Uncovering Groups via Heterogeneous Interaction Analysis

Lei Tang
Computer Science & Engineering
Arizona State University
Tempe, Arizona 85287
Email: L.Tang@asu.edu

Xufei Wang
Computer Science & Engineering
Arizona State University
Tempe, Arizona 85287
Email: Xufei.Wang@asu.edu

Huan Liu
Computer Science & Engineering
Arizona State University
Tempe, Arizona 85287
Email: Huan.Liu@asu.edu

***Abstract*—With the pervasive availability of Web 2.0 and social networking sites, people can interact with each other easily through various social media. For instance, popular sites like Del.icio.us, Flickr, and YouTube allow users to comment shared content (bookmark, photos, videos), and users can tag their own favorite content. Users can also connect to each other, and subscribe to or become a fan or a follower of others. These diverse individual activities result in a *multi-dimensional network* among actors, forming cross-dimension group structures with group members sharing certain similarities. It is challenging to effectively integrate the network information of multiple dimensions in order to discover cross-dimension group structures. In this work, we propose a two-phase strategy to identify the hidden structures shared across dimensions in multi-dimensional networks. We extract structural features from each dimension of the network via modularity analysis, and then integrate them all to find out a robust community structure among actors. Experiments on synthetic and real-world data validate the superiority of our strategy, enabling the analysis of collective behavior underneath diverse individual activities in a large scale.**

***Keywords*-Heterogeneous Interaction, Multi-Dimensional Networks, Community Detection, Heterogeneous Network, Cross-Dimension Network Validation**

I. Introduction

The recent boom of online social networking sites (e.g., Del.icio.us, Flickr, YouTube, Facebook, MySpace, Twitter, etc.) facilitate human beings to interact with each other more conveniently than ever. It enables traditional social network analysis from hundreds of subjects to hundreds of thousands, and even more. With the readily availability of large-scale interaction networks in social media, it is gaining increasing attentions from a variety of disciplines to study the modeling and prediction of human collective behavior, including sociology, anthropology, economics, epidemics, business marketing, and behavioral science.

One fundamental task in social network analysis to understand human collective behavior is to identify actors' social positions or cohesive subgroups (a.k.a. communities) whose group members interact with each other more frequently than those outside the group [1]. This group information can be utilized for post-analysis or other related tasks such as visualization [2], group evolution [3], [4] or detecting stable clusters across temporal changes [5], group formation [6], group profiling [7], viral marketing [8], and relational learning, behavior modeling and prediction [9], [10].

A plethora of approaches have been proposed to address community discovery in social networks. However, most existing works focus on only one dimension of interaction among people (i.e., social networks of single-type interaction). In reality, people interact with each other in assorted forms of activities leading to multiple networks among the same set of actors, or a *multi-dimensional network*[1] with each dimension representing one type of interaction. For instance, at popular photo and video sharing sites (Flickr and YouTube), a user can connect to his friends through email invitation or the provided "add as contacts" function; users can also tag/comment on the social contents like photos and videos; a user at YouTube can upload a video to respond to a video post by another user; and a user can also become a fan of another user by subscription to the user's contributions of social contents. An interaction network can be constructed based on each form of activity, representing one facet of human interactions.

Indeed, any directed network can also be considered as a 2-dimensional network. Take email communication as an example. People can act in two different roles: senders and receivers. These two roles are not interchangeable. Spammers in the network send an overwhelming number of emails to normal users but seldom receive responses from them. The sender and receiver roles essentially represent two different interaction patterns, and can be represented using interaction networks derived from the directed communication network. The follower network in the popular micro-blogging site Twitter also shares a similar spirit.

When a multi-dimensional network with heterogeneous interactions is available, it might be insufficient to extract actors' group membership accurately if only one type of interaction is utilized. In social media, certain type of interaction can be incomplete due to users' privacy concern.

[1] Some practitioners also use *multi-relational network*. In social science, multi-relational network tends to refer to the case that multiple different relations exist between two actors. While in computer science domain, multi-relational network tends to refer a network with heterogeneous entities interacting with each other, which actually corresponds to a multi-mode network [4]. Here, we use multi-dimensional network to emphasize that actors are involved in distinct interactions.

The interaction can also be noisy since it is relatively much easier to get connected to another user in social media than in the physical world. No wonder some users have thousands of online friends whereas this is hardly true in reality. For instance, one user in Flickr connects to more than 19,000 friends[2]. For this kind of users, it is really fuzzy to mine the real community he's involved in given the friend network alone. On the other hand, a substantial number of users in the network might have only one or two contacts. With this noisy and highly imbalanced interactions, relying on one type of interaction alone might miss the true user community subscription. Instead, integrating assorted forms of interaction information can compensate the incomplete information at each dimension as well as reducing the noise and obtaining a more reliable community structure.

Intuitively, with a multi-dimensional network, one can use richer information to infer more accurate latent community structures among actors. However, idiosyncratic personalities lead to varied local correlations between dimensions. Some people interact with other members within the same group in one form of activity consistently, but may be inactive in another. It thus becomes a challenge to identify groups in multi-dimensional networks as we have to fuse the information from all the dimensions for joint analysis.

In this work, we first review modularity maximization [11], [12], a recently developed measure to quantify community partitions in social networks. We discuss its application in one-dimensional networks, and then introduce simple strategies to extend modularity maximization from one-dimensional (1-D) networks to multi-dimensional (M-D) networks. Since the straightforward extensions are potentially sensitive to noise, we propose a two-phase strategy to handle community detection in multi-dimensional networks. We first extract potential structural features from each dimension via modularity analysis. In the second phase, we concatenate these features to find groups. Typically, a real-world network does not have full information for the ground truth about the group membership. So a novel cross-dimension validation procedure is proposed to compare the clustering results obtained from different approaches. Our experiments on both synthetic and real-world network data validate the superiority of our proposed approach. Moreover, our approach can be easily paralleled and thus applicable for large-scale networks.

