A New Node Centrality Evaluation Model for Multi-community Weighted Social Networks

Jingru Li[†], Li Yu[†], Jia Zhao[‡] and Chaozhun Wen[†]
[†]School of Electronic Information and Communications,
Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[‡]Wuhan Zhongyuan Electronics Group Co., Ltd., Wuhan, Hubei 430074, China
Email: li1jingru@163.com, hustlyu@mail.hust.edu.cn, zhaojiakitty@163.com, U201313686@hust.edu.cn

***Abstract*—Identifying key nodes is an important research issue in social networks. Most of current social networks are weighted networks as well as consist of multiple communities. A suitable centrality measure for weighted social networks should be capable of finding most important nodes in each community. However, based on the existing centrality measure for weighted social networks, the most influential nodes are closely gathered in one community or distribute in a portion of all communities. In this paper, we propose a Tie Strength Matrix based Principal Component Centrality (TSM-based PCC), which extends PCC, a centrality measure for unweighted networks, to weighted social networks. Experiment results show that, based on TSM-based PCC influential nodes can be picked out accurately in real social network datasets. Furthermore, TSM-based PCC outperforms other centrality measures in identifying important nodes in each community. Hence the proposed TSM-based PCC is feasible and effective in weighted social networks.**

I. Introduction

Identifying important nodes is a key research issue in social networks. We always need to find the most active and influential nodes, which play crucial roles in many phenomena such as cascading, synchronizing, information spreading, social advertising and so on. In social network analysis, centrality is used to represent the importance of nodes in networks [1]. Many centrality measures have been proposed for unweighted social networks [2], [3], [4], [5], [6], which ignore weights of edges between nodes. However, as the development of network science, the research on weighted networks has more and more significance [7]. Newman stated that, the connections in many networks are not merely binary, either present or not, but have associated weights that record their strengths [8]. Ignoring weights will lose richer information, which can help us better understand characteristics of networks. Thus it is necessary to propose suitable centrality measures for weighted social networks.

In recent years, there has been a growing interest in studying centrality for weighted networks [10], [11], [12]. As proposed by Newman, by a simple mapping from a weighted network to an unweighted multigraph, standard centrality measures for unweighted networks can be extended to weighted ones [8]. Degree, closeness and betweenness centrality measures were extended to weighted networks respectively [13], [14], [15]. Newman also suggested the generalization of EVC to a weighted network for ranking search results in citation networks [16]. Different from the way of generalization, Xingqin Qi et al. proposed directly a new centrality measure, Laplacian centrality, for weighted networks [17]. Comparing with the standard centrality measures, Laplacian centrality is an intermediate measuring between global and local characterization of the importance of a node, and is reflected by the degree to which the Laplacian energy of the network drops when the node is deleted.

In general, social networks consist of multiple communities. Given this, most influential nodes are probably distributed in different communities. Thus, identifying influential nodes in multi-community weighted social networks is a significant research issue. However, the current centrality measures for weighted networks cannot solve this problem. Classical centrality measures (i.e. degree, closeness and betweenness) for weighted networks do not point out clearly the relation of centrality and community structure. Based on EVC, most influential nodes gather in a community of the whole network, which does not fit for the actual situation. As an intermediate between global and local structural characterization, Laplacian centrality can not solve the above problem either, since global information between communities is required [17].

Based on leading P eigenvectors of adjacent matrix, the foremost nodes picked out by PCC are belong to different communities of unweighted networks [5]. In this paper we improve this promising centrality measurement to recognize influential nodes in multi-community weighted social networks. And we make the following key contributions: (1) We firstly propose Tie Strength Matrix (TSM) based PCC, which generalizes PCC for unweighted networks to weighted social networks; (2) Experiment results show that, based on TSM-based PCC the most influential nodes can be picked out accurately in real social network datasets; (3) TSM-based PCC outperforms other centrality measures in identifying important nodes in each community.

The rest of this paper proceeds as follows. In Section II, the derivation on the extension of PCC from unweighted to weighted networks is presented. TSM-based centrality for weighted social networks is proposed in Section III. Then the experiment results and analysis are given in Section IV. Finally, we conclude the paper in Section V.

II. Extension of PCC from unweighted to weighted networks

A. PCC for unweighted networks

Let $\mathbf{A} = (a_{ij})$ denote the adjacency matrix of a unweighted graph $G(V, E)$ consisting of the set of nodes $V = \{v_1, v_2, v_3, \cdots, v_N\}$ of size N and the set of undirected edges E, and

$$a_{ij} = a_{ji} = \begin{cases} 1, \text{when a link exists betweent } v_i \text{ and } v_j \\ 0, \text{otherwise} \end{cases} .$$

Let N eigenvalues of the adjacency matrix $\mathbf{A}$ are ranked in descending order of magnitude $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|$, then the corresponding eigenvectors are $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$. Then PCC of a node is defined as its Euclidean distance from the origin in the P-dimensional eigenspace, whose basis vectors are $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_P$ of the adjacency matrix $\mathbf{A}$ [5], [18].

