Exploring Groups from Heterogeneous Data via Sparse Learning

Huawen Liu[1,3], Jiuyong Li[2], Lin Liu[2], Jixue Liu[2],
Ivan Lee[2], and Jianmin Zhao[1]

[1] Zhejiang Normal University, Jinhua 321004, China
[2] University of South Australia, Adelaide, SA5095, Australia
[3] Academy of Mathematics and Systems Science, CAS, Beijing 100190, China

Abstract. Complexity networks, such as social networks, biological networks and co-citation networks, are ubiquitous in reality. Identifying groups from data is critical for network analysis, for it can offer deep insights in understanding the structural properties and functions of complex networks. Over the past decades, many endeavors from interdisciplinary fields have been attempted to identify groups from data. However, little attention has been paid on exploring groups and their relationships from different views. In this work, we address this issue by using canonical correlation analysis (CCA) to analyze groups and their interplays in the networks. To further improve the interpretability of results, we solve the optimization problem with sparse learning, and then propose a generalized framework of group discovery from heterogeneous data. This framework enables us to find groups and explicitly model their relationships from diverse views simultaneously. Extensive experimental studies conducted on both synthetic and DBLP datasets demonstrate the effectiveness of the proposed method.

Keywords: Group discovery, complexity network, canonical correlation analysis, LASSO, Sparse learning.

1 Introduction

It is well known that everything in the universe is relevant to others and nothing independently exists. According to different functions and properties, entities intertwined with each other form groups, resulting in complex networks [14]. Typical examples include functional regulation models of proteins in biology [10], trophic pathways of species in ecology [3,8], communities of people in sociology [16], interlinks of web pages in World Wide Web [5], collaboration relationships of authors in bibliography [11], and many others. Since complex networks are ubiquitous in reality, network learning now gains increasing attentions from a variety of disciplines including computer science, physics, economics, business marketing, biology, engineering, epidemiology, social and behavioral science [4].

Exploring the structures, functions, as well as the interactions of networks is very important, because it may provide us an insightful understanding how the

networks work [16,4,20]. For example, identifying collaboration relationships of scientists provides us an indication to which topics are popular and what the research trends will be potentially studied in future [11]; uncovering functional modules of proteins helps us understand which causes lead to certain diseases and how the proteins are co-regularized [10]; revealing social communities of people aids us in finding out what the interests and opinions of people are [23].

An essential yet challenging task of network learning is to discover functional units in complex networks [4]. In literature, the functional unit is also known as community, module, clique, cluster or coalition, depending on the specific contexts or applications at hand [16]. These terms cover the entities only. Here we call the functional unit as *group*, for the interplays between the entities have also been taken into account. Since group discovery is helpful to analyze the network structures and further capture knowledge about the functions and properties of the network systems, it is not surprised that it has attracted many attentions from different domains [20,17,19].

Over the last decades, group discovery has been extensively investigated and dozens of discovery methods, including graph partition models, clique based models, clustering models, modularity maximization models, and so on, have been developed [4]. However, most of them place emphases merely on identifying function units. To the best of our knowledge, the interrelations of groups, which are very common in nature, have not been fully exploited. In some cases, identifying the interrelations is more crucial to understand the structural and functional properties, for it may offer us an insightful perspective to the networks.

Another issue is that the existing methods only take the entities with the same type into consideration and ignore other information with different types. Empirical studies show that the network structures in real world are complex and often involve the entities derived from heterogeneous data sources. The information with different types may bring benefits to explore and analyze the complex networks if it had been taken into account in network learning.

In this paper, we present a generalized framework to explicitly address the problems mentioned above. It adopts canonical correlation analysis (CCA) to analyze groups and their interrelations simultaneously. It not only allows to handle the groups with different types, but also provides an effective solution to scale their interrelations in a quantitative manner. Furthermore, we turn the objective function of CCA into a LASSO penalized least square problem (i.e., ℓ_1-norm penalty) by using complex linear algebra equivalent transformations. Consequently, the proposed method can obtain an optimal sparse solution for large-scale complex networks. Specifically, the contributions of this work are twofold: 1) Our method can handle the entities from the networks with different types, and is also extensible to the multi-view or multi-slice situations after some revisions have been made. 2) Since our method adopts LASSO (ℓ_1-norm penalty) to uncover groups and their interrelations, it has sparse property and the final results can be interpreted easily.