II. Modularity for 1-D Networks

In this section, we briefly review the concept of modularity in the context of 1-D networks. In large-scale social networks, three patterns are frequently observed [13]: 1) small-world phenomenon, i.e., the distance among any pair of nodes in a network is small; 2) scale-free property, or alternatively, the node degree in a network follows a power law distribution; and 3) strong community structure. Modularity [11] is proposed specifically to measure the strength of a community partition for real-world networks by taking into account the degree distribution of nodes. It is shown to be effective in various kinds of complex networks [12], [14]. For later presentation convenience, a succinct description of modularity is included below.

Consider dividing the interaction matrix A of n vertices and m edges into k non-overlapping communities. Let s_i denote the community membership of vertex v_i, d_i represents the degree of vertex i. Modularity is like a statistical test that the null model is a uniform random graph model, in which one actor connects to others with uniform probability. For two nodes with degree d_i and d_j respectively, the expected number of edges between the two in a uniform random graph model is $d_i d_j / 2m$. Modularity measures how far the within-group interaction deviates from a uniform random graph with the same degree distribution. It is defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{d_i d_j}{2m} \right] \delta(s_i, s_j) \tag{1}$$

where $\delta(s_i, s_j) = 1$ if $s_i = s_j$, and 0 otherwise. A larger modularity indicates more frequent within-group interaction. Note that Q could be negative if the vertices are split into bad clusters. $Q > 0$ indicates the clustering captures certain degree of community structure. In general, one aims to find a community structure such that Q is maximized.

The modularity in Eq. (1) can also be represented in a matrix form. Let $\mathbf{d} \in Z_+^n$ denotes the degree of each node, where Z_+ denotes the set of non-negative integers, $S \in \{0, 1\}^{n \times k}$ be a community indicator matrix defined as follows:

$$S_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ belongs to community } j \\ 0 & \text{otherwise} \end{cases}$$

and modularity matrix defined as

$$B = A - \frac{\mathbf{d}\mathbf{d}^T}{2m} \tag{2}$$

The modularity can be reformulated as

$$Q = \frac{1}{2m} Tr(S^T B S) \tag{3}$$

The discreteness of S poses the modularity maximization problem as NP-hard [15], but with a spectral relaxation to allow S to be continuous, the optimal S can be computed as the top-k eigenvectors of the modularity matrix B [12].

Contrast to the sparse interaction matrix A, the modularity matrix B is dense and cannot be computed out and held in memory if n is large (which is typically true for real-world social networks). To calculate the top eigenvectors, power method or Lanczos method can be applied as they rely only on the basic matrix-vector multiplication without holding the full matrix. Since B is the difference of a sparse matrix (A)

[2] http://www.flickr.com/people/22711787@N00

and a rank-one matrix $(\mathbf{d}\mathbf{d}^T/2m)$, the multiplication of matrix B and a vector $\mathbf{x}$ can be calculated as:

$$B\mathbf{x} = A\mathbf{x} - \frac{\mathbf{d}^T\mathbf{x}}{2m}\mathbf{d}$$

The same trick can be applied to any structured matrix similar to B (a sparse matrix plus low-rank update). This strategy is employed later in our baseline approaches as well.

The degree of freedom of k clusters is $k-1$, so we can compute the top $k-1$ eigenvectors to form a low-dimensional embedding of the interaction network into a Euclidean space. Then a post-processing optimization step like k-means can be applied to find out a discrete community assignment [14].

Occasionally, interactions are weighted rather than boolean. It is trivial to extend modularity to handle weighted networks. Instead of counting the number of edges, we can set the degree d_i of one node v_i and the total number of degrees $2m$ in Eq. (3) as follows:

$$d_i = \sum_{j=1}^{n} A_{ij}, \quad 2m = \sum_{i=1}^{n} d_i$$

III. Modularity for M-D Networks

In the previous section, we have reviewed the scheme of modularity maximization to identify communities in 1-D networks. Here, we extend the modularity analysis to multi-dimensional networks. A d-dimensional network is represented as

$$\mathcal{A} = \{A_1, A_2, \cdots, A_d\}$$

A^i represents the interaction among actors in the i-th dimension satisfying

$$A_i \in \mathcal{R}_+^{n\times n}, \quad A_i = (A_i)^T, \quad i = 1, 2, \cdots, d$$

where n is the total number of actors involved in the network. Here, we concentrate on symmetric networks[3].

In a multi-dimensional network, the interactions of actors are represented in various forms. In certain scenarios, a latent community structure exists among these actors, which explains these interactions. The goal of this work is to *infer the shared latent community structure among the actors given a multi-dimensional network.* In particular, we attempt to find out a community assignment such that Q_i is maximized for $i = 1, \cdots, d$. Different extensions following this general criterion are derived as presented below.

A. Average Modularity Maximization (AMM)

A simple strategy to handle a multi-dimensional network is to treat it as single-dimensional. One straightforward approach is to calculate the average interaction network among social actors:

$$\bar{A} = \frac{1}{d}\sum_{i=1}^{d} A_i \tag{4}$$

Correspondingly,

$$\bar{m} = \frac{1}{d}\sum_{i=1}^{d} m_i, \quad \bar{\mathbf{d}} = \frac{1}{d}\sum_{i=1}^{d} \mathbf{d}_i \tag{5}$$

With $\bar{A}$, this boils down to classical community detection in a single-dimensional network. That is, we can maximize the modularity as below:

$$\max Q = \frac{1}{2\bar{m}} Tr\left(S^T \left[\bar{A} - \frac{\bar{\mathbf{d}}\bar{\mathbf{d}}^T}{2\bar{m}}\right] S\right) \tag{6}$$

In reality, social actors often participate in different dimensions of network with varied intensity. Even within the same group, the interaction can be very sparse in one dimension but relatively more observable in another dimension. So if there is one dimension with intensive interactions, simply averaging all the dimensions would overwhelm the structural information in other dimensions.

