Learning Social Circles in Ego-Networks Based on Multi-View Network Structure

Chao Lan, Yuhao Yang, Xiaoli Li, Bo Luo, and Jun Huan

Abstract—Automatic social circle detection in ego-networks is a fundamentally important task for social network analysis. So far, most studies focused on how to detect overlapping circles or how to detect based on both network structure and node profiles. This paper asks an orthogonal research question: how to detect circles by leveraging multiple views of the network structure? As a first step, we crawl ego networks from Twitter and model them by six views, including user relationships, user interactions, and user content. We then apply both standard and our modified multi-view spectral clustering techniques to detect circles on these ego-networks. By extensive automatic and manual evaluations, we deliver two major findings: first, multi-view clustering techniques detect better circles than single-view clustering methods; second, our modified clustering technique which presumes sparse networks are incomplete detects better circles than the standard clustering technique which ignores such potential incompleteness. In particular, the second finding makes us conjecture a direct application of standard clustering on potentially incomplete networks may yield biased results. We lightly investigate this issue by deriving a bias upper bound that integrates theories of spectral clustering and matrix perturbation, and discussing how the bound may be affected by several network characteristics.

Index Terms—Social circle detection, privacy protection, multi-view spectral clustering, graph perturbation

1 Introduction

Online social network has been rising as a new and very popular platform for modern socialization - Facebook had recorded one billion active user accounts by late 2012, with about 10 million messages posted every hour and 46 percent of young users checking their Facebook as a first thing in the morning[1]. What lies behind this tremendous popularity, on the other hand, is a rich source of network information that could be properly integrated and analyzed for better understanding and promoting the modern online socialization, fulfilling the values of social network analysis.

In social network analysis, a fundamental and important task is to detect *social circles* in a user's ego-network (or, as we abbreviate as *ego-net*) [40]. Here, a user's ego-net is a sub-network that contains only her friends as nodes—the user is called the *ego*, each friend is called an *alter*, and a *social circle* is a subset of the alters who are similar under certain measurement. As suggested in [40], social circle has many potential applications, including content filtering and group recommendation. We also notice its particular application in the privacy and HCI research communities for controlling information boundary [52], [54], in a sense that an ego could have some new posts only visible to friends in designated social circles, which could reduce the risk of revealing her (private) information to untargeted friends. Indeed, it has been shown a user's information such as location could be inferred from her posts that contain local restaurants [35] or location-indicating words like "Time Square" [8], [12].

While the notion of social circle has been commercialized in several products including the Google+ circle and the Facebook custom list, it seems not well-received by users. As argued in [40], a main reason is most products require manual labeling of these circles, which is usually tedious and labor-intensive. To push the practice of social circle, it hence remains an important task to design methods that could automatically and effectively detect them in ego-nets.

Tracing this line of research, we notice the literature has been focused on addressing two questions, namely, how to detect circles that overlap and how to detect circles based on network node attributes (e.g., [7], [40], [64]); there is also an attempt to improve circle detection in a target ego-net by leveraging circle information from other ego-nets [16]. While these studies have advanced the practice of social circle in various directions, they all consider only a single view of the network structure. In reality, however, the ego-net structure may be described by multiple views—one view may show the friend relationship between alters while another may show their interaction frequencies. This simple observation motivates us to ask an orthogonal research question in this paper, i.e., *how to effectively leverage the (usually present) multiple views of ego-net structure for better social circle detection?*

To investigate the question, we first crawl ego-nets from Twitter and employ classic techniques to model the ego-net structure from six views, namely, two relationship views regarding the friendship and common friends between alters, three interaction views regarding the replies, co-replies and re-tweets of alters, and one content view regarding alters' post similarities. We do not use alter profiles (e.g., education, age or hobbies) as most studies do, considering alters may not provide these information due to privacy concerns.

1. http://www.statisticbrain.com/facebook-statistics/

- *C. Lan, X. Li, B. Luo, and J. Huan are with the EECS Department, University of Kansas, Lawrence, KS 66045. E-mail: {clan, xiaolili, bluo, jhuan}@ittc.ku.edu.*
- *Y. Yang is with Microsoft, Redmond, WA 98052. E-mail: yangyuhao05@gmail.com.*

Manuscript received 7 Dec. 2015; revised 3 Jan. 2017; accepted 6 Mar. 2017. Date of publication 21 Mar. 2017; date of current version 5 July 2017. Recommended for acceptance by S. Yan. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TKDE.2017.2685385

Then, we examine and compare several clustering techniques in their performance of detecting social circles based on the constructed multi-view ego-net structure. The examination includes a most common single-view clustering technique based solely on the friendship view [63], a benchmark single-view clustering technique that naively integrates all views into one and performs clustering, a standard multi-view clustering technique that fully transfers information across views [29], and our modification of this technique which now selectively transfers information across views. Based on extensive experimental evaluations, we have come to two major findings: first, multi-view techniques generally outperform single-view techniques in the qualities of detected circles; second, our modified multi-view clustering technique outperforms the standard multi-view technique.

The second observation raises our particular interest, as it suggests more careful interpretation and treatment of the sparse ego-net structures. Indeed, we have observed that 1) some views of an ego-net structure are very sparse and 2) our modified multi-view technique that selectively transfers information from sparse views to other views outperforms the standard multi-view technique that fully transfers information across views. Our conjecture for the co-occurrence of both phenomena is that the sparse ego-net structures may in fact be interpreted as incomplete structures (e.g., due to the limited time for data collection), and standard clustering techniques that ignore such 'hidden' incompleteness may output a result which deviates significantly from the optimal one. Note we are not blaming the sparseness for the performance degeneration, but the incompleteness that induces the sparsity. To better understand the issue, we also derive a performance deviation upper bound by integrating theories of spectral clustering and matrix perturbation, and discuss how it may be affected by several network characteristics. The obtained implications are supported in simulations.

In summary, the contributions of this paper[2] are unfolded in two phases. First, we propose to effectively leverage multiple views of the network structure for better automatic social circle detection in ego-nets. To that end, we introduce multi-view spectral clustering techniques and demonstrate they superior circle detection performance, as compared with common single-view clustering techniques. Second, we propose to interpret the sparseness of ego-net structure as incompleteness, and conjecture the ignorance of such hidden incompleteness may result in performance bias. To that end, we first derive an upper bound for the performance bias, with implications supported in simulations; we then propose a modified multi-view clustering technique which selectively transfers information from sparse views, and demonstrate its superior circle detection performance as compared with the standard multi-view clustering technique which fully transfers information across views. Finally, extensive experimental evaluations are done based on the ego-nets we crawled from Twitter.

The rest of this paper is organized as follows: Section 2 introduces the notations and problem setting; Section 3 introduces the multi-view ego-net structure we modeled; Section 4 presents the examined multi-view spectral clustering techniques as well as our interpretation of the network sparseness; Section 5 presents the experimental evaluations; related works are reviewed in section 6 and discussions in Section 7; Section 8 concludes the studies.

2. This paper is a journal extension of our previous study [65].

2 Notations and Problem Setting

For a matrix M, let M_{ij} be its entry at row i and column j, $M_{:j}$ be its j_{th} column and $M_{i:}$ be its i_{th} row; let M^T be its transpose, $||M||$ be its operator norm and $||M||_F$ be its Frobenius norm; when M is associated with view t, we denote it by $M^{(t)}$. Let I be an identity matrix properly sized by the context. For two matrices M, M' (of the same size), let $\succeq$ and $\succ$ be the Loewner partial orders such that $M \succeq M'$ if $M - M'$ is positive semi-definite and $M \succ M'$ if $M - M'$ is positive definite; let $M \circ M'$ be their Hadamard product. Finally, define $[\ell] := \{1, 2, \ldots, \ell\}$ for an integer $\ell > 0$.

Recall the structure of an ego-net could be described from multiple views, where each view corresponds to one type of connections between network nodes (i.e., alters). We characterize the view t of an ego-net structure by a similarity matrix $K^{(t)}$, such that $K^{(t)}_{ij}$ is some pre-defined similarity between alter i and alter j. (When referring to an arbitrary view, however, as we do in theoretical analysis, the superscript t may be omitted in notation.)

Now, given an ego-net consisting of n alters and characterized by multiple views $\{K^{(t)}\}$ where $t \in [T]$, our task is to automatically detect social circles based *solely* on $\{K^{(t)}\}$.

3 A Multi-View Ego-Net Structure

3.1 A Motivating Example

An advantage of considering multiple views of the ego-net structure is that different views may provide complementary information for more effective discovery of hidden social circles. Fig. 1 shows a sub-sample of the ego-net structure we crawled from Twitter, which consists of six alters (denoted by A, B, C, D, E, F respectively) and described from five views—(a) shows two relation views indicating the friend relations between alters and their common friend numbers; (b) shows two interaction views indicating the numbers of replies and retweets between alters; (c) shows a content view indicating similarities between alters' posts.

We see different types of views are partly consistent in suggesting the alters similarities, e.g., alters A and B not only have strong connections in the relation view, but also interact frequently based on the interaction view; on the other hand, although alters C and D are not friend (yet), it may still be helpful to group them since they have many friends in common and highly similar posts (i.e., they may still find a lot to talk with each other and thus promote the network information flow).

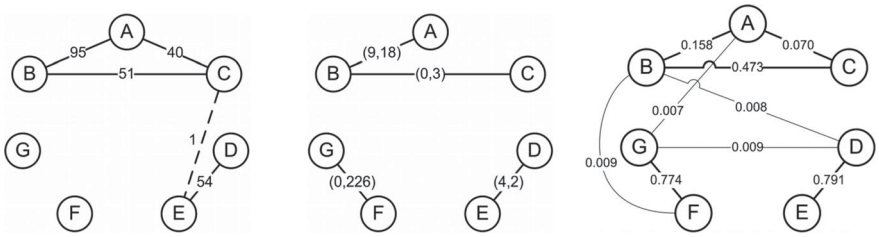
3.2 View Modeling

In this study, we crawl data from Twitter and employ classic techniques to model six views of its ego-net structures. These models are explained as below.

Friendship. This view characterizes the friend relation between alters by a similarity matrix $K^{(1)}$, where $K^{(1)}_{ij}$=1 if alters i and j follow each other on Twitter and $K^{(1)}_{ij}$=0 otherwise. It is a most common view for social circle detection.

Common Friend. This view characterizes the number of common friends between alters by a similarity matrix $K^{(2)}$, where $K^{(2)}_{ij} = m$ if alters i and j have m friends in common (excluding the alters i and j themselves).

Reply. This view characterizes the reply frequency between alters by a similarity matrix $K^{(3)}$, where $K^{(3)}_{ij} = m$ if alters i and j reply to one or another by m times in total.



(a) Structural Views (b) Interaction Views (C) Content View

Fig. 1. A Real-World Online Social Sub Ego-Net. In (a), a solid line indicates the connected alters are friends, while a dash line indicates they are not friends; the line label m indicates the two alters have m common friends. In (b), a line labeled by (m, n) indicates the connected alters have m replies and n re-tweets to each other in total. In (c), a line labeled by p indicates the connected alters have their posts similar by degree p (as explained in section 3.2), and we only show connections with label greater than 0.0065.

