

Adaptive weights community detection

Franz Besold, Vladimir Spokoiny

submitted: August 25, 2022

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: franz.besold@wias-berlin.de
vladimir.spokoiny@wias-berlin.de

No. 2951
Berlin 2022



2020 *Mathematics Subject Classification.* 62H30, 62G10.

Key words and phrases. Adaptive weights, community detection, stochastic block model.

Financial support by German Ministry for Education via the Berlin Center for Machine Learning (01IS18037I) is gratefully acknowledged. The research was also supported by the Russian Science Foundation grant No. 18-11-00132.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: <http://www.wias-berlin.de/>

Adaptive weights community detection

Franz Besold, Vladimir Spokoiny

Abstract

Due to the technological progress of the last decades, Community Detection has become a major topic in machine learning. However, there is still a huge gap between practical and theoretical results, as theoretically optimal procedures often lack a feasible implementation and vice versa. This paper aims to close this gap and presents a novel algorithm that is both numerically and statistically efficient. Our procedure uses a test of homogeneity to compute adaptive weights describing local communities. The approach was inspired by the Adaptive Weights Community Detection (AWCD) algorithm by [2]. This algorithm delivered some promising results on artificial and real-life data, but our theoretical analysis reveals its performance to be suboptimal on a stochastic block model. In particular, the involved estimators are biased and the procedure does not work for sparse graphs. We propose significant modifications, addressing both shortcomings and achieving a nearly optimal rate of strong consistency on the stochastic block model. Our theoretical results are illustrated and validated by numerical experiments.

1 Introduction

1.1 Community detection

Community detection has become a major topic in modern statistics with applications in various fields. A very illustrative example is social graphs. Originally discussing relatively small examples such as the famous Zachary's network of karate club members [40] consisting of only 34 vertices, it is nowadays possible to process huge data with millions of vertices such as the Facebook graph [37, 12], the amazon purchasing network [8], mobile phone networks [7] or most recently, data of COVID-19 infections [39]. Moreover, community detection has various applications in biology and bioinformatics [20]. Other popular examples are citation networks [34] and the world wide web [10]. For a more exhaustive list of applications we refer to [28, 14, 17].

The topics of community detection and clustering are clearly related. In fact, after embedding the nodes in a metric space, the problem of community detection can be reduced to the problem of clustering. Nevertheless, the underlying data of a graph is fundamentally different from a point cloud in $\mathbb{R}^n$ and there is no canonical method to embed the vertices in a metric space. Consequently, the theoretical analysis and underlying models differ for each field. This has motivated the development of many methods that deal specifically with community detection.

Similar to the topic of clustering, the task of community detection lacks a clear definition leading to a vast amount of algorithms with varying objectives. Most generally, the goal of community detection can be described as recovering groups of vertices with a similar connection pattern from a given graph. The graph may be weighted or directed. The most common case is that vertices inside each community are more densely connected to each other than to vertices of other communities. The set of groups may be a partition, a set of overlapping communities or a hierarchical structure. One of the earliest and

best-known methods is the Kernighan-Lin algorithm [21] which aims to find a bisection of a graph with a minimal number of edges connecting the two components through successively swapping vertices between the two communities. Applying it iteratively, this method produces a partition of the vertices. The number and size of the communities need to be known. A similar but faster method is the spectral graph bisection [13, 5, 32]. However, the *cut size*, i.e. the number of edges connecting the different components, is not a suitable quality measure of a community structure where the number and size of the communities are unknown. Instead, the so-called *modularity* [27] has become the most popular quality function. It measures the discrepancy between the given number of edges between different communities and the expectation of this term for a similar random graph without the community structure. Originally, it was introduced as a stopping criterion for the algorithm of [16]. However, there have been introduced many algorithms since that directly maximize modularity, starting with the greedy optimization [25, 8, 7] and including for example spectral optimization [26] and simulated annealing [18]. Modularity can be extended to the case of weighted graphs [24], directed graphs [3] and overlapping communities [35]. There are plenty of other methods available besides modularity maximization, such as the *Clique Percolation Method* [29] constructing each community as a union of heavily overlapping cliques and thus allowing for overlapping communities, or most recently, also neural networks [36]. Hierarchical methods can be separated into *agglomerative* and *divisive* methods. Divisive algorithms iteratively split the communities into smaller communities. An example is the algorithm by [16] starting with the original graph and successively removing edges based on a certain measure of *betweenness*, e.g. of the number of shortest paths containing a given edge. Conversely, agglomerative algorithms iterative merge communities into larger communities. An example is the greedy optimization of modularity proposed by [25] starting by removing all edges from the graph and successively adding edges based on the impact on the modularity. Considering the communities as connected components, a hierarchical structure can be obtained. For a comprehensive survey on community detection algorithms, we refer to [14].