B. Total Modularity Maximization (TMM)

Another variant is to maximize the total modularity among all the dimensions. That is,

$$\begin{aligned}\max \bar{Q} &= \frac{1}{d}\sum_{i=1}^{d} Q_i = \frac{1}{d}\sum_{i=1}^{d} Tr\left(S^T \frac{B_i}{2m_i} S\right)\\ &= Tr\left(S^T\left[\frac{1}{d}\sum_{i=1}^{d}\left\{\frac{A_i}{2m_i} - \frac{\mathbf{d}\mathbf{d}^T}{(2m_i)^2}\right\}\right] S\right)\end{aligned}$$

where Q_i is the modularity in the i-th dimension. Akin to the modularity matrix in 1-D networks, the central matrix in the above equation is dense, thus cannot be computed out directly. To compute the top eigenvectors, the involved matrix-vector multiplication is calculated as:

$$\frac{1}{d}\left[\sum_{i=1}^{d} \frac{A_i}{2m_i}\mathbf{x} - \sum_{i=1}^{d} \frac{\mathbf{d}_i^T\mathbf{x}}{(2m_i)^2}\mathbf{d}_i\right]$$

This strategy considers the degree distribution in each dimension separately whereas the previous average modularity maximization does not distinguish degree distributions in any dimension. It is not clear whether the modularity is directly comparable between dimensions and this total modularity maximization can work in practice.

IV. Principal Modularity Maximization (PMM)

In the previous section, we have described two baseline strategies (denoted as AMM and TMM, respectively) to integrate multiple dimensions of a network. Now we shall show principal modularity maximization (PMM) which circumvents the comparability problem of different dimensions.

[3] Directed network can be converted into undirected networks through certain schemes as shown in later parts.

As shown in Figure 1, it consists of two steps: structural feature extraction and cross-dimension integration.

A. Structural Feature Extraction

Structural features are the network-extracted dimensions that are indicative of community structure. Recall that in Section II, to maximize the modularity, we compute a low-dimensional embedding using the top eigenvectors of the modularity matrix. In other words, those selected eigenvectors represent possible community partitions. Thus, the eigenvectors can be treated as the important structural features extracted from the network.

One concern with previous AMM and TMM is that they are not robust to noisy dimensions of a network. This motivates us to consider denoising each dimension of the network first. Since those eigenvectors with negative or small eigenvalues contribute marginally to the modularity and are very likely to be noise, they should be abandoned. For a multi-dimensional network, we can extract social features from each dimension of the network. Only those eigenvectors with a positive eigenvalue should be kept. We can also retain just some top-ranking community indicators to reduce noise.

B. Cross-Dimension Integration

Assuming a latent community structure is shared across dimensions in a multi-dimensional network, it is expected that the extracted structural features should be similar. However, the features based on modularity maximization are not unique. Dissimilar structural features do not suggest that the corresponding community structures are drastically different. Let S be the top-ℓ eigenvectors that maximize Q, and V an orthonormal matrix such that

$$V \in R^{\ell \times \ell}, \; VV^T = I_\ell, \; V^T V = I_\ell$$

It can be verified that SV also maximize Q:

$$\begin{aligned} & \frac{1}{2m} tr((SV)^T B(SV)) \\ = & \frac{1}{2m} tr(S^T BSVV^T) \\ = & \frac{1}{2m} tr(S^T BS) = Q_{\max} \end{aligned}$$

Essentially, SV and S are equivalent under an orthogonal transformation. In the simplest case, $S' = -S$ is also a valid solution. Instead, we expect the structural features of different dimensions to be highly correlated after transformation. To capture the correlations between multiple sets of variables, (generalized) canonical correlation analysis (CCA) [16], [17], is the standard statistical technique. CCA attempts to find a transformation for each set of variables such that the pairwise correlations are maximized. Here we briefly illustrate one scheme of generalized CCA which turns out to equal to principal component analysis (PCA) in our specific case.

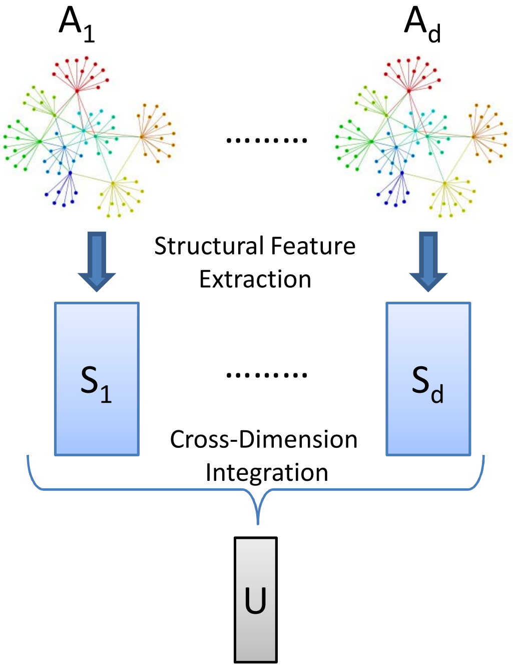


Figure 1. Overview of Principal Modularity Maximization

Let $S_i \in R^{n \times \ell_i}$ denote the structural features extracted from the i-th dimension of the network, and $w_i \in R^{\ell_i}$ be the linear transformation applied to structural features of dimension i. The correlation between two dimensions after transformation is

$$R(i,j) = (S_i w_i)^T (S_j w_j) = w_i^T (S_i^T S_j) w_j = w_i^T C_{ij} w_j$$

with $C_{ij} = S_i^T S_j$ representing the covariance between the structural features of the i-th and the j-th dimensions. Generalized CCA attempts to maximize the summation of pairwise correlations as in the following form:

$$\max \quad \sum_{i=1}^{d} \sum_{j=1}^{d} w_i^T C_{ij} w_j \tag{7}$$

$$s.t. \quad \sum_{i=1}^{d} w_i^T C_{ii} w_i = 1 \tag{8}$$

Using standard Lagrange multiplier and setting the derivatives respect to w_i to zero, we obtain equation below:

