Local Fiedler Vector Centrality for Detection of Deep and Overlapping Communities in Networks

Pin-Yu Chen and Alfred O. Hero III, *Fellow, IEEE*

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA

Email : {pinyu,hero}@umich.edu

Abstract—In this paper, a new centrality called local Fiedler vector centrality (LFVC) is proposed to analyze the connectivity structure of a graph. It is associated with the sensitivity of algebraic connectivity to node or edge removals and features distributed computations via the associated graph Laplacian matrix. We prove that LFVC can be related to a monotonic submodular set function that guarantees that greedy node or edge removals come within a factor $1 - 1/e$ of the optimal non-greedy batch removal strategy. Due to the close relationship between graph topology and community structure, we use LFVC to detect deep and overlapping communities on real-world social network datasets. The results offer new insights on community detection by discovering new significant communities and key members in the network. Notably, LFVC is also shown to significantly outperform other well-known centralities for community detection.

I. INTRODUCTION

In social, biological and technological networks [1]–[3], communities are defined as tightly connected modules embedded in a graph [4]–[6]. Communities play an important role in determining collective behaviors. Recently, community detection has attracted a great deal of interest across different fields [5], [7]. Conventional community detection techniques such as the edge betweenness method [1] and the modularity method [8] presume that each node in a graph is affiliated with only one community of a network. However, in reality an individual node can have multiple memberships resulting in overlapping community structure [9]–[11].

Communities can be cast as the collection of remaining connected components/subgraphs when a subset of nodes or edges are removed based on centrality measures (e.g., the isolated community defined in [6]). In [1], edge removal based on edge betweenness centrality is proposed to detect hierarchical community structures. In [12], node removal based on node degree is proposed to improve the performance of the modularity method. In this paper, we are interested in investigating new centrality measures such that their removals come within a factor $1 - 1/e$ of the optimal non-greedy batch removal strategy to aid community detection.

Since current community detection methods must assign one membership to each node in the network, detection performance of such methods may suffer from the presence of "outlier nodes" that prevent reliable community detection. Consequently, significant communities may be neglected or disguised during the detection process in the presence of outlier nodes resulting in biased or inaccurate social network analysis. The goal of this paper is to propose a systematic

approach to detect significant communities and key members in the network, which we refer as deep and overlapping community detection. For instance, when our proposed approach is applied to the network scientist coauthorship dataset we show that a zoologist is correctly identified as an outlier node during the detection process since the authors are mostly physicists, thus leading to revelation of new community structures.

Local Fiedler vector centrality (LFVC) is proposed to evaluate the connectivity structure of a graph based on spectral graph theory [13]. LFVC is associated with an upper bound on algebraic connectivity [14] when a subset of nodes or edges are removed from a graph. We show that LFVC relates to a monotonic submodular set function such that greedy node or edge removals can be employed with bounded performance loss relative to the optimal non-greedy batch removal strategy. Moreover, LFVC can be computed in a distributed manner and it is applicable to large-scale network analysis. We apply this method to real-world social network datasets and compare to the modularity method and other well-known centralities.

II. SPECTRAL GRAPH THEORY AND CENTRALITIES

A. Graph Laplacian and algebraic connectivity

Consider an undirected and unweighted graph $G = (V, E)$ without self loops and multiple edges between any node pair. V is the node set with $|V| = n$ and E is the edge/link set with $|E| = m$. The connectivity structure is characterized by an n-by-n binary and symmetric adjacency matrix A, where $A_{ij} = 1$ if $(i, j) \in E$, otherwise $A_{ij} = 0$. Let $d_i = \sum_{j=1}^{n} A_{ij}$ denote the degree of node i. The degree matrix $D = \text{diag}(\mathbf{d})$ is a diagonal matrix with degree pattern $(d_1, d_2, \ldots d_n)$ on its diagonal. The graph Laplacian of G is defined as $L = D - A$. Let $\lambda_i(L)$ denote the ith smallest eigenvalue of L, $\mathbf{1}$ denote the all one vector, and N_i denote the set of neighboring nodes of i, i.e., $N_i = \{j : A_{ij} = 1\}$. It is well-known [13], [15] that

$$\mathbf{x}^T L \mathbf{x} = \frac{1}{2} \sum_{i,j \in V} A_{ij}(x_i - x_j)^2 = \frac{1}{2} \sum_{i \in V} \sum_{j \in N_i} (x_i - x_j)^2 \quad (1)$$

and $L\mathbf{1} = (D - A)\mathbf{1} = 0$, therefore $\lambda_1(L) = 0$ and L is a positive semidefinite (PSD) matrix.

