Unifying Virtual and Physical Worlds: Learning Toward Local and Global Consistency

XIANG WANG, National University of Singapore
LIQIANG NIE and XUEMENG SONG, ShanDong University
DONGXIANG ZHANG, University of Electronic Science and Technology of China
TAT-SENG CHUA, National University of Singapore

4

Event-based social networking services, such as Meetup, are capable of linking online virtual interactions to offline physical activities. Compared to mono online social networking services (e.g., Twitter and Google+), such dual networks provide a complete picture of users' online and offline behaviors that more often than not are compatible and complementary. In the light of this, we argue that joint learning over dual networks offers us a better way to comprehensively understand user behaviors and their underlying organizational principles. Despite its value, few efforts have been dedicated to jointly considering the following factors within a unified model: (1) local user contextualization, (2) global structure coherence, and (3) effectiveness evaluation. Toward this end, we propose a novel dual clustering model for community detection over dual networks to jointly model local consistency for a specific user and global consistency of partitioning results across networks. We theoretically derived its solution. In addition, we verified our model regarding multiple metrics from different aspects and applied it to the application of event attendance prediction.

CCS Concepts: • **Information systems** → **Clustering**; *Social networking sites*; • **Computing methodologies** → *Cluster analysis;*

Additional Key Words and Phrases: Event-based social networks, global consistency, local consistency

ACM Reference Format:
Xiang Wang, Liqiang Nie, Xuemeng Song, Dongxiang Zhang, and Tat-Seng Chua. 2017. Unifying virtual and physical worlds: Learning toward local and global consistency. ACM Trans. Inf. Syst. 36, 1, Article 4 (April 2017), 26 pages.
DOI: http://dx.doi.org/10.1145/3052774

1. INTRODUCTION

The past decades have witnessed the proliferation of conventional online social networking services such as Facebook[1] and Renren,[2] which break the barrier of physical distance and allow interactions among friends in a virtual world. As a new branch of social networking services, event-based social networks (EBSNs) are taking root and

[1]http://www.facebook.com/.
[2]http://www.renren.com/.

This work was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its IRC@SG Funding Initiative.
Authors' addresses: X. Wang and T.-S. Chua, AS6, #5-25 Lab for Media Search, School of Computing, National University of Singapore, 13 Computing Drive, Singapore 117417; emails: xiangwang@u.nus.edu, dcscts@nus.edu.sg; L. Nie and X. Song, No. 1500, Shunhua Road, Jinan, Shandong Province, P.R. China 250101; emails: {nieliqiang, sxmustc}@gmail.com; D. Zhang, Qingshuihe Campus: No. 2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu, Sichuan, P.R. China; email: zhangdo@uestc.edu.cn.

DOI: http://dx.doi.org/10.1145/3052774

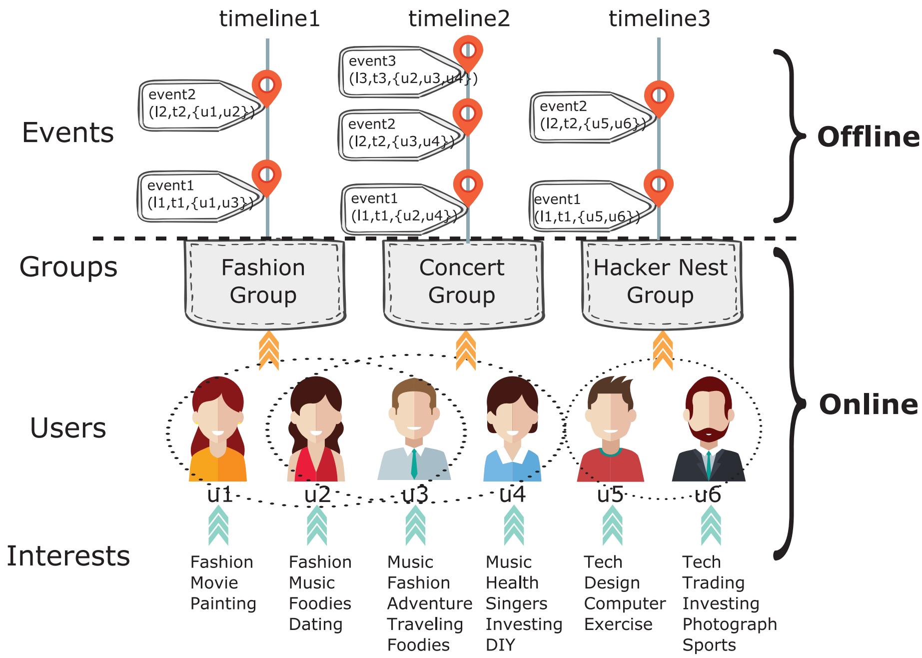


Fig. 1. Demonstration of EBSNs, where users u can be involved in multiple online groups represented by dash circles and attend the specific physical event at the specific location l and timestamp t.

exerting tremendous fascination on people. Distinct from conventional online social networking services, which usually connect acquaintances, EBSNs not only help individuals find like-minded event attendees and further form lasting, influential, and local community groups, but they also facilitate them to regularly meet face-to-face in various physical locations. In other words, EBSNs emphasize more on organizing "regional" physical events, such as "1 hour of bubble soccer" and "bubble bump sports at Kovan Sports Centre Pte Ltd" among local communities. To date, the unique and interesting services of EBSNs have attracted a sheer volume of users. The most representative examples are Meetup,[3] Plancast,[4] and Facebook Event.[5] Take Meetup as an example: as of August 2015, it claimed to have 27.7 million members in 180 countries and 210,240 groups [Liu et al. 2012; Pham et al. 2016].

As demonstrated in Figure 1, EBSNs are able to present online and offline social networking services simultaneously. In the online part, users are associated with several interest tags, such as fashion, music, and technique [Li et al. 2015]. They can be concurrently involved in multiple online groups according to their personal interests [Wang et al. 2011; Song et al. 2015b]. Within the same groups, users can interact with each other online by exchanging thoughts and sharing experiences on topics of common interests [Li et al. 2015]. In the offline part, EBSNs also support and record user interactions in the physical world, where users can create and invite others to join social events, such as dining out and jogging, at specific places and times [Zhang et al. 2016]. As we can see, each network may contain some knowledge that the other does not have; therefore, they can be jointly employed to comprehensively and accurately

[3]https://www.meetup.com.
[4]https://www.eventbrite.com.
[5]https://www.facebook.com/events/upcoming/.

describe user behaviors. In fact, user behaviors across online and offline networks are inseparable and mutually reinforced [Tang and Liu 2009; Li et al. 2012; Nie et al. 2016; Halcrow et al. 2016]. On one hand, online group users sharing common interests hold a better than average chance of attending the same offline events. On the other hand, coparticipating in the offline events can in turn strengthen their social ties and further propel the engagement of online interactions [Foley et al. 2015; Nie et al. 2016]. In summary, online and offline user behaviors are usually compatible and complementary to each other rather than independent. We thus argue that, as compared to pure online or offline network, appropriate aggregation of them can provide us a better way to comprehensively understand user behaviors and their underlying organizational principles.

In this article, we address one particular and fundamental task, namely community detection in EBSNs. Typical community detection aims at grouping persons who are connected to each other by relatively durable social relations to form a tight and cohesive social entity due to the presence of a "unity of will" or "sharing of common values" [Ng et al. 2011; James et al. 2012; Nie et al. 2016]. Community detection by jointly considering online and offline networks will offer opportunities to gain new insights into user behaviors and their organizational structures, and further support new services and applications. However, this task is nontrivial and poses a set of tough challenges:

(1) *User contextualization*: Due to various complex reasons, some users may be active in online interactions but rarely attend physical events or vice versa. This leads to the problem of inconsistent and unbalanced interactive behaviors of the same users across online and offline networks. Moreover, the same users probably own different close neighborhoods, namely social circles, in these different networks. Therefore, how to design a suitable approach to characterizing the invariant essence of users across heterogeneous networks remains largely untapped.
(2) *Structure coherence*: From an overall perspective, online and offline networks are composed of the same set of users; meanwhile, they interpret their behaviors from different but coherent angles. Accordingly, we believe that the inherent and hidden community structures embedded in these two networks should be similar. However, how to properly model the consensus structures to reinforce the performance of community detection is a challenging issue.
(3) *Effectiveness evaluation*: How to quantitatively validate the result of community detection and verify its applicability in real-world problems is another challenge we face.

Community detection across online and offline networks can be treated as a particular task of either multiview clustering or community detection in multilayer graphs. These two clustering technologies both exploit and fuse information from multiple views [Leung et al. 2011; Ng et al. 2011; Tang et al. 2012; Li et al. 2014]. It is worthwhile mentioning that a wide range of approaches have been proposed, with strong theoretical underpinnings and great practical success. In general, prior efforts can be roughly summarized into two categories. One is to concatenate a set of features extracted from multiple views before applying any conventional off-the-shelf clustering algorithms on these features [Blaschko and Lampert 2008; Chaudhuri et al. 2009]. The other is to derive clustering structures from individual views or individual layer graphs and reconcile them based on the principle of consensus [Bruno and Marchand-Maillet 2009; Gao et al. 2013a; Kumar and Daumé 2011; Kumar et al. 2011; Liu et al. 2012; Zhou and Burges 2007; Boden et al. 2012; Zeng et al. 2006]. However, the approaches in the first category usually ignore the relatedness among internetworks and frequently suffer from the curse of dimensionality and incompatible feature space. The second category is to derive clustering structures from individual views or layer graphs and reconcile them based on the principle of consensus. The approaches in this category usually model the global consensus across views or layers and consider more comprehensive representations of users compared with that of the first category. However, rare efforts are dedicated to involving the local consensus, which is more beneficial to user contextualization [Yang et al. 2010]. In addition, few of them have been applied to solve the community detection over EBSNs.

To tackle the preceding challenges, we propose CLEVER—a novel dual CLustering model for community dEtection oVer dual nEtwoRks. Rather than learning from online and offline separately, CLEVER is capable of reinforcing community detection over online and offline networks by coregularizing local and global consistency in a unified model. To be more specific, local consistency is to encourage the invariants of the same users across different networks, whereby users are contextualized by their surrounding neighbors. This is inspired by the proverb "a leopard never changes his spots in different environments" and the principle of "neighborhood preserving" [Ronen et al. 2014; Jiang et al. 2014; Nie et al. 2016]. Comparatively speaking, global consistency aims to maximize the agreement on community detection results in separate networks, which intelligently ensures the similar underlying structures over the same set of users. We coregularize local and global consistency in our proposed CLEVER model at the same time. We verify our model on a publicly accessible dataset and apply it to a real-world application: event attendance prediction. By conducting experiments on a representative dataset crawled from Meetup, we demonstrate that our proposed model yields significant gains in EBSNs' community detection and achieves fairly satisfactory results for the application.

The main contributions of this article are threefold:

(1) We propose a novel dual clustering model for community detection over virtual-physical networks in EBSNs. It enhances the community detection performance by harvesting the compatible and complementary information cues with local and global consistency.
(2) We theoretically optimize our proposed model by the novel and effective Crank-Nicolson–like update scheme, which tackles the quadratic programming optimization problem with orthogonal constraints.
(3) We validate the effectiveness of our proposed CLEVER model on the application of event attendance prediction. In addition, we have released our data, code, and parameter settings to facilitate other researchers to repeat our experiments and verify their own models.[6]

The remainder of the article is organized as follows. Section 2 reviews the related work. Sections 3 and 4 respectively detail our proposed CLEVER model and present experimental results and analyses. Section 5 introduces the application, followed by our conclusion and future work in Section 6.

2. RELATED WORK

Multiview clustering and multilayer graph clustering methods are suitable for dual clustering over dual networks in EBSNs. The basic idea of these approaches is to partition objects into clusters based on multiple representations of the object from different views. In recent years, multiview and multilayer graph learning has received increasing attention, and existing algorithms can be classified into two categories: early fusion [Blaschko and Lampert 2008; Chaudhuri et al. 2009] and joint learning [Bruno and Marchand-Maillet 2009; Gao et al. 2013a; Kumar and Daumé 2011; Kumar et al.

[6]http://lms-dataming.wix.com/clever.

2011; Liu et al. 2012; Zhang et al. 2015; Zeng et al. 2006; Boden et al. 2012; Cai et al. 2013; He et al. 2014; Li et al. 2014].

The early fusion methods concatenate all views into a single view and then adapt any conventional off-the-shelf clustering algorithms, such as k-means [Blaschko and Lampert 2008; Chaudhuri et al. 2009]. For instance, Chaudhuri et al. [2009] constructed a lower-dimensional feature subspace from multiple views of data by canonical correlation analysis (CCA) and then applied single linkage clustering on the projections. Blaschko and Lampert [2008] utilized kernel-CCA to simultaneously learn linear projections from multiple spaces into a common latent space and then proposed a generalization of spectral clustering based on the latent space. However, these approaches generally overlook the obvious fact that each view has its own specific statistical property and ignore the structural relatedness among views. Additionally, these approaches suffer from the overfitting problem in case of insufficient training samples.

The other paradigm of joint learning aims to derive clustering structures and attributes from individual views and reconcile them based on the principle of consensus. This line of research can be divided into several subdirections. First, the normalized cut, generalized from a single view to multiple views, finds a cut that is close to the optimal one on each graph. Zhou and Burges [2007] generalized the normalized cut approach to multiple views via a random walk formulation. Second, cotraining is to limit clustering result in each view to agree with those in other views. Kumar and Daumé [2011] incorporated a cotraining framework with multiview spectral clustering, where the graph structure of one view is constrained and modified by the eigenvectors of the Laplacian in other views. Third, joint nonnegative matrix factorization (NMF) for multiview clustering jointly factorizes the multiple matrices through coregularization. He et al. [2014] applied NMF on multiview data to obtain coefficient matrices derived from factorizations of different views and further regularized them toward a common consensus. Moreover, Ni et al. [2015] developed a flexible and robust framework based on NMF that clusters multiple domain-specific networks sharing multiple underlying clustering structures. Fourth, coregularization encourages that corresponding data points in each view should have the same cluster membership via coregularizing the clustering hypotheses. Kumar and Daumé [2011] coregularized the clustering hypotheses from multiple views and further enforced these hypotheses to be consistent across all views via a disagreement measurement. Cai et al. [2011] proposed an effective multimodal spectral clustering method via learning a commonly shared graph Laplacian matrix by unifying different views. Fifth, pattern mining focuses on attribute and structural information of multiple graphs. Boden et al. [2012] proposed a multilayer graph learning method considering both aspects of structural density and attribute similarity of nodes. Liu et al. [2012] employed an extended Fiedler method to incorporate the heterogeneity between online and offline networks during the community detection process. Although existing approaches focus on the principle of global consensus across views, they ignore local consistency of nodes' contextualization. In addition, few of them are applied to the situation of community detection over online and offline networks.