A big challenge in Community Detection is the gap between theoretical results and practically relevant algorithms. Many of those lack a rigorous statistical analysis even on the most basic models, while statistically efficient algorithms are often not numerically feasible. For example, the rate-optimal procedure suggested by [41] offers no implementation of polynomial complexity, whereas surprisingly little is known about the performance of the popular Louvain algorithm [7] on the stochastic block model despite some recent progress [9].

1.2 Stochastic block model

The *stochastic block model (SBM)* [19] is the simplest and by far the most studied model for community detection. Under the SBM, edges are generated by independent Bernoulli variables, with the parameters only depending on the corresponding communities for each pair of vertices. The minimax rates are well known for different forms of recovery [1]. In this paper we will only discuss *exact recovery*, also known as *strong consistency*, i.e. we want to recover the entire partition with large probability and without any misclassified vertices.

The SBM is of course not a very realistic model, as the degree distribution inside communities is usually not uniform in applications. It has for example been shown, that the SBM provides a poor fit for the famous karate club network [6]. However, there has been recently a lot of progress on the theoretical foundations of community detection that go beyond the SBM, such as the study of *degree-corrected block models* [15] or *graphons* [22, 30]. And also for practical simulation, there have been various models introduced that are much more realistic than the standard SBM such as the LFR

benchmark [23] or [4] where the degree distribution follows a power law.

Nonetheless, the study of the SBM for a new method can be very useful, as this model already captures a major information theoretic bottleneck for random graphs. We will start discussing a very simple version of the SBM, which is also called symmetric SBM. We consider the size of each community to be deterministic.

Definition 1. Suppose $K \in \mathbb{Z}_{\geq 2}$ is fixed, $n > 0$ and $0 \leq \rho < \theta \leq 1$. By $SBM(n, K, \theta, \rho)$ we denote a random graph with nK vertices that are divided into K communities of size n and whose edges are generated according to independent Bernoulli variables of mean θ inside the communities and mean ρ between different communities.

We will later also discuss the generalization of our results to a more general SBM.

1.3 AWCD revisited

This paper follows up on a proposal by Larisa Adamyan, Kirill Efimov and Vladimir Spokoiny for a novel community detection method based on a hypothesis test of homogeneity for an SBM called *adaptive weights community detection* (AWCD) [2]. The idea of this test originates from a likelihood-ratio test for local homogeneity [31] with applications to image processing. Later, the same test was also used for the problem of adaptive clustering [11]. Unfortunately, no theoretical analysis was provided for the AWCD algorithm. We will provide the first theoretical study in this paper and demonstrate that significant modifications are necessary to achieve a good performance on the SBM. We start by introducing the original algorithm and the idea behind it: Let us consider a general SBM with two disjoint communities C_1^* and C_2^* . The communities may have different sizes. The edges are independent and follow Bernoulli distributions. For $i = 1, 2$, edges inside community C_i^* follow a Bernoulli distribution with parameter θ_i , the remaining edges between the two communities follow a parameter ρ . We consider the null hypothesis

$$H_0: \theta_1 = \theta_2 = \rho \in [0, 1]$$

against the alternative

$$H_1: \theta_1, \theta_2, \rho \in [0, 1].$$

The likelihood-ratio test statistic turns out to be

$$T = N_{11}\mathcal{K}(\tilde{\theta}_{11}, \tilde{\theta}_{1\vee 2}) + N_{22}\mathcal{K}(\tilde{\theta}_{22}, \tilde{\theta}_{1\vee 2}) + N_{12}\mathcal{K}(\tilde{\theta}_{12}, \tilde{\theta}_{1\vee 2}),$$

where

- $\tilde{\theta}_{ii}$ denotes MLE of θ_i under H_1 ,
- $\tilde{\theta}_{12}$ denotes MLE of ρ under H_1 ,
- N_{ij} denotes the sample size of the respective MLEs and
- $\tilde{\theta}_{1\vee 2}$ denotes MLE of ρ under H_0 .

Given data in form of an adjacency matrix $\mathcal{Y}$, the AWCD algorithm applies this test iteratively on *local communities* $\mathcal{C}_i$ that are computed for each node i and represented by a weight matrix W such that $\mathcal{C}_i = \{j : W_{ij} = 1\}$. An exact description is given in Algorithm 1. [2] propose to use the usual graph theoretic neighborhood $\mathcal{Y}$ as a starting guess. Ideally, for some tuning parameter λ , the local community structure W , which is updated at each step of the procedure, converges towards the true underlying community structure W^* . Note that the proposed starting guess violates the setup of the likelihood ratio test w.r.t. the following points:

- Local Communities are *overlapping*.
- Local Communities are *small* compared to true communities.
- Local Communities are of *low precision*, i.e. they contain many false members.

As we will discuss in the following, these violations lead to practical and theoretical limitations of the procedure and motivated us to propose substantial modifications.