The algebraic connectivity of a graph is defined as the second smallest eigenvalue of L, i.e., $\lambda_2(L)$. A graph is connected if and only if $\lambda_2(L) > 0$, and, for any non-complete graph [14], $\lambda_2(L)$ is an lower bound on node/edge connectivity, where node/edge connectivity is the least number of node/edge removals that disconnects the graph. Therefore, a graph with larger algebraic connectivity is more resilient to node and edge removals. For any connected graph, we can represent the algebraic connectivity as $\lambda_2(L) = \min_{\|\mathbf{x}\|_2 = 1, \ \mathbf{x} \perp \mathbf{1}} \mathbf{x}^T L \mathbf{x}$

using the Courant-Fischer theorem [16] and the fact that $\mathbf{1}$ is an eigenvector of $\lambda_1(L)$.

The Fiedler vector is the eigenvector of $\lambda_2(L)$ [14], which can be distributedly computed via local information exchange over the graph Laplacian matrix [17]. It has been widely used in graph partitioning, image segmentation and data clustering [18]–[22]. The Fiedler vector facilitates community detection by separating the nodes in the graph according to the signs of the corresponding Fiedler vector elements. Hierarchical community structure can be detected by successively evaluating the Fiedler vector on the successively discovered communities.

Similarly, the modularity method [8], [23] uses the largest eigenvector of a matrix B for community detection, where $B_{ij} = A_{ij} - \frac{d_i d_j}{2m}$ is the number of excessive edges relative to the associated random graph. The modularity is defined as $\max_{s \in \{1,-1\}^n} \frac{1}{4m} s^T B s$, where s is the community indicator vector. To divide a network into more than two communities, Newman proposes that one should compute the incremental modularity of further dividing each community into two communities and select the partition that has the largest modularity increment. It is shown in [23] that the modularity method tends to give more accurate results compared with community detection based on the standard Fielder vector approach. However, unlike the proposed method, these methods can not be applied to overlapping community detection.

B. Some examples of centralities

Centralities can be classified into two categories. Global centralities require complete topological information whereas local centralities only require local information from neighboring nodes for their computation.

Betweenness [24]: betweenness is a global node centrality defined as $\text{betweenness}(i) = \sum_{k \neq i} \sum_{j \neq i, j > k} \frac{\sigma_{kj}(i)}{\sigma_{kj}}$, where σ_{kj} is the total number of shortest paths from k to j and $\sigma_{kj}(i)$ is the number of such shortest paths passing through i.
Closeness [25]: let $\rho(i, j)$ denote the shortest path distance between node i and node j in a connected graph. $\text{closeness}(i) = 1/\sum_{j \in V, j \neq i} \rho(i, j)$ and it is a global node centrality.
Eigenvector centrality (Eigen centrality) [26]: eigenvector centrality of node i refers to the ith entry of largest eigenvector of adjacency matrix A and it is a global node centrality.
Degree (d_i): degree is the simplest local node centrality measure which accounts for the number of neighboring nodes.
Ego centrality [27]: Ego centrality can be viewed as a local version of betweeness centrality that computes the shortest paths between its neighboring nodes. Let $A(i)$ be the local adjacency matrix of i and I be an identity matrix. Since $[A^2(i)]_{kj}$ is the number of two-hop walks between k and j and $A^2(i)[I - A(i)]_{kj}$ is the total number of two-hop shortest paths between k and j, ego centrality is defined as $\text{ego}(i) = \sum_k \sum_{j > k} 1/A^2(i)[I - A(i)]_{kj}$.

