Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Community-based immunization in opportunistic social networks

Peiyan Yuan\textsuperscript{a,}\textsuperscript{*}, Shaojie Tang\textsuperscript{b}

\textsuperscript{a} School of Computer and Information Engineering, Henan Normal University, Xinxiang, Henan, China
\textsuperscript{b} Department of Information Systems, University of Texas at Dallas, Dallas, TX, USA

\textbf{H I G H L I G H T S}

\begin{itemize}
  \item The community structure helps to delay the outbreak of epidemic.
  \item Nodes with high local importance can remarkably suppress the epidemic.
  \item It is not “the more central, the better” for the implementation of control strategy.
\end{itemize}

\textbf{A R T I C L E  I N F O}

\textbf{Article history:}
Received 7 May 2014
Received in revised form 20 October 2014
Available online 4 November 2014

\textbf{Keywords:}
Opportunistic social networks
Epidemic spreading
Community-based immunization
Local importance

\textbf{A B S T R A C T}

Immunizing important nodes has been shown to be an effective solution to suppress the epidemic spreading. Most studies focus on the globally important nodes in a network, but neglect the locally important nodes in different communities. We claim that given the temporal community feature of opportunistic social networks (OSN), this strategy has a biased understanding of the epidemic dynamics, leading us to conjecture that it is not “the more central, the better” for the implementation of control strategy. In this paper, we track the evolution of community structure and study the effect of community-based immunization strategy on epidemic spreading. We first break the OSN traces down into different communities, and find that the community structure helps to delay the outbreak of epidemic. We then evaluate the local importance of nodes in communities, and show that immunizing nodes with high local importance can remarkably suppress the epidemic. More interestingly, we find that high local importance but non-central nodes play a big role in epidemic spreading process, removing them improves the immunization efficiency by 25\% to 150\% at different scenarios.

\textcopyright 2014 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid pervasive of a new generation of smart devices, it is possible and necessary to disseminate content in networks by exploiting the human mobility and intermittent device-to-device contacts. Networks with such intermittent device-to-device contacts are generally called delay-tolerant [1] or opportunistic [2]. Until recently, a variety of opportunistic social networks (OSN) have been studied, such as pocket switched networks [3], publish/subscribe systems [4,5], and human contact networks [6]. Many interesting phenomena have been observed, including the heavy-tailed distribution of contact times and node degree [7,8], the small world phenomenon [9], the dynamics of epidemic spreading [10,11], the high clustering of aggregated social contact statistics [12], etc.
The social contact feature makes nodes in OSN vulnerable to infection. Therefore, when an infectious disease appears in a population (e.g., the severe acute respiratory syndrome (SARS) and the H7N9 virus), designing effective immunization strategies has become very important. To achieve this goal, several immunization strategies have been recently developed, ranging from ring immunization [13] to targeted immunization [14–18]. The targeted strategy has shown to be more effective than the ring strategy to delay the outbreak of epidemic. The basic idea of targeted immunization is that it first ranks importance of nodes and then removes them, from highest importance to lowest, to observe their impact on epidemic spreading speed. The importance of nodes is generally measured by node’s degree, closeness or betweenness centrality in the network [19]. To further improve the immunization efficiency, the authors of Ref. [20] suggested that node importance should be recalculated after every step of node removal.

Previous work mainly concentrates on the global measures of nodes in a network, while ignoring the local importance of nodes in different groups. A number of recent studies indicate that opportunistic networks have a high clustered property [10,21] and show a temporal community structure [22]. Networks with properties at the level of community are quite different from their properties at the entire network level [23]. Studies thus paying more attention to the whole network topology and neglecting temporal community structure may miss many interesting features. For example, in real world it is found that people have different average number of contacts in different social cliques. The same person in one clique may be sociable, having many contacts with others, while in another clique he/she may be more taciturn. Such contact behaviors have also been seen in opportunistic social networks, where people are more frequently to contact their family or friends, while they meet accidentally with strangers [24]. If one tried to characterize such a network by statistics of the mean number of contacts a person has, one would be missing the features of the network, such as the dynamics of epidemic, resulting in a biased understanding of the epidemic spreading process. As shown in Fig. 1, suppose an infectious disease occurs in a community $C_1$, and unfortunately, Alice is infected. In order to delay the epidemic spreading, Bob should be protected first, because of his close contact to Alice and other people. In this situation, although Carl has a high social status in the whole network, the infectious disease would break out in the community and then spread to the entire network if he was immunized first.

To this end, we investigate the evolution of community structure in opportunistic social networks, and analyze the effect of community-based immunization strategy on epidemic spreading. We make the following contributions.

We find that the spreading speed of epidemic within one community is faster than that across different communities. This result is encouraging as it indicates that the outbreak of epidemic could be delayed, if one could further break down the OSN traces into lots of small communities by removing some special nodes.

We observe that the most efficient immunization strategy on epidemic spreading is to remove nodes with high local importance in communities. More interestingly, we find that high local importance but non-central nodes contribute more to epidemic spreading process, leading us to conjecture that it is not “the more central, the better” for the implementation of control strategy.

We study the role of nodes’ local importance in epidemic spreading, experimentally and analytically. We do so not only by excluding nodes with high local importance from OSN traces but also by developing an analytic model which formally characterizes the relationship between nodes’ local importance and the community cohesion. We find that the community cohesion is heavily dependent on local importance of nodes. Removing locally important nodes sharply lowers the community cohesion, and thus helps to suppress the epidemic.

We organize the remainder of this paper as follows. Section 2 reviews related works. Section 3 introduces the OSN traces and network model. The next two sections present our solutions to evaluate node’s local importance and cluster nodes, respectively. We analyze the effect of community-based immunization on epidemic spreading in Section 6. Finally, in Section 7 we conclude our paper.

2. Related work

Immunization strategies in general can be classified into two categories: the ring strategy [13] and targeted strategy [14–18]. The targeted immunization strategy, in which nodes playing an important role in the network are removed first, has shown to be very effective [16]. Pastor-Satorras et al. [14] first studied the targeted immunization strategy and found that immunizing nodes with strong connectivity is more effective than that of randomly selected nodes. Holme et al. considered a more challenging scenario, where only neighborhood information of nodes can be used. They observed that the most efficient strategy is to iteratively immunize the neighbor of nodes with big out-degree [15]. Subsequently, Zhang et al. [17] developed a more precise model by taking the immunization willingness of individuals into account. In addition, Schneider [18] tested the role of node’s betweenness centrality in epidemic spreading, and obtained the similar result.