3. STEPWISE MODEL DEMONSTRATION

We first declare some notations. In particular, we use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to denote matrices and vectors, respectively. We use $\mathbf{I}$ to denote the identity matrix. If not clarified, all vectors are in column forms. We employ nonbold letters (e.g., x) to represent scalars and Greek letters (e.g., β) as parameters. We denote the Frobenius norm and the trace of matrix $\mathbf{X}$ as $\|\mathbf{X}\|_F$ and $tr(\mathbf{X})$, respectively. Moreover, let $X(i, j)$, $\mathbf{X}(i, :)$, and $\mathbf{X}(:, j)$ respectively denote the entry in row i and column j, the i-th row, and the j-th column of $\mathbf{X}$. We utilize the superscripts v and p to indicate the virtual (online) and physical (offline) networks, respectively, as well as subscript i to represent the specific user i.

3.1. Problem Statement

Suppose that we have a set of N users. Let us respectively denote their representations in online and offline networks as $\mathbf{X}^v = [\mathbf{x}_1^v, \ldots, \mathbf{x}_N^v] \in \mathbb{R}^{D_v \times N}$ and $\mathbf{X}^p = [\mathbf{x}_1^p, \ldots, \mathbf{x}_N^p] \in \mathbb{R}^{D_p \times N}$, where D_v and D_p are the corresponding feature dimensions. In particular, we characterize the user online and offline representations (i.e., $\mathbf{X}^v$ and $\mathbf{X}^p$) by using their interest tags and physical event attendance records, respectively. D_v and D_p separately denote the number of all interest tags and event records.

The basic insight of community detection is to encourage similar users to have similar community assignments [Pham et al. 2016]. Toward this end, we need to reveal the pairwise similarities among users and then construct the networks, which explore the implicit topological structure among users. In this work, we employ the Gaussian similarity function on user representations to obtain online and offline social affinity graphs. It is noted that some off-the-shelf clustering methods, such as k-means, applied on user representations rather than social networks, hardly lead themselves to the explicit topology and compact communities. In particular, we construct online and offline social affinity graphs $\mathbf{A}^v \in \mathbb{R}^{N \times N}$ and $\mathbf{A}^p \in \mathbb{R}^{N \times N}$ by calculating their pairwise similarities with the Gaussian similarity function [Nie et al. 2011a, 2015, 2016],

$$\begin{cases} A^v(i,j) = \exp\left(-\frac{\left\|\mathbf{x}_i^v - \mathbf{x}_j^v\right\|^2}{(\theta^v)^2}\right) \\ A^p(i,j) = \exp\left(-\frac{\left\|\mathbf{x}_i^p - \mathbf{x}_j^p\right\|^2}{(\theta^p)^2}\right) \end{cases}, \tag{1}$$

where the radius parameters θ^v and θ^p are simply set as the median of the Euclidean distances of all user pairs in online and offline networks separately.

Our research objective is to obtain compact and cohesive partitions $\mathcal{P}^v = \{\mathcal{P}_1^v, \ldots, \mathcal{P}_C^v\}$ and $\mathcal{P}^p = \{\mathcal{P}_1^p, \ldots, \mathcal{P}_C^p\}$ over the virtual and physical networks simultaneously, where $\mathcal{P}_i^v$ ($\mathcal{P}_i^p$) refers to the i-th community identified from the virtual (physical) network and C denotes the number of clusters. Meanwhile, $\mathcal{P}^v$ and $\mathcal{P}^p$ should be as consensus as possible from the perspective of local user level and global structure level.

For the ease of formulation, inspired by the work in Dhillon et al. [2004], we define two scaled assignment matrices $\mathbf{G}^v = [G^v(i,c)] \in \mathbb{R}^{N \times C}$ and $\mathbf{G}^p = [G^p(i,c)] \in \mathbb{R}^{N \times C}$ to respectively represent $\mathcal{P}^v$ and $\mathcal{P}^p$. In particular, they are formulated as follows:

$$\begin{cases} G^v(i,c) = \begin{cases} \frac{1}{\sqrt{|\mathcal{P}_c^v|}}, & \text{if } \mathbf{x}_i^v \in \mathcal{P}_c^v \\ 0, & \text{otherwise} \end{cases}, \\ G^p(i,c) = \begin{cases} \frac{1}{\sqrt{|\mathcal{P}_c^p|}}, & \text{if } \mathbf{x}_i^p \in \mathcal{P}_c^p \\ 0, & \text{otherwise} \end{cases}. \end{cases} \tag{2}$$

According to Equation (2), we can derive the orthogonal properties of $\mathbf{G}^v$ and $\mathbf{G}^p$ as follows:

$$\begin{cases} \mathbf{G}^{v\top}\mathbf{G}^v = \mathbf{I} \\ \mathbf{G}^{p\top}\mathbf{G}^p = \mathbf{I} \end{cases}. \tag{3}$$

Rather than forcing $\mathbf{G}^v = \mathbf{G}^p$ (i.e., the same community structure over dual networks), we remain the different but coherent structures over different networks (i.e., $\mathbf{G}^v$ and

$\mathbf{G}^p$), considering that users may have inconsistent online and offline behaviors. Hence, how to bridge the gap between these structures plays a pivotal role in propagating interdependent information implicitly and further boosting the clustering performance.

3.2. Clustering over Individual Network

Clustering [Chi et al. 2007; Wang and Davidson 2010; Wauthier et al. 2012; Liu et al. 2015] over a single network thus far has been well studied. Thereinto, spectral clustering [Ng et al. 2001] has become one of the most popular modern clustering algorithms. Theoretically, it is simple to implement via utilizing the eigenvectors of the Laplacian matrix derived from the data. In practice, it can be solved efficiently by a standard linear algebra software and generally outperforms traditional clustering methods such as k-means. Following the standard spectral clustering framework, we can separately perform clustering over online and offline networks via the following formulations:

$$\begin{cases} \min_{\mathbf{G}^v} tr(\mathbf{G}^{v\top}\mathcal{L}^v\mathbf{G}^v), \quad \text{subject to } \mathbf{G}^{v\top}\mathbf{G}^v = \mathbf{I} \\ \min_{\mathbf{G}^p} tr(\mathbf{G}^{p\top}\mathcal{L}^p\mathbf{G}^p), \quad \text{subject to } \mathbf{G}^{p\top}\mathbf{G}^p = \mathbf{I} \end{cases}, \tag{4}$$

where $\mathcal{L}^v$ and $\mathcal{L}^p$ are the normalized Laplacian matrices over online and offline networks, respectively, and

$$\begin{cases} \mathcal{L}^v = \mathbf{I} - (\mathbf{D}^v)^{-\frac{1}{2}}\mathbf{A}^v(\mathbf{D}^v)^{-\frac{1}{2}} \\ \mathcal{L}^p = \mathbf{I} - (\mathbf{D}^p)^{-\frac{1}{2}}\mathbf{A}^p(\mathbf{D}^p)^{-\frac{1}{2}} \end{cases}, \tag{5}$$

where $\mathbf{D}^v$ and $\mathbf{D}^p$ are the diagonal degree matrices. Despite huge empirical success of spectral clustering methods over a single network, they are unable to leverage the compatible and complementary relatedness among multiple networks to enhance the partition performance. Hence, their current stage is suboptimal to community detection over EBSNs.

3.3. Clustering over Dual Networks Toward Global Consistency

To achieve improved community detection performance over EBSNs, we simultaneously harvest the information from online and offline views. The basic idea is that these two networks admit the shared underlying clustering structure. In other words, corresponding users should have consistent cluster memberships across dual networks. Here, we encode the cluster memberships of users with the scaled assignment matrices $\mathbf{G}^v$ and $\mathbf{G}^p$, which can be viewed as new user representations (with the i-th row standing for the i-th user representation). With such new representations at hand, we aim to encourage *global consistency*, which encourages all pairwise similarities of users under the new representations to be similar over dual networks. This amounts to enforcing online and offline clustering structures to be consistent in a global view.

Inspired by the disagreement measurement function presented in Kumar et al. [2011] and Tang et al. [2012], we propose minimizing the following function over dual networks to implement global consistency:

$$D_{global}(\mathbf{G}^v, \mathbf{G}^p) = \left\| \frac{\mathbf{G}^v\mathbf{G}^{v\top}}{\|\mathbf{G}^v\|_F^2} - \frac{\mathbf{G}^p\mathbf{G}^{p\top}}{\|\mathbf{G}^p\|_F^2} \right\|_F^2, \tag{6}$$

where $\mathbf{G}^v(i,:)$ and $\mathbf{G}^p(i,:)$ refer to online and offline cluster memberships for the same user i, respectively. We treat the scaled community assignment matrices $\mathbf{G}^v$ and $\mathbf{G}^p$ as the new user representations [Kumar et al. 2011]. Accordingly, $\mathbf{G}^v\mathbf{G}^{v\top}$ and $\mathbf{G}^p\mathbf{G}^{p\top}$ represent online and offline pairwise similarities among users, respectively. The similarity matrices are normalized by their Frobenius norms to make the same users comparable across networks. We note that $||\mathbf{G}^v||_F^2 = tr(\mathbf{G}^{v\top}\mathbf{G}^v) = C$ and $||\mathbf{G}^p||_F^2 = tr(\mathbf{G}^{p\top}\mathbf{G}^p) = C$, where C is the number of communities, and the orthogonal constraints can make them comparable in nature.

We observed that some users have somehow different behaviors across the physical and virtual societies. This may result in users having different grouping distributions. To address such a problem, we intentionally do not enforce the grouping information over online and offline networks to be equal. Instead, we leverage the pairwise similarity among users to illustrate global consistency, as we found that users' relations are relatively stable across networks.

Combining global consistency with the spectral clustering objectives mentioned in Equations (4) and (5), we obtain the following joint optimization problem:

$$\begin{aligned} \min_{\mathbf{G}^v,\mathbf{G}^p} & tr(\mathbf{G}^{v\top}\mathcal{L}^v\mathbf{G}^v) + tr(\mathbf{G}^{p\top}\mathcal{L}^p\mathbf{G}^p) \\ & + \frac{\lambda_1}{2}\left\|\frac{\mathbf{G}^v\mathbf{G}^{v\top}}{\|\mathbf{G}^v\|_F^2} - \frac{\mathbf{G}^p\mathbf{G}^{p\top}}{\|\mathbf{G}^p\|_F^2}\right\|_F^2, \\ \text{subject to} \quad & \mathbf{G}^{v\top}\mathbf{G}^v = \mathbf{I}, \quad \mathbf{G}^{p\top}\mathbf{G}^p = \mathbf{I}. \end{aligned} \tag{7}$$

3.4. Clustering over Dual Networks Toward Local Consistency

Apart from global consistency, we further utilize neighborhood-preserving matching to explore the consensus over dual networks. In particular, neighborhood preserving assumes that the same users in different networks may have similar ego networks and social circles in nature [Jiang et al. 2014; Yang et al. 2014; Zhang et al. 2013, 2016a]. In this work, for each user, we contextualize the social circles by his or her nearest neighbors. We formulate this idea as the problem of "local consistency."

For a given user i, we calculate his or her similarities with others based on Equation (1). Hereafter, we select the K users with highest similarities as the nearest neighbors to construct his or her online and offline social circles. We denote them as $\mathcal{N}_K(\mathbf{x}_i^v)$ and $\mathcal{N}_K(\mathbf{x}_i^p)$,

$$\begin{cases} \mathcal{N}_K(\mathbf{x}_i^v) = \{i_1^v, \ldots, i_K^v\} \\ \mathcal{N}_K(\mathbf{x}_i^p) = \{i_1^p, \ldots, i_K^p\} \end{cases}, \tag{8}$$

where i_k^v and i_k^p indicate the k-th nearest neighbor of the i-th user in online and offline settings, respectively. We thus can define two selection matrices, $\mathbf{S}_i^v = [S_i^v(j,k)] \in \mathbb{R}^{N\times K}$ and $\mathbf{S}_i^p = [S_i^p(j,k)] \in \mathbb{R}^{N\times K}$, for each user to select its neighbors based on Equation (8):

$$\begin{cases} S_i^v(j,k) = \begin{cases} 1, & \text{if } j = i_k^v \in \mathcal{N}_K(\mathbf{x}_i^v) \\ 0, & \text{otherwise} \end{cases}, \\ S_i^p(j,k) = \begin{cases} 1, & \text{if } j = i_k^p \in \mathcal{N}_K(\mathbf{x}_i^p) \\ 0, & \text{otherwise} \end{cases}. \end{cases} \tag{9}$$

We hence can obtain the social circles $\mathbf{X}_i^v \in \mathbb{R}^{D_v\times K}$ and $\mathbf{X}_i^p \in \mathbb{R}^{D_p\times K}$ of the user i as follows:

$$\begin{cases} \mathbf{X}_i^v = \mathbf{X}^v\mathbf{S}_i^v \\ \mathbf{X}_i^p = \mathbf{X}^p\mathbf{S}_i^p \end{cases}. \tag{10}$$

