Parameterized centrality metric for network analysis

Rumi Ghosh and Kristina Lerman

A Parameterized Centrality Metric for Network Analysis

Rumi Ghosh* and Kristina Lerman†

USC Information Sciences Institute

4676 Admiralty Way, Marina del Rey, CA 90292

A variety of metrics have been proposed to measure the relative importance of nodes in a network. One of these, alpha-centrality [P. Bonacich, Am. J. Sociology 92, 1170 (1987)], measures the number of attenuated paths that exist between nodes. We introduce a normalized version of this metric and use it to study network structure, for example, to rank nodes and find community structure of the network. Specifically, we extend the modularity-maximization method for community detection to use this metric as the measure of node connectivity. Normalized alpha-centrality is a powerful tool for network analysis, since it contains a tunable parameter that sets the length scale of interactions. Studying how rankings and discovered communities change when this parameter is varied allows us to identify locally and globally important nodes and structures. We apply the proposed metric to several benchmark networks and show that it leads to better insight into network structure than alternative metrics.

I. INTRODUCTION

Centrality measures the degree to which network structure contributes to the importance, or status, of a node in a network. Over the years many different centrality metrics have been defined. One of the more popular metrics, betweenness centrality [1], measures the fraction of all shortest paths in a network that pass through a given node. Other centrality metrics include those based on random walks [2–5] and path-based metrics. The simplest path-based metric, degree centrality, measures the number of edges that connect a node to others in a network. According to this measure, the most important nodes are those that have the most connections. However, a node's centrality depends not only on how many others it is connected to but also on the centralities of those nodes [6, 7]. This measure is captured by the total number of paths linking a node to other nodes in a network. One such metric, alpha-centrality [6, 8], measures the total number of paths from a node, exponentially attenuated by their length. The attenuation parameter in alpha-centrality sets the length scale of interactions. Unlike other centrality metrics, which do not distinguish between local and global structure, a parameterized centrality metric can differentiate between locally connected nodes, i.e., nodes that are linked to other nodes which are themselves interconnected, and globally connected nodes that link and mediate communication between poorly connected groups of nodes. Studies of human [9–11] and animal [12] populations suggest that such 'bridges' or 'brokers' play a crucial role in the information flow and cohesiveness of the entire group. Recent studies [13] showed the utility of this metric also in identifying influential nodes in online social networks.

One difficulty in applying alpha-centrality in network analysis is that its key parameter is bounded by the spectrum of the corresponding adjacency matrix of the network. As a result, the metric diverges for larger values of this parameter. We address this problem by defining *normalized alpha-centrality*. We show that the new metric avoids the problem of bounded parameters while retaining the desirable characteristics of alpha-centrality, namely its ability to differentiate between local and global structures.

In addition to ranking nodes, parameterized centrality can be used to identify communities within a network [15]. In this paper, we generalize modularity maximization-based approach [16, 17] to use normalized alpha-centrality. Rather than find regions of the network that have greater than expected number of edges connecting nodes [18], our approach looks for regions that have greater than expected number of attenuated paths connecting nodes. One advantage of this method is that the attenuation parameter can be varied to identify local vs. global communities.

Normalized alpha-centrality is a powerful tool for network analysis. By differentiating between locally and globally connected nodes, it provides a simple alternative to previous attempts to quantify fine-grained structure of complex networks, such as the motif-based [19, 20] and role-based [21, 22] descriptions. The former measures the relative abundance of subgraphs of a certain type, while latter classifies nodes according to their connectivity within and outside of their community. Applying either of these descriptions to real networks is computationally expensive: role-based analysis, for example, requires the network to be decomposed into distinct communities first. Normalized alpha-centrality, on the other hand, measures node connectivity at different length scales, allowing us to resolve network structure in a computationally efficient manner.

We use normalized alpha-centrality to study the structure of several benchmark networks, as well as a real-

* rumig@usc.edu
† lerman@isi.edu

world online social network. We show that this parameterized centrality metric can identify locally and globally important nodes and communities, leading to a more nuanced understanding of network structure.

II. CENTRALITY AND NETWORK STRUCTURE

Bonacich [6, 8] defined *alpha-centrality* $C_{i,j}(\alpha, \beta, t)$ as the total number of attenuated paths between nodes i and j, with β and α giving the attenuation factors [46] along direct edges (from i) and indirect edges (from intermediate nodes) in the path from i to j, respectively, with t specifying the length of the longest path.

Given the adjacency matrix of the network A, alpha-centrality matrix is defined as follows:

$$C(\alpha, \beta, t) = \beta A + \beta \alpha_1 A^2 + \cdots + \beta \big(\prod_{k=1}^{t} \alpha_k \big) A^{t+1} \quad (1)$$

The first term gives the number of paths of length one (edges) from i to j, the second gives the number of paths of length two, etc. Although α_k along different edges in a path could in principle be different, for simplicity, we take them all to be equal: $\alpha_k = \alpha, \forall k$. In this case, the series converges to $C(\alpha, \beta, t \to \infty) = \beta A (I - \alpha A)^{-1}$, which holds while $\alpha < 1/\lambda_1$, where λ_1 is the largest characteristic root of A [23]. The computation of λ_1 is difficult, especially for large networks, which include most complex real-world networks.

To get around this difficulty, we define *normalized alpha-centrality* matrix as:

$$\mathcal{N}(\alpha, \beta, t \to \infty) = \frac{C(\alpha, \beta, t \to \infty)}{\sum_{ij} C_{ij}(\alpha, \beta, t \to \infty)} \quad (2)$$

As we show in the Appendix, in contrast to alpha centrality, normalized alpha-centrality is not bounded by λ_1. Also, we prove that, assuming $|\lambda_1|$ is strictly greater than any other eigenvalue, $\lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}(\alpha, \beta, t \to \infty)$ exists; and as α increases, $\mathcal{N}(\alpha, \beta, t \to \infty)$ remains finite and converges to this value.

Just like the original alpha-centrality, normalized alpha-centrality contains a tunable parameter α that sets the length scale of interactions. For $\alpha = 0$, (normalized) alpha-centrality takes into account direct edges only. As α increases, $\mathcal{N}(\alpha, \beta, t \to \infty)$ becomes a more global measure, taking into account ever larger network components. The expected path length, or radius of centrality, is $(1 - \alpha)^{-1}$.

A. Node Ranking

Much of the analysis done by social scientists considered local structure, i.e., the number [24] and nature [9, 10, 25] of an individual's ties. By focusing on

local structure, however, traditional theories fail to take into account the macroscopic structure of the network. Many metrics proposed and studied over the years deal with this shortcoming, including PageRank [3] and random walk centrality [5]. These metrics aim to identify nodes that are 'close' in some sense to other nodes in the network, and are therefore, more important. PageRank, for example, gives the probability that a random walk initiated at node i will reach j, while random-walk centrality computes the number of times a node i will be visited by walks from all pairs of nodes in the network.

Normalized alpha-centrality, $\mathcal{N}_i(\alpha, \beta, t \to \infty) = \sum_j \mathcal{N}_{ij}(\alpha, \beta, t \to \infty)$, also measures how 'close' node i is to other nodes in a network and can be used to rank the nodes accordingly. The presence of a tunable parameter turns normalized alpha-centrality into a powerful tool for studying network structure and allows us to seamlessly connect the rankings produced by well-known local and global centrality metrics. For $\alpha = 0$, normalized alpha-centrality takes into account local interactions that are mediated by direct edges only, and therefore, reduces to *degree centrality*. As α increases and longer range interactions become more important, nodes that are connected by longer paths grow in importance. For $\alpha < 1/\lambda_1$, the rankings produced by normalized alpha-centrality are equivalent to those produced by alpha-centrality. Also as shown in the Appendix, for symmetric matrices, as $\alpha \to 1/|\lambda_1|$, normalized alpha-centrality converges to *eigenvector centrality* [6]. The rankings no longer change as α increases further, since α has reached some fundamental length scale of the network.