Recently, the concept of node importance has been introduced to the fields of routing and message diffusion in opportunistic social networks. For example, the authors of Refs. [25,12,26] exploited the node importance to make routing decisions. They found that forwarding messages to nodes with high centrality could increase the message delivery ratio. Similarly, the authors of Refs. [21,22] observed that nodes wandering from one community to another (such nodes thus have a high contact frequency with others in the network) contributed more to message diffusion.

All of the above works focus on the central nodes in a network, they evaluated node’s importance either by logical centrality metric or by physical contact behavior among nodes. Our work supplements the previous results, exploring the role
of potential non-central, locally important nodes in epidemic spreading. Furthermore, our understanding on the effect of locally important nodes motivates studies of applications beyond epidemic spreading, such as publish/subscribe systems [4,5], multicasting [27] and location of rumor source [28].

3. Preliminaries

3.1. Real and synthetical traces

We use real and synthetic traces in this study, both of them have their own advantages and can be complementary to each other. The former helps to observe the real epidemic spreading behavior; the latter provides an opportunity to study the dynamics of epidemic spreading in a large scale. The details of each kind of trace are discussed below.

**Real trace (NCSU).** The NCSU trace [29] is taken by twenty students who live in a campus dormitory. Every week, these students carried Garmin GPS 60CSx handheld receivers, which are WAAS (Wide Area Augmentation System) capable with a position accuracy of better than three meters 95% of the time, for their daily regular activities, and altogether thirty-five trajectories were gathered from 2006-08-26 to 2006-11-16. The GPS receivers perform a device discovery every 10 s to record their current positions. To reduce GPS errors, the authors of Ref. [29] calibrate a position at every 30 s by averaging three samples over the past thirty-second period. According to the position information in each trajectory, we assume that two nodes have a contact if their distance is less than 150 m, a realistic range for WiFi transmissions.

**Synthetical trace (SLAW).** Considering the limitation on scalability of real trace, we need a proper mobility model that can characterize the nature of human mobility, to investigate the dynamics of epidemic spreading, both in width and depth. Although many random mobility models, such as Random Walk and Random Way Point, have been widely used in opportunistic social networks for evaluating routing performance or even the epidemic dynamics [30,31], they cannot reflect the main features of human mobility, including the truncated power-law flights and pause-times, the heterogeneously bounded mobility areas of different nodes, etc. The recent work [32] proposed a new mobility model called SLAW (Self-similar Least Action Walk) that can produce synthetical trace incorporating various features of human mobility. We therefore use it to produce synthetical human traces for investigation on epidemic spreading. The simulation area is $2000 \times 2000$ m$^2$, the number of nodes is varying from 100 to 500 with a step 100. Table 1 summarizes all SLAW model parameters (for detail meanings of these parameters, please refer to Ref. [32]).

| Parameter                      | Value       |
|--------------------------------|-------------|
| Distance alpha                | 3           |
| Number of nodes               | 100–500     |
| Simulation area               | $2000 \times 2000$ m$^2$ |
| Number of waypoints           | 2000        |
| Hurst parameter               | 0.75        |
| Simulation duration           | 5 h         |
| Clustering range              | 50 m        |
| Levy exponent for pause time  | 1           |
| Minimum/maximum pause time    | 60/3600 s   |

3.2. Network model

Traditional methods to model opportunistic social network mainly include time expanded graph and binary graph. The time expanded graph catches each snapshot of the original network, that is, new edges connecting a node and its copy
are added at the next snapshot with the edges having been existed in the last snapshot. It hence incurs to the scalability issue [33]. On the other hand, the binary graph can alleviate the storage overhead. It however neglects the duration of each contact and its decayed age. This method only characterizes the time-varying network at a coarse-grained level [12]. Considering these facts, we model an OSN as a Decayed Aggregation Graph DAG = (V, E), where V denotes the set of nodes (|V| = n) and E denotes the set of edges. Let \( W(t) = (w_{uv}(t))_{n \times n} \) denote its adjacency matrix and \( N_{uv}(t) = ((on_i, off_i), i = 1, 2, \ldots, N) \) denote the contact series between nodes \( u \) and \( v \) in the interval [0, t], where the tuple \( (on_i, off_i) \) denotes the start moment and end moment of the \( ith \) contact respectively, and \( N \) is the number of contacts. We formulate the contact strength between nodes (i.e., the value of \( w_{uv}(t) \)) as a Decayed Sum Problem [34]. Decayed Sum Problem includes two components. The first one is weighted function \( f(i) \), and the second one is decayed function \( g(T - off_i) \), as shown in the following equation.

**Definition 1 (Decayed Sum).** Given the contact series \( N_{uv}(t) \), the goal is to estimate the decayed sum at any current time \( T \)

\[
w_{uv}(T) = \sum_{i=1}^{N} f(i)g(T - off_i)
\]

where \( f(i) = off_i - on_i \) denotes the \( ith \) contact duration (i.e., the weighted part of Decayed Sum Problem) and \( g(T - off_i) \) denotes the decayed part. We set \( g(T - off_i) = e^{-(T - off_i)} \) as the inter-contact time between nodes generally obeys an exponential decay in OSN [35]. Hence, Eq. (1) can be reformulated as

\[
w_{uv}(T) = \sum_{i=1}^{N} (off_i - on_i)e^{-(T - off_i)}.
\]

We next analyze the space complexity of DAG. Exact tracking of \( w_{uv}(T) \) needs \( O(N) \) storage bits. Considering scalability issue (in general, \( N \gg n \)), we should further reduce the storage overhead while keeping the same calculation precision.

Let \( h(t) = off_i - on_i \) if and only if \( t \) is equal to \( off_i \), otherwise, \( h(t) = 0 \), we obtain the following theorem (please see Appendix for proof).