III. LOCAL FIEDLER VECTOR CENTRALITY (LFVC)

A. Edge-LFVC

Consider the graph $\widetilde{G} = (V, E \cup (i, j))$ by adding an edge (i, j) to G, where $(i, j) \notin E$ and we denote the resulting graph Laplacian by $\widetilde{L}(i, j)$. Let $\mathbf{e}_i$ be a zero vector except that its ith is equal to 1. We have $\widetilde{L} = L + \Delta L$ and $\Delta L = \Delta D - \Delta A$, where ΔD and ΔA are the augmented degree and adjacency matrices, respectively. We have $\Delta D = \mathbf{e}_i \mathbf{e}_i^T + \mathbf{e}_j \mathbf{e}_j^T$ and $\Delta A =$

$\mathbf{e}_i \mathbf{e}_j^T + \mathbf{e}_j \mathbf{e}_i^T$ and therefore $\widetilde{L}(i, j) = L + (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$. That is, The resulting graph Laplacian $\widetilde{L}(i, j)$ after adding an edge (i, j) to G is the original Laplacian matrix L perturbed by a rank one matrix. Similarly, when a $(i, j) \in E$ is removed from G, we have $\widetilde{L}(i, j) = L - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$.

Consider removing an edge $(i, j) \in E$ in G and following the definition of $\lambda_2(L) = \min_{\|\mathbf{x}\|_2 = 1, \mathbf{x} \perp \mathbf{1}} \mathbf{x}^T L \mathbf{x}$. Let y denote the Fiedler vector of L, computing $\mathbf{y}^T \widetilde{L}(i, j) \mathbf{y}$ gives an upper bound on $\lambda_2(\widetilde{L}(i, j))$

$$\lambda_2(\widetilde{L}(i, j)) \leq \mathbf{y}^T \widetilde{L}(i, j) \mathbf{y} = \lambda_2(L) - (y_i - y_j)^2. \quad (2)$$

It is worth mentioning that for any connected graph G, there exists at least one edge removal such that the inequality $\lambda_2(\widetilde{L}(i, j)) < \lambda_2(L)$ holds, otherwise $y_i = y_j$ for all $i, j \in V$ and this violates the constraints that $\|\mathbf{y}\|_2 = 1$ and $\sum_{i=1}^n y_i = 0$. Consequently, there exists at least one edge removal that leads to a decrease in algebraic connectivity.

Similarly, when we remove a subset of edges $E_R \subset E$ from G, where $|E_R| = h$. The upper bound becomes

$$\lambda_2(\widetilde{L}(E_R)) \leq \lambda_2(L) - \sum_{k=(i,j) \in E_R} (y_i - y_j)^2. \quad (3)$$

We define the local Fiedler vector edge centrality as

$$\text{edge-LFVC}(i, j) = (y_i - y_j)^2. \quad (4)$$

From (2) and (3), the top h edge removals which lead to the most decrease in algebraic connectivity of the remaining graph are the top h edges with the highest edge centralities.

B. Node-LFVC

When a node $i \in V$ is removed from G, all the edges attached to i will be removed from G. Following (2), the resulting graph Laplacian $\widetilde{L}(i)$ can be regarded as a rank d_i perturbation to L. Since $L - \widetilde{L}(i) = \sum_{j \in N_i} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$,

$$\lambda_2(\widetilde{L}(i)) \leq \lambda_2(L) - \sum_{j \in N_i} (y_i - y_j)^2. \quad (5)$$

Similarly, for any connected graph, there exists at least one node removal that leads to a decrease in algebraic connectivity.