Co-Reply. This view characterizes the co-reply frequency of alters by a similarity matrix $K^{(4)}$, where $K^{(4)}_{ij} = m$ if alters i and j co-reply to m posts on Twitter.

Re-Tweet. This view characterizes the re-tweet frequency between alters by a similarity matrix $K^{(5)}$, where $K^{(5)}_{ij} = m$ if alters i and j retweet each other for m times in total.

Topic. This view characterizes the post similarity between alters by a similarity matrix $K^{(6)}$, where $K^{(6)}_{ij}$ is the cosine similarity between the normalized topic vectors of alter i and alter j. These vectors are obtained by first getting a topic vector for each alter by uploading his/her posts to the online annotation tool TagMe [18], and then normalizing the returned vectors by the TF-IDF technique.

3.3 Cluster Assumption

Our cluster assumption is similar to [40] but extended from its single view setting to a multi-view setting. Specifically, we assume alters in the same social circles should have high similarity (as compared with alters in different circles) from multiple views. This means within-circle alters are more likely to be friends, to share more common friends, to retweet or reply to each other more often, to co-reply to more posts and to post more similarity tweets.

4 Clustering on Multi-View Ego-Net

Based on the multi-view ego-net structure presented in the previous section, we propose to detect social circles by multi-view spectral clustering techniques (e.g., [29], [31]), which have been shown effective in clustering multi-view graphs.[3] Specifically, we employ co-trained spectral clustering [29], which is briefly reviewed in Section 4.1.

During investigation, however, we noticed some views of the crawled ego-nets were extremely sparse, picturing very scarce connections between alters. This had raised our attention, because if we simply took these views as what they appeared and threw them into a standard multi-view clustering algorithm that fully transfers information across views, we may end up finding too few circles (to be useful) as strong connections in dense views may be largely suppressed by weak connections in sparse views. This has motivated us to consider a variant of the standard technique which could *selectively* transfer information across views, as we develop and present in Section 4.2.

What makes things more interesting is we later realized the above problem might be much deeper than it appeared – if an ego-net is indeed intrinsically sparse, there should be nothing wrong with detecting few social circles. But how to explain that our modified technique (which assumes graphs are not sparse but inherently incomplete) did show performance improvement over the standard technique (which assumes graphs are sparse by nature)? This has motivated us to probe a deeper question: is the ego-net truly sparse by nature, or is it just inherently incomplete[4] which induces its sparseness? and how bad could it be when one clusters a graph while ignoring its inherent incompleteness (as done by most studies)? We discuss these issues in Section 4.3.

The discussions in this section involve a number of new notations. In Table 1, we summarize the major notations for an arbitrarily fixed view.

4.1 Co-Trained Spectral Clustering: A Brief Review

This section briefly reviews *co-trained spectral clustering* [29], a popular multi-view spectral clustering technique we propose to apply for social circle detection.

Spectral clustering [45] is a classic technique to group nodes of a graph based *solely* on the graph topology. The topology is usually characterized by a node similarity matrix, from which a graph Laplacian matrix is constructed. It is shown eigen-vectors of this Laplacian matrix contain discriminative information for node clustering, and spectral clustering uses these vectors as latent node features on which standard attribute-based clustering techniques such as K-means are performed to group nodes.

In general, multi-view spectral clustering is an extension of the classic spectral clustering from the single-view setting to a multi-view setting, where clustering information in one view is used to modify the clustering tasks in other views so that different views would reach some consistency in results. The co-trained spectral clustering technique [29] alternately uses eigen-vectors of an examined view to refine similarity matrices of other views, by first projecting and then reconstructing those matrices in a new space spanned by eigen-vectors of the examined view. Take view t for example, $U^{(t)}_{:[k]}$ be a matrix whose columns are the k principal eigen-vectors of the normalized Laplacian matrix of $K^{(t)}$. Then, the similarity matrix of another view t' is refined by

$$K^{(t')} = U^{(t)}_{:[k]} \left(U^{(t)}_{:[k]}\right)^T K^{(t')}. \quad (1)$$

Authors showed (1) could encourage consistent clustering across views, by throwing away grouping information within each cluster in each view. Finally, when the alternate

3. In this paper, a network (structure) is viewed as a graph and these two terms are used interchangeably.

4. By 'inherently incomplete', we mean one does not really know whether a graph is incomplete or not, nor does he know which part is incomplete (as assumed known in most studies on incomplete graphs).

TABLE 1
Major Notations (of an Arbitrary View)

Notation	Interpretation
K	similarity matrix (of an arbitrarily fixed view)
D	diagonal matrix with $D_{ii} = \sum_j K_{ij}$
σ_k	k_{th} principal eigenvalue of D
L	normalized Laplacian matrix of K by (6)
λ_k	k_{th} principal eigenvalue of L
U	matrix with each column an eigenvector of L
$\mathbb{P}$	orthogonal projection onto U's range space of
C	matrix of a clustering result, with $C_{ij} = 1$ if alters i, j are grouped and $C_{ij} = 0$ otherwise
$\mathcal{V}$	matrix of a clustering result, with $\mathcal{V}_{ik} = 1$ if alter i is assigned to cluster k and $\mathcal{V}_{ik} = 0$ otherwise

update converges, eigen-vectors of all views (or, some dominant view) are concatenated to form latent node features on which standard K-means is performed to group nodes.

4.2 Selective Co-Trained Spectral Clustering

While the standard co-trained spectral clustering has been shown effective for graph clustering, it ignores the inherent incompleteness of different ego-net views. This means results on sparse views may not be very reliable (in a sense that alters assigned to different groups may not be truly distance in relations), and fully transferring them to other views may mislead their clustering performance. Since we do not know whether a sparse view is incomplete or not, a safe strategy is to *only* transfer its assignments on pairs of alters whose connections are observed.

In this section, we present a heuristic which modifies co-trained spectral clustering so that clustering results in sparse views are selectively transferred to refine other views. The heuristic is twofold: a view is considered incomplete if its fraction of observed connections is below some threshold, and (in an incomplete view) alters are considered to have observed connections if their similarities are non-zero.

The algorithm of our proposed technique is presented in Algorithm 1, where the involved functions are defined as:

- $eig(K, k)$ returns an n-by-k matrix whose columns are the k principal eigen-vectors of the normalized Laplacian of the (similarity) matrix K.
- $clust(U, k)$ returns an n-by-n matrix C obtained by performing k-means clustering on sample matrix U (where each row is one example), such that C_{ij}=1 if examples i and j are assigned to the same group and C_{ij}=-1 otherwise.
- $update(t)$ is defined for view t as

$$update(t) = \exp \sum_{t' \in [T], t' \neq t} C^{(t')} \circ (\mathbf{1}\{K^{(t')} \neq \mathbf{0}\})^{\delta_{t'}}, \qquad (2)$$

where $C^{(t')}$ is the output matrix of $clust(U^{(t')}, k)$, $\mathbf{0}$ is a matrix of zeros same sized as $K^{(t')}$, $\mathbf{1}$ is an element-wise indicator function, and $\delta_{t'}$ is a binary function outputting 1 if $K^{(t')}$ is sufficiently sparse (i.e., its fraction of observed entries is below some threshold) and 0 otherwise.

The core of our modified algorithm is $update(t)$, which is used to update the similarity matrix of view t based on information selectively transferred from other views. We slightly elaborate its design in the following, assuming the case of two-view clustering:

Algorithm 1. Selective Co-Trained Spectral Clustering

Input: Similarity matrices of T views $K_1, K_2, \ldots, K_T$
Initialize: $\forall \in [T]$, $U^{(t)} = eig(K^{(t)}, k)$, $C^{(t)} = clust(U^{(t)}, k)$
for $i = 1$ **to** $rounds$ **do**
  for $t \in [T]$ **do**
    2: Refine similarity matrix $K^{(t)} = update(t) \circ K^{(t)}$
    3: Update $U^{(t)} = eig(K^{(t)}, k)$
    4: Update $C^{(t)} = clust(U^{(t)}, k)$
  end for
end for
Output: apply K-means on the concatenated feature matrix $U = [U^{(k_1)}, \ldots, U^{(k_\ell)}]$ of dominant views $k_1, \ldots, k_\ell$.

For $C^{(t')}$. if alters i, j are assigned to the same group in view t', we have $C^{(t')}_{ij} = 1$. This *could* result in $update(t)_{ij} > 1$ and consequently the increase of these two alters' similarity in view t through $K^{(t)}_{ij} = K^{(t)}_{ij} \cdot update(t)_{ij}$. Note, however, $C^{(t')}_{ij} = 1$ does not guarantee the increase of $K^{(t)}_{ij}$, because in order to get $update(t)_{ij} > 1$, we also need the indicator function to output 1 if view t is incomplete (i.e., the connection between two alters needs to be observed in view t).

For $K^{(t')}$. if alters i and j have observed connection, we have $K^{(t')}_{ij} \neq 0$ and thus $(\mathbf{1}\{K^{(t')} \neq \mathbf{0}\})_{ij} = 1$. This *could* allow their clustering result $C^{(t')}_{ij}$ be transferred to other views through $update(t)$ (specifically, $C^{(t')}_{ij} \cdot (\mathbf{1}\{K^{(t')} \neq \mathbf{0}\})^{\delta_{t'}}_{ij}$). Of course, the whole selective mechanism is valid only when view t' is considered incomplete (i.e., $\delta_{t'} = 1$).

For $\delta_{t'}$. if view t' is considered incomplete, we have $\delta_{t'} = 1$. This will activate the selective mechanism for $K^{(t')}$ (as we described above); otherwise, the selective mechanism is de-activated and all clustering results in view t' will be transferred to modify view t through $update(t)$.

The following example demonstrates how the proposed algorithm may leverage the clustering result in one view to refine another other view. Suppose the result in view one is

$$C^{(1)} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}, \qquad (3)$$

which indicates alters 1 and 3 are grouped whereas 2 and 3 are separated. Suppose the similarity matrix of view two is

$$K^{(2)} = \begin{bmatrix} 1 & 0.4 & 0.1 \\ 0.4 & 1 & 0.6 \\ 0.1 & 0.6 & 1 \end{bmatrix}. \qquad (4)$$

The algorithm will refine $K^{(2)}$ using $C^{(1)}$ through $update(1)$ and, when results are fully transferred, we have an update

$$K^{(2)} = \begin{bmatrix} 2.7 & 0.2 & 0.3 \\ 0.2 & 2.7 & 0.2 \\ 0.3 & 0.2 & 2.7 \end{bmatrix}. \qquad (5)$$

In the updated $K^{(2)}$, it is clear similarities between alters grouped in view one are increased and vice versa. Note, however, that grouping in view one does not necessarily lead to grouping in view two – for instance, alters 1 and 3 still have low similarity even though they were grouped in one view—the updated matrix is a compromise between results in other views and observations in the current view. When

results are selectively transferred, we expect the algorithm would converge faster as less consistency needs to be compromised between views. This seems indeed the case as is evident from our experimental studies. Finally, note since $C^{(1)}$ is symmetric, $K^{(2)}$ would remain symmetric after update (and thus its corresponding Laplacian matrix remains positive semi-definite which admits positive eigenvalues).

