C.F. van Oosten

c.f.van.oosten@umail.leidenuniv.nl

Community Detection in a Time Dynamic Stochastic Block Model

Master thesis

Thesis supervisor: Dr. J. Schmidt-Hieber

Date of publication: March 10, 2017

Mathematical Institute of Leiden University

1 Summary

In network data analysis, community detection is a fundamental statistical problem. A popular approach to analyse such data networks is the stochastic block model. In this model connection probabilities between nodes only depend on the communities they belong to. This thesis is concerned with a two-stage method that achieves optimal misclassification proportion in the stochastic block model. Even though the stochastic block model is simple and models the group structure, it lacks flexibility and is not realistic. The reason is that models generated with the stochastic block model miss the generation of hubs which are nodes with a high degree. Hubs are however often observed in network data. Indeed, lots of real life models have a cumulative advantage where "the rich get richer". The preferential attachment model models this advantage by adding a degree dependent term in the connection probabilities, but does not include any group structure. This thesis contributes to a previous result for misclassification proportion in stochastic block models by adding in a degree dependence to the connection probability which changes over time. This allows models different from the stochastic block model to be used. In the thesis we propose a model which combines the stochastic block model with the preferential attachment model to allow the modeling of hubs in networks. For this model we prove that under some conditions the average edge degrees in this model are comparable to those in the stochastic block model. We achieve this by bounding the expected degree of the time dynamic model in terms of the stochastic block model which requires techniques like martingale concentration inequalities and probabilistic approximations.

First we define the stochastic block model in Section 2. In Section 3 we summarize previous results which uses the stochastic block model as a basis. After that we introduce the preferential attachment model in Section 4 which we use to make our own time dynamic stochastic block model in Section 5.

Contents

2 Networks and the Stochastic Block Model

In network data analysis [17, 7] the goal is to infer structures in the generating process of observed networks. One of those structures are communities: partitions of graph nodes which can be grouped together in a suitable sense. Several authors have proposed methods [6, 8, 15] and researched theoretical understanding on community recovery ([2, 18, 14, 13, 1] among others). To model network data with community structure we focus on the stochastic block model. This models simple, unweighted networks where the vertices are partitioned in distinct communities. The probability of an edge between two vertices does not depend on the nodes themselves but on the communities the nodes belong to. The stochastic block model generates a network based on the given labels and connection probabilities which are assumed to be Bernoulli distributed.

We start with the basic definitions of a network. A network is a graph with a set of nodes $V = \{1, 2, ..., n\}$ and a set of edges $E \subseteq V \times V$ which are connections between nodes. In social networks for example nodes often represent people while edges represent friendships between two persons. In general we describe networks by using a connectivity matrix $A = (A_{ij})_{i,j \in V}$ where $A_{ij} = 1$ if edge (i, j) is in E and $A_{ij} = 0$ otherwise. This has all the information about the nodes and edges needed to construct the network.

We recall the definition of the stochastic block model proposed by [10]: Let k be the number of communities in the model. The stochastic block model is characterized by a symmetric connectivity matrix $B \in [0, 1]^{k \times k}$ and a label vector $\sigma \in [k]^n$ where $[k] = \{1, 2, ..., k\}$. The connectivity matrix B denotes the connection probabilities between groups and the label vector σ labels every node to a specific group. An adjacency matrix $A \in \{0, 1\}^{n \times n}$ is generated using the stochastic block model by defining

- $A_{uv} = A_{vu} \sim \text{Bern}(B_{\sigma(u)\sigma(v)})$ for all $u < v \in [n]$,
- $A_{uu} = 0$ for every $u \in [n]$,

which gives an adjacency matrix for a simple graph. To gain some heuristic insight into this model consider two communities. If the connection probabilities are very different it is easy to separate the two communities, while if they are close to each other the groups are very similar and hard to distinguish.

Given an adjacency matrix A we can generate a network. An example of such a network based on the stochastic block model is shown in Figure 1.

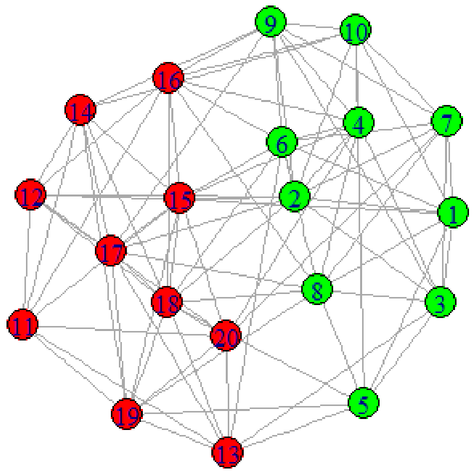


Figure 1: Randomly generated graph with $n = 20$ nodes, $k = 2$ communities and $B_{ii} = 0.66, B_{ij} = 0.33$ for all $i \neq j$. Nodes 1 to 10 belong to the green group while nodes 11 to 20 belong to the red one.

3 A Two-step Community Detection Algorithm

Before coming to the time dynamic stochastic block model we first review the method for community detection in stochastic block models proposed by Gao et al. [5] and Zhang and Zhou [19]. We first look at the algorithm and then study theoretical properties of the method.

3.1 First Step: Spectral Clustering Methods

In the stochastic block model, connection probabilities only depend on the group to which a node belongs to. Therefore the rank of the connection probability matrix $P = (P_{uv})$ is at most the amount of communities k. By clustering eigenvectors of P we can therefore reduce the data from $n \times n$ to $n \times k$. This is refered to as spectral clustering [16]. Since A_{uv} is a Bernoulli variable with $E[A_{uv}] = P_{uv}$ we apply the clustering to the adjacency matrix A which can be computed from data.

We discuss two different spectral clustering methods: unnormalized spectral clustering and normalized spectral clustering. Unnormalized spectral clustering (USC) takes the eigenvectors of the adjacency matrix A without any normalization. Normalized spectral clustering (NSC) uses the eigenvectors of the Laplacian $L(A) = ([L(A)]_{uv})$ with $[L(A)]_{uv} = d_u^{-1/2} d_v^{-1/2} A_{uv}$. When these methods are applied to sparse graphs they do not perform well since A and $L(A)$ are poor estimators for P and $L(P)$, cf. [12]. In order to obtain better estimators we regularize both spectral clustering methods. To regularize USC we introduce a trimming operator T_τ which replaces the u-th row and u-th column of A by zero whenever $d_u > \tau$. Nodes of high degree are therefore discarded. We define USC(τ) as the regularized USC with threshold$\tau \in [0, \infty)$. Note that USC(∞) is the same as unregularized USC. For NSC we define the adjusted adjacency matrix $A_\tau = A + \frac{\tau}{n}\mathbf{1}\mathbf{1}^T$ where $\mathbf{1} = (1, 1, ..., 1)^T \in \mathbb{R}^n$. Define NSC($\tau$) as clustering of the eigenvectors of the Laplacian $L(A_\tau)$. Note that NSC(0) is the same as unregularized NSC. The following initial community assignment algorithm applies spectral clustering to community detection:

Algorithm 1 (initial group labelling) ([5], Algorithm 2)

Input : Data matrix $\hat{U} \in \mathbb{R}^{n \times k}$, leading eigenvectors of either $T_\tau(A)$ or $L(A_\tau)$, number of communities k, critical radius $r = \mu\sqrt{\frac{k}{n}}$ with some constant $\mu > 0$

Output: Community assignment $\hat{\sigma}$

Set $S = [n]$

for i = *1 to* k **do**

  Let $t_i = \text{argmax}_{u \in S} |\{v \in S : \|\hat{U}_{v*} - \hat{U}_{u*}\| < r\}|$;

  Set $\hat{\mathcal{C}}_i = \{v \in S : \|\hat{U}_{v*} - \hat{U}_{t_i*}\| < r\}$;

  Label $\hat{\sigma}(u) = i$ for all $u \in \hat{\mathcal{C}}_i$;

  Update $S \leftarrow S \backslash \hat{\mathcal{C}}_i$.

end

If $S \neq \emptyset$, then for any $u \in S$, set $\hat{\sigma}(u) = \text{argmin}_{i \in [k]} \frac{1}{\hat{\mathcal{C}}_i} \sum\limits_{v \in \hat{\mathcal{C}}_i} \|\hat{U}_{u*} - \hat{U}_{v*}\|$.

3.2 Second Step: Refined Group Labelling

Based on the initial group labelling we present in this section a refinement method that was also proposed in [5]. Denote by σ^0 the initial group assignment that we obtained from Algorithm 1. For every node u, we apply Algorithm 1 on the adjacency matrix leaving node u and its edges out. This gives initial label assignments $\sigma_u^0(v)$ for all $v \neq u$. We define $\hat{\sigma}_u(v) = \sigma_u^0(v)$ for all $v \neq u$ and choose $\hat{\sigma_u(u)}$ as the community that has the most neighbouring nodes in common with u. Using this we obtain label vectors $\hat{\sigma}_u$ for every node u. To make a unified assignment vector these label vectors have to be combined. However, due to permutations in the assignment vectors the community labels can be different. We illustrate this with the following example:

	1	2	3	4
1	**1**	**1**	2	2
2	**1**	**1**	**1**	2
3	1	**2**	**2**	**2**
4	2	2	**1**	**1**

The u-th row corresponds to the label vector $\hat{\sigma}_u$. The bold numbers are the nodes which have the same label as u in $\hat{\sigma}_u$. To obtain a unified community assignment we compare each row with $\hat{\sigma}_1$. To decide which permutation is most likely we look at which community in $\hat{\sigma}_1$ has the most overlap with the community of u in $\hat{\sigma}_u$. For $u = 1$ this gives $\hat{\sigma}(1) = \hat{\sigma}_1(1) = 1$. For $u = 2$ the community of u has an overlap of two nodes with community 1 while only one node overlap with community 2. Therefore $\hat{\sigma}(2) = 1$. For $u = 3$ the community of u has an overlap of one node with community 1 and an overlap of two nodes with community 2. Therefore $\hat{\sigma}(3) = 2$. For $u = 4$ the community of u has no overlap with community 1 and an overlap of two nodes with community 2. Therefore $\hat{\sigma}(4) = 2$. Thus the final community assignment becomes $\hat{\sigma} = (1, 1, 2, 2)$.

Let us now describe the algorithm in more detail. Define $A_{-u} \in \{0, 1\}^{n-1}$ as the adjacency matrix A leaving row u and column u out. Applying the initial group label method on A_{-u} for every u gives initial assignment vectors $\sigma_u^0 = \left(\sigma_u^0(v)\right)_{v \in [n] \setminus \{u\}}$. We define $\hat{\sigma}_u(v) = \sigma_u^0(v)$ for all $v \neq u$ and set $\hat{\sigma}_u(u) = 0$ to make vectors of length n. For equal community sizes the maximum likelihood estimator for $\hat{\sigma}_u(u)$ given labels $(\hat{\sigma}_u(v))_{v \neq u}$ [3] is

$$\operatorname*{argmax}_{l} \left\{ \sum_{v} A_{uv} \mathbb{1}\left(\hat{\sigma}_u(v) = l\right) \right\}.$$

When community sizes are different nodes belonging to a larger community are more common. Therefore the maximum likelihood estimator has to compensate for larger communities by adding a penalty term $\rho_u \sum_v \mathbb{1}(\hat{\sigma}_u(v) = l)$. Here ρ_u depends on the average connection probabilities in and between communities as labeled by $\hat{\sigma}_u$. For the assignment of $\hat{\sigma}_u(u)$ we then get

$$\hat{\sigma}_u(u) = \underset{l}{\operatorname{argmax}} \left\{ \sum_v A_{uv} \mathbb{1}\left(\hat{\sigma}_u(v) = l\right) - \rho_u \sum_v \mathbb{1}\left(\hat{\sigma}_u(v) = l\right) \right\}.$$

In the next step of the algorithm we account for different permutations in label vectors to make one unified community assignment $\hat{\sigma}$. Define $\hat{\sigma}(1) = \hat{\sigma}_1(1)$ and for $u = 2, 3, ..., n$ we set $\hat{\sigma}(u)$ as the community in $\hat{\sigma}_1$ which has the most overlap with community $\hat{\sigma}_u(u)$ in $\hat{\sigma}_u$. This means for $u = 2, 3, ..., n$ we take

$$\hat{\sigma}(u) = \underset{l \in [k]}{\operatorname{argmax}} \left|\{v : \hat{\sigma}_1(v) = l\} \cap \{v : \hat{\sigma}_u(v) = \hat{\sigma}_u(u)\}\right|.$$

This gives the community assignment $\hat{\sigma}$.