In this paper, we address the current gap between theoretical and practical results in Community Detection. Our contributions include:

- A novel procedure based on [11] with the following modifications:
 - Reducing the bias of the involved estimators.
 - Increasing the size of the initial communities.
- Rates of strong consistency of the new algorithm on the SBM: Contrary to [11], the rate is nearly optimal in the most common case where the quotient between two Bernoulli parameters of a symmetric stochastic block model is constant.
- Numerical illustrations and validation for our theoretical results.

The rest of the paper is organized as follows. In section 2 we present our main results. We start in subsection 2.1 by discussing rates of strong consistency for both the original version of the algorithm as well as a modified version that removes bias terms. In subsection 2.2 we will show that these results can be extended to sparse graphs by increasing the size of the initial starting guess for each neighborhood. In particular, this extension of the algorithm achieves a nearly optimal rate of strong consistency for stochastic block models using two Bernoulli parameters having a constant quotient. For simplicity, these results are stated for the symmetric stochastic block model. We discuss the generalization to a more general stochastic block model in subsection 2.3. In the following section 3 we present numerical results illustrating the main results of section 2. All proofs are collected in section 4. In appendix A we discuss further details on the presented rates of consistency.

2 Results

2.1 Improvement of rates via bias correction

For simplicity, we start by considering the very simple model $\text{SBM}(n, K, \rho, \theta)$ introduced in Definition 1. As we are interested in the asymptotics $n \rightarrow \infty$, the parameters θ and ρ may depend on n .

Algorithm 1 AWCD algorithm from [2]

```

1: input: adjacency matrix  $\mathcal{Y}$ , number of iterations  $l_{\max}$  and a test threshold  $\lambda \in \mathbb{R}$ 
2: initialize the weight matrix  $W^{(0)} = \mathcal{Y}$ 
3: for  $l$  from 1 to  $l_{\max}$  do
4:    $S = (W^{(0)}\mathcal{Y}W^{(0)})_{i,j}$ 
5:    $N_{ij} = |\mathcal{C}_i|(|\mathcal{C}_j| - \mathbb{1}(i = j))$ 
6:    $\tilde{\theta}_{ij} = \frac{S_{ij}}{N_{ij}}$ 
7:    $\tilde{\theta}_{i \vee j} = \frac{S_{ii} + 2S_{ij} + S_{jj}}{N_{ii} + 2N_{ij} + N_{jj}}$ 
8:   compute test statistic  $T = N_{11}\mathcal{K}(\tilde{\theta}_{11}, \tilde{\theta}_{1 \vee 2}) + N_{22}\mathcal{K}(\tilde{\theta}_{22}, \tilde{\theta}_{1 \vee 2}) + N_{12}\mathcal{K}(\tilde{\theta}_{12}, \tilde{\theta}_{1 \vee 2})$ 
9:   update weight matrix  $W = (\mathbb{1}(T_{ij} \leq \lambda))_{i,j}$ 
10:  update initial weight matrix  $W^{(0)} = W$ 
11: end for
12: output: matrix of binary weights  $W$ 

```

However, we will not use the index n explicitly to simplify the notation. In contrast to θ and ρ , we consider the number of communities K to be fixed. We denote the corresponding adjacency matrix by $\mathcal{Y}$. Moreover, we will only consider the first step of algorithm 1 for our theoretical analysis, i.e. $l_{\max} = 1$.

The first modification of the algorithm that we propose is related to the definition of S_{ij} . This sum is supposed to count the number of edges connecting the initial communities $\mathcal{C}_i$ and $\mathcal{C}_j$. In particular, after conditioning on $\mathcal{C}_i$ and $\mathcal{C}_j$, we would like to have a sum of independent Bernoulli variables. However, if we also condition on $\mathcal{Y}_{ij} = 1$, the sum contains a deterministic part of size $|\mathcal{C}_i| + |\mathcal{C}_j| - 1$ which is with large probability of order θn :

$$\begin{aligned}
S_{ij} &= \sum_{k, l \notin \{i, j\}} \mathcal{Y}_{il} \mathcal{Y}_{lk} \mathcal{Y}_{kj} + \mathcal{Y}_{ij} \sum_{k \notin \{i, j\}} (\mathcal{Y}_{ik} \mathcal{Y}_{ki} + \mathcal{Y}_{jk} \mathcal{Y}_{kj}) + \mathcal{Y}_{ij}^3 \\
&= \underbrace{\sum_{k, l \notin \{i, j\}} \mathcal{Y}_{il} \mathcal{Y}_{lk} \mathcal{Y}_{kj}}_{\mathcal{O}(\theta^3 n^2)} + \mathbb{1}(\mathcal{Y}_{ij} = 1) \underbrace{\left(|\mathcal{C}_i| + |\mathcal{C}_j| - 1 \right)}_{\mathcal{O}(\theta n)}
\end{aligned}$$