4.3 When Ego-Net is Inherently Incomplete

As we mentioned before, a sparse ego-net may be inherently incomplete. Such incompleteness distinguishes itself from most previous studies on incomplete graphs which assume prior knowledge on whether the graph is indeed incomplete and which part of the graph is incomplete. None of these is known, however, for an inherently incomplete graph. Then, what could we say about clustering such a graph?

This section is an attempt to answer the above question through the derivation and discussion of an upper bound on the possible performance bias when one performs standard clustering on an inherently incomplete graph while ignoring its potential incompleteness (as most studies do). To derive the bound, we integrate a classic spectral clustering theory [26] with a recent result in matrix perturbation theory [66], and employ several properties of the Loewner partial orders (e.g., [23, Chapter 7]). We then discuss the implications of our bound, with a focus on how it could be affected by various ego-net characteristics; the implications seem supported in later simulations. Our discussion will focus on single-view clustering as it is the backbone of multi-view clustering techniques (e.g., co-trained spectral clustering could be regarded as single-view spectral clustering on a dominant view which has been refined by other views).

4.3.1 Preliminaries

First, we make a *free-approximation* assumption to simplify discussion. It is well known that spectral clustering is an approximated solution to the optimal normalized cut problem, and there is a rich literature studying the approximation error (e.g., [51], [67]). Although we apply spectral clustering and evaluate results under the optimal cut framework, such approximation is not our focus. We thus assume the approximation error is zero, which could be satisfied if the k principal eigen-vectors of the graph Laplacian matrix are piece-wise constant with respect to the optimal normalized cut result on the graph [26]; when the assumption is not satisfied, our analysis could be generalized by simply adding an error term for the approximation.

Next, recall we have an ego-net consisting of n alters and characterized by multiple n-by-n similarities matrices $\{K^{(t)}\}$, each representing one view of the ego-net structure. Since all analysis in this section applies to an arbitrary single view, the superscript t (i.e., view index) will be omitted.

Consider the task of k-partitioning the n alters based on a *complete* ego-net characterized by similarity matrix K. Let

$$L = D^{-1/2}KD^{-1/2}, \tag{6}$$

be the normalized Laplacian matrix,[5] where D is an n-by-n diagonal matrix where $D_{ii} = \sum_{j\in[n]} K_{ij}$. Let σ_k and λ_k be the k_{th} principal eigenvalues of D and L respectively.

5. A classic definition is $L = I - D^{-1/2}KD^{-1/2}$. Ours is from [45], which admits the same eigen-vectors but facilitates discussion.

Let $\tilde{K}$ denote an inherently incomplete observation of K, with observed entries indexed by set Ω, such that $\tilde{K}_{ij} = K_{ij}$ if $(i,j) \in \Omega$ and $\tilde{K}_{ij} = 0$ otherwise. Note $\tilde{K}_{ij} = 0$ may imply K_{ij} is unobserved or K_{ij} is observed but has value 0. Similar to L, let $\tilde{L} = \tilde{D}^{-1/2}\tilde{K}\tilde{D}^{-1/2}$ be the normalized Laplacian of $\tilde{K}$, where $\tilde{D}$ is diagonal matrix with $\tilde{D}_{ii} = \sum_{j\in[n]} \tilde{K}_{ij}$.

The resulted k-partition of n alters based on K is represented as an n-by-k matrix $\mathcal{V}$, defined as $\mathcal{V}_{ij} = 1$ if alter i is assigned to cluster j and $\mathcal{V}_{ij} = 0$ otherwise. Similarly, the partition result based on $\tilde{K}$ is represented as matrix $\tilde{\mathcal{V}}$. We then evaluate the difference between these two results using the metric employed in [26, Formula (2)], i.e.,

$$d(\mathcal{V}, \tilde{\mathcal{V}}) = \frac{1}{2}\left\| \sum_{j\in[k]} \frac{\mathcal{V}_{:j}(\mathcal{V}_{:j})^T}{(\mathcal{V}_{:j})^T\mathcal{V}_{:j}} - \sum_{j'\in[k]} \frac{\tilde{\mathcal{V}}_{:j'}(\tilde{\mathcal{V}}_{:j'})^T}{(\tilde{\mathcal{V}}_{:j'})^T\tilde{\mathcal{V}}_{:j'}} \right\|_F^2. \tag{7}$$

Intuitively, metric (7) counts the pairs of alters assigned to different clusters in two results, each weighted by the corresponding cluster size. Indeed, note $(\mathcal{V}_{:j}(\mathcal{V}_{:j})^T)_{i_1 i_2}$ equals 1 if alters i_1 and i_2 are both assigned to cluster j and equals 0 otherwise; and $(\mathcal{V}_{:j})^T\mathcal{V}_j$ is the size of cluster j. Also note the metric is bounded when each cluster contains at least one alter, which could be easily guaranteed by proper algorithm design. Finally, it is generally hard to give a threshold under which a bias $d(\mathcal{V}, \tilde{\mathcal{V}})$ could be considered acceptable, as it depends on the cluster sizes and the applications. Nevertheless, one could get more sense through simple calculations: for example, suppose n nodes are equally partitioned into k clusters, and results based on complete and incomplete graphs differ on p fraction of node pairs, then $d(\mathcal{V}, \tilde{\mathcal{V}}) = k^2 p$.

4.3.2 A Bias Bound and Its Implications

Based on notations introduced in the previous section, our derived bias bound is stated as follows.

Proposition 1. *Let $\mathcal{V}, \tilde{\mathcal{V}}$ be the k-partitioning result matrices of the optimal normalized cuts based on $K, \tilde{K}$ respectively. Let $\tilde{\sigma}_1$ be the principal eigenvalue of $D - (D\tilde{D})^{1/2}$ and denote $\tilde{\Delta} = L - \tilde{L}$. Then*

$$d(\mathcal{V}, \tilde{\mathcal{V}}) \le \frac{\sigma_1}{\sigma_n}\left(\frac{\tilde{\sigma}_1}{\sigma_n} + \frac{2\min\{\sqrt{k}||\tilde{\Delta}||_2, ||\tilde{\Delta}||_F\}}{\lambda_k - \lambda_{k+1}}\right). \tag{8}$$

The bound has several implications.

First, note $\sigma_1, \sigma_n, \lambda_k, \lambda_{k+1}$ are constants when an (underlying and complete) ego-net is given. Their impacts on the bias bound could be interpreted as follows. For σ_1 and σ_n, note they respectively describe the overall behaviors of the most and least active alters[6] in the ego-net, since they are respectively the largest and smallest row sums of K. Since $\sigma_1/\sigma_n \ge 1$ in (8), we see standard spectral clustering (which ignores the potential graph incompleteness) may suffer less performance bias if alters are equally active in the ego-net (in which case σ_1/σ_n would be close to 1).

Second, the bound suggests ways of choosing k to lower the risk of performance bias. Based on the term $\sqrt{k}$ in (8), we see detecting fewer social circles could generally reduce

6. The notion of 'active' is open to interpretation: in an interaction view, for example, an active alter is someone who interacts frequently with others; while in a relation view, an active alter is someone who has a lot of friends (an indicator of his active socialization).

the risk of suffering performance bias; based on $\lambda_k - \lambda_{k+1}$, on the other hand, we see one may choose k at the 'steepest' place of the graph spectrum. For instance, when alters have equally active (normalized) behaviors, the graph spectrum is flat and we may choose a large k to maximize $\lambda_k - \lambda_{k+1}$.

Finally, the bound sheds light on how bias may decrease as more connections are observed on the ego-net: as observations increase, it is clear that $\tilde{K} \to K$[7] and thus $\tilde{D} \to D$ and $\tilde{L} \to L$. The latter further implies $\tilde{\sigma}_1 \to 0$ and $\Delta \to 0$, which based on (8) implies a small bias bound. In particular, if the ego-net is fully observed, we have $\tilde{\sigma}_1 = 0$ and $\Delta = 0$, which results in a zero bias bound and hence no performance bias.

4.3.3 Proof of Proposition 1

Notations. First, recall K is an n-by-n similarity matrix with an associated Laplacian L and a diagonal D, and λ_k, σ_k are the k_{th} principal eigenvalues of L, D respectively. The k-partitioning result is stored in an n-by-k matrix $\mathcal{V}$.

Let $L = U\Lambda U^T$ be the eigen-decomposition of L such that Λ is an n-by-n diagonal matrix with Λ_{ii}=λ_i ($\lambda_1 \geq \lambda_2 \geq ...$) and U is an n-by-n unitary matrix where $U_{:i}$ is the eigenvector for λ_i. Let $U_{[k]}$ be an n-by-k sub-matrix of U where $(U_{[k]})_{:i} = U_{:i}$ for $i = 1, \ldots, k$. Since $U_{[k]}$ is orthonormal, we have $\mathbb{P}_k = U_{[k]}U_{[k]}^T$ as the *orthogonal projection* onto the range space of $U_{[k]}$ (e.g.,[22, Chapter 2]).

All the above notations apply to $\tilde{K}$ and its associated variables, yet capped with notation '$\sim$'. For instance, $\tilde{\mathbb{P}}_k$ is the orthogonal projection onto the range space of $\tilde{U}_{[k]}$, which contains the k principal eigenvectors of the Laplacian $\tilde{L}$.

Finally, let $\succeq, \succ$ be the Loewner partial orders such that $A \succeq B$ if $A - B$ is positive semi-definite (PSD) and $A \succ B$ if $A - B$ is positive definite. Note $A \succeq 0$ implies A is PSD.

Proof Sketch. The strategy of our proof is as follows: we first bound $d(\mathcal{V}, \tilde{\mathcal{V}})$ by two new terms using the triangular inequality; then we bound the first term using a recent result in perturbation theory [66], and bound the second term using several Loewner partial order properties (e.g., [23, Chapter 7]); we also borrow some results from [26]. □

Step 1: bound $d(\mathcal{V}, \tilde{V})$. By [2, Formula (3)] we have

$$d(\mathcal{V}, \tilde{\mathcal{V}}) \leq \frac{\sigma_1}{\sigma_n} \cdot d(D^{1/2}\mathcal{V}, D^{1/2}\tilde{\mathcal{V}}). \tag{9}$$

Further, by triangular inequality it follows

$$d(D^{1/2}\mathcal{V}, D^{1/2}\tilde{\mathcal{V}}) \leq d_w(\mathcal{V}, \tilde{\mathcal{V}}) + d(\tilde{D}^{1/2}\tilde{\mathcal{V}}, D^{1/2}\tilde{\mathcal{V}}), \tag{10}$$

where we define $d_w(\mathcal{V}, \tilde{\mathcal{V}}) = d(D^{1/2}\mathcal{V}, \tilde{D}^{1/2}\tilde{\mathcal{V}})$.