Algorithm 2 ([5], Algorithm 1)

Input : $A \in \{0,1\}^{n\times n}$, number of communities k, initial community detection method σ^0

Output: Community assignment $\hat{\sigma}$

Penalized neighbor voting: for *u = 1 to n* **do**

Apply σ^0 on A_{-u} to obtain $\sigma_u^0(v)$ for all $v \neq u$ and let $\sigma_u^0(u) = 0$

Define $\tilde{\mathcal{C}}_i^u = \{v : \sigma_u^0(v) = i\}$ for all $i \in [k]$; let $\tilde{\mathcal{E}}_i^u$ be the set of edges within $\tilde{\mathcal{C}}_i^u$, and let $\tilde{\mathcal{E}}_{ij}^u$ be the set of edges between $\tilde{\mathcal{C}}_i^u$ and $\tilde{\mathcal{C}}_i^u$ when $i \neq j$;

Define

$$\hat{B}_{ii}^u = \frac{|\mathcal{E}_i^u|}{\frac{1}{2}|\tilde{\mathcal{C}}_i^u|(|\tilde{\mathcal{C}}_i^u| - 1)},\ \hat{B}_{ij}^u = \frac{|\mathcal{E}_{ij}^u|}{|\tilde{\mathcal{C}}_i^u||\tilde{\mathcal{C}}_j^u|},\ \forall i \neq j \in [k]$$

and let

$$\hat{a}_u = n \min_{i\in[k]} \hat{B}_{ii}^u \text{ and } \hat{b}_u = n \max_{i\neq j\in[k]} \hat{B}_{ij}^u$$

Define $\hat{\sigma}_u : [n] \to [k]$ by $\hat{\sigma}_u(v) = \sigma_u^0(v)$ for all $v \neq u$ and

$$\hat{\sigma}_u(u) = \operatorname*{argmax}_{l\in[k]} \sum_{\sigma_u^0(v)=l} A_{uv} - \rho_u \sum_{v\in[n]} \mathbb{1}_{\{\sigma_u^0(v)=l\}}$$

where for

$$t_u = \frac{1}{2}\log\left(\frac{\hat{a}_u(1-\frac{\hat{b}_u}{n})}{\hat{b}_u(1-\frac{\hat{a}_u}{n})}\right)$$

we define

$$\rho_u = -\frac{1}{2t_u}\log\left(\frac{\frac{\hat{a}_u}{n}e^{-t_u} + 1 - \frac{\hat{a}_u}{n}}{\frac{\hat{b}_u}{n}e^{t_u} + 1 - \frac{\hat{b}_u}{n}}\right)$$

end

Step 2: Define $\hat{\sigma}(1) = \hat{\sigma}_1(1)$. For $u = 2, ..., n$, define

$$\hat{\sigma}(u) = \operatorname*{argmax}_{l\in[k]} |\{v : \hat{\sigma}_1(v) = l\} \cap \{v : \hat{\sigma}_u(v) = \hat{\sigma}_u(u)\}|.$$

3.3 Theoretical Properties

To analyse the stochastic block model, we first recall the parameter spaces used in [5]: The space $\Theta_0(n, k, a, b, \beta)$ consists of all (B, σ) such that

- $\sigma : [n] \to [k]$ is a label function
- community sizes are comparable:

$$|\{u \in [n] : \sigma(u) = i\}| \in [\frac{n}{\beta k} - 1, \frac{\beta n}{k} + 1],\ \forall i \in [k]$$

- $B = (B_{ij}) \in [0, 1]^{k \times k}$ are connection probabilities such that

$$B_{ii} = \frac{a}{n} \text{ for all } i \text{ and } B_{ij} = \frac{b}{n} \text{ for all } i \neq j$$

Note that $n \in \mathbb{N}$ is the number of nodes, $k \in \mathbb{N}$ the amount of communities, $a, b \in \mathbb{R}^+$ constants which depend on n and $\beta \geq 1$ constant. This parameter space consists of label functions σ with comparable community sizes and constant connection probability $B_{ii} = \frac{a}{n}$ within communities and $B_{ij} = \frac{b}{n}$ between communities. Assuming constant connection probabilities for all communities is very restrictive, since communities in real world networks generally vary in size. Therefore [5] also introduces the parameter space $\Theta(n, k, a, b, \lambda, \beta, \alpha)$ which consists of all (B, σ) such that

- $\sigma : [n] \to [k]$ is a label function
- community sizes are comparable:

$$|\{u \in [n] : \sigma(u) = i\}| \in [\frac{n}{\beta k} - 1, \frac{\beta n}{k} + 1],\ \forall i \in [k]$$

- $B = B^T = (B_{ij}) \in [0, 1]^{k \times k}$ such that

$$\frac{a}{n} \leq \min_i B_{ii} \leq \max_i B_{ii} \leq \frac{\alpha a}{n}$$
$$\frac{b}{\alpha n} \leq \frac{1}{k(k-1)} \sum_{i \neq j} B_{ij} \leq \max_{i \neq j} B_{ij} \leq \frac{b}{n}$$

- $\lambda_k(P) \geq \lambda$ with $P = (P_{uv}) = (B_{\sigma(u)\sigma(v)})$ and $\lambda_k(P)$ the k^{th} singular value of P.

Here B_{ii} can vary between $\frac{a}{n}$ and $\frac{\alpha a}{n}$ for all i. The connection probabilities between communities B_{ij} is bounded from above by $\frac{b}{n}$ for all $i \neq j$ and the average over all B_{ij} is bounded from below by $\frac{b}{\alpha n}$. To ensure that $B_{ij} \in (0,1)$ and that $B_{ij} < B_{ii}$ for all $i \neq j$ we assume that $0 < \frac{b}{n} < \frac{a}{n} < 1$ and $\alpha \geq 1$ constant. Note that when λ is small enough we have $\Theta_0(n,k,a,b,\beta) \subset \Theta(n,k,a,b,\lambda,\beta,\alpha)$.

In order to classify the error of a label vector $\hat{\sigma}$ compared to the true label vector σ we recall the Hamming loss

$$l(\hat{\sigma},\sigma) = \min_{\pi \in S_k} \frac{1}{n} \sum_{u=1}^{n} \mathbb{1}(\pi(\hat{\sigma}(u)) \neq \sigma(u)).$$

The minimum over all permutations in the symmetric group S_k is taken because community labels are interchangeable. This loss function gives the proportion of label assignments that are different between $\hat{\sigma}$ and σ.

To prove consistency of the refined group labelling algorithm, Gao et al. [5] provide three main results. The first theorem shows that the initialization step via spectral clustering satisfies a minor consistency property:

Theorem 3.1. ([5], Theorem 3) *Assume $e \leq a \leq C_1 b$ for some constant $C_1 > 0$ and*

$$\frac{ka}{\lambda_k^2} \leq c, \tag{1}$$

for some sufficiently small $c \in (0,1)$ where $\lambda_k = \lambda_k(P)$ the k-th singular value of P. Consider USC(τ) with a sufficiently small constant μ in Algorithm 1 and $\tau = C_2 \overline{d}$ for some sufficiently large constant $C_2 > 0$, where $\overline{d} = \frac{1}{n} \sum_{u \in [n]} d_u$ is the average degree. For any constant $C' > 0$, there exists some $C > 0$ only depending on C', C_1, C_2 and μ such that

$$l(\hat{\sigma},\sigma) \leq C \frac{a}{\lambda_k^2}$$

with probability at least $1 - n^{-C'}$.

A similar consistency property also holds for NSC(τ), cf. [5], Theorem 4. The asymptotic minimax risk under the Hamming loss is derived in the following theorem:

Theorem 3.2. ([19], Theorem 1.1) *When $\frac{(a-b)^2}{ak \log k} \to \infty$, we have*

$$\inf_{\hat{\sigma}} \sup_{(B,\sigma) \in \Theta} \mathbb{E}_{B,\sigma} l(\hat{\sigma},\sigma) = \begin{cases} \exp\left(-(1+\eta)\frac{nI^*}{2}\right), & k = 2; \\ \exp\left(-(1+\eta)\frac{nI^*}{\beta k}\right), & k \geq 3; \end{cases}$$

for both $\Theta = \Theta_0(n,k,a,b,\beta)$ *and* $\Theta = \Theta(n,k,a,b,\lambda,\beta;\alpha)$ *with any* $\lambda < \frac{a-b}{2\beta k}$ *and any* $\beta \in [1, \sqrt{5/3})$, *where* $\eta = \eta_n \to 0$ *is some sequence tending to 0 as* $n \to \infty$ *and*

$$I^* = -2\log\left(\sqrt{\frac{a}{n}}\sqrt{\frac{b}{n}} + \sqrt{1-\frac{a}{n}}\sqrt{1-\frac{b}{n}}\right) \tag{2}$$

is the Rényi divergence of order $\frac{1}{2}$ *between Bern* $\left(\frac{a}{n}\right)$ *and Bern* $\left(\frac{b}{n}\right)$.