B. Community Detection

Newman & Girvan [18] proposed *modularity* as a metric for evaluating community structure of a network. The modularity-optimization class of community detection algorithms [16, 17, 26] finds a network division that maximizes the modularity, which is defined as $Q =$ (connectivity within community)-(expected connectivity), where connectivity is measured by the density of edges. We extend this definition to use normalized alpha-centrality as the measure of network connectivity. According to this definition, in the best division of a network, there are more weighted paths connecting nodes to others within their own community than to nodes in other communities. Modularity can, therefore, be written as:

$$Q(\alpha) = \sum_{ij} [\mathcal{N}_{ij}(\alpha, t \to \infty) - \overline{\mathcal{N}}_{ij}(\alpha, t \to \infty)] \delta(s_i, s_j)$$

$$(3)$$

$\mathcal{N}_{ij}(\alpha, t \to \infty)$ is given by Eq. (2). We vary α between 0 to 1. Since β factors out of modularity, without loss of generality we take $\beta = 1$. $\overline{\mathcal{N}}_{ij}(\alpha, t \to \infty)$ is the expected normalized alpha-centrality, and s_i is the index of the community i belongs to, with $\delta(s_i, s_j) = 1$ if

$s_i = s_j$; otherwise, $\delta(s_i, s_j) = 0$. We round the values of $\mathcal{N}_{ij}(\alpha, t \to \infty)$ to the nearest integer.

To compute $\overline{\mathcal{N}_{ij}}(\alpha, t \to \infty)$, we consider a graph, referred to as the null model, which has the same number of nodes and edges as the original graph, but in which the edges are placed at random. To make the derivation below more intuitive, instead of normalized alpha-centrality, we talk of the number of attenuated paths. When all the nodes are placed in a single group, then axiomatically, $Q(\alpha) = 0$. Therefore $\sum_{ij}[\mathcal{N}_{ij}(\alpha, t \to \infty) - \overline{\mathcal{N}_{ij}}(\alpha, t \to \infty)] = 0$, and we set $W = \sum_{ij} \overline{\mathcal{N}_{ij}}(\alpha, t \to \infty) = \sum_{ij} \mathcal{N}_{ij}(\alpha, t \to \infty)$. Therefore, according to the argument above, the total number of paths between nodes in the null model $\sum_{ij} \overline{\mathcal{N}_{ij}}(\alpha, t \to \infty)$ is equal to the total number of paths in the original graph, $\sum_{ij} \mathcal{N}_{ij}(\alpha, t \to \infty)$. We further restrict the choice of null model to one where the expected number of paths reaching node j, W_j^{in}, is equal to the actual number of paths reaching the corresponding node in the original graph. $W_j^{in} = \sum_i \overline{\mathcal{N}_{ij}}(\alpha, t \to \infty) = \sum_i \mathcal{N}_{ij}(\alpha, t \to \infty)$. Similarly, we also assume that in the null model, the expected number of paths originating at node i, W_i^{out}, is equal to the actual number of paths originating at the corresponding node in the original graph $W_i^{out} = \sum_j \overline{\mathcal{N}_{ij}}(\alpha, t \to \infty) = \sum_j \mathcal{N}_{ij}(\alpha, t \to \infty)$. W, W_i^{out} and W_j^{in} are then rounded to the nearest integers.

Next, we reduce the original graph G to a new graph G' that has the same number of nodes as G and total number of edges W, such that each edge has weight 1 and the number of edges between nodes i and j in G' is $\mathcal{N}_{ij}(\alpha, \beta, t \to \infty)$. Now the expected number of paths between i and j in graph G could be taken as the expected number of the edges between nodes i and j in graph G' and the actual number of paths between nodes i and j in graph G can be taken as the actual number of edges between node i and node j in graph G'. The equivalent random graph G'' is used to find the *expected* number of edges from node i to node j. In this graph the edges are placed in random subject to constraints: (i) The total number of edges in G'' is W; (ii) The out-degree of node i in $G'' = $ out-degree of node i in $G' = W_i^{out}$; (iii) The in-degree of a node j in graph $G'' = $ in-degree of node j in graph $G' = W_j^{in}$. Thus in G'' the probability that an edge will emanate from a particular node depends only on the out-degree of that node; the probability that an edge is incident on a particular node depends only on the in-degree of that node; and the probabilities of the two nodes being the two ends of a single edge are independent of each other. In this case, the probability that an edge exists from i to j is given by *edge in G' emanates from i · edge in G' incident on j*=$(W_i^{out}/W)(W_j^{in}/W)$. Since the total number of edges is W in G'', therefore the expected number of edges between i and j is $W \cdot (W_i^{out}/W)(W_j^{in}/W) = \overline{\mathcal{N}_{ij}}(\alpha, \beta, t \to \infty)$, the expected α centrality in G.

Once we compute $Q(\alpha)$, we have to select an algorithm to divide the network into communities that maximize $Q(\alpha)$. Brandes et al. [27] have shown that the decision version of modularity maximization is NP-complete. Like others [17, 28], we use the leading eigenvector method to obtain an approximate solution. In this method, nodes are assigned to either of two groups based on a single eigenvector corresponding to the largest positive eigenvalue of the modularity matrix. This process is repeated for each group until modularity does not increase further upon division.

While computational complexity of normalized alpha-centrality (described below) is greater than that of edge density, using an approximate algorithm [14] or parallel programming techniques, e.g., map-reduce [29], can improve the scalability of the modularity-based community detection method that uses alpha-centrality.

C. Relation to Other Centrality Measures

We can generalize the centrality metric presented above to a notion of path-based connectivity and relate it to other centrality metrics. Let $q = (q_{ij})$ be a $n \times n$ matrix such that n is the number of nodes in the network and q_{ij}^t is the number of paths of length t connecting nodes i and j. The number of paths of length one connecting i and j is $q_{ij}^1 = A_{ij}$; the number of paths of length two is $q_{ij}^2 = (A \times A)_{ij}$, etc. The expected number of paths connecting nodes i and j is $E(q_{ij})$ where:

$$\mathbf{E(q)} = (\mathbf{W_1} \cdot \mathbf{q^1} + \mathbf{W_2} \cdot \mathbf{q^2} + \ldots + \mathbf{W_t} \cdot \mathbf{q^t} + \ldots),$$

where $\mathbf{W_k}$ can be a scalar or a vector. $E(q_{ij})$ can be thought of as the proximity score, used to find out how connected or close two nodes in a network are.

Several path-based centrality metrics can be expressed in terms of $E(q_{ij})$, including random walk models [3, 5, 30–32], degree centrality, Katz score [7], as well as alpha-centrality. In a random walk model, a particle starts a random walk at node i, and iteratively transitions to its neighbors with probability proportional to the corresponding edge weight. At each step, the particle returns to i with some restart probability $(1 - c)$. The proximity score is defined as the steady-state probability $r_{i,j}$ that the particle will reach node j [31].

- If $W_k = c^k \cdot D^{-k}$ where c is a constant and D is a diagonal $n \times n$ matrix with $D_{ij} = \sum_{j=1}^t A_{ij}$ if $i = j$ and 0 otherwise; then, $E(q_{ij})$ reduces to proximity score in random walk models [30, 31].