If a subset of nodes $R \subset V$ are removed from G, where $|R| = q$. Then

$$L - \widetilde{L}(R) = \sum_{i \in R} \sum_{j \in N_i} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \quad (6)$$
$$- \frac{1}{2} \sum_{i \in R} \sum_{j \in R} A_{ij}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T,$$

where the last term accounts for the edges that are attached to the removed nodes at both ends. Consequently, we obtain an upper bound for multiple node removals

$$\lambda_2(\widetilde{L}(R)) \leq \lambda_2(L) - \sum_{i \in R} \sum_{j \in N_i} (y_i - y_j)^2 \quad (7)$$
$$+ \frac{1}{2} \sum_{i \in R} \sum_{j \in R} A_{ij}(y_i - y_j)^2.$$

We define the local Fiedler vector node centrality as

$$\text{node-LFVC}(i) = \sum_{j \in N_i} (y_i - y_j)^2, \quad (8)$$

which is the sum of the square terms of the Fielder vector element differences between node i and its neighboring nodes. Similar to the modularity method [8], the computational complexity for node-LFVC is $O((m+n)n)$.

A deep community is defined as a nonsingleton maximal connected subgraph (i.e., a connected component) in conjunction with the removed nodes that were previously adjacent to the subgraph before their removal. In other words, consider a nonsingleton maximal connected subgraph S with node set V_S when a set of nodes R is removed from G. The deep community is the set $V_S \cup \{i \in R : A_{ij} = 1 \text{ for some } j \in S\}$.

C. Monotonic submodular set function and greedy removals

Consider the problem of finding the optimal node removal set with cardinality q that maximizes loss in algebraic connectivity. The computational complexity of this batch removal problem is of combinatorial order $\binom{n}{q}$. Here we show that greedy removal, whose computation is only linear in n, is almost as good as the combinatorial algorithm in terms of achieving, within a multiplicative constant $(1-1/e)$, the same algebraic connectivity. Let

$$f(R) = \sum_{i \in R} \sum_{j \in N_i} (y_i - y_j)^2 - \frac{1}{2} \sum_{i \in R} \sum_{j \in R} A_{ij}(y_i - y_j)^2 \quad (9)$$

such that $\lambda_2(\widetilde{L}(R)) \leq \lambda_2(L) - f(R)$. Note that when $|R| = 1$, $f(R)$ degenerates to node-LFVC. $f(R)$ is nonnegative since

$$f(R) = \sum_{i \in R} \sum_{j \in N_i} (y_i - y_j)^2 - \frac{1}{2} \sum_{i \in R} \sum_{j \in V} A_{ij}(y_i - y_j)^2 \quad (10)$$
$$+ \frac{1}{2} \sum_{i \in R} \sum_{j \in V/R} A_{ij}(y_i - y_j)^2$$
$$= \frac{1}{2} \sum_{i \in R} \left(\sum_{j \in N_i} (y_i - y_j)^2 + \sum_{j \in V/R} A_{ij}(y_i - y_j)^2 \right) \geq 0.$$

The following theorem shows submodularity of $f(R)$. This implies that greedy node removals based on node-LFVC are almost as effective as the combinatorially complex batch algorithm. More specifically, it implies the inequalities in (15).

Theorem 1. *$f(R)$ is a monotonic submodular set function.*

Proof: Consider two removal sets $R_1 \subset R_2 \subset V$, since

$$f(R_2) - f(R_1)$$
$$= \sum_{i \in R_2/R_1} \sum_{j \in N_i} (y_i - y_j)^2 - \sum_{i \in R_1} \sum_{j \in R_2/R_1} A_{ij}(y_i - y_j)^2$$
$$- \frac{1}{2} \sum_{i \in R_2/R_1} \sum_{j \in R_2/R_1} A_{ij}(y_i - y_j)^2$$
$$\geq \sum_{i \in R_2/R_1} \sum_{j \in V} A_{ij}(y_i - y_j)^2 - \sum_{i \in R_1} \sum_{j \in R_2/R_1} A_{ij}(y_i - y_j)^2$$
$$- \sum_{i \in R_2/R_1} \sum_{j \in R_2/R_1} A_{ij}(y_i - y_j)^2$$
$$= \sum_{i \in R_2/R_1} \left(\sum_{j \in V} A_{ij}(y_i - y_j)^2 - \sum_{j \in R_2} A_{ij}(y_i - y_j)^2 \right)$$
$$= \sum_{i \in R_2/R_1} \sum_{j \in V/R_2} A_{ij}(y_i - y_j)^2 \geq 0, \quad (11)$$