Let $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2\, \cdots\, \mathbf{x}_N]$, and $\Lambda = [\lambda_1 \lambda_2 \cdots \lambda_N]^T$, then $\mathbf{X}_{N\times P}$ denotes the submatrix of $\mathbf{X}$ consisting of the first P columns, and $\Lambda_{P\times 1}$ denotes the vector consisting of the first P elements of Λ. Then PCC [5] for unweighted networks can be expressed as

$$\mathbf{C}_P = \sqrt{((\mathbf{A}\mathbf{X}_{N\times P}) \odot (\mathbf{A}\mathbf{X}_{N\times P}))\, 1_{P\times 1}} \tag{1}$$

or

$$\mathbf{C}_P = \sqrt{(\mathbf{X}_{N\times P} \odot \mathbf{X}_{N\times P})\, (\Lambda_{P\times 1} \odot \Lambda_{P\times 1})} \tag{2}$$

where the '$\odot$' operator is the Hadamard (or entrywise product) operator and $1_{P\times 1}$ is a vector of 1s of length P.

B. Extension of PCC from unweighted to weighted networks

1) Derivation of the closed-form expression of PCC: This subsection is for explaining the extension of PCC from unweighted to weighted networks. Suppose node v_i is represented by the vector $(a_{i1}\ a_{i2}\ \cdots\ a_{in})^T$, and its coordinate in the P-dimensional eigenspace is $({x'}_{i1}\ {x'}_{i2}\ \cdots\ {x'}_{iP})^T$. Then

$${x'}_{ij} = \mathbf{x}_j^T (a_{i1}\ a_{i2}\ \cdots\ a_{in})^T = a_{i1}x_{j1} + a_{i2}x_{j2} + \cdots + a_{in}x_{jn}, \tag{3}$$

where $j = 1, 2, \cdots, P$. Suppose PCC score of node v_i is c_i, and the i-th row vector of $\mathbf{A}$ is $\mathbf{A}_i$. Then according to the definition of PCC,

$$\begin{aligned} c_i &= \sqrt{\sum_{j=1}^{P} {{x'}_{ij}}^2} = \sqrt{\sum_{j=1}^{P} (\mathbf{A}_i \mathbf{x}_j)^2} \\ &= \sqrt{\left((\mathbf{A}_i\mathbf{x}_1)^2\ (\mathbf{A}_i\mathbf{x}_2)^2\ \cdots\ (\mathbf{A}_i\mathbf{x}_p)^2\right) 1_{P\times 1}} \\ &= \sqrt{((\mathbf{A}_i\mathbf{X}_{N\times P}) \odot (\mathbf{A}_i\mathbf{X}_{N\times P}))\, 1_{P\times 1}} \end{aligned} \tag{4}$$

Equation (4) can be rewritten as equation (1). And equation (2) can be deduced from equation (1).

2) Extension of PCC (integer weight case): With nonnegative integer weights, weighted networks and the corresponding multigraphs have the same "weighted adjacent matrix" whose entries are weights on edges. Newman proposed that, a weighted network can be mapped to an unweighted multigraph, that is, the edge with weight n can be replaced by n parallel edges with weight 1 each [8].

Let $G(V, E_W)$ be a weighted network with the node set $V = \{v_1, v_2, v_3, \cdots, v_N\}$ and weighted edge set E_W where each edge $e = (v_i, v_j)$ has a corresponding weight w_{ij}. Let $\mathbf{W} = (w_{ij})$ be the weight matrix.

Assume N eigenvalues of $\mathbf{W}$ are ranked in descending order of magnitude $|\lambda_{w,1}| \geq |\lambda_{w,2}| \geq \cdots \geq |\lambda_{w,N}|$, and the corresponding eigenvectors are $\mathbf{x}_{w,1}, \mathbf{x}_{w,2}, \cdots, \mathbf{x}_{w,N}$. Let $\mathbf{X}_{w,N\times P} = [\mathbf{x}_{w,1}\, \mathbf{x}_{w,2}\, \cdots\, \mathbf{x}_{w,P}]$, and $\Lambda_{w,P\times 1} = [\lambda_{w,1}\lambda_{w,2}\cdots\lambda_{w,P}]^T$, where $1 \leq P \leq N$.