- If $W_k = \Pi_{j=1}^k \alpha_j$, where the scalar α_j is the attenuation factor along the j-th link in the path, then $E(q_{ij})$ reduces to alpha-centrality score from i to j (normalizing leads to normalized alpha-centrality). For ease of computation, we have taken $\alpha_1 = \beta$ and $\alpha_i = \alpha$, $\forall i \neq 1$. Alpha-centrality holds for $\alpha < 1/\lambda_1$ (λ_1 is the largest eigenvalue of A). However, as shown in the Appendix, normalized alpha-centrality holds for all values of α.

- When $\beta = \alpha$, this in turn reduces to the Katz status score [7].

- If $W_k = \alpha$, and adjacency matrix A is symmetric, then as $\alpha \to 1/\lambda_1$, $E(q)$ is proportional to the inner product of the eigenvector corresponding to λ_1, with itself. It would lead to eigenvector centrality as shown in Appendix.

- When $W_1 = 1$ and $W_k = 0$, $\forall k > 1$, then $E(q_{ij})$ is the degree centrality used in modularity-maximization approaches [16].

We have implemented [13] a simple iterative algorithm to compute alpha-centrality:

$$C(\alpha, \beta, t+1) = \beta A + \alpha C(\alpha, \beta, t) A \qquad (4)$$

For any given value of α, this method iteratively computes $C(\alpha, \beta, t \to \infty)$ and consequently $\mathcal{N}(\alpha, \beta, t \to \infty)$ until convergence. Experimentally we have observed that for $\alpha < 1/|\lambda_1|$ this method converges after fewer than 10 iterations. Considering a network with n nodes and m links, each iteration of this algorithm (for a given value of α) has a runtime complexity of $O(mn)$ and space complexity of $O(m)$.

In order to study how normalized alpha-centrality varies with α before convergence (i.e., for $\alpha < 1/|\lambda_1|$), we must choose a small enough increment as α step size. According to Gershgorin circle theorem [33], $|\lambda_1| \leq \min(d_{max}^{out}, d_{max}^{in})$, where d_{max}^{out} and d_{max}^{in} are the maximum out- and in-degrees of the network respectively. Therefore, α step size is of order $\propto 1/\min(d_{max}^{out}, d_{max}^{in})$. For undirected networks, $d_{max}^{out} = d_{max}^{in}$.

Normalized alpha-centrality can be implemented using the map-reduce paradigm [29], guaranteeing the scalability of this algorithm. In addition, [14] recently presented an approximate algorithm that computes alpha-centrality scores that are guaranteed to be within the specified error δ of the exact values, and proved that the algorithm has runtime complexity of $O(\frac{n}{\delta} d_{max}^{out})$. Using the approximate algorithm, we were able to compute alpha-centrality for a network with 70,000 nodes and 30 million edges on a single desktop computer.

III. EMPIRICAL RESULTS

We apply the formalism developed above to benchmark networks studied in literature and a network extracted from the social photosharing site Flickr.

A. Karate Club Network

First, we study the friendship network of Zachary's karate club [34] shown in Figure 1. During the course of the study, a disagreement developed between the administrator and the club's instructor, resulting in the division of the club into two factions, represented by circles and squares in Figure 1. We find community division of this network for different values of α. The first bisection of the network results in two communities, regardless of the value of α, which are identical to the two factions observed by Zachary. However, when the algorithm runs to termination (no more bisections are possible), different groups are found for different values of α. For $\alpha = 0$, the method reduces to edge-based modularity maximization [26] and leads to four groups [15, 35] (Figure 1(a)). For $0 < \alpha < 0.14$ it discovers three groups (Figure 1(b)), and for $\alpha > 0.14$, two groups that are identical to the factions found by Zachary (Figure 1(c)). Thus, increasing α allows local groups to merge into more global communities. Note that the minimum value of α for which two groups are found is close to the reciprocal of the largest eigenvalue of the adjacency matrix of this network, which is 0.1487.

Figure 2 shows how the normalized alpha-centrality scores of nodes change with α. Varying α allows us to smoothly transition from a local measure of centrality to a global measure. For $\alpha = 0$, normalized alpha-centrality reproduces the rankings given by degree centrality. As we show in the appendix, the final rankings produced by normalized alpha-centrality for this symmetric matrix are the same as those given by eigenvector centrality. This is confirmed by Fig. 3, which plots (converged) normalized alpha-centrality and eigenvector centrality scores of karate club nodes. Note that the scale of eigenvector centrality is given by the right hand y-axis. The values overlap almost perfectly after rescaling the axis.

Nodes 34 and 1 have the highest centrality scores, especially at lower α values. These are the *leaders* of their communities. It was the disagreement between these nodes, the club administrator (node 1) and instructor (node 34), that led to the club's division. Nodes 33 and 2 also have high centrality and hold leadership positions. All these nodes are also scored highly by betweenness centrality and PageRank. Note that centrality scores of these nodes decrease with α, indicating that they are far more important locally than globally.

A node may also have high centrality if it is connected to many nodes from different communities. Such nodes, which *bridge* communities, are crucially important to maintaining cohesiveness and facilitating communication flow in both human [9, 10] and animal [12] groups. We can identify these nodes because their normalized alpha-centrality increases with α, i.e., they become more important as longer paths become more important. Centrality of nodes 3, 14, 9, 31, 8, 20, 10, etc., increases with α from moderate to relatively high values. While most of these nodes are directly connected to both communities, some are only indirectly connected by longer paths. Other centrality metrics fail to detect such nodes. For example, betweenness centrality of these nodes is low, though non-zero.

Nodes 25, 26 and 17 have low centrality which decreases with α. These are *peripheral* members. Between-

(a) $\alpha = 0$ (b) $0 < \alpha < 0.14$ (c) $\alpha \geq 0.14$
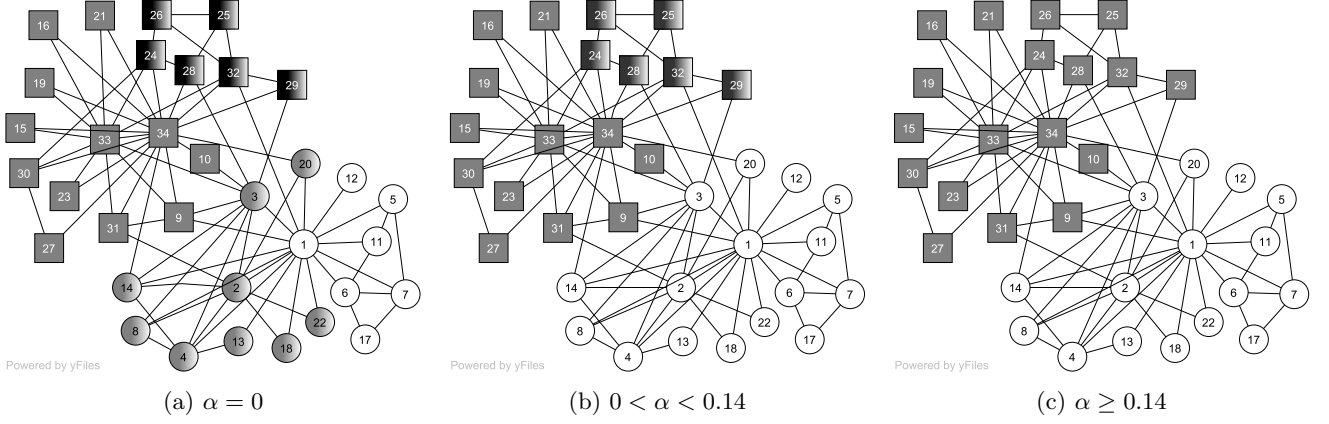
FIG. 1. Zachary's karate club data. Circles and squares represent the two actual factions, while colors stand for discovered communities as the strength of ties increases: (a) $\alpha = 0$, (b) $0 < \alpha < 0.14$, (c) $\alpha \geq 0.14$.
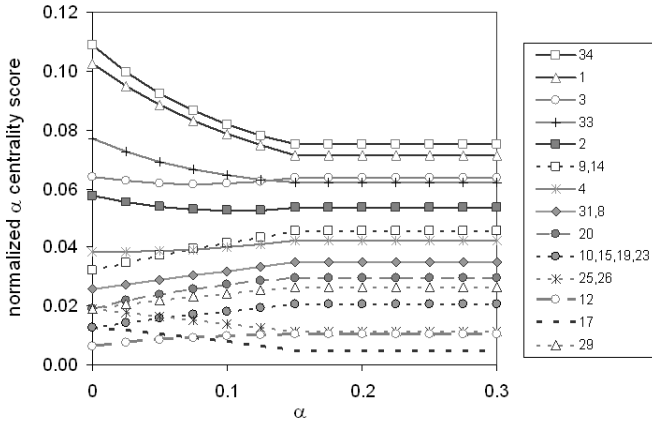