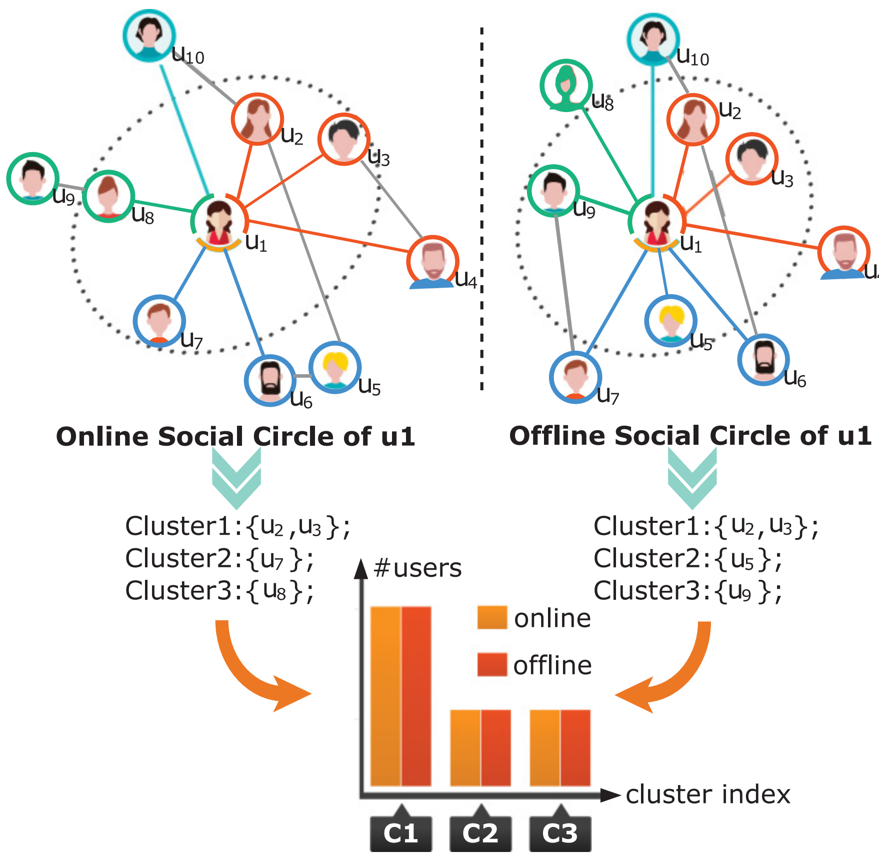


Fig. 2. Illustration of local consistency over dual networks for a specific user u_1. Dashed circles represent his or her social circles with the size of four, and different colored edges denote different communities.

Given a specific user i, and his or her local social circles $\mathcal{N}_K(\mathbf{x}_i^v)$ and $\mathcal{N}_K(\mathbf{x}_i^p)$, we can select the corresponding K distribution vectors over C clusters from $\mathbf{G}^v$ and $\mathbf{G}^p$, and we then can define the local scaled cluster assignment matrices $\mathbf{G}_i^v \in \mathbb{R}^{K\times C}$ and $\mathbf{G}_i^p \in \mathbb{R}^{K\times C}$ as follows:

$$\begin{cases} \mathbf{G}_i^v = \mathbf{S}_i^{v\,T}\mathbf{G}^v \\ \mathbf{G}_i^p = \mathbf{S}_i^{p\,T}\mathbf{G}^p \end{cases}. \tag{11}$$

Given the social circles across dual-networks, as illustrated in Figure 2, we aim to regularize *local consistency*, which encourages the cluster distributions of a specific user's social circle over the C clusters to be close across dual networks. This amounts to enforcing online and offline contextualization of the specific user to be consistent in a local view.

It is noticeable that online and offline social circles of the same user may involve different neighbors, namely $\mathcal{N}_K(\mathbf{x}_i^v) \neq \mathcal{N}_K(\mathbf{x}_i^p)$. For instance, the k-th closest neighbors of the i-th user in dual networks are not necessarily the same person, and hence they can fall into different clusters. However, for the same cluster c, it is assumed to have similar weights over the K neighbors. In other words, the difference between $\sum_k^K G_i^v(k,c)$ and $\sum_k^K G_i^p(k,c)$ should be as small as possible. As such, we measure local consistency by

$$\begin{aligned} D_{local}(\mathbf{G}_i^v, \mathbf{G}_i^p) &= \sum_{c=1}^{C} \left(\sum_{k=1}^{K} \mathbf{G}_i^v(k,c) - \sum_{k=1}^{K} \mathbf{G}_i^p(k,c) \right)^2 \\ &= \left\| (\mathbf{G}_i^v - \mathbf{G}_i^p)^\top \mathbf{e} \right\|^2, \end{aligned} \tag{12}$$

where $\mathbf{e} \in \mathbb{R}^K$ is the all-ones' vector.

Furthermore, we can estimate the consistency of all users' online and offline social circles as follows:

$$D_{local}(\mathbf{G}^v, \mathbf{G}^p) = \sum_{i=1}^{n} \left\| (\mathbf{G}_i^v - \mathbf{G}_i^p)^\top \mathbf{e} \right\|^2. \tag{13}$$

Analogous to the clustering over dual networks toward global consistency, we have the following joint optimization problem toward local consistency:

$$\begin{aligned} \min_{\mathbf{G}^v, \mathbf{G}^p} \; & tr(\mathbf{G}^{v\top} \mathcal{L}^v \mathbf{G}^v) + tr(\mathbf{G}^{p\top} \mathcal{L}^p \mathbf{G}^p) \\ & + \frac{\lambda_1}{2} \sum_{i=1}^{n} \left\| (\mathbf{G}_i^v - \mathbf{G}_i^p)^\top \mathbf{e} \right\|^2, \\ \text{subject to} \quad & \mathbf{G}^{v\top}\mathbf{G}^v = \mathbf{I}, \quad \mathbf{G}^{p\top}\mathbf{G}^p = \mathbf{I}. \end{aligned} \tag{14}$$

3.5. CLEVER Model Toward Global and Local Consistency

As discussed in Section 3.2, the objective functions of clustering over an individual network can be expressed in the form of $tr(\mathbf{G}^{v\top}\mathcal{L}^v\mathbf{G}^v)$ and $tr(\mathbf{G}^{p\top}\mathcal{L}^v\mathbf{G}^p)$, where $\mathcal{L}^v$ and $\mathcal{L}^p$ are the normalized Laplacian matrices defined based on the affinity matrices $\mathbf{A}^v$ and $\mathbf{A}^p$ in Equation (1), respectively. However, the conventional Laplacian matrix only considers the topological structure of the user networks and overlooks the discriminative community information. As proposed in Yang et al. [2010] and Yu and Shi [2003], the discriminative information is beneficial for exhibiting the manifold structure in the networks and has a positive influence on community detection performance. To refine the Laplacian matrices, we therefore resort to utilize the social circles across dual networks and further uncover the intrinsic manifold structures of social communities.

To explore the discriminative information in each social circle, we turn to the Fisher criterion [Friedman and Kandel 1999], which is widely used in image processing and computer vision. Intuitively, to obtain the better community partitions, the distance between users from different communities should be as large as possible, whereas the distance between users within the same community should be the smallest. Therefore, for a given social circle of the i-th user, we define the total scatter matrices $\mathbf{T}_i^v$ and $\mathbf{T}_i^p$, which sum up all pairwise distances between the near neighbors, and the between-cluster scatter matrices $\mathbf{B}_i^v$ and $\mathbf{B}_i^p$, which sum up all distances between near neighbors from different clusters:

$$\begin{cases} \mathbf{T}_i^v = \sum_{j \in \mathcal{N}_K(\mathbf{x}_i^v)} (\mathbf{x}_j^v - \mu_i^v)(\mathbf{x}_j^v - \mu_i^v)^\top & = \mathbf{X}_i^v \mathbf{H}\mathbf{H}\mathbf{X}_i^{v\top}, \\ \mathbf{B}_i^v = \sum_{c=1}^{C} n_c^v (\mu_i^v - \mu_i^{v,c})(\mu_i^v - \mu_i^{v,c})^\top & = \mathbf{X}_i^v \mathbf{H}\mathbf{G}_i^v \mathbf{G}_i^{v\top} \mathbf{H}\mathbf{X}_i^{v\top} \\ \mathbf{T}_i^p = \sum_{j \in \mathcal{N}_K(\mathbf{x}_i^p)} (\mathbf{x}_j^p - \mu_i^p)(\mathbf{x}_j^p - \mu_i^p)^\top & = \mathbf{X}_i^p \mathbf{H}\mathbf{H}\mathbf{X}_i^{p\top}, \\ \mathbf{B}_i^p = \sum_{c=1}^{C} n_c^p (\mu_i^p - \mu_i^{p,c})(\mu_i^v - \mu_i^{p,c})^\top & = \mathbf{X}_i^p \mathbf{H}\mathbf{G}_i^p \mathbf{G}_i^{p\top} \mathbf{H}\mathbf{X}_i^{p\top} \end{cases}, \tag{15}$$

where μ_i^v and μ_i^p respectively denote the means of the nearest neighbors in $\mathcal{N}_K(\mathbf{x}_i^v)$ and $\mathcal{N}_K(\mathbf{x}_i^p)$; $\mu_i^{p,c}$ and $\mu_i^{p,c}$ separately represent the mean of the nearest neighbors within the c-th cluster; $\mathbf{H} = \mathbf{I} - \frac{1}{K}\mathbf{E} \in \mathbb{R}^{K\times K}$ is a constant matrix; and $\mathbf{E} \in \mathbb{R}^{K\times K}$ is the all-ones' matrix, whereby K is the number of the nearest neighbors. To better characterize the underlying structure of local circles, we define a new concept, called the *local discriminant score*, as follows:

$$\begin{aligned} F(\mathbf{G}_i^v) &= tr(\mathbf{G}_i^{v\top}\mathbf{H}\mathbf{G}_i^v - (\mathbf{T}_i^v + \lambda\mathbf{I})^{-1}\mathbf{B}_i^v) \\ &= tr(\mathbf{G}^{v\top}\mathbf{S}_i^v\mathbf{H}(\mathbf{H}^\top\mathbf{X}_i^{v\top}\mathbf{X}_i^v\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}\mathbf{S}_i^{v\top}\mathbf{G}^v) \\ &= tr(\mathbf{G}^{v\top}\mathcal{L}_i^{v*}\mathbf{G}^v), \end{aligned} \tag{16}$$

where $\mathcal{L}_i^{v*} = \mathbf{S}_i^v\mathbf{H}(\mathbf{H}^\top\mathbf{X}_i^{v\top}\mathbf{X}_i^v\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}\mathbf{S}_i^{v\top}$. Similar to Equation (16), we have $F(\mathbf{G}_i^p) = tr(\mathbf{G}^{p\top}\mathcal{L}_i^{p*}\mathbf{G}^p)$, where $\mathcal{L}_i^{p*} = \mathbf{S}_i^p\mathbf{H}(\mathbf{H}^\top\mathbf{X}_i^{p\top}\mathbf{X}_i^p\mathbf{H}+\lambda\mathbf{I})^{-1}\mathbf{H}\mathbf{S}_i^{p\top}$. A smaller local discriminant score indicates that the neighbors in the social circles from different clusters are better separated.

Considering all users' social circles over dual networks, the relations $\mathbf{G}_i^v = \mathbf{S}_i^{v\,T}\mathbf{G}^v$, and $\mathbf{G}_i^p = \mathbf{S}_i^{p\,T}\mathbf{G}^p$ in Equation (11), we reformulate the conventional standard spectral clustering methods with the local discriminant model in Equation (16) as the following optimization problems:

$$\begin{cases} \min\limits_{\mathbf{G}^v} \sum_{i=1}^N F(\mathbf{G}_i^v) = \min\limits_{\mathbf{G}^v} tr(\mathbf{G}^{v\top}\mathcal{L}^{v*}\mathbf{G}^v) \\ \min\limits_{\mathbf{G}^p} \sum_{i=1}^N F(\mathbf{G}_i^p) = \min\limits_{\mathbf{G}^p} tr(\mathbf{G}^{p\top}\mathcal{L}^{p*}\mathbf{G}^p) \end{cases}, \tag{17}$$

where $\mathcal{L}^{v*} = \sum_{i=1}^N \mathcal{L}_i^{v*}$ and $\mathcal{L}^{p*} = \sum_{i=1}^N \mathcal{L}_i^{v*}$ can be viewed as the modified Laplacian matrices by employing discriminant information and manifold information over online and offline networks, respectively.

At last, integrating the refined spectral clustering framework in Equation (17), global consistency in Equation (6), and local consistency in Equation (13) over dual networks, we ultimately reach the objective function of our proposed CLEVER model as follows:

$$\begin{aligned} &\min_{\mathbf{G}^v,\mathbf{G}^p} tr(\mathbf{G}^{v\top}\mathcal{L}^{v*}\mathbf{G}^v) + tr(\mathbf{G}^{p\top}\mathcal{L}^{p*}\mathbf{G}^p) \\ &+\frac{\lambda_1}{2}\sum_{i=1}^n \left\|(\mathbf{G}_i^v - \mathbf{G}_i^p)^\top\mathbf{e}\right\|^2 + \frac{\lambda_2}{2}\left\|\frac{\mathbf{G}^v\mathbf{G}^{v\top}}{\|\mathbf{G}^v\|_F^2} - \frac{\mathbf{G}^p\mathbf{G}^{p\top}}{\|\mathbf{G}^p\|_F^2}\right\|_F^2, \\ &\text{subject to} \quad \mathbf{G}^{v\top}\mathbf{G}^v = \mathbf{I}, \quad \mathbf{G}^{p\top}\mathbf{G}^p = \mathbf{I}. \end{aligned} \tag{18}$$

Overall, our proposed CLEVER model has two parameters, as shown in Equation (18). Parameters λ_1 and λ_2 respectively regularize local and global consistency. We will detail the parameter tuning procedure in the experiments.

