Consistent Recovery of Communities from Sparse Multi-relational Networks: A Scalable Algorithm with Optimal Recovery Conditions

Sharmodeep Bhattacharyya[1] and Shirshendu Chatterjee[2]

[1] Oregon State University, Corvallis, OR, 97330, USA
bhattash@science.oregonstate.edu
[2] City University of New York, New York, NY, 10031, USA
shirshendu@ccny.cuny.edu

Abstract. *Multi-layer* and *multiplex* networks show up frequently in many recent network datasets. We consider the problem of identifying the common *community membership structure* of a finite sequence of networks, called *multi-relational networks*, which can be considered a particular case of multiplex and multi-layer networks. We propose two *scalable* spectral methods for identifying communities within a finite sequence of networks. We provide theoretical results to quantify the performance of the proposed methods when individual networks are generated from either the *stochastic block model* or the *degree-corrected block model*. The methods are guaranteed to recover communities consistently when either the number of networks goes to infinity arbitrarily slowly, or the expected degree of a typical node goes to infinity arbitrarily slowly, even if all the individual networks have fixed size and are sparse. This condition on the parameters of the network models mentioned above is both sufficient for consistent community recovery using our methods and also necessary to have any consistent community detection procedure. We also give some simulation results to demonstrate the efficacy of the proposed methods.

Keywords: Spectral Clustering · Community Detection · Multi-relational Networks · Multi-layer Networks · Stochastic block model · Degree-corrected block model.

1 Introduction

In this paper, we focus on the problem of identifying common community structure present in a finite sequene of (possibly incredibly sparse) networks. The community detection problem can be thought of as a particular case of vertex clustering problem, in which the goal is to divide the set of vertices of a given network (or a finite sequence of networks) into groups based on some common properties of the vertices. The primary objective in the community detection problem is to identify groups of vertices of a given network (or a finite sequence of networks) so that the average number of connections within the groups are *significantly more* than those between the groups. Several random graph models have been proposed in the literature for generating networks with community structure. Examples of random graph models for a single network with community structure include stochastic block model [7], degree-corrected block model [8] and random dot product model [22].

1.1 Community detection methods for a single network

Many methods have been proposed in the statistics and machine learning literature to identify the community structure (see [5] for a review) within a given single network. An important class of methods for detecting communities within a given network, which we refer to as *spectral methods*, involve the spectrum of various matrices (e.g. the adjacency matrix, the laplacian matrix) associated with the network. Spectral methods for community detection was introduced in [4], and analyzed in many subsequent papers (see [1, 10, 11, 13, 17, 18]). In addition to being model agnostic, the spectral methods are highly scalable, as the main numerical procedure involved in these methods is matrix factorization, and many scalable implementations of matrix factorization algorithms have been developed in the numerical analysis literature. The accuracy of some spectral methods in recovering communities within a given single network has been proven theoretically if the network is dense and is generated from some form of exchangeable random graph models [17, 19]. But, to the best of our knowledge, no known community detection algorithm is scalable and has been proved to perform consistently to identify communities within several kinds of sparse network.

1.2 Existing community detection methods from multiple networks

Several approaches have been put forward to develop statistical frameworks for inference on temporal and multi-layer network models. Although most of such methods have not been developed with the goal of community detection, many of them can be used for such a purpose. For example, the methods developed in [21], [12] and [24] can be used to perform model-based community detection, and the authors of [6] and [15] use likelihood-based methods (e.g., profile-likelihood) to identify communities in networks generated from multi-layer network models. Various other authors have proposed model agnostic procedures (see, e.g., [20], [9], [3] and [2]) for detecting communities in multi-layer networks. Spectral algorithms have also been used to find communities from a finite sequence of networks [14, 16]. However, most of these works lack quantitative estimates evaluating the performances of the proposed methods and theoretical results which guarantee the consistent recovery of communities. Also, most of these methods including the existing spectral methods do not work when individual networks as well as an aggregated version of the multi-relational networks, are both sparse.

1.3 Our Contribution

Realizing the above limitations of the existing approaches for performing community detection on a single (resp. multiple) network (resp. networks), and recognizing the advantages of using spectral methods (e.g., scalability and model agnostic nature) for a given single network, we propose and analyze two spectral algorithms for finding the common community structure within a given finite sequence of networks.