As the sum S_{ij} is of order $\theta^3 n^2$, it can only be informative as long as $\theta \gg n^{-\frac{1}{2}}$. This problem can be easily fixed by redefining

$$\begin{aligned}
S_{ij} &:= (W^{(0)}\mathcal{Y}W^{(0)})_{i,j} - \mathbb{1}(\mathcal{Y}_{ij} = 1) (|\mathcal{C}_i| + |\mathcal{C}_j| - 1) \\
&= \sum_{\mathcal{Y}_{il}=1, \mathcal{Y}_{kj}=1, k, l \notin \{i, j\}} \mathcal{Y}_{kl}.
\end{aligned}$$

Let us call this debiased version of the algorithm AWCD_1 and the original version AWCD_1° . Indeed, this modification improves the rate of $\theta - \rho$ significantly:

Theorem 1. *For both versions of the algorithm we consider the asymptotics $\frac{\theta}{\rho} \equiv C$ as well as the more general condition $\frac{\theta}{\rho} \leq C$ for the symmetric $\text{SBM}(n, K, \theta, \rho)$ with a fixed number of communities K . Consistent exact recovery, i.e.*

$$\mathbb{P}(\forall i, j : W_{ij} = W_{ij}^*) \rightarrow 1,$$

is achieved after the first step as long as $\theta \leq \frac{1}{2}$ and

	$\frac{\theta}{\rho} \equiv C$	$\frac{\theta}{\rho} \leq C$
$AWCD_1^\circ$	$\theta \gg n^{-\frac{1}{2}}$	$\theta - \rho \gg \max\{n^{-\frac{1}{3}}\theta^{\frac{1}{3}}, n^{-\frac{1}{6}}\theta^{\frac{5}{6}}(\log n)^{\frac{1}{6}}\}$
$AWCD_1$	$\theta \gg n^{-\frac{2}{3}}(\log n)^{\frac{1}{3}}$	$\theta - \rho \gg \max\{n^{-\frac{1}{3}}\theta^{\frac{1}{2}}(\log n)^{\frac{1}{6}}, n^{-\frac{1}{6}}\theta^{\frac{5}{6}}(\log n)^{\frac{1}{6}}\}.$

Remark 1. In the above table, we use the notation $x \gg y$ to describe the following: There exists a constant c such that the statement of the theorem holds as long as $cx > y$.

Remark 2. The constant $\frac{1}{2}$ may be replaced by any positive constant smaller than 1.

We have visualized the consistency regime for $\frac{\theta}{\rho} \leq C$ in a $\log_n$ - $\log_n$ plot of the quotient $\frac{\theta-\rho}{\theta}$ against the parameter θ in Figure 1. Note that in the case $\frac{\theta}{\rho} \equiv C$, the minimax rate of consistent exact recovery is $(\log n)n^{-1}$, matching the minimal rate ensuring consistent connectivity of the graph, c.f. [1].

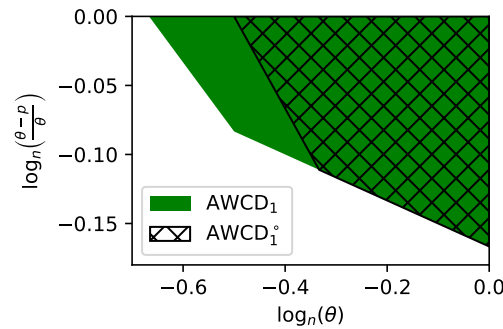


Figure 1: A $\log_n$ - $\log_n$ plot of the regime of strong consistency for $AWCD_1$ and $AWCD_1^\circ$

2.2 Increase of starting neighborhood for sparse graphs

Although the introduced modification improves the rate of strong consistency, the obtained rate is still far from optimal. As previously mentioned, the algorithm violates the likelihood-ratio test setup with respect to the assumptions on the starting guess. In particular, the local communities are relatively small and also contain many members from different communities. The following observation shows, that a better starting guess does indeed improve the rate further.

Observation 1. Suppose instead of $W^{(0)} = \mathcal{Y}$ we are given an improved starting guess for each i in form of a random subset of its true community of the same size $\asymp \theta n$ as before. We assume $\frac{\theta}{\rho} \equiv C$ and $\theta < \frac{1}{2}$. Then $AWCD_1$ achieves consistent exact recovery after the first step as long as

$$\theta \gg n^{-\frac{2}{3}}(\log n)^{\frac{1}{3}}.$$

If the size of the above starting guess is increased to $\asymp \theta^{\frac{1}{2}} n$, then this rate improves to