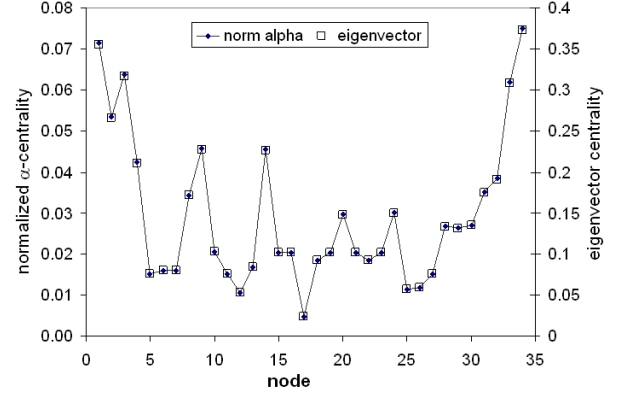FIG. 2. Centrality scores of Zachary club members vs. α.

FIG. 3. Comparison of eigenvector centrality and converged normalized alpha-centrality scores of nodes in the Zachary's karate club network.

ness centrality of 17 is zero, as expected, but 25 and 26 have scores similar to 31. PageRank scores of these peripheral nodes are higher than nodes 21, 22, 23, which are connected to central nodes, and comparable to scores of the bridging nodes 20 and 31. While both betweenness centrality and PageRank correctly pick out leaders, they do not distinguish between locally and globally connected nodes.

Guimera and collaborators [21] proposed a role-based description of complex networks as an alternative to the 'average description' approach, which characterizes network structure in terms of average degree or degree distribution. They define a role in terms of the relative within-community degree z (which measures how well the node is connected to other nodes in its community) and participation coefficient P (which measures how well the node is connected to nodes in other communities). They propose a heuristic classification scheme to assign roles to nodes based on where they fall in the z–P plane and find similar patterns of role-to-role connectivity among networks with similar functional needs and growth mech-

anisms [22].

Figure 4 shows the positions of nodes in the karate club network in the z–P plane. Shaded regions demarcate the boundaries of different roles according to Guimera et $al.$'s classification scheme. Nodes separate into provincial hubs (34, 1), peripheral (33, 2, 28, 14, 31, 29, 20, 3, 9, 10) and ultra-peripheral nodes (rest of the nodes). No special role is assigned to the bridging nodes, such as 9. Even if the boundary of non-hub connectors is shifted to slightly less than $P = 0.5$ in order to identify nodes 3, 9, 10 as serving a special role, the method would still miss node 14, whose position in the network is very similar to node 9. This is because the method takes into account direct links only, rather than complete connectivity between nodes. The method also requires one to first identify communities in the network, which is a computationally expensive for large networks. Our method, on the other hand, provides a computationally scalable way to characterize local and global network structure without the added expense of community identification.
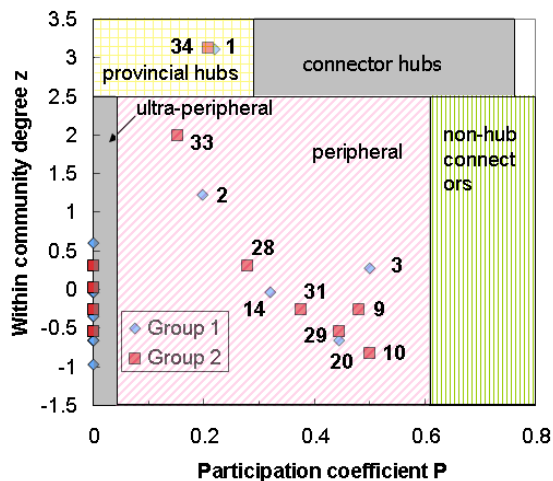
FIG. 4. (Color online) Classification of karate club nodes according to the roles scheme proposed by Guimera et al. [21]: (i) non-hubs ($z < 2.5$) are divided into *ultra-peripheral*, *peripheral*, and *connector* nodes (*kinless* nodes whose links are homogeneously distributed among all communities are not shown); (ii) hubs ($z \geq 2.5$) are subdivided into *provincial* (majority of link within their own community), *connector hubs* (many links to other communities). *Global hubs* whose links are homogeneously distributed among all communities are not shown.

B. Florentine Families

Padgett [36] studied the structure of political and business relationships among the elite families of Renaissance Florence. The two rival factions during this period were the *oligarchs*, composed of the patrician families, and the *Mediceans*, who formed close ties with the "new men," the newly powerful businessmen. Before the rise of Medicis, the oligarchs dominated Florentine politics and economics and cemented their power through marriage. They were less willing to enter into business relationships with the "new men." The Medicis, on the other hand, consolidated their power through business and marriage relationships. Scholars have studied the business and marriage networks of Renaissance Florence to explain the outcomes of the power struggles between the factions and the rise of the Medici family during this important period of Western European history.

We applied our community detection algorithm to the heterogenous network containing both marriage and business ties shown in Figure 5. The marriage ties are shown by solid lines. The dashed lines show the different business relations. The marriage ties are asymmetric, with the wife-giving family being considered superior to the wife-receiving family. All relations are weighted equally. We symmetrized the resulting adjacency matrix by adding it to its transpose. We studied community division of this network for difference values of α. For $0 \leq \alpha \leq 0.05$ we found seven distinct groups, shown

in Fig. 5(a), for $0.05 \leq \alpha \leq 0.1$ we found six groups (Fig. 5(b)), and for $0.1 \leq \alpha \leq 0.25$ five groups (Fig. 5(c)). In every division of the network, two of the groups are small and disconnected from the rest of the network. The first of these is composed of Guadigni, Fioravanti and Bischeri families and the second of Orlandini and Davazati families. These small groups are not shown in Fig. 5.

