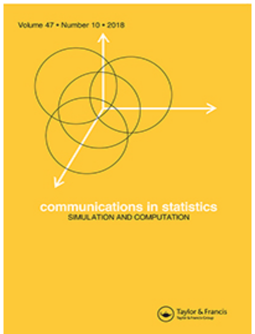


Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Semidefinite programming based community detection for node-attributed networks and multiplex networks

Fengqin Tang , Chunning Wang , Jinxia Su & Yuanyuan Wang

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2020.1847291

Semidefinite programming based community detection for node-attributed networks and multiplex networks

Fengqin Tang[a], Chunning Wang[b], Jinxia Su[b], and Yuanyuan Wang[b]

[a]School of Mathematics Sciences, Huaibei Normal University, Huaibei, China; [b]School of Mathematics and Statistics, Lanzhou University, Lanzhou, China

ABSTRACT

Community detection is an effective exploration technique for analyzing networks. Most of the network data not only describes the connections of network nodes but also describes the properties of the nodes. In this paper, we propose a community detection method collects relevant evidences from the information of node attributes and the information of network structure to assist the community detection task on node-attributed networks. We find communities in the framework of the semidefinite programming (SDP) method. In practical applications, the distribution of some node attributes may be uncorrelated with the network structure or the network itself may contain no communities as in a random graph. A sparse attribute self-adjustment mechanism is introduced to determine the relative importance of each source of information. As a by product, our method is also effective for community detection of multilayer networks that allow for multiple kinds of relations over the same set of nodes. Experimental results demonstrate the effectiveness of the proposed method.

ARTICLE HISTORY
Received 19 November 2019
Accepted 2 November 2020

KEYWORDS
Community detection; Multiplex networks; Semidefinite programming; Stochastic block model

MATHEMATICS SUBJECT CLASSIFICATION
62-07; 62P25

1. Introduction

Mining network information for identifying communities is an important issue across a number of research fields including but not limit to social science, biology, physics, engineering, etc (Girvan and Newman 2002; Guimera and Amaral 2005; Ahn et al. 2010; Dorogovtsev and Mendes 2013; Guédon and Vershynin 2016). In applications, most network data are represented by undirected graphs with edges describing interactions between nodes. Communities (groups or clusters) exist in most real-world networks and identifying these groups helps us better understand the behavior of the underlying system (Fortunato 2010; Fortunato and Hric 2016). Besides structural connections, nodes in real-world networks may have several attributes describing their properties. For example, users in a social network usually come with personal information like age, gender, birthplace, and so on. One should infer communities from the networks by considering both sources of information.

Recently, a large number of effective works have been done on identifying communities for node-attributed networks. Günnemann et al. (2013) propose a variant of EM algorithm based on spectral clustering for identifying communities with node attributes and detect each cluster favoring a subset of attributes. Zhang et al. (2016) introduce a joint criterion for community detection (JCDC) with node attributes. The authors elaborate that the contributions of different attributes may differ across communities. Neville, Adler, and Jensen (2003) and Steinhaeuser and Chawla (2010) use the

CONTACT Fengqin Tang tfq05@163.com School of Mathematics Sciences, Huaibei Normal University, Huaibei, China

similarity provided from node attributes as the weights of the original network edge, then apply the suitable community detection methods for weighted networks. Falih et al. (2017) and Jia et al. (2017) introduce a structural-based similarity between nodes and combine the similarity calculated according to nodes attributes, then, use some classic distance-based clustering algorithms to detect communities. Weng and Feng (2016) propose a flexible network model incorporating nodal information, and develop likelihood-based inference methods. Binkiewicz, Vogelstein, and Rohe (2017) propose covariate-assisted spectral clustering (CASC) where the degree of contributions between the information from structure and the information from node attributes are adjusted by a tuning parameter. Yan and Sarkar (2016) consider sparse networks in conjunction with finite dimensional sub-gaussian mixtures as covariates and propose an optimization framework based on semidefinite programming relaxations (SDP-comb). The authors add the attribute kernel function to the network structure cluster objective as a K-means type penalty and balance the structural information and node attribute information by a tuning parameter. All of the above works either uses node attributes unless there exist structural relationships between nodes, or treat all attributes fairly. However, many practices have shown that different node attributes have potentially different contributions for community detection ignoring the existence of structure relations between nodes. See Razaee (2017), Tang et al. (2019), Tang and Ding (2019), Zhou, Cheng, and Yu (2009), Yang, McAuley, and Leskovec (2013), Zanghi, Volant, and Ambroise (2010) surveys of this area.

In this paper, we propose a joint community detection approach for node-attributed networks from the perspective of multilayer networks. By multilayer networks, we mean the networks contain different aspects of interactions among nodes where each layer represents one type of interaction. For simplicity, we assume that all layers are defined on the same node set and this kind of multilayer networks are known as multiplex networks. Multiplex networks allow for multiple kinds of relations between the same set of nodes and capture more complexity real-world interactions than monolayer networks. We seek for a unified representation of the network structure and node attributes instead of regarding the node-attributed network as an augmented one. We do so by using different node attributes to generate different layers of the network. For each network layer constructed by node attributes, the edge weights between nodes are determined by the similarities along the specified node attribute. Parallelly, the original unweighted network is converted into a weighted network by imposing the Gaussian similarity function on the adjacency matrix. The generated network admits the same community structure of all layers to be the same, except for unrelated layers, which may have no community structure. We introduce an attribute weight-adjustment mechanical inspired by Sparse K-means (Witten and Tibshirani 2010) to determine the relative importance of the topology structure and individual attributes.

In fact, there is no need to manually generate multiplex networks as described above, and there are many natural multiplex networks in the real world. For example, a group of users may have connections on multiple social-media such as Facebook, Twitter, and LinkedIn. Much attention has been paid to the study of community detection for multiplex networks. For example, Huang et al. (2012) propose an affinity aggregation spectral clustering (AASC) algorithm which extends spectral clustering to a setting with multiplex affinities available; Wang et al. (2014) suggest to find a universal feature embedding for multi feature clustering and propose a multiview spectral clustering based minimax optimization method (MVSCW). Some other related works including maximizing the multiplex versions of normalized cut or generative model based approaches can be found in Schein et al. (2015), Valles-Catala et al. (2016), Zhou and Burges (2007) and Schein et al. (2016), among many others. Most of above works either collapse multilayer networks into a single-layer network or extend the methods for single-layer networks by using consensus clustering. Huang et al. (2012) and Wang et al. (2014) have considered the connection among various layers, and they believe that all network layers are more or less useful. One might expect that the ground truth communities present in the data differ only with respect

Table 1. Notation definitions.