$$\theta \gg (\log n)^{\frac{1}{2}} n^{-1}.$$

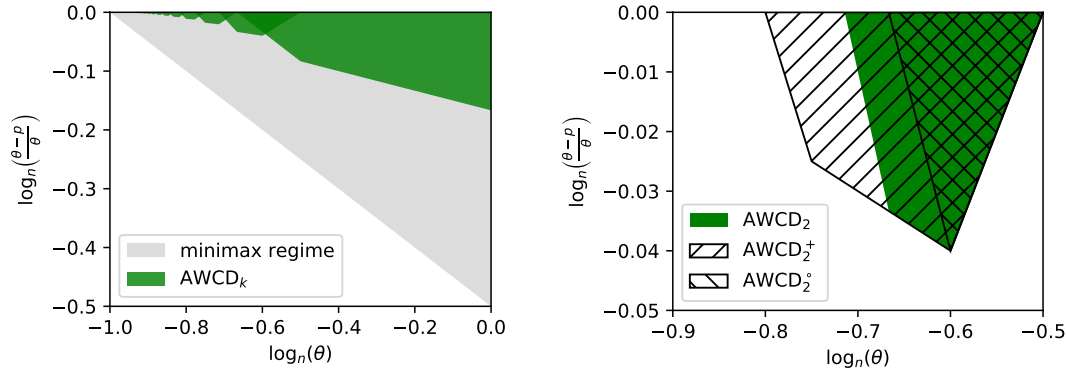


Figure 2: Left: A $\log_n$ - $\log_n$ plot of the strong consistency regime of AWCD_k under a proper choice of the neighborhood parameter k in comparison to the minimax regime [41]. Right: A $\log_n$ - $\log_n$ plot compares the consistency regime of the three versions of the algorithm in case $k = 2$.

Considering that the reconstruction of the community structure from such a good starting guess is trivial, it is surprising, that the rate from Theorem 1 does not improve from increased precision of the starting guess (at least in the case $\frac{\theta}{\rho} \equiv C$). However, we observe the expected improvement after additionally increasing the size of the starting guess. The reason for this phenomenon is that for the computation of each test statistic T_{ij} we only use a part of the available data $\mathcal{Y}$. The larger the starting guess is, the larger part of the data we use. If the size of the starting guess is too small, we cannot expect the algorithm to recover the community structure correctly, even if the starting guess is a subset of the true community.

As we will see in the following, the algorithm also benefits from an increased starting guess if the precision is not increased or even slightly worse. A simple way to increase the size of the starting guess is to take into account members that are connected via a path of minimal path length k . We call this set of members k -neighborhood of i and denote it by $\mathcal{C}_i^k$. Similarly, we can also work with the neighborhood $\mathcal{C}_i^{\leq k}$ of members that are connected via a path of length k or smaller. If the graph is sparse enough, these neighborhoods are almost of the same size and we expect very similar results. To simplify some technical details, we will focus on using the neighborhood $\mathcal{C}_i^k$.

Moreover, an alternative but similar approach to increase the part of the data $\mathcal{Y}$ used for the computation of test statistic T_{ij} is to instead (or additionally) change the way how we count the number of connections between two different local communities: Instead of only counting the edges that directly connect the two local communities, we can also count connections via paths of a certain (maximum) length. This approach should lead to very similar results to those presented in the following.

Let us modify the definitions in the first step of algorithm 1 to

$$S_{ij} := \sum_{v_1 \in \mathcal{C}_i^k, v_2 \in \mathcal{C}_j^k} \mathcal{Y}_{v_1 v_2} \text{ and} \\ N_{ij} := |\mathcal{C}_i^k| (|\mathcal{C}_j^k| - \mathbb{1}(i = j)).$$

We will call this version of the algorithm AWCD_k^o . Moreover, we also consider an analogous bias correction term as before:

$$S_{ij} := \sum_{v_1 \in \mathcal{C}_i^k, v_2 \in \mathcal{C}_j^k} \mathcal{Y}_{v_1 v_2} - \mathbb{1}(i = j) (|\mathcal{C}_i^k| + |\mathcal{C}_j^k|)$$

We denote the corresponding algorithm by AWCD_k . We will show that the bias correction does significantly improve the rates. However, this is only an approximate bias correction. A completely unbiased version AWCD_k^+ can for example be defined via

$$S_{ij} := \sum_{v_1 \in \mathcal{C}_i^k \setminus \mathcal{C}_j^{k-1}, v_2 \in \mathcal{C}_j^k \setminus \mathcal{C}_i^{k-1}} \mathcal{Y}_{v_1 v_2} \text{ and}$$

$$N_{ij} := \sum_{v_1 \in \mathcal{C}_i^k \setminus \mathcal{C}_j^{k-1} \neq v_2 \in \mathcal{C}_j^k \setminus \mathcal{C}_i^{k-1}} 1.$$

leading to even better rates. Unfortunately, we do not know if there exists an efficient implementation. Conversely, the other two versions can be implemented efficiently using matrix multiplications: The starting neighborhood guesses $\mathcal{C}_i^k$ can be represented by the weight matrix

$$W^{(0)} = \left(\mathbb{1} \left(\left(\sum_{l=1}^k \mathcal{Y}^l \right)_{ij} > 0 \right) - \mathbb{1} \left(\left(\sum_{l=1}^{k-1} \mathcal{Y}^l \right)_{ij} > 0 \right) \right)_{i,j}$$

allowing us to compute S as previously as a product of matrices $S = W^{(0)} \mathcal{Y} W^{(0)}$. Computation of N and the bias correction only requires $\mathcal{O}(n^2)$ operations. Thus, the overall complexity is the same as the complexity of the involved (sparse) matrix multiplications. Consequently, AWCD_k is the main algorithm that we propose in this paper. The other two versions are just considered for the sake of comparison.