**Theorem 1.** At each time slot \( t = 0, 1, 2, \ldots, T \), the value of \( w_{uv}(T) \) can be maintained easily using

\[
w_{uv}(T) = h(T) + e^{-1}w_{uv}(T - 1).
\]

From Theorem 1, each node only carries a single counter to exactly track the contact strength between itself and any other node, which forms the row vector \( w_u \) of matrix \( W \). We use this matrix to cluster nodes in the next section.

4. Local importance of nodes

This paper mainly concentrates on the local importance of nodes and its effect on epidemic spreading. As discussed in Sections 1 and 2, the local importance of nodes reflects their social status in a community, rather than in the whole network.\(^1\) Traditional solutions for evaluating node importance mainly include the node degree, closeness and betweenness. All of them are not applicable, due to the unknown number of neighbors (for degree measure) and vulnerable end-to-end path (for closeness and betweenness measures) in opportunistic social networks. For example, if we use the betweenness to measure node importance, we first need to collect the shortest paths for each node pairs in an offline way, and then count the number of times that each node appears in these shortest paths. Furthermore, the three methods have a bias towards a global measure of nodes. To deal with these issues, we use the technology of Principal Component Analysis (PCA) [36] to evaluate node’s local importance, which provides an online method for evaluating node importance.

**Principal component analysis:** Principal component analysis is a powerful tool to extract relevant information from a data set by filtering noise and redundant data. This relevant information reveals the hidden, simplified structures underlying the data set. We generalize the principle of PCA as follows.

Suppose that a node \( u \) has built the matrix \( W \) (please refer to Section 3.2), and the matrix \( W \) has been centralized (i.e., subtract the corresponding mean from each column). Let \( C_W = W^TW/(n - 1) \) denote the covariance matrix of \( W \). Let us further diagonalize the \( C_W \) as

\[
P^T C_W P = \Lambda
\]

where \( \Lambda = \text{diag}(1, 2, \ldots, n) \) and \( P \) is a normalized orthogonal matrix. Let \( \lambda_i \) be the eigenvectors of \( C_W \) and \( \lambda_i \) the corresponding eigenvalues, and \( \lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \). We can see from Fig. 2 that the row vector \( \alpha_u (\alpha_{u1}, \alpha_{u2}, \ldots, \alpha_{un}) \)

\(^1\) We use the term “local importance” to specifically denote node’s social status in a community and the term “global importance” to denote that in the whole network.
Fig. 2. The spectral space of $W$ and its vector representation.

### Table 2
The main notations used in the paper.

| Notation | Explanation |
|----------|-------------|
| $w_u$    | The row vector of matrix $W$ |
| $C_W$    | The covariance matrix of $W$ |
| $P_{k+1}$ | The noise components of $W$ |
| $P_k$    | The principal components of $W$ |
| $P$      | The eigenvector decomposition of $C_W$ |
| $W_k$    | The eigenvector decomposition of $C_W$ |
| $C_{W_k}$ | The covariance matrix of $W_k$ |
| $A$      | The diagonal matrix of $C_W$ |
| $x_i$    | The $i$th eigenvector of $C_W$ |
| $x^T_i$  | The matrix $x_i$ |
| $\alpha^{+}_u$ | The distribution of node $u$ in the $n$-dimensional spectral space |
| $\alpha^{+1..k}_u$ | The signal distribution of $\alpha^+_u$ |
| $\alpha^+_{u,1..n}$ | The noise distribution of $\alpha^+_u$ |

denotes the distribution of node $u$ in the $n$-dimensional spectral space, and the column vector $x_i(\alpha^+_1, \alpha^+_2, \ldots, \alpha^+_n)$ denotes the coordinates of all of the nodes in the $i$th dimension of the spectral space. In addition, once we get the orthogonal matrix $P$, we generally select the top $k$-dimensional spectral space $(x_1, x_2, \ldots, x_k)$ as the principal component of $W$, since the corresponding top $k$ eigenvalues dominate the spectral graph features [37]. Algorithm 1 describes the above computation process and Table 2 lists the main notations used in the paper.

### Algorithm 1 PCA

1: Input: an adjacency matrix $W$ of DAG
2: Output: orthogonal matrix $P$ and diagonalized matrix $A$
3: $W = \text{centralized}(W)$
4: $C_W = \text{cov}(W)$
5: $[P, A] \leftarrow \text{eigs}(C_W, n)$

Mathematically, let $\Lambda_k$ denote $\text{diag}(1, 2, \ldots, k)$ and the matrix $P_k = (x_1, x_2, \ldots, x_k)$. Let $\alpha^+_u$ represent $(|\alpha^{+1}_u|, |\alpha^{+2}_u|, \ldots, |\alpha^{+n}_u|)$, where $|\alpha^{+u}_i|$ denotes the absolute value of $\alpha^{+u}_i$, we have

**Lemma 1.** For a given decayed aggregation graph DAG with $k$ communities, the matrix $P_k$ is the projection matrix, elements of the vector $\alpha^{+}_u$ are the projected values of node $u$ in such $k$ communities.

**Proof.** Let $W_k$ denote the dimensionality reduction matrix of $W$ and the matrix $C_{W_k}$ be the covariance matrix of $W_k$. Based on the theory of PCA, $C_{W_k}$ should be diagonalized as well, we have

$$C_{W_k} = \frac{W_k^TW_k}{n-1} = \Lambda_k. \quad (5)$$

On the other hand, from Eq. (4), we get

$$P^T C_W P = A \Rightarrow P_k^T C_W P_k = \Lambda_k. \quad (6)$$

Replace $\Lambda_k$ with Eq. (5) and $C_W$ with $W^TW/(n-1)$, respectively, we have

$$\frac{W_k^TW_k}{n-1} = P_k^T C_W P_k \Rightarrow \frac{W_k^TW_k}{n-1} = \frac{P_k^TW^TW P_k}{n-1}.$$
Multiply both sides by \((n - 1)\) and use the substitution of \(P_k^T W = (WP_k)^T\), we get
\[
WP_k = W_k.
\]
Hence, we conclude that the matrix \(P_k\) is the projection matrix.