The three largest groups within the connected component are: (1) families aligned with Medici, (2) families aligned with the oligarchs, such as Strozzi, Peruzzi, and (3) mostly oligarch families with split loyalties, like the Alibizzi. For lower values of α, the oligarchs are split into two groups (Fig. 5(a), (b)). The rift within the oligarchs detected by our algorithm is corroborated by historic events. When a lottery randomly produced too many Medici officeholders in the Signoria (1433), Rinaldo Albizzi, the titular head of the oligarchs, sent out a word to assemble troops in order to forcibly seize Signoria from the Medicis. However, his repeated efforts to assemble troops (especially from Palla Strozzi) were frustrated by other supporters' changing their minds and drifting away [36], indicating factional split within the oligarchs. In contrast, Medicis could immediately and effectively mobilize their supporters, as the result of which no military action ensued and Cosimo Medici took over the budding Florentine state.

As we increase the scale of interactions by increasing α, the five groups within the connected component gradually coalesce into three distinct, as shown in Fig. 5. First, the group comprising of Guasconi, Da-Uzzano and Ardinghelli merge with the Medicis (Fig. 5(b)). When α is further increased, the group comprising of Rondinelli, Solosmei and Della Casa merge with the oligarchs (Fig. 5(c)).

Figure 6 shows how the normalized alpha-centrality scores of the families in the heterogeneous business-marriage network change with α. Guasconi family has the highest centrality score across all values of α. This is not surprising given this family's central position in the network bridging the oligarchs and the Mediceans. This observation is corroborated by the findings that cross-pressurised (by the Mediceans and the oligarchs) Guasconis were split in their partisan loyalties [36]. Similarly, the Medicis, who were able to expertly exploit both business and marriage connections, increase in importance as α increases. On the other hand, the oligarchs such as Strozzi and Peruzzi families, who were patrician to the core and had few business relations outside of their faction, see their centrality decrease with α. Thus, the historic ascendance of the Mediceans can be observed already in the business and marriage networks they created.

Although the heterogenous network of Florentine families is not undirected, since it contains asymmetric marriage relations, we find that rankings produced by alpha-centrality as $\alpha \to 1/|\lambda_1|$ are well correlated with those produced by eigenvector centrality. However, as observed by Bonacich [6], for asymmetric marriage network, there

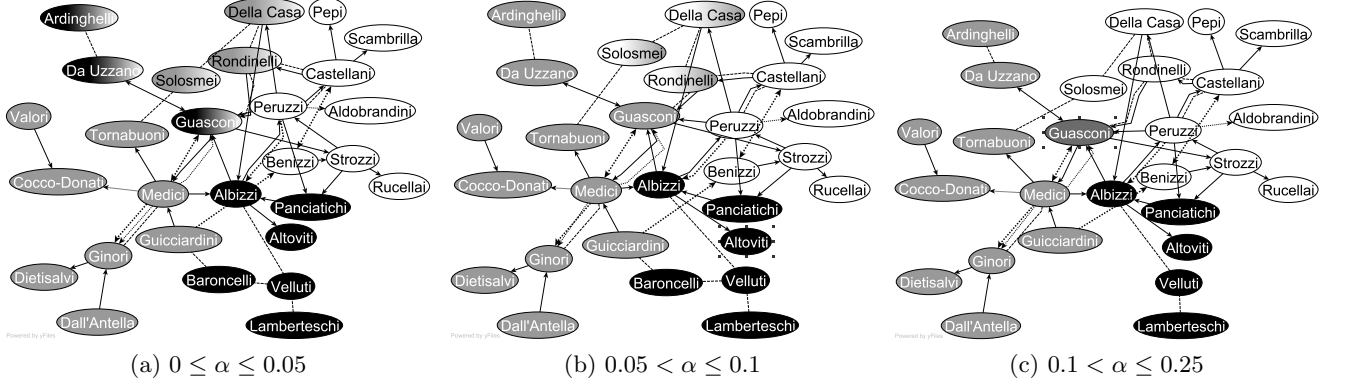(a) $0 \leq \alpha \leq 0.05$ (b) $0.05 < \alpha \leq 0.1$ (c) $0.1 < \alpha \leq 0.25$

FIG. 5. The groups in the Florentine family data set: (a) $0 \leq \alpha \leq 0.05$ (b) $0.05 < \alpha \leq 0.1$ (c) $0.1 < \alpha \leq 0.25$. Two disconnected groups with two and three families are not shown. Solid lines show the marriage ties, while dashes lines show different types of business ties.
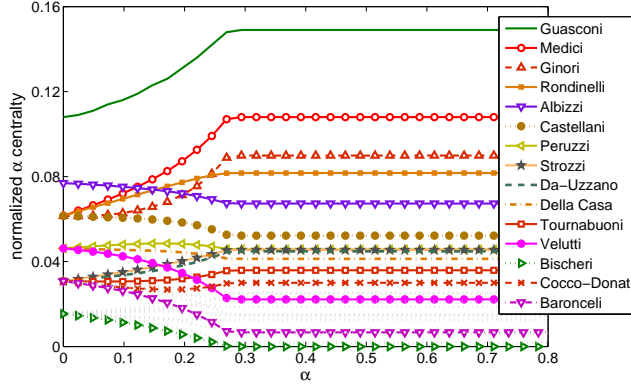

FIG. 6. (Color online) Normalized alpha-centrality scores of the families within the business-marriage network. Some of the nodes are identified, with the rest shown as dashed light grey lines.

C. Other Real-World Networks

In addition to the social networks described above, we evaluated the performance of our community division algorithm on three other real-world networks: the US College football and the political books networks, as well as the social network retrieved from the social photosharing site Flickr. We were not able to evaluate rankings due to the lack of ground truth for these data sets.

The first network represents the schedule of Division 1 games for the 2001 season where the nodes represent teams and the edges represent the regular season games between teams [37]. The teams are divided into conferences containing 8 to 12 teams each. Games are more frequent between members of the same conference, though inter-conference games also take place. This leads to an intuition, that the natural communities may be larger than conferences.

The political books network represents books about US politics sold by the online bookseller Amazon[47]. Edges represent frequent co-purchasing by the same buyers, as indicated by the "customers who bought this book also bought these other books" feature of Amazon. The nodes were labeled *liberal*, *neutral*, or *conservative* by Mark Newman on a reading their descriptions and reviews on Amazon[48]. We take these labels as communities.

To collect the final data set, we sampled Flickr's social network by identifying roughly 2000 users interested in one of three topics: *portraiture*, *wildlife*, and *technology*. We used the Flckr API to perform a tag search using relevant keywords to retrieve 500 'most interesting' images for each topic and extracted the names of users who uploaded these images [49]. Further, we identified four users (eight for the *wildlife* topic) who were interested in each topic by studying their profiles, specifically group membership and user's tags. Groups such as "Big Cats", "Zoo", "The Wildlife Photography", etc. pointed to user's interest in *wildlife*. In addition, tags that users attached to their images could also help identify their interests. Users who used nature and macro tags were probably interested *wildlife* rather than *technology*. Similarly, users interested in human, rather than animal, *portraiture* tagged their images with baby and family. We then used Flickr API to retrieve these users' contacts, as well as their contacts' contacts, and labeled all by the topic through which they were discovered. We reduced this network to an undirected network of mutual contacts only, resulting in a network of 5747 users,

are important differences between the rankings of eigenvector centrality and alpha-centrality. For instance, the eigenvector centrality scores of Bischeri, Guadigni and Orlandini are zero, even though they were wife-giving families. The reason for this is their segregation from the large connected component. This anomaly is corrected by the normalized alpha-centrality, though the scores it gives to Bischeri, Guadigni and Orlandini are very small.

with 1620, 1337 and 2790 users labeled *technology*, *portraiture* and *wildlife* respectively. Although we did not verify that all the users were interested in the topics they were labeled with, we use these 'soft' labels to evaluate the discovered communities.