This extension of the algorithm adds a tuning parameter k to the procedure. However, this does not increase the parameter space much, as k will only take very few different values in practice, e.g. $k \in \{1, 2, 3, 4\}$. Moreover, it may be estimated from the sparsity of the given adjacency matrix.

[2] demonstrated, that starting from the 1-neighborhood, multiple iterations of the algorithm can improve the quality of the result further if the first step already produces a reasonably good estimation of the community structure. This idea of using the output of the previous iteration as a new starting guess can of course also be applied to the algorithms introduced above. However, this approach will not improve the rate any further, see section 3.

In the following theorem, we collect the rates of strong consistency for the cases $\frac{\theta}{\rho} \equiv C$ and $\frac{\theta}{\rho} \leq C$ for each of the three algorithms introduced above. The rates are visualized in a $\log_n$ - $\log_n$ plot in Figure 2.

Theorem 2. Suppose $k \geq 2$. Consistent exact recovery is achieved on $\text{SBM}(n, K, \theta, \rho)$ after the first step as long as

	$\frac{\theta}{\rho} \equiv C$	$\frac{\theta}{\rho} \leq C$
AWCD_k°	$n^{-\frac{k}{k+1}} \ll \theta \ll n^{-\frac{k-1}{k}}$	$\theta - \rho \gg \max\{C_1, C_2^\circ\}$
AWCD_k	$n^{-\frac{2k+1}{2k+3}} (\log n)^{\frac{1}{2k+3}} \ll \theta \ll n^{-\frac{k-1}{k}}$	$\theta - \rho \gg \max\{C_1, C_2\}$
AWCD_k^+	$n^{-\frac{2k}{2k+1}} (\log n)^{\frac{1}{2k+1}} \ll \theta \ll n^{-\frac{k-1}{k}}$	$\theta - \rho \gg \max\{C_1, C_2^+\}$

for

$$\begin{aligned} C_1 &= \theta^{\frac{3k+1}{2k+1}} n^{\frac{k-1}{2k+1}} + \theta^{\frac{4k+1}{4k+2}} n^{-\frac{1}{4k+2}} (\log n)^{\frac{1}{4k+2}}, \\ C_2^\circ &= \theta^{\frac{k}{2k+1}} n^{-\frac{k}{2k+1}}, \\ C_2 &= \theta^{\frac{2k-1}{4k+2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{4k+2}} \text{ and} \\ C_2^+ &= \theta^{\frac{1}{2}} n^{-\frac{k}{2k+1}} (\log n)^{\frac{1}{4k+2}}. \end{aligned}$$

Remark 3. In particular, if k is tuned correctly, the algorithm AWCD_k nearly achieves the optimal rate of strong consistency for the case $\frac{\theta}{\rho} \equiv C$ as the regimes for different k are overlapping.

2.3 General SBM

We suspect that any of the above results may be generalized to an SBM having communities of different sizes and more Bernoulli parameters. Indeed, continuing the study of AWCD_1 , we can extend the results of Theorem 1 to more general stochastic block models.

Proposition 1. We consider the SBM with communities of different sizes $n_1, \dots, n_K$ and only two Bernoulli parameters $\theta > \rho$ as before. We denote the minimal and maximal community size by $n_{\min}$ and $n_{\max}$. Then AWCD_1 achieves consistent exact recovery after the first step (under the asymptotics $n_{\min} \rightarrow \infty$ and $\frac{\theta}{\rho} \leq C$) as long as

$$\theta - \rho \gg \left(\frac{n_{\max}}{n_{\min}} \right)^{\frac{2}{3}} \max \left\{ n_{\min}^{-\frac{1}{3}} \theta^{\frac{1}{2}} (\log n_{\min})^{\frac{1}{6}}, n_{\min}^{-\frac{1}{6}} \theta^{\frac{5}{6}} (\log n_{\min})^{\frac{1}{6}} \right\}$$

and $\theta < \frac{1}{2}$ (or any constant smaller than 1).