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1d} \\ C_{21} & C_{22} & \cdots & C_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ C_{d1} & C_{d2} & \cdots & C_{dd} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} \tag{9}$$

$$= \lambda \begin{bmatrix} C_{11} & 0 & \cdots & 0 \\ 0 & C_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_{dd} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} \tag{10}$$

```
Algorithm: Principal Modularity Maximization
    Input: $Net = \{A_1, A_2, \cdots, A_d\}$,
           number of communities $k$,
           number of structural features to extract $\ell$;
    Output: community assignment $idx$.
  1. Compute top $\ell$ eigenvectors of the modularity matrix
     as in Eq (2) for each $A_i$ via Lanczos method;
  2. Select the vectors with positive eigenvalues as $S_i$;
  3. Compute slim SVD of $X = [S_1, S_2, \cdots S_d] = UDV^T$;
  4. Obtain lower-dimensional embedding $\widetilde{U} = U(:, k-1)$;
  5. Normalize the rows of $\widetilde{U}$ to unit length;
  6. Calculate the cluster $idx$ with k-means on $\widetilde{U}$.
```

Figure 2. PMM for Multi-Dimensional Networks

Recall that our structural features extracted from each dimension is essentially the top eigenvectors of the modularity matrix satisfying $S_i^T S_i = I$. Thus, matrix $diag(C_{11}, C_{22}, \cdots, C_{dd})$ in Eq. (10) becomes an identity matrix. Hence $\mathbf{w} = [w_1, w_2, \cdots, w_d]^T$ corresponds the top eigenvectors of the full covariance matrix in Eq. (9), which is equivalent to PCA applied to data of the following form:

$$X = [S_1, S_2, \cdots, S_d] \tag{11}$$

To compute the $(k-1)$-dimension embedding, we just need to project the above data onto the top $(k-1)$ principal vectors. Suppose $X = UDV^T$ is the SVD of X, it follows that the top $(k-1)$ vectors of U are the lower-dimensional embedding.

The detailed algorithm is summarized in Figure 2. In summary, we first extract structural features from each dimension of the network via modularity maximization; then PCA is applied on the concatenated data as in Eq. (11) to select the top eigenvectors. Thus, we name our approach as *Principal Modularity Maximization* (PMM). After projecting the data onto the principal vectors, we obtain a lower-dimensional embedding which captures the principal pattern across all the dimensions of the network. Then we can perform k-means on this embedding to find out the discrete community assignment.

V. Experiments on Synthetic Data

In this section, we evaluate and compare different strategies applied to multi-dimensional networks. Typically, a real-world network does not provide the ground truth information of community membership, so we first resort to synthetic data to conduct some controlled experiments. The synthetic data has 3 clusters, with each having 50, 100, 200 members respectively. There are 4 dimensions of interactions among these 350 social actors. For each dimension, group members connect with each other following a random generated within-group interaction probability. The interaction probability differs with respect to groups at distinct dimensions. After that, we add some noise to the network by randomly connecting any two actors with low probability. Normalized mutual information (NMI) [18] is adopted to measure the clustering performance in the controlled experiments. NMI is a measure between 0 and 1. NMI=1 when two clusters are exactly the same.

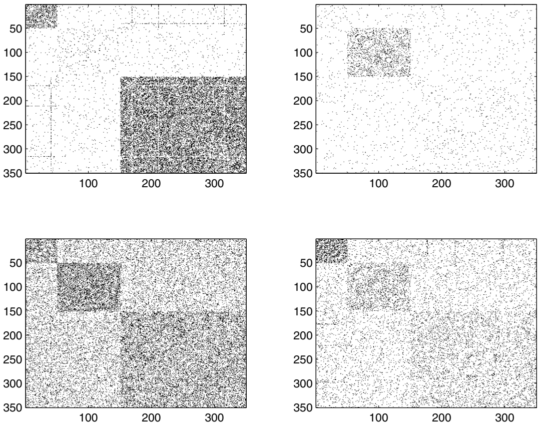
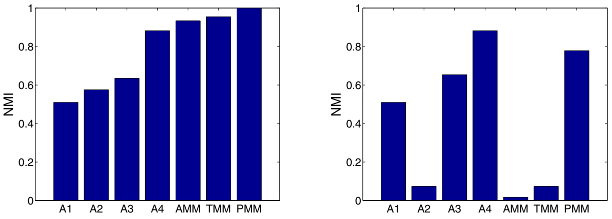


Figure 3. Example of Synthetic 4-Dimensional Network



(a) Comparable Dimensions

(b) Incomparable Dimensions

Figure 4. Performance of Various Strategies

Figure 3 shows one example of the generated multi-dimensional network. Clearly, different dimensions demonstrate different interaction patterns. Figure 4a shows the clustering performance in terms of NMI. A_1, A_2, A_3 and A_4 denote the performance based on a single dimension while the remaining 3 bars show that of AMM, TMM and PMM, respectively. Clearly, the 3 methods which consider all dimensions in the network outperform those on single-dimensional networks. This could be easily explained by the patterns represented in Figure 3. The first dimension of the network actually only shows two groups, and the second dimension involves only one group with the other two hidden behind the noise. Thus, using a single view is very unlikely to recover the correct latent community structure. This is indicated by the low NMI of the first two dimensions. Utilizing all the dimensions helps reduce the noise and uncover the shared community structure.