Step 2: bound $d_w(\mathcal{V}, \tilde{\mathcal{V}})$. It is known that spectral clustering is a relaxation of the optimal normalized cut problem. In [2, Formula (1)], the relaxation error is measured by the difference between the orthogonal projections for the two problems. In our context, these two projections (associated with matrix K) are $\mathbb{P}_k$ and

$$\Pi_k := \sum\nolimits_{j\in[k]} D^{1/2}\mathcal{V}_j\mathcal{V}_j^T D^{1/2}/(\mathcal{V}_j^T D\mathcal{V}_j). \tag{11}$$

Then, our free-approximation assumption implies $\mathbb{P}_k = \Pi_k$. Similarly, the orthogonal projections associated with $\tilde{K}$ are $\tilde{\mathbb{P}}_k$ and $\tilde{\Pi}_k := \sum_{j\in[k]} \tilde{D}^{1/2}\tilde{\mathcal{V}}_j\tilde{\mathcal{V}}_j^T\tilde{D}^{1/2}/(\tilde{\mathcal{V}}_j^T\tilde{D}\tilde{\mathcal{V}}_j)$, and that $\tilde{\mathbb{P}}_k = \tilde{\Pi}_k$. Since by definition $d_w(\mathcal{V}, \tilde{\mathcal{V}}) = ||\Pi_k - \tilde{\Pi}_k||_F^2$, we have

$$d_w(\mathcal{V}, \tilde{\mathcal{V}}) = ||\mathbb{P}_k - \tilde{\mathbb{P}}_k||_F^2. \tag{12}$$

Now, our task becomes bounding $||\mathbb{P}_k - \tilde{\mathbb{P}}_k||_F^2$ instead. A classic technique is the *Davis-Kahan* theorem (e.g.,[66, Theorem 1]), which would give

$$||\mathbb{P}_k - \tilde{\mathbb{P}}_k||_F \leq \frac{||\Delta||_F}{\kappa}, \tag{13}$$

where $\kappa = \inf\{|\lambda_i - \tilde{\lambda}_j|; 1 \leq i \leq k, k < j \leq n\}$[8]. While this bound is seminal, it contains an implicit dependency on Δ (through parameter $\tilde{\lambda}$), whereas we prefer a bound that has more explicit dependency on Δ (for easier interpretation). To this end, we employ a recent generalization of the Davis-Kahan theorem, which is stated as follows.

Theorem 1 ([66, Theorem 2]). *Let $L, \tilde{L} \in \mathbb{R}^{n\times n}$ be two symmetric matrices with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ and $\tilde{\lambda}_1 \geq \ldots \geq \tilde{\lambda}_n$ respectively. Fix $1 \leq r \leq s \leq n$ and assume $\min\{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}\} > 0$, where $\lambda_0 := \infty$ and $\lambda_{n+1} := -\infty$. Let $d := s - r + 1$, and let $U_k = [u_r, u_{r+1}, \ldots, u_s] \in \mathbb{R}^{n\times d}$ and $\tilde{U}_d = [\tilde{u}_r, \tilde{u}_{r+1}, \ldots, \tilde{u}_s] \in \mathbb{R}^{n\times k}$ have orthonormal columns satisfying $Lu_j = \lambda_j u_j$ and $\tilde{L}\tilde{u}_j = \lambda_j\tilde{u}_j$ for $j = r, r+1, \ldots, s$. Then*

$$||\mathbb{P}_k - \tilde{\mathbb{P}}_k||_F \leq \frac{2\min(\sqrt{d}||\Delta||_2, ||\Delta||_F)}{\min\{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}\}}. \tag{14}$$

By (12) and Theorem 1 (with $r = 1$ and $s = k$), we have

$$d_w(\mathcal{V}, \tilde{\mathcal{V}}) \leq \frac{4\min(\sqrt{k}||\Delta||_2^2, ||\Delta||_F^2)}{(\lambda_k - \lambda_{k+1})^2}. \tag{15}$$

Step 3: bound $d_+ := d(\tilde{D}^{1/2}\tilde{\mathcal{V}}, D^{1/2}\tilde{\mathcal{V}})$. First, it is easy to verify the following lemma by algebraic arguments.

Lemma 1. *For any orthogonal matrix $M \in \mathbb{R}^{n\times p}$,*

$$\sum\nolimits_{j\in[p]} M_{:j}M_{:j}^T/(M_{:j}^T M_{:j}) = M(M^T M)^{-1}M^T. \tag{16}$$

Note $D^{1/2}\tilde{\mathcal{V}}$ and $\tilde{D}^{1/2}\tilde{\mathcal{V}}$ are orthogonal. Then, jointly applying Lemma 1 and the alternative expression in [2, Page 6] for $||M(M^TM)^{-1}M^T - \tilde{M}(\tilde{M}^T\tilde{M})^{-1}\tilde{M}^T||_F^2$, we have

$$d_+ = \mathrm{tr}\{\mathcal{T}^{-1/2}(\mathcal{T} - \tilde{\mathcal{N}})\mathcal{T}^{-1/2}\}, \tag{17}$$

where $\mathcal{T} = \tilde{\mathcal{V}}^T D\tilde{\mathcal{V}}$ and

$$\tilde{\mathcal{N}} = \tilde{\mathcal{V}}^T(D\tilde{D})^{1/2}\tilde{\mathcal{V}}(\tilde{\mathcal{V}}^T\tilde{D}\tilde{\mathcal{V}})^{-1}\tilde{\mathcal{V}}^T(\tilde{D}D)^{1/2}\tilde{\mathcal{V}}. \tag{18}$$

It the sequel, we bound $\mathcal{T}^{-1/2}$ and $\mathcal{T} - \tilde{\mathcal{N}}$ separately.

Note $\tilde{\mathcal{V}}$ and $D^{1/2}\tilde{\mathcal{V}}$ have linearly independent columns (since $\tilde{\mathcal{V}}$ indicates a partition of alters and $D^{1/2}$ does not change such indication). Then, by [23, Theorem 7.2.10]

$$\mathcal{T} = \tilde{\mathcal{V}}^T D\tilde{\mathcal{V}} \succ 0 \quad \text{and} \quad \tilde{\mathcal{V}}^T\tilde{\mathcal{V}} \succ 0. \tag{19}$$

Further, since σ_n is the smallest diagonal entry of D, it is easy to verify $D \succeq \sigma_n I$. This implies

7. Notation $A \to B$ means $A - B$ approaches zero matrix.

8. The original theorem bounds the angles between subspaces, which equals the difference of their orthogonal projectors, e.g., [15, Page 10].

$$\mathcal{T} \succeq \sigma_n \tilde{\mathcal{V}}^T \tilde{\mathcal{V}}. \tag{20}$$

Based on (19) and (20), by [23, Corollary 7.7.4] we have

$$\sigma_n^{-1/2} (\tilde{\mathcal{V}}^T \tilde{\mathcal{V}})^{-1/2} \succeq \mathcal{T}^{-1/2}. \tag{21}$$

To bound $\mathcal{T} - \tilde{\mathcal{N}}$, notice $(D\tilde{D})^{1/2} \succeq \tilde{D}$, which implies

$$(\tilde{\mathcal{V}}^T \tilde{D} \tilde{\mathcal{V}})^{-1} \succeq (\tilde{\mathcal{V}}^T (D\tilde{D})^{1/2} \tilde{\mathcal{V}})^{-1}. \tag{22}$$

Plugging (22) in (18), we have

$$\tilde{\mathcal{N}} \succeq \tilde{\mathcal{V}}^T (D\tilde{D})^{1/2} \tilde{\mathcal{V}}. \tag{23}$$

Combining (23) and (19), we have

$$\tilde{\sigma}_1 \tilde{\mathcal{V}}^T \tilde{\mathcal{V}} \succeq \tilde{\mathcal{V}}^T (D - (D\tilde{D})^{1/2}) \tilde{\mathcal{V}} \succeq \mathcal{T} - \tilde{\mathcal{N}}, \tag{24}$$

where $\tilde{\sigma}_1$ is the largest diagonal entry of $D - (D\tilde{D})^{1/2}$. Further combining (21) and (24) gives

$$\sigma_n^{-1} \tilde{\sigma}_1 I \succeq \mathcal{T}^{-1/2} (\mathcal{T} - \tilde{\mathcal{N}}) \mathcal{T}^{-1/2}, \tag{25}$$

and taking trace on both sides yields

$$\sigma_n^{-1} \tilde{\sigma}_1 \geq d_+. \tag{26}$$

Step 4. Combining ((9), (10), (15), (26)) proves the proposition.

5 Experimental Study

In this section, we perform extensive experimental evaluations, with a focus on examining our set of hypotheses that better social circles in ego-nets could be detected (1) based on multiple views of the ego-net structure (as opposed to current studies that consider only a single view), and (2) by applying multi-view spectral clustering (as opposed to single-view clustering), and (3) by selectively transferring information from sparse views to others in multi-view clustering (as opposed to standard multi-view techniques that fully transfer those information). Implications obtained from the bias bound are also examined.

5.1 Data Preparation

We will experiment on Twitter, which is one of the most popular online social network platforms. To this end, we first implemented a crawler to collect Twitter data using its API, which can return any user's profile, follower/following lists and tweets. The user profile consists of user name, screen name, user id, profile create time, description (a personal statement), location and time zone. The tweets information consists of tweet id, post time, tweet location, in-reply-to user id, in-reply-to status id, list of re-tweets (user id and tweet id) and tweet content. For each user, we only collected his/her most recent 2,000 tweets due to many constraints. It is also noted not all the attributes are available and accurate for all users—user location in user profiles is self-generated textual description, where we have seen "Worldwide" and "Coming Soon Everywhere" etc; meanwhile, tweet locations are accurate latitudes and longitudes, but they are missing from most of the tweets; besides, Twitter has enforced mandatory limits on the crawling rate, especially for crawling account-specific information. Finally, we have collected 92 data sets—92 seed users and all their friends. In our data set, each seed user has 245 friends on average. In total, we have collected information of more than 22 K users, with approximately 3 million friendship links, and more than 27 million tweet messages. It should be mentioned the seed users were collected following the standard random walk sampling technique, which has demonstrated both efficiency and effectiveness in analyzing online social networks [38], [46]. (We notice this technique also falls into the category of snowball sampling, which has been reported as one major sampling approach on Twitter [19].)