Remark 4. The lower bound no longer simplifies to a single term in the case $\frac{\theta}{\rho} \equiv C$ but rather to

$$\theta \gg \max \left\{ \left(\frac{n_{\max}}{n_{\min}} \right)^{\frac{4}{3}} n_{\min}^{-\frac{2}{3}} (\log n_{\min})^{\frac{1}{3}}, \left(\frac{n_{\max}}{n_{\min}} \right)^4 n_{\min}^{-1} \log n_{\min} \right\}.$$

Only for $n_{\max}^{10} \lesssim n_{\min}^{11}$ does the condition further simplify to

$$\theta \gg \left(\frac{n_{\max}}{n_{\min}} \right)^{\frac{4}{3}} n_{\min}^{-\frac{2}{3}} \log n_{\min},$$

similarly to the rate of Theorem 1.

Proposition 2. We consider a stochastic block model with two blocks of block sizes n_1 and n_2 with parameters $\theta_1, \theta_2 > \rho$ under the asymptotics $n_{\min} \rightarrow \infty$. Then AWCD_1 achieves consistent exact recovery after the first step as long as

$$\theta_{\min} - \rho \gg \left(\frac{\theta_{\max}}{\theta_{\min}} \right)^{\frac{7}{3}} \left(\frac{n_{\max}}{n_{\min}} \right)^2 \max \left\{ n_{\min}^{-\frac{1}{3}} \theta_{\max}^{\frac{1}{2}} (\log n_{\min})^{\frac{1}{6}}, n_{\min}^{-\frac{1}{6}} \theta_{\max}^{\frac{5}{6}} (\log n_{\min})^{\frac{1}{6}} \right\}$$

and $\theta_{\max} < \frac{1}{2}$ (or any constant smaller than 1).

Remark 5. Note that we do in general allow for $\frac{n_{\max}}{n_{\min}} \rightarrow \infty$ or $\frac{\theta_{\max}}{\theta_{\min}} \rightarrow \infty$. If $\frac{n_{\max}}{n_{\min}}$ and $\frac{\theta_{\max}}{\theta_{\min}}$ are bounded by a constant M , then the rates in Proposition 1 and 2 are exactly the same as in Theorem 1. However, the respective implicit constant factor in the lower bound depends on M .

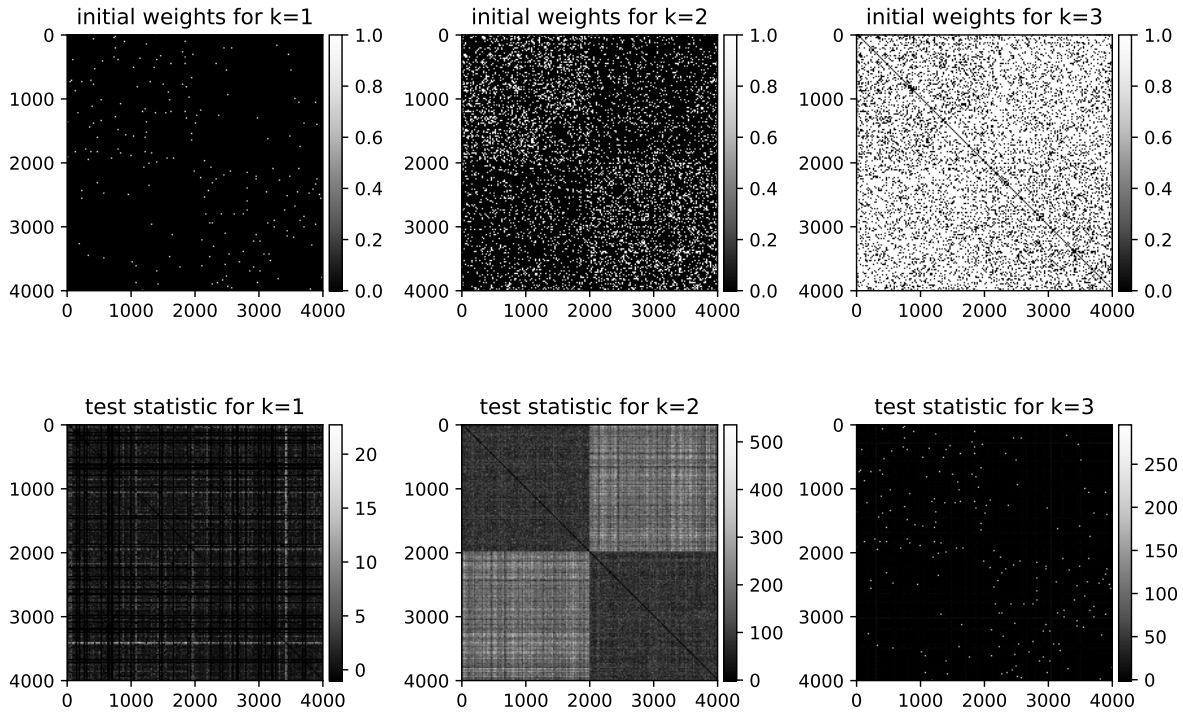


Figure 3: We consider a realization of $\text{SBM}(2000, 2, 0.01, 0.0025)$. Top: Adjacency matrix of the starting guess $W^{(0)}$ of the community structure from AWCD_k for $k = 1, 2, 3$. Bottom: The corresponding test statistics.