$f(R)$ is a monotonic increasing set function (i.e., $f(R_2) \geq f(R_1)$ for all $R_1 \subset R_2 \subset V$). Furthermore, $f(R)$ is a submodular set function [28], [29] since for any node $v \in V, v \notin R_2$,

$$f(R_1 \cup v) - f(R_1) = \sum_{j \in N_v} (y_v - y_j)^2 - \sum_{j \in R_1} A_{vj}(y_v - y_j)^2$$
$$\geq \sum_{j \in N_v} (y_v - y_j)^2 - \sum_{j \in R_2} A_{vj}(y_v - y_j)^2$$
$$= f(R_2 \cup v) - f(R_2). \quad (12)$$

The set function $f(R)$ has diminishing gain property and therefore $f(R)$ is a submodular set function. ∎

When $|R| = q$, let R_{opt} be the optimal node removal set that maximizes $f(R)$ and R_k be the greedy node removal set at kth stage with $|R_k| = k$. By submodularity, there exists a $v \in R_{\text{opt}}/R_k$ such that

$$f(R_k \cup v) - f(R_k) \geq \frac{1}{q} \left(f(R_{\text{opt}}) - f(R_k) \right). \quad (13)$$

We have $f(R_{\text{opt}}) - f(R_{k+1}) \leq (1 - \frac{1}{q})(f(R_{\text{opt}}) - f(R_k))$ and

$$f(R_{\text{opt}}) - f(R_q) \leq (1 - \frac{1}{q})^q f(R_{\text{opt}}) \leq \frac{1}{e} f(R_{\text{opt}}). \quad (14)$$

The submodularity of the function f implies that after q greedy iterations the loss in performance is within a factor $1/e$ of optimal batch removal [28]. In other words, when removing R_q from G, the algebraic connectivity is guaranteed to decrease by at least $(1 - e^{-1})f(R_{\text{opt}})$ of its original value.

$$\lambda_2(\widetilde{L}(R_q)) \leq \lambda_2(L) - f(R_q)$$
$$\leq \lambda_2(L) - (1 - e^{-1})f(R_{\text{opt}}). \quad (15)$$

Consequently, identifying the top q nodes affecting algebraic connectivity can be regarded as a submodular set function maximization problem, and the greedy algorithm can be applied iteratively to find the node with the highest node-LFVC and successively remove it from the graph.

IV. PERFORMANCE EVALUATION

In this section, we use node-LFVC to perform community detection on several datasets collected from real-world social networks and compare the results with the modularity method [8], [23]. The community structures are revealed by iteratively removing the node with the highest centrality from the largest community. Due to Theorem 1 and (15) after q iterations, this greedy node removal strategy is almost as good as the non-greedy batch removal of the best q nodes.