Notations	Definition	Explanations
G, V, E	–	network (graph), node set, edge set
n, K	–	number of nodes, number of clusters
n_k	–	size of the k-th community
$A = (A_{ij}) \in \mathbb{R}^{n\times n}$	$A_{ij} = 1$ nodes i, j connect, otherwise $A_{ij} = 0$	adjacency matrix
$X \in \mathbb{R}^{n\times m}$	–	attribute matrix
$Z \in \{0,1\}^{n\times K}$	$Z = (Z_{ij})$	clustering indicator matrix
$Y = ZZ^T$	–	clustering matrix
$B \in [0,1]^{K\times K}$	$B = (B_{kl})$	link probability matrix of SBM
$X_{i\cdot} \in \mathbb{R}^m$	–	the attribute vector of the i-th node
$X_{\cdot j} \in \mathbb{R}^n$	–	the j-th attribute
π_i	$\frac{n_i}{n}$	proportion of nodes in the i-th clsuter
$\hat{n}$	$\min\{n_1, ..., n_k\}$	smallest cluster size
$\check{n}$	$\max\{n_1, ..., n_k\}$	largest cluster size
I_n	–	diagonal matrix of size $n \times n$
1_n	–	all one n-vector
E_n	–	all one matrix with size $n \times n$
C_k	$\{i : Z_{i,k} = 1\}$	node set for k-th cluster
$\mu_k = (\mu_{k1}, ..., \mu_{km})$	–	mean for $X_{i\cdot}$ if $i \in C_k$
d_{kl}	$\|\|\mu_k - \mu_l\|\|$	distance between clusters for attributes
$f(x) : \mathbb{R}_+ \to [0,1]$	$\exp(-x/\iota_i^2)$	kernel function for the i-th attribute
$M_i \in [0,1]^{n\times n}$	$M_i(k,l) = f(\|\|X_{ik} - X_{il}\|\|_2^2)$	kernel matrix for the i-th attribute
$Y_0 \in \mathbb{R}^{n\times n}$	$Zdiag(1/n_1, ..., 1/n_K)Z^T$	ground truth clustering matrix

to a small fraction of the layers, and will be masked if one uses all the layers fairly. The proposed method is devised based on these observations to address the problem of community detection over multilayer networks and node-attributed networks by a sparse layer (attribute) self-adjustment mechanism.

The paper is organized as follows. In Sec. 2, we show the notations, some existing works of semidefinite programming (SDP) and sparse clustering. Our contributions are established in Sec. 3. Section 4 shows the consistency analysis of our method. Section 5 shows the experimental results on simulations and real-world network data. We give a discussion in Sec. 6. The proofs of the consistency analysis of the proposed method are shown in Appendix A.

2. An overview of SDP and sparse clustering

2.1. Notation

Let B^T represent the transpose of the matrix B. For two matrices $B, H \in \mathbb{R}^{m\times n}$, the inner product is $\langle B, H\rangle = \text{trace}(B^T H)$. Let $\vec{\omega} = \{\omega_1, ..., \omega_m\}$ be a vector, $||\vec{\omega}||^2 = \sum_{k=1}^m |\omega_k|^2, ||\vec{\omega}||_1 = \sum_{k=1}^m |\omega_k|$. We summarize the notations used in this paper in Table 1.

2.2. Related works

Recently, Amini and Levina (2018) propose a unified SDP solution to the problem of fitting the stochastic block model (SBM). The authors maximize $\langle A, Y\rangle$ with respect to $Y = ZZ^T$, where A and Z stand for the adjacency matrix and the clustering indicator matrix, respectively. More related works can be found in Cai and Li (2014) and Chen and Xu (2016). The SDP relaxations can be further used for K-means type loss by maximizing $\langle M, Y\rangle$, where $M = (M_{ij})$ is the similarity matrix. One can choose $M_{ij} = X_i^T X_j$ for classical K-means whereas can choose $\exp\left(\frac{\|X_i - X_j\|_2^2}{\sigma}\right)$ for K-means in the kernel space (Dhillon et al. 2004; Schein et al. 2015; Mixon, Villar, and Ward 2016).

There has been a considerable work in the field of sparse clustering, such as Li et al. (2014), Wang and Zhu (2008), and Witten, Tibshirani, and Hastie (2009). Witten and Tibshirani (2010) propose a novel approach named sparse K-means which aims at optimizing the following objective function

$$\max_{\vec{\omega}\Theta} \sum_{i=1}^{m} \omega_i g_i(X_{\cdot i}, \Theta) \tag{1}$$

subject to $||\vec{\omega}||^2 \leq 1, ||\vec{\omega}||_1 \leq \kappa, \omega_i \geq 0, \forall i$. Here $\vec{\omega} = (\omega_1, ..., \omega_m)$, ω_i is the weight of the i-th attribute and κ is a positive constant. For K-means, $g_i(X_{\cdot i}, \Theta)$ represents the between-cluster sum of square along the i-th attribute $X_{\cdot i}$ and Θ stands for a partition of the observations into K disjoint sets.

Keeping Θ fixed, the optimization of Eq. (1) with respect to $\vec{\omega}$

$$\max_{\vec{\omega}} \{\vec{\omega}^T g(X, \Theta)\} \text{subject to} ||\vec{\omega}||^2 \leq 1, ||\vec{\omega}||_1 \leq \kappa, \omega_{\mathrm{i}} \geq 0, \tag{2}$$

is followed by the Proposition of Witten and Tibshirani (2010), here $g(X, \Theta) = \{g_1(X_{\cdot 1}, \Theta), ..., g_m(X_{\cdot m}, \Theta)\}, g_i(X_{\cdot i}, \Theta)$ is a convex function and Θ is a parameter set. We restate the Proposition of Witten and Tibshirani (2010) as follows.

Proposition A. The solution for the convex problem (2) is $\vec{\omega} = \frac{S((g(X,\Theta))_+, \Delta)}{||S((g(X,\Theta))_+, \Delta)||^2}$, where $(g(X, \Theta))_+$ denotes the positive part of $g(X, \Theta)$ and where $\Delta = 0$ if that results in $||\vec{\omega}||_1 \leq \kappa$; otherwise, $\Delta > 0$ is chosen to yield $||\vec{\omega}||_1 = \kappa, S(x, c) = \mathrm{sign}(x)(|x| - c)_+$.

3. Methodology

3.1. The proposed method

We consider the setting where the network is generated from the stochastic block model (SBM). Assume that the estimated clusters are consistent with the latent membership in the attribute space. For $i < j$, under the SBM, the A_{ij} are Bernoulli random variables, with

$$\Pr(A_{ij} = 1|Z) = Z_i^T B Z_j$$

where Z_i is the ith row of Z.

For node attributes, we focus on a general case where the data are sampled from a noisy version of manifold. For each attribute i,

$$X_{ji} = \sum_{k=1}^{K} Z_{jk}\mu_{ki} + W_j, 1 \leq j \leq n$$

where W_j are subgaussian random variables with mean zero, covariance σ_k^2 and subgaussian norm ψ_k for $j \in C_k$.

A wide range of investigations show that there exist disagreements between certain node attributes and the network structure. For example, students' gender attribute seems to play an important role for grouping students in a school friendship network, while the height attribute may not correlate with the groups we are interested in. On the other hand, the network may be a random graph and contains no communities while node attributes may have meaningful partition information. Thus, we need to learn the contributions of the network and individual attributes for community detection. We propose a general SDP relaxation (G-SDP) as:

Algorithm 1 General-SDP(G-SDP)

Input:
M_A, M_i: similarity matrices; K: cluster number; κ, ι, ι_i, ρ: tuning parameters;
Output: $\mathbf{C}$: cluster assignment; $\overrightarrow{\omega}$: weights of the network and attributes;
Algorithm:

(1) Initialize $\overrightarrow{\omega}, \omega_0^{(0)} = \omega_1^{(0)} = \ldots = \omega_m^{(0)} = \frac{1}{\sqrt{m+1}}$;

(2) Iterate until convergence.