558   H. Liu et al.

The rest of this paper is organized as follows. Section 2 briefly recalls previous related work. We presents the basic concept of CCA in Section 3. Section 4 proposes a new group discovery method by using sparse learning. The experimental results and discussions on artificial and real world networks have been provided in Section 5, followed by the conclusions in Section 6.

2 Related Work

Group discovery is a hot topic in network learning. Over the past years, a considerable number of discovery methods have been witnessed. Here only the latest methods will be discussed briefly. Interested readers can refer to good survey literature (see e.g., [4]) and references therein to get more information.

Since complex networks are often represented as graphs, graph analysis, which has solid mathematical and theoretical fundamentals, has been extensively investigated in network learning. This kind of discovery methods apply graph theory to explore groups, which tightly connected with each other by edges [4]. Typical examples of such kind include clique-based, graph partition-based, ratio cut-based, normalized cut-based and max-flow-min-cut detection approaches. Note that graph partitioning is a NP-hard problem. Moreover, it is also need the number of groups and even their sizes, which are usually unknown in advance.

Clustering is another solution for group discovery. The clustering techniques, such as hierarchical clustering, partitional clustering and spectral clustering, have been taken to identify groups from data. For instance, Chi et al. [2] disclosed communities and evaluated their evolution process by using the spectral clustering, which usually partitions nodes in a graph into clusters in terms of the eigenvectors of its matrix representation. Like the graph-based methods, the limitation of clustering is its relatively high computational cost, and the number of clusters should also be pre-specified in some situations.

Modularity is widely used as a stopping measure for clustering, resulting in the prevalence of the modularity maximization-based community discovery methods [13]. The great success of this framework relies on the modularity assumption, that is, the higher a modularity is, the better its corresponding partition is. This implies that the partition with maximum modularity on a given graph is the best group. As a typical example, Jiang and McQuay [7] exploited modularity Laplacian to discover communities by optimizing the modularity functions with additional nonnegative constraint. However, the modularity optimization is also NP-complete. Additionally, in real world the assumption of modularity maximum is not always true [4].

Statistical inference is a powerful tool to deduce properties of data in machine learning, and has also been used to model and analyze graph topological structures. The discovery methods based on block modeling, Bayesian inference, latent Dirichlet allocation and model selection belong to this kind of representative cases. For example, Yang et al. [22] estimated parameters with a Bayesian

treatment in modeling networks and then developed a dynamic stochastic block model to find communities and their evolution in dynamic situations.

More recent studies of group discovery focus on exploring the evolution or behaviors of groups from the multi-slice or multi-dimension prospective. The representative examples of such framework have been illustrated in [14,22,12,18]. Pons and Latapy [15] integrated the communities discovered by different methods in a post-processing way. Tang et al. [20] exploited four integration strategies, i.e., network interactions, utility functions, structural features and community partitions, to fuse the communities derived from multi-dimension sources. Lancichinetti et al. [9] took edge directions, edge weights, overlapping communities, hierarchies and community dynamics into account to identify the significant communities.

It is worthy to notice that most of discovery algorithms mentioned above can not handle data from heterogeneous sources and the interrelations of groups. The networks in reality, however, often encounter groups with different types. Thus identifying groups with different types and their interrelations is important, because it may bring more information and provide us a deep insight to understand the working mechanisms of the networks. Recently, author-topic model (ATM) has gained much attraction in information retrieval [1]. It mainly adopts graph-based (e.g., LDA and pLDA), semantic-based (e.g., PCA, LSI and pLSI) or their extensions with other techniques (e.g., Gibbs sampling and HMM) to reveal the groups of authors and topics [1]. However, ATM only qualitatively describes the relationships of groups. As far as we are aware, little attention has been put on measuring the interrelations of groups in a quantitative way.