The main contributions of our work can be summarized as follows.

(a) We propose and analyze two scalable and model agnostic methods, for identifying communities within a multi-relational network having a common community structure. These methods
  - can be used to identify communities within a single network too.
  - are flexible enough to accommodate both sparse and dense networks.

(b) We prove theoretically that our methods outperform existing methods when the given network is generated from either the stochastic block model or the degree-corrected block model or their extensions in a multi-relational setup.

(c) We also prove analytically that, under the mildest (necessary) parametric condition, the proposed methods identify communities in the networks generated from single or multi-layer stochastic block models and degree-corrected block models consistently. We show that in the multi-relational network setting, our spectral clustering methods can recover the common community structure consistently even if each of the individual networks has fixed size and is highly sparse (e.g., has a constant average degree) and has connectivity below the community detectability threshold.

2 Community Detection Algorithms

A *multi-relational network* can be considered as an *edge-colored multi-graph*, where different colors correspond to edge sets of different network snapshots. The t-th snapshot $G_n^{(t)}$ is represented by the corresponding adjacency matrix $\mathbf{A}_{n\times n}^{(t)}$. *Let* $\mathbf{Z}_{n\times K}$ *denote the actual common community membership matrix of the nodes in each of the graphs* $G_n^{(t)}$*, where,* $\mathbf{Z}_{ik}=1$ *if the* i*-th node belongs to the* k*-th community for all* $G_n^{(t)}$*. The goal is to estimate* $\mathbf{Z}$. The algorithms are given in Algorithm 1 and 2.

Let $[n] := \{1,2,\ldots,n\}$ for $n \in \mathbb{N}$, $\mathscr{M}_{m,n}$ be the set of all $m\times n$ matrices which have exactly one 1 and $n-1$ 0's in each row. $||\cdot||_2, ||\cdot||, ||\cdot||_F$ denote Euclidean ℓ_2-norm, operator norm and Frobenius norm respectively. $\lambda_i(\cdot)$ denotes the i-th largest eigenvalue. For the truncation parameter δ in Algorithm 1, any small positive value is a good choice. In our implementation, we used $\delta = 0.01$.

Algorithm 1: Spectral Clustering of the Sum of the Adjacency Matrices
Input: Adjacency matrices $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(T)}$; number of communities K; approximation parameter ϵ, truncation parameter δ.
Output: Membership matrix $\hat{\mathbf{Z}}_0$.

1. Obtain the **sum of the adjacency matrices**, $\mathbf{A}_0 = \sum_{t=1}^{T} \mathbf{A}^{(t)}$.
2. Get $\bar{d} := \frac{1}{nT}\mathbf{1}_n^T \mathbf{A}_0 \mathbf{1}_n$. Let n' be the number of rows (having indices $1 \leqslant k_1 < k_2 < \cdots < k_{n'} \leqslant n$) of $\mathbf{A}_0$ having row sum at most $e(T\bar{d})^{1+\delta}$.
3. Let $\mathbf{A} \in \mathbb{R}^{n' \times n'}$ be the submatrix of $\mathbf{A}_0$: $A_{i,j} = (A_0)_{k_i,k_j}, i, j \in [n']$.
4. Obtain $\hat{U} \in \mathbb{R}^{n' \times K}$ consisting of the leading K eigenvectors of $\mathbf{A}$ corresponding to its largest absolute eigenvalues.
5. Use $(1+\epsilon)$ approximate K-means clustering algorithm on the row vectors of $\hat{\mathbf{U}}$ to obtain $\hat{\mathbf{Z}} \in \mathscr{M}_{n',K}$ and $\hat{\mathbf{X}} \in \mathbb{R}^{K \times K}$ satisfying

$$||\hat{\mathbf{Z}}\hat{\mathbf{X}} - \hat{\mathbf{U}}||_F^2 \leqslant (1+\epsilon) \min_{\boldsymbol{\Gamma} \in \mathscr{M}_{n' \times K}, \mathbf{X} \in \mathbb{R}^{K \times K}} ||\boldsymbol{\Gamma}\mathbf{X} - \hat{\mathbf{U}}||_F^2. \tag{2.1}$$

6. Extend $\hat{\mathbf{Z}}$ to obtain $\hat{\mathbf{Z}}_0 \in \mathscr{M}_{n,K}$ as follows. $(\hat{\mathbf{Z}}_0)_{j,*} = \hat{\mathbf{Z}}_{i,*}$ (resp. $(1, 0, \ldots, 0)$) for $j = k_i$ (resp. $j \notin \{k_1, \ldots, k_{n'}\}$).
7. $\hat{\mathbf{Z}}_0$ is the estimate of $\mathbf{Z}$.