The condition $\beta \in [1, \sqrt{5/3})$ restricts the difference in size of communities. The next result in [5] guarantees the accuracy of the parameter estimation in Algorithm 2 which is required to derive upper bounds for the performance of the refinement scheme:

Lemma 3.3. ([5], Lemma 1) *Let* $\Theta = \Theta(n,k,a,b,\lambda,\beta;\alpha)$. *Suppose as* $n \to \infty$, $\frac{(a-b)^2}{ak} \to \infty$ *and there exist constants* $C_0, \delta > 0$ *and a positive sequence* $\gamma = \gamma_n$ *such that*

$$\inf_{(B,\sigma)\in\Theta} \min_{u\in[n]} P_{B,\sigma}\{l(\sigma,\sigma_u^0) \le \gamma\} \ge 1 - C_0 n^{-(1+\delta)} \tag{3}$$

with $\gamma = o(\frac{1}{k\log k})$ *and* $\gamma = o(\frac{a-b}{ak})$ *and* σ_u^0 *as in Algorithm 2. Then there is a sequence* $\eta = \eta_n \to 0$ *as* $n \to \infty$ *and a constant* $C > 0$ *such that*

$$\min_{u\in[n]} \inf_{(B,\sigma)\in\Theta} P\left\{\min_{\pi\in S_k} \max_{i,j\in[k]} |\hat{B}_{ij}^n - B_{\pi(i)\pi(j)}| \le \eta\left(\frac{a-b}{n}\right)\right\} \ge 1 - Cn^{-(1+\delta)} \tag{4}$$

This means that we can consistently estimate all connection probabilities with high probability. Note that the spectral clustering algorithm USC(τ) satisfies (3) under the assumptions of Theorem 3.1. With this it is then possible to show that the refined group labelling algorithm returns an estimator that is asymptotically minimax:

Theorem 3.4. ([5], Theorem 4) *Suppose as* $n \to \infty$, $\frac{(a-b)^2}{ak\log k} \to \infty$, $a \asymp b$ *and condition* 3 *is satisfied for* $\gamma = o\left(\frac{1}{k\log k}\right)$ *and* $\Theta = \Theta_0(n,k,a,b,\beta)$. *Then there is a sequence* $\eta \to 0$ *such that*

$$\begin{cases} \sup_{(B,\sigma)\in\Theta} \mathbb{P}_{B,\sigma}\left\{l(\sigma,\hat{\sigma}) \ge \exp\left(-(1-\eta)\frac{nI^*}{2}\right)\right\} \to 0, & \text{if } k = 2, \\ \sup_{(B,\sigma)\in\Theta} \mathbb{P}_{B,\sigma}\left\{l(\sigma,\hat{\sigma}) \ge \exp\left(-(1-\eta)\frac{nI^*}{\beta k}\right)\right\} \to 0, & \text{if } k \ge 3, \end{cases} \tag{5}$$

where I^* *is the Rényi divergence of order* $\frac{1}{2}$ *as defined in* (2).
If in addition (3) *is satisfied for* $\gamma = o\left(\frac{1}{k\log k}\right)$ *and* $\gamma = o\left(\frac{a-b}{ak}\right)$ *and* $\Theta = \Theta(n,k,a,b,\lambda,\beta;\alpha)$, *then the conclusion* (5) *also holds for* $\Theta = \Theta(n,k,a,b,\lambda,\beta;\alpha)$.

4 Preferential Attachment

The preferential attachment model works differently than the stochastic block model by generating nodes iteratively. This model includes a degree dependent component in the connection probability, which gets larger when the degree of a node is higher. The motivation behind this model is to reproduce degree distributions with powerlaw behaviour, often observed in large real-world networks.

A popular preferential attachment model is the Barabási-Albert model ([11, p. 173]). The (BA) model works as follows: Start with an initial graph $G^{(0)}$ with $N_v^{(0)}$ vertices and $N_e^{(0)}$ edges. Then in every iteration $t = 1, 2, ...$ the current graph $G^{(t-1)}$ is modified to create a new graph $G^{(t)}$ by adding a new vertex v. Each new vertex v is connected to $m < N_v^{(0)}$ existing vertices in $G^{(t-1)}$. The probability that v connects to a given vertex u in $G^{(t-1)} = (V^{(t-1)}, E^{(t-1)})$ is given by

$$\frac{d_u}{\sum_{u' \in V^{(t-1)}} d_{u'}}.$$

This probability describes the preference of the new vertex to connect with existing vertices with higher degrees.

Note that this model has the condition to add exactly m edges in every step. To remove this restriction we generalize this model in the following way: We remove the restriction of having to add nodes with a given degree m and for each potential edge decide on including an edge by a Bernoulli random variable dependent on the degree of the existing vertex. Let v be the v-th iteration of the model. Then for all existing vertices $u < v$ the probability of adding an edge between u and v is a Bernoulli random variable with parameter

$$P(A_{uv} = 1 | d_u(v-1)) = f(d_u(v-1)) \, , u < v$$

where $f : \mathbb{N} \to [0, 1]$ is an increasing function and $d_u(v-1)$ is the degree of note u after $v - 1$ iterations. This allows every edge to be added independent of any other edges added in the same timestep. After adding edges based on these connection probabilities the next node $v + 1$ is added and the process is repeated until the graph has reached a predetermined size n. This model generates hubs, which are nodes with high degrees, by giving a higher probability of new nodes being connected to these hubs. Note that the preferential attachment model does not have any group structure.

5 A Time Dynamic Stochastic Block Model

In this section we propose a new model that combines the preferential attachment model with the stochastic block model. Let $P = (P_{uv}) = (B_{\sigma(u)\sigma(v)})$ and $\{q_u\}_{u\in[n]}$ with $0 < q_u < 1$ for every $u \in [n]$. A graph is iteratively generated by adding a single node v in each step which has connection probabilities

$$P(A_{uv} = 1 | d_u(v-1) = d_u) = P_{uv} f(d_u(v-1)) q_v \, , u < v \tag{6}$$

where $f(d_u) = 1 - (1-q_u)e^{-d_u}$ and $d_u(v-1)$ is the degree of node u after $v-1$ iterations. Note that f is non decreasing in d_u and has range $[q_u, 1]$ for $d_u \geq 0$. The choice for this function f stimulates the creation of hubs by increasing the connection probabilities for high degree nodes.

As an example, assume $k = 2$ communities and set $q_u = \frac{1}{2}$ for all nodes u. Let $B_{11} = B_{22} = \frac{2}{3}$ and $B_{12} = B_{21} = \frac{1}{3}$. The network is then built up as follows:

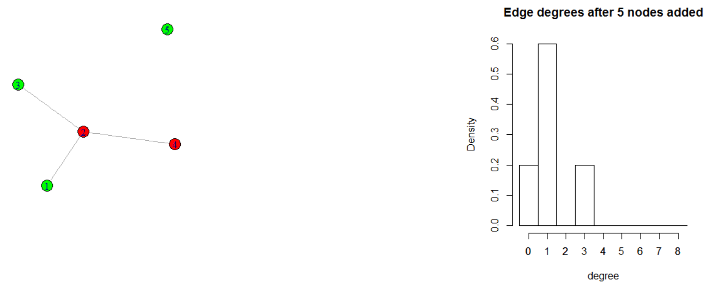
- Node 1 of group 1 is added, no other nodes are available to connect with.
- Node 2 of group 2 is added.
- Edge $(2, 1)$ is added with probability $B_{21} \cdot (1 - \frac{1}{2}e^{-0}) \cdot \frac{1}{2} = \frac{1}{12}$.
- Node 3 of group 1 is added.
- Edge $(3, 1)$ is added with probability

$$\begin{cases} B_{11} \cdot (1 - \frac{1}{2}e^{-1}) \cdot \frac{1}{2} = \frac{1}{3}(1 - \frac{1}{2e}) \approx 0.27 & \text{if edge } (3,1) \text{ is added} \\ B_{11} \cdot (1 - \frac{1}{2}e^{-0}) \cdot \frac{1}{2} = \frac{1}{6} \approx 0.17 & \text{otherwise} \end{cases} .$$

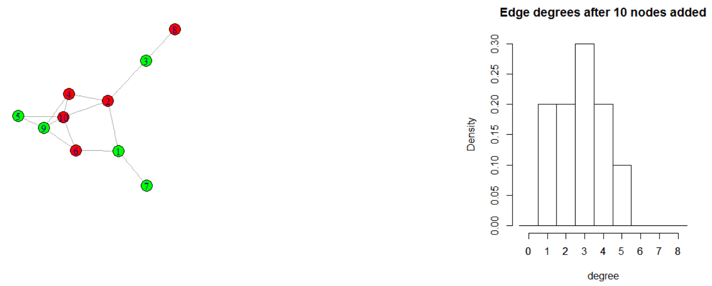
- Edge $(3, 2)$ is added with probability

$$\begin{cases} B_{12} \cdot (1 - \frac{1}{2}e^{-1}) \cdot \frac{1}{2} = \frac{1}{6}(1 - \frac{1}{2e}) \approx 0.14 & \text{if edge } (3,2) \text{ is added} \\ B_{12} \cdot (1 - \frac{1}{2}e^{-0}) \cdot \frac{1}{2} = \frac{1}{12} \approx 0.08 & \text{otherwise} \end{cases} .$$

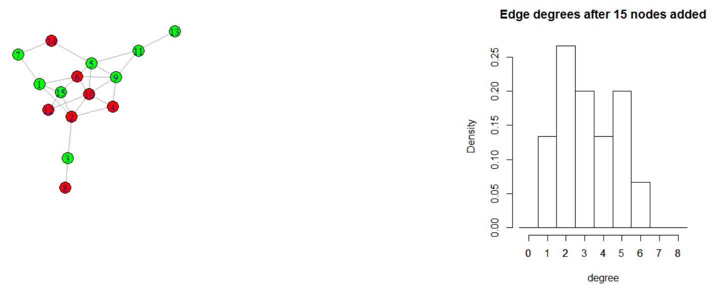
This shows that nodes added get a higher connection probability with nodes of high degree and nodes of the same group. Note that for lower values of q the variation in degree becomes a bigger factor in the connection probability as can be observed in the skewed degree distributions in Figure 3.



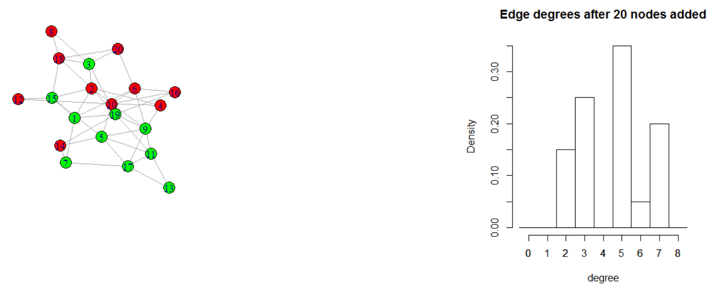
(a) After 5 timesteps



(b) After 10 timesteps

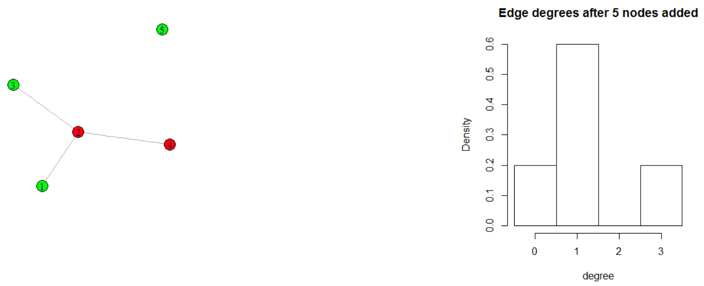


(c) After 15 timesteps

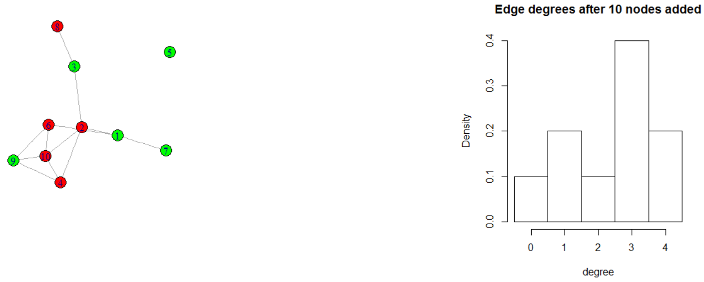


(d) After 20 timesteps

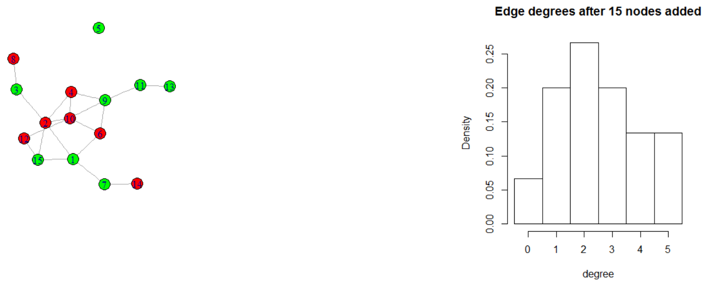
Figure 2: Timeline of the time dynamic stochastic block model with $k = 2$, $q_u = q = \frac{1}{2}$ and $B_{ii} = 0.66, B_{ij} = 0.33$ for all $i \neq j$. Communities are color coded. The even numbers are in one group and the odd numbers form the second group.