Based on the method given by Newman [8], we can calculate the Euclidean distance of nodes from the origin in the P-dimensional eigenspace whose basis vectors are $\mathbf{x}_{w,1}, \mathbf{x}_{w,2}, \cdots, \mathbf{x}_{w,P}$, by repeating steps of equations (3)-(4). Then we can get the similar form of expression of PCC for weighted networks

$$\mathbf{C}_{w,P} = \sqrt{((\mathbf{W}\mathbf{X}_{w,N\times P}) \odot (\mathbf{W}\mathbf{X}_{w,N\times P}))\, 1_{P\times 1}} \tag{5}$$

or

$$\mathbf{C}_{w,P} = \sqrt{(\mathbf{X}_{w,N\times P} \odot \mathbf{X}_{w,N\times P})\, (\Lambda_{w,P\times 1} \odot \Lambda_{w,P\times 1})}. \tag{6}$$

3) Extension of PCC (Non-integer weight case): We hope the equations (5)-(6) still can be suitable for non-integer weights. Obviously, $\exists$ positive integer k, s.t. all entries of matrix $\mathbf{W}' = k\mathbf{W}$ are non-negative integers. When all weights are multiplied by the same positive coefficient k, characteristics of the network may have proportional change. Since centrality focuses on the relative size of nodes' importance, it's probable that the ranking results based on centrality measure keep still when substituting $\mathbf{W}'$ instead of the original $\mathbf{W}$. The proof process is as follows.

Theorem 1. *Assume λ_0 is an eigenvalue of the square matrix $\mathbf{A}$ corresponding to the eigenvector $\mathbf{x}$, i.e. $\lambda_0\mathbf{x} = \mathbf{A}\mathbf{x}$. Then for any real number k, $k\lambda_0$ is an eigenvalue of the matrix $k\mathbf{A}$ corresponding to the eigenvector $\mathbf{x}$, i.e. $(k\lambda_0)\mathbf{x} = (k\mathbf{A})\mathbf{x}$.*

Theorem 1 can be easily proved based on theoretical basis of matrix theory. Suppose that N eigenvalues of $\mathbf{W}'$ are ranked in descending order of magnitude $\left|\lambda'_{w,1}\right| \geq \left|\lambda'_{w,2}\right| \geq \cdots \geq \left|\lambda'_{w,N}\right|$, and the corresponding eigenvectors are ${\mathbf{x}'}_{w,1}, {\mathbf{x}'}_{w,2}, \cdots, {\mathbf{x}'}_{w,N}$. By theorem 1, if one eigenvalue/eigenvector pair of $\mathbf{W}$ is $\lambda_{w,i}/\mathbf{x}_{w,i}$, then $\mathbf{W}'$ must have one eigenvalue/eigenvector pair as $k\lambda_{w,i}/\mathbf{x}_{w,i}$. Obviously, there exist $\lambda'_{w,i} = k\lambda_{w,i}$, and ${\mathbf{x}'}_{w,i} = \mathbf{x}_{w,i}$, where $i = 1, 2, \cdots, N$.

Based on integer matrix $\mathbf{W}'$, compute the corresponding PCC using equation (6) (or (5)) as follows

$$\begin{aligned}\mathbf{C}'_{w,P} &= \sqrt{\left(\mathbf{X}_{w,N\times P}\odot\mathbf{X}_{w,N\times P}\right)\left(\left(k\Lambda_{w,P\times 1}\right)\odot\left(k\Lambda_{w,P\times 1}\right)\right)}\\ &= k\sqrt{\left(\mathbf{X}_{w,N\times P}\odot\mathbf{X}_{w,N\times P}\right)\left(\Lambda_{w,P\times 1}\odot\Lambda_{w,P\times 1}\right)}\\ &= k\mathbf{C}_{w,P}\end{aligned} \tag{7}$$

Since k is positive, the ranking results of nodes' importance based on $\mathbf{C}'_{w,P}$ and $\mathbf{C}_{w,P}$ are the same. Due to $\mathbf{W}' = k\mathbf{W}$, it's reasonable that the weighted PCC vector for $\mathbf{W}$ is $\mathbf{C}_{w,P}$, which is obtained by equation (5) or (6) directly from the matrix $\mathbf{W}$ containing fractional weights. Thus, the expressions for weighted PCC are equations (5)-(6).

4) Illustration on the reasonableness of weighted PCC: In the above derivation process, as coefficients of intermediate quantities, w_{ij} represents the strength of ties. For same intermediate quantities, higher w_{ij} influences the final computing results of nodes' centrality more than lower w_{ij}, which means that strong ties influence nodes' centrality more than weak ties. Thus it's reasonable to deduce the centrality using w_{ij} instead of a_{ij}.