Algorithm 2: Spherical Spectral Clustering of the Sum of the Adjacency Matrices
Input: Adjacency matrices $A^{(1)}, A^{(2)}, \ldots, A^{(T)}$; number of communities K; approximation parameter ϵ, truncation parameter δ.
Output: Membership matrix $\check{\mathbf{Z}}$.

1. Perform till Step 4 of Algorithm 1.
2. Let n_+ be the number of nonzero rows of $\hat{\mathbf{U}}$. Obtain $\hat{\mathbf{U}}^+ \in \mathbb{R}^{n_+ \times K}$ consisting of normalized nonzero rows of $\hat{\mathbf{U}}$, i.e. $\hat{\mathbf{U}}^+_{i,*} = \hat{\mathbf{U}}_{i,*} / \left\|\hat{\mathbf{U}}_{i,*}\right\|_2$ for i such that $\left\|\hat{\mathbf{U}}_{i,*}\right\|_2 > 0$.
3. Use $(1+\epsilon)$ approximate K-median clustering algorithm on the row vectors of $\hat{\mathbf{U}}^+$ to obtain $\check{\mathbf{Z}}^+ \in \mathscr{M}_{n_+,K}$ and $\check{X} \in \mathbb{R}^{K \times K}$ satisfying

$$\left\|\check{\mathbf{Z}}^+\check{\mathbf{X}} - \hat{\mathbf{U}}^+\right\|_F \leqslant (1+\epsilon) \min_{\boldsymbol{\Gamma} \in \mathscr{M}_{n'' \times K}, \mathbf{X} \in \mathbb{R}^{K \times K}} \left\|\boldsymbol{\Gamma}\mathbf{X} - \hat{\mathbf{U}}^+\right\|_F. \tag{2.2}$$

4. Extend $\check{\mathbf{Z}}^+$ to obtain $\check{\mathbf{Z}}$ by (arbitrarily) adding $n' - n_+$ many canonical unit row vectors at the end, like in Step 6 of Algorithm 1.
5. Extend $\check{\mathbf{Z}}$ to obtain $\check{\mathbf{Z}}_0 \in \mathscr{M}_{n,K}$ as follows. $(\hat{\mathbf{Z}}_0)_{j,*} = \hat{\mathbf{Z}}_{i,*}$ (resp. $(1, 0, \ldots, 0)$) for $j = k_i$ (resp. $j \notin \{k_1, \ldots, k_{n'}\}$).
6. $\check{\mathbf{Z}}_0$ is the estimate of $\boldsymbol{Z}$.

3 Theoretical results about the performance of the algorithms

We consider two different models for a *multi-relational network* generation. The first one is *Multi-layer stochastic block model* with (i) the latent membership vector $\boldsymbol{z} = (z_1, \ldots, z_n)$, where each $z_i \in [K]$, (ii) the set of T, $K \times K$ connectivity probability matrices $\{\mathbf{B}^{(t)}\}_{t=1}^{T}$ and (iii) the $K \times 1$ probability vector of allocation in each community, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$.

$$\mathbf{z}_1, \ldots, \mathbf{z}_n \overset{iid}{\sim} \text{Mult}(1; (\pi_1, \ldots, \pi_K)), \mathbb{P}\left(A_{ij}^{(t)} = 1 \middle| \mathbf{z}_i, \mathbf{z}_j\right) = B_{\mathbf{z}_i \mathbf{z}_j}^{(t)}, \text{ so} \tag{3.1}$$

$$A_{ij}^{(t)} \sim Bernoulli(P_{ij}^{(t)}), \text{ where } \mathbf{P}^{(t)} := \mathbf{Z}\mathbf{B}^{(t)}\mathbf{Z}^T. \tag{3.2}$$