(a) After 5 timesteps



(b) After 10 timesteps



(c) After 15 timesteps

(d) After 20 timesteps

Figure 3: Timeline of the time dynamic stochastic block model with $k = 2$, $q_u = q = \frac{1}{100}$ and $B_{ii} = 0.66, B_{ij} = 0.33$ for all $i \neq j$. Communities are color coded. The even numbers are in one group and the odd numbers form the second group.

5.1 Main Result

To preserve consistency of Algorithm 1 for the time dynamic stochastic block model, we use the following theorem to show that the edge degrees of this model are comparable with the edge degrees in the stochastic block model:

Theorem 5.1. *Assume $k = 2$ and let $q_u = \frac{1}{2}$ for all $u \in [n]$. Then*

$$\left(1 - \mathcal{O}\left(\frac{1}{n}\right)\right) \sum_{v=1}^{n} \frac{P_{uv}}{2} \leq E[\hat{d}_u] \leq \sum_{v=1}^{n} \frac{P_{uv}}{2}.$$

Proof: From the assumptions $k = 2$ and $q_u = \frac{1}{2}$ we find that

$$P(A_{uv} = 1 | d_u = d_u(v-1)) = \frac{P_{uv}}{2} f(d_u).$$

Since $q_u \leq f(d_u) \leq 1$ we get $P(A_{uv} = 1 | d_u = d_u(v-1)) \leq \frac{P_{uv}}{2}$ and $P(A_{uv} = 1 | d_u = d_u(v-1)) = \frac{P_{uv}}{2} f(d_u) \geq \frac{P_{uv}}{4}$. Using the upper bound we find that

$$E[\hat{d}_u] \leq \sum_{v=1}^{n} \frac{P_{uv}}{2}, \tag{7}$$

which gives an upper bound for the expected degree. To find a lower bound we need the following Lemma:

Lemma 5.2. ***(Bernstein inequality for martingales)[9, p. 144]:*** *Let $(M_n)_{n \geq 0}$ be a $(\mathcal{F}_n)$-martingale such that for all $k \in [2, \infty), E[|M_{i+1} - M_i|^k | \mathcal{F}_i] \leq c^k k^k$ for some constant $c > 0$. Then*

$$\forall t \geq 0 : P(|M_n| \geq t) \leq 2 \exp\left(-\frac{t^2}{ce(2cn + t)}\right). \tag{8}$$

Let

$$V_k = \sum_{v=1}^{k} A_{uv} - p_{uv}$$

with $p_{uv} = E[A_{uv} | d_u(v-1)] = \frac{P_{uv}}{2}(1 - \frac{1}{2} e^{-d_u(v-1)})$. Then we find

$$\begin{aligned} E[V_k | V_1, ..., V_{k-1}] &= V_{k-1} + E[A_{uk} - p_{uk} | V_1, ..., V_{k-1}] \\ &= V_{k-1} + E[A_{uk} - p_{uk} | V_{k-1}] \\ &= V_{k-1} + E[A_{uk} | V_{k-1}] - E[p_{uk} | V_{k-1}] \\ &= V_{k-1} + E[A_{uk} | d_u(k-1)] - p_{uk} \\ &= V_{k-1} + p_{uk} - p_{uk} = V_{k-1} \end{aligned}$$

Note that

$$E[|V_k|] = E\left[\left|\sum_{v=1}^{k} A_{uv} - p_{uv}\right|\right] \leq \sum_{v=1}^{k} E\left[|A_{uv} - p_{uv}|\right] \leq \sum_{v=1}^{k} 1 = k < \infty,$$

so $(V_k)_k$ is a martingale with filtration $\mathcal{F}_k = \sigma(V_1, ..., V_{k-1})$. Checking the condition for Lemma 5.2 gives

$$\begin{aligned}
&E[|V_i - V_{i-1}|^k | V_1, ..., V_i] \\
&= E[|A_{ui} - p_{ui}|^k | V_1, ..., V_{i-1}] \\
&= E[|A_{ui} - p_{ui}|^k | d_u(i-1)] \\
&= |1 - p_{ui}|^k P(A_{ui} = 1 | d_u(i-1)) + |-p_{ui}|^k P(A_{ui} = 0 | d_u(i-1)) \\
&\leq |1 - p_{ui}|^k 1 + |p_{ui}|^k 1 \leq 2.
\end{aligned}$$

Thus for $c = 1$ the condition for the Bernstein inequality for Martingales holds. Lemma 5.2 therefore gives

$$\forall t \geq 0 : P(|V_k| \geq t) \leq 2 \exp\left(-\frac{t^2}{e(2k+t)}\right).$$

Choose

$$t^* := 2\sqrt{2e}\sqrt{\log n}\sqrt{k},$$

where $k \geq m = K \log(n)$. Then for $K > 2e$ we find

$$\begin{aligned}
t^* &= 2\sqrt{2e}\sqrt{\log(n)}\sqrt{k} \\
&= 2\sqrt{2e \log(n)}\sqrt{k} \\
&< 2\sqrt{K \log(n)}\sqrt{k} \\
&\leq 2\sqrt{k}\sqrt{k} = 2k.
\end{aligned}$$

This means

$$-\frac{(t^*)^2}{e(2k+t)} \leq -\frac{(t^*)^2}{e(4k)} = -\frac{8e \log(n) k}{4ek} = -2 \log n.$$

Substituting this in Bernstein's inequality for Martingales (8) we find for $k = m = K \log(n)$

$$P\left(\left|\sum_{v=1}^{m} A_{uv} - p_{uv}\right| \geq t^*\right) \leq 2 \exp\left(-2 \log n\right) = \frac{2}{n^2}.$$

Thus we find that

$$\left|\sum_{v=1}^{m} A_{uv} - p_{uv}\right| \leq 2\sqrt{2e}\sqrt{\log n}\sqrt{m} = 2\sqrt{2e}\sqrt{\log n}\sqrt{K \log n} = 2\sqrt{2e}\sqrt{K} \log(n),$$

with probability at least $1 - \frac{2}{n^2}$.

Let $p_{\min} \in (0,1)$ such that $0 < p_{\min} \leq p_{uv} < 1$ for all $u, v \in [n], u \neq v$. Let $0 < c^* \leq p_{\min}$. Then $c^* \in \mathbb{R}$ is a constant such that $\sum_{v=1}^{m} p_{uv} \geq \sum_{v=1}^{m} p_{\min} \geq K\log(n)c^*$. Since $\sum_{v=1}^{m} p_{uv} \geq K\log(n)c^*$, choosing $Kc^* \geq 1 + 2\sqrt{2eK}$ gives

$$\hat{d}_u(m) = \sum_{v=1}^{m} A_{uv} \geq \log n \text{ with probability at least } 1 - \frac{2}{n^2}. \tag{9}$$

Looking in the general case when a node $v > m+1$ gets added we find using (6) and $q_u = \frac{1}{2}$ for all u that

$$\begin{aligned} & P(A_{uv} = 1 | d_u(v-1))\mathbb{1}(d_u(v-1) \geq \log n) \qquad (10) \\ &= \frac{P_{uv}}{2}\left(1 - \frac{1}{2}e^{-d_u(v-1)}\right)\mathbb{1}(d_u(v-1) \geq \log n) \\ &\geq \frac{P_{uv}}{2}\left(1 - \frac{1}{2}e^{-\log(n)}\right) \\ &= \frac{P_{uv}}{2}\left(1 - \frac{1}{2n}\right). \qquad (11) \end{aligned}$$

For the expected degree of a node u we find

$$\begin{aligned} E[\hat{d}_u] = E[\sum_{j=1}^{n} A_{uj}] &= \sum_{j=1}^{n} E[A_{uj}] = \sum_{j=1}^{n} E[E[A_{uj}|d_u(j-1)]] \\ &= \sum_{j=1}^{n} E[P(A_{uj}|d_u(j-1))] \\ &\geq \sum_{j=m}^{n} E[P(A_{uj}|d_u(j-1))] \\ &\geq \sum_{j=m}^{n} P(d_u(j-1) \geq \log n)P(A_{uj}|d_u(j-1))\mathbb{1}(d_u(j-1) \geq \log n) \\ &\overset{(9),(10)}{\geq} \sum_{j=m}^{n} \left(1 - \frac{2}{n^2}\right)\frac{P_{uj}}{2}\left(1 - \frac{1}{2n}\right) \\ &\geq \sum_{j=m}^{n} \frac{P_{uj}}{2}\left(1 - \mathcal{O}\left(\frac{1}{n}\right)\right). \end{aligned}$$

Combining this with the upper bound in equation (7), we obtain

$$\left(1 - \mathcal{O}\left(\frac{1}{n}\right)\right)\sum_{j=m}^{n} \frac{P_{uj}}{2} \leq E[\hat{d}_u] \leq \sum_{j=1}^{n} \frac{P_{uj}}{2}.$$

□

6 Discussion

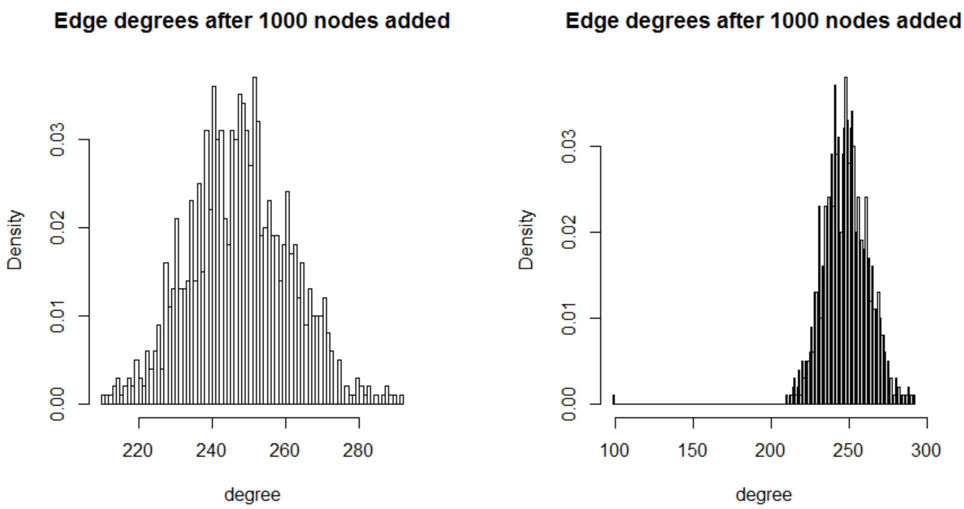
In this section we discuss a few important issues related to the time dynamic stochastic block model that could be studied in future research.

6.1 Assumptions of Theorem 5.1 and further work

For Theorem 5.1 on the edge degrees in the time dynamic stochastic block model, we worked under the assumption of two communities. Moreover, we assumed $q_u = \frac{1}{2}$. These assumptions are very restrictive and could be relaxed at the expense of more technical proofs. The assertion of Theorem 5.1 shows that the edge degrees for the stochastic block model and our time dynamic stochastic block model are the same up to a much smaller order. It is therefore plausible that Algorithm 1 and theory in Gao et al. [5] can be extended with minor modifications to the time dynamic stochastic block model. A formal proof is beyond the scope of this thesis and postponed for further research.