Based on the scaled assignment matrices $\mathbf{G}^v$ and $\mathbf{G}^p$, we can obtain the individual partition results corresponding to online and offline networks. Furthermore, we conducted the following equation on $\mathbf{G}^v$ and $\mathbf{G}^p$ to get the final unified community $\tilde{\mathbf{G}}$:

$$\tilde{\mathbf{G}} = \frac{1}{2}\tilde{\mathbf{G}}^v + \frac{1}{2}\tilde{\mathbf{G}}^p, \quad \tilde{\mathbf{G}}^v = \mathbf{G}^v(\mathbf{G}^v \otimes \mathbf{G}^v)^{-\frac{1}{2}}, \quad \tilde{\mathbf{G}}^p = \mathbf{G}^v(\mathbf{G}^p \otimes \mathbf{G}^p)^{-\frac{1}{2}}, \tag{19}$$

where $\tilde{\mathbf{G}}^v$ and $\tilde{\mathbf{G}}^p$ denote the assignment matrices with their element as the probability of the i-th user to the j-th community, and $\tilde{\mathbf{G}}$ represents the final unified assignment matrix that combines online and offline parts and stands for the final probability distributions over all communities. Thereafter, for each user, we can assign her or him to the community with the highest probability. From the final partition process, we can find that it can guarantee the exact and clear partition result. It is noted that by setting the flexible probability threshold, our framework can allow the users to drop into multiple communities simultaneously.

3.6. Optimization

We can equally transform the third term in Equation (18) (i.e., the local consistency term) into a trance-norm form as follows:

$$tr(\mathbf{G}^{v\top}\mathbf{U}^v\mathbf{G}^v) + tr(\mathbf{G}^{p\top}\mathbf{U}^p\mathbf{G}^p) - 2tr(\mathbf{G}^{p\top}\mathbf{U}^{v,p}\mathbf{G}^v), \tag{20}$$

where $\mathbf{U}^v = \sum_{i=1}^n \mathbf{S}_i^v\mathbf{E}\mathbf{S}_i^{v\top} \in \mathbb{R}^{N\times N}$, $\mathbf{U}^p = \sum_{i=1}^n \mathbf{S}_i^p\mathbf{E}\mathbf{S}_i^{p\top} \in \mathbb{R}^{N\times N}$, and $\mathbf{U}^{v,p} = \sum_{i=1}^n \mathbf{S}_i^p\mathbf{E}\mathbf{S}_i^{v\top} \in \mathbb{R}^{N\times N}$.

Similarly, we can restate the fourth term (i.e., the global consistency term) in Equation (18) with its equivalent trace-norm form as follows:

$$\frac{1}{C^2}tr(\mathbf{G}^v\mathbf{G}^{v\top}\mathbf{G}^v\mathbf{G}^{v\top} - 2\mathbf{G}^v\mathbf{G}^{v\top}\mathbf{G}^p\mathbf{G}^{p\top} + \mathbf{G}^p\mathbf{G}^{p\top}\mathbf{G}^p\mathbf{G}^{p\top}) = \frac{2}{C} - \frac{2}{C^2}tr(\mathbf{G}^v\mathbf{G}^{v\top}\mathbf{G}^p\mathbf{G}^{p\top}).$$

As such, by ignoring the constant additive and scaling terms, our objective function in Equation (18) can be rewritten as follows:

$$\begin{aligned} \min_{\mathbf{G}^v,\mathbf{G}^p} \mathcal{F}(\mathbf{G}^v, \mathbf{G}^p) = \min_{\mathbf{G}^v,\mathbf{G}^p} & tr(\mathbf{G}^{v\top}\tilde{\mathcal{L}}^v\mathbf{G}^v) + tr(\mathbf{G}^{p\top}\tilde{\mathcal{L}}^p\mathbf{G}^p) \\ & - \lambda_1 tr(\mathbf{G}^{p\top}\mathbf{U}^{v,p}\mathbf{G}^v) - \lambda_2 tr(\mathbf{G}^v\mathbf{G}^{v\top}\mathbf{G}^p\mathbf{G}^{p\top}), \\ \text{subject to} \quad & \mathbf{G}^{v\top}\mathbf{G}^v = \mathbf{I}, \quad \mathbf{G}^{p\top}\mathbf{G}^p = \mathbf{I}, \end{aligned} \tag{21}$$

where $\tilde{\mathcal{L}}^v = \mathcal{L}^{v*} + \frac{\lambda_1}{2}\mathbf{U}^v$ and $\tilde{\mathcal{L}}^p = \mathcal{L}^{p*} + \frac{\lambda_1}{2}\mathbf{U}^p$. They can be regarded as the new Laplacian matrices.

As can be seen, the preceding objective function is not easy to optimize because of its nonconvex terms. Even worse, the orthogonality constraints are not only nonconvex but also numerically expensive to preserve during iterations, although they play a key role in joint clustering. Therefore, it is unsuitable to transform the problem into the standard spectral clustering objective function [Kumar et al. 2011], which can be easily optimized with the eigenvalue decomposition method like Nie et al. [2011b]. However, Wen and Yin [2013] proposed using a Crank-Nicolson–like update scheme to preserve the constraints; based on it, they developed curvilinear search algorithms with lower per-iteration cost. Inspired by this, in this work we integrate this constraint-preserving update scheme with the alternative optimization strategy to solve $\mathbf{G}^v$ and $\mathbf{G}^p$.

We alternatively optimize one variable while fixing the other in each iteration. In particular, given an initial $\mathbf{G}^p_{(0)}$, we alternatively solve the following optimization problems,

$$\mathbf{G}^v_{(t+1)} = \arg\min_{\mathbf{G}^v} \mathcal{F}(\mathbf{G}^v, \mathbf{G}^p_{(t)}), \quad \text{subject to } \mathbf{G}^{v\top}\mathbf{G}^v = \mathbf{I}, \tag{22}$$

and

$$\mathbf{G}^p_{(t+1)} = \arg\min_{\mathbf{G}^p} \mathcal{F}(\mathbf{G}^v_{(t+1)}, \mathbf{G}^p), \quad \text{subject to } \mathbf{G}^{p\top}\mathbf{G}^p = \mathbf{I}, \tag{23}$$

until convergence, where t is the iteration counter. Since the optimization in Equations (22) and (23) are symmetric, we only show how to efficiently solve Equation (22), and the other can be solved in a similar way. The details are summarized in Algorithm 1.

The subroutine in Algorithm 1 is detailed in Algorithm 2. It accepts input $\mathbf{G}^v_{(t)}$ with $\mathbf{G}^p_{(t)}$ fixed and outputs $\mathbf{G}^v_{(t+1)}$ after the i-th iteration. Particularly, we apply the curvilinear search with the Barzilai-Borwein step method [Wen and Yin 2013] on Equation (22)

ALGORITHM 1: Alternative Optimization to Solve $\mathbf{G}^v$ and $\mathbf{G}^p$ with a Given Cluster Number C and Local Circle Size K

Input: Objective function $\mathcal{F}(\mathbf{G}^v, \mathbf{G}^p)$,
parameters $\varphi = \{\lambda_1, \lambda_2, \rho, \eta, \delta, \tau, \tau_m, \tau_M\}$
Output: $\mathbf{G}^v$, $\mathbf{G}^p$
1: Initialize $t = 0$;
2: Initialize $\mathbf{G}^v_{(t)}$, $\mathbf{G}^p_{(t)}$ with random cluster assignment;
3: **while** $\mathbf{G}^v_{(t)}$ and $\mathbf{G}^p_{(t)}$ do not converge **do**
4: update $\mathbf{G}^v_{(t+1)}$ with $\mathbf{G}^p_{(t)}$ fixed by Algorithm (2);
5: update $\mathbf{G}^p_{(t+1)}$ with $\mathbf{G}^v_{(t+1)}$ fixed by Algorithm (2);
6: update $t \leftarrow t + 1$;
7: **end while**
8: Return $\mathbf{G}^v \leftarrow \mathbf{G}^v_{(t)}$ and $\mathbf{G}^p \leftarrow \mathbf{G}^p_{(t)}$;

ALGORITHM 2: Curvilinear Search Method

Input: $\mathbf{G}^v_{(t)}$ and function $\mathcal{F}$ according to Equation (22),
parameters $\varphi = \{\lambda_1, \lambda_2, \rho, \eta, \delta, \tau, \tau_m, \tau_M, \epsilon\}$
Output: $\mathbf{G}^v_{(t+1)}$
1: Initialize $j = 0$, $\mathbf{W}_j = \mathbf{G}^v_{(t)}$, $C_j = 0$ and $Q_j = 1$;
2: **while** $\|\nabla\mathcal{F}(\mathbf{W}_j)\| > \epsilon$ **do**
3: $\nabla\mathcal{F}(\mathbf{W}_j) = 2\tilde{\mathcal{L}}^v\mathbf{W}_j - \lambda_1\mathbf{U}^{v,p}\mathbf{G}^p_{(t)} - 2\lambda_2\mathbf{G}^p_{(t)}\mathbf{G}^{p\,\top}_{(t)}\mathbf{W}_j$;
$\mathbf{A}_j = \nabla\mathcal{F}(\mathbf{W}_j)\mathbf{W}_j^\top - \mathbf{W}_j\nabla\mathcal{F}(\mathbf{W}_j)^\top$;
$\mathbf{Y}_j(\tau_j) = (\mathbf{I} + \frac{\tau_j}{2}\mathbf{A}_j)^{-1}(\mathbf{I} - \frac{\tau_j}{2}\mathbf{A}_j)\mathbf{W}_j$;
4: **while** $\mathcal{F}(\mathbf{Y}_j(\tau)) \geq C_j + \rho\tau\mathcal{F}'_\tau(\mathbf{Y}_j(0))$ **do**
5: $\tau = \delta\tau$,
$\mathbf{Y}_j(\tau) = \left(\mathbf{I} + \frac{\tau}{2}\mathbf{A}_j\right)^{-1}\left(\mathbf{I} - \frac{\tau}{2}\mathbf{A}_j\right)\mathbf{W}_j$;
6: **end while**
7: update $\mathbf{W}_{j+1} \leftarrow \mathbf{Y}_j(\tau)$;
update $Q_{j+1} \leftarrow \eta Q_j + 1$;
update $C_{j+1} \leftarrow \frac{(\eta Q_j C_j + \mathcal{F}(\mathbf{W}_{j+1}))}{Q_{j+1}}$;
update $\tau_{j+1,1} = \frac{tr\left((\mathbf{W}_{j+1}-\mathbf{W}_j)^\top(\mathbf{W}_{j+1}-\mathbf{W}_j)\right)}{\left|tr\left((\mathbf{W}_{j+1}-\mathbf{W}_j)^\top(\nabla\mathcal{F}(\mathbf{W}_{j+1})-\nabla\mathcal{F}(\mathbf{W}_j))\right)\right|}$;
8: set $\tau = \max\left\{\min\left\{\tau_{k+1,1}, \tau_M\right\}, \tau_m\right\}$;
9: update $j \leftarrow j + 1$;
10: **end while**
11: Return $\mathbf{G}^v_{(t+1)} \leftarrow \mathbf{W}_j$.

to update $\mathbf{G}^v_{(t)}$ until convergence. For simplicity, we construct new sequences $\{\mathbf{W}_j\}$ in each iteration to update $\mathbf{G}^v_{(t)}$ by using the curvilinear search method (CSM) as follows:

$$\mathbf{W} = \arg\min_{\mathbf{W}} \left\{\mathcal{F}(\mathbf{W}) = \mathcal{F}(\mathbf{W}, \mathbf{G}^p_{(t)})\right\}. \tag{24}$$

Given a feasible point $\mathbf{W}_j$ at the j-th iteration of *CSM*, $\mathbf{W}_{j+1}$ can be generated by the new trail point $\mathbf{Y}_j(\tau_j)$ according to the proposed approach in Wen and Yin [2013],

$$\mathbf{W}_{j+1} = \mathbf{Y}_j(\tau_j), \tag{25}$$

Table I. Statistics of the Regional Meetup Dataset

City	Users (#)	Interest Tags (#)	Events (#)	Avg. Tags Per User	Avg. Events Per User
California (CA)	5,904	2,664	6,106	23.42	13.55
New York City (NYC)	6,440	2,630	7,054	17.44	12.79

where $\mathbf{Y}_j(\tau_j)$ is determined by the Crank-Nicolson–like scheme via utilizing the gradient $\nabla\mathcal{F}(\mathbf{W}_j)$ and a skew-symmetric matrix $\mathbf{A}_j$,

$$\begin{cases} \nabla\mathcal{F}(\mathbf{W}_j) = 2\tilde{\mathcal{L}}^v\mathbf{W}_j - \lambda_1\mathbf{U}^{v,p}\mathbf{G}^p_{(t)} - 2\lambda_2\mathbf{G}^p_{(t)}\mathbf{G}^{p\ \top}_{(t)}\mathbf{W}_j, \\ \mathbf{A}_j \qquad\quad = \nabla\mathcal{F}(\mathbf{W}_j)\mathbf{W}_j^{\top} - \mathbf{W}_j\nabla\mathcal{F}(\mathbf{W}_j)^{\top}, \\ \mathbf{Y}_j(\tau_j) \quad\ = (\mathbf{I} + \frac{\tau_k}{2}\mathbf{A}_j)^{-1}(\mathbf{I} - \frac{\tau_j}{2}\mathbf{A}_j)\mathbf{W}_j, \end{cases} \tag{26}$$

where τ_j is the linear combination of the Barzilai-Borwein step size.

4. EXPERIMENTS

All experiments were conducted over a server equipped with an Intel Core i7-4970 CPU at 3.60GHz on 32G RAM, four cores, and a 64-bit Windows 10 operating system.