Comparing the three community detection strategies to handle multi-dimensional networks, our proposed principal modularity maximization outperforms the other two. To show the possible drawback of TMM and AMM, we insert some strong noise (with interaction weights ranging from

Table I
AVERAGE PERFORMANCE OVER 100 RUNS

	Strategy	Performance
Single-Dimensional	A_1	0.7237 ± 0.1924
	A_2	0.6798 ± 0.1888
	A_3	0.6672 ± 0.1848
	A_4	0.6906 ± 0.1976
Multi-Dimensional	AMM	0.7946 ± 0.1623
	TMM	0.9157 ± 0.1137
	PMM	$\mathbf{0.9351 \pm 0.1059}$

0 to 20) to the interaction matrix of the second dimension. Note that after this change, the performance of using the second dimension alone is decreasing from 0.5 to 0.1. That is, this dimension actually does not help identify the latent structure. With such a dominant dimension, both AMM and TMM fail. On the contrary, our proposed PMM still achieves reasonable good performance. This implies that PMM is more robust to noisy dimensions in multi-dimensional networks.

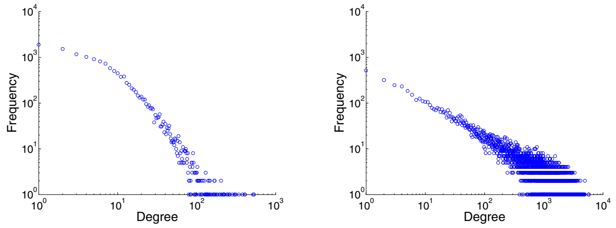
Figure 4 just shows one example. We regenerate 100 different synthetic data sets and report the average performance of each method plus its standard deviation in Table I. Clearly, multi-dimensional outperforms single-dimensional community detection method with lower variance. Due to the randomness of each run, it is not surprising that single-dimensional method shows larger variance. Among the three multi-dimensional modularity maximization strategies, PMM, with lowest variance, outperforms the other two and is more stable.

VI. EXPERIMENTS ON SOCIAL MEDIA DATA

In the previous section, we compare different strategies on synthetic data with clear ground truth information. Here, we examine our approach on real-world social media. A big challenge for evaluation is that the community membership information is often unknown in reality. To manually verify and label the community membership for each user is acceptable for a small network but hardly can it scale to large online social networks. To address this issue, we first describe a cross-dimension network validation procedure following the idea of cross validation as in conventional data mining. After that, the detail of the data and experiment results are presented.

A. Cross-Dimension Network Validation

Since a latent community structure is shared across different dimensions, a good community structure extracted from some dimensions should match the interaction at other dimensions. Akin to cross validation, we can perform cross-dimension network validation as follows: Given a multi-dimensional network $Net = \{A_i | 1 \leq i \leq d\}$, we can use $d-1$ dimensions for training and the remaining one as test data. That is, we learn a community structure from $d-1$ dimensions of the network. Based on the obtained community structure, we measure the modularity on the test dimension. It verifies how the learned community structure from other dimensions matches with the test dimension. A larger modularity implies more accurate community structure is discovered using the training data.



(a) Contact Network
(b) Favorite Network

Figure 5. Power law distribution on Different Dimensions

Table II
THE SPARSITY OF EACH DIMENSION

Network	Dimension	Density
A_1	contact	6.74×10^{-4}
A_2	co-contact	1.71×10^{-2}
A_3	co-subscription	4.90×10^{-2}
A_4	co-subscribed	1.97×10^{-2}
A_5	favorite	3.34×10^{-2}

B. YouTube Data Collection

YouTube[4] is currently the most popular video sharing web site. It is reported to “attract 100 million video views per day”[5]. As of March 17th, 2008, there have been 78.3 million videos uploaded, with over 200, 000 videos uploaded per day[6]. This social networking site allows users to interact with each other in various forms such as contacts, subscriptions, sharing favorite videos, etc. We use YouTube Data API[7] to crawl the contacts network, subscription network as well as each user’s favorite videos. To avoid sample selection bias, we choose 100 authors of recently uploaded videos as the seed set, and expand the network via their contacts and subscriptions. We crawled a small portion of the whole network, with $30,522$ user profiles reaching in total $848,003$ contacts and $1,299,642$ favorite videos. After removing those users who decline to share their contact information, we have $15,088$ active user profiles in the network.

One issue is that the collected subscription network is directional while modularity is proposed for undirected networks. For such case, simply ignoring the direction mixes the two roles of the directional interaction (similar to the email communication example in the introduction). Instead, we decompose the asymmetric interaction A into

[4] http://www.youtube.com/
[5] http://www.usatoday.com/tech/news/2006-07-16-youtube-views_x.htm
[6] http://ksudigg.wetpaint.com/page/YouTube+Statistics?t=anon
[7] http://code.google.com/apis/youtube/overview.html

two unidirectional interactions:

$$A' = A * A^T; \tag{12}$$

$$A'' = A^T * A. \tag{13}$$

Essentially, if two social actors both subscribe to the same set of users, it is likely that they are similar and share the same community; On the other hand, if two are referred by the same set of actors, their similarity tends to be higher than that of random pairs. This is similar to the two roles of hub and authority of web pages as mentioned in [19]. It is also adopted for semi-supervised learning on directed graphs [20].