A. Coauthorship among network scientists

Consider the coauthorship network in network science studied by Newman [23]. Nodes represent network scientists and edges represent the existence of coauthorship. The coauthorship network is fragile in the sense that it can be divided into 2 deep communities by 1 node removal and into 6 deep communities by 2 node removals. The first node with the highest node-LFVC is Yamir Moreno, who is a network scientist in Spain but has many collaborators outside Spain. The local (two-hop) coauthorship network of Yamir Moreno is shown in Fig. 1. The red square community mainly represents the network scientists in Spain and Europe, whereas the blue triangle community represents the rest of the network
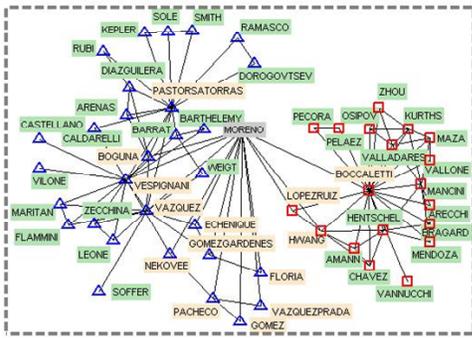
Fig. 1. Yamir Moreno's local 2-hop coauthorship network (from part of the network of coauthorship among network scientists [23] having $n = 379$ nodes and $m = 914$ edges). Moreno has 14 coauthors (marked by light orange color) and his coauthors have 35 coauthors. The modularity method [23] detects that Moreno is a member of only one large community (dashed box in gray). The proposed LFVC method detects Moreno as belonging to two separate communities indicated by red and blue nodes, respectively.
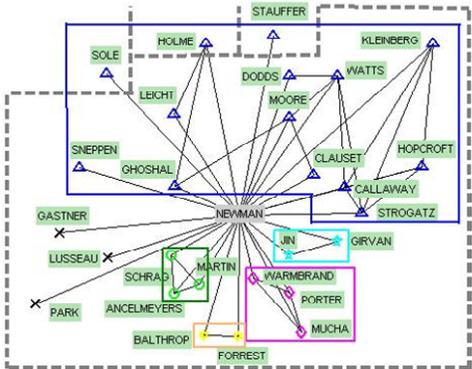

Fig. 2. Mark Newman's local 1-hop coauthor network in the network scientist coauthorship graph [23]. The proposed LFVC method detects Newman as belonging to 5 communities (marked by different vertex shapes and colors in solid boxes) and being associated with 3 outlier scholars (marked by black X label). For instance, Lusseau is detected as an outlier node since his research area is primarily in zoology. As shown in gray dashed box, the modularity method [23] detects 25 out of 28 scholars as being in a single community, and the top left 3 scholars as belonging to 3 different communities.

scientists. After removing Yamir Moreno from the network, the node with the highest node-LFVC in the remaining largest community is Mark Newman, who is associated with 5 community memberships and 3 outlier nodes as shown in Fig. 2. Each community can be related to certain relationship such as colleagues, students and research institutions. For instance, although Lusseau has coauthorship with Newman, his research area is primarily in zoology and therefore he has no interactions with other network scientists in the dataset since other network scientists are mainly specialists in physics. This also explains why Lusseau is detected as an outlier node in the community detection process in Fig. 2. Consequently, using node-LFVC for overlapping community detection and outlier nodes exclusion reveals new community structures that are not identified by the modularity method.

B. Friendship in hamsterster.com

Hamsterster.com is an online social website for hamster owners. We use the friendship dataset collected in [30] for community detection based on node-LFVC and other well-
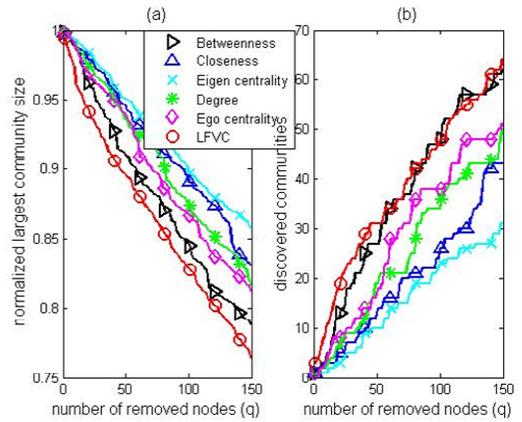

Fig. 3. Community detection on friendship in Hamsterster.com [30] with $n = 2000$ nodes and $m = 16097$ edges. (a) Normalized largest community size. (b) Discovered communities. In addition to minimizing largest community size, node-LFVC is capable of detecting more communities.

known centralities introduced in Sec. II-B. Two quantities, the normalized largest community size and the number of discovered communities with respect to node removals, are used to evaluate the performance of community detection when different centralities are applied. These two quantities reflect the effectiveness of graph partition and the capability of community detection.