(a) Holding $\overrightarrow{\omega}$ fixed, optimize (3) with respect to M_A and M_i using ADMM, while not converge do:

(a1) $Y^{(k+1)} = \text{Proj}_1\left(\frac{1}{2}(O^{(k)} - U^{(k)} + H^{(k)} - V^{(k)}) + \frac{1}{\rho}\sum_{i=1}^{m}\omega_i^{(k)}M_i + \omega_0^{(k)}M_A\right)$;

(a2) $O^{(k+1)} = \max(0, Y^{(k+1)} + U^{(k)}), O^{(k+1)} = \min(1/\hat{n}_k, O^{(k+1)})$;

(a3) $H^{(k+1)} = \text{Proj}_2(Y^{(k+1)} + V^{(k)})$;

(a4) $U^{(k+1)} = U^{(k)} + Y^{(k+1)} - O^{(k+1)}$;

(a5) $V^{(k+1)} = U^{(k)} + Y^{(k+1)} - H^{(k+1)}$;

(a6) Return $Y^{(k)}$, perform K-means on $Y^{(k)}$ and obtain $\mathbf{C}^{(k)}$.

(b) Holding $\mathbf{C}^{(k)}$ fixed, optimize (3) with respect to $\overrightarrow{\omega}^{(k)}$ by Proposition 2.1.

(3) The final clusters and weights are given by $\mathbf{C}$ and $\overrightarrow{\omega}$.

Figure 1. Pseudo-code of G-SDP alogorithm.

$$\begin{aligned}&\arg\ \max_{Y,\vec{\omega}}(\omega_0\langle M_A, Y\rangle + \sum_{i=1}^{m}\omega_i\langle M_i, Y\rangle)\\&\text{subject to } Y\geqslant 0, 0 \leq Y \leq 1\hat{n}, YI_n = I_n, \text{trace}(Y) = K,\\&\qquad ||\vec{\omega}||^2 \leq 1, ||\vec{\omega}||_1 \leq \kappa, \omega_i \geq 0, i = 0, 1, ..., m.\end{aligned} \tag{3}$$

Here M_A represents the kernel similarity matrix related to the original network, for example, we may set $M_A = \exp(A/\iota^2), M_i(1 \leq i \leq m)$ represent the kernel similarity matrix of the i-th attribute, for example, $(M_i)_{lk} = \exp\left(\frac{-\|X_{li}-X_{ki}\|^2}{\iota_i^2}\right)$, and ι, ι_i are tuning parameters in the Gaussian kernel function. The parameters ω_0 and $\omega_i(1 \leq i \leq m)$ are the weights of the network structure and the i-th attribute. For the simplicity of the notation, we still use $\vec{\omega}$ to represent $\{\omega_0, \omega_1, ..., \omega_m\}$. For model (3), the constraints on Y are similar to Peng and Wei (2007) and the constraints on $\vec{\omega}$ are the same as Sparse K-means constraints of Witten and Tibshirani (2010).

The optimization for (3) is processed by an iterative manner: keeping $\vec{\omega}$ or Y held fixed, we optimize (3) with respect to the other one. We use the alternating direction method of multipliers (ADMM) for the optimization step while keeping ω fixed. The iterative algorithm for maximizing (3) is presented in Figure 1.

Let Γ_i be a symmetric matrix with 1 in the off-diagonal elements of the i-th column and row, and 0 everywhere else. Denote $\Upsilon(Y)_i = \langle\Gamma_i, Y\rangle, i \in [n], \Upsilon(Y)_{n+1} = \langle I_n, Y\rangle$. Then, in Figure 1 $(a1)$, $\text{Proj}_1(Y) = Y - \Upsilon^*(\Upsilon\Upsilon^*)^{-1}(\Upsilon(Y) - \vec{b})$ where $(\Upsilon\Upsilon^*)^{-1} = \begin{bmatrix}\frac{1}{2n}(I_n - \frac{n-2}{2n(n-1)}E_n) & \frac{1}{2n(1-n)}1_n\\ \frac{1}{2n(1-n)}I_n & \frac{1}{n-1}\end{bmatrix}, \vec{b} = (2, ..., 2, K) \in \mathbb{R}^{n+1}$ and Υ^* is the adjoint of Υ. Denote the eigendecomposition of Y by $Y = \Psi\Lambda\Psi^T$ with $\Lambda = diag(\lambda_1, ..., \lambda_n)$, then, in Figure 1 $(a3)$, $\text{Proj}_2(Y) = \Psi diag(\max(0, \lambda_1), ..., \max(0, \lambda_n))\Psi^T$.

Algorithm 2 Selection of κ

(1) Obtain permuted data list $(M_A, M_1, \ldots, M_m)_1, (M_A, M_1, \ldots, M_m)_2, \ldots, (M_A, M_1, \ldots, M_m)_L$ by independently permuting the observations within M_A and each M_i.

(2) For each candidate tuning parameter value κ:

- **(2a)** Apply the proposed G-SDP method to compute $\Pi(\kappa) = \omega_0\langle M_A, Y\rangle + \sum_{i=1}^{m} \omega_i\langle M_i, Y\rangle$ on the original data.
- **(2b)** Compute $\Pi_l(\kappa)(1 \leq l \leq L)$, the same quality as in step (2a), but on the permuted data list obtained in step (1).
- **(2c)** Calculate $Gap(\kappa) = \log(\Pi(\kappa)) - \frac{1}{L}\sum_{l=1}^{L} \log(\Pi_l(\kappa))$

(3) Return $\hat{\kappa} = \arg\max_{\kappa} Gap(\kappa)$.

Figure 2. Pseudo-code of selection of κ.

3.2. Selection of tunning parameters

3.2.1. Selection of κ

We see that κ is the L_1 bound on $\vec{\omega}$ in model (3). The selection of κ is motivated by the method introduced in Witten and Tibshirani (2010), which builds on a permutation approach that is inspired by the gap statistic for selecting the number of clusters in K-means clustering (Tibshirani et al., 2001). Details can be found in Figure 2. An exhaustive search for κ requires L runs of the G-SDP algorithm. We set $L = 25$ as in the code from Witten and Tibshirani (2010).

3.2.2. Selection of $\imath, \imath_i(1 \leq i \leq m)$

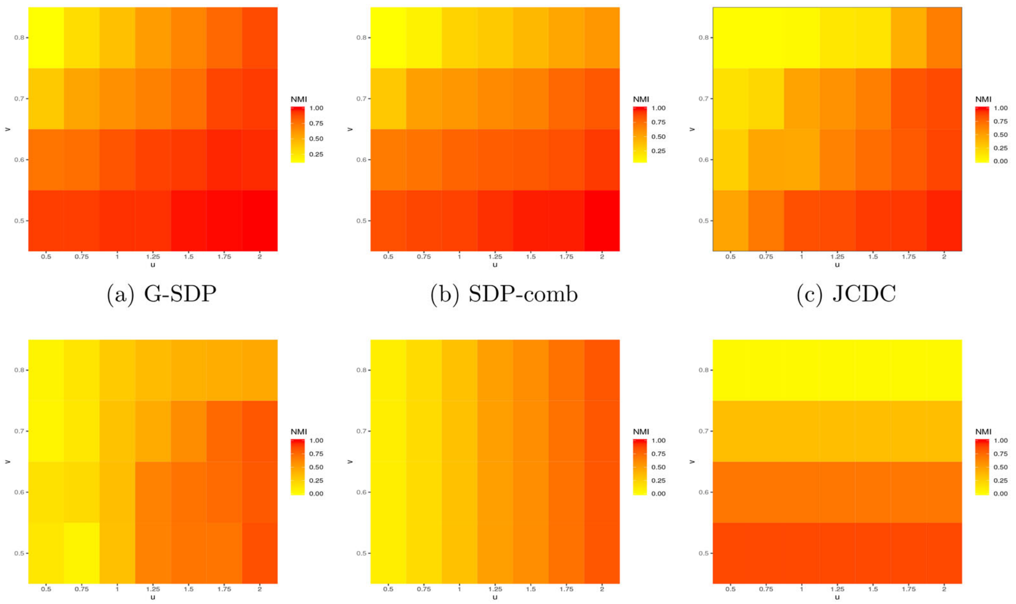
Note that $\imath, \imath_i(1 \leq i \leq m)$ are tuning parameters in the Gaussian similarity function controlling the width of the neighborhoods. In many real-world settings, there are several rules of thumb used to choose $\imath, \imath_i$, for example, Luxburg (2007) shows that one can choose $\imath, \imath_i$ in the order of the mean distance of a point to its k-th nearest neighbor with $k \sim log(n) + 1$, or can choose $\imath, \imath_i$ as the length of the longest edge in a minimal spanning tree of the fully connected graph on the data points. In the proposed method, we use the latter scheme to choose the tuning parameters and Tang et al. (2019) show the effectiveness of the selection of the parameters. However, it is generally hard to possess the theoretical foundation for adjusting parameter selection.