Another reasonableness that PCC can be used in weighted networks is that, PCC was inspired by KLT [5], which is a signal transforming algorithm operated on covariance matrices, where an entry with a large magnitude is representative of a strong relationship between two random variables. This is consistent with our research on social networks, whose adjacent matrix has a large entry for a strong tie. And PCC may be more suitable for weighted networks than unweighted ones, for tie strength matrix is more alike with covariance matrix than unweighted adjacent matrix.

III. Tie Strength Matrix (TSM) based centrality for weighted social networks

How to quantify the weights of edges is the key question of weighted networks without explicit weights on edges. Some research work has been done on the correlation relationship between nodes [7], [9]. In social networks, tie strength is a perfect tool to weigh the edges or social ties. Tie strength was first introduced in 1973 by Granovetter, who suggested that the strength of a relationship relied on four components: the frequency of contact, the length or history of the relationship, contact duration, and the number of transactions [19]. Scholars extended upon these measures and proposed seven components in all to measure tie strength: Frequency, Intimacy/Closeness, Longevity, Reciprocity, Recency, Multiple social context and Trust [20]. Researchers can select different measures according to the network conditions.

In this section, we quantify the tie strength based on three factors: frequency, closeness and recency. The more frequent the interaction on a tie is, the stronger the tie is; the closer the interaction on a tie is, the stronger the tie is. To describe the recency of tie strength, we introduce a forgetting factor, which is used to quantify the fact that the influence of the old contacts on current tie strength will decrease as time goes on. Apparently, the weight of an older contact is less than that of a recent contact, reflecting the time-varying nature of ties, which has not been taken into account by traditional centrality measures [20].

These three factors are presented in detail as follows, where frequency factor and closeness factor are evaluated according to an evidence-based strategy, and the trust in a piece of evidence is measured as a ratio of supporting and contradicting evidences [20].

(1) Frequency factor

Frequency factor is evaluated based on the frequency of encounters between nodes, and is given by

$$FF_i\left(j\right) = \frac{f\left(j\right)}{F\left(i\right) - f\left(j\right)}, \tag{8}$$

where $FF_i\left(j\right)$ is the frequency factor of node i to node j, $f\left(j\right)$ is the number node i encountered node j and $F\left(i\right)$ is the total number node i encountered other nodes.

(2) Closeness factor

Closeness factor is related with the time a node has connected to another node, and is expressed as

$$CF_i\left(j\right) = \frac{d\left(j\right)}{D\left(i\right) - d\left(j\right)}, \tag{9}$$

where $CF_i\left(j\right)$ is the closeness factor of node i to node j, $d\left(j\right)$ is the total time node i has connected to node j , and $D\left(i\right)$ is the total time node i has been connected with all other encountered nodes.

(3) Forgetting factor

Forgetting factor is defined for recency, which means that recent interactions between nodes are more important than older ones. According to the decreasing influence of old interactions on current tie strength, K interactions at a previous time t_1 is equivalent to $K\beta^{\Delta t}$ interactions at a later time t_2, where $\Delta t = t_2 - t_1$, and $\beta\left(0 < \beta \le 1\right)$ is referred to as the *forgetting factor*. Assume there are Δ_K additional interactions between time t_1 and t_2. Then, at time t_2, K is updated to $K\beta^{\Delta t} + \Delta_K$. Analogously, the updated results of $f\left(j\right), F\left(i\right), d\left(j\right)$ and $D\left(i\right)$ after time interval Δt are shown as follows.

$$f\left(j\right) \to f\left(j\right)\beta^{\Delta t} + \Delta_{f(j)}, \tag{10}$$

$$F\left(i\right) \to F\left(i\right)\beta^{\Delta t} + \Delta_{F(i)}, \tag{11}$$

$$d\left(j\right) \to d\left(j\right)\beta^{\Delta t} + \Delta_{d(j)}, \tag{12}$$

$$D\left(i\right) \to D\left(i\right)\beta^{\Delta t} + \Delta_{D(i)}. \tag{13}$$

The overall tie strength is defined as

$$\text{TieStrength}_i\left(j\right) = FF_i\left(j\right) + CF_i\left(j\right). \tag{14}$$

Substituting results (10)-(13) into equation (8) and (9), then the updated result of $\text{TieStrength}_i(j)$ is given as

$$
\begin{aligned}
&\text{TieStrength}_i^*(j)\\
=&\frac{f(j)\beta^{\Delta t}+\Delta_{f(j)}}{\left[F(i)\beta^{\Delta t}+\Delta_{F(i)}\right]-\left[f(j)\beta^{\Delta t}+\Delta_{f(j)}\right]}\\
&+\frac{d(j)\beta^{\Delta t}+\Delta_{d(j)}}{\left[D(i)\beta^{\Delta t}+\Delta_{D(i)}\right]-\left[d(j)\beta^{\Delta t}+\Delta_{d(j)}\right]}.
\end{aligned}
\tag{15}
$$