Let \(c_i\) denote the principal direction of the \(ith\) eigenvector \(x_i\), Based on the singular value decomposition (SVD) of \(W_k\) \([38]\), both \(c_i\) and \(x_i\) are satisfying:
\[
W_k^T c_i = \lambda_i c_i, \quad W_k^T W_k x_i = \lambda_i x_i.
\]
After some algebra, we obtain \(x_i = \frac{W_k^T c_i}{\lambda_i} / \lambda_i\), which shows that elements of \(x_i\) are the projected values of nodes on the principal direction \(c_i\). Hence, \(|\alpha_{ui}|\) is the projection length of node \(u\) in community \(i\).

**Theorem 2** (Node \(u\)'s Local Importance). Let \(L_u^i\) denote the local importance of node \(u\) with respect to a community \(i\), we have
\[
L_u^i = |\alpha_{ui}| \lambda_i.
\]

**Proof.** From Lemma 1, we know that the projection length of node \(u\) in community \(i\) is \(|\alpha_{ui}|\), and from the spectral graph theory \([37]\), it has been shown that the eigenvalue \(\lambda_i\) indicates the strength of community \(i\) in a graph. Hence, we get
\[
L_u^i = |\alpha_{ui}| \lambda_i.
\]
Obviously, the \(L_u^i\) is mathematically equivalent to \(|\alpha_{ui}|\) if we ignore the factor \(\lambda_i\) (considering the fact that local importance of each node in community \(i\) has the common factor \(\lambda_i\)). We specifically use the above equation to denote node's local importance, this is mainly because the product of the two parts \((|\alpha_{ui}|\) and \(\lambda_i\)) has a special physical significance, it reflects the contribution of community \(i\) to node \(u\)'s global importance \(G_u\), that is,
\[
G_u = \sum_{i=1}^{k} |\alpha_{ui}| \lambda_i.
\]

With the above two equations\(^2\), we evaluate the local and global importance of each node, and plot their empirical distributions in Fig. 3, where the subgraphs correspond to the case in which the local importance of nodes is measured with respect to the largest community (Fig. 3(a)), the second (Fig. 3(b)), the third (Fig. 3(c)) and the fourth largest community (Fig. 3(d)), respectively. We observe that, for most nodes, the global importance shows a strong correlation with the local importance. Furthermore, the correlation decreases with decreasing community size, as small communities play a weak role in a graph. Another interesting observation is that the two social metrics of nodes have different increasing rates, which result in some central nodes having a relatively low importance in a community, and vice versa. In Section 6, we study their effects on epidemic spreading.

### 5. Community detection

Cutting a graph into small clusters has been studied widely. We use the \(k\)-means, one of the most well-known clustering algorithms \([39]\), to detect the temporal community structure. The advantage of the \(k\)-means algorithm compared to other methods such as CNM \([40]\) and \(k\)-clique \([41]\) is that it does not need to know the neighbor relationship between nodes, it only requires the adjacent matrix of a weighted graph such as the DAG, while the CNM and \(k\)-clique are more appropriate to a binary graph. In addition, based on the technology of PCA discussed above, it is confident to determine the number of communities, the initial elements for each community and the termination condition, three issues strongly affecting the performance of \(k\)-means. We next discuss how to detect the community structure based on the refined \(k\)-means.

\(^2\) For the proof of Eq. (9), please refer to Appendix.
(1) Determining \( k \), the number of communities: PCA provides a roadmap to reduce a confusing data set to a lower dimension that retains the main features of the original data set. The rationale behind this is that the eigenvalues of a network, play a big role in many important graph features. It has been shown that the maximum degree, clique number, and even the randomness of a graph are all related to \( \lambda_1 \). In general, we select the top \( k \) eigenvectors to denote the main structures of the graph, where the value of \( k \) satisfies
\[
\sum_{i=1}^{k} \lambda_i / \sum_{j=1}^{n} \lambda_j \geq R. \tag{10}
\]

In this paper, we set \( R = 0.85 \), an experimental value\(^3\) belonging to the default interval \([0.7, 0.9]\) [37].

(2) Excluding the noise nodes: In opportunistic social networks, there commonly exist some nodes with few contacts to other nodes. We call them noise nodes in this study. Excluding the noise nodes has little effect on epidemic spreading speed. Furthermore, it helps to reduce the problem size and algorithm complexity. We now discuss how to use PCA to identify the noise nodes.

PCA divides a network into two different parts: (1) the principal components \( P_k \), and (2) the opposite \( P_{k+1} \), where the \( P_{k+1} = (x_{k+1}, x_{k+2}, \ldots, x_{n}) \), as shown in Fig. 2. We call the latter noise components of the network. Accordingly, we divide the row vector \( \alpha_u \) by \( \alpha_u^{1:k} (\alpha_u, \alpha_{u2}, \ldots, \alpha_{un}) \) and \( \alpha_u^{k+1:n} (\alpha_{u,k+1}, \alpha_{u,k+2}, \ldots, \alpha_{un}) \), the signal and noise of the node \( u \). From the theory of PCA, if node \( u \) is dominated by its noise components, that is, \( \alpha_u \) is dominated by the \( \alpha_u^{k+1:n} \), we can exclude this node from the graph. We use signal-to-noise ratio to identify which component dominates a node.

**Definition 2** (Node \( u \)'s Signal-to-Noise Ratio \( SNR_u \) = \( \sum_{i \in [1,k]} (\lambda_i |\alpha_{ui}|)^2 / \sum_{j \in [k+1,n]} (\lambda_j |\alpha_{uj}|)^2 \). The SNR is the ratio of signal energy over that of noise.

From Theorem 2, we know that the node \( u \)'s local importance relative to community \( i \) is \( |\alpha_{ui}| \lambda_i \), which is also the amplitude of node \( u \)'s signal in the \( i \)th dimensional spectral space. Hence, the signal energy \( e_{\text{signal}}^u \) of node \( u \) can be presented as \( e_{\text{signal}}^u = \sum_{i \in [1,k]} (\lambda_i |\alpha_{ui}|)^2 = \sum_{i \in [1,k]} (\lambda_i |\alpha_{ui}|)^2 \). The noise strength \( e_{\text{noise}}^u \) is equal to \( \sum_{j \in [k+1,n]} (\lambda_j |\alpha_{uj}|)^2 \).