We use Wallace criterion [38] to evaluate the quality of discovered communities. This metric, which we call *purity*, measures the fraction of correct pairs (belonging to the same community) that are assigned to the same group by the algorithm. We found this to be a very useful metric, particularly in cases for which the exact ground truth is not known and only a "soft" community division is provided; for example, in the football data set a community could be a union of two or more conferences. Note that the purity would be maximum when the entire network is taken as one single community. This would be problematic if the metric was used as the criterion to find groups. However, since we generally find multiple groups using an independent method, this does not pose a problem.

TABLE I. The number and purity of communities discovered at different values of α

karate club			football			flickr		
α	grps	Pu	α	grps	Pu	α	grps	Pu
0.00	4	0.505	0.00	8	0.715	0.000	4	0.501
0.12	3	0.736	0.02	8	0.723	0.001	3	0.565
0.14	2	1.000	0.04	8	0.723	0.002	3	0.567
florentine			0.06	7	0.723	0.003	3	0.567
0.00	7	0.34	0.08	7	0.723	0.004	3	0.567
0.05	6	0.34	0.10	7	0.791	0.005	3	0.568
0.10	5	0.42	0.12	6	0.803	0.006	3	0.570
political books			0.14	6	0.813	0.007	3	0.571
0.00	4	0.633	0.16	6	0.813	0.008	3	0.572
0.04	3	0.805	0.18	4	0.862	0.009	3	0.574
0.08	2	0.917						

The number and purity of the communities found in networks as a function of the parameter α are shown in Table I. The $\alpha = 0$ case corresponds to the standard modularity method. As α increases, the number of groups discovered in all networks goes down, while their purity increases. This is consistent with our hypothesis that using smaller values of α allows us to identify more local network structure, while larger values of α lead to more global structure. In the Karate club network, for example, at $\alpha = 0$, there are four small communities, as shown in Fig. 1(a). These local communities coalesce into two large groups as α increases (Fig. 1(c)), which are identical to the groups identified by Zachary [34].

To evaluate the communities discovered in the Florentine families network, we use the tight constraint of party loyalty. Hence the families could be either Medicean, oligarch, or have split loyalties. We note that this is a very conservative evaluation criterion, since there were factions present within the parties themselves [36]. Since families with split loyalties would be correctly classified

as belonging to either of the two parties, we remove them from purity calculation, focusing instead on identifying community of party loyalists only. Purity is further reduced by the presence of the two isolated groups. However, purity of discovered communities increases with α. The small local communities found at lower value of α could indicate factions within parties.

IV. RELATED WORK

A variety of metrics have been proposed to measure node's centrality in a network [1, 3, 5–7, 24, 39, 40], yet few studies systematically evaluated their performance on real-world networks. Liben-Nowell and Kleinberg [41] compared the performance of several commonly used centrality metrics on the link prediction task and found Katz score [7] to be the most effective measure for this task, outperforming PageRank [3] and its variants. The alpha-centrality metric modifies the Katz score by introducing a parameter α, that gives a weight to indirect links and also sets the length scale of interactions in the network. We showed recently [13] that normalized alpha-centrality outperforms other centrality metrics on the task of predicting influential nodes in an online social network.

Guimera and collaborators [21, 22] proposed role-based description of complex networks. They define a role in terms of the relative within-community degree z (which measures how well the node is connected to other nodes in its community) and participation coefficient P (which measures how well the node is connected to nodes in other communities). They proposed a heuristic classification scheme based on where the nodes lie in the z–P plane. This classification scheme is similar to the local vs. globally-connected distinction we are making, with connector nodes being more globally connected nodes while provincial hubs and peripheral nodes are more locally connected. Role-based analysis requires community decomposition of the network to be performed first. This is a computationally expensive procedure for most real-world networks. Our approach, on the other hand, allows us to differentiate between roles of nodes in a more computationally efficient way.

Community detection is another active area in networks research (see [15] for a comprehensive review). Like us, Arenas et al. [42] have generalized modularity to find correlations between nodes that go beyond nearest neighbors. Their approach relies on the presence of motifs [19, 20], i.e., connected subgraphs such as cycles, to identify communities within a network. For example, higher than expected density of triangles implies presence of a community, and a triangle modularity may be defined to identify it. The motif-based modularity uses the size of the motif to impose a limit on the proximity of neighbors. Our method, on the other hand, imposes no such limit. The measure of global correlation computed using alpha-centrality is equal to the weighted average of correlations for motifs of different sizes. Our method

enables us to easily calculate this complex term.

V. CONCLUSION

In this paper, we introduced normalized alpha-centrality as a metric to study network structure. Like the original alpha-centrality [6] on which it is based, this metric measures the number of paths that exist between nodes in a network, attenuated by their length with the attenuation parameter α. This parameter sets the length scale of the interaction. When $\alpha = 0$, the centrality metric takes into account direct edges only and is equivalent to degree centrality. As α increases, the metric takes into consideration more distant network interactions, becoming a more global measure. Normalized alpha-centrality allows us to smoothly interpolate between local metrics, such as degree centrality, and global metrics, such as eigenvector centrality [6]. Unlike the original alpha-centrality, which bounds α to be less than the reciprocal of the largest eigenvalue of the adjacency metric of the network, normalized alpha-centrality sets no such limit.

We used normalized alpha-centrality to study the structure of networks, specifically, identify important nodes and communities within the network. We extended the modularity maximization class of algorithms [37] to use (normalized) alpha-centrality, rather than edge density, as a measure of network connectivity. For small values of α, smaller, more locally connected communities emerge, while for larger values of α, we observe larger globally connected communities. We also used this metric to rank nodes in a network. By studying changes in rankings that occur when parameter α is varied, we were able to identify locally important 'leaders' and globally important 'bridges' or 'brokers' that facilitate communication between different communities. We applied this approach to benchmark networks studied in literature and found that it results in network division in close agreement with the ground truth. We can easily extend this definition to multi-modal networks that link entities of different types, and use approach described in this paper to study the structure of such networks [43].

ACKNOWLEDGMENTS

Appendix A: Proofs of Convergence

In this section we prove some properties of the normalized alpha-centrality metric proposed in this paper. The *alpha-centrality matrix* $C(\alpha, \beta, t)$ $\forall \alpha \in [0, 1]$ is defined as:

$$C(\alpha, \beta, t) = \beta A(I + \alpha A + \alpha^2 A^2 + \cdots + \alpha^t A^t)$$

$$= \beta A \sum_{k=0}^{t} \alpha^k A^k \tag{A1}$$

The *normalized alpha-centrality matrix* is then given by:

$$\mathcal{N}(\alpha, \beta, t) = \frac{C(\alpha, \beta, t)}{\sum_{i,j} (C(\alpha, \beta, t))_{ij}} \tag{A2}$$

The normalized alpha-centrality vector is $\mathcal{N}_i(\beta, \alpha, t \to \infty) = e\mathcal{N}(\alpha, \beta, t \to \infty)$ where e is a $1 \times n$ unit vector and n is the number of nodes in the network.

If λ is an *eigenvalue* of A, then

$$(I - \frac{1}{\lambda}A)x = 0 \tag{A3}$$

Invertibility of $(I - \frac{1}{\lambda}A)$ would lead to the trivial solution of eigenvector x ($x = 0$). Hence for computation of eigenvalues and eigenvectors, we require that no inverse of $(I - \frac{1}{\lambda}A)$ should exist, i.e.

$$Det(I - \frac{1}{\lambda}A) = 0. \tag{A4}$$

Equation (A4) is called the *characteristic equation*, solving which gives the *eigenvalues* and *eigenvectors* of the adjacency matrix A.