Define the following matrix

$$\mathbf{T}=(T_{ij})=\left(\text{TieStrength}_i^*(j)\right), \tag{16}$$

as *tie strength matrix* (TSM for short). $\mathbf{T}$ is equivalent to the quantified weighted matrix $\mathbf{W}$ for weighted social networks, and will replace $\mathbf{W}$ in equation (5)-(6) when used to calculate PCC for weighted social networks. Then the real-time TSM-based PCC for weighted social networks are as follows

$$\mathbf{C}_{t,P}=\sqrt{\left(\left(\mathbf{T}\mathbf{X}_{t,N\times P}\right)\odot\left(\mathbf{T}\mathbf{X}_{t,N\times P}\right)\right)1_{P\times 1}} \tag{17}$$

or

$$\mathbf{C}_{t,P}=\sqrt{\left(\mathbf{X}_{t,N\times P}\odot\mathbf{X}_{t,N\times P}\right)\left(\Lambda_{t,P\times 1}\odot\Lambda_{t,P\times 1}\right)} \tag{18}$$

where $\mathbf{X}_{t,N\times P}$ and $\Lambda_{t,P\times 1}$ are defined similarly with weight matrix case, and $\mathbf{C}_{t,P}$ is the PCC vector based on TSM $\mathbf{T}$. When applied in the experiments, all tie strength will be normalized into the interval [0,1].

IV. Experiments and analysis

In this section, the accuracy and effectiveness of PCC on real weighted social networks are verified. The experiments are implemented on several classical weighted social networks. From the experiment results and analysis, the central nodes belong to multiple communities can be successfully detected.

A. Verification on accuracy and effectiveness of PCC on real weighted social networks

1) Freeman's EIES network: Freeman's EIES dataset includes three networks recording interactive relationship among scholars studying social network analysis (SNA) [17]. we choose the third weighted network where the weight of each edge is the number of messages sent among 32 researchers on an electronic communication tool.

We calculate the centrality values of the 32 scholars based on the five existing centrality measures for weighted networks (i.e. degree, betweenness, closeness, Laplacian, EVC) and weighted PCC simultaneously. The results are shown in Table I, where the leftmost column are scientists' labeled numbers from 1 to 32, and they are ranked in descending order of 32 scientists' PCC scores. In this order results by six centrality measures are in the remaining columns, where the top four scores for each centrality are highlighted in bold.

From Table I, based on weighted PCC ($P_{app}=3$), the top two scholars (labeled as 1 and 29) are the same as other weighted centrality methods. And the top four scientists detected by weighted PCC are consistent with four out of five other methods. These results verify the utility and good accuracy of weighted PCC for weighted social networks from one aspect.

index	Degree	Betwe-enness	Closeness	Laplacian	EVC	PCC
1	**3449**	**864**	**62.8947**	**0.6544**	**770.1061**	**951.1978**
29	**2221**	**384**	**59.545**	**0.3772**	**572.6664**	**668.9834**
8	**1721**	**64**	**55.0109**	**0.2670**	**470.7864**	**562.214**
2	**1500**	0	**56.2145**	**0.2517**	**461.6073**	**538.4736**
32	1157	**5**	51.6036	0.1801	373.1667	425.0129
31	1324	0	49.532	0.1725	332.6431	336.5168
11	869	0	47.7976	0.1183	258.8136	286.9808
24	783	0	47.4057	0.1017	225.4388	227.1564
30	423	0	44.4079	0.0563	141.8275	204.1509
27	470	0	39.238	0.0548	135.2292	136.7744
9	392	0	37.0054	0.0454	117.1301	131.0102
15	276	0	42.1726	0.0462	115.7569	125.0094
10	446	0	37.9357	0.0457	112.8001	116.475
16	345	0	35.619	0.0443	113.9675	114.5087
17	318	0	35.619	0.0372	95.0837	103.7415
4	377	1	33.5708	0.0386	97.3441	98.9814
22	284	0	39.2548	0.0365	95.4588	98.8085
6	274	0	34.5777	0.0346	89.7482	93.9882
26	171	0	37.8311	0.0282	72.194	89.6212
18	246	0	34.7922	0.0306	77.9836	89.2844
12	206	0	36.5929	0.0271	73.3645	79.4205
14	214	0	34.3294	0.0274	71.5568	72.085
7	124	0	27.2235	0.0167	44.8164	49.6149
19	175	0	22.8723	0.0169	43.6222	49.2356
25	149	0	24.6484	0.0153	40.5792	46.7062
13	148	0	23.8952	0.0161	43.6112	44.1341
5	189	0	19.2466	0.0155	41.7317	41.9812
21	112	0	25.9092	0.0143	39.4467	41.1839
23	125	0	21.6725	0.0151	40.501	40.7236
20	101	0	21.6725	0.013	34.8506	36.0787
3	101	0	20.2504	0.0111	29.9272	32.0587
28	78	0	18.7309	0.0097	27.1466	28.7889