4. Consistency analysis

For fixed $w_i(i = 0, 1, \ldots, m)$, we consider the consistency issue with respect to following object function

$$\begin{aligned} &Y_{comb} = \arg\max_{Y} \left\{ \omega_0\langle M_A, Y\rangle + \sum_{i=1}^{m} \omega_i\langle M_i, Y\rangle \right\} \\ &\text{subject to } Y \succcurlyeq 0, 0 \leq Y \leq 1\hat{n}, YI_n = \frac{n}{K}I_n, \text{trace}(Y) = K. \end{aligned} \tag{4}$$

Related works on consistency issues of the semidefinite optimization including: Guédon and Vershynin (2016) and Mixon et al. (2016) obtain a relative error bound of the estimated matrix from the ideal clustering matrix; Yan and Sarkar (2016) investigate community detection in networks in the presence of node covariates and show the relative error of the solution to the combined SDP matrix from the ideal clustering matrix. Based on these, we further provide an upper



(a) G-SDP (b) SDP-comb (c) JCDC

(d) CASC (e) K-means (f) SPC

Figure 3. The average NMI values of different community detection methods as u and v vary.

bound of $||Y_{comb} - Y_0||_F^2$ with Y_{comb} standing for the optimal solution of object function (4). We show that this upper bound is a tighter quantity than a simple linear combination of the separations stemming from the network structure and each attribute.

Let $\Delta_k^{(i)} = c_k^{(i)}\psi_k\sqrt{m}$ for some positive constant $c_k^{(i)}, 1 \le k \le K, 1 \le i \le n$(Yan and Sarkar, 2016). For each attribute $X_{\cdot i}$, set $G_k^{(i)} = \{j \in C_k : ||X_{ji} - \mu_{ki}|| \le \Delta_k^{(i)}\}$ as the group of ideal clustering nodes which close to their population mean in the k-th cluster. Define $t_k^{(i)} = f(2\Delta_k^{(i)}), r_k^{(i)} = \max_{k \ne s} f(d_{ks} - \Delta_k^{(i)} - \Delta_s^{(i)}), v_k^{(i)} = t_k^{(i)} - r_k^{(i)}$. Let $Y_0 = Z\Lambda\Lambda^T Z^T$ be the ground truth clustering matrix with $\Lambda = diag\left(\frac{1}{\sqrt{n_1}}, \ldots, \frac{1}{\sqrt{n_K}}\right)$.

A common choice for reference matrix of the adjacency matrix A is $\mathbb{E}[A|Z]$. For M_A, we define the reference matrix Ω_A as

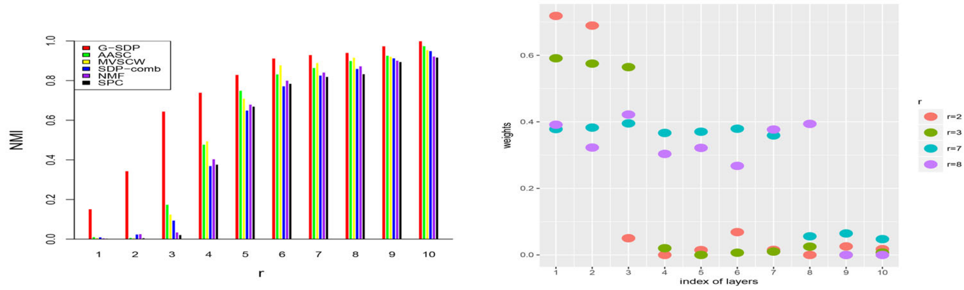
$$(\Omega_A)_{jl} = \begin{cases} B_{kk}\left(e^{\frac{1}{t^2}} - 1\right) + 1 & \text{if } j, l \in C_k \\ B_{ks}\left(e^{\frac{1}{t^2}} - 1\right) + 1 & \text{if } j \in C_k, l \in C_s, k \ne s. \end{cases}$$

Inspired by Yan and Sarkar (2016), we set the reference matrix of M_i as

$$(\Omega_i)_{jl} = \begin{cases} f(2\Delta_k^{(i)}) & \text{if } j, l \in C_k \\ \min\{f(d_{ks} - \Delta_k^{(i)} - \Delta_s^{(i)}), (M_i)_{jl}\} & \text{if } j \in C_k, l \in C_s, k \ne s. \end{cases}$$

Denote $\gamma = \min_k(\omega_0(B_{kk} - \max_{s \ne k} B_{ks})\left(e^{\frac{1}{t^2}} - 1\right) + \sum_{i=1}^m \omega_i(f(2\Delta_k^{(i)}) - \max_{s \ne k} f(d_{ks} - \Delta_k^{(i)} - \Delta_s^{(i)})))$ for some $\Delta_k^{(i)}$ such that $\gamma \ge 0$.

Theorem 1. *Let* $g = \frac{2}{n-1}\sum_{i<j} Var(e^{A_{ij}}) \ge 9, n_k = n\pi_k, \hat{n} = n\pi_{min},$ *and* $\pi_0^{(i)} = \sum_k (n_k \exp\{-\Delta_k^{(i)2}/5\psi_k^2\} + \sqrt{n_k \log n_k/2})$. *If* $\gamma > 0$ *and* $\pi_{min} = \Theta(1)$, *then, with probability tending to one,*



(a) (b)

Figure 4. (a) Average NMI values of different community detection methods as r varies. (b) Estimated weights obtained by G-SDP as r varies.

$$\|Y_{comb} - Y_0\|_F^2 \leq \frac{4K_G}{\pi_{min}^2\gamma}\left(6\omega_0\frac{\sqrt{g}}{n} + \sum_{i=1}^{m}\omega_i(2\frac{\pi_0^{(i)}}{n} + \sum_k \pi_k^2(1 - f(2\Delta_k^{(i)})))\right)$$

The proof of Theorem 1 is given in Appendix A.