Based on Definition 2, we call node \( u \) a noise node if its \( SNR_u \) satisfies \( SNR_u < 1 \).

(3) Determining the initial elements for each community: After we have ascertained the number of communities and excluded the noise nodes, the next step is to determine the initial centroid \( m_i \) (\( i = 1, 2, \ldots, k \)) for each community \( C_i \). We select the node \( u \), s.t. \( \max |\alpha_{ui}| \) (\( u = 1, 2, \ldots, n \)) for each eigenvector \( x_i \), as the initial node of community \( i \), and set \( m_i = \alpha_u \). Algorithm 2 describes this procedure.

**Algorithm 2** The Initial Centroid

1: Input: \( P_k \), \( \text{maxValue} \leftarrow 0 \), \( v \leftarrow 0 \)
2: Output: \( m_i; i = 1, 2, \ldots, k \)
3: for \( i = 1 \) to \( k \) do
4: \( \text{maxValue} \leftarrow |\alpha_{ui}| \)
5: \( v \leftarrow 1 \) (Tracking who is the maximum)
6: for \( u = 2 \) to \( n \) do
7: if \( |\alpha_{ui}| > \text{maxValue} \) then \( u \) is not a noise node then
8: \( \text{maxValue} \leftarrow |\alpha_{ui}|, v \leftarrow u \)
9: end if
10: end for
11: \( C_i \leftarrow C_i \cup \{v\}, m_i \leftarrow \alpha_v \)
12: end for

(4) Termination condition of \( k \)-means: Suppose all of the non-noise nodes have been clustered, and the \( m_i \) is updated by
\[
m_i = \left( \sum_{u \in C_i} \alpha_u \right) / n_i \tag{11}
\]

where \( n_i \) is the number of nodes belonging to \( C_i \). \( k \)-means is characterized by minimizing the sum of squared errors,
\[
J = \sum_{i=1}^{k} \sum_{u \in C_i} (\alpha_u - m_i)^2. \tag{12}
\]

\(^3\) We tried different values of \( R \) and observed no significant difference for ranges of \( 0.7-0.9 \) in the number of communities. For the NCSU trace, the number of communities varies from 5 to 7 at moment \( t = 12,000 \) s; For the SLAW trace (\( n = 500 \)), the number of communities varies from 10 to 13 at the same moment.
It has been shown that the standard iterative method to $k$-means suffers seriously from the local minima problem, because of the greedy nature of the update strategy. Fortunately, Theorem 3 guarantees the PCA-based $k$-means is immune to this problem.

**Theorem 3** (Theorem 3.2 in Ref. [42]). Minimizing $J$ is equivalent to maximizing trace($P^T C_W P$) (please refer to Eq. (19) of Ref. [42]), and max trace($P^T C_W P$) = $\lambda_1 + \lambda_2 + \cdots + \lambda_k$.

In other words, the PCA-based $k$-means has reached the optimal performance once we cluster all of the non-noise nodes for the first time.

**5) Clustering nodes:** For any node $u$, we compute the distance between itself and the centroid $m_i$, $\text{dist}(\alpha_u, m_i)$, and select $i$, s.t. min $\text{dist}(\alpha_u, m_i)$ ($i = 1, 2, \ldots, k$) as the community node $u$ belongs to, where,

$$
\text{dist}(\alpha_u, m_i) = \theta(u, i) = \arccos \frac{\alpha_u^T m_i}{\|\alpha_u\|_2 \|m_i\|_2}
$$

and $\theta(u, i)$ denotes the angle between $\alpha_u$ and $m_i$. After node $u$ joins the community $i$, we update the centroid $m_i$ by Eq. (11) so as to select the next node. Algorithm 3 describes the clustering procedure.

6. **Results and analysis**

6.1. **Temporal community structure underlying the traces**

We define the temporal community structure as a series of snapshots of communities underlying the traces. We take a snapshot every 120 s for the communities. Compared to the five hours duration of experiment, the snapshot interval is chosen to be relatively small so as to obtain a detailed view of the community evolution and to make an unbiased understanding of epidemic spreading as much as possible.

**Fig. 4** plots the number of communities hidden behind the traces at different snapshots, where the term “S(number)” denotes the SLAW trace with different number of nodes. We observe that the topology is volatile over time. At NCSU, the number of communities varies from 4 to 10 with a mean of 7.3. The number of communities at $S(100)$ varies between 8 and 16 with a mean of 11.7. **Table 4** summarizes the statistics of community structure for all the traces.

**Fig. 5** shows size of the top 4 communities at the SLAW trace with 500 nodes. We find that each community is not stable over time as well. The size of the largest community varies between 13% and 35% during the experiment, and 11%–20% of the nodes belong to the second largest community. The third and fourth largest community are much closer, varying from 9% to 17% and from 8% to 15%, respectively. In summary, the top 4 communities cover almost 60% nodes. Other smaller communities share the rest nodes.

These results suggest that nodes in opportunistic social networks do not belong to a single, stable community. Instead, the network is made of many temporal clusters. We analyze their effects on epidemic spreading in the next section.
6.2. Role of temporal community in epidemic spreading

Having shown that the temporal communities are built on node’s social contacts, it is significant to understand the role of these communities in epidemic spreading. To this end, we record the epidemic spreading time within one community and that across different communities. Table 3 summarizes the results. We find that the spreading speed of epidemic within one community is faster than that from one community to another. At NCSU, the mean spreading time within one community is 19 min, and 45 min between communities. The SLAW traces show similar phenomena, especially at S(300), where the inter-community spreading time is almost four times of that of intra-community. This is mainly because there exist many small communities at this scenario (as shown in Table 4, the S(300) trace has the most number of communities and the community structure changes dramatically). The small communities are more effective to refrain the epidemic spreading than the big ones. This result is encouraging as it indicates that the outbreak of epidemic could be delayed, if one could break down the OSN traces into lots of small communities by removing some special nodes.