TABLE I. Centrality scores of 32 scholars in Freeman's EIES network based on various centrality measures.

2) Zachary's karate club network: Unweighted network is the special case of weighted network. In this part we use the unweighted "karate club network" of Zachary [17] to test the performance of weighted PCC. Zachary observed 34 members of a karate club over two years, and the club members split into two groups because of the disagreement between the administrator of the club and the clubs instructor. In the karate club network, the administrator of the club and the clubs instructor are label by 34 and 1 respectively.

The centrality values of the 34 members based on six centrality measures are shown in Table II, where the 34 members' labeled numbers in the leftmost column are ranked in descending order of 34 PCC scores, and results by other centrality measures are also in this order. Moreover, the top two scores for each centrality are highlighted in bold. We can find that, PCC along with degree, Laplacian and EVC successfully detect two most important leading members, but betweenness and closeness fail. What's more, $P_{app} = 2$ for PCC of the karate club network, illustrating that PCC can accurately identify the number of major communities of social networks.

index	Degree	Betwe-enness	Closeness	Laplacian	EVC	PCC
34	**17**	147	0.5667	**0.3187**	**2.5111**	**3.1157**
1	**16**	**547**	**0.5862**	**0.2997**	**2.3909**	**3.0698**
33	12	276	0.5313	0.2032	2.0758	2.6303
2	9	32	0.5	0.1418	1.7888	2.2341
3	10	**314**	**0.5763**	0.1769	2.1333	2.231
4	6	0	0.4789	0.098	1.4203	1.898
14	5	8	0.5313	0.1067	1.5232	1.6638
9	5	47	0.5313	0.1082	1.5294	1.5538
8	4	0	0.4533	0.0746	1.1498	1.5503
24	5	34	0.4048	0.0804	1.0097	1.4786
32	6	78	0.5574	0.1096	1.2848	1.3819
30	4	9	0.3953	0.0673	0.9077	1.3715
31	4	0	0.4722	0.0775	1.1754	1.2701
20	3	0	0.5152	0.0702	0.9948	1.0349
28	4	39	0.4722	0.0658	0.8977	1.0339
23	2	0	0.382	0.0468	0.682	0.9738
21	2	0	0.382	0.0468	0.682	0.9738
19	2	0	0.382	0.0468	0.682	0.9738
16	2	0	0.382	0.0468	0.682	0.9738
15	2	0	0.382	0.0468	0.682	0.9738
29	3	0	0.4658	0.057	0.8816	0.9455
22	2	0	0.3864	0.0409	0.6215	0.9035
18	2	0	0.3864	0.0409	0.6215	0.9035
7	4	2	0.3953	0.0512	0.5346	0.9007
6	4	61	0.3953	0.0512	0.5346	0.9007
13	2	0	0.382	0.0365	0.5667	0.8547
11	3	0	0.3908	0.0424	0.5109	0.8397
5	3	0	0.3908	0.0424	0.5109	0.8397
27	2	0	0.3736	0.0351	0.5083	0.7691
10	2	0	0.4474	0.0439	0.6906	0.7309
26	3	18	0.3864	0.0292	0.3982	0.5475
12	1	0	0.3778	0.0249	0.3555	0.5254
25	3	2	0.3864	0.0278	0.3837	0.4755
17	2	0	0.2931	0.0161	0.159	0.3319

TABLE II. Centrality scores of 34 members in Zachary's karate club network based on various centrality measures.

B. Effectiveness in identifying influential nodes in each community in weighted social networks

1) Experiment datasets: This part analyzes the centrality evaluation algorithm by using MIT Reality Mining dataset [21], which was obtained from Reality Mining project at MIT Media Lab. The experiment in this project used 104 Nokia cellphones, generating the data including calllogs, celltowerIDs, Bluetooth data, phone status, and application usage. Bluetooth data is used in order to identify direct contacts between nodes, to analyze the effectiveness of TSM-based PCC in weighted social network. 94 pieces of valid user data are picked out to implement the experiment. This network contains two main communities: (1) group 1 consists of 26 first-year business school students, (2) group 2 consists of 68 individuals working together in the same Media Lab building. The friend graph and community structure partition is shown in Fig. 1, where nodes with orange color and blue color represent group 1 and group 2 respectively.