6.2 Generalizing the model

Theorem 5.1 on the edge degrees and simulations show that in the limit there are no hubs in the time dynamic stochastic block model (see Figure 4). This shows that this model does not include enough preferential attachment to generate hubs in large networks. Therefore a model with more influence of preferential attachment would be needed to model large networks with hubs. This could be attained by choosing q_u dependent on the node.



(a) $q = \frac{1}{2}$

(b) $q = \frac{1}{100}$

Figure 4: Edge degree distribution after 1000 nodes are added using the time dynamic stochastic block model using the same values as in Figure 2 (a) and Figure 3 (b).

7 Theoretical Properties

Below we replicate slight modifications of the proofs of Theorem 3.1 and Lemma 3.3, which are due to [5].

7.1 Proof of Theorem 3.1

To prove Theorem 3.1 we need the following lemmas which are proven by Gao et al. [5].

Lemma 7.1. **([5], Lemma 5)** *Consider a symmetric adjacency matrix* $A \in \{0,1\}^{n\times n}$ *and a symmetric matrix* $P \in [0,1]^{n\times n}$ *satisfying* $A_{uu} = 0$ *for all* $U \in [n]$ *and* $A_{uv} \sim Bern(P_{uv})$ *independently for all* $u > v$. *For any* $C' > 0$, *there exists some* $\hat{C} > 0$ *such that*

$$\|T_\tau(A) - P\|_{op} \leq \hat{C}\sqrt{np_{max} + 1},$$

with probability at least $1 - n^{-C'}$ *uniformly over* $\tau \in [C_1(np_{max}+1), C_2(np_{max}+1)]$ *for some sufficiently large constants* C_1, C_2, *where* $p_{max} = max_{u\geq v}P_{uv}$.

Lemma 7.2. ([5], Lemma 6) *For $P = (P_{uv}) = (B_{\sigma(u)\sigma(v)})$, we have singular value decomposition $P = U\Lambda UT$, where*

$$U = Z\Delta^{-1}W,$$

with $\Delta = diag(\sqrt{n_1}, ..., \sqrt{n_k}), Z \in \{0,1\}^{n\times k}$ is a matrix with exactly one nonzero entry in each row at $(i, \sigma(i))$ taking value 1 and $W \in O(k,k)$ where $O(k_1,k_2) = \{V \in \mathbb{R}^{k_1\times k_2} : V^TV = I_{k_2}\}$ for $k_1 \geq k_2$.

Proof of Theorem 3.1: By assumption, $\tau = C_2\overline{d}$ and thus $E[\tau] \in [C_1'a, C_2'a]$ for sufficiently large constants C_1' and C_2'. With Bernstein's inequality this gives $\tau \in [C_1^*a, C_2^*a]$ with probability at least $1 - e^{-C'n}$ for sufficiently large constants C_1^* and C_2^*. When (1) holds we have

$$\lambda_k \geq \frac{\sqrt{k}\sqrt{a}}{\sqrt{c}} \geq 2\hat{C}\sqrt{a},$$

where the last inequality holds for sufficiently small $c \in (0,1)$. Lemma 7.1 gives

$$\|T_\tau(A) - P\|_{op} \leq \hat{C}\sqrt{np_{max}+1} \leq \hat{C}\sqrt{n\frac{\alpha a}{n}+1} = \hat{C}\sqrt{\alpha a+1} \leq \hat{C}\sqrt{a}, \quad (12)$$

where $\hat{C} > 0$. Then

$$\lambda_k(T_\tau(A)) \geq \lambda_k(P) - \|T_\tau(A) - P\|_{op} \geq 2\hat{C}\sqrt{a} - \hat{C}\sqrt{a} = \hat{C}\sqrt{a} > 0.$$

Therefore there exists $c_1 \in (0,1)$ such that the k^{th} singular value of $T_\tau(A)$ is lower bounded by $c_1\lambda_k(P)$ with probability at least $1 - n^{-C'}$. With Davis-Kahan's sin-theta theorem [4] we then obtain

$$\|\hat{U} - UW_1\|_F \leq C\frac{\sqrt{k}}{\lambda_k}\|T_\tau(A) - P\|_{op},$$

where $W_1 \in O(k,k)$ and $C > 0$ is a constant. Using Lemma 7.2 we have $U = Z\Delta^{-1}W_2$ for some Z as in Lemma 7.2, $\Delta = \text{diag}(\sqrt{n_1}, ..., \sqrt{n_k})$ with n_i the size of community i and $W_2 \in O(k,k)$. This gives $UW_1 = Z\Delta^{-1}W_2W_1 = Z\Delta^{-1}W_3$ where $W_3 = W_2W_1 \in O(k,k)$. Let $V = Z\Delta^{-1}W_3$. Then we have

$$\|\hat{U} - V\|_F \leq C\frac{\sqrt{k}}{\lambda_k}\|T_\tau(A) - P\|_{op}. \quad (13)$$

Substituting (12) into (13) gives

$$\|\hat{U} - V\|_F \leq C\frac{\sqrt{k}\sqrt{a}}{\lambda_k} \quad (14)$$

with probability at least $1 - n^{-C'}$. Let $Q = \Delta^{-1} W_3 \in \mathbb{R}^{k \times k}$. Then $V = ZQ$. Note that $(V_{ui})_{i\in[k]} = (Q_{\sigma(u)i})_{i\in[k]}$ for all $u \in [n]$ by definition of Z. Therefore it follows that

$$\begin{aligned}\|(Q_{\sigma(u)i})_{i\in[k]} - (Q_{\sigma(v)i})_{i\in[k]}\| &= \|(V_{ui})_{i\in[k]} - (V_{vi})_{i\in[k]}\| \\ &= \begin{cases} \sqrt{\frac{1}{n_{\sigma(u)}} + \frac{1}{n_{\sigma(v)}}} & \text{if } \sigma(i) \neq \sigma(j), \\ 0 & \text{if } \sigma(i) = \sigma(j), \end{cases}\end{aligned}$$

where the terms of W_3 cancel because $W_3^T W_3 = \text{Id}$. Because of the parameter space Θ we have $n_{\sigma(u)} \in [\frac{n}{\beta k} - 1, \frac{\beta n}{k} + 1]$ for all $\sigma(u) \in [k]$. Therefore

$$\frac{1}{n_{\sigma(u)}} \geq \frac{1}{\frac{\beta n}{k} + 1} \approx \frac{1}{\frac{\beta n}{k}} = \frac{k}{\beta n}.$$

Thus if $\sigma(u) \neq \sigma(v)$ we have

$$\|(Q_{\sigma(u)i})_{i\in[k]} - (Q_{\sigma(v)i})_{i\in[k]}\| \geq \sqrt{\frac{2k}{\beta n}}. \tag{15}$$

Let

$$T_i = \left\{ u \in \sigma^{-1}(i) : \|(\hat{U}_{uj})_{j\in[k]} - (Q_{ij})_{j\in[k]}\| < \frac{r}{2} \right\}, \; i \in [k]$$

be the sets of nodes within critical radius $r = \mu\sqrt{\frac{k}{n}}$ from $(Q_{ij})_{j\in[k]}$. Note that when $\mu < \sqrt{\frac{2}{\beta}}$, the sets $\{T_i\}_{i\in[k]}$ are disjoint because of (15). Then

$$\cup_{i\in[k]} T_i = \left\{ u \in [n] : \|(\hat{U}_{uj})_{j\in[k]} - (V_{uj})_{j\in[k]}\| < \frac{r}{2} \right\}.$$

For the amount of nodes not included in $\cup_{i\in[k]} T_i$ we find

$$\frac{r^2}{4} |(\cup_{i\in[k]} T_i)^c| \leq \sum_{u\in[n]} \|(\hat{U}_{uj})_{j\in[k]} - (V_{uj})_{j\in[k]}\|^2 \leq \frac{C^2 ka}{\lambda_k^2},$$

where the last inequality holds because of (14). Substituting the critical radius $r = \mu\sqrt{\frac{k}{n}}$ gives the upper bound

$$|(\cup_{i\in[k]} T_i)^c| \leq \frac{4C^2 ka}{\mu^2 \lambda_k^2 \frac{k}{n}} = \frac{4C^2 na}{\mu^2 \lambda_k^2}. \tag{16}$$

By means of contradiction we prove that

$$|T_i| \geq |\sigma^{-1}(i)| - |(\cup_{i\in[k]} T_i)^c|. \tag{17}$$

Suppose $|T_i| < |\sigma^{-1}(i)| - |(\cup_{i\in[k]} T_i)^c|$ for some $i \in [k]$. Because the sets $\{T_i\}_{i\in[k]}$ are disjoint we find

$$| \cup_{i\in[k]} T_i| = \sum_{i\in[k]} |T_i| < n - |(\cup_{i\in[k]} T_i)^c| = | \cup_{i\in[k]} T_i|.$$

This is a contradiction, so (17) holds. This means the size of T_i is lower bounded by

$$|T_i| \geq |\sigma^{-1}(i)| - |(\cup_{i\in[k]} T_i)^c| \geq \frac{n}{\beta k} - \frac{4C^2 na}{\mu^2 \lambda_k^2} > \frac{n}{2\beta k} \text{ for all } i \in [k], \tag{18}$$

where we use Assumption (1) for the last inequality with c chosen small enough such that $c < \frac{\mu^2}{8C^2\beta}$.

Thus most data points in $\{(\hat{U}_{uj})_{j\in[k]}\}_{u\in[n]}$ are within critical radius of the points $\{(Q_{ij})_{j\in[k]}\}_{i\in[k]}$. Note that $\{T_i\}_{i\in[k]}$ as well as $\{\hat{C}_i\}_{i\in[k]}$ in Algorithm 1 are defined by the critical radius $r = \mu\sqrt{\frac{k}{n}}$. When $\mu < \frac{1}{4}\sqrt{\frac{2}{\beta}}$, every set T_i can intersect with at most one set C_j.

We use induction to prove that there exists a permutation $\pi \in S_k$, with S_k the symmetric group, such that for $\hat{C}_i$ as defined in Algorithm 1 we have

$$\hat{C}_i \cap T_{\pi(i)} \neq \varnothing \text{ and } |\hat{C}_i| \geq |T_{\pi(i)}| \text{ for each } i \in [k]. \tag{19}$$

For $i = 1$ we have $|\hat{C}_1| \geq \max_{i\in[k]} |T_i|$ because of the way $\hat{C}_1$ is constructed in Algorithm 1. Suppose now that there exists no set T_i that overlaps with $\hat{C}_1$. Then

$$|(\cup_{i\in[k]} T_i)^c| \geq |\hat{C}_1| \geq \max_{i\in[k]} |T_i| \geq \frac{n}{2\beta k}, \tag{20}$$

where the last inequality is given in (18). However from (16) with Assumption (1) we find for $c < \frac{\mu^2}{8C^2\beta}$ that

$$|(\cup_{i\in[k]} T_i)^c| \leq \frac{4C^2 na}{\mu^2 \lambda_k^2} < \frac{n}{2\beta k},$$

which contradicts with (20). Thus there must exist a set T_j that overlaps with $\hat{C}_1$. Write $\pi(1) = j$. We find that

$$\begin{aligned}
|\hat{C}_1^c \cap T_{\pi(1)}| &= |T_{\pi(1)}| - |T_{\pi(1)} \cap \hat{C}_1| \\
&\leq |\hat{C}_1| - |T_{\pi(1)} \cap \hat{C}_1| \\
&= |\hat{C}_1 \cap T_{\pi(1)}^c| \\
&\leq |(\cup_{i\in[k]} T_i)^c|,
\end{aligned}$$

where we use that $T_{\pi(i)} \cap \hat{C}_1 = \varnothing$ for all $i \neq 1$. With (16) we find

$$|\hat{C}_i^c \cap T_{\pi(i)}| \leq \frac{4C^2 na}{\mu^2 \lambda_k^2}, \tag{21}$$

for $i = 1$. Thus (19) holds for $i = 1$.