Previous work has suggested that removing central nodes is an effective way to delay the epidemic outbreak. This conclusion is somewhat inconsistent with our aforementioned results, since some central nodes have relatively low importance in communities and removing them does little damage to the community structure. Therefore, we conjecture that it may be inefficient to suppress the epidemic spreading by removing central nodes. To validate this, we classify nodes into different categories according to their global and local importance. Specifically, we select as High LoCal imporTance nodes (HOT) with the top 10% nodes that have the highest local importance with respect to a community, and the rest as Low LoCal imporTance nodes (LOT). In addition, we define a central node as a node that belongs to the top 10% nodes with the highest importance in the whole network, the remaining nodes are called non-central nodes (similar selection/definition has been used in Ref. [43]). Our basic idea here is to understand the role of each kind of node in epidemic spreading.

We first calculate the epidemic spreading time for each trace including all nodes. We then repeat the same experiment by removing each of the four node categories (we remove the same amount of nodes for each category in order to make a fair comparison). Fig. 6 presents the results. The y-axis denotes the immunization efficiency compared to the base case (the case including all nodes), i.e., higher bar means higher efficiency. The first and counter-intuitive phenomenon is that HOT nodes, instead of the central, play a big role in epidemic spreading (as shown in Fig. 6(a)). Compared to removing the central nodes, the immunization efficiency increases 50% on average when the HOT nodes are removed. More interestingly, we observe that non-central HOT nodes are responsible for most of the epidemic spreading in opportunistic social networks (Fig. 6(b) and (c)). Removing them improves the immunization efficiency by 25%–150% at all traces. In contrast, removing central LOT nodes shows a more limited improvement. This phenomenon experimentally validates that is not “the more central, the better” for the implementation of control strategy. We next explore the reason behind this phenomenon.
6.3. Theory analysis

Section 6.2 indicates that the spreading speed of epidemic heavily depends on the temporal communities. As a result, even though HOT nodes have on average lower global importance than the central, they do great damage to the community structure when removed, and thus help to suppress the epidemic. The goal of this section is to formally characterize the relationship between the cohesive community and node’s importance. We generally use the community density to denote its cohesion. Let $D(C_i)$ represent density of community $i$, we have the following theorem (see Appendix for proof).

**Theorem 4.** The community density $D(C_i)$ can be expressed as

$$D(C_i) = \lambda^{-3/2} \sum_{i=1}^{n} (U_i)^2.$$  

(13)

We use this equation to evaluate the impact of removing nodes on community density, and plot the changing of community density in Fig. 7, where the $Y$-axis denotes the ratio of community density after some nodes are removed from a community over the initial density of that community. We find that community density is heavily dependent on node’s importance. However, there exists a large difference of impact between the local and global importance of nodes. Removing in rank order the most important to the least nodes in a community leads to a faster decline in community cohesion. In contrast, removing first the central nodes will shrink the community but with a slower speed. Taken together, HOT nodes appear to be crucial for implementing the immunization strategy, because of their large impact on community structure when removed.

7. Conclusion

In this paper, we improve our understanding of immunization strategy on epidemic spreading in opportunistic social networks. We observe that a temporal community structure helps to control the epidemic spreading. This phenomenon is encouraging as it indicates that the outbreak of epidemic could be delayed, if we could further break down the OSN traces into lots of small communities by removing some special nodes. Motivated by this observation, we separate nodes into different behavioral classes from a community viewpoint. We show that HOT nodes can remarkably suppress the epidemic spreading when removed. More interestingly, we find that non-central HOT nodes are responsible for most of the epidemic spreading. These results reveal a counter-intuitive conclusion: it is not “the more central, the better” for the implementation of control strategy.
Acknowledgments

We acknowledge the support of the National Natural Science Foundation of China under Grant Nos. U1404602, U1304607, the Science and Technology Foundation of Henan Educational Committee under Grant Nos. 14A520031, 14A520068. We also wish to thank the CRAWDAD archive project for making the DTN traces available to research community.

Appendix

Proof of Theorem 1. Let us split the interval $[0, T]$ into two disjoined parts $T_1$ and $T_2$, where $T_1 = \cup_{i=1}^{N} t_i (t_i = [on_i, off_i])$ and $T_1 \cup T_2 = [0, T]$. We have

$$\sum_{t \leq T} h(t) e^{-\lambda t} = \sum_{t \in T_1} h(t) e^{-\lambda t} + \sum_{t \in T_2} h(t) e^{-\lambda t}.$$  

For any $t \in T_2$, since $t \neq off_i$ (note that $off_i \in T_1$ and $T_1 \cap T_2 \equiv \emptyset$), we have $h(t) = 0$. Hence,

$$\sum_{t \leq T} h(t) e^{-\lambda t} = \sum_{t \in T_1} h(t) e^{-\lambda t} = \sum_{i=1}^{N} \sum_{t \in t_i} h(t) e^{-\lambda t} = \sum_{i=1}^{N} (off_i - on_i) e^{-\lambda (T - off_i)}.$$  

From Eq. (2) p. 4, we get,

$$w_{uv}(T) = \sum_{t \leq T} h(t) e^{-\lambda t} = h(T) e^{-\lambda T} + \sum_{t \leq T-1} h(t) e^{-\lambda t} = h(T) + \sum_{t \leq T-1} h(t) e^{-\lambda (T-1-t+1)} = h(T) + e^{-\lambda} \sum_{t \leq T-1} h(t) e^{-\lambda (T-1-t)} = h(T) + e^{-\lambda} w_{uv}(T-1).$$  

Proof of Eq. (9). We first give the following lemma.

Lemma 2. Nodes from $k$ communities form $k$ quasi-orthogonal lines in the spectral space.

Proof. From the clustering process mentioned above, we know that the centroid $m_i$ ($m_{1i}, m_{2i}, \ldots, m_{ni}$) can approximately represent the line formed by nodes within the $i$th community (please refer to Eq. (11) p. 7). On the other hand, the virtual centroid vector $\tilde{m}_i$ should be close to eigenvector $\tilde{x}_i$. This is mainly because $m_i \approx \tilde{m}_i = (\sum_{u \in C_i} \alpha_{ui}) / n_i$, as $\alpha_{ui}$ is the dominant part of $\alpha_u$. Hence, $\tilde{m}_i$ locates in the line formed by the eigenvector $\tilde{x}_i$. We get the conclusion as different eigenvectors are linearly independent.  