3   Canonical Correlation Analysis

Canonical correlation analysis (for short, CCA) proposed by Hotelling is a well-known multivariate technique [6]. Let $X = \{x_1, ..., x_p\}$ and $Y = \{y_1, ..., y_q\}$ be two sets of variables, and both of them are centralized, i.e., $\sum_{i=1}^{p} x_i = 0$ and $\sum_{i=1}^{q} y_i = 0$. CCA aims at obtaining two weighted linear combinations ω_X and ω_Y of X and Y, respectively, such that their correlation is maximal, i.e.,

$$\rho(\omega_X, \omega_Y) = \max \frac{<\omega_X, \omega_Y>}{\|\omega_X\|\|\omega_Y\|}, \tag{1}$$

where $\omega_X = Xu$ and $\omega_Y = Yv$ are canonical variates with the weight vectors $u' = (u_1, ..., u_p)$ and $v' = (v_1, ..., v_q)$.

The intuitive meanings of CCA is that it projects two sets of variables into a lower-dimensional space in which they are maximally correlated. Since solving the maximal value of $\rho(\omega_X, \omega_Y)$ is invariant to the scaling of u and v either together or independently, Eq. 1 can be rewritten as follows:

$$\begin{aligned} \arg\max_{u,v} \; & u'X'Yv \\ & s.t. \; u'X'Xu = 1, v'Y'Yv = 1. \end{aligned} \tag{2}$$

For the optimization problem in Eq. 2, one of frequently used solutions is to formulate it as a Lagrangian optimization form. Due to the limitation of space, here

we will not provide the details of inferences step by step. Eventually, the CCA formulation seeks for solving the eigenvectors and eigenvalues of the following generalized eigenvalue problem:

$$X'Y(Y'Y)^{-1}Y'Xu = \lambda X'Xu$$
$$Y'X(X'X)^{-1}X'Yv = \lambda Y'Yv$$

(3)

where λ is the eigenvalue, and u and v are its corresponding eigenvectors with respect to X and Y, respectively. If $X'X$ is invertible (i.e., non-singular), the first formula of Eq. 3 can be further formulated as a standard eigenvalue problem of $(X'X)^{-1}X'Y(Y'Y)^{-1}Y'Xu = \lambda u$. Otherwise, other strategies, such as generalized inverses and regularization, should be considered to obtain the inverse of $X'X$ or $Y'Y$.

4   Group Relationship Discovery

4.1   Problem Statement

Assume $X = \{x_1, x_2, ..., x_p\} \in \Re^{n \times p}$ and $Y = \{y_1, y_2, ..., y_q\} \in \Re^{n \times q}$ are two sets of variables (or features) representing n instances (or samples), where $x_i \cap y_j = \emptyset$ for $i = 1, ..., p$ and $j = 1, ..., q$. They can be treated as the n instances observed from two different perspectives.

In this paper, the purpose of group discovery is twofold. The first one is, for each set of variables, e.g., X, to identify a set of groups $G_X = \{G_{Xi}|G_{Xi} \subseteq X, i = 1, ..., k\}$, such that the elements are highly relative to each other in the same group G_{Xi}, while irrelative to those in other groups G_{Xj} $(i \neq j)$. Similar operations can be performed on Y to obtain G_Y. The second task of this paper is to scale the relationships between groups derived from different types, e.g., $G_{Xi} \in G_X$ and $G_{Yj} \in G_Y$, in a quantitative manner.

Since the group interrelations are often hidden and implicitly observed through a large number of variables, it is not appropriate to evaluate correlations between pairs of variables individually, or simply calculate the accumulative total of dependencies between variables from groups with different views. A desirable solution is to take the groups as a whole, rather than their individual variables, into account in extracting the group interrelations.

4.2   Obtaining the First Group

Given two sets of variables X and Y, CCA can effectively obtain their canonical variates such that the correlations between them are maximal. However, one of the CCA problems is that the computational cost of matrix decomposition is relatively high, especially when the quantity of variables exceeds tens of thousands. The non-singular property of matrix is the second issue that should also be taken into consideration. Here we go further and efficiently solve it via ℓ_1-norm regularization.

For the Eq. 2, we have the following property:

Property 1. The optimization problem of CCA (i.e., Eq. 2) is equivalent to a distance minimization problem between two matrices, i.e.,

$$\arg\min_{u,v} f(u,v) = \|Xu - Yv\|^2$$
$$s.t.\ u'X'Xu = 1, v'Y'Yv = 1. \tag{4}$$

One may observe that this formulation is a least square if either Xu or Yv is fixed, i.e.,

$$f(u) = \|Xu - \alpha\|^2$$
$$f(v) = \|\beta - Yv\|^2 \tag{5}$$

To solve this optimization problem above, Partial Least Squares (PLS) seems to be an effective technique. The off-the-shelf method of PLS performs a regression operation on each formulation within Eq. 5 alternately, and ultimately obtains u and v. However, a possible drawback of PLS is that interpreting the derived results becomes impossible, since u and v are weighted combinations of all available variables in X and Y, respectively.

