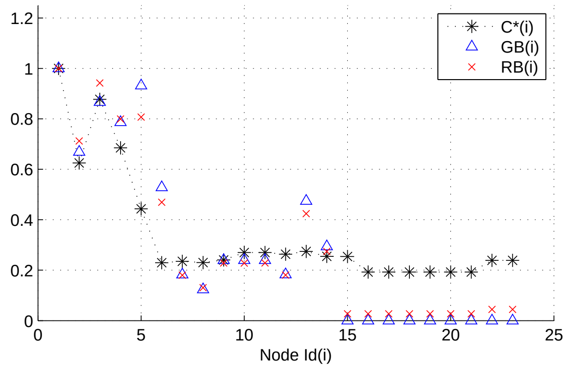


Figure 7: Max-normalized values of structural centrality, geodesic and random-walk betweennesses for core, gateway and some other nodes.

Table 1: Sensitivity of global structural descriptors to local perturbations, $\overline{X} = 1/n \sum_i^n X(i)$: Average node centrality for a network.

#	Structural Descriptor	PERT-I	PERT-II
1	$\mathcal{K}^*(G)$ (or $\overline{\mathcal{C}^*}$)	$\downarrow$	$\uparrow$
2	$R_1(G)$	$\uparrow$	$\leftrightarrow$
3	$\overline{GC}$	$\downarrow$	$\uparrow$
4	$\overline{SC}$, $\overline{GB}$	$\uparrow$	$\downarrow$
5	$\overline{RB}$	$\uparrow$	$\uparrow$

improves local connectivities in the sub-network. However, $\Delta R_1(G_1 \rightarrow G_2) = 0$ (and average SC decreases) while $\Delta\mathcal{K}^*(G_1 \rightarrow G_2) = 0.036$ which once again shows the efficacy of Kirchoff index as a measure of global connectedness of networks.

8. Conclusion and Future Work

In this work we introduced random eccentricity, an index for node centrality in complex networks, which is associated with an overhead incurred in random detours over the network. We established a relationship between the detour overhead and the triangle inequality of effective resistances and provided interpretations in terms of a Euclidean embedding and the set of rooted spanning forests. Through empirical evaluations over real life networks, we presented a comparison with other centrality indices popular in literature. The observations made in this work clearly show that random eccentricity captures the true significance of a node's structural connectedness to the network. As future work, we propose to extend the use of such

Table 2: Taxonomy and computational complexities of centrality measures.

#	Measure	Paths covered	Complexity
1.	Degree	-	$O(m)$
2.	GC, GB	Geodesic paths	$O(n^3)$
3.	$\mathcal{C}^*$	All paths	$O(n^3)$
4.	SC	All paths	$O(n^3)$
5.	RWB	All paths	$O(m+n)n^2$
6.	FB	All paths	$O(m^2 n)$

methods for developing metrics which can help compare the structural robustness of two different networks of comparable sizes and volumes.

References

[1] www.stanford.edu/services/internet2/abilene.html.

[2] R. Albert, H. Jeong, and A. L. Barábasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[3] A. L. Barábasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[4] A. Ben-Israel and T. Greville. *Generalized Inverses: Theory and Applications,* 2^{nd} *edition.* Springer-Verlag, 2003.

[5] B. Bollobas and P. Erdos. Graphs of extremal weights. *Ars Combin.*, 50:225–233, 1998.

[6] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications.* Springer, 1997.

[7] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. In *Proc. of Annual ACM Symposium on Theory of Computing*, pages 574–586, 1989.

[8] P. Chebotarev and E. Shamis. The matrix-forest theorem and measuring relations in small social groups. *Automation and Remote Control*, 58(9):1505–1514, 1997.

[9] P. Chebotarev and E. Shamis. On a duality between metrics and S-proximities. *Automation and Remote Control*, 59(4):608–612, 1998.

[10] P. Chebotarev and E. Shamis. On proximity measures for graph vertices. *Automation and Remote Control*, 59(10):1443–1459, 1998.

[11] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling,* 2^{nd} *edition.* Chapman and Hall/CRC, 2001.

[12] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *Science*, 301:827–829, 2003.

[13] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks.* The Math. Assoc. of America, 1984.

[14] E. Estrada and J. A. Rodríguez-Velázquez. Subgraph centrality in complex networks. *Phys. Review E*, 71, 2005.

[15] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proc. of the ACM SIGCOMM*, pages 251–262, 1999.

[16] I. J. Farkas, I. Derenyi, A. L. Barábasi, and T. Vicsek. Spectra of real world graphs: Beyond the semicircle law. *Physical Review E*, 64(2), 2001.

[17] M. Fiedler. Moore-Penrose involutions in the classes of laplacians and simplices. *Lin. Multilin. Algebra*, 39:171–178, 1995.

[18] F. Fouss, A. Pirotte, J. M. Renders, and M. Saerens. A novel way of computing similarities between nodes of a graph, with application to collaborative recommendation. In *Proc. 2005 IEEE/WIC/ACM Int'l Joint Conf. Web Intelligence*, pages 550–556, 2005.

[19] F. Fouss, A. Pirotte, J. M. Renders, and M. Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 19, 2007.

[20] L. C. Freeman. A set of measures of centrality based upon betweenness. *Sociometry*, 40:35–41, 1977.

[21] L. C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1979.

[22] L. C. Freeman, S. P. Borgatti, and D. R. White. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, 13:141–154, 1991.

[23] F. Gobel and A. Jagers. Random walks on graphs. *Stochastic Processes and Their Applications*, 2:311–336, 1974.

[24] K. I. Goh, E. Oh, B. Kahng, and D. Kim. Betweenness centrality correlation in social networks. *Physical Review E*, 67(017101), 2003.

[25] C. M. Grinstead and J. L. Snell. *Introduction to Probability.* American Mathematical Society, 2006.

[26] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1985.

[27] S. Karlin and H. M. Taylor. *A First Course in Stochastic Processes.* Academic Press, New York, second edition, 1975.

[28] J. G. Kemeny, J. L. Snell, and A. W. Knapp. *Denumerable Markov Chains.* Van Nostrand, New York, 1966.

[29] D. J. Klein and M. Randic. Resistance distance. *J. Math. Chemistry*, 12:81–95, 1993.

[30] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the internet's router-level topology. In *Proc. of the ACM SIGCOMM*, 2004.

[31] S. Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[32] K. Nakako. Distribution of measures of centrality: Enumerated distributions of freeman's graph centrality measures. *Connections*, 13(3):10–22, 1990.

[33] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.

[34] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Preprint Physics/0605087*, 2006.

[35] J. D. Noh and H. Rieger. Random walks on complex networks. *Phys. Rev. Lett.*, 92, 2004.

[36] M. Randic. On characterization of molecular branching. *J. Amer. Chem. Society*, 97:6609–6615, 1975.

[37] M. Saerens, F. Fouss, L. Yen, and P. Dupont. The principal components analysis of a graph, and its relationships to spectral clustering. In *Proc. 15th European Conf. Machine Learning (ECML '04)*, pages 371–383, 2004.

[38] K. A. Stephenson and M. Zelen. Rethinking centrality: Methods and examples. *Social Networks*, 11:1–37, 1989.

[39] P. Tetali. Random walks and effective resistance of networks. *Journal of Theoretical Probability*, pages 101–109, 1991.

[40] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.