4.1. Data Description

We conducted experiments on Meetup,[7] a publicly benchmarked dataset [Pham et al. 2015]. We utilized user profiles consisting of personal interest labels and event attendance records to construct the online and offline networks, respectively. We detail the constructions of these two networks as follows:

(1) We extracted Meetup users from the dataset and obtained 9,095 and 8,339 original users with their profiles and event attendance records from the state of California (CA) and New York City (NYC), respectively.
(2) We filtered out users with fewer than five interest tags and five event attendance records to guarantee that all remaining users are relatively active across dual networks. We ultimately obtained 5,904 and 6,440 users in CA and NYC, respectively. The details are summarized in Table I.
(3) We characterized the user online and offline representations (i.e., $\mathbf{X}^v$ and $\mathbf{X}^p$) by using their interest tags and physical event attendance records, respectively. In particular, for the i-th user in CA, we transformed her or his distribution over all interest tags and events into 2,664- and 6,106-dimensional feature vectors to separately represent $\mathbf{x}_i^v$ and $\mathbf{x}_i^p$, whereas there were 2,630- and 7,054-dimensional vectors for the users in NYC.
(4) After obtaining the original user representations, we constructed dual networks (i.e., $\mathbf{A}^v$ and $\mathbf{A}^p$) where we treated users as vertices and their online and offline pairwise relationships as edges. In particular, we leveraged the pairwise similarities based on Equation (1) to stand for the relationships.

To quantitatively evaluate the community detection performance by different methods on dual networks, we adopted the following metrics:

(1) The normalized Davies-Bouldin index (ndbi) [Davies and Bouldin 1979] measures the uniqueness of clusters with respect to the unified similarity measure [Liu et al. 2012; Cheng et al. 2013; Zhou and Liu 2013],

$$ndbi(\mathcal{P}) = \frac{\sum_{i=1}^{C}\min_{j\neq i}\frac{d(\mathbf{c}_i,\mathbf{c}_j)+d(\mathbf{c}_j,\mathbf{c}_i)}{\sigma_i+\sigma_j+d(\mathbf{c}_i,\mathbf{c}_j)+d(\mathbf{c}_j,\mathbf{c}_i)}}{C}, \tag{27}$$

[7] https://www.ntu.edu.sg/home/gaocong/datacode.htm.

where $\mathbf{c}_i$ is the centroid of community $\mathcal{P}_i \in \mathcal{P}$, $d(\mathbf{c}_i, \mathbf{c}_j)$ is the distance between centroids $\mathbf{c}_i$ and $\mathbf{c}_j$, and σ_i is the average distance between elements in $\mathcal{P}_i$ and centroid $\mathbf{c}_i$. A higher ndbi value indicates a more cohesive community.

(2) The silhouette index (sil) [Rousseeuw 1987] validates clustering performance based on the pairwise difference of between- and within-cluster distances [Liu et al. 2010; Cheng et al. 2013],

$$sil(\mathcal{P}) = \frac{1}{C}\sum_{i=1}^{C}\left(\frac{1}{|\mathcal{P}_i|}\sum_{u\in\mathcal{P}_i}\frac{b(u)-a(u)}{\max\{b(u), a(u)\}}\right), \tag{28}$$

where $a(u) = \frac{1}{|\mathcal{P}_i|-1}\sum_{v\in\mathcal{P}_i, u\neq v} d(u, v)$ and $b(u) = \min_{j\neq i}(\frac{1}{|\mathcal{P}_j|}\sum_{v\in\mathcal{P}_j} d(u, v))$. A higher sil corresponds to a better clustering result.

(3) Normalized mutual information (nmi) [Nguyen and Caruana 2007; Li et al. 2014], as a consensus metric, gives the mutual information between online and offline clustering results normalized by the cluster entropies [Kumar et al. 2011; Nguyen and Caruana 2007; Ni et al. 2015; Cheng et al. 2013]. We utilized it to measure similarities of online and offline communities. As such,

$$nmi(\mathcal{P}^v, \mathcal{P}^p) = \frac{mi(\mathcal{P}^v, \mathcal{P}^p)}{\sqrt{H(\mathcal{P}^v)H(\mathcal{P}^p)}}, \tag{29}$$

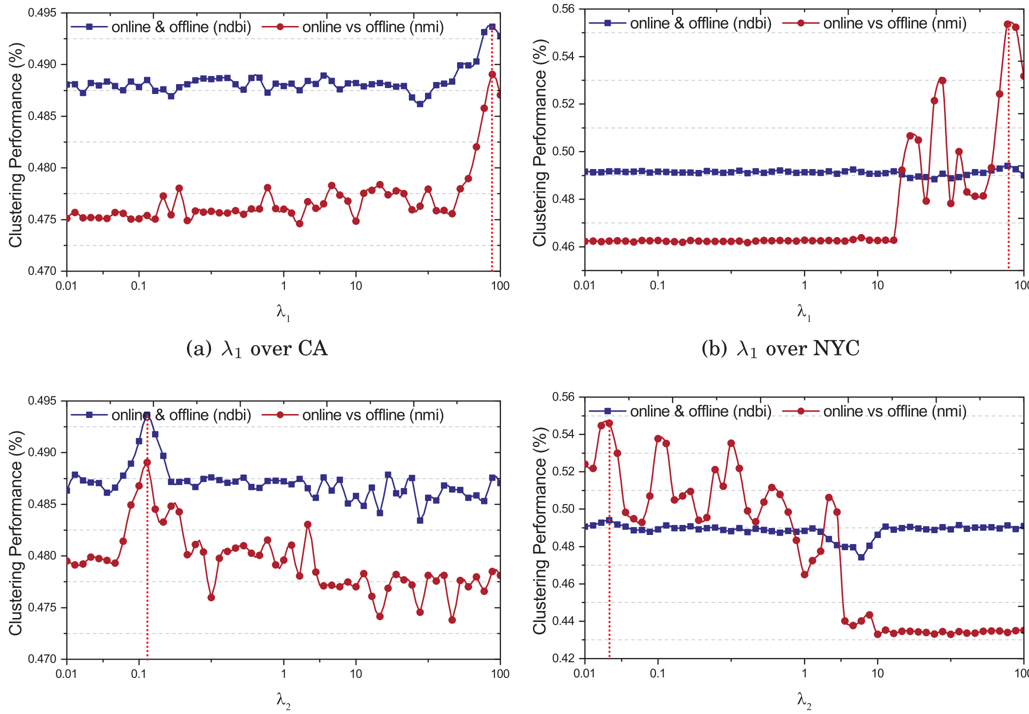
where $mi(\mathcal{P}^v, \mathcal{P}^p) = \sum_{i,j=1}^{C} P(i, j)\log\frac{P(i,j)}{P(i)P(j)}$ denotes the mutual information between $\mathcal{P}^v$ and $\mathcal{P}^p$, and $H(\mathcal{P}^v) = -\sum_{i=1}^{C} P(i)\log P(i)$ represents the entropy of $\mathcal{P}^v$. We defined $P(i) = \frac{|\mathcal{P}_i^v|}{N}$ and $P(i, j) = \frac{|\mathcal{P}_i^v\cap\mathcal{P}_j^p|}{N}$. The value of nmi ranges between 0 and 1, with a higher value indicating a closer match between online and offline clustering results.

It is noted that the solution of CLEVER is not unique, as the step size of our algorithm is adaptive for different initialization values and global minimization cannot be guaranteed. To study the sensitivity of our model to the initialization, we randomly generated 10 different initialization values (i.e., $\mathbf{G}_{(0)}^v$ and $\mathbf{G}_{(0)}^p$) and then fed them into our CLEVER model. For other competitors, the initialization procedure is analogous to ensure a fair comparison. Thereafter, we performed paired t-tests between our model and each of baselines over the 10-round results.

4.2. Parameter Tuning and Sensitivity

We have two implicit parameters, namely the number of communities C and the size of social circle K, as shown in Equation (21). Since there is no prior knowledge of community structures in EBSNs, we verified our models and all baselines with different numbers of communities $C \in \{20, 40, 80, 100, 150, 200, 250, 300\}$. Regarding the size of social circle K, it plays a pivotal role in the local consistency term and affects the refined Laplacian matrices as shown in Equations (13) and (17). In this work, we set various sizes of social circle $K \in [10, 100]$ with step size 10.

Apart from the implicit parameters, we have two explicit parameters λ_1 and λ_2 in Equation (21). Grid search was adopted to select the optimal parameters between 10^{-2} to 10^2 with small but adaptive step sizes. In particular, the step size was set to 0.01, 0.05, 1, and 5 for the ranges [0.01, 0.1], [0.1, 1], [1, 10], and [10, 100], respectively. The parameters corresponding to the best ndbi were used to report the final results. For other comparing systems, the procedures to tune the parameters are analogous to ensure fair comparison.



(a) λ_1 over CA (b) λ_1 over NYC

(c) λ_2 over CA (d) λ_2 over NYC

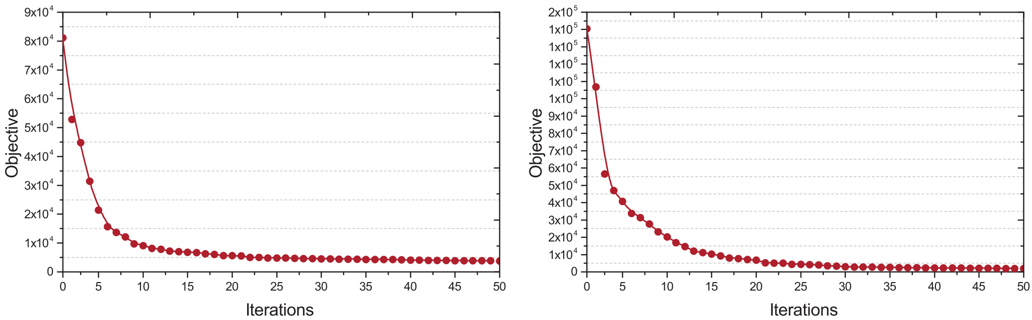
Fig. 3. Illustration of parameter tuning over CA and NYC datasets by varying one and fixing the other. The optimal setting of each parameter is marked by the red dotted line. We can see that the performance regarding ndbi and nmi is nonsensitive near the optimal parameters.

Taking the tuning procedure over the CA dataset as an example, we observed that our model reached optimal performance when $\lambda_1 = 0.15$ and $\lambda_2 = 95$. Figure 3 illustrates the performance of our model over CA and NYC datasets regarding these two parameters, respectively. Jointly analyzing the four subfigures, we can safely draw two conclusions. First, the performance of our proposed model changes within small ranges near the optimal settings justifies that our model is not sensitive to the parameters around their optimal configuration. Second, the curves of ndbi and nmi have similar trends, which further indicates that the consensus of online and offline networks has an immediate impact on community detection.

4.3. Convergence Analysis

It is shown that the iterative solver of our proposed CLEVER, namely the curvilinear search with Barzilar-Borwein step method, can guarantee the iterations to converge to a stationary point [Wen and Yin 2013]. It therefore is a convergent monotone curvilinear search algorithm.

Moreover, we recorded the values of the objective function in Equation (18) along with the iteration times using the optimal parameter settings. Since each of the steps in Algorithms 1 and 2 decreases the objective function that has a lower bound of 0, the convergence of the alternating optimization is guaranteed [Gao et al. 2013b]. Figure 4 shows the convergence process with respect to the number of iterations.



(a) CLEVER on CA (b) CLEVER on NYC

Fig. 4. Convergence process illustration of the CLEVER model on CA and NYC, respectively. We can see that the most effective updates occurred in the first 20 iterations.

4.4. Scalability Discussion

Due to the increasing popularity of EBSNs, scalability has become a major concern regarding community detection across dual networks. We hence discuss the computational cost of each iteration in our proposed model and illustrate its efficiency in this section.

The computational cost in each iteration of Algorithm 2 mainly comes from three parts:

(1) *Computation of* $\nabla\mathcal{F}(\mathbf{W}_j)$: It is worth mentioning that the cost of the specific matrix multiplication $\mathbf{G}^p_{(t)}\mathbf{G}^{p\top}_{(t)}$ and that of constructing the constant matrices $\tilde{\mathcal{L}}^v$ and $\mathbf{U}^{v,p}$ remain the same for all iterations in the curvilinear search method. We thus can save much practical time costs by caching the results. The complexity of $\nabla\mathcal{F}(\mathbf{W}_j)$ is $O(N^2C)$ since the matrix multiplication $\mathbf{G}^p_{(t)}\mathbf{G}^{p\top}_{(t)}\mathbf{W}_j$.
(2) *Computation of* $\mathbf{A}_j$: The time cost of $\mathbf{A}_j$ is $O(N^2C)$.
(3) *Computation of* $\mathbf{Y}_j(\tau_j)$: The speed bottleneck of computing $\mathbf{Y}_j(\tau_j)$ lies in computing the inversion of $(\mathbf{I}+\frac{\tau}{2}\mathbf{A}_j)$, which has the complexity of $O(N^3)$. We thus can estimate the total cost for deriving $\mathbf{Y}_j(\tau_j)$ is $O(N^3)$.

To sum up, the computational complexity of our model is $O(N^3)$, which indeed reflects the unsatisfactory scalability. However, using the method of Coppersmith and Winograd, the cost can be bounded by $O(N^{2.376})$ [Zhai et al. 2012]. Although the computational complexity of our algorithm is the major bottleneck, the scale of users involved into one region or one local community cannot be very large in reality due to the restrictions of the regional population capacity. According to the running time of different methods in Table II, we can see that the computational cost of our proposed CLEVER is acceptable. Furthermore, we can implement a parallel and distributed version of our algorithm to speed up it [Ghoting et al. 2011; Chierichetti et al. 2014]. For example, we can distribute users within the same regions to the same compute node via MapReduce and furthermore optimize the matrix multiplications in Algorithm 2.

We evaluated the scalability of our proposed CLEVER algorithm by varying the scale of users in dual networks. As Figure 5 illustrates, CLEVER is efficient and capable of handling relative large graphs with massive vertices in dual networks.

4.5. Overall Model Performance Evaluation

To justify the effectiveness of our proposed CLEVER model, we compared it to several state-of-the-art competitors:

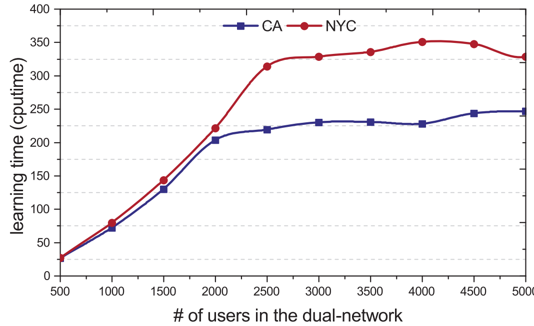


Fig. 5. Scalability of the CLEVER model by varying the number of users over dual networks in CA and NYC, respectively.