The interpretability of the obtained results is very important for the practical applications. It directly affects users understanding data, such as providing evidences for decision-makers, boosting product promotion for businessmen, uncovering pathogenies of diseases for doctors. An effective strategy is to make the results sparse via variable selection.

We resort to a least absolute shrinkage and selection operator (for short, LASSO) penalized model [21] to fulfill the purpose of variable selection. The underlying is that LASSO enables us estimating the objective function and achieving variable selection simultaneously in one stage, where variables will be selected by assigning zeros to the weights of variables with very small coefficients. Specifically, the first objective function $f(u)$ in Eq. 5 has been transformed to the following form after a ℓ_1-norm constraint has been performed on u.

$$f(u) = \|Xu - \alpha\|^2 + \lambda_u \sum |u_i| \tag{6}$$

where λ_u ($\lambda_u \geq 0$) is a tuning constant for u. Under this constraint, the weights of some variables become zero if λ_u is enough small. Specifically, each weight coefficient u_i decreases after comparing with a threshold. If u_i is lower than the threshold it will be set to zero, otherwise it will be modified or preserved. Thus the purpose of sparse solution can be achieved. Given λ_u, u_i is determined via the soft-thresholding strategy, i.e.,

$$u_i = sgn((\alpha'X)_i)(|(\alpha'X)_i| - \lambda_u)_+ \tag{7}$$

where $sgn(z)$ is the sign function of z, and $(z)_+$ is defined to z if $z > 0$ and 0 if $z \leq 0$. Similarly, $f(v)$ can also be handled by performing another ℓ_1-norm penalty regularization λ_v.

After u and v have been obtained, the new coordinate systems of X and Y are formed and represented as Xu and Yv, respectively. Since u and v are sparse, Xu and Yv have good interpretability and represent the first group derived from X and Y, respectively. More importantly, this pair of groups has maximal correlation, which can be measured in a quantitative way, i.e., $\rho = < Xu, Yv >$, after normalized.

4.3   Obtaining the Rest Groups

Assume the same instances are observed from two different views. Obtaining only one group from data is not enough, since it is unlikely to describe all relationships of variables, especially in high-dimensional settings. Therefore, additional groups should also be uncovered from data. To untie this knot, an alternative solution is to minimize the criterion of discovering the first group (i.e., Eq. 4) repeatedly, each time on the residual variables obtained by wiping off the information of the groups found previously.

To achieve the goal, a trick is available, where the original data will be updated according to the information of the obtained groups. The central idea is similar to the *whitening* step in image processing, which has often been used to lessen dependencies or correlations of images. Specifically, after the first group Xu has been disclosed, the original data X is extended as $X'_n = [X' \quad \frac{u}{\sqrt{\rho}}]$, where ρ is the canonical correlation. This extension aims to lessen the effect of Xu, so as not to frustrate the process of discovering other groups in succession. Y can also be handled in a similar way. Indeed, this extension process is lossless, because the original and extension matrices have the same canonical variates. Under the context of X_n (or Y_n), we can obtain its first canonical variate $X_n u_2$ (or $Y_n v_2$), which actually is the second group of X (or Y). Akin to u_2, other groups can be identified by updating X continuously, until the extension matrix contain no more useful information.

4.4   Group Discovery Algorithm

Based on the analysis above, we present a new group discovery algorithm shown as Alg. 1. It consists of two main loops, one nested within the other. The inner iteration aims to identify one group, whereas the outer loop obtains all groups hidden behind data.

The algorithm starts at initializing relative parameters, such as normalizing X and Y. In the outer loop, the initial weights of variables v_k (or u_k) is set, guaranteeing the inner iteration convergence. For simplicity, here we take the first right singular vector of Y_k as the initial value of v_k. Once the vector has been assign an initial value, the inner iteration can solve the optimization problem of CCA by using ℓ_1-norm penalty. The process is repeated until the residual matrices have no more useful information, or k is larger than a pre-specified threshold K.