Let d be the maximum expected degree of a node, λ be the average of the smallest eigenvalues of normalized probability matrices $\{\mathbf{B}^{(t)}\}_{t=1}^{T}$ and

$$\lambda = \frac{n}{Td} \sum_{t \in [T]} \lambda_K(\mathbf{B}^{(t)}) > 0 \tag{3.3}$$

Theorem 1. *For any $\epsilon, \eta, \delta > 0$ and $c \in (0,1)$, there are constants $C_1 = C_1(\epsilon, c, \delta), C_2 = C_2(c, \delta) > 0$ such that if $Td \geqslant C_2(K/\lambda)^{1+\delta}, n \geqslant 3K$ and if $n_{min} > 2/\lambda$, then the proportion of misclassified nodes in Algorithm 1 is*

$$\leqslant \left(\frac{n_{min}}{n}\right)^{-1} e^{-(1-c)Td} + C_1 \left(\frac{n_{min}}{n} - e^{-(1-c)Td}\right)^{-2} K\lambda^{-2}(Td)^{-1+2\eta+2\delta}$$

with probability $\geqslant 1 - 5\exp(-\min\{cTd\lambda, \frac{1}{5}(Td)^{2\eta}\log n\})$. n_{min} = smallest community size. Therefore, in the special case, when (i) K is a constant and (ii) the community sizes are balanced, i.e. $n_{max}/n_{min} = O(1)$, then the proportion of misclassified nodes in $\hat{\mathbf{Z}}_0$ goes to zero with probability $1 - o(1)$ if $Td\lambda \to \infty$.

The other model is *multi-layer degree-corrected block model* with (i) the latent membership vector $\boldsymbol{z} = (z_1, \ldots, z_n)$, where each $z_i \in [K]$, (ii) the set of T, $K \times K$ connectivity probability matrices $\{\mathbf{B}^{(t)}\}_{t=1}^{T}$, (iii) a set of *degree parameters*

$$\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n) \text{ satisfying } \max_{i \in \mathcal{C}_k} \psi_i = 1 \text{ for all } k \in \{1, 2, \ldots, K\} \tag{3.4}$$

and (iv) the $K \times 1$ probability vector of community allocation $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$.

$$\mathbf{z}_1, \ldots, \mathbf{z}_n \overset{iid}{\sim} \text{Mult}(1; (\pi_1, \ldots, \pi_K)), \mathbb{P}\left(A_{ij}^{(t)} = 1 \middle| \mathbf{z}_i, \mathbf{z}_j\right) = \psi_i \psi_j B_{\mathbf{z}_i \mathbf{z}_j}^{(t)}, \text{ so} \tag{3.5}$$

$$A_{ij}^{(t)} \sim Bernoulli(\tilde{P}_{ij}^{(t)}), \text{ where } \tilde{\mathbf{P}}^{(t)} := Diag(\boldsymbol{\psi})\mathbf{Z}\mathbf{B}^{(t)}\mathbf{Z}^T Diag(\boldsymbol{\psi}). \tag{3.6}$$

For $k \in [K]$, let $\tilde{n}'_k := \sum_{i \in \mathcal{C}_k \cap \{k_1, \ldots, k_{n'}\}} \psi_i^2$ and $\tau_k := \sum_{i \in \mathcal{C}_k} \psi_i^2 \sum_{i \in \mathcal{C}_k} \psi_i^{-2}$ be a measure of heterogeneity of $\boldsymbol{\psi}$.

Theorem 2. *For any $\epsilon, \eta, \delta > 0$ and $c \in (0,1)$, there are constants $C_1(\epsilon, c, \delta), C_2(c, \delta) > 0$ such that if $Td \geqslant C_2(K/\lambda)^{1+1/\delta}$ and $n \geqslant 3K$ is large enough, then the total number of misclassified nodes in Algorithm 2 is*

$$\leqslant \frac{n}{e^{(1-c)Td}} + C_1 \left[\frac{K\tilde{n}'_{max}}{(\psi_{min}\lambda\tilde{n}'_{min})^2} + \frac{\sqrt{K\sum_{k\in[K]}\tau_k n(Td)^{-(1/2)+\delta+\eta}}}{\lambda\tilde{n}'_{min}} \right] \tag{3.7}$$

with probability at least $1 - 5\exp(-\min\{cTd\lambda, \frac{1}{5}(Td)^{2\eta}\log n\})$.