We now prove Eq. (9). Let variable $E_u$ denote the event measuring global importance of node $u$ (i.e., $G_u$) in the whole network. Let variable $E_{u}^i$ denote the event that measures node $u$'s local importance in a community $i$. We have

$$P(E_u) = P(\text{measuring } G_u \text{ in the whole network}) = P(\text{measuring } G_u \text{ across all communities}) = P\left(\bigcup_{i=1}^{k} E_{u}^i\right) = \sum_{i=1}^{k} P(E_{u}^i) \text{ from the Lemma 2.}$$  

That is, measuring node $u$'s global importance is equal to first measuring its local importance in different communities, and then put all the components together.  

Proof of Lemma 2. From the clustering process mentioned above, we know that the centroid $m_i$ ($m_{1i}, m_{2i}, \ldots, m_{ni}$) can approximately represent the line formed by nodes within the $i$th community (please refer to Eq. (11) p. 7). On the other hand, the virtual centroid vector $\tilde{m}_i$ should be close to eigenvector $\tilde{x}_i$. This is mainly because $m_i \approx \tilde{m}_i = (\sum_{u \in C_i} \alpha_{ui}) / n_i$, as $\alpha_{ui}$ is the dominant part of $\alpha_u$. Hence, $\tilde{m}_i$ locates in the line formed by the eigenvector $\tilde{x}_i$. We get the conclusion as different eigenvectors are linearly independent.  

We now prove Eq. (9). Let variable $E_u$ denote the event measuring global importance of node $u$ (i.e., $G_u$) in the whole network. Let variable $E_{u}^i$ denote the event that measures node $u$'s local importance in a community $i$. We have

$$P(E_u) = P(\text{measuring } G_u \text{ in the whole network}) = P(\text{measuring } G_u \text{ across all communities}) = P\left(\bigcup_{i=1}^{k} E_{u}^i\right) = \sum_{i=1}^{k} P(E_{u}^i) \text{ from the Lemma 2.}$$  

That is, measuring node $u$'s global importance is equal to first measuring its local importance in different communities, and then put all the components together.  

Proof of Eq. (9). We first give the following lemma.

Lemma 2. Nodes from $k$ communities form $k$ quasi-orthogonal lines in the spectral space.

Proof. From the clustering process mentioned above, we know that the centroid $m_i$ ($m_{1i}, m_{2i}, \ldots, m_{ni}$) can approximately represent the line formed by nodes within the $i$th community (please refer to Eq. (11) p. 7). On the other hand, the virtual centroid vector $\tilde{m}_i$ should be close to eigenvector $\tilde{x}_i$. This is mainly because $m_i \approx \tilde{m}_i = (\sum_{u \in C_i} \alpha_{ui}) / n_i$, as $\alpha_{ui}$ is the dominant part of $\alpha_u$. Hence, $\tilde{m}_i$ locates in the line formed by the eigenvector $\tilde{x}_i$. We get the conclusion as different eigenvectors are linearly independent.  

We now prove Eq. (9). Let variable $E_u$ denote the event measuring global importance of node $u$ (i.e., $G_u$) in the whole network. Let variable $E_{u}^i$ denote the event that measures node $u$'s local importance in a community $i$. We have

$$P(E_u) = P(\text{measuring } G_u \text{ in the whole network}) = P(\text{measuring } G_u \text{ across all communities}) = P\left(\bigcup_{i=1}^{k} E_{u}^i\right) = \sum_{i=1}^{k} P(E_{u}^i) \text{ from the Lemma 2.}$$  

That is, measuring node $u$'s global importance is equal to first measuring its local importance in different communities, and then put all the components together.
Proof of Theorem 4. Consider the division of an unweighted graph $U$ into $k$ non-overlapped communities $C_1, C_2, \ldots, C_k$. Let $v_i = (v_{i1}, v_{i2}, \ldots, v_{in})$ be the index vector of community $C_i$, and $v_{ui}$ is equal to 1 if node $u$ belongs to $C_i$ and 0 otherwise. For community $C_i$, its density can be expressed as

$$D(C_i) = \frac{\text{number of edges in } C_i}{\text{number of nodes in } C_i} = \frac{v_i^T A v_i}{v_i^T v_i}$$

where $A$ is the adjacency matrix of graph $U$. Based on Wielandt’s theory [44], we can relax $v_{ui} \in \{0, 1\}$ to real space and take $v_i$ to $x_i$. Therefore, the density $D(C_i)$ in a weighted graph like DAG can be presented as follows.

$$D(C_i) = \frac{x_i^T W x_i}{x_i^T x_i}.$$  

Replace $W$ with its singular value decomposition [38]: $W = \sum \lambda_i 1/2 x_i x_i^T$, we have

$$D(C_i) = \frac{x_i^T \left( \sum_{i=1}^{k} \lambda_i 1/2 x_i x_i^T \right) x_i}{x_i^T x_i} = \frac{x_i^T (\lambda_i 1/2 x_i x_i^T) x_i}{x_i^T x_i} = \lambda_i 1/2 x_i x_i^T = \lambda_i 1/2 \sum u a_{ui}^2.$$  

From Eq. (8) p. 6, we get $l_{ui} = |v_{ui}| \lambda_i$. Replace $a_{ui}$ with $l_{ui}/\lambda_i$, we get $D(C_i) = \lambda_i^{-3/2} \sum u (l_{ui})^2$.

References

[1] K. Fall, S. Farrell, DTN: an architectural retrospective, IEEE J. Sel. Areas Commun. 26 (5) (2008) 828–836.

[2] M. Conti, M. Kumar, Opportunities in opportunistic computing, Computer 43 (1) (2010) 42–50.

[3] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, in: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking, ACM, 2005, pp. 244–251.

[4] P. Costa, C. Mascolo, M. Musolesi, G.P. Picco, Socially-aware routin for publish–subscribe in delay-tolerant mobile ad hoc networks, IEEE J. Sel. Areas Commun. 26 (5) (2008) 748–760.

[5] W. Gao, G. Cao, A. Iyengar, M. Srivatsa, Supporting cooperative caching in disruption tolerant networks, in: 2011 31st International Conference on Distributed Computing Systems (ICDCS), IEEE, 2011, pp. 151–161.