5   Simulation Experiments

5.1   Experiments on Synthetic Datasets

The synthetic data consists of two datasets. Each one represents the same 100 samples observed from different views. The first dataset X has 60 variables

---

Algorithm 1. Identifying groups and their relationships

```
1). Initialize parameters, e.g., X₁=X and Y₁=Y.
```
2). For $k = 1, ..., K$:
```
    2.1) Initialize v and normalize it.
    2.2) Repeat until convergence
        2.2.1) Obtain u by virtue of Eq. 7.
```
        2.2.2) Normalize u as $u = \frac{u}{\|u\|}$.
```
        2.2.3) Obtain v by virtue of Eq. 7.
```
        2.2.4) Normalize v as $v = \frac{v}{\|v\|}$.
    2.3) Estimate the correlation $\rho(Xu, Yv)$.
```
    2.4) Update the residual matrices Xₖ and Yₖ.
```

$\{x_1, ..., x_{60}\}$ that belongs to six groups $G_X = \{xg_1, ..., xg_6\}$ with 10 variables in each group. The second dataset Y represents the same samples from another view. It contains 150 variables $\{y_1, ..., y_{150}\}$ also organized into six groups $G_Y = \{yg_1, ..., yg_6\}$ with the mean number of variables. Links between G_X and G_Y are generated as follows. For each sample in each dataset, it is randomly assigned to one of groups and the probabilities of variables corresponding to the group are larger than 0.2. For other groups, the values are less than 0.1.

The experimental results on the artificial datasets are presented in Table 1, where $|xg|$ denotes the number of variables contained within the i-th group xg_i. From this table, we know that the proposed method has excellent performance. It not only identified all groups from both datasets, but also correctly measured the interrelations between them. Additionally, the correlations of the groups were also simultaneously calculated in a quantitative way after the groups were obtained. For example, the group xg_5 (the 6th line) obtained from X was exactly matched with the group yg_5 from Y, and their correlation coefficient was 0.832. From the perspective of individual groups, the performance of the pro-

Table 1. Experimental results on the synthetic data

No.	Group pair	$|yg|$	$|xg|$	Correlation coefficient
1	(xg_1, yg_1)	23	9	0.950
2	(xg_2, yg_2)	23	9	0.879
3	(xg_6, yg_6)	25	10	0.877
4	(xg_4, yg_4)	24	10	0.870
5	(xg_3, yg_3)	24	10	0.841
6	(xg_5, yg_5)	25	10	0.832

posed method is also quite well (see the $|xg|$ and $|yg|$ columns in Table 1). It successfully identified all groups from the artificial data. More importantly, the disclosed groups contain the right members, and are subsets of the corresponding assumption ones. For instance, the xg_2 group is a subset of the assumption one in X, for it contains all members, i.e., $x_{21}, ..., x_{30}$, except x_{22}. Besides, the size of each disclosed group also approaches that of the original one.

To further demonstrate the effectiveness, we calculated the similarity of the disclosed results to the assumption ones. Since the distributions of groups are known in advance, we took normalized mutual information (NMI) and Jaccard coefficient as our measurements in the simulation experiments. In our experiments, the values of NMI of the discovered groups $XG = \{xg_1, .., xg_6\}$ and $YG = \{yg_1, ..., yg_6\}$ to the assumption ones G_X and G_Y are 0.992 and 0.991, respectively. Correspondingly, the Jaccard coefficients are 0.967 and 0.96, respectively. These factors indicate that the proposed method is good at identifying groups from data and the groups identified by our method are quite similar the assumption ones.

5.2   Experiments on DBLP Datasets

A typical application of group discovery is science bibliography, where the groups of keywords and authors are quite ubiquitous. Here we also carried out experiments on the DBLP database [19]. For the convenience of discussion, we downloaded the DBLP database[1] and extracted 1071 papers published in five international conferences from 2000 to 2004. In our experiment, the papers were organized into two datasets, where the first one A involved 2022 authors, while the second one T contained 1065 terms extracted from paper titles.

After performing the proposed method on A and T, we obtained 143 groups of authors G_A and terms G_T, where the authors in the ith group G_{Ai} are associated with the terms within the ith group G_{Ti}. Table 2 lists 10 over 143 disclosed groups, where CC denotes the correlation coefficient of G_{Ai} to G_{Ti}, while $\#P, \#A$ and $\#T$ indicate how many papers, authors and terms were involved in each group. As an example, the 8th group of authors G_{A8} (G_{T8}) covers two papers involving 7 authors (7 terms). The topic of this pair mainly concerns *semantic* information by *parsing structures* of texts.