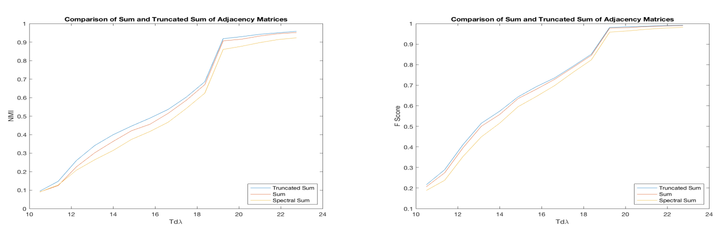
Therefore, in the special case, when (i) K is a constant, (ii) the community sizes are balanced, i.e. $n_{max}/n_{min} = O(1)$ and (iii) $\psi_i = \alpha_i / \max\{\alpha_j : z_i = z_j\}$, where $(\alpha_i)_{i=1}^n$ are i.i.d. positive weights, then consistency holds for $\check{\mathbf{Z}}_0$ with probability $1 - o(1)$ if $\mathbb{E}[\max\{\alpha_1^2, \alpha_1^{-2}\}] < \infty$ and $Td\lambda \to \infty$.

Remark 3 *The condition "$Td\lambda \to \infty$" is necessary and sufficient in order to have a consistent estimator of $\mathbf{Z}$. Theorem 1 and 2 proves the sufficiency. The necessity of the condition follows from the work of [23].*

Remark 4 *Note that the assertion of Theorem 1 and 2 are non-asymptotic results, so, the asymptotic result on consistent label recovery can hold for different conditions like - (i) constant T and $n \to \infty$; (ii) constant n and $T \to \infty$; (iii) $K \to \infty$ and suitable conditions on n, d and T and so on.*

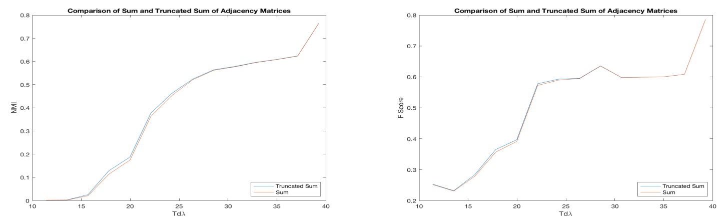
4 Simulation Results

We simulate multiple stochastic block model with $n = 40,000$, $K = 4$ and $T = 10$, but varying $\{\mathbf{B}^{(t)}\}_{t=1}^T$ such that $Td\lambda$ increases. We simulate multiple degree-corrected block model with $n = 20,000$, $K = 4$ and $T = 10$, but varying $\{\mathbf{B}^{(t)}\}_{t=1}^T$ such that $Td\lambda$ increases. The degree parameters in multiple degree-corrected block model are generated from $U(0.5, 1)$.



(a) (b)

Fig. 4.1. Comparison of (i), (ii) and Algorithm 1 (Truncated Sum) (a) using normalized mutual information and (b) using F-score.



(a) (b)

Fig. 4.2. (a) Comparison of (iii) and Algorithm 2 (Truncated Sum) (a) using normalized mutual information and (b) using F-score.

We implement five different algorithms - (i) *Sum*: spectral clustering with sum of adjacency matrices without truncation; (ii) *Spectral sum*: clustering the rows of sum of eigen-spaces $\sum_{t=1}^{T} \hat{\mathbf{U}}^{(t)}$ of each network snapshot (where, $\hat{\mathbf{U}}^{(t)}_{n\times K}$ is the matrix formed by the eigenvectors of top K eigenvectors of $A^{(t)}$); (iii) *Sum (Spherical)*: spherical spectral clustering with sum of adjacency matrices without truncation; (iv) Algorithm 1; (v) Algorithm 2. For models generated under multiple stochastic block model, we compare the algorithms (i), (ii) and (iv). For models generated under multiple degree-corrected block model, we compare the algorithms (iii) and (v). For metric of success, we use normalized mutual information and F-score. We can see from Figures 4.1 and 4.2 that for $Td\lambda$ between $(10, 20)$ for multiple stochastic block model and $(10, 40)$ for multiple degree-corrected block model, Algorithm 1 and 2 out-performs all other algorithms. The simulation results are in concert with the theoretical results in Theorem 1 and Theorem 2.