5. Empirical studies

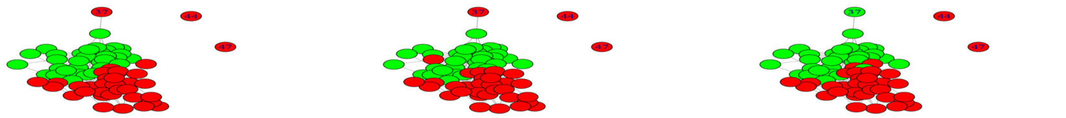
5.1. Numerical simulations

5.1.1. Simulation 1

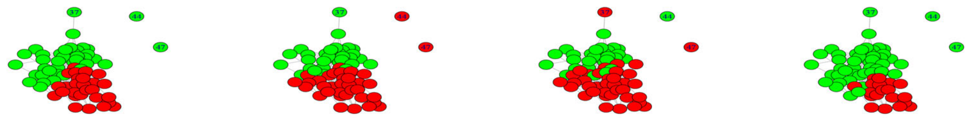
The first simulation demonstrates the performance of different methods as the information of the network structure and the information of attributes vary. We create a data set of $n = 350$ nodes following the degree-corrected stochastic block model (DCSBM) with 2 clusters, where one cluster has 150 nodes and the other has 200 nodes. The within-group probability is $\theta_i\theta_j p$ while the cross-group probability is $\theta_i\theta_j \nu p$ with $p = 0.1$, $\nu \in [0.5, 0.8]$. We set 95% of the nodes in each cluster with the degree-corrected parameter $\theta_i = 1$ and 5% of the nodes with $\theta_i = 10$. Each node has five attributes, four of which follow multivariate normal distributions $\mathcal{N}((u, 0, 0, 0), \Sigma_1)$ and $\mathcal{N}(-(u, 0, 0, 0), \Sigma_2)$ in the two clusters respectively, where $u \in [0.5, 2]$. The remaining attribute is uniform distributed for all nodes.

We present a qualitative comparison of G-SDP with five community detection methods CASC, JCDC, SDP-comb, spectral clustering (SPC), and K-means. Among these methods, G-SDP, JCDC, CASC, and SDP-comb use both structural information and node attributes information. SPC and K-means depend on either the structural adjacency matrix or node attributes. We use normalized mutual information (NMI, Fortunato (2010)) to measure the clustering quality of all comparison algorithms.

We repeat 50 independent runs each algorithm. Figure 3 shows the heatmaps of average NMI for all methods as ν and u vary. Since K-means uses only node attributes, its performance is only affected by u (Figure 3(e)). Analogously, SPC uses only the network structure information, thus its performance is only affected by ν (Figure 3(f)). For the cases where the community structure and node attributes are informative for identifying communities (small ν and large u), SDP-comb can obtain nearly perfect clustering results as G-SDP does. However, in the case where the structural information gets confusing, i.e., ν becomes large, SDP-comb performs worse than G-SDP does. This is probably because G-SDP considers the different contributions of network structure and each node attribute rather than weighing the relative importance of the network structure and overall node attributes via a tuning parameter.



(a) Status (b) G-SDP, NMI=0.68 (c) SDP-comb, NMI=0.58



(d) JCDC, NMI=0.44 (e) CASC, NMI=0.49 (f) K-means, NMI=0.42 (g) SPC, NMI=0.4

Figure 5. (a) Network colored by status, where green is partner and red is associate. (b)–(g) Results on the lawyers friendship network estimated using different methods. Colors match those in (a) in the best way possible.

The employment of the node attributes depends on the existence of the structural interactions in the JCDC algorithm. Thus, as v gets large, the potential communities get mixed, we see poorer community detection results returned by JCDC than expected. Community sizes and hub nodes appear to affect the clustering quality of CASC because it inherits the nature of spectral clustering. We range over cases where the community structure is relatively easy to infer to those that is hard to infer by varying u and v, and find that G-SDP achieves the highest NMI in each case.

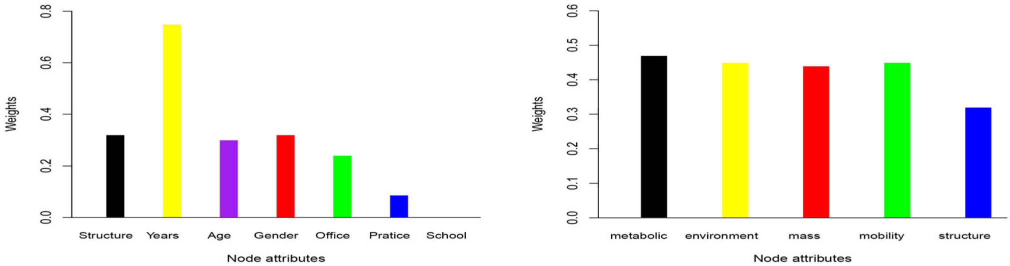
5.1.2. Simulation 2

In this part, we evaluate the performance of G-SDP for identifying communities in multiplex networks with redundant layers. The community structure of all informative layers is assumed to be the same. The comparison methods include MVSCW, AASC, SDP-comb, SPC, and NMF. The NMF algorithm is a nonnegative matrix factorization method that works on the flattened single-layered network of the multiplex networks (Lee and Seung 2001). Specially, SDP-comb, SPC, and NMF are single-view methods, and for comparison, we aggregate the multiplex network into a single one with equal weights for each layer.

The network contains ten layers and each layer has 200 nodes. The first r layers are generated by DCSBM with within-group connection probability $0.1\theta_i\theta_j$ and cross-group connection probability $0.07\theta_i\theta_j$. Here θ_i are defined as in Simulation 5.1.1. Each of these r layers has two equal-sized clusters. The remaining $10 - r$ layers are independent Erdős-Rényi (ER) random graphs with connection probability $p = 0.1$. We manually vary r from 1 to 10 in increments 1. For each r, we run G-SDP and the competition methods on 50 replications. The quality of all methods is measured by NMI as well and the results are shown in Figure 4.

Figure 4(a) shows that for all values of r, the average NMI values of G-SDP are substantially larger than those of other competing methods. Even though in the case where the layers with signal community structure are heavily interfered by the noisy ER layers ($r < 4$), G-SDP can still identify the noisy layers by the sparse weight-adjustment mechanism. However, the performances of MVSCW and AASC are not significantly better than those of NMF and SPC as r takes small values.

To address the ability of the proposed algorithm to identify the relative importance of each layer for the community detection problem in a multiplex network, we record the estimated weights of the layers as r varies. To conserve space, we show the results of four instances where $r = 2, 3, 7, 8$ in Figure 4(b). By the plot, we see that G-SDP has the advantage to learn the contributions of different layers by assigning them reasonable weights. Overall, the average NMI values imply that G-SDP outperforms other methods even in cases where only a fraction of the network layers are informative.



(a) Attribute Weights (b) Attribute Weights

Figure 6. (a) Attribute weights returned by G-SDP in lawyers friendship. (b) Attribute weights returned by G-SDP in Weddell sea network.

5.2. Real-world data set

5.2.1. The lawyers friendship network

The first data set describes the law partnership in a Northeastern US corporate (Lazega 2001). Here we focus on the friendship among lawyers. Additionally, each lawyer has seven attributes describing their personal information, including formal status, office, gender, practice, law school, age, and year with the firm. The ground truth communities of the network are unknown. Inspired by Zhang et al. (2016), we take the status partition as the ground-truth communities. The other attributes are taken as auxiliary information for finding communities. The final attributed network provides a symmetrical adjacency matrix describing the friendship among 71 lawyers with 399 undirected edges.