Suppose that (19) and (21) hold for $i \in [l-1]$ with $l \leq k$. Since we know that each $\hat{C}_i$ overlaps with at most one T_j we find that

$$\left| \left(\bigcup_{i \in [l-1]} \hat{C}_i \right) \cap \left(\bigcup_{j \in \left([k] \setminus \{\pi(i)\}_{i \in [l-1]} \right)} T_j \right) \right| = \varnothing.$$

After $l-1$ steps in Algorithm 1 we therefore have $\bigcup_{j \in \left([k] \setminus \{\pi(i)\}_{i \in [l-1]} \right)} T_j \subseteq S$. Since $\hat{C}_l$ is defined in the algorithm by the set of points that are critical radius r apart while the sets T_j are constructed with critical radius $\frac{r}{2}$ we find that

$$|\hat{C}_l| \geq \max_{j \in \left([k] \setminus \{\pi(i)\}_{i \in [l-1]} \right)} |T_j| \geq \frac{n}{2\beta k}.$$

Suppose that $\hat{C}_l$ overlaps with $T_{\pi(i)}$ for some $i \in [l-1]$. Since (15) holds, the sets $\{T_i\}_{i \in [k]}$ are seperated from each other in such a way that $\hat{C}_l$ can only overlap with one T_i. Thus $T_{\pi(i)}$ is the only set from $\{T_i\}_{i \in [k]}$ that overlaps $\hat{C}_l$. Therefore

$$|\hat{C}_l| \leq |\hat{C}_l \cap T_{\pi(i)}| + |(\cup_{i \in [k]} T_i)^c|. \tag{22}$$

Since $\hat{C}_l$ gets constructed in a later step in Algorithm 1 than $\hat{C}_{\pi(i)}$, we have $|\hat{C}_l \cap T_{\pi(i)}| \leq |\hat{C}_i^c \cap T_{\pi(i)}|$. Substituting this in (22) with upper bound (21) and applying (16) we find that

$$|\hat{C}_l| \leq \frac{8C^2 na}{\mu^2 \lambda_k^2}. \tag{23}$$

Note that this contradicts with the upper bound $|\hat{C}_l| \geq \frac{n}{2\beta k}$ found in (20). Therefore we must have $\hat{C}_l \cap T_{\pi(i)} = \varnothing$ for all $i \in [l-1]$. Suppose now that $\hat{C}_l \cap T_{\pi(i)} = \varnothing$ for all $i \in [k]$. Then we have

$$|(\cup_{i \in [k]} T_i)^c| \geq |\hat{C}_l| \geq \frac{n}{2\beta k},$$

which contradicts (16). Therefore $\hat{C}_l \cap T_{\pi(l)} \neq \varnothing$ for some $\pi(l) \in [k] \setminus \cup_{i=1}^{l-1} \{\pi(i)\}$ and (19) holds for $i = l$. Using the same argument that is used to prove (21) for

$i = 1$ we have

$$\begin{aligned}|\hat{C}_l^c \cap T_{\pi(l)}| &= |T_{\pi(l)}| - |T_{\pi(l)} \cap \hat{C}_l| \\ &\leq |\hat{C}_l| - |T_{\pi(l)} \cap \hat{C}_l| \\ &= |\hat{C}_l \cap T_{\pi(l)}^c| \\ &\leq |(\cup_{i\in[k]} T_i)^c|,\end{aligned}$$

which proves (21) for $i = l$ given that (19) and (21) hold for $i \in [l-1]$. Therefore by induction there exists a permutation $\pi \in S_k$, with S_k the symmetric group, such that for $\hat{C}_i$, as defined in Algorithm 1, (19) and (21) hold.

Since we know that $\hat{C}_j \cap T_{\pi(j)} \neq \varnothing$, no other set $T_{\pi(i)}$ with $i \neq j$ can have overlap with $\hat{C}_j$. This means $\left(\bigcup_{j\neq i} \hat{C}_j\right) \cap T_{\pi(i)} = \varnothing$. Thus $T_{\pi(i)} \subseteq \left(\bigcup_{j\neq i} \hat{C}_j\right)^c$. Since $T_{\pi(i)} \cap \hat{C}_i^c \subseteq \hat{C}_i^c$ by default we get

$$T_{\pi(i)} \cap \hat{C}_i^c \subseteq \left(\bigcup_{j\in[k]} \hat{C}_j\right)^c.$$

This holds for all $i \in [k]$, so we find that

$$\bigcup_{i\in[k]} \left(T_{\pi(i)} \cap \hat{C}_i^c\right) \subseteq \left(\bigcup_{j\in[k]} \hat{C}_j\right)^c.$$

Because the sets $\{T_i\}_{i\in[k]}$ do not overlap we obtain

$$\sum_{i\in[k]} |T_{\pi(i)} \cap \hat{C}_i^c| = \left|\bigcup_{i\in[k]} \left(T_{\pi(i)} \cap \hat{C}_i^c\right)\right| \leq |(\cup_{j\in[k]} \hat{C}_j)^c|. \tag{24}$$

Note that the sets $\{\hat{C}_i\}_{i\in[k]}$ do not overlap. Therefore we find using $|\hat{C}_i| \geq |T_{\pi(i)}|$ for each $i \in [k]$ from (19) that

$$|(\cup_{i\in[k]} \hat{C}_i)^c| = n - \sum_{i\in[k]} |\hat{C}_i| \leq n - \sum_{i\in[k]} |T_i| = |(\cup_{i\in[k]} T_i)^c|. \tag{25}$$

Because of (16) we then have

$$|(\cup_{i\in[k]} T_i)^c| \leq \frac{4C^2 na}{\mu^2 \lambda_k^2},$$

which means

$$\sum_{i\in[k]} |T_{\pi(i)} \cap \hat{C}_i^c| \leq \frac{4C^2 na}{\mu^2 \lambda_k^2}. \tag{26}$$

Note that any $u \in T_{\pi(i)} \cap \hat{C}_i$ has not been mislabeled under permutation π, so $\hat{\sigma}(u) = \pi^{-1}(\sigma(u))$. Therefore we can bound the misclassification proportion from above as follows:

$$\begin{aligned} l(\hat{\sigma}, \pi^{-1}(\sigma)) &\leq \frac{1}{n} \left| \left(\cup_{i\in[k]} (T_{\pi(i)} \cap \hat{C}_i) \right)^c \right| \\ &\leq \frac{1}{n} \left(\left| \left(\cup_{i\in[k]} (T_{\pi(i)} \cap \hat{C}_i) \right)^c \cap (\cup_{i\in[k]} T_i) \right| + |(\cup_{i\in[k]} T_i)^c| \right) \\ &\leq \frac{1}{n} \left(\sum_{i\in[k]} |T_{\pi(i)} \cap \hat{C}_i^c| + |(\cup_{i\in[k]} T_i)^c| \right) \end{aligned}$$

Applying the upper bounds found in (16) and (26) we find that

$$l(\hat{\sigma}, \pi^{-1}(\sigma)) \leq \frac{8C^2 na}{\mu^2 \lambda_k^2},$$

which proves Theorem 3.1. □

7.2 Proof of Lemma 3.3

We follow closely the proof of Lemma 1 in the paper by Gao et al. [5]. Let $\Theta = \Theta(n, k, a, b, \lambda, \beta; \alpha)$. For any community assignments σ_1 and σ_2, recall the Hamming loss

$$l(\sigma_1, \sigma_2) = \frac{1}{n} \sum_{u=1}^{n} \mathbb{1}(\sigma_1(u) \neq \sigma_2(u))$$

which computes the proportion of misclassified labels between σ_1 and σ_2. In the first step of the proof we choose (B, σ) and $u \in [n]$ constant. We work under the event $E_u = \{l(\pi_u(\sigma), \sigma_u^0 \leq \gamma)\}$ where π_u is a permutation on $[k]$ and σ_u^0 is the community assignment obtained from applying Algorithm 1 to A_{-u} and setting $\sigma_u^0(u) = 0$ as in Algorithm 2. For the rest of the proof we assume that $\pi_u = \mathrm{Id}$ to simplify equations. By Assumption 3 event E_u occurs with high probability. On event E_u the maximum amount of misclassified nodes is γn. Choose any $i \in [k]$ and let $C_{i,\sigma}$ be the set of nodes with community label i. Using the notation $n_i := |C_{i,\sigma}|$ we find

$$n_i \geq |\tilde{C}_{i,\sigma}^u \cap C_{i,\sigma}| \geq n_i - \gamma n \text{ and } |\tilde{C}_{i,\sigma}^u \cap C_{i,\sigma}^c| \leq \gamma n. \tag{27}$$

Let $C_i' \subseteq [n]$ such that (27) holds with C_i' in place of $\tilde{C}_{i,\sigma}^u$. At most γn nodes with true community i can be mislabelled as community $j \neq i$ and at most γn nodes with true community $j \neq i$ can be wrongly labelled as community i. Therefore there are at most $\sum_{l=0}^{\gamma n} \binom{n_i}{l} \sum_{m=0}^{\gamma n} \binom{n-n_i}{l}$ different subsets $C_i' \subseteq [n]$ with this property. Using Stirling's formula we find

$$\begin{aligned}
\sum_{l=0}^{\gamma n} \binom{n_i}{l} \sum_{m=0}^{\gamma n} \binom{n-n_i}{l} &\leq (\gamma n+1)^2 \left(\frac{en_i}{\gamma n}\right)^{\gamma n} \left(\frac{en}{\gamma n}\right)^{\gamma n} \\
&\overset{\frac{n_i}{n}\leq 1}{\leq} (\gamma n+1)^2 \left(\frac{e}{\gamma}\right)^{\gamma n} \left(\frac{e}{\gamma}\right)^{\gamma n} \\
&= \exp\left\{2\log(\gamma n+1) + 2\gamma n \log\left(\frac{e}{\gamma}\right)\right\} \\
&\leq \exp\left\{C_1 \gamma n \log\left(\frac{1}{\gamma}\right)\right\},
\end{aligned}$$

where $C_1 > 0$ is a constant. Define $\mathcal{E}_i'$ as the set of edges between nodes in C_i'. The nodes in C_i' can be divided into two groups: nodes with true community i and nodes with a different true community $j \neq i$. By equation (27) there are at least $n_i - \gamma n$ nodes with true community i. By definition of the parameter space Θ we know that $n_i \geq \frac{n}{\beta k} - 1$. Combining both, the proportion of nodes in C_i' with true community i is at least

$$\frac{n_i - \gamma n}{n_i} = 1 - \frac{\gamma n}{n_i} \geq 1 - \frac{\gamma n}{\frac{n}{\beta k} - 1} = 1 - \frac{\gamma n \beta k}{n - \beta k} \overset{n\to\infty}{\approx} 1 - \beta\gamma k$$

This means at least a $(1-\beta\gamma k)^2$ proportion of $|\mathcal{E}_i'|$ contains edges between two nodes with true community i. These follow the $\text{Bern}(B_{ii})$ distribution. Similarly, the proportion of nodes of C_i' with true community $j \neq i$ is at most $\beta\gamma k$. This means that at most $(\beta\gamma k)^2$ proportion of $|\mathcal{E}_i'|$ are edges between two nodes of the same community $j \neq i$. These have distributions that are stochastically smaller than $\text{Bern}\left(\frac{\alpha a}{n}\right)$ and stochastically larger than $\text{Bern}\left(\frac{a}{n}\right)$ by definition of the parameter space Θ. There are also edges between nodes of different communities. We find that there are at most $2(\beta\gamma k)(1-\beta\gamma k) = 2\beta\gamma k - 2(\beta\gamma k)^2 \leq 2\beta\gamma k$ edges of this type. Edges between different communities follow a distribution stochastically smaller than $\text{Bern}\left(\frac{b}{n}\right)$. Then in the worst case scenario there is a proportion of $(1-\beta\gamma k)^2$ edges with distribution $\text{Bern}(B_{ii})$, $2\beta\gamma k - 2(\beta\gamma k)^2$ edges with distribution stochastically smaller than $\text{Bern}\left(\frac{b}{n}\right)$ and $(\beta\gamma k)^2$ edges with distribution stochastically larger than $\text{Bern}\left(\frac{a}{n}\right)$.