References

[1] Bhattacharyya, S., Bickel, P.J.: Community detection in networks using graph distance. arXiv preprint arXiv:1401.3915 (2014)

[2] Chen, P.Y., Hero, A.O.: Multilayer spectral graph clustering via convex layer aggregation: Theory and algorithms. IEEE Transactions on Signal and Information Processing over Networks 3(3), 553–567 (2017)

[3] Dong, X., Frossard, P., Vandergheynst, P., Nefedov, N.: Clustering with multi-layer graphs: A spectral perspective. IEEE Transactions on Signal Processing 60(11), 5820–5831 (2012)

[4] Fiedler, M.: Algebraic connectivity of graphs. Czechoslovak Math. J. 23(98), 298–305 (1973)

[5] Fortunato, S.: Community detection in graphs. Physics reports 486(3-5), 75–174 (2010)

[6] Han, Q., Xu, K., Airoldi, E.: Consistent estimation of dynamic and multi-layer block models. In: International Conference on Machine Learning. pp. 1511–1520 (2015)

[7] Holland, P.W., Laskey, K.B., Leinhardt, S.: Stochastic blockmodels: First steps. Social networks 5(2), 109–137 (1983)

[8] Karrer, B., Newman, M.E.: Stochastic blockmodels and community structure in networks. Physical Review E 83(1), 016107 (2011)

[9] Kumar, A., Rai, P., Daumé III, H.: Co-regularized spectral clustering with multiple kernels (2010)

[10] Lei, J., Rinaldo, A., et al.: Consistency of spectral clustering in stochastic block models. The Annals of Statistics 43(1), 215–237 (2015)

[11] von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. Ann. Statist. 36(2), 555–586 (2008), http://dx.doi.org/10.1214/009053607000000640

[12] Matias, C., Miele, V.: Statistical clustering of temporal networks through a dynamic stochastic block model. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79(4), 1119–1141 (2017)

[13] Ng, A.Y., Jordan, M.I., Weiss, Y., et al.: On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems 2, 849–856 (2002)

[14] Paul, S., Chen, Y.: Spectral and matrix factorization methods for consistent community detection in multi-layer networks. arXiv preprint arXiv:1704.07353 (2017)

[15] Paul, S., Chen, Y., et al.: Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. Electronic Journal of Statistics 10(2), 3807–3870 (2016)

[16] Pensky, M., Zhang, T., et al.: Spectral clustering in the dynamic stochastic block model. Electronic Journal of Statistics 13(1), 678–709 (2019)

[17] Rohe, K., Chatterjee, S., Yu, B.: Spectral clustering and the high-dimensional stochastic blockmodel. Ann. Statist. 39(4), 1878–1915 (2011), http://dx.doi.org/10.1214/11-AOS887

[18] Shi, J., Malik, J.: Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(8), 888–905 (2000)

[19] Sussman, D.L., Tang, M., Fishkind, D.E., Priebe, C.E.: A consistent adjacency spectral embedding for stochastic blockmodel graphs. Journal of the American Statistical Association 107(499), 1119–1128 (2012)

[20] Tang, W., Lu, Z., Dhillon, I.S.: Clustering with multiple graphs. In: Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on. pp. 1016–1021. IEEE (2009)

[21] Xu, K.S., Hero, A.O.: Dynamic stochastic blockmodels for time-evolving social networks. IEEE Journal of Selected Topics in Signal Processing 8(4), 552–562 (2014)

[22] Young, S.J., Scheinerman, E.R.: Random dot product graph models for social networks. In: International Workshop on Algorithms and Models for the Web-Graph. pp. 138–149. Springer (2007)

[23] Zhang, A.Y., Zhou, H.H., et al.: Minimax rates of community detection in stochastic block models. The Annals of Statistics 44(5), 2252–2280 (2016)

[24] Zhang, X., Moore, C., Newman, M.E.: Random graph models for dynamic networks. The European Physical Journal B 90(10), 200 (2017)