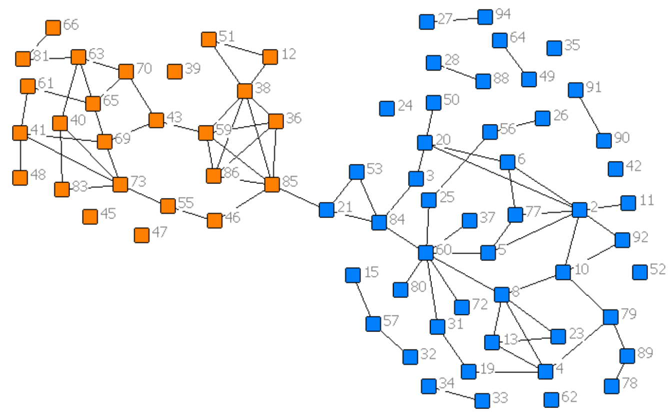
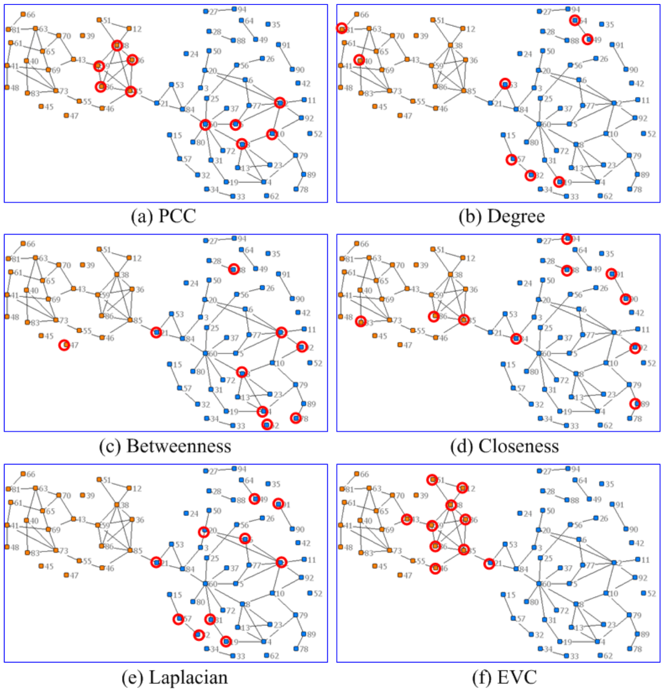


Fig. 1. User friendship graph with community structure partition

2) Experiment results and analysis: It is expected that a suitable centrality measure can identify the most influential nodes in both communities. Based on six centrality measures for weighted social networks, we compute the centrality values for all nodes and sort them in descending order. Then the top-10 nodes are regarded as the most influential nodes in the network. Red circles are used to label these influential nodes. The experimental results are shown in Fig. 2, from which we can observe the distribution of most important nodes in different communities of the whole network.

It can be seen that, the top-10 influential nodes identified by TSM-based PCC distribute among two different communities. And in each community, all the influential nods occupy important positions. In contrast, the other centrality measures cannot achieve such desired effect. Specifically,

1) The important nodes recognized by degree centrality measures are dispersed over the whole network. However, there is no obvious relationship between these nodes and the network community structure. This is because that degree centrality only consider local structural information, but does not take into account the network's global structure, including multi-community property.
2) Betweenness and closeness centralities for weighted networks are based on shortest path algorithm. In this weighted MIT social network, the weights on edges are computed based on Bluetooth interactions between nodes. Some weak ties are taken into account to evaluate the centrality. As a result, based on betweenness



(a) PCC (b) Degree

(c) Betweenness (d) Closeness

(e) Laplacian (f) EVC

Fig. 2. Influential nodes distribution based on six centrality measures for weighted social networks

and closeness centralities, the relation of centrality and community structure cannot be shown clearly either.

3) By Laplacian and EVC centrality, most influential nodes gather in a community of the network. EVC inherits its weakness in unweighted case, and tends to find the centers of whole network instead of each community. Laplacian centrality as an intermediate between global and local structural characterization in weighted networks, has advantages when nodes to community (not the whole network) relationships need to be understood [17]. However, Laplacian centrality can not be used to solve the problem of finding most important nodes in each community in the whole network either, for global information between communities is needed to solve this problem.

In conclusion, TSM-based PCC outperforms other centrality measures in identifying the most important nodes in each community. Hence the proposed TSM-based PCC is effective in weighted social networks.