The community detection results for all methods are shown in Figure 5. There are two isolated nodes, numbered 44 and 47, in the network, thus, SPC, which is a structural-based algorithm, can not assign them into the correct group. JCDC fails to find the correct division for nodes 44 and 47 because it can not use node attributes without the existence of the structural relationship between nodes. Since K-means only relies on the node attributes and treats them equally, it can not assign node 44 correctly. Figure 5(a) shows that node 37 has only one connection, however it contains rich attribute information, thus its partition depends heavily on the attribute information. The methods SDP-comb, JCDC, CASC, and SPC all fail to recover the status partition of node 37 because of its confusing structural information. SDP-comb improves upon semidefinite programming relaxations by using node attributes. Due to two isolated nodes, the average NMI value of JCDC is lower than that of CASC. The weights of the node attributes obtained by G-SDP are shown in Figure 6(a). In summary, G-SDP achieves the highest NMI value and learns the relative importance of all information in the network as well.

5.2.2. Weddell sea dataset

We consider a ecological data set extracted in Jacob et al. (2011), which describes directed predator–prey interactions between 488 marine species of the marine ecosystem of Weddell Sea, a large bay off the coast of Antarctica. In the network, species are represented by nodes, the predator–-prey interactions are represented by directed edges. There are several node attributes contained in the dataset including prey types, feeding mode, metabolic type, mobility category, zone within the ocean, the average adult body mass, and others. As discussed in the previous work of Newman and Clauset (2016), here we compare partitions found by all the competing methods to the prey types. The attributes–metabolic type, environment, the mean body mass, and mobility type are appended to the predator–prey interactions between species.

G-SDP achieves the highest NMI value by using the node attributes reasonably as shown in Table 2. The application of node attributes in JCDC algorithm depends on the network structure,

Table 2. Average NMI obtained by different methods in Weddell sea dataset.

Methods	NMI	Methods	NMI
G-SDP	**0.58**	CASC	0.32
SDP-comb	0.48	K-means	0.3
JCDC	0.29	SPC	0.22

thus JCDC performs similarly to SPC, none of them can distinguish hebivore from omnivorous. K-means only relies on the node attributes and miscluters the herbivores and omnivores. CASC is based on spectral clustering and improves its performance by combining node attributes. The performance of SDP-comb is slight worse than that of G-SDP because SDP-comb uses all node attributes in the same way. We further establish the ability of learning the importance of the structure and individual attributes of the proposed G-SDP algorithm by listing the weights of the attributes in Figure 6(b).

5.2.3. Au-CS network

The dataset describes online and offline relationships among 61 employees of a Computer Science research department in Austria (Wilson et al. 2017). The employees are represented by nodes. There are five different relationships among the employees, such as Facebook, leisure, work, coauthor, and lunch, which are represented by five layers of the network. The number of edges in each layer of the network is shown in Table 3, and the number of edges varies from several hundred to several tens. We compare the performance of G-SDP with MVSCW and AASC. There is no additional information in the data to determine the potential labels of the nodes or the cluster number, following the pre-processing steps discussed in Wilson et al. (2017), we set the cluster number $K = 6$ for all methods. We assume that the network community structure of all layers is the same. Additionally, we use the density metric at each layer of the network to evaluate the quality of different methods. The density is defined as follows:

$$density(\{C_k\}_{k=1}^K) = \sum_{k=1}^{K} \frac{|\{(v_p, v_q)|v_p, v_q \in C_k, (v_p, v_q) \in E\}|}{|E|}.$$

The performance of all methods is illustrated in Figure 7. Figure 7(a,b) suggests that G-SDP provides a natural visualization of the lunch layer and the work layer, and one can see almost six communities identified by the method. These findings suggest that there is a close co-occurrence of work and lunch interactions among the employees. Figure 7(c) shows that the network structure of the leisure layer is a bit blurry. One possible explanation is that the partners usually have working lunch together and their entertainment time may cross, but not too much. Due to the sparseness of the Facebook layer, the proposed method only finds one community at this layer (Figure 7(d)), which is even worse in the coauthor layer (Figure 7(e)). As shown in Table 3, the weights of the lunch layer and work layer obtained by G-SDP are relative large (i.e., 0.66, 0.52), which coincides with the performance of G-SDP on these two layers. The other two methods, MVSCW and AASC are also able to identify the community structure of the signal layers, such as the lunch and the work layers (Figure 7(f)–(o)). The density values of different methods are shown in Table 3. In summary, the proposed method G-SDP significantly outperforms the other two methods.

5.2.4 C. elegans multiplex connectome

The *Caenorhabditis elegans* (*C. elegans*) connectome dataset is a multiplex neuronal network describing the full mapped brain of a living organism. The network consists of 279 neurons and more than three thousand connections among them (Chen et al. 2006; De Domenico, Porter, and Arenas 2015). The multiplex network includes three layers corresponding to different synaptic

Table 3. AU-CS network.

	Lunch	Work	Leisure	Facebook	Coauthor
Number of edges	386	388	176	248	42
Weights returned by G-SDP	0.66	0.52	0.36	0.32	0.22
Density returned by G-SDP	0.82	0.59	0.71	0.50	0.95
Density returned by MVSCW	0.79	0.59	0.70	0.49	0.95
Density returned by AASC	0.75	0.55	0.68	0.53	0.9

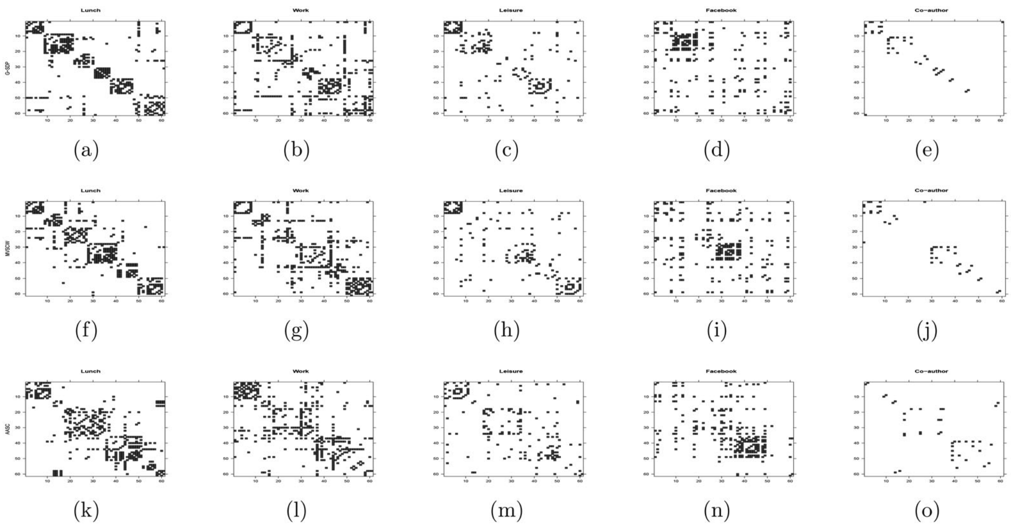


Figure 7. The AU-CS multiplex network. (a)–(e) Estimated results on five layers of the network returned by G-SDP. The vertices have been ordered based on the six communities identified by G-SDP. (f)–(j) Estimated results on five layers of the network returned by MVSCW. The vertices have been ordered based on the six communities identified by MVSCW. (k)–(o) Estimated results on five layers of the network returned by AASC. The vertices have been ordered based on the six communities identified by AASC.

Table 4. *C. elegans* multiplex connectome.