(1) *PCAKM*: K-means is a traditional clustering method for detecting communities [Qi et al. 2012]. In this experiment, to boost performance and improve efficiency of community detection, we first applied PCA on the original feature spaces and then adopted k-means to partition online and offline networks, separately.
(2) *SC*: Spectral clustering [Ng et al. 2001] is one of the most popular modern clustering algorithms, which can detect communities over a single network. In this experiment, we separately employed this method on the standard online and offline Laplacian matrices as shown in Equation (5).
(3) *RMKMC*: The novel and robust multiview k-means clustering method proposed in Cai et al. [2013] is able to uncover the consensus pattern and detect communities across multiple networks. We applied it on dual networks and obtained a shared clustering result.
(4) *MMSC*: This multimodal spectral clustering method introduced in Cai et al. [2011] learns a commonly shared graph Laplacian matrix by unifying different views, where each modal stands for a feature type from one single view. We applied it on dual networks and obtained a shared clustering result.
(5) *CoNMF*: CoNMF is a coregularized NMF model presented in He et al. [2014] that extends NMF for multiview clustering [Liu et al. 2012]. It jointly factorizes the multiple matrices through coregularization. We applied it on dual networks and then obtained a shared clustering result.

Table II summarizes the performance comparison between our model and the baselines in terms of ndbi, nmi, and sil measures. It is worthwhile highlighting that PCAKM, SC, and our model respectively output two partition results corresponding to online and offline networks; nevertheless, the outputs of RMKMC, MMSC, and CoNMF are shared unified partitions across dual networks. Due to the output nature, not all methods can be measured by the three metrics. From Table II, we have the following observations. First, with regard to the separate partition results over online and offline networks, CLEVER achieves the best performance in terms of ndbi across two cities. This clearly demonstrates that information encoded in different networks is complementary and able to reinforce the separate partitions. Second, our proposed model performs better than PCAKM and SC in terms of nmi, which considers all possible community matching between online and offline community detection results. This justifies that our local and global consistency constrains are reasonable. Third, our CLEVER model is superior to RMKMC, MMSC, and CoNMF models in terms of ndbi and sil measures. To compare to them, we linearly fused the community detection results of our model on online and offline networks. Noticeably, we are unable to fuse the clustering results of PCAKM and SC, as it is hard to align their clusters across dual

Table II. Performance Comparison Among Various Methods in Terms of ndbi, nmi, and sil

Method		PCAKM	SC	RMKMC	MMSC	CoNMF	**CLEVER**
CA	Online (ndbi)	$0.4700 \pm 2e\text{-}5$	$0.4571 \pm 8e\text{-}6$	$0.4472 \pm 1e\text{-}5$	$0.3886 \pm 3e\text{-}4$	$0.4152 \pm 7e\text{-}5$	**0.4719 ± 7e-6**
	Offline (ndbi)	$0.4506 \pm 1e\text{-}4$	$0.4600 \pm 6e\text{-}6$	$0.4732 \pm 2e\text{-}5$	$0.4102 \pm 5e\text{-}4$	$0.4424 \pm 5e\text{-}5$	**0.4756 ± 8e-6**
	Online vs. offline (nmi)	$0.3844 \pm 3e\text{-}5$	$0.3763 \pm 6e\text{-}6$	1	1	1	**0.4891 ± 1e-3**
	Online & offline (ndbi)	—	—	$0.4830 \pm 1e\text{-}5$	$0.4129 \pm 4e\text{-}4$	$0.4399 \pm 7e\text{-}5$	**0.4937 ± 6e-6**
	Online & offline (sil)	—	—	$-0.3790 \pm 4e\text{-}4$	$-0.1228 \pm 3e\text{-}4$	$-0.2569 \pm 2e\text{-}4$	**−0.1427 ± 3e-3**
	Cputime	$4.2993e2$	$2.5074e2$	$4.1844e2$	$5.1466e2$	$5.6756e4$	**4.7297e2**
	p-Value	—	—	$2.7334e\text{-}6$	$4.5006e\text{-}7$	$2.1256e\text{-}8$	-
NYC	Online (ndbi)	$0.4670 \pm 3e\text{-}5$	$0.4566 \pm 7e\text{-}6$	$0.4463 \pm 6e\text{-}5$	$0.3726 \pm 4e\text{-}4$	$0.4475 \pm 5e\text{-}5$	**0.4736 ± 7e-6**
	Offline (ndbi)	$0.3297 \pm 1e\text{-}4$	$0.4620 \pm 2e\text{-}5$	$0.4588 \pm 7e\text{-}5$	$0.3741 \pm 5e\text{-}4$	$0.4536 \pm 4e\text{-}5$	**0.4748 ± 8e-6**
	Online vs. offline (nmi)	$0.3375 \pm 1e\text{-}5$	$0.4335 \pm 1e\text{-}5$	1	1	1	**0.5536 ± 5e-3**
	Online & offline (ndbi)	—	—	$0.4705 \pm 6e\text{-}5$	$0.3856 \pm 5e\text{-}4$	$0.4669 \pm 4e\text{-}5$	**0.4940 ± 7e-6**
	Online & offline (sil)	—	—	$-0.3455 \pm 4e\text{-}4$	$-0.1384 \pm 3e\text{-}3$	$-0.2542 \pm 4e\text{-}4$	**−0.1169 ± 3e-4**
	Cputime	$4.6399e2$	$2.5966e2$	$4.8520e2$	$4.8902e2$	$5.9326e4$	**5.1588e2**
	p-Value	—	—	$5.1240e\text{-}4$	$6.5857e\text{-}8$	2.7505-6	—

Note: We reported the results with variance on two cities (CA and NYC) with $C = 300$ and $K = 20$. Thereinto, we illustrate the learning performance regarding various metrics over different networks, such as denoting the performance withe respect to ndbi over the unified dual networks as "Online & offline (ndbi)." The significance test is based on the ndbi over online and offline networks.

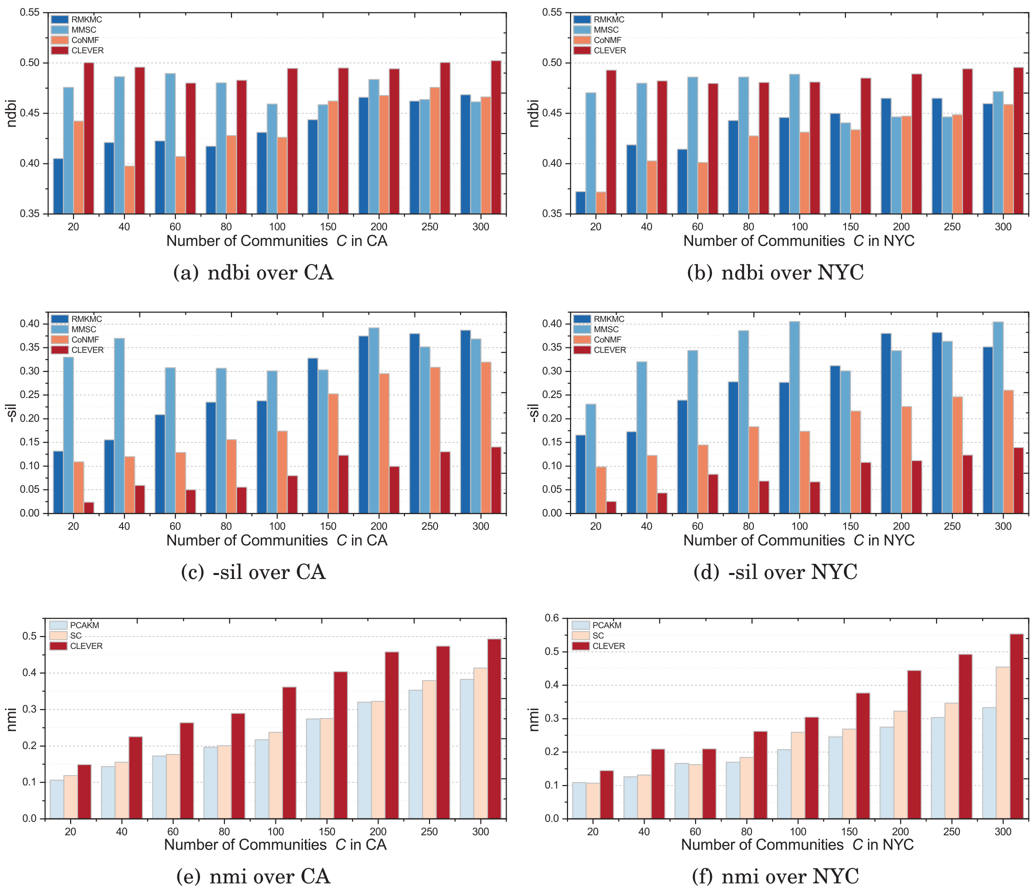
networks. Although RMKMC, MMSC, and CoNMF consider global consistency across the networks, they ignore local consistency. This further confirms the importance of local consistency in community detection over dual networks. Fourth, partition performance over CA is stably better than NYC, which implies that online and offline behaviors of CA citizens are more consistent.

In addition, we compared our method to all baselines by varying the number of communities. As illustrated in Figure 6, our model significantly outperforms the other baselines over the two cities with respect to sil and nmi. Meanwhile, our model can achieve remarkable and comparable performance with regard to ndbi when $C \leq 100$ while showing its superiority with the increasing number of communities C. In addition, we analyzed the sensitivity of our model to the social circle size on the NYC data. The results are demonstrated in Figure 7. It is observed that our model is nonsensitive to the social circle size K. Moreover, we observed that the performance of unified community detection significantly outperforms these separate ones, and this observation reflects the essential importance of unifying virtual and physical networks.

4.6. Componentwise Model Evaluation

The main insight of our proposed CLEVER model is to conduct dual clustering over dual networks. Its significant features are to preserve global consistency of networks and local consistency of social circles, as well as to refine the novel Laplacian matrices. In addition, it enables two flexible community structures rather than one over dual networks. Hence, to investigate the significance of each component, we employed the componentwise model evaluation. In particular, we disabled some terms of our objective function in Equation (21) as follows:

(1) *Separate*: To verify that the joint modeling indeed helps, we overlooked all consensus terms by setting $\lambda_1 = \lambda_2 = 0$. Actually, the method is similar to SC, in which we conduct spectral clustering over individual networks separately.
(2) *Equal*: In this method, we verified that the exactly equal community structures somewhat degrease learning performance due to the overblown consensus. In particular, we set $\mathbf{G}^v = \mathbf{G}^p$, which is equivalent to set $\lambda_2 \rightarrow \infty$.
(3) *Global*: In this method, we only considered global consistency over dual networks by setting $\lambda_1 = 0$ in Equation (21).

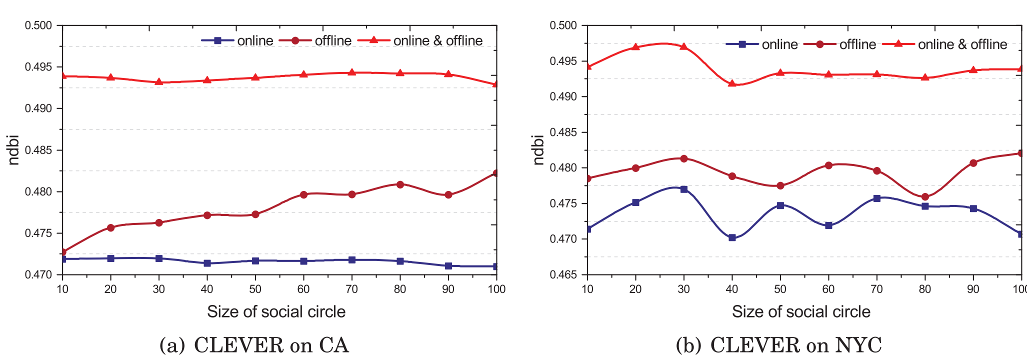


(a) ndbi over CA (b) ndbi over NYC

(c) -sil over CA (d) -sil over NYC

(e) nmi over CA (f) nmi over NYC

Fig. 6. Performance comparison among various methods by varying the number of communities C. The partition performance over CA and NYC is respectively measured in terms of ndbi, sil, and nmi metrics.



(a) CLEVER on CA (b) CLEVER on NYC

Fig. 7. Performance of the CLEVER model in terms of ndbi by varying the social circle size K on CA and NYC. Thereinto, ndbi represents learning performance with regard to ndbi over online, offline, and unified networks, respectively.