V. Conclusion

Identifying the most influential nodes is crucial in weighted social networks. Among the existing centrality measures for weighted social networks, seldom works have been focused on identifying the most influential nodes in each community. To solve this problem, we generalize PCC, which is a centrality measure for unweighted networks, to weighted social networks, and propose TSM-based PCC. Experiment results show that, by our model the most influential nodes can be detected accurately in real social network datasets. Moreover, TSM-based PCC outperforms other centrality measures in recognizing the most important nodes in each community. Hence the proposed TSM-based PCC is feasible and effective in weighted social networks.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NSFC) (No. 61231010), and National High Technology Research and Development Program (No. 2015AA015901).

References

[1] S. Wasserman, F. Faust, "Social Network Analysis: Methods and Applications," in Cambridge, U.K. (Cambridge Univ. Press, 1994)

[2] L. Tang, H. Liu, "Community Detection and Mining in Social Media," Synthesis Lectures on Data Mining and Knowledge Discovery, 2010, **2**, pp. 1-137

[3] M. Kitsak, L. Gallos, S. Havlin, *et al.*, "Identification of influential spreaders in complex networks," Nature Physics, 2010, **6**, (11), pp 888-893

[4] E. Estrada, J. Rodriguez-Velazquez, "Subgraph centrality in complex networks," Physical Review E, 2005, **71**, (5), pp. 056103

[5] M. Ilyas, H. Radha, "A KLT-inspired node centrality for identifying influential neighborhoods in graphs." Conference on Information Sciences and Systems (CISS), Princeton, NJ, Mar. 2010, pp. 1–7

[6] D. Chen, and L. Lü, M. Shang, *et al.*, "Identifying influential nodes in complex networks," Physica A: Statistical mechanics and its applications, 2012, **391**, (4), pp 1777–1787

[7] X. Deng, Y. Wu, M. Dong, C. Li and Y. Pan, "A Weighted Network Model Based on the Correlation Degree between Nodes," 2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN), Shenzhen, 2015, pp. 216-221.

[8] M. Newman, "Analysis of weighted networks," Physical Review E, 2007, **70**, (5), pp. 56131

[9] C. Zhang, K. Ota and M. Dong, "Cooperative Positioning Optimization in Mobile Social Networks," 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, 2016, pp. 1-5.

[10] S. Segarra and A. Ribeiro, "Stability and Continuity of Centrality Measures in Weighted Graphs," IEEE Transactions on Signal Processing, 2016, **64**, (3), pp 543-555

[11] Z. Sun, B. Wang, J. Sheng, Y. Hu, Y. Wang, J. Shao, "Identifying Influential Nodes in Complex Networks based on Weighted Formal Concept Analysis," IEEE Access, 2017, **PP**, (99), pp 1-1

[12] X. Tang, J. Wang, J. Zhong and Y. Pan, "Predicting Essential Proteins Based on Weighted Degree Centrality," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2014, **11**, (2), pp 407-418

[13] A. Barrat, M. Barthelemy, R. Pastor-Satorras, *et al.*, "The architecture of complex weighted networks," Proceedings of the National Academy of Sciences, 2004, **101**, (11), pp 3747-3752

[14] M. Newman, "Scientific collaboration networks. Part II: shortest paths, weighted networks, and centrality," Physical Review E, 2001, **64**, pp. 016132

[15] U. Brandes, "A faster algorithm for betweenness centrality," Journal of Mathematical Sociology, 2001, **25**, pp. 163-177

[16] S. Redner, "How popular is your paper? An empirical study of the citation distribution," The European Physical Journal B, 1998, **4**, (2), pp. 131-134

[17] X. Qi, E. Fuller, Q. Wu, *et al.*, "Laplacian centrality: A new centrality measure for weighted networks," Information Sciences, 2012, **194**, pp 240–253

[18] U. Muhammad, M. Shafiq, X. Alex, *et al.*, "A Distributed and Privacy Preserving Algorithm for Identifying Information Hubs in Social Networks." INFOCOM, Apr. 2011, pp. 561-565

[19] M. Granovetter, "The Strength of Weak Ties," The Am. J. Sociology, May., 1973, **78**, (6), pp. 1360-1380

[20] E.Dal y, M. Haahr, "Social Network analysis for information flow in disconnected delay-tolerant manets," IEEE Transactions on Mobile Computing, 2009, **8**, (5), pp. 606-621

[21] N. Eagle, A. Pentland, D. Lazer, "Inferring Social Network Structure using Mobile Phone Data." Proceedings of the National Academy of Sciences, Sep., 2009, pp. 15274-15278