Table 2. Ten groups discovered from the DBLP corpus

No	CC	#P	#A	#T	Authors	Terms
Gp1	0.999	3	5	11	S Donoho, S Zhu, ...	option, markets, ...
Gp2	0.998	3	5	9	D Cau, S Acid, ...	effects, causality, ...
Gp4	0.994	3	6	6	C Borgelt, K Lou, ...	expectation, fuzzy,...
Gp5	0.993	4	6	5	ZW Ras, R Wong, ...	profit, rules, ...
Gp6	0.993	2	4	6	S Zhang, Y Zhu, ...	elastic, burst, ...
Gp8	0.991	2	7	7	S Pradhan, L Lloyd, ...	structure, parsing, ...
Gp11	0.990	2	3	4	M Sauban, BF White,...	profiling, document, ...
Gp25	0.986	2	3	5	J Adibi, E Oja,...	hidden, markov, ...
Gp38	0.979	3	3	6	PW Eklund, G Stumme,...	discovery, ontologies, ...
Gp107	0.946	3	9	6	P Kim, J Ye, ...	media, stream, ...

---

[1] `http://www.public.asu.edu/~{}ltang9/data/dblp.tar.gz`

An interesting discovery made by the proposed method is that co-authors were not always be included into the same author group. For instance, the three papers in the 107th author group had 10 authors totally, while only 9 was included within this author group. The missing co-author is *M. Vazirgiannis*. In fact, *M. Vazirgiannis* published more papers about *clustering* than *semantics*. This means that the results identified by our method rely on keywords in each group on the whole, rather simply collect co-occurrence information.

The interrelations between G_{Ai} and G_{Ti} were also estimated by the proposed method and illustrated as a dot graph (see Fig. 1), where the x-axis denotes the group pairs $Gp_i = (G_{Ai}, G_{Ti})$. The large values of correlation coefficients elucidate the effectiveness of our method, for in each group pair the disclosed groups with different types were tightly related with each other. Among the 143 coefficients, only six of them were lower than 0.90, whereas the minimal one was still larger than 0.81. This, however, is reasonable for the last six group pairs were identified upon the residual information the previously discovered groups left and covered more authors and terms.
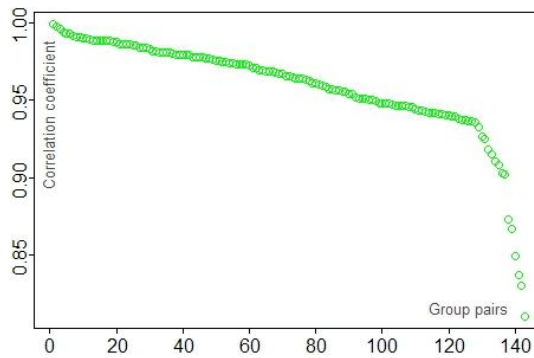

Fig. 1. The correlation coefficients of group pairs

Identifying hot topics is one of major tasks extensively studied in the field of author-topic models. This goal can also be achieved by our method. We obtained the hot topics in terms of the quantities of papers covered by the group pairs for the sake of simplification. Fig. 2 provides us an intuitive observation about the popularity of topics during the five years. From this graph, one may easily observe what topics are hot during the past years. For instance, *Gp97* followed by *Gp*140 and *Gp*135 was the most popular topic comprising 15 research papers. It concerned the study of *category*, *mining* and *relationship*, and included six authors, e.g., *P Yu*, *R Hilderman*, *H Hamilton*, etc. Another interesting fact is that identifying the groups and their relationships can also offer some insights into the research trends and the collaboration relationships between different research teams if the latest information is available.

Fig. 2. The cloud graph of groups

6   Conclusions

Generally, additional information from heterogeneous sources is helpful to analyze and understand network structures and functions. In this paper, we propose a statistical learning framework to discover groups and capture their interrelations from different data sources. The central idea of our method is to formulate group discovery as an optimization problem of CCA, and then extend it to a LASSO problem to achieve the sparse purpose. Within this framework, group discovery and their relationship measurement are turned out to be easily fulfilled in one stage. Simulation experiments were conducted on both carefully designed synthetic datasets and the DBLP corpus. The experimental results show that the proposed method tends to identify accurate group information and reveal useful insights in a given network from two different views.