To utilize all aspects of information in our collected data, we construct a 5-dimensional network:

A_1: contact network: the contact network among those $15,088$ active users;

A_2: co-contact network: two active users are connected if they both add another user as contact; This is constructed based on all the reachable 848,003 users (excluding those active ones) in our collected data following Eq. (12).

A_3: co-subscription network: the connection between two users denotes they subscribe to the same user; constructed following Eq. (12);

A_4: co-subscribed network: two users are connected if they are both subscribed by the same user; constructed following Eq. (13);

A_5: favorite network: two users are connected if they share favorite videos.

The interactions in all dimensions are weighted[8]. Table II shows the connection density of each dimension. Contact dimension is the most sparse one, while the other dimensions, due to the construction, are denser. Figure 5 shows the degree distribution in contacts network and favorite network. Both follow a power law pattern as expected.

C. Comparative Study

AMM, TMM and PMM as well as single-dimensional modularity maximization methods are compared. We cluster the active users involved in the network into different number of communities ranging from 10 to 100. The clustering performance of single-dimensional and multi-dimensional methods when $k = 20$, 40 and 60 are presented in Tables III-V. We omit the detailed results for other cases as a similar trend is observed. In these tables, the rows represent methods and the columns denote the dimensions used as test data. The bold face denotes the optimal performance in each column. Note that in our cross-dimension network validation procedure, the test dimension is not available during training, thus the diagonal entries for single-dimensional methods are not shown.

Table III
PERFORMANCE WHEN K=20

Methods	A_1	A_2	A_3	A_4	A_5
A_1	—	.0007	.0008	.0008	.0002
A_2	.1548	—	.0133	.0361	.0076
A_3	.0712	.0275	—	.0446	.0140
A_4	.0584	.0569	.0186	—	.0108
A_5	.0314	.0135	.0095	.0180	—
AMM	.1096	.0001	.0018	.0053	.0070
TMM	.3740	.1856	.1246	.1800	.0706
PMM	**.4085**	**.2063**	**.1307**	**.1844**	**.0947**

Table IV
PERFORMANCE WHEN K=40

Methods	A_1	A_2	A_3	A_4	A_5
A_1	—	.0071	.0080	.0100	.0010
A_2	.2091	—	.0283	.0700	.0153
A_3	.1718	.0801	—	.0776	.0204
A_4	.0636	.0580	.0189	—	.0108
A_5	.0529	.0207	.0038	.0103	—
AMM	.1540	.0611	.0019	.0070	.0157
TMM	.2880	.1448	.0799	.1236	.0521
PMM	**.3514**	**.1521**	**.0808**	**.1309**	**.0574**

Table V
PERFORMANCE WHEN K=60

Methods	A_1	A_2	A_3	A_4	A_5
A_1	—	.0811	.0139	.0214	.0076
A_2	.2058	—	.0190	.0521	.0173
A_3	.1163	.0581	—	.0498	.0165
A_4	.1161	.0816	.0346	—	.0200
A_5	.0669	.0323	.0094	.0268	—
AMM	.1281	.0511	.0022	.0061	.0225
TMM	.2790	.1145	.0604	**.1208**	.0415
PMM	**.3296**	**.1329**	**.0656**	.1101	**.0417**

PMM is clearly the winner most of the time. The other dimensions except A_1 are quite noisy due to their construction process, hence yielding consistently lower modularity. A closer examination reveals that utilizing information of all the dimensions (TMM and PMM) outperforms single-dimensional clustering. AMM does not work well, because our network are weighted and simple average blurs the latent community structure information presented at each dimension. Comparing all the multi-dimensional clustering approaches, both AMM and TMM are not comparable to our proposed PMM. Typically, AMM $<$ TMM $<$ PMM. Part of the reason is that, TMM considers degree distribution in separate dimensions. Our algorithm, by removing noise in each dimension, achieves the most accurate community structure among all the methods. This is evident in the contact dimension. Figure 6 shows the performance of the multi-dimensional clustering methods with respect to number of clusters (k) on A_1. No matter how many clusters we set, PMM outperforms the other two methods with a significant margin.

One observation is that the modularity decreases for almost all the methods when the number of communities

[8] In a preliminary version [21] of this work, the interactions are represented as boolean values, which miss some information of the similarity among actors.

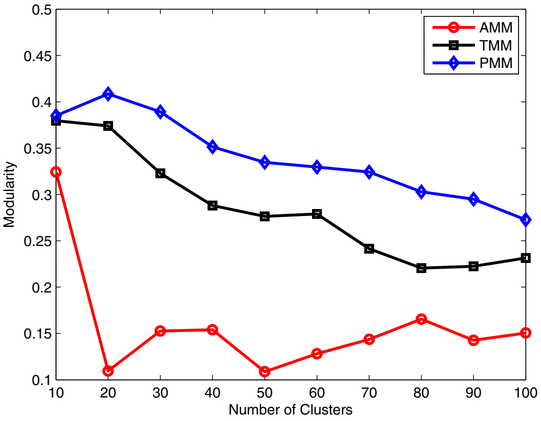


Figure 6. Performance comparison on Contact Dimension

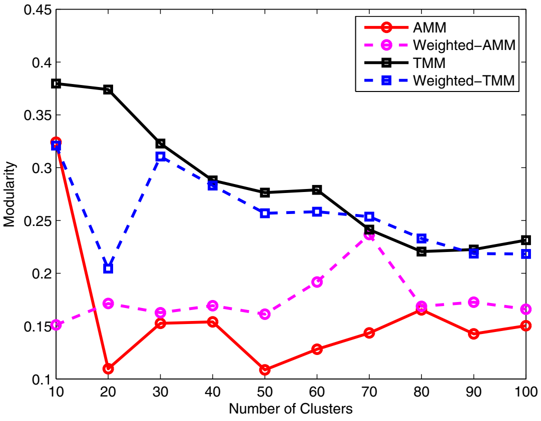


Figure 7. Performance of Weighted AMM & TMM

multiplies. This is partly due to the resolution limit of modularity [22]. It is shown that modularity measure favors network partitions with groups of modules combined into larger communities, which explains the decay of modularity with respect to increasing number of modules in the experiment. However, here we focus on the comparison of different methods and the number of communities is fixed for all methods. The resolution limit of modularity does not invalidate the conclusions about the superiority of different methods.