Combining these observations we obtain

$$\begin{aligned}
(1-\beta\gamma k)^2 B_{ii} + (\beta\gamma k)^2 \frac{a}{n} &\leq E\left[\frac{|\mathcal{E}_i'|}{\frac{1}{2}|C_i'|(|C_i|-1)}\right] \\
&\leq \max_{t\in[0,\beta\gamma k]} \left\{(1-t)^2 B_{ii} + t^2 \frac{\alpha a}{n} + 2t\frac{b}{n}\right\}. \qquad (28)
\end{aligned}$$

The assumption $\gamma = o(\frac{1}{k \log k})$ in Lemma 3.3 gives $\beta\gamma k = o(\frac{\beta}{\log k}) = o(1)$. We can further refine the left hand side of the previous inequality by using that

$$(\beta\gamma k)^2 \frac{a}{n} = \beta\gamma k B_{ii} o(1),$$

which gives

$$\begin{aligned}(1-\beta\gamma k)^2 B_{ii} + (\beta\gamma k)^2 \frac{a}{n} &= (1 - 2\beta\gamma k + \beta\gamma k o(1)) B_{ii} + \beta\gamma k B_{ii} o(1) \\ &= (1 - (2 + o(1))\beta\gamma k) B_{ii}.\end{aligned}$$

For $t = \beta\gamma k$ we find that the right hand side of (28) results in

$$\begin{aligned}&(1-\beta\gamma k)^2 B_{ii} + (\beta\gamma k)^2 \frac{\alpha a}{n} + 2\beta\gamma k \frac{b}{n} \\ &= B_{ii} + \beta\gamma k \cdot o(1) \left(B_{ii} + \frac{\alpha a}{n}\right) + 2\beta\gamma k \left(\frac{b}{n} - B_{ii}\right) \\ &\overset{n\to\infty}{\approx} B_{ii} + 2\beta\gamma k \left(\frac{b}{n} - B_{ii}\right) \leq B_{ii}\end{aligned}$$

where the last inequality holds since $\frac{b}{n} < B_{ii}$. For the right hand side of (28), the derivative of $(1-t)^2 B_{ii} + t^2 \frac{\alpha a}{n} + 2t\frac{b}{n}$ is

$$-2(1-t)B_{ii} + 2t\frac{\alpha a}{n} + 2\frac{b}{n} = 2t\left(B_{ii} + \frac{\alpha a}{n}\right) + 2\left(\frac{b}{n} - B_{ii}\right).$$

Note that $\frac{b}{n} < \frac{a}{n} \leq B_{ii} \leq \frac{\alpha a}{n}$, so for $t = 0$ the derivative is $2(\frac{b}{n} - B_{ii}) < 0$. Then $(1-t)^2 B_{ii} + t^2 \frac{\alpha a}{n} + 2t\frac{b}{n}$ is a quadratic equation with negative derivative at $t = 0$ and the value at $t = 0$ is greater than the value at $t = \beta\gamma k$. Therefore the maximum of $(1-t)^2 B_{ii} + t^2 \frac{\alpha a}{n} + 2t\frac{b}{n}$ on $[0, \beta\gamma k]$ is attained at $t = 0$ with value B_{ii}.

From (28) it thus holds that

$$\left| E\left[\frac{|\mathcal{E}_i'|}{\frac{1}{2}|C_i'|(|C_i| - 1)}\right] - B_{ii}\right| \leq C\beta\gamma k \frac{\alpha a}{n} \leq \eta' \left(\frac{a-b}{n}\right). \tag{29}$$

Because we assume that $\gamma = o(\frac{a-b}{ak})$ in Lemma 3.3, the last inequality holds for some $\eta' \leq \hat{C}\beta\alpha\frac{a-b}{n}$ that depends on a, k, α, β and γ.

Since each of the edges between nodes in community i are chosen independently, we find that

$$\big||\mathcal{E}_i'| - E[\mathcal{E}_i']\big| = \left|\sum_{j\in\mathcal{E}_i'} X_j - \sum_{j\in\mathcal{E}_i'} E[X_j]\right| = \left|\sum_{j\in\mathcal{E}_i'} (X_j - E[X_j])\right|,$$

where X_j are Bernoulli distributed random variables. Note that $X_j \in \{0,1\}$ and $E[X_j] \in [0,1]$, so $|X_j - E[X_j]| \leq 1$. Using Bernstein's inequality we find

$$
\begin{aligned}
P(\big||\mathcal{E}_i'| - E|\mathcal{E}_i'|\big| \geq t) &= P(|\sum_{j\in\mathcal{E}_i'}(X_j - E[X_j])| \geq t) \\
&\leq \exp\left\{-\frac{t^2/2}{\sum_{j\in\mathcal{E}_i'} \mathrm{Var}(X_j) + \frac{t}{3}}\right\} \\
&= \exp\left\{-\frac{t^2}{2\sum_{j\in\mathcal{E}_i'} \mathrm{Var}(X_j) + \frac{2t}{3}}\right\}. \qquad (30)
\end{aligned}
$$

Because X_j are Bernoulli distributed with parameter $p_j \leq \frac{\alpha a}{n}$ for all $j \in C_i'$, we have

$$
\mathrm{Var}(X_j) = p_j(1-p_j) \leq p_j \leq \frac{\alpha a}{n},
$$

so

$$
\begin{aligned}
\sum_{j\in\mathcal{E}_i'} \mathrm{Var}(X_j) &\leq \sum_{j\in\mathcal{E}_i'} \frac{\alpha a}{n} \\
&= \frac{|C_i'|(|C_i'|-1)}{2}\frac{\alpha a}{n} \\
&\leq \frac{1}{2}(n_i + \gamma n)^2 \frac{\alpha a}{n}.
\end{aligned}
$$

Substituting this in (30) gives

$$
P(\big||\mathcal{E}_i'| - E|\mathcal{E}_i'|\big| \geq t) \leq \exp\left\{-\frac{t^2}{(n_i+\gamma n)^2\frac{\alpha a}{n} + \frac{2t}{3}}\right\}. \qquad (31)
$$

Choose $(t^*)^2$ as

$$
\begin{aligned}
(t^*)^2 =& (n_i + \gamma n)^2 \frac{\alpha a}{n}(C_1 \gamma n \log \gamma^{-1} + (3+\delta)\log n) \\
&\vee (2C_1\gamma n \log\gamma^{-1} + 2(3+\delta)\log n)^2 \qquad (32)
\end{aligned}
$$

where $\vee$ denotes the maximum of the expressions on the left and right hand side. Define

$$
c := C_1 \gamma n \log \gamma^{-1} + (3+\delta)\log n.
$$

Then

$$
(t^*)^2 = (n_i+\gamma n)^2 \frac{\alpha a}{n} c \vee 4c^2.
$$

We can distinquish two situations: either (I) $(n_i+\gamma n)^2\frac{\alpha a}{n} \leq 4c$ or (II) $(n_i+\gamma n)^2\frac{\alpha a}{n} > 4c$. In case (I) we find

$$\begin{aligned}(t^*)^2 &= (n_i+\gamma n)^2\frac{\alpha a}{n}(C_1\gamma n\log\gamma^{-1} + (3+\delta)\log n)\\ &\lesssim \left(\frac{n}{k}\right)^2 a\gamma\log\gamma^{-1}.\end{aligned}$$

Here we use the fact that $\frac{\log x}{x}$ is a monotone decreasing function in x which means $\gamma\log\gamma^{-1} \geq \frac{1}{n}\log n$ for any $\gamma \geq \frac{1}{n}$. Note that in the situation where $\gamma < \frac{1}{n}$, γ approaches 0 as $n\to\infty$ which means the initialization already predicted the true community labels. Therefore we only look at the situation where $\gamma \geq \frac{1}{n}$ which means $\gamma n\log\gamma^{-1} \lesssim \log n$. Recall that $n_i \leq \frac{\beta n}{k}+1$ because of the parameter space Θ. With the assumption in Lemma 3.3 that $\gamma = o(\frac{1}{k\log k})$ we find $(n_i+\gamma n)^2 \lesssim \left(\frac{n}{k}\right)^2$.

In case (II) we see that

$$\begin{aligned}(t^*)^2 &= 4\left(C_1\gamma n\log\gamma^{-1} + (3+\delta)\log n\right)^2\\ &\lesssim (\gamma n\log\gamma^{-1})^2,\end{aligned}$$

where we again use that $\gamma n\log\gamma^{-1} \lesssim \log n$ for any $\gamma \geq \frac{1}{n}$.