	ElectrJ	MonoSyn	PolySyn
Number of edges	1031	888	1703
Weights returned by G-SDP	0.3	0.52	0.81
Density returned by G-SDP	0.82	0.83	0.87
Density returned by MVSCW	0.81	0.8	0.77
Density returned by AASC	0.78	0.83	0.85

junctions: electric (ElectrJ), chemical monadic (MonoSyn), and polyadic (PolySyn). The details of the network are shown in Table 4. Since the ground truth communities are typically unknown in the network, we take density as the evaluation metric used in Sec. 5.2.3. The cluster number is chosen by eigengap heuristic. Precisely, if the eigenvalues $\lambda_1, \ldots, \lambda_K$ of the graph Laplacian are small, but λ_{K+1} is relatively large, then the cluster number is K. Here, we take $K = 2$. Table 4 shows the clustering results of the methods G-SDP, MVSCW, and AASC. We can see that the proposed G-SDP method obtains the highest density for each layer. The weights estimated by G-SDP imply that the PolySyn layer plays a more important role for detecting communities than the other two layers.

6. Discussion

The proposed G-SDP method can learn the relative importance of the network and each attribute using a feature selection framework inspired by sparse K-means. Furthermore, the proposed method can handle the community detection problem for multiplex networks and identify the contribution of each layer. We demonstrate the performance of our method on synthetic networks and real-world networks. The estimated results reveal superiority of the proposed method for community detection in node-attributed networks and multiplex networks. Our current constrained method assumes homogeneous structures among layers. Further direction includes considering multiplex networks with intralayer and interlayer heterogeneous structures and overlapping intralayer edges.

Acknowledgements

We thank Yuan Zhang and Yan B for generously sharing their code. We thank Professor Bingyi Jing for discussions related to this paper.

Appendix A

We start with a lemma from Yan and Sarkar (2016), which bounds the spectral norm of the difference between the solution to an SDP with input matrix M to its ideal clustering matrix.

Lemma 1. *Let $\hat{Y}$ be the optimal solution of the following SDP:*

$$\begin{aligned}&\max_Y \langle M, Y\rangle\\ &\text{subject to } Y \succcurlyeq 0, 0 \leq Y \leq 1\hat{n}, YI_n = \frac{n}{K} I_n, \text{trace}(Y) = K.\end{aligned}$$

Let Ω be a matrix, where $\Omega_{ij} = \begin{cases} \alpha_k & \text{if} i,j \in C_k; \\ \Omega_{ij} \geq \beta_k, & \text{if} i \in C_k, j \in C_l, l \neq k. \end{cases}$ Let Y_0 be the ideal clustering matrix. If $\min_k(\alpha_k - \beta_k) \geq 0$, then

$$||\hat{Y} - Y_0||_F^2 \leq \frac{2\langle M - \Omega, \hat{Y} - Y_0\rangle}{\hat{n}\min_k(\alpha_k - \beta_k)}.$$

The second lemma is from Guédon and Vershynin (2016).

Lemma 2 (Grothendieck inequality). Let $\Xi = \{Z : Z \succcurlyeq 0, diag(Z)\, I_n\}, A = (A_{ij}) \in \mathbb{R}^{n\times n}$ be a symmetric matrix whose diagonal entries equal 0, and $A_{ij} (i < j)$ are independent random variables in $[0, 1]$. Let $\mathbb{E}_A = \mathbb{E}[A|Z]$. Assume that $\bar{p} := \frac{1}{n(n-1)}\sum_{i<j} Var(A_{ij}) \geq \frac{9}{n}$. Then, with probability at least $1 - e^3 5^{-n}$, we have

$$\max_{Z\in\Xi} |\langle A - \mathbb{E}_A, Z\rangle| \leq K_G||A - \mathbb{E}_A||_{l_\infty\to l_1} \leq 3K_G\bar{p}^{1/2}n^{3/2},$$

where K_G is the Grothendieck constant.

Proof **of Theorem 1.** Using Lemma 1 with $\Omega = \Omega^{(A)} + \sum_{i=1}^m \Omega^{(i)}$ yields that

$$||Y_{comb} - Y_0||_F^2 \leq \frac{2}{\hat{n}\gamma}(\omega_0\langle M_A - \Omega^{(A)}, Y_{comb} - Y_0\rangle + \sum_{i=1}^m \omega_i\langle M_i - \Omega^{(i)}, Y_{comb} - Y_0\rangle).$$

Yan and Sarkar (2016) have proved that, for each i,

$$\|M_i - \Omega^{(i)}\|_{l_\infty\to l_1} \leq \sum_k n_k^2(1 - f(2\Delta_k^{(i)})) + 2n_c^{(i)}n \tag{5}$$

where $n_c^{(i)} = \sum_{k=1}^m (n_k - |G_k^{(i)}|)$. By a simple use of Grothendieck inequality and combining with Eq. (5) yield that

$$\begin{aligned}||Y_{comb} - Y_0||_F^2 &\leq \frac{2K_G}{\hat{n}^2\gamma}\left(\omega_0\|M_A - \Omega^{(A)}\|_{l_\infty\to l_1} + \sum_{i=1}^m \omega_i\|M_i - \Omega^{(i)}\|_{l_\infty\to l_1}\right)\\ &\leq \frac{4K_G}{\hat{n}^2\gamma}\left(6\omega_0 n\sqrt{g} + \sum_{i=1}^m \omega_i(2n_c^{(i)}n + \sum_k n_k^2(1 - f(2\Delta_k^{(i)})))\right).\end{aligned}$$

Furthermore, Yan and Sarkar (2016) point out that with probability tending to one,

$$n_c^{(i)} \leq \sum_k \left(n_k \exp\{-(\Delta_k^{(i)})^2/5\psi_k^2\} + \sqrt{n_k \log n_k/2}\right).$$

Thus,

$$\begin{aligned}||Y_{comb}-Y_0||_F^2 &\le \frac{4K_G}{\hat{n}^2\gamma}\left(6\omega_0 n\sqrt{g}+\sum_{i=1}^{m}\omega_i(2n_c^{(i)}n+\sum_k n_k^2(1-f(2\Delta_k^{(i)})))\right)\\ &\le \frac{4K_G}{\pi_{min}^2\gamma}\left(6\omega_0\frac{\sqrt{g}}{n}+\sum_{i=1}^{m}\omega_i(2\frac{\pi_0^{(i)}}{n}+\sum_k \pi_k^2(1-f(2\Delta_k^{(i)})))\right).\end{aligned}$$

Funding

This work is supported by the National Natural Science Foundation of China (11971214), Natural Science Foundation of the AnHui Higher Education Institutions of China (KJ2020A0023, KJ2020A0035, KJ2020ZD008).

ORCID

Fengqin Tang http://orcid.org/0000-0001-7234-1419
Jinxia Su http://orcid.org/0000-0002-2338-4786

References

Ahn, Y.-Y., J. P. Bagrow, and S. Lehmann. 2010. Link communities reveal multiscale complexity in networks. *Nature* 466 (7307):761–64. doi:10.1038/nature09182.

Amini, A. A., and E. Levina. 2018. On semidefinite relaxations for the block model. *Annals of Statistics* 46 (1): 149–79. doi:10.1214/17-AOS1545.

Binkiewicz, N., J. T. Vogelstein, and K. Rohe. 2017. Covariate-assisted spectral clustering. *Biometrika* 104 (2): 361–77. doi:10.1093/biomet/asx008.