[6] N. Sastry, D. Manjunath, K. Sollins, J. Crowcroft, Data delivery properties of human contact networks, IEEE Trans. Mob. Comput. 10 (6) (2011) 868–880.

[7] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, Impact of human mobility on opportunistic forwarding algorithms, IEEE Trans. Mob. Comput. 6 (6) (2007) 606–620.

[8] A.G. Miklas, K.K. Golli, K.K. Chan, S. Saroiu, K.P. Gummadi, E. De Lara, Exploiting social interactions in mobile systems, in: UbiComp 2007: Ubiquitous Computing, Springer, 2007, pp. 409–428.

[9] J. Tang, S. Sollato, M. Musolesi, C. Mascolo, V. Latora, Small-world behavior in time-varying graphs, Phys. Rev. E 81 (5) (2010) 055101.

[10] D.J. Klein, J. Hespanha, U. Madhow, A reaction–diffusion model for epidemic routing in sparsely connected manets, in: 2010 Proceedings IEEE INFOCOM, IEEE, 2010, pp. 1–9.

[11] X. Shi, L.A. Adamic, M.J. Strauss, Networks of strong ties, Physica A 378 (1) (2007) 33–47.

[12] P. Hui, J. Crowcroft, E. Yoneki, Bubblerap: Social-based forwarding in delay-tolerant networks, IEEE Trans. Mob. Comput. 10 (11) (2011) 1576–1589.

[13] J. Mather, B. Schönfisch, M. Kirkilionis, Ring vaccination, J. Math. Biol. 41 (2) (2000) 143–171.

[14] P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, Impact of human mobility on opportunistic forwarding algorithms, IEEE Trans. Mob. Comput. 6 (6) (2007) 606–620.

[15] A.-K. Pietilänen, C. Diot, Dissemination in opportunistic social networks: the role of temporal communities, in: Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2012, pp. 165–174.

[16] M.E. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (3) (2006) 036104.

[17] R. Pastor-Satorras, A. Vespignani, Immunization of complex networks, Phys. Rev. E 65 (3) (2002) 036104.

[18] P. Holme, Efficient local strategies for vaccination and network attack, Europhys. Lett. 68 (6) (2004) 908.

[19] Y. Chen, G. Paul, S. Havlin, F. Liljeros, H.E. Stanley, Finding a better immunization strategy, Phys. Rev. Lett. 101 (5) (2008) 058701.

[20] H. Zhang, J. Zhang, C. Zhou, M. Small, B. Wang, Hub nodes inhibit the outbreak of epidemic under voluntary vaccination, New J. Phys. 12 (2) (2010) 023015.

[21] C.M. Schneider, T. Mihaljev, S. Havlin, H.J. Herrmann, Suppressing epidemics with a limited amount of immunization units, Phys. Rev. E 84 (6) (2011) 061911.

[22] D. Mogenti, Tearing down the Internet, IEEE J. Sel. Areas Commun. 21 (6) (2003) 949–960.

[23] N.P. Nguyen, T.N. Dinh, S. Tokala, M.T. Thai, Overlapping communities in dynamic networks: their detection and mobile applications, in: Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, ACM, 2011, pp. 85–96.

[24] A.-K. Pietiläinen, C. Diot, Dissemination in opportunistic social networks: the role of temporal communities, in: Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2012, pp. 165–174.

[25] M.E. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (3) (2006) 036104.

[26] P.-Y. Yuan, H.-D. Ma, P.-R. Duan, Impact of strangers on opportunistic routing performance, J. Comput. Sci. Tech. 28 (3) (2013) 574–582.

[27] E.M. Daly, M. Haahr, Social network analysis for routing in disconnected delay-tolerant manets, in: Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2007, pp. 32–40.

[28] W. Gao, Q. Li, B. Zhao, G. Cao, Multicasting in delay tolerant networks: a social network perspective, in: Proceedings of the Tenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2009, pp. 299–308.

[29] D. Shah, T. Zaman, Finding rumor sources on random graphs. ArXiv Preprint arXiv:1110.6230.

[30] T. Spyropoulos, K. Psounis, C.S. Raghavendra, Efficient routing in intermittently connected mobile networks: the multiple-copy case, IEEE/ACM Trans. Netw. 16 (1) (2008) 77–90.

[31] X. Zhang, G. Neglia, J. Kurose, D. Towsley, Performance modeling of epidemic routing, Comput. Netw. 51 (10) (2007) 2867–2891.
[32] K. Lee, S. Hong, S.J. Kim, I. Rhee, S. Chong, SLAW: Self-similar least-action human walk, IEEE/ACM Trans. Netw. (TON) 20 (2) (2012) 515–529.
[33] B. George, S. Shekhar, Time-aggregated graphs for modeling spatio-temporal networks, in: Advances in Conceptual Modeling-Theory and Practice, Springer, 2006, pp. 85–99.
[34] E. Cohen, M. Strauss, Maintaining time-decaying stream aggregates, in: Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, ACM, 2003, pp. 223–233.
[35] T. Karagiannis, J.-Y. Le Boudec, M. Vojnovic, Power law and exponential decay of intercontact times between mobile devices, IEEE Trans. Mob. Comput. 9 (10) (2010) 1377–1390.
[36] I. Jolliffe, Principal Component Analysis, Wiley Online Library, 2005.
[37] F.R. Chung, Spectral Graph Theory, Vol. 92. American Mathematical Soc., 1997.
[38] G.H. Golub, C.F. Van Loan, Matrix Computations, Vol. 3. JHU Press, 2012.
[39] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (3) (2010) 75–174.
[40] A. Clauset, M.E. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (6) (2004) 066111.
[41] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
[42] C. Ding, X. He, K-means clustering via principal component analysis, in: Proceedings of the Twenty-First International Conference on Machine Learning, ACM, 2004, p. 29.
[43] J. Tang, T. Lou, J. Kleinberg, Inferring social ties across heterogenous networks, in: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, ACM, 2012, pp. 743–752.
[44] G.W. Stewart, J.-g. Sun, Matrix Perturbation Theory.