Combining (I) and (II) gives

$$\begin{aligned}(t^*)^2 &= (n_i+\gamma n)^2\frac{\alpha a}{n}c \vee 4c^2\\ &\lesssim \left(\frac{n}{k}\right)^2 a\gamma\log\gamma^{-1} + \left(\gamma n\log\gamma^{-1}\right)^2\\ &\lesssim \left(\frac{n}{k}\sqrt{a\gamma\log\gamma^{-1}} + \gamma n\log\gamma^{-1}\right)^2.\end{aligned}$$

Substituting $t^* = \sqrt{(t^*)^2}$ from (32) in Bernstein's inequality (31) gives in case (I) $(n_i+\gamma n)^2\frac{\alpha a}{n} \leq 4(C_1\gamma n\log\gamma^{-1} + (3+\delta)\log n)$ that

$$\begin{aligned}&P\left(\left||\mathcal{E}_i'| - E|\mathcal{E}_i'|\right| \geq C_{\alpha,\beta,\delta}\left(\frac{n}{k}\sqrt{a\gamma\log\gamma^{-1}} + \gamma n\log\gamma^{-1}\right)\right)\\ &\leq \exp\left\{-\frac{(n_i+\gamma n)^2\frac{\alpha a}{n}(C_1\gamma n\log\gamma^{-1}+(3+\delta)\log n)}{(n_i+\gamma n)^2\frac{\alpha a}{n} + \frac{2}{3}\sqrt{(n_i+\gamma n)^2\frac{\alpha a}{n}(C_1\gamma n\log\gamma^{-1}+(3+\delta)\log n)}}\right\}\\ &\leq \exp\left\{-\frac{(n_i+\gamma n)^2\frac{\alpha a}{n}(C_1\gamma n\log\gamma^{-1}+(3+\delta)\log n)}{4(C_1\gamma n\log\gamma^{-1}+(3+\delta)\log n) + \frac{2}{3}\sqrt{4(C_1\gamma n\log\gamma^{-1}+(3+\delta)\log n)^2}}\right\}\\ &\lesssim \exp\left\{-(n_i+\gamma n)^2\frac{\alpha a}{n}\right\},\end{aligned}$$

where $C_{\alpha,\beta,\delta}$ is a constant depending on α, β and δ. Substituting t^* in (31) gives in case (II) $(n_i + \gamma n)^2 \frac{\alpha a}{n} > 4(C_1 \gamma n \log \gamma^{-1} + (3+\delta)\log n)$ that

$$
\begin{aligned}
&P\left(\left||\mathcal{E}_i'| - E|\mathcal{E}_i'|\right| \geq C_{\alpha,\beta,\delta}\left(\frac{n}{k}\sqrt{a\gamma\log\gamma^{-1}} + \gamma n \log\gamma^{-1}\right)\right) \\
&\leq \exp\left\{-\frac{(2C_1\gamma n\log\gamma^{-1} + 2(3+\delta)\log n)^2}{(n_i+\gamma n)^2\frac{\alpha a}{n} + \frac{2}{3}(2C_1\gamma n\log\gamma^{-1} + 2(3+\delta)\log n)}\right\} \\
&\leq \exp\left\{-\frac{(2C_1\gamma n\log\gamma^{-1} + 2(3+\delta)\log n)^2}{4(C_1\gamma n\log\gamma^{-1} + (3+\delta)\log n) + \frac{2}{3}(2C_1\gamma n\log\gamma^{-1} + 2(3+\delta)\log n)}\right\} \\
&\lesssim \exp\{-C_1\gamma n\log\gamma^{-1} - (3+\delta)\log n\} \\
&= \exp\{-C_1\gamma n\log\gamma^{-1}\} n^{-(3+\delta)}
\end{aligned}
$$

Thus we get

$$
\begin{aligned}
&P\left(\left||\mathcal{E}_i'| - E|\mathcal{E}_i'|\right| \geq C_{\alpha,\beta,\delta}\left(\frac{n}{k}\sqrt{a\gamma\log\gamma^{-1}} + \gamma n \log\gamma^{-1}\right)\right) \\
&\lesssim \exp\{-C_1\gamma n\log\gamma^{-1}\} n^{-(3+\delta)}.
\end{aligned}
$$

This means with probability at least $1 - \exp\{-C_1\gamma n\log\gamma^{-1}\}n^{-(3+\delta)}$ we have

$$
\begin{aligned}
\left|\frac{|\mathcal{E}_i'|}{\frac{1}{2}|C_i'|(|C_i'|-1)} - \frac{E|\mathcal{E}_i'|}{\frac{1}{2}|C_i'|(|C_i'|-1)}\right| &\leq \frac{C_{\alpha,\beta,\delta}(\frac{n}{k}\sqrt{a\gamma\log\gamma^{-1}} + \gamma n\log\gamma^{-1})}{\frac{1}{2}|C_i'|(|C_i'|-1)} \\
&\cong \hat{C}_{\alpha,\beta,\delta}\left(\frac{k}{n}\sqrt{a\gamma\log\gamma^{-1}} + \frac{k^2\gamma\log\gamma^{-1}}{n}\right),
\end{aligned}
$$

where we use that $|C_i| \cong \frac{n}{k}$. Note that because $\sqrt{ak} \ll a-b$ under the assumption that $\frac{(a-b)^2}{ak} \to \infty$ and $k\gamma\log\gamma^{-1} = O(1)$ which gives $k^2\gamma\log\gamma^{-1} \lesssim k \ll \frac{(a-b)^2}{a} \lesssim a-b$. So

$$
\hat{C}_{\alpha,\beta,\delta}\left(\frac{k}{n}\sqrt{a\gamma\log\gamma^{-1}} + \frac{k^2\gamma\log\gamma^{-1}}{n}\right) \leq \eta''\left(\frac{a-b}{n}\right),
$$

where $\eta'' \overset{n\to\infty}{\to} 0$ is a constant that depends on $a, k, \alpha, \beta, \gamma$ and δ. Therefore

$$
\left|\frac{|\mathcal{E}_i'|}{\frac{1}{2}|C_i'|(|C_i'|-1)} - \frac{E|\mathcal{E}_i'|}{\frac{1}{2}|C_i'|(|C_i'|-1)}\right| \leq \eta''\left(\frac{a-b}{n}\right). \tag{33}
$$

Combining (29) and (33) we find with probability at least $1 - \exp\{-C_1\gamma n\log\gamma^{-1}\}n^{-(3+\delta)}$ that

$$
\left|\frac{|\tilde{\mathcal{E}}_i'|}{\frac{1}{2}|\tilde{C}_i'|(|\tilde{C}_i'|-1)} - B_{ii}\right| \leq \eta\left(\frac{a-b}{n}\right), \tag{34}
$$

where $\eta \overset{n\to\infty}{\to} 0$ depends on $a, k, \alpha, \beta, \gamma$ and δ. Note that $\frac{|\tilde{\mathcal{E}}_i'|}{\frac{1}{2}|\tilde{C}_i'|(|\tilde{C}_i'|-1)} = \hat{B}_{ii}$ in Algorithm 2. The proof for $\hat{B}_{ij}$ in place of $\hat{B}_{ii}$ follows the same structure.

Thus with probability at least $1 - \exp\{-C_1 \gamma n \log \gamma^{-1}\} n^{-(3+\delta)}$ we have

$$\left| \hat{B}_{ij} - B_{ij} \right| \leq \eta \left(\frac{a-b}{n} \right) \text{ for all } i,j \in [k].$$

Recall Boole's inequality (also known as the union bound), where for a countable set of events $A_1, A_2, \ldots$ it holds that

$$P\left(\bigcup_i A_i \right) \leq \sum_i P(A_i).$$

This gives

$$\begin{aligned}
&P\left(\max_{i,j\in[k]} |\hat{B}^n_{ij} - B_{ij}| > \eta \left(\frac{a-b}{n} \right) \right) \\
&= P\left(\exists i,j \in [n] : |\hat{B}^n_{ij} - B_{ij}| > \eta \left(\frac{a-b}{n} \right) \right) \\
&\leq \sum_{i,j\in[n]} P\left(|\hat{B}^n_{ij} - B_{ij}| > \eta \left(\frac{a-b}{n} \right) \right) \\
&\leq n(n-1) \exp\{-C_1 \gamma n \log \gamma^{-1}\} n^{-(3+\delta)} \\
&= \exp\{-C_1 \gamma n \log \gamma^{-1}\} n^{-(1+\delta)}.
\end{aligned}$$

Therefore

$$P\left(\max_{i,j\in[k]} |\hat{B}^n_{ij} - B_{ij}| \leq \eta \left(\frac{a-b}{n} \right) \right) \geq 1 - \exp\{-C_1 \gamma n \log \gamma^{-1}\} n^{-(1+\delta)},$$

which is independent of the choice of u and (B, σ). This proves Claim (4) in Lemma 3.3. $\square$

8 Code

The following R code generates a graph according to the stochastic block model for Figure 1:

```
library(igraph)
set.seed(2016)
N = 20 #nr of nodes
k = 2 #nr of communities
sigma = c(rep(c(1),N/k),rep(c(2),N/k))#set communities
                                      #for every group
#sigma = c(rep(c(1,2),N/k))
B = matrix(data = c(0.66, 0.33,0.33,0.66), k) #community adjacency matrix
A = matrix(0,N, N) #initialize adjacency matrix

for(i in 1:(N-1)) { #for every node added
  for(j in (i+1):N) { $check for all previous nodes previous nodes
    A[i,j] = rbinom(1,1,B[sigma[i], sigma[j]]) #add edge with connection
    A[j,i] = A[i,j] #prob between groups and reverse the edge if needed
  }
}
#generate a graph from adjacency matrix
G = graph_from_adjacency_matrix(A, mode = "undirected")
#set group color for every node
V(G)$color <- ifelse(sigma[V(G)] == 1, "green", "red")
plot(G) #plot the graph
```

The following R code generates a graph or histogram according to the time dynamic stochastic block model:

```
library(igraph) #load graph package
set.seed(20171234) #fix seed for same picture
N = 5 #nr of nodes
k = 2 #nr of communities
#Q = c(rep(c(1,N)))
Q = c(rep(c(1/2),N))  #set q_u for every node to 1/2
#sigma = c(rep(c(1),N/k),rep(c(2),N/k + N %% 2))
sigma = c(rep(c(1,2),N/k), c(rep(1,N %% 2 ))) #set communities
                                               #for every group
B = matrix(data = c(0.66, 0.33,0.33,0.66), k) #community adjacency matrix
```

```
A = matrix(0,N, N) #initialize adjacency matrix
C_old = integer(N) #initialize degree vector old
C_new = integer(N) #initialize degree vector new

f <- function(d,q) {
  1-(1-q)*exp(-d) #define our function f as in (4)
}
for(i in 2:N) { #for every node added
  for(j in 1:(i-1)) { #check for all previous nodes
    P = 1/2*B[sigma[i],sigma[j]]*f(C_old[j],Q[j]) #probability of an edge
                                                   #between node i and j
    A[i,j] = rbinom(1,1,P) #with prob P add an edge between node i and j
    A[j,i] = A[i,j] #reverse the edge
    C_new[i] = C_new[i]+A[i,j] #increase edge degree if an edge
    C_new[j] = C_new[j]+A[i,j] #gets added to node i or node j

  }
  C_old = C_new #copy the new edge degrees for the new node
}
#generate a graph from adjacency matrix
G = graph_from_adjacency_matrix(A, mode = "undirected")
#set group color for every node
V(G)$color <- ifelse(sigma[V(G)] == 1, "green", "red")
plot(G) #plot the graph
#make a histogram of the edge degrees
hist(C_new, freq = FALSE, breaks = seq(-0.5,8.5,1),
main = paste("Edge degrees after", N, "nodes added"), xlab = "degree" )
```

References

[1] E. Abbe, A. S. Bandeira, and G. Hall. Exact Recovery in the Stochastic Block Model. *ArXiv e-prints*, May 2014.

[2] Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

[3] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.*, 17(1):882–938, January 2016.

[4] Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[5] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving Optimal Misclassification Proportion in Stochastic Block Model. *ArXiv e-prints*, May 2015.

[6] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[7] Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010.

[8] Mark S. Handcock, Adrian E. Raftery, and Jeremy M. Tantrum. Model-based clustering for social networks, 2007.

[9] Marc Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, 79(1):135 – 163, 1999.

[10] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137, 1983.

[11] Eric D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[12] C. M. Le, E. Levina, and R. Vershynin. Sparse random graphs: regularization and concentration of the Laplacian. *ArXiv e-prints*, February 2015.

[13] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC '14, pages 694–703, New York, NY, USA, 2014. ACM.

[14] E. Mossel, J. Neeman, and A. Sly. Stochastic Block Models and Reconstruction. *ArXiv e-prints*, February 2012.

[15] M. E. J. Newman and E. A. Leicht. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Science*, 104:9564–9569, June 2007.

[16] Ulrike von Luxburg. A tutorial on spectral clustering. *CoRR*, abs/0711.0189, 2007.

[17] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

[18] Yunpeng Zhao, Elizaveta Levina, and Ji Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292, 2012.

[19] Harrison H. Zhou and Anderson Y. Zhang. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.