Table III. Effectiveness Evaluation of Global Consistency, Local Consistency, and the Refined Laplacian Matrix in Our Proposed CLEVER Model over Online and Offline Networks of Meetup in CA and NYC When $C = 300$ and $K = 20$

	Method	Separate	Equal	Global	Local	CLEVER-LM	CLEVER
CA	Online (ndbi)	$0.4610 \pm 8e\text{-}6$	$0.4669 \pm 6e\text{-}6$	$0.4713 \pm 3e\text{-}5$	$0.4715 \pm 4e\text{-}3$	$0.4643 \pm 1e\text{-}5$	**$0.4719 \pm 7e\text{-}6$**
	Offline (ndbi)	$0.4621 \pm 6e\text{-}6$	$0.4737 \pm 1e\text{-}5$	$0.4744 \pm 2e\text{-}5$	$0.4738 \pm 6e\text{-}4$	$0.4738 \pm 5e\text{-}6$	**$0.4756 \pm 8e\text{-}6$**
	Online vs. offline (nmi)	$0.4034 \pm 6e\text{-}6$	1	$0.4030 \pm 6e\text{-}6$	$0.4018 \pm 7e\text{-}4$	$0.4810 \pm 4e\text{-}2$	**$0.4891 \pm 1e\text{-}3$**
	Online & offline (ndbi)	$0.4785 \pm 5e\text{-}6$	$0.4805 \pm 1e\text{-}6$	$0.4894 \pm 1e\text{-}4$	$0.4896 \pm 1e\text{-}5$	$0.4893 \pm 9e\text{-}6$	**$0.4937 \pm 6e\text{-}6$**
	Online & offline (sil)	$-0.1898 \pm 2e\text{-}5$	$-0.2378 \pm 6e\text{-}5$	$-0.1979 \pm 5e\text{-}4$	$-0.1954 \pm 3e\text{-}4$	$-0.1701 \pm 8e\text{-}4$	**$-0.1427 \pm 3e\text{-}3$**
	p-Value	$3.1136e\text{-}6$	$7.1420e\text{-}7$	$1.3722e\text{-}6$	$3.1496e\text{-}6$	$3.9276e\text{-}5$	-
NYC	Online (ndbi)	$0.4579 \pm 7e\text{-}6$	$0.4720 \pm 1e\text{-}4$	$0.4677 \pm 8e\text{-}6$	$0.4706 \pm 2e\text{-}5$	$0.4655 \pm 8e\text{-}6$	**$0.4736 \pm 7e\text{-}6$**
	Offline (ndbi)	$0.4616 \pm 2e\text{-}5$	$0.4733 \pm 5e\text{-}5$	$0.4746 \pm 6e\text{-}6$	$0.4751 \pm 2e\text{-}4$	$0.4733 \pm 6e\text{-}6$	**$0.4748 \pm 8e\text{-}6$**
	Online vs. offline (nmi)	$0.4513 \pm 1e\text{-}5$	1	$0.4506 \pm 4e\text{-}4$	$0.4503 \pm 5e\text{-}4$	$0.4889 \pm 6e\text{-}6$	**$0.5536 \pm 5e\text{-}3$**
	Online & offline (ndbi)	$0.4883 \pm 3e\text{-}5$	$0.4900 \pm 4e\text{-}5$	$0.4905 \pm 5e\text{-}6$	$0.4910 \pm 3e\text{-}5$	$0.4910 \pm 2e\text{-}5$	**$0.4940 \pm 7e\text{-}6$**
	Online & offline (sil)	$-0.1258 \pm 1e\text{-}6$	$-0.1319 \pm 5e\text{-}5$	$-0.1584 \pm 4e\text{-}6$	$-0.1447 \pm 1e\text{-}4$	$-0.1284 \pm 1e\text{-}5$	**$-0.1169 \pm 3e\text{-}4$**
	p-Value	$1.3872e\text{-}7$	$1.5607e\text{-}5$	$9.2401e\text{-}5$	$7.366e\text{-}6$	$5.1167e\text{-}6$	—

Note: We also provide the variance. The significance test is based on the ndbi over online and offline networks.

(4) *Local*: In this method, we only took local consistency over dual networks into consideration. This is accomplished by setting $\lambda_2 = 0$ in Equation (21).

(5) *CLEVER-LM*: To verify the feasibility and efficacy of our refined Laplacian matrix in Equation (17), we replaced it with the standard Laplacian matrix in Equation (5).

Table III displays results of the componentwise evaluation of the CLEVER method. From the table, we can make the following observations. First, no matter what type of consistency terms we dropped, it hurt the performance of our model. This verifies the importance of global and local consistency over dual networks. Second, the Separate method achieves the worst performance. This signals that the joint modeling plays a pivotal role, which can effectively transfer the underlying information between two networks. Third, meanwhile, the Equal method has poor performance. It verifies that the exactly equal structures somewhat overemphasize the consensus and overlook the flexibility. Fourth, CLEVER-LM can obtain the comparable performance regarding ndbi. However, its clustering quality is worse in comparison with CLEVER. This confirms that the refined Laplacian matrices are more suitable than the standard ones. It hence verifies the effectiveness of the exploited manifold structures of social circles. We can therefore draw a safe conclusion that jointly modeling global and local consistency with the refined Laplacian matrices are beneficial to user contextualization and community detection over dual networks.

5. APPLICATION

Aside from the general validation, we also validated our model on a real-world problem: event attendance prediction [Liu et al. 2012; Foley et al. 2015]. Formally, given an event, a small set of early bird registers, a whole set of attendee candidates, and their online interaction records, we aim to predict who will probably attend this event and how many users will be present eventually. A good approach to this problem will remarkably facilitate a wide range of event organizers, such as workshop initiators, conference logistics chairs, and even presidential election committees, by enabling them to design better proactive coping strategies in advance.

However, it is nontrivial to predict attendance of the given event due to the following reasons. First, we have the *micro factor*, in which individual behaviors are often affected by one's social interactions [Cui et al. 2011]. To be more specific, if one's close friends confirm to attend a coming event, there exists a better than average chance that he or she will show up as well. Second, we have the *macro factor*, in which communities are usually interest driven [Ronen et al. 2014]. If the given event matches the taste of one community, the users within the community have higher probabilities of attending the event than those from unmatched communities. For instance, researchers from the data mining community prefer participating in the SIGIR conference over the SIGGRAPH conference. Moreover, traditional classifiers, such as SVM, suffer from insufficient training samples (i.e., usually, only a few early birds are available for a given event), and then they are hard to capture the information propagations among users. In this work, we adapted our CLEVER model to solve this problem, as it is able to uncover an optimal and cohesive community structure among users.

5.1. Experiments

5.1.1. Experimental Settings. Given a happened event e, we can build the ground truth by collecting its real attendees $\mathcal{U} = \{u_1, \ldots, u_{N_e}\}$, where N_e stands for the number of users attending e. To simulate the real case, we randomly selected M attendees from $\mathcal{U}$ to construct the set $\mathcal{M}$, who act as early bird registers. Based on the partition results of our model, assume that user u_i is assigned into the c-th community, which holds a set of members $\mathcal{P}_c$. We can estimate the probability of user u_i attending event e by jointly considering the macro and micro factors as follows:

$$\begin{aligned} p(u_i|\mathcal{M}, \mathcal{P}_c) &= p_{macro}(u_i|\mathcal{M}, \mathcal{P}_c) \times p_{micro}(u_i|\mathcal{M}, \mathcal{P}_c) \\ &= \left(\frac{1}{|\mathcal{P}_c|}\frac{|\mathcal{P}_{c,M}|}{|\mathcal{M}|}\right) \times \left(\frac{\sum_{u_j \in \mathcal{P}_{c,M}} sim(u_i, u_j)}{|\mathcal{P}_{c,M}|}\right) \\ &= \frac{\sum_{u_j \in \mathcal{P}_{c,M}} sim(u_i, u_j)}{|\mathcal{P}_c| \times |\mathcal{M}|}, \end{aligned} \tag{30}$$

where $\mathcal{P}_{c,M} = \mathcal{P}_c \cap \mathcal{M}$ denotes the attendees from the c-th community, and $p_{macro}(u_i|\mathcal{M}, \mathcal{P}_c)$ and $p_{micro}(u_i|\mathcal{M}, \mathcal{P}_c)$ denote the probability of u_i attending the event affected by the public environment and his or her social connections.

5.1.2. Event Attendance Prediction Evaluation. To measure the effectiveness of event attendance prediction from different angles, we adopted two widely used metrics. One is $F1$ [Li et al. 2014; Song et al. 2015a, 2016]:

$$F1 = 2\frac{precision \cdot recall}{precision + recall}; \tag{31}$$

the other is normalized mean squared error [Zhang et al. 2016b; He et al. 2016]:

$$nMSE = \frac{\sum_{e=1}^{N}(pred_e - real_e)^2}{N \times \bar{pred} \times \bar{real}}, \tag{32}$$

where N, $pred_e$, $real_e$, $\bar{pred}$, and $\bar{real}$ denote the number of testing events, the number of predicted attendees, the number of real attendees for event e, and the average number of predicted and real attendees, respectively.

We chose RMKMC, MMSC, and CoNMF as the baselines, as they output unified partition results over dual networks. Performance comparison in terms of $F1$ and $nMSE$ are reported in Tables IV and V, respectively. From Table IV, we observe that our model stably outperforms the baselines on two cities with different numbers of early bird registers. This actually reflects that the remarkable improvement of our model in predicting whether a user will attend the given event is much stronger. In other words, our model shows superiority in predicting who will attend the desired events in most cases. Table V illustrates that the number of predicted attendees by our model is much closer to the number of real attendees compared to other methods.

Table IV. Performance Comparison of Event Attendance Prediction Among Various Methods in Terms of $F1$ score by Varying the Number of Early Bird Registers, {10, ..., 50}, for the Given e

Method		RMKMC	MMSC	CoNMF	Separate	Equal	CLEVER
CA	$\|\mathcal{M}\| = 10$	$0.1587 \pm 4e$-4	$0.1103 \pm 6e$-5	$0.1372 \pm 9e$-5	$0.1619 \pm 1e$-5	$0.1623 \pm 2e$-5	**0.1832 ± 7e-4**
	$\|\mathcal{M}\| = 20$	$0.2810 \pm 1e$-6	$0.2023 \pm 6e$-5	$0.2220 \pm 2e$-4	$0.2244 \pm 1e$-5	$0.2462 \pm 4e$-5	**0.3216 ± 2e-3**
	$\|\mathcal{M}\| = 30$	$0.3290 \pm 5e$-5	0.2725 ± 1-4	$0.2737 \pm 4e$-4	$0.2643 \pm 2e$-5	$0.4126 \pm 5e$-5	**0.4252 ± 3e-3**
	$\|\mathcal{M}\| = 40$	$0.3530 \pm 8e$-5	$0.3208 \pm 2e$-4	$0.3101 \pm 2e$-4	$0.3638 \pm 2e$-4	$0.4874 \pm 7e$-4	**0.4968 ± 5e-3**
	$\|\mathcal{M}\| = 50$	$0.4719 \pm 6e$-5	$0.4615 \pm 1e$-4	$0.4316 \pm 5e$-4	$0.4289 \pm 2e$-4	$0.5055 \pm 2e$-4	**0.5639 ± 6e-3**
	p-value	$2.3881e$-3	$2.6238e$-5	$1.0288e$-5	$2.8812e$-3	$2.300e$-2	—
NYC	$\|\mathcal{M}\| = 10$	$0.1501 \pm 2e$-6	$0.1253 \pm 5e$-5	$0.1126 \pm 1e$-4	$0.1515 \pm 6e$-6	$0.1502 \pm 9e$-6	**0.1551 ± 3e-5**
	$\|\mathcal{M}\| = 20$	$0.2625 \pm 2e$-4	$0.2094 \pm 6e$-5	$0.1936 \pm 1e$-4	$0.2414 \pm 2e$-5	$0.2404 \pm 6e$-5	**0.2743 ± 2e-4**
	$\|\mathcal{M}\| = 30$	$0.3527 \pm 5e$-6	$0.2512 \pm 4e$-5	$0.2439 \pm 1e$-4	$0.3501 \pm 1e$-4	$0.3841 \pm 1e$-4	**0.3662 ± 8e-4**
	$\|\mathcal{M}\| = 40$	$0.4125 \pm 5e$-5	0.3775 ± 4-5	$0.38801 \pm 9e$-5	$0.3823 \pm 2e$-4	$0.4420 \pm 1e$-4	**0.4498 ± 1e-3**
	$\|\mathcal{M}\| = 50$	$0.4279 \pm 2e$-6	0.4104 ± 2-5	$0.4289 \pm 4e$-4	$0.4064 \pm 4e$-4	$0.4526 \pm 2e$-4	**0.5046 ± 1e-3**
	p-value	$3.4562e$-2	$2.2938e$-8	$5.4205e$-7	$5.0421e$-3	$2.8812e$-3	—

Note: We reported the results with variance on two cities (CA and NYC) with $C = 300$ and $K = 20$. The significance test is based on the $F1$ score at $|\mathcal{M}| = 50$.

Table V. Performance Comparison Among Various Methods for the Number of Attendees Prediction

	Event ID	Real Attendees(#)	RMKMC	MMSC	CoNMF	Equal	Separate	CLEVER
CA	1	108	44	20	112	38	37	**107**
	2	85	77	18	127	73	61	**98**
	3	105	69	19	106	63	49	**70**
	4	86	44	12	116	38	50	**81**
	5	82	44	10	150	64	56	**111**
	nMSE	—	0.3344	4.1138	0.1283	0.3668	0.4548	**0.0511**
NYC	6	113	46	44	130	37	49	**106**
	7	119	132	48	189	54	42	**123**
	8	120	138	39	171	65	39	**134**
	9	120	100	42	170	41	49	**164**
	10	116	75	37	196	47	47	**139**
	nMSE	—	0.1223	1.1616	0.1658	0.8374	0.9929	**0.0348**

Note: For each event listed, the number of early bird registers is 20.

Hence, our model achieves significant improvements in predicting how many attendees will be present eventually.

6. CONCLUSION AND FUTURE WORK

In this article, we present a novel dual clustering model for community detection over dual networks, which are an abstract expression of EBSNs in nature. This model is able to enhance network partition performance by integrating online and offline interactions among users and preserving local and global consistency across networks. In addition, it strengthens the model by an improved Laplacian definition. We have theoretically derived its solution. The model and its components were verified on two real-world datasets. Meanwhile, we applied our model to the application of event attendance prediction. It is worth emphasizing that our model is applicable to other applications, such as group recommendation and event suggestion.

In the future, we plan to extend our model to handle multiple networks instead of dual networks.

REFERENCES

Matthew B. Blaschko and Christoph H. Lampert. 2008. Correlational spectral clustering. In *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*. 1–8.

Brigitte Boden, Stephan Günnemann, Holger Hoffmann, and Thomas Seidl. 2012. Mining coherent subgraphs in multi-layer graphs with edge labels. In *Proceedings of the 2012 ACM Conference on Knowledge Discovery and Data Mining*. 1258–1266.

Eric Bruno and Stéphane Marchand-Maillet. 2009. Multiview clustering: A late fusion approach using latent models. In *Proceedings of the 2009 International Conference on Research and Development in Information Retrieval*. 736–737.