A crawled Twitter ego-net structure was modeled by six views, as introduced in Section 3.2. Specifically, each view was implemented as follows: for the *friendship* view, two alters were marked as friend if each is in both the follower and following lists of the other, and the *common friend* view counted the number of such friends shared by alters; for the *reply* view, the reply number from alter Nancy to alter Bob was obtained by scanning through Bob's tweets and counting the replies from Nancy, and vice versa; for the *co-reply* view, the co-reply number of two alters was obtained by scanning through all tweets in the crawled ego-net and counting those they both replied; for the *re-tweet* view, the retweet number from alter Nancy by alter Bob was obtained by scanning through Nancy's tweets and counting those retweeted by Bob; for the *content* view, we first obtained a topic vector for each alter by uploading her tweets and profile to TagMe [18], removed those returned topics whose relevance scores are below 0.2 (between $[0, 1]$) by the Pareto principle, and normalized all topic vectors by TF-IDF.[9]

5.2 Experimented Techniques

In experiment, we examined the performance of four clustering techniques that rely *only* on the ego-net structure.

SCAN [63] is a classic and popular clustering technique designed to detect social circles based solely on the friendship view of the network structure. We employed it as a representative single-view clustering technique for social circle detection in experiments.

Spectral Clustering (sc) [45] is a classic single-view clustering technique which groups instances based solely on their similarities. Although spectral clustering has not been specifically applied for social circle detection, we employed it in experiments as another representative of the single-view clustering techniques. Specifically, we first separately applied this technique on dominant views to learn their latent feature matrices $U^{(t)}$'s (i.e., eigenvectors of the normalized Laplacian matrix of each view), and then concatenated these matrices in a column-wise manner to form an integrated latent feature matrix on which standard k-means clustering was applied to obtain the final grouping result. This approach could be interpreted as the standard multi-view spectral clustering but without cross-view information transfer, a common design to evaluate the effectiveness of multi-view learning techniques.

Co-Trained Spectral Clustering (CSC). [29] is a popular multi-view spectral clustering technique which we have reviewed in Section 4.1 and employed in experiments as a

9. After a similarity matrix is obtained, we normalize it into $[0, 1]$ by dividing all entries by the maximum entry and fix diagonal entries to 1, indicating self-similarity is always the largest. Note when the maximum entry is zero, we do not perform normalization. (This, however, rarely occurred in experiments.) Also note the normalization may slightly change the interpretation of a similarity matrix, but should ideally not affect the clustering result based on it.

representative of the multi-view clustering technique. Similar to the previous application of (single-view) spectral clustering, when the co-trained spectral clustering algorithm converged, we concatenated the obtained latent feature matrices of dominant views to form an integrated feature matrix, on which K-means was performed to cluster alters.

Selective Co-Trained Spectral Clustering (scsc). is the modified multi-view spectral clustering algorithm we proposed in Algorithm 1, which selectively transfers clustering results in sparse views to refine other views.

It should be mentioned we considered the *friendship*, *common friend* and *topic* as three dominant views, not only because they were generally denser (hence more complete) but also because they demonstrated stable and better performance in experiments. Note, however, although the non-dominant views were not directly used (as part of the integrated latent feature matrices) for clustering, they were helpful in that their information had been transferred to the dominant views by the multi-view algorithms. In addition, we fixed $k = 5$ for examined techniques (except SCAN which automatically determines k), since we had observed similar trends in their performance as k increased from 3 to 10. Hyper-parameters of SCAN were set as default.

5.3 Evaluation 1: Cluster Compactness

We first evaluated the quality of detected clusters based on its compactness, which is a most common measurement.

5.3.1 Evaluation Metric

In our problem, the unavailability of both cluster ground truth and alter feature matrix has precluded the use of most standard cluster evaluation metrics, including both external ones such as the random index and F-measure and internal ones such as the Davies-Bouldin index and Dunn index. We hence presented and used the following metric to evaluate the compactness of a clustering result (recall in Section 4.2 we introduced an n-by-n indicator matrix C to represent the result such that $C_{ij} = 1$ if alters i and j are grouped and $C_{ij} = 0$ otherwise):

$$\gamma = \Big(\sum\nolimits_t S_w^{(t)}\Big) / \Big(\sum\nolimits_t S_b^{(t)}\Big), \tag{27}$$

where $t \in [T]$ (as there are T views in total),

$$S_w^{(t)} = \frac{\sum_{(i,j)} K_{ij}^{(t)} \cdot 1\{C_{ij} > 0\}}{\sum_{(i,j)} 1\{C_{ij} > 0\}}, \tag{28}$$

and

$$S_b^{(t)} = \frac{\sum_{(i,j)} K_{ij}^{(t)} \cdot 1\{C_{ij} < 0\}}{\sum_{(i,j)} 1\{C_{ij} < 0\}}. \tag{29}$$

where $(i, j) \in [n] \times [n]$ (as there are n alters in the ego-net).

Taking spirit from the classic discriminant analysis (e.g., [4, Formula (4)]), we name (27) the *total similarity ratio*, where $S_w^{(t)}$ is the *within-circle similarity* that measures the average similarity between clustered alters in view t and $S_b^{(t)}$ is the *between-circle similarity* that measures the average similarity between separate alters in view t. It is clear a compact set of clusters should have high within-cluster similarity yet low between-cluster similarity (thus a large total similarity ratio), consistent with a common argument that alters within a circle should have high similarities and vice versa. Finally, to evaluate results for a single view, we use metric

$$\gamma^{(t)} = S_w^{(t)} / S_b^{(t)}; \tag{30}$$

and when $S_b^{(t)} = 0$ (as is often the case when view t is very sparse), we add a small constant to it in the metric.

While by design our metric is an echo of the Fisher ratio in classic discriminant analysis, it finds itself connected to several cluster quality metrics in the literature. For instance, $S_w^{(t)}$ and $S_b^{(t)}$ could be interpreted as the *homogeneity index* and *separation index* in [50] respectively, except we directly have alter similarities instead of computing them using alter fingerprints; their ratio is also related to the *weighted inter-intra index* [56] and *Calinski-Harabasz index* [6], in a sense that for equally sized clusters $S_w^{(t)}/S_b^{(t)}$ differs from these indices mainly by constants (depending on the cluster size and sample size). These connections could be easily verified, and would largely remain valid for clusters of different sizes (e.g., by relaxing it to the case of equally-sized clusters).

5.3.2 Results and Discussions

To better understand the performance of spectral clustering techniques, we first experimented on one single ego-net, which contains 386 alters. Recall the techniques are single-view spectral cluster (*sc*), co-trained spectral clustering (*csc*) and selective co-trained spectral clustering (*scsc*). For *csc* and *scsc*, we updated their view refinements for 20 rounds (by which both were observed converged generally) and reported the similarity ratios of their clustering results on each view in Fig. 2. We saw *csc* generally improved with more rounds of update, which is consistent with the spirit of co-trained style learning algorithms. However, its convergence rate was slow (as compared with that of *scsc*), and its performance improvements over *sc* were not significant on the topic and reply views and were little on the friendship and co-reply views. As we explained before, this may be due to the ignorance of *csc* on the inherent incompleteness of sparse ego-net views. Comparatively, our proposed *scsc* converged fast (usually within one or two rounds of update) and improved *sc* consistently and significantly on all views.

Next, we examined the sizes of clusters output by different spectral clustering techniques on an ego-net with 102 alters. For *csc* and *scsc*, these sizes were reported at the update rounds where they respectively achieved their best clustering performance (i.e., the highest total similarity ratios). The statistics were summarized in Table 2. It appears *csc* encouraged more balanced clusters, while both *sc* and *scsc* output a big cluster. This may be because *csc* enforced stronger view consistency so that a sparse view, for instance, could require a dense (dominant) view to 'break down' its inherently big clusters. It should be pointed out imbalanced clusters make sense in many practices, e.g., a family circle is usually much smaller than a friend circle.

Then, we examined the cluster qualities of *sc*, *csc* and *scsc* over 92 ego-nets we crawled from Twitter. The total similarity ratio of each technique on each ego-net is shown in Fig. 3. We saw *scsc* consistently outperformed the other two techniques, and standard *csc* had the worst performance. These were consistent with our previous observations in Fig. 2, and our earlier discussions on the limitation of *csc*: as it bindly transfers results on an inherently incomplete view to other views, the clustering tasks on those views may be misled.

Finally, on the same 92 ego-nets we compared the cluster qualities of all examined techniques on the friendship view (as SCAN was designed based specifically on this view).

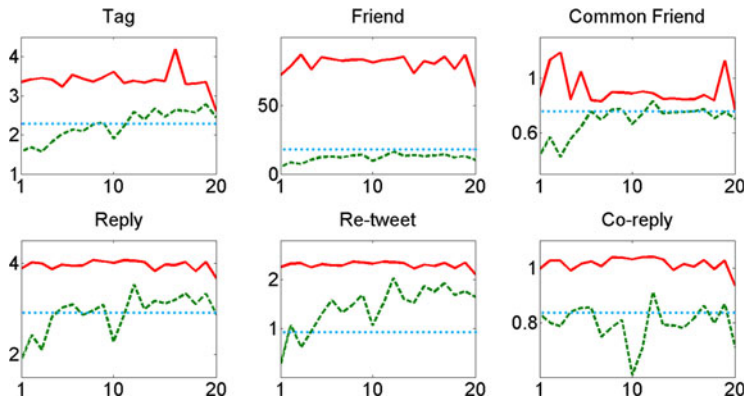


Fig. 2. The similarity ratios of examined spectral clustering techniques on six views of a Twitter ego-net. In each figure, the vertical axis represents the similarity ratio 30 and the horizontal axis represents the round of algorithm updates. The blue and dotted lines represent the performance of sc, the green and dash lines represent that of csc, and the red and solid lines represent that of the proposed $scsc$.

Since SCAN would remove outliers and reported results based solely on the remaining alters, for fair comparison we also reported performance of other techniques based on the same set of remaining alters on each ego-net. These results were shown in Fig. 4. We saw results similar to the previous examination, and that SCAN performed similarly to sc but not as good as $scsc$. This results suggest that neither single-view clustering or multi-view clustering with improper (full) cross-view information transfer is sufficient for detecting high quality clusters.

5.4 Evaluation 2: Quality of Boundary Alters

In this experiment, we manually evaluated the qualities of detected circles, with a focus on the performance of $scsc$ on *boundary alters*, that is, alters that are most distant from the centroids of their assigned clusters. (We focused on these alters because they were most likely to be mis-clustered, and it was too expensive to manually evaluate all alters).

Given a clustering result, our general idea is to let human evaluator (without any information on this result) assign a boundary alter to one of the following two detected circles: (1) its actual assigned circle and (2) its nearest neighbor circle, defined as the circle (not assigned for the alter) whose members have the smallest distance to the alter on average. To be specific, for each boundary alter we first selected 10 tested alters, with 5 randomly from its assigned circle and another 5 randomly from its nearest neighbor circle. Then, a human evaluator would score from 1 to 5 on how much he/she agreed the boundary alter should be clustered with each of these tested alters (based on their profiles and tweets) – 1 for strongly disagree, 2 for somewhat disagree, 3 for neutral, 4 for somewhat agree and 5 for strongly agree—and the two scores averaged over both tested alters from the assigned circle and tested alters from the neighbor circle were separately reported.