References

1. Blei, D.: Introduction to probabilistic topic models. Comm. of the ACM 55(4), 77–84 (2012)
2. Chi, Y., Song, X., Zhou, D., Hino, K., Tseng, B.: Evolutionary spectral clustering by incorporating temporal smoothness. In: Proc. of the 13th ACM SIGKDD Int'l Conf. on Know. Disc. and Data Mining, pp. 153–162. ACM (2007)
3. Chiua, G., Westveld, A.: A unifying approach for food webs, phylogeny, social networks, and statistics. Proc. Natl. Acad. Sci. USA 108(38), 15881–15886 (2011)
4. Fortunato, S.: Community detection in graphs. Phy. Rept. 486(3), 75–174 (2010)
5. Getoor, L., Diehl, C.: Link mining: a survey. SIGKDD Explor. Newsl. 7, 3–12 (2005)

6. Hotelling, H.: Relations between two sets of variables. Biometrika 28(3/4), 312–377 (1936)
7. Jiang, J.Q., McQuay, L.J.: Modularity functions maximization with nonnegative relaxation facilitates community detection in networks. Phys. A 391(2), 854–865 (2012)
8. Krause, A., Frank, K., Mason, D., Ulanowicz, R., Taylor, W.: Compartments revealed in food-web structure. Nature 426(6964), 282–285 (2003)
9. Lancichinetti, A., Radicchi, F., Ramasco, J., Fortunato, S.: Finding statistically significant communities in networks. PLoS ONE 6(4), e18961 (2011)
10. Liu, B., Liu, L., Tsykin, A., Goodall, G., Green, J., Zhu, M., Kim, C., Li, J.: Identifying functional mirna-mrna regulatory modules with correspondence latent dirichlet allocation. Bioinformatics 26(24), 3105–3111 (2010)
11. Michal, R.Z., Chaitanya, C., Thomas, G., Padhraic, S., Mark, S.: Learning author-topic models from text corpora. ACM Trans. Inf. Syst. 28(1), Article 4 (2010)
12. Mucha, P., Richardson, T., Macon, K., Porter, M., Onnela, J.P.: Community structure in time-dependent, multiscale, and multiplex networks. Science 328, 876–878 (2010)
13. Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69(2), 026113 (2004)
14. Palla, G., Barabási, A.L., Vicsek, T.: Quantifying social group evolution. Nature 446(7136), 664–667 (2007)
15. Pons, P., Latapy, M.: Post-processing hierarchical community structures: Quality improvements and multi-scale view. Theo. Comp. Sci. 412(8), 892–900 (2011)
16. Scott, J.: Social Network Analysis: A Handbook. SAGE Publications, London (2000)
17. Serrour, B., Arenas, A., Gómez, S.: Detecting communities of triangles in complex networks using spectral optimization. Comp. Comm. 34(5), 629–634 (2011)
18. Shen, H.W., Cheng, X.Q., Fang, B.X.: Covariance, correlation matrix, and the multiscale community structure of networks. Phy. Rev. E 82(1), 016114 (2010)
19. Tang, L., Liu, H., Zhang, J., Nazeri, Z.: Community evolution in dynamic multi-mode networks. In: Proc. of the 14th ACM SIGKDD Intl' Conf. on Knowl. Disc. and Data Mining, pp. 677–685. ACM (2008)
20. Tang, L., Wang, X., Liu, H.: Community detection via heterogeneous interaction analysis. Know. Dis. Dat. Min. 25(1), 1–33 (2012)
21. Tibshirani, R.: Regression shrinkage and selection via the lasso: a retrospective. J. R. Statist. Soc. B 73(3), 273–282 (2011)
22. Yang, T., Chi, Y., Zhu, S., Gong, Y., Jin, R.: Detecting communities and their evolutions in dynamic social networks- a bayesian approach. Mach. Learn. 82(2), 157–189 (2011)
23. Yang, Z., Tang, J., Li, J.: Social community analysis via factor graph model. IEEE Intelligent Sys. 26(3), 58–65 (2011)