As shown in Fig. 3 (a), the normalized largest community size decays linearly with respect to the number of node removals. Among all node centralities, node-LFVC has the steepest decaying rate, which suggests that it is capable of detecting significant communities embedded in the largest community more rapidly than the other centralities. Furthermore, using node-LFVC can discover more communities, as shown in Fig. 3 (b), especially during the first 50 node removals. The only node centrality that is comparable to node-LFVC is betweenness centrality. It it worth mentioning that computing betweenness centrality is a centralize approach that requires global shortest path information from every node pair in the network, whereas node-LFVC only requires local information from the standard Fiedler vector elements and it can be computed in a distributed manner, which is more scalable to large-scale network data. Notably, node-LFVC significantly outperforms other local node centrality such as node degree and ego centrality.

V. CONCLUSION

Based on the sensitivity of algebraic connectivity to node or edge removals, we propose a centrality called local Fiedler vector centrality (LFVC) for network connectivity analysis. We prove that the sensitivity on algebraic connectivity is a monotonic submodular set function. Therefore, despite its greedy nature, LFVC can be applied to identify the most vulnerable nodes or edges with bounded performance loss. Applying LFVC to community detection on real-world social network datasets, the results are self-contained in the sense that our method is capable of discovering significant communities and key members compared with conventional methods. In addition, LFVC is shown to have comparable performance to global centrality such as betweeness and it significantly outperforms other distributed centralities in terms of the largest community size and the number of discovered communities.

REFERENCES

[1] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[2] K.-C. Chen, M. Chiang, and H. Poor, "From technological networks to social networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 548–572, 2013.

[3] A. Bertrand and M. Moonen, "Seeing the bigger picture: How nodes can learn their place within a complex ad hoc network topology," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 71–82, 2013.

[4] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[5] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[6] M. Coscia, F. Giannotti, and D. Pedreschi, "A classification for community discovery methods in complex networks," *Statistical Analysis and Data Mining*, vol. 4, no. 5, 2011.

[7] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E*, vol. 80, p. 056117, Nov 2009.

[8] M. E. J. Newman, "Modularity and community structure in networks," *Proc. National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[9] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, June 2005.

[10] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706–1712, 2009.

[11] Z.-Y. Zhang, Y. Wang, and Y.-Y. Ahn, "Overlapping community detection in complex networks using symmetric binary matrix factorization," *Phys. Rev. E*, vol. 87, p. 062803, Jun 2013.

[12] H. Wen, E. A. Leicht, and R. M. D'Souza, "Improving community detection in networks by targeted node removal," *Phys. Rev. E*, vol. 83, p. 016114, Jan 2011.

[13] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.

[14] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 98, pp. 298–305, 1973.

[15] R. Merris, "Laplacian matrices of graphs: a survey," *Linear Algebra and its Applications*, vol. 197-198, pp. 143–176, 1994.

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[17] A. Bertrand and M. Moonen, "Distributed computation of the Fiedler vector with application to topology inference in ad hoc networks," *Signal Processing*, vol. 93, no. 5, pp. 1106–1117, 2013.

[18] A. Pothen, H. D. Simon, and K.-P. Liou, "Partitioning sparse matrices with eigenvectors of graphs," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 3, pp. 430–452, May 1990.

[19] D. A. Spielman and S.-H. Teng, "Spectral partitioning works: Planar graphs and finite element meshes," *Linear Algebra and its Applications*, vol. 421, no. 2-3, pp. 284–305, 2007.

[20] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[21] U. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[22] S. E. Schaeffer, "Graph clustering," *Computer Science Review*, vol. 1, no. 1, pp. 27–64, 2007.

[23] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, p. 036104, Sep 2006.

[24] L. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, pp. 35–41, 1977.

[25] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[26] M. Newman, *Networks: An Introduction*. Oxford University Press, Inc., 2010.

[27] M. Everett and S. P. Borgatti, "Ego network betweenness," *Social Networks*, vol. 27, no. 1, pp. 31–38, 2005.

[28] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions I," *Mathematical Programming*, vol. 14, pp. 265–294, 1978.

[29] S. Fujishige, *Submodular Functions and Optimization*. Annals of Discrete Math., North Holland, 1990.

[30] Hamsterster full network dataset - KONECT, Oct 2013.