TABLE 2
The Size of Five Social Circles Detected by Different Spectral Clustering Techniques

Cluster	1	2	3	4	5	std
SC	8	12	13	25	44	14.6
CSC	16	17	20	24	25	4.04
SCSC	10	10	13	15	54	18.9

std *is the standard deviation of the sizes over five circles.*

We had summarized the above two scores for 60 randomly selected boundary alters, and found the averaged results were 2.63 for alters from the actually assigned circles and 2.52 for alters from the nearest neighbor circles. This suggested the performance of $scsc$ is relatively consistent with human—it managed to assign boundary alters to circles where they had *tighter* connections. The observation that both scores were low, on the other hand, suggested the intrinsic difficulty of social circles detection—it could be circles were generated based on less visible profiles, or they overlapped by nature which confused human evaluator.

5.5 Evaluation 3: Keywords of Detected Clusters

To better interpret the detected social circles, we extracted and examined their related tags (as returned by the TagMe tool) in this experiment. Our examination focused on tags that were *discriminative* across circles and, most importantly, *representative* of the content posted in each circle.

The representative tags for each circle were extracted as follows. Let T_i be the set of tags returned by TagMe for alter i, and $tf(i,t)$ be the frequency of tag t appeared in the posts of alter i. The representativeness of tag t for a circle $\mathcal{S}$ (which is an index set of its assigned alters) was measured by

$$\Pr\{t|\mathcal{S}\} = \frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} \widetilde{tf}(i,t), \tag{31}$$

where

$$\widetilde{tf}(i,t) = tf(i,t)/(\max\{tf(i,t)|t \in T_i\}). \tag{32}$$

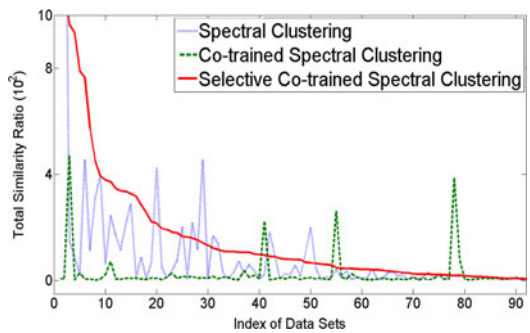


Fig. 3. The total similarity ratios of different spectral clustering techniques on 92 Twitter ego-nets. The ratios averaged over these ego-nets are: 108.8 for sc, 24.8 for csc, and 187.4 for $scsc$.

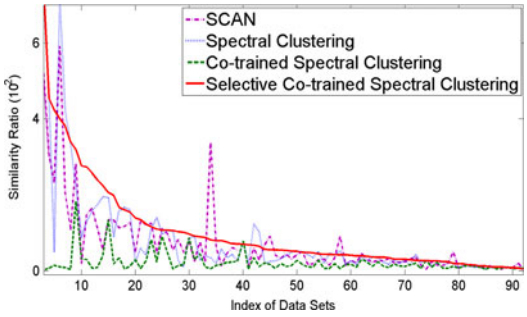


Fig. 4. The similarity ratios of all examined clustering techniques on the friendship view on 92 Twitter ego-nets. The ratios averaged over these ego-nets are: 71.9 for SCAN, 79.5 for cs, 20.9 for csc, and 100.6 for $scsc$.

TABLE 3
Representative Tags for Detected Circles in an Ego-Net

Circle	Representative Tags
$\mathcal{S}_1$	Human, Sleep
$\mathcal{S}_2$	Valentine's Day, Dance, Sport
$\mathcal{S}_3$	Ireland, Beer, Coffee
$\mathcal{S}_4$	Social media, Health, Cancer
$\mathcal{S}_5$	Yahoo!, WHATS'On (Software), Android

Roughly speaking, $\Pr\{t|\mathcal{S}\}$ is the averaged frequency of tag t appeared in circle $\mathcal{S}$. Then, tags with largest $\Pr\{t|\mathcal{S}\}$ were deemed most representative and extracted for examination.

The discriminative tags across circles were extracted as follows. Let $\mathcal{S}_k$ be the circle indexed by k, and $\mathcal{K} = \{k\}$ be the set of all circle indices. The discriminative degree of a tag t for circle k was measured as the KL-divergence

$$\mathbb{D}(t) = \sum_{k\in\mathcal{K}} \left(P(t;k) \ln \frac{P(t;k)}{Q(t;k)} \right), \tag{33}$$

where

$$P(t;k) = \Pr\{t|\mathcal{S}_k\} / \left(\sum_{k\in\mathcal{K}} \Pr\{t|\mathcal{S}_k\} \right), \tag{34}$$

and

$$Q(t;k) = 1/|\mathcal{K}|. \tag{35}$$

Intuitively, both $P(t;k)$ and $Q(t;k)$ could be interpreted as the probability mass of tag t in circle k (as over all detected circles) – $P(t;k)$ was the actually mass and $Q(t;k)$ was an ideal mass when t was uniformly distributed; then, $\mathbb{D}(t)$ said how much the actual distribution of tag t deviated from the uniform (thus non-informative) distribution. In our experiment, tags with the highest $\mathbb{D}(t)$ were deemed most discriminative and selected for examination.

The two types of tags extracted from circles detected by $scsc$ on a randomly selected ego-net were summarized in Table 3. They were obtained by first applying (33) to extract most discriminative tags over all circles, and then applying (31) among these tags to extract most representative ones for each circle. It was clear different circles had different semantic focuses: for instance, circle 2 had more interest in entertainment, while circle 4 seemed more concerned about health care and circle 5 talked about technology frequently. (We also skimmed through the tweets posted in these circles and had consistent findings.) This suggested circles detected by $scsc$ could be pretty interpretable in terms of topics.

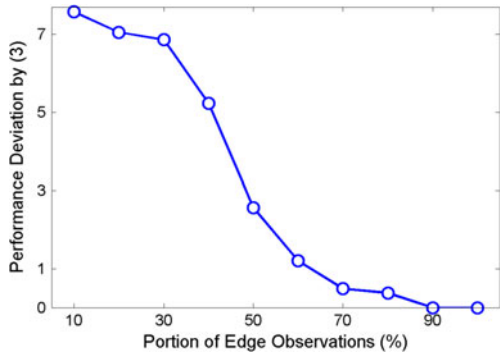


Fig. 5. Performance deviation based on graph K and its inherently incomplete observation $\tilde{K}$ under metric (7).

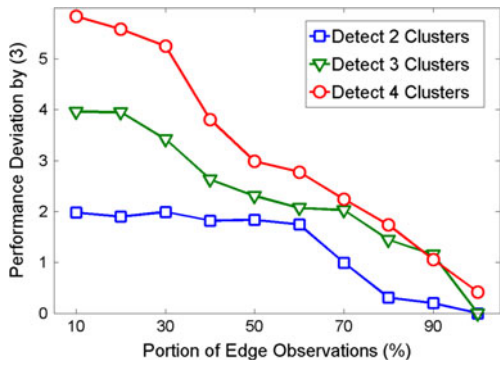


Fig. 6. Performance deviation under different number of clusters.

5.6 Evaluation 4: Cluster Inherently Incomplete Graph

In this section, we examined several implications obtained from Proposition 1. Since in reality it is impossible to know whether a social network has been fully observed, we presented simulations for examination.

Consider a set of 200 nodes partitioned into k_* groups, each containing $\lfloor 200/k_* \rfloor$ nodes. (The last group also contains the residual nodes.) We constructed a binary graph by building an edge between each pair of nodes, with probably p if they were from the same group and with probability $1-p$ otherwise. The resulted graph was considered as the underlying complete graph. Note this graph ccould be fully characterized by its adjacent matrix $K \in \mathbb{R}^{200\times 200}$ such that $K_{ij} = 1$ with probability p if nodes i and j were from the same group and with probability $1-p$ otherwise.

To simulate inherently incomplete observations of the graph, we randomly hid a portion of its edges by flipping a portion of entry 1's to 0's in K (without recording which entries were flipped). Let δ be the portion of un-flipped 1's and $\tilde{K}$ be the resulted adjacent matrix. Then, we applied the standard spectral clustering (e.g., [45]) on both K and $\tilde{K}$ and evaluated the difference of their performance based on metric (7). To minimize the performance variation of the K-means clustering method (mostly induced from its selection of initial cluster centers), we fixed one node for each group to form the initial centers.

In Fig. 5, we showed the performance deviation as the observations increase with $k_* = 5$. It is clear the deviation decreases as more observations are obtained, which is consistent with the implication of our bound.

In Fig. 6, we showed the performance deviation under different numbers of detected clusters, which was controlled by

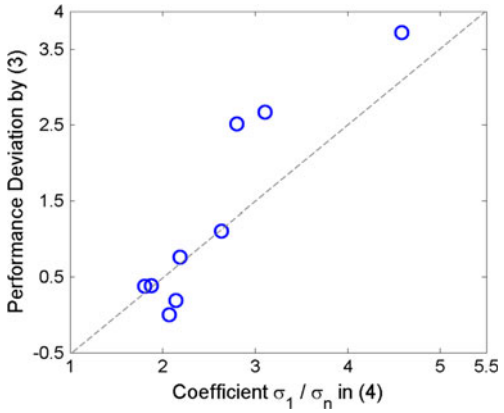


Fig. 7. Performance deviation versus coefficient σ_1/σ_n under different choices of edge-generation probability p' for one group of clusters.

parameter k in (8). For each choice, the initial cluster centers were chosen by a standard 'cluster' setting in Matlab (i.e., 10 percent of the nodes were randomly sampled to perform clustering first, and centers of the resulted clusters were used as the initial centers for the final clustering algorithm). It is clear that smaller choice of k suffers smaller performance deviation, and this is consistent with the implication of our bias bound in (8) which decreases as k decreases.

Finally, we examined the impact of coefficient σ_1/σ_n on performance deviation. To this end, we first fixed $k_* = 5$, $k = 5$ and $\delta = 30$ percent so the graph contained 5 clusters by nature. The edge-generation probability remained largely the same as before, except for the firs cluster we used another probability p' which varied from 0.1 to 0.9 with step size 0.1. Note different choices of p' would generate different graphs, and for each choice we applied spectral clustering on both its generated $K, \tilde{K}$ and evaluated the performance deviation. Meanwhile, we recorded the coefficient σ_1/σ_n for each K. Finally, the deviation-coefficient pair for each choice of p' was plotted as point in Fig. 7.

We had three major observations from the figure. First, in general a larger coefficient σ_1/σ_n corresponded to larger performance deviation, which was consistent with the implication of our bound. Second, the relation between the coefficient and deviation was roughly linear, which also coincided with our result. It was noted, however, such relation did not cross the origin as suggested by the bound; instead, it was biased by nearly a constant factor. We believed such bias was reasonable and should largely corresponded to the approximation error of ignored for spectral clustering. (Recall this error was ignored by the free-approximation assumption.) We hence do not claim the bound tight. However, its multiple implications were verified and may still provide useful insights for algorithm designs.