Cai, T. T., and X. Li. 2014. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Annals of Statistics* 43:5–24.

Chen, B. L., D. H. Hall, and D. B. Chklovskii. 2006. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences of the United States of America* 103 (12):4723–28. doi:10.1073/pnas.0506806103.

Chen, Y., and J. Xu. 2016. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research* 17:882–938.

De Domenico, M., M. A. Porter, and A. Arenas. 2015. MuxViz: A tool for multilayer analysis and visualization of networks. *Journal of Complex Networks* 3 (2):159–76. doi:10.1093/comnet/cnu038.

Dhillon, I. S., Y. Guan, and B. Kulis. 2004. Kernel k-means: Spectral clustering and normalized cuts. Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Seattle, WA, 551–56.

Dorogovtsev, S. N., and J. F. F. Mendes. 2013. Evolution of networks: From biological nets to the Internet and WWW. London: Oxford University Press.

Falih, I., N. Grozavu, R. Kanawati, and Y. Bennani. 2017. ANCA: Attributed network clustering algorithm. In *Complex networks and their applications VI*, eds. C. Cherifi, H. Cherifi, M. Karsai, and M. Musolesi, 241–52. Cham: Springer.

Fortunato, S. 2010. Community detection in graphs. *Physics Reports* 486 (3-5):75–174. doi:10.1016/j.physrep.2009.11.002.

Fortunato, S., and D. Hric. 2016. Community detection in networks: A user guide. *Physics Reports* 659:1–44. doi:10.1016/j.physrep.2016.09.002.

Girvan, M., and M. E. J. Newman. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 7821–26.

Guédon, O., and R. Vershynin. 2016. Community detection in sparse networks via Grothendieck's inequality. *Probability Theory and Related Fields* 165 (3-4):1025–49. doi:10.1007/s00440-015-0659-z.

Guimera, R., and L. A. N. Amaral. 2005. Functional cartography of complex metabolic networks. *Nature* 433:895.

Günnemann, S., I. Färber, S. Raubach, and T. Seidl. 2013. Spectral subspace clustering for graphs with feature vectors. 13th International Conference on Data Mining, Austin, Texas, 231–40.

Huang, H. C., Y. Y. Chuang, and C. S. Chen. 2012. Affinity aggregation for spectral clustering. IEEE conference on computer vision and pattern recognition, Providence, Rhode Island, 773–80.

Jacob, U., A. Thierry, and U. Brose. 2011. The role of body size in complex food webs: A cold case. *Advances in Ecological Research* 45:181–223.

Jia, C., Y. Li, M. B. Carson, X. Wang, and J. Yu. 2017. Node attribute-enhanced community detection in complex networks. *Scientific Reports* 7 (1):2626. doi:10.1038/s41598-017-02751-8.

Lazega, E. 2001. *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford: Oxford University Press on Demand.

Lee, D. D., and H. S. Seung. 2001. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems* 13:556–62.

Li, Z., J. Liu, Y. Yang, X. Zhou, and H. Lu. 2014. Clustering-guided sparse structural learning for unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering* 26:2138–50.

Luxburg, V. 2007. A tutorial on spectral clustering. *Statistics and Computing* 17:395–416.

Mixon, D. G., S. Villar, and R. Ward. 2016. Clustering subgaussian mixtures by semidefinite programming. arXiv preprint arXiv:1602.06612.

Neville, J., M. Adler, and D. Jensen. 2003. Clustering relational data using attribute and link information. In *Proceedings of the text mining and link analysis workshop, 18th international joint conference on artificial intelligence*, Stockholm, Sweden, 9–15.

Newman, M. E. J., and A. Clauset. 2016. Structure and inference in annotated networks. *Nature Communications* 7:11863. doi:10.1038/ncomms11863.

Peng, J., and Y. Wei. 2007. Approximating k-means-type clustering via semidefinite programming. *SIAM Journal on Optimization* 18 (1):186–205. doi:10.1137/050641983.

Razaee, Z. 2017. *Community detection in networks with node covariates*. Ph. D. diss. Los Angeles, CA: UCLA.

Schein, A., J. Paisley, D. M. Blei, and H. Wallach. 2015. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Sydney, NSW, Australia, 1045–54.

Schein, A., M. Zhou, D. M. Blei, and H. Wallach. 2016. Bayesian poisson tucker decomposition for learning the structure of international relations. arXiv preprint arXiv:1606.01855.

Steinhaeuser, K., and N. V. Chawla. 2010. Identifying and evaluating community structure in complex networks. *Pattern Recoganition Letters* 31 (5):413–21.

Tang, F., and W. Ding. 2019. Community detection with structural and attribute similarities. *Journal of Statistical Computation and Simulation* 89 (4):668–85. doi:10.1080/00949655.2019.1568435.

Tang, F., Y. Wang, J. Su, and C. Wang. 2019. Spectral clustering-based network community detection with node attributes. *Statistics and Its Interface* 12 (1):123–33. doi:10.4310/SII.2019.v12.n1.a11.

Tibshirani, R., G. Walther, and T. Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2):411–23. doi:10.1111/1467-9868.00293.

Valles-Catala, T., F. A. Massucci, R. Guimera, and M. Sales-Pardo. 2016. Multilayer stochastic block models reveal the multilayer structure of complex networks. *Physical Review X* 6:011036.

Wang, H., C. Weng, and J. Yuan. 2014. Multi-feature spectral clustering with minimax optimization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, 4106–13.

Wang, S., and J. Zhu. 2008. Variable selection for model-based high-dimensional clustering and its application to microarray data. *Biometrics* 64 (2):440–48. doi:10.1111/j.1541-0420.2007.00922.x.

Weng, H., and Y. Feng. 2016. Community detection with nodal information. arXiv preprint arXiv:1610.09735.

Wilson, J. D., J. Palowitch, S. Bhamidi, and A. B. Nobel. 2017. Community extraction in multilayer networks with heterogeneous community structure. *The Journal of Machine Learning Research* 18:5458–506.

Witten, D. M., and R. Tibshirani. 2010. A framework for feature selection in clustering. *Journal of the American Statistical Association* 105 (490):713–26. doi:10.1198/jasa.2010.tm09415.

Witten, D. M., R. Tibshirani, and T. Hastie. 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics (Oxford, England)* 10 (3):515–34. doi:10.1093/biostatistics/kxp008.

Yan, B., and P. Sarkar. 2016. Covariate Regularized Community Detection in Sparse Graphs. arXiv preprint arXiv:1607.02675.

Yang, J., J. McAuley, and J. Leskovec. 2013. Community detection in networks with node attributes. Data mining IEEE 13th international conference, Dallas, Texas, 1151–56.

Zanghi, H., S. Volant, and C. Ambroise. 2010. Clustering based on random graph model embedding vertex features. *Pattern Recognition Letters* 31 (9):830–36. doi:10.1016/j.patrec.2010.01.026.

Zhang, Y., E. Levina, and J. Zhu. 2016. Community detection in networks with node features. *Electronic Journal of Statistics* 10 (2):3153–78. doi:10.1214/16-EJS1206.

Zhou, D., and C. J. C. Burges. 2007. Spectral clustering and transductive learning with multiple views. Proceedings of the 24th International Conference on Machine Learning, Corvalis, Oregon, USA. 1159–66.

Zhou, Y., H. Cheng, and J. X. Yu. 2009. Graph clustering based on structural/attribute similarities. Proceedings of the VLDB Endowment, Lyon, France, 2, 718–29.