The adjacency matrix A can be written as:

$$A = X\Lambda X^{-1} = \sum_{i=1}^{n} \lambda_i Y_i \tag{A5}$$

where X is a matrix whose columns are the eigenvectors of A, and Λ is a diagonal matrix whose diagonal elements are the eigenvalues of A, $\Lambda_{ii} = \lambda_i$, arranged according to the ordering of the eigenvectors in X. Without loss of generality we assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. The matrices Y_i can be determined from the product

$$Y_i = X Z_i X^{-1} \tag{A6}$$

where Z_i is the *selection matrix* having zeros everywhere except for element $(Z_i)_{ii} = 1$ [44].

Adjacency matrix A raised to kth power is given by

$$A^k = X\Lambda^k X^{-1} = \sum_{i=1}^{n} \lambda_i^k Y_i \tag{A7}$$

Using Equation (A7), A1 reduces to

$$C(\alpha, \beta, t) = \beta A \sum_{i=1}^{n} \sum_{k=1}^{t} \alpha^k \lambda_i^k Y_i$$

$$= \beta A \sum_{i=1}^{n} \lambda_i \frac{(-1)^{p_i}(1 - \alpha^{t+1}\lambda_i^{t+1})}{(-1)^{p_i}(1 - \alpha\lambda_i)} Y_i \tag{A8}$$

where $p_i = 0$ if $\alpha |\lambda_i| < 1$, and $p_i = 1$ if $\alpha |\lambda_i| > 1$. For the equations A1 and A8 to hold non-trivially, $\alpha \neq 1/|\lambda_i|$, $\forall i \in 1, 2 \cdots, n$.

We characterize the series $\{\mathcal{N}(\alpha, \beta, t \to \infty)\}$ for $\alpha \in [0, 1]$ as follows:

1. $\alpha \ll \frac{1}{|\lambda_1|}$: If $\alpha \ll \frac{1}{|\lambda_1|}$, $C(\alpha, \beta, t \to \infty)$ (and $\mathcal{N}(\alpha, \beta, t \to \infty)$) would be independent of α, since

$$C(\alpha, \beta, t \to \infty) \approx \beta A$$
$$\mathcal{N}(\alpha, \beta, t \to \infty) \approx \frac{A}{\sum_{ij}(A)_{ij}} \qquad (A9)$$

2. $\alpha < \frac{1}{|\lambda_1|}$: The sequence of matrices $\{C(\alpha, \beta, t)\}$ converges to $C(\alpha, \beta)$ as $t \to \infty$ if all the sequences $\{(C(\alpha, \beta, t))_{ij}\}$ for every fixed i and j converge to $(C(\alpha, \beta))_{ij}$ [45]. If $\alpha < \frac{1}{|\lambda_1|}$, $C(\alpha, \beta, t)$ converges to $C(\alpha, \beta)$.

$$C(\alpha, \beta, t \to \infty) = \beta A \sum_{i=0}^{n} \frac{\lambda_i}{1 - \alpha \lambda_i} Y_i = \beta A (I - \alpha A)^{-1}$$
$$= C(\alpha, \beta) \qquad (A10)$$
$$\mathcal{N}(\alpha, \beta, t \to \infty) = \frac{C(\alpha, \beta)}{\sum_{ij}(C(\alpha, \beta))_{ij}} \qquad (A11)$$

3. $\alpha > \frac{1}{|\lambda_1|}$ and $t \to \infty$, $\beta \alpha^t A^{t+1}$ dominates in the Equation (A8).

$$C(\alpha, \beta, t \to \infty) \approx \beta \alpha^t A^{t+1}$$
$$\mathcal{N}(\alpha, \beta, t \to \infty) \approx \frac{A^{t+1}}{\sum_{i,j} A_{ij}^{t+1}} \qquad (A12)$$

Theorem 1 *The induced ordering of nodes due to normalized alpha-centrality is equal to the induced ordering of nodes due to alpha-centrality for $\alpha < 1/|\lambda_1|$.*

Proof. Since centrality score due to alpha-centrality is $eC(\alpha, \beta, t \to \infty)$ and that due to normalized alpha-centrality is $e\mathcal{N}(\alpha, \beta, t \to \infty)$, where e is a unit vector, from equations A9 and A11, the ordering of nodes produced by alpha-centrality ($\alpha < 1/|\lambda_1|$) would be equal to the ordering of nodes produced by normalized alpha-centrality ($\alpha < 1/|\lambda_1|$).

Theorem 2 *The value of normalized alpha-centrality matrix remains the same $\forall \alpha \in (1/|\lambda_1|, 1]$ ($\mathcal{N}(\alpha > 1/|\lambda_1|, \beta, t \to \infty) = \mathcal{N}'(\alpha, \beta)$).*

Proof. As can be seen from equation A12 when $\alpha > 1/|\lambda_1|$ and $t \to \infty$, $\mathcal{N}(\alpha > 1/|\lambda_1|, \beta, t \to \infty)$ reduces to $A^{t+1}/\sum_{i,j} A_{ij}^{t+1} = \mathcal{N}'(\alpha, \beta)$ and is independent of α.

The remaining theorems hold under the condition that $|\lambda_1|$ is strictly greater than any other eigenvalue, which is true in most real life cases studied.

Theorem 3 $\lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}_i(\alpha, \beta, t \to \infty)$ *exists and* $\lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}_i(\alpha, \beta, t \to \infty) = \mathcal{N}_i'(\alpha, \beta) = \mathcal{N}_i(\alpha > 1/|\lambda_1|, \beta, t \to \infty) = eAY_1/\sum_{i,j}(AY_1)_{ij}$.

Proof. Under the assumption that $|\lambda_1|$ is strictly greater than any eigenvalue, as $\alpha \to 1/|\lambda_1^-|$, Equation (A11) reduces to

$$C(\alpha \to 1/|\lambda_1^-|, \beta, t \to \infty) \approx \frac{\beta \lambda_1}{1 - \alpha \lambda_1} AY_1 \qquad (A13)$$

This is because all other eigenvectors shrink in importance as $\alpha \to 1/|\lambda_1^-|$ [6]. Therefore, as $\alpha \to 1/|\lambda_1^-|$, we have

$$\mathcal{N}(\alpha \to 1/|\lambda_1^-|, \beta, t \to \infty) \approx \frac{AY_1}{\sum_{i,j}(AY_1)_{ij}} \qquad (A14)$$

Under the assumption that $|\lambda_1|$ is strictly greater than any other eigenvalue, $\beta \alpha^t \lambda_1^n AY_1$ dominates in the Equation (A8), A12.

$$C(\alpha \to 1/|\lambda_1^+|, \beta, t \to \infty) \approx \beta \alpha^t \lambda_1^n AY_1$$
$$\mathcal{N}(\alpha \to 1/|\lambda_1^+|, \beta, t \to \infty) \approx \frac{AY_1}{\sum_{i,j}(AY_1)_{ij}} \qquad (A15)$$

Hence from equation A15, as $\alpha \to 1/|\lambda_1^+|$, we have

$$\mathcal{N}(\alpha \to 1/|\lambda_1^+|, \beta, t \to \infty) \approx \frac{AY_1}{\sum_{i,j}(AY_1)_{ij}} \qquad (A16)$$

Since,

$$\lim_{\alpha \to 1/|\lambda_1^-|} \mathcal{N}(\alpha, \beta, t \to \infty) = \lim_{\alpha \to 1/|\lambda_1^+|} \mathcal{N}(\alpha, \beta, t \to \infty)$$
$$= \frac{AY_1}{\sum_{i,j}(AY_1)_{ij}},$$

therefore, the limit $\lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}(\alpha, \beta, t \to \infty)$ exists and

$$\lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}(\alpha, \beta, t \to \infty) = \frac{AY_1}{\sum_{i,j}(AY_1)_{ij}}$$
$$= \mathcal{N}_i'(\alpha, \beta)$$

Since $\mathcal{N}_i(\alpha, \beta, t \to \infty) = e\mathcal{N}(\alpha, \beta, t \to \infty)$, therefore, $\lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}_i(\alpha, \beta, t \to \infty) = \mathcal{N}_i'(\alpha, \beta) = \mathcal{N}_i(\alpha > 1/|\lambda_1|, \beta, t \to \infty) = eAY_1/\sum_{i,j}(AY_1)_{ij}$.