6 Related Work

6.1 Social Circle Clustering

A social circle is a group of people with certain type of social intimacy. Identifying social circles in a user's online social network provides an important way for the user to exert appropriate access control over his/her information dissemination [14], [53]. However, manual identification is usually tedious and exhausting, which triggered the study of automatic social circle identification [20], [25]. The idea is to algorithmically cluster people into groups so that those in the same group are similar under proper metric [1], [43].

Existing social circle detection techniques can be roughly categorized into graph-based and content-based. Graph-based detection methods use only topological structure or linkage information of the social network, including graph partitioning [27], hierarchical clustering [43], likelihood maximization [44] and matrix factorization [68]. On the other hand, content-based detection methods use semantic information on social network such as email, tweet messages and documents [9], [37], [58], [69], [70]. In addition, user interaction on social network has also been used for detection [47], as well as user profiles [40], [54]. It should be mentioned, however, these techniques focused on a single view of the network structure in social circle identification.

6.2 Multi-View Spectral Clustering

Multi-view spectral clustering is a framework that effectively combines information from multiple sources under the view consistency idea. It has demonstrated superior performance in many applications such as document categorization, digit classification and image annotation (see [62] and reference therein). However, to our knowledge this framework has not been applied for detecting social circles, and our work is the first effort in this direction.

In this paper, we have employed the co-trained spectral clustering algorithm [29], but also realized its potential limitation of ignoring the inherent incompleteness of sparse views. The modified algorithm we presented in this paper is an attempt to lift this limitation and we demonstrated its advantage in experiments.

6.3 Clustering on Incomplete Graph

Clustering on incomplete graphs is not a new topic. See studies in [10], [34], [49], [57] for example. Most of these works provided only algorithmic solutions and a few theoretical studies assumed prior knowledge on which part of the graph is missing. However, none of them address our concern on the performance deviation between clustering a graph and its incomplete observation.

A work more related to ours is [26], which analyzes how much the spectral clustering solution on a *complete* graph may deviate from the optimal normalized cut solution. Our analysis focuses on a fundamentally different problem, i.e., how much would two spectral clustering solutions deviate, with one based on a complete graph and the other based on its incomplete observation. Technically, we use the same evaluation metric as [26] and borrow some of its results, while additionally introducing perturbation theories to incorporate the incomplete observation.

Another related work is [24], which analyzes the effect of graph perturbation on spectral clustering. We study the same research question, but our analysis is fundamentally different from theirs in at least three aspects. First, the problem settings are different: they study only bi-partitioning based on the second principal eigen-vector, while we study multi-partitioning based on the k learning eigen-vectors. Second, the evaluation metrics are different: their metric does not consider the cluster sizes while ours does. Third, the proving techniques are different: they use a water-filling argument whereas we largely rely on the fundamental properties of Loewner partial orders; we also borrow a latest perturbation result from [66] and some results in [26].

7 Discussions

While this paper has focused on initiating and verifying the idea that one could exploit multiple views of the ego-net structure (e.g., by multi-view spectral clustering techniques) for better social circle detection in ego-nets, we have realized certain orthogonal directions that could further the study.

7.1 Problem Setting

The presented study focused on detecting *disjoint* social circles based solely on *network structure*, while we had mentioned other studies that focused on detecting overlapping circles and using alter profiles.

Detecting disjoint social circles is a common setting in the literature (e.g., [21], [42]). Here, our adoption was particularly motivated from a user privacy protection perspective – a major proposal in the privacy research community is to protect user privacy by drawing and controlling information boundaries in online social networks, so that an ego's posts are spread only within designated circles [55]; in this case, if an alter is assigned to multiple circles, then her re-actions (e.g., 'like' or 're-tweet') in assigned designated circles may be easily observed in other assigned non-designated circles. We admit, however, social circles may overlap in reality. In that case, first notice the two settings are convertible—two overlapping circles $\mathcal{S}_1$ and $\mathcal{S}_2$ generally admit three disjoint circles $\mathcal{S}_1 \cap \mathcal{S}_2$, $\mathcal{S}_1 \setminus \mathcal{S}_2$ and $\mathcal{S}_2 \setminus \mathcal{S}_1$; and three disjoint circles $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$ could be merged into two overlapping ones $\mathcal{C}_1 \cup \mathcal{C}_2$ and $\mathcal{C}_1 \cup \mathcal{C}_3$. This allows a direct technical extension of our study for the case of overlapping circles. In addition, one could also extend more sophisticated techniques (e.g.,[40]) from their single-view settings to multi-view settings. While the technical extensions may not be particularly difficult, a more challenging question is how to balance circle overlapping and privacy protection (as we mentioned above).

Using network structure to detect social circle is also a common setting [63], and our adoption was again motivated by privacy protection—that alters may be reluctant to share or even fill in their true profiles on online social networks due to privacy concerns. When this is not a serious concern, however, and alter profiles are largely available, our study could be directly extended by building an extra view on the profile similarities between alters (e.g., similarity between two alters is the inner product of their profile vectors).

7.2 Network Modeling

The network modeling techniques presented in this paper were largely based on classic techniques and where chosen for their simplicity or based on our experience.

As one example, the interaction similarity between two alters was obtained by summarizing their similarities in both directions (e.g., the reply number between Bob and Nancy was obtained by adding the number of replies from Bob to Nancy and that from Nancy to Bob). This is a classic technique that symmetrizes directed social networks into un-directed ones for analysis (e.g., [39], [41], [48]), but ignores the directional information which may be integrated in a *finer* manner for further performance improvement. In particular, how to effectively integrate direction modeling with multi-view learning remains an open challenge. (For instance, one may think of modeling each direction as one view but would suffer more sparsity in each new view.)

As another example, the topic view was normalized by the classic TF-IDF technique to highlight the importance of tags for each alter. We chose this pre-processing technique for it had been successfully applied in a similar study [47] as well as other tasks of social network analysis based on our own experience (e.g.,[11], [17]). However, the technique itself is not without any limitation—in our application, for instance, it may assign a low TF-IDF weight to an alter's truly interested topic if she only connects with people who talk about that topic. It then remains an open question that whether and how such lower weights may have adverse impact on social circle detection (in particular, under the multi-view clustering framework presented in this paper).

7.3 Circle Clustering

The major social circle detection techniques examined in this paper belong to the family of multi-view spectral clustering (e.g.,[29], [31]). We chose this family for it is a union of spectral clustering [45] and multi-view learning [5], two ancestors with high reputations. It is noted, however, each ancestor has its own challenges which could have been inherited by multi-view spectral clustering.

As one example, standard spectral clustering [45] eventually groups instances by directly applying K-means, which is a tremendously popular data clustering technique. However, K-means has a well-known challenge in *manually* choosing a proper group number [28], which is passed down to spectral clustering and now to the multi-view spectral clustering techniques we applied for social circle detection. One could clearly address the problem from the very origin, say, by applying new K-means [28] that could automatically determine the group number. However, it would perhaps be more interesting to integrate such determination with multi-view learning in the context of social circle detection.

As another example, the multi-view learning family has been largely built on the *view-consensus* assumption, i.e., the label assignments on different instance views should largely agree [62], which has been a key to theoretically justify its success and sample efficiency (e.g., [3], [32]). However, the assumption may not always hold, say, due the presence of noise (e.g.,[13]), in which case enforcing view consistency may result in performance degradation—an effect usually referred as *negative transfer*. Negative transfer has been broadly studied in multi-task learning and collective matrix factorization (e.g., [30], [33]), but its discussion in multi-view learning seems scarce. The selective transfer mechanism presented in this paper was an attempt to mitigate the problem, but we believe there remains spaces for improvements. For instance, we have ignored results on noisy observations during while transferring information across views, and it remains an open question how noisy observations may be leveraged to improve performance.

7.4 Application

A potential application of social circle is to draw information boundary between different circles, so that a message could be delivered only to designated circles (e.g., [54], [55]). Note the final social circle construction does not have to be entirely automatic, and the ego may manually modify the detected groups. In this case, circle detection still significantly reduces the effort of human labeling. Another issue is the information boundaries may not be completely secure if the social networking sites allows breaches in privacy protection (e.g., alters could 're-share' their received private posts). These are, however, beyond the scope of the paper.

The discovered social circles could also be used to improve the efficiency of ad delivery, targeted advertising,

and opinion mining in social groups. (See [40] for more discussions.) Social circles could also be used to study users' socialization behavior and social network information flow. When the temporal information of data is available, our methods may be further extended to detect circles in evolving ego-nets. (See some latest progress on dynamic social network analysis in [59], [60] for instance.) In addition, when information of the ego-net is available from other domains (e.g., [36], [61]), it is possible to further improve our work by considering cross-domain cross-view social circle detection.

8 Conclusion

In this paper, we proposed to automatically detect social circles of an ego-net based on its multi-view network structure. We crawled and modeled Twitter ego-nets by six views, and showed multi-view spectral clustering outperformed the commonly adopted single-view clustering on these ego-nets. We also showed, by treating sparse views as inherently incomplete ones and selectively transferring information across views, our modified multi-view clustering technique outperformed the standard multi-view clustering technique. The performance bias of standard clustering on inherently incomplete networks was briefly studied.

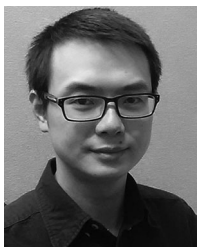
Acknowledgments

This work is supported in part by the US National Science Foundation under NSF CNS-1422206, NSF CNS-1337899, and NSF DGE-1565570.