D. Weighted AMM & TMM

AMM and TMM both treat the interaction of each dimension equivalently. If one dimension's community structure is more prominent, it seems reasonable to trust that dimension more, which asks for a weighted summation of the interaction or modularity. Since modularity calibrates the community effect of a network, one hypothesis is that whether we can use the modularity at each dimension as a guide to do the weighted average.

Let Q_i denotes the modularity computed for each dimension. For weighted AMM, the average interaction matrix in Eq. (4) becomes

$$\bar{A} = \sum_{i=1}^{d} \frac{Q_i}{\sum_{j=1}^{d} Q_j} A_i \tag{14}$$

In a similar vein, we compute the eigenvectors of the following matrix for weighted TMM:

$$\sum_{i=1}^{d} \frac{Q_i}{\sum_{j=1}^{d} Q_j} \left\{ \frac{A_i}{2m_i} - \frac{\mathbf{d}\mathbf{d}^T}{(2m_i)^2} \right\} \tag{15}$$

Figure 7 shows the results with different weights associated with each dimension as stated in Eq. 14) and (15). The weighted extension helps for AMM but does not help for TMM. It requires more insightful understanding upon the dimensions to assign proper weights over each dimension. After all, all their performance is still not comparable to our proposed PMM, which does not require weight specification.

E. Discussions

Among all the multi-dimensional methods, AMM is most efficient as it only requires a single-dimensional clustering procedure on the average interaction matrix. Our proposed PMM needs to extract the important structural features from each network dimension, thus the total computation is more expensive. But this step can be easily paralleled with a multi-core CPU or clusters. The number of structural features ℓ extracted from each dimension is normally much smaller than the number of actors n, i.e., $\ell << n$, resulting in a narrow data set X in Eq. (11) of size $n \times \ell d$. So the subsequent SVD computation to find out the shared latent structure is acceptable. Another advantage we want to emphasize is that PMM not only finds the shared latent community structure, but also allows community identification in a specific dimension by utilizing the extracted structural features. This cannot be accomplished by AMM or TMM.

Note that the PMM framework presented in Figure 1 is easy to generalize. Other variants, such as graph Laplacian, multi-dimensional scaling can also be utilized to extract structural features. As long as the structural features are orthonormal, the cross-dimension integration remains the same: concatenate all the features and perform PCA. In addition, if any additional features are available about the nodes in the network, it is easy to combine them with structural features for joint analysis.

VII. Related Work

Some works attempt to address unsupervised learning with multiple data sources or clustering results, such as cluster ensemble [18], [23], [24] and consensus clustering [25]–[28]. Most of the algorithms aim to find a robust clustering based on multiple clustering results, which are prepared via feature or instance sampling or disparate clustering algorithms. A similar idea is applied to community detection in social networks [29]. A small portion of connections between nodes are randomly removed before each run, leading to multiple different clustering results. Those clusters occurring repeatedly are considered more stable, and are deemed to reflect the natural communities in reality. However, all the cluster ensemble methods concentrate on either attribute-based data or one-dimensional networks.

Another related field is multi-view clustering. Bickel and Scheffere [30] propose co-EM and an extension of k-means and hierarchical clustering to handle data with two conditional independent views. Sa [31] creates a bipartite based on the two views and tries to minimize the disagreement. Different spectral frameworks with multiple views are studied in [32] and [33]. The former defines a weighted mixture of random walk over each view to identify communities. The latter assumes clustering membership of each view is provided and finds an optimal community pattern via minimizing the divergence of the transformed optimal pattern and the community membership of each view. As for real-world social networks, one striking observation is that spectral clustering always finds tight and small-scale but almost trivial communities (say, the community is connecting to the remaining network via one edge) [34]. Modularity maximization, on the other hand, tends to find modules composed of small-scale communities [22]. A comparison between spectral clustering and modularity maximization within a large-scale multi-dimensional network is worthy of future work.

Some theoretical analysis of multi-view clustering via canonical correlation analysis is presented in [35]. It shows that under the assumption that the views are uncorrelated given the cluster label, a much weaker condition is required for CCA to separate clusters successfully. But the conclusion is based on two views with each being attributes. How to generalize the theoretical result to networks of multiple heterogeneous interactions requires further research.

Unsupervised multiple kernel learning [36] is relevant if we deem each dimension of the network as a similarity or kernel matrix. Multiple kernel learning aims to find a combination of kernels to optimize for classification or clustering. Unfortunately, its limited scalability hinders its application even to a medium-size network.

VIII. Conclusions

Multi-dimensional networks commonly exist in many social networking sites, reflecting diverse individual activities. In this work, we propose to detect the latent communal structure in a multi-dimensional network. We formally describe the community detection problem in multi-dimensional networks and discuss two straightforward extensions of modularity maximization from single-dimensional to multi-dimensional networks: Average modularity maximization (AMM) and total modularity maximization (TMM) .We show that both methods are not robust to handle networks with noise. Principal modularity maximization (PMM) is proposed to overcome this limitation. We extract structural features from each dimension of the network, which also effectively removes the noise in that dimension; and then apply cross-dimension analysis on the constructed data to find the lower-dimensional embedding such that the features extracted from all the dimensions are highly correlated to each other. PMM has been empirically shown to outperform single-dimensional clustering as well as AMM and TMM in both synthetic and YouTube data. Its superiority is most observable when a certain dimension of the network is rather noisy.