Theorem 4 *For symmetric matrices, the induced ordering of nodes due to eigenvector centrality C_E is equivalent to the induced ordering of nodes given by normalized centrality $\mathcal{N}_i'(\alpha, \beta) = \lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}_i(\alpha, \beta, t \to \infty) = \mathcal{N}_i(\alpha > 1/|\lambda_1|, \beta, t \to \infty) = eAY_1/\sum_{i,j}(AY_1)_{ij}$.*

Proof. For symmetric matrices

$$A = X\Lambda X^{-1} = X\Lambda X^T \qquad (A17)$$

Therefore equation A6 reduces to $Y_i = XZ_iX^T = X_iX_i^T$, where X_i is the column of X representing the eigenvector corresponding to λ_i. Hence, for symmetric matrices

$$\begin{aligned}
\mathcal{N}'_i(\alpha,\beta) &= \mathcal{N}_i(\alpha > 1/|\lambda_1|, \beta, t \to \infty) \\
&= \lim_{\alpha \to 1/|\lambda_1|} \mathcal{N}_i(\alpha, \beta, t \to \infty) \\
&= \frac{eAY_1}{\sum_{i,j}(AY_1)_{ij}} \\
&= c_1 eAX_1X_1^T = c_2 X_1^T \qquad (A18)
\end{aligned}$$

where $c_1 = \frac{1}{\sum_{i,j}(AY_1)_{ij}}$ and $c_2 = c_1 eAX_1$.

Since X_1^T corresponds to the eigenvector centrality vector C_E, hence for symmetric matrices, the induced ordering of nodes given by eigenvector centrality C_E is equivalent to the induced ordering of nodes given by normalized centrality $\mathcal{N}'_i(\alpha,\beta) = \lim_{\alpha \to 1/|\lambda_1|}\mathcal{N}_i(\alpha,\beta,t\to\infty) = \mathcal{N}_i(\alpha > 1/|\lambda_1|,\beta,t\to\infty) = eAY_1/\sum_{i,j}(AY_1)_{ij}$.

[1] L. C. Freeman, Social Networks **1**, 215 (1979).
[2] K. Stephenson and M. Zelen, Social Networks **11**, 1 (1989).
[3] L. Page, S. Brin, R. Motwani, and T. Winograd, Tech. Rep., Stanford Digital Library Technologies Project (1998).
[4] J. D. Noh and H. Rieger, Phys. Rev. E **66**, 066127 (2002).
[5] M. Newman, Social Networks **27**, 39 (2005).
[6] P. Bonacich, Social Networks **23**, 191 (2001).
[7] L. Katz, Psychometrika **18**, 39 (1953).
[8] P. Bonacich, Am. J. Sociology **92**, 1170 (1987).
[9] M. Granovetter, Am. J. Sociology **78**, 1360 (1973).
[10] G. Simmel, *The Sociology of Georg Simmel* (Free Press, 1950), chap. Individual and Society.
[11] P. Csermely, Trends in Biochemical Sciences **33**, 569 (2008).
[12] D. Lusseau and M. E. J. Newman, Proc. Royal Society of London. Series B: Biological Sciences **271**, S477 (2004).
[13] R. Ghosh and K. Lerman, in *Proc. KDD workshop on Social Network Analysis (SNA-KDD)* (2010).
[14] R. Ghosh, K. Lerman, T. Surachawala, K. Voevodski, and S.-H. Teng, in *submitted to KDD* (2011), 1102.4639.
[15] S. Fortunato, Phys. Reports **486**, 75 (2010).
[16] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).
[17] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).
[18] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[19] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[20] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 1538 (2004).
[21] R. Guimera and L. A. N. Amaral, Nature **433**, 895 (2005).
[22] R. Guimera, M. Sales-Pardo, and L. A. N. Amaral, Nat Phys **3**, 63 (2007).
[23] W. L. Ferrar, *Finite Matrices* (Oxford Univ. Press, 1951).
[24] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge Univ.Press, 1994).
[25] R. S. Burt, *Structural Holes: The Structure of Competition* (Harvard University Press, Cambridge, MA, 1992).
[26] M. E. J. Newman, The European Physical Journal B **38**, 321 (2004).
[27] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, IEEE Trans. on Knowl. and Data Eng. **20**, 172 (2008), ISSN 1041-4347.
[28] E. A. Leicht and M. E. J. Newman, Phys. Rev. Letters **100**, 118703 (2008).
[29] J. Dean and S. Ghemawat, Commun. ACM **51**, 107 (2008), ISSN 0001-0782.
[30] H. Tong, C. Faloutsos, and J. Pan, Data Mining, 2006. ICDM '06. Sixth International Conference on pp. 613–622 (2006), ISSN 1550-4786.
[31] H. Tong, S. Papadimitriou, P. S. Yu, and C. Faloutsos, in *Proc. SIAM Conference on Data Mining* (2008), pp. 704–715.
[32] H. Zhou, Phys. Rev. E **67**, 041908 (2003).
[33] G. Strang, *Linear Algebra and Its Applications* (Academic Press Inc., Orlando, FL, 1980), second edition ed.
[34] W. W. Zachary, J. Anthropological Research **33**, 452 (1977).
[35] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104+ (2005).
[36] J. F. Padgett and C. K. Ansell, Am. J. Sociology **98**, 1259 (1993).
[37] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA. **99**, 7821 (2002).
[38] D. L.Wallace, Journal of the American Statistical Association **383**, 569 (1983).
[39] C. Hubbel, Sociometry **28**, 377 (1965).
[40] P. Bonacich, J. of Mathematical Sociology **2**, 113 (1972).
[41] D. Liben-Nowell and J. Kleinberg, J. Am. Soc. Inf. Sci. Technol. **58**, 1019 (2007), ISSN 1532-2882.
[42] A. Arenas, A. Fernandez, S. Fortunato, and S. Gomez, Mathematical Systems Theory **41**, 224001 (2008).
[43] R. Ghosh and K. Lerman, in *Proc. 1st IEEE SIGCOM Social Computing Conference (SocialCom09)* (2009).
[44] F. Gebali, Analysis of Computer and Communication Networks p. 65:122 (2008).
[45] P. Dienes, Quart. J. of Math. (Oxford) **3**, 253 (1932).
[46] For some types of networks, e.g., commodity exchange networks, Bonacich allows $\alpha < 0$. In communication and information networks we are considering, $\alpha > 0$.
[47] http://www.orgnet.com/.
[48] http://www-personal.umich.edu/~mejn/netdata/.
[49] The keywords used for image search were (a) newborn for the *portraiture* topic, (b) tiger and beetle for the *wildlife* topic, and (c) apple for the *technology* topic.