References

[1] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Netw.*, vol. 25, pp. 211–230, 2003.
[2] F. R. Bach and M. I. Jordan, "Learning spectral clustering," *Comput. Sci.*, Tech. Rep. UCB/CSD-03-1249, 2003.
[3] M.-F. Balcan, A. Blum, and K. Yang, "Co-training and expansion: Towards bridging theory and practice," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 89–96.
[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[5] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. Conf. Learning Theory*, 1998, pp. 92–100.
[6] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Statist.-Theory Methods*, vol. 3, no. 1, pp. 1–27, 1974.
[7] T. Chakraborty, S. Patranabis, P. Goyal, and A. Mukherjee, "On the formation of circles in co-authorship networks," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 109–118.
[8] H.-W. Chang, D. Lee, M. Eltaher, and J. Lee, "@phillies tweeting from philly? predicting twitter user locations with spatial word usage," in *Proc. Conf. Adv. Social Netw. Anal. Mining*, 2012, pp. 111–118.
[9] J. Chang, J. Boyd-Graber, and D. M. Blei, "Connections between the lines: augmenting social networks with text," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 169–178.
[10] Y. Chen, A. Jalali, S. Sanghavi, and H. Xu, "Clustering partially observed graphs via convex optimization," *J. Mach. Learning Res.*, vol. 15, no. 1, pp. 2213–2238, 2014.
[11] Y. Chen, N. Yu, B. Luo, and X.-w. Chen, "iLike: Integrating visual and textual features for vertical search," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 221–230.
[12] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 759–768.
[13] C. Christoudias, R. Urtasun, and T. Darrell, "Multi-view learning in the presence of view disagreement," in *Proc. Conf. Uncertainty Artif. Intell.*, 2012, pp. 88–96.
[14] D. M. boyd and N. B. Ellison, "Social network sites:definition, history and scholarship," *J. Comput.-Mediated Commun.*, vol. 13, pp. 210–230, 2008.
[15] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation," *J. Numerical Anal.*, vol. 6, pp. 159–173, 1970.
[16] K. Dykstra, J. Lijffijt, and A. Gionis, "Covering the egonet: A crowdsourcing approach to social circle discovery on Twitter," in *Proc. AAAI Conf. Web Social Media*, 2015, pp. 606–609.
[17] H. Fei, R. Jiang, Y. Yang, B. Luo, and J. Huan, "Content based social behavior prediction: A multi-task learning approach," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 995–1000.
[18] P. Ferragina and U. Scaiella, "Fast and accurate annotation of short texts with Wikipedia pages," *IEEE Softw.*, vol. 29, no. 1, pp. 70–75, Jan./Feb. 2012.
[19] C. Gerlitz and B. Rieder, "Mining one percent of Twitter: Collections, baselines, sampling," *M/C J.*, vol. 16, no. 2, 2013.
[20] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proc. Conf. Human Factors Comput. Syst.*, 2009, pp. 211–220.
[21] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proc. Natl. Academy Sci. United State America*, vol. 99, no. 12, pp. 7821–7826, 2002.
[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins University Press, 2012.
[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 2012.
[24] L. Huang, D. Yan, N. Taft, and M. I. Jordan, "Spectral clustering with perturbed data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 705–712.
[25] S. Jones and E. O'Neill, "Feasibility of structural network clustering for group-based privacy control in social networks," in *Proc. ACM Symp. Usable Privacy Secur.*, 2010, Art. no. 9.
[26] F. Jordan and F. Bach, "Learning spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, pp. 305–312.
[27] B. W. Kernighan and S. Lin, "An efficient heuristic procedure for partitioing graphs," *Bell Syst. Tehnical J.*, vol. 49, pp. 291–307, 1970.
[28] B. Kulis and M. I. Jordan, "Revisiting k-means: New algorithms via bayesian nonparametrics," in *Proc. 29th Int. Conf. Mach. Learning*, 2012, pp. 513–520.
[29] A. Kumar and H. Daume, "A co-training approach for multi-view spectral clustering," in *Proc. 28th Int. Conf. Mach. Learning*, 2011, pp. 393–400.
[30] A. Kumar and H. Daume, "Learning task grouping and overlap in multi-task learning," in *Proc. 29th Int. Conf. Mach. Learn.*, 2012, pp. 1383–1390.
[31] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, 2011, pp. 1413–1421.
[32] C. Lan and J. Huan, "Reducing the unlabeled sample complexity of semi-supervised multi-view learning," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 627–634.
[33] C. Lan, J. Wang, and J. Huan, "Towards a theoretical understanding of negative transfer in collective matrix factorization," in *Proc. 32nd Conf. Uncertainty Artif. Intell.*, 2016, pp. 367–376.
[34] S.-Y. Li, Y. Jiang, and Z.-H. Zhou, "Partial multi-view clustering," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 1968–1974.
[35] W. Li, P. Serdyukov, A. P. de Vries, C. Eickhoff, and M. Larson, "The where in the Tweet," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 2473–2476.
[36] S. Liu, S. Wang, and F. Zhu, "Structured learning from heterogeneous behavior for social identity linkage," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 7, pp. 2005–2019, Jul. 2015.
[37] Y. Liu, A. N. Mizil, and W. Gryc, "Topic-link lda: Joint models of topic and author community," in *Proc. 26th Annu. Int. Conf. Mach. Learning*, 2009, pp. 665–672.
[38] J. Lu and D. Li, "Sampling online social networks by random walk," in *Proc. ACM Int. Workshop Hot Topics Interdisciplinary Soc. Netw. Res.*, 2012, pp. 33–40.
[39] S. Mahadevan, M. Maggioni, K. Ferguson, and S. Osentoski, "Learning representation and control in continuous Markov decision processes," in *Proc. 21st Natl. Conf. Artif. Intell.*, 2006, pp. 1194–1199.
[40] J. Mcauley and J. Leskovec, "Discovering social circles in ego networks," *ACM Trans. Knowl. Discovery Data*, vol. 8, no. 1, 2014, Art. no. 4.
[41] M. Meilă and W. Pentney, "Clustering by weighted cuts in directed graphs," in *Proc. SIAM Conf. Data Mining*, 2007, pp. 135–144.

[42] M. E. Newman, "Modularity and community structure in networks," *Proc. Natl. Academy Sci. United State America*, vol. 103, no. 23, pp. 8577–8582, 2006.
[43] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, 2004, Art. no. 026113.
[44] M. E. Newman and E. A. Leicht, "Mixture models and exploratory analysis in networks," *Proc. Natl. Academy Sci. United State America*, vol. 104, no. 23, pp. 9564–9569, 2007.
[45] A. Y. Ng, et al., "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 849–856.
[46] M. Papagelis, G. Das, and N. Koudas, "Sampling online social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 3, pp. 662–676, Mar. 2013.
[47] Y. Ruan, D. Fuhry, and S. Parthasarathy, "Efficient community detection in large networks using content and links," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 1089–1098.
[48] V. Satuluri and S. Parthasarathy, "Symmetrizations for clustering directed graphs," in *Proc. Int. Conf. Extending Database Technol.*, 2011, pp. 343–354.
[49] W. Shao, X. Shi, and P. S. Yu, "Clustering on multiple incomplete datasets via collective kernel learning," in *Proc. IEEE 13th Int. Conf. Data Mining*, 2013, pp. 1181–1186.
[50] R. Sharan, A. Maron-Katz, and R. Shamir, "Click and expander: A system for clustering and visualizing gene expression data, *Bioinf.*, vol. 19, no. 14, pp. 1787–1799, 2003.
[51] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[52] P. Shi, H. Xu, and Y. Chen, "Using contextual integrity to examine interpersonal information boundary on social network sites," in *Proc. Conf. Human Factors Comput. Syst.*, 2013, pp. 35–38.
[53] M. M. Skeels and J. Grudin, "When social networks cross boundaries: A case study of workplace use of Facebook and Linkedin," in *Proc. Int. Conf. Supporting Group Work*, 2009, pp. 95–104.
[54] A. Squicciarini, S. Karumanchi, D. Lin, and N. DeSisto, "Identifying hidden social circles for advanced privacy configuration," *Comput. Secur.*, vol. 41, pp. 40–51, 2013.
[55] A. Squicciarini, D. Lin, S. Karumanchi, and N. DeSisto, "Automatic social group organization and privacy management," in *Proc. 8th Int. Conf. Collaborative Comput.: Netw., Appl. Worksharing*, 2012, pp. 89–96.
[56] A. Strehl, "Relationship-based clustering and cluster ensembles for high-dimensional data mining," PhD thesis, The University of Texas at Austin, 2002.
[57] R. K. Vinayak, S. Oymak, and B. Hassibi, "Graph clustering with missing data: Convex algorithms and analysis," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2996–3004.
[58] X. Wang, H. Liu, and W. Fan, "Connecting users with similar interests via tag network inference," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 1019–1024.
[59] B. Wu, W.-H. Cheng, Y. Zhang, and T. Mei, "Time matters: Multi-scale temporalization of social media popularity," in *Proc. ACM Multimedia Conf.*, 2016, pp. 1336–1344.
[60] B. Wu, T. Mei, W.-H. Cheng, and Y. Zhang, "Unfolding temporal dynamics: Predicting social media popularity using multi-scale temporal decomposition," in *Proc. AAAI Publ. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 272–278.
[61] S.-H. Wu, H.-H. Chien, K.-h. Lin, and P. Yu, "Learning the consistent behavior of common users for target node prediction across social networks," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 298–306.
[62] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.
[63] X. Xu, N. Yuruk, Z. Feng, and T. A. J. Schweiger, "SCAN: A structural clustering algorithm for networks," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2007, pp. 824–833.
[64] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *Proc. IEEE Int. Conf. Data Mining*, 2013, pp. 1151–1156.
[65] Y. Yang, C. Lan, X. Li, B. Luo, and J. Huan, "Automatic social circle detection using multi-view clustering," in *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2014, pp. 1019–1028.
[66] Y. Yu, T. Wang, and R. J. Samworth, "A useful variant of the davis–kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, 2015.
[67] H. Zha, X. He, C. Ding, M. Gu, and H. D. Simon, "Spectral relaxation for k-means clustering," in *Proc. Adv. Neural Inform. Process. Syst.*, 2002, pp. 1057–1064.
[68] Y. Zhang and D.-Y. Yeung, "Overlapping community detection via bounded nonnegative matrix tri-factorization," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 606–614.
[69] D. Zhou, E. Manavoglu, J. Li, C. L. Giles, and H. Zha, "Probabilistic models for discovering e-communities," *Proc. 15th Int. Conf. World Wide Web*, 2006, pp. 173–182.
[70] W. Zhou, H. Jin, and Y. Liu, "Community discovery and profiling with social messages," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 388–396.



Chao Lan received the BE and ME degrees from the Nanjing University of Post and Telecommunication, China, in 2008 and 2011, respectively. He is currently working toward the PhD degree in computer science at the University of Kansas. His research area is machine learning.

Yuhao Yang received the BEng degree from Nanjing University, in 2009 and the PhD degree from the University of Kansas, in 2014. During his doctorate study, he worked with Dr. Bo Luo in the area of online social network user privacy investigation and protection. He is currently a software engineering at Software Engineer at Microsoft, Universal Store, Membership Privacy.

Xiaoli Li received the MS degree in mechanics engineering from the Dalian University of Technology, China, in 2001. She is currently working toward the PhD degree in computer science at the University of Kansas. Her research interests include classical multi-task multi-view learning and developing Bayesian nonparametric models for online multi-task multi-view learning.

Bo Luo received the BE degree from the University of Science and Technology of China in 2001, the MPhil degree from the Chinese University of Hong Kong in 2003, and the PhD degree from Pennsylvanian State University, in 2008. He is currently an associate professor with the EECS Department, University of Kansas (KU). He is also the director of the Information Assurance Laboratory of the Information and Telecommunication Technology Center at KU. His research interests lie in the intersection of security and privacy, and data science. In particular, he is interested in: information security and privacy, smart grid and IoT/CPS security, information retrieval, Web, and online social networks.

Jun Huan received the PhD degree in computer science from the University of North Carolina. He is a professor in the Department of Electrical Engineering and Computer Science, the University of Kansas. He directs the Data Science and Computational Life Sciences Laboratory at the KU Information and Telecommunication Technology Center (ITTC). He works on data science, machine learning, data mining, big data, and interdisciplinary topics including bioinformatics. He has published more than 120 peer-reviewed papers in leading conferences and journals and has graduated more than 10 graduate students including seven PhDs.