In current work, we attempt to extract the shared latent community structure beneath heterogeneous interactions. It would be exciting as well to study the specific group structure of each dimension. Are there different group structures in each dimension? How are they correlated? Could we utilize dimension information for more effective viral marketing? All these questions are worthy of further research.

Acknowledgments

This work is, in part, supported by AFOSR-FA95500810132 and ONR-N000140810477.

References

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

[2] H. Kang, L. Getoor, and L. Singh, "Visual analysis of dynamic group membership in temporal social networks," *SIGKDD Explorations, Special Issue on Visual Analytics*, vol. 9, no. 2, pp. 13–21, dec 2007.

[3] S. Asur, S. Parthasarathy, and D. Ucar, "An event-based framework for characterizing the evolutionary behavior of interaction graphs," in *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2007, pp. 913–921.

[4] L. Tang, H. Liu, J. Zhang, and Z. Nazeri, "Community evolution in dynamic multi-mode networks," in *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2008, pp. 677–685.

[5] N. Bansal, F. Chiang, N. Koudas, and F. W. Tompa, "Seeking stable clusters in the blogosphere," in *VLDB '07: Proceedings of the 33rd international conference on Very large data bases*, 2007, pp. 806–817.

[6] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2006, pp. 44–54.

[7] L. Tang, H. Liu, J. Zhang, N. Agarwal, and J. J. Salerno, "Topic taxonomy adaptation for group profiling," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 4, pp. 1–28, 2008.

[8] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *KDD*, 2002, pp. 61–70.

[9] L. Tang and H. Liu, "Relational learning via latent social dimensions," in *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2009, pp. 817–826.

[10] ——, "Scalable learning of collective behavior based on sparse social dimensions," in *The 18th ACM Conference on Information and Knowledge Management*, 2009.

[11] M. Newman, "Modularity and community structure in networks," *PNAS*, vol. 103, no. 23, pp. 8577–8582, 2006.

[12] ——, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 74, no. 3, 2006. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.74.036104

[13] D. Chakrabarti and C. Faloutsos, "Graph mining: Laws, generators, and algorithms," *ACM Comput. Surv.*, vol. 38, no. 1, p. 2, 2006.

[14] S. White and P. Smyth, "A spectral clustering approaches to finding communities in graphs," in *SDM*, 2005.

[15] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "Maximizing modularity is hard," *Arxiv preprint physics/0608255*, 2006.

[16] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936. [Online]. Available: http://dx.doi.org/10.2307/2333955

[17] J. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, pp. 433–451, 1971.

[18] A. Strehl and J. Ghosh, "Cluster ensembles — a knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2003.

[19] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[20] D. Zhou, J. Huang, and B. Schölkopf, "Learning from labeled and unlabeled data on a directed graph," in *ICML '05: Proceedings of the 22nd international conference on Machine learning*. New York, NY, USA: ACM, 2005, pp. 1036–1043.

[21] L. Tang and H. Liu, "Uncovering cross-dimension group structures in multi-dimensional networks," in *SDM workshop on Analysis of Dynamic Networks*, 2009.

[22] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *PNAS*, vol. 104, no. 1, pp. 36–41, January 2007. [Online]. Available: http://dx.doi.org/10.1073/pnas.0605965104

[23] A. Topchy, A. K. Jain, and W. Punch, "Combining multiple weak clusterings," in *ICDM '03: Proceedings of the Third IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2003, p. 331.

[24] X. Z. Fern and C. E. Brodley, "Solving cluster ensemble problems by bipartite graph partitioning," in *ICML '04: Proceedings of the twenty-first international conference on Machine learning*. New York, NY, USA: ACM, 2004, p. 36.

[25] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data," *Mach. Learn.*, vol. 52, no. 1-2, pp. 91–118, 2003.

[26] T. Hu and S. Y. Sung, "Consensus clustering," *Intell. Data Anal.*, vol. 9, no. 6, pp. 551–565, 2005.

[27] N. Nguyen and R. Caruana, "Consensus clusterings," in *ICDM '07: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 607–612.

[28] A. Goder and V. Filkov, "Consensus clustering algorithms: Comparison and refinement," in *SDM*, 2008.

[29] J. Hopcroft, O. Khan, B. Kulis, and B. Selman, "Natural communities in large linked networks," in *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2003, pp. 541–546.

[30] S. Bickel and T. Scheffere, "Multi-view clustering," in *ICDM '04: Proceedings of the Fourth IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 19–26.

[31] V. R. de Sa, "Spectral clustering with two views," in *Proceedings of Workshop of Learning with Multiple Views*, 2005.

[32] D. Zhou and C. J. C. Burges, "Spectral clustering and transductive learning with multiple views," in *ICML '07: Proceedings of the 24th international conference on Machine learning*. New York, NY, USA: ACM, 2007, pp. 1159–1166.

[33] B. Long, P. S. Yu, and Z. M. Zhang, "A general model for multiple view unsupervised learning," in *SDM*, 2008.

[34] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Statistical properties of community structure in large social and information networks," in *WWW '08: Proceeding of the 17th international conference on World Wide Web*. New York, NY, USA: ACM, 2008, pp. 695–704.

[35] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*. New York, NY, USA: ACM, 2009, pp. 1–8.

[36] H. Valizadegan and R. Jin, "Generalized maximum margin clustering and unsupervised kernel learning," in *NIPS*, 2007.