Xiao Cai, Feiping Nie, and Heng Huang. 2013. Multi-view k-means clustering on big data. In *Proceedings of the 2013 International Joint Conference on Artificial Intelligence*. 2598–2604.

Xiao Cai, Feiping Nie, Heng Huang, and Farhad Kamangar. 2011. Heterogeneous image feature integration via multi-modal spectral clustering. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*. 1977–1984.

Kamalika Chaudhuri, Sham M. Kakade, Karen Livescu, and Karthik Sridharan. 2009. Multi-view clustering via canonical correlation analysis. In *Proceedings of the 2009 International Conference on Machine Learning*. 129–136.

Wei Cheng, Xiang Zhang, Zhishan Guo, Yubao Wu, Patrick F. Sullivan, and Wei Wang. 2013. Flexible and robust co-regularized multi-domain graph clustering. In *Proceedings of the 2013 ACM Conference on Knowledge Discovery and Data Mining*. 320–328.

Yun Chi, Xiaodan Song, Dengyong Zhou, Koji Hino, and Belle L. Tseng. 2007. Evolutionary spectral clustering by incorporating temporal smoothness. In *Proceedings of the 2007 ACM Conference on Knowledge Discovery and Data Mining*. 153–162.

Flavio Chierichetti, Nilesh N. Dalvi, and Ravi Kumar. 2014. Correlation clustering in MapReduce. In *Proceedings of the 2014 International Conference on Knowledge Discovery and Data Mining*. 641–650.

Peng Cui, Fei Wang, Shaowei Liu, Mingdong Ou, Shiqiang Yang, and Lifeng Sun. 2011. Who should share what? Item-level social influence prediction for users and posts ranking. In *Proceedings of the 2011 International Conference on Research and Development in Information Retrieval*. 185–194.

David L. Davies and Donald W. Bouldin. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1, 2, 224–227.

Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. 2004. Kernel k-means: Spectral clustering and normalized cuts. In *Proceedings of the 2004 ACM Conference on Knowledge Discovery and Data Mining*. 551–556.

John Foley, Michael Bendersky, and Vanja Josifovski. 2015. Learning to extract local events from the Web. In *Proceedings of the 2015 International Conference on Research and Development in Information Retrieval*. 423–432.

Menahem Friedman and Abraham Kandel. 1999. *Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approaches*. Series in Machine Perception and Artificial Intelligence, Vol. 32. World Scientific.

Jing Gao, Jiawei Han, Jialu Liu, and Chi Wang. 2013a. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 International Conference on Data Mining*. 252–260.

Yue Gao, Meng Wang, Zheng-Jun Zha, Jialie Shen, Xuelong Li, and Xindong Wu. 2013b. Visual-textual joint relevance learning for tag-based social image search. *IEEE Transactions on Image Processing* 22, 1, 363–376.

Amol Ghoting, Prabhanjan Kambadur, Edwin P. D. Pednault, and Ramakrishnan Kannan. 2011. NIMBLE: A toolkit for the implementation of parallel data mining and machine learning algorithms on MapReduce. In *Proceedings of the 2011 International Conference on Knowledge Discovery and Data Mining*. 334–342.

Caroline A. Halcrow, Leslie Carr, and Susan Halford. 2016. Using the SPENCE model of online/offline community to analyse sociality of social machines. In *Proceedings of the 2016 International World Wide Web Conferences*. 769–774.

Xiangnan He, Min-Yen Kan, Peichu Xie, and Xiao Chen. 2014. Comment-based multi-view clustering of Web 2.0 items. In *Proceedings of the 2014 International World Wide Web Conferences*. 771–782.

Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. 2016. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 2016 International Conference on Research and Development in Information Retrieval*. 549–558.

Paul James, Yaso Nadarajah, Karen Haive, and Victoria Stead. 2012. *Sustainable Communities, Sustainable Development: Other Paths for Papua New Guinea*. University of Hawaii Press, Honolulu, HI.

Meng Jiang, Peng Cui, Fei Wang, Wenwu Zhu, and Shiqiang Yang. 2014. Scalable recommendation with social contextual information. *IEEE Transactions on Knowledge and Data Engineering* 26, 11, 2789–2802.

Abhishek Kumar and Hal Daumé III. 2011. A co-training approach for multi-view spectral clustering. In *Proceedings of the 2011 International Conference on Machine Learning*. 393–400.

Abhishek Kumar, Piyush Rai, and Hal Daumé III. 2011. Co-regularized multi-view spectral clustering. In *Proceedings of the 2011 Annual Conference on Neural Information Processing Systems*. 1413–1421.

Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang-Chien Lee. 2011. CLR: A collaborative location recommendation framework based on co-clustering. In *Proceedings of the 2011 International Conference on Research and Development in Information Retrieval*. 305–314.

Xutao Li, Gao Cong, Xiaoli Li, Tuan-Anh Nguyen Pham, and Shonali Krishnaswamy. 2015. Rank-GeoFM: A ranking based geographical factorization method for point of interest recommendation. In *Proceedings of the 2015 International Conference on Research and Development in Information Retrieval*. 433–442.

Xutao Li, Michael K. Ng, and Yunming Ye. 2012. HAR: Hub, authority and relevance scores in multi-relational data for query search. In *Proceedings of the 2012 SIAM International Conference on Data Mining*. 141–152.

Xutao Li, Michael K. Ng, and Yunming Ye. 2014. Multicomm: Finding community structure in multi-dimensional networks. *IEEE Transactions on Knowledge and Data Engineering* 26, 4, 929–941.

Hongfu Liu, Tongliang Liu, Junjie Wu, Dacheng Tao, and Yun Fu. 2015. Spectral ensemble clustering. In *Proceedings of the 2015 ACM Conference on Knowledge Discovery and Data Mining*. 715–724.

Xingjie Liu, Qi He, Yuanyuan Tian, Wang-Chien Lee, John McPherson, and Jiawei Han. 2012. Event-based social networks: Linking the online and offline social worlds. In *Proceedings of the 2012 ACM Conference on Knowledge Discovery and Data Mining*. 1032–1040.

Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, and Junjie Wu. 2010. Understanding of internal clustering validation measures. In *Proceedings of the 2010 International Conference on Data Mining*. 911–916.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. 2001. On spectral clustering: Analysis and an algorithm. In *Proceedings of the 2001 Annual Conference on Neural Information Processing Systems*. 849–856.

Michaek Kwok-Po Ng, Xutao Li, and Yunming Ye. 2011. MultiRank: Co-ranking for objects and relations in multi-relational data. In *Proceedings of the 2011 ACM Conference on Knowledge Discovery and Data Mining*. 1217–1225.

Nam Nguyen and Rich Caruana. 2007. Consensus clusterings. In *Proceedings of the 2007 International Conference on Data Mining*. 607–612.

Jingchao Ni, Hanghang Tong, Wei Fan, and Xiang Zhang. 2015. Flexible and robust multi-network clustering. In *Proceedings of the 2015 ACM Conference on Knowledge Discovery and Data Mining*. 835–844.

Feiping Nie, Zinan Zeng, Ivor W. Tsang, Dong Xu, and Changshui Zhang. 2011b. Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering. *IEEE Transactions on Neural Networks* 22, 11, 1796–1808.

Liqiang Nie, Xuemeng Song, and Tat-Seng Chua. 2016. *Learning from Multiple Social Networks*. Morgan & Claypool.

Liqiang Nie, Meng Wang, Zheng-Jun Zha, Guangda Li, and Tat-Seng Chua. 2011a. Multimedia answering: Enriching text QA with media information. In *Proceedings of the 2011 International Conference on Research and Development in Information Retrieval*. 695–704.

Liqiang Nie, Yi-Liang Zhao, Mohammad Akbari, Jialie Shen, and Tat-Seng Chua. 2015. Bridging the vocabulary gap between health seekers and healthcare knowledge. *IEEE Transactions on Knowledge and Data Engineering* 27, 2, 396–409.

Tuan-Anh Nguyen Pham, Xutao Li, Gao Cong, and Zhenjie Zhang. 2015. A general graph-based model for recommendation in event-based social networks. In *Proceedings of the IEEE International Conference on Data Engineering*. 567–578.

Tuan-Anh Nguyen Pham, Xutao Li, Gao Cong, and Zhenjie Zhang. 2016. A general recommendation model for heterogeneous networks. *IEEE Transactions on Knowledge and Data Engineering* 28, 12, 3140–3153.

Guo-Jun Qi, Charu C. Aggarwal, and Thomas S. Huang. 2012. Community detection with edge content in social media networks. In *Proceedings of the 2012 IEEE International Conference on Data Engineering*. 534–545.

Inbal Ronen, Ido Guy, Elad Kravi, and Maya Barnea. 2014. Recommending social media content to community owners. In *Proceedings of the 2014 International Conference on Research and Development in Information Retrieval*. 243–252.

Peter J. Rousseeuw. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, 1, 53–65.

Xuemeng Song, Zhaoyan Ming, Liqiang Nie, Yi-Liang Zhao, and Tat-Seng Chua. 2016. Volunteerism tendency prediction via harvesting multiple social networks. *ACM Transactions on Information Systems* 34, 2, 10.

Xuemeng Song, Liqiang Nie, Luming Zhang, Mohammad Akbari, and Tat-Seng Chua. 2015a. Multiple social network learning and its application in volunteerism tendency prediction. In *Proceedings of the 2015 International Conference on Research and Development in Information Retrieval*. 213–222.

Xuemeng Song, Liqiang Nie, Luming Zhang, Maofu Liu, and Tat-Seng Chua. 2015b. Interest inference via structure-constrained multi-source multi-task learning. In *Proceedings of the 2015 International Joint Conference on Artificial Intelligence*. 2371–2377.

Lei Tang and Huan Liu. 2009. Uncovering cross-dimension group structures in multi-dimensional networks. In *Proceedings of the 2009 SDM Workshop on Analysis of Dynamic Networks*. 568–575.

Lei Tang, Huan Liu, and Jianping Zhang. 2012. Identifying evolving groups in dynamic multimode networks. *IEEE Transactions on Knowledge and Data Engineering* 24, 1, 72–85.

Chi Wang, Rajat Raina, David Fong, Ding Zhou, Jiawei Han, and Greg J. Badros. 2011. Learning relevance from heterogeneous social network and its application in online targeting. In *Proceedings of the 2011 International Conference on Research and Development in Information Retrieval*. 655–664.

Xiang Wang and Ian Davidson. 2010. Flexible constrained spectral clustering. In *Proceedings of the 2010 ACM Conference on Knowledge Discovery and Data Mining*. 563–572.

Fabian L. Wauthier, Nebojsa Jojic, and Michael I. Jordan. 2012. Active spectral clustering via iterative uncertainty reduction. In *Proceedings of the 2012 ACM Conference on Knowledge Discovery and Data Mining*. 1339–1347.

Zaiwen Wen and Wotao Yin. 2013. A feasible method for optimization with orthogonality constraints. *Mathematical Programming* 142, 1–2, 397–434.

Yi Yang, Dong Xu, Feiping Nie, Shuicheng Yan, and Yueting Zhuang. 2010. Image clustering using local discriminant models and global integration. *IEEE Transactions on Image Processing* 19, 10, 2761–2773.

Yuhao Yang, Chao Lan, Xiaoli Li, Bo Luo, and Jun Huan. 2014. Automatic social circle detection using multi-view clustering. In *Proceedings of the 2014 International Conference on Information and Knowledge Management*. 1019–1028.

Stella X. Yu and Jianbo Shi. 2003. Multiclass spectral clustering. In *Proceedings of the 2003 International Conference on Computer Vision*. 313–319.

Zhiping Zeng, Jianyong Wang, Lizhu Zhou, and George Karypis. 2006. Coherent closed quasi-clique discovery from large dense graph databases. In *Proceedings of the 2006 ACM Conference on Knowledge Discovery and Data Mining*. 797–802.

Deming Zhai, Hong Chang, Shiguang Shan, Xilin Chen, and Wen Gao. 2012. Multiview metric learning with global consistency and local smoothness. *ACM Transactions on Intelligent Systems and Technology* 3, 3, 53.

Chao Zhang, Guangyu Zhou, Quan Yuan, Honglei Zhuang, Yu Zheng, Lance M. Kaplan, Shaowen Wang, and Jiawei Han. 2016. GeoBurst: Real-time local event detection in geo-tagged tweet streams. In *Proceedings of the 2016 International Conference on Research and Development in Information Retrieval*. 513–522.

Hanwang Zhang, Fumin Shen, Wei Liu, Xiangnan He, Huanbo Luan, and Tat-Seng Chua. 2016b. Discrete collaborative filtering. In *Proceedings of the 2016 International Conference on Research and Development in Information Retrieval*. 325–334.

Hanwang Zhang, Zheng-Jun Zha, Yang Yang, Shuicheng Yan, Yue Gao, and Tat-Seng Chua. 2013. Attribute-augmented semantic hierarchy: Towards bridging semantic gap and intention gap in image retrieval. In *Proceedings of the 2013 ACM Conference on Multimedia*. 33–42.

Jianglong Zhang, Liqiang Nie, Xiang Wang, Xiangnan He, Xianglin Huang, and Tat-Seng Chua. 2016a. Shorter-is-better: Venue category estimation from micro-video. In *Proceedings of the 2016 ACM Conference on Multimedia*. 1415–1424.

Yutao Zhang, Jie Tang, Zhilin Yang, Jian Pei, and Philip S. Yu. 2015. COSNET: Connecting heterogeneous social networks with local and global consistency. In *Proceedings of the 2015 ACM Conference on Knowledge Discovery and Data Mining*. 1485–1494.

Dengyong Zhou and Christopher J. C. Burges. 2007. Spectral clustering and transductive learning with multiple views. In *Proceedings of the 2007 International Conference on Machine Learning*. 1159–1166.

Yang Zhou and Ling Liu. 2013. Social influence based clustering of heterogeneous information networks. In *Proceedings of the 2013 ACM Conference on Knowledge Discovery and Data Mining*. 338–346.

Received July 2016; revised November 2016; accepted January 2017