3 Experiments

[2] already demonstrated a state-of-the-art performance of the original algorithm on the LFR benchmark [23] for a total sample size of $n = 1000$ as well as on smaller real-life examples such as the famous Zachary’s karate club network [40]. Moreover, it has been shown that optimizing modularity is a reasonable method to choose the tuning parameter λ . In this section, we focus on validating our theoretical results and consider only the simple symmetric stochastic block model $\text{SBM}(n, 2, \theta, \rho)$ consisting of two communities of identical size n and having two Bernoulli parameters $\theta > \rho$. We want to study two questions:

- 1 How does the performance of the procedure depend on the starting neighborhood via the parameter k
- 2 Can we observe the stated rates of consistency?

We do not study the effect of the bias correction term. This aspect of the algorithm is important for strong consistency. However, in terms of the accuracy of the final weight matrix, also known as *Rand index* [33], the effect is rather small: If the matrix is relatively dense, then the bias term is also relatively small, whereas the bias correction only applies to a relatively small fraction of the test statistics T_{ij} , if the matrix is sparse.

We start by studying one realization of the SBM as described above with the parameters $n = 2000$, $\theta = 0.01$ and $\rho = 0.0025$. In Figure 3 we see the original adjacency matrix, the corresponding 2- and 3-neighborhood adjacency matrix as well as the test statistics of the algorithm AWCD_k for $k = 1, 2, 3$.

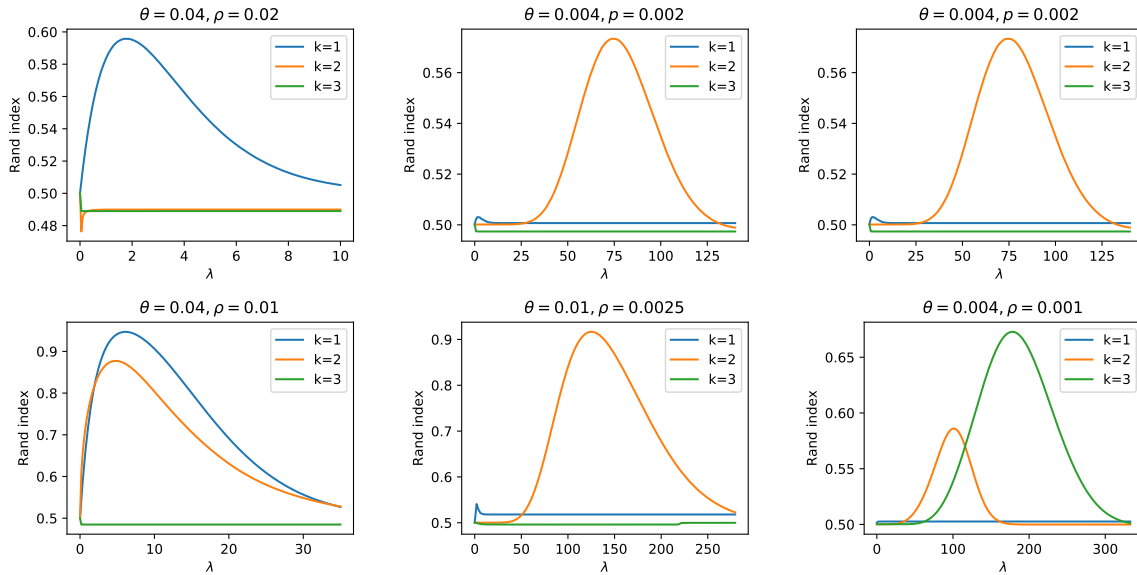


Figure 4: Rand index achieved by AWCD_k on $\text{SBM}(2000, 2, \theta, \rho)$ for different algorithm parameters (k, λ) and for several Bernoulli parameter combinations (θ, ρ)

We can see that the test statistic is only informative for $k = 2$: For $k = 1$, each sum S_{ij} consists only of a few summands, so the test statistic cannot be reliable, whereas, for $k = 3$, each starting neighborhood contains almost all members of the network. This is confirmed by a plot of the Rand index of the final weight matrices corresponding to these test matrices and different tuning parameters λ in Figure 4 (bottom, middle): Only for $k = 2$ does the algorithm correctly identify the community structure. [2] have demonstrated in the case $k = 1$ that multiple iterations of the algorithm can improve the final weight matrix further if enough information on the community structure is recovered after the first step. However, in this example, we do not benefit from these additional iterations, as the data is too sparse and the weight matrix after the first step is not informative enough. This is demonstrated in Figure 5.

We repeated the above experiment for several other parameter combinations (θ, ρ) and plotted the resulting Rand indices in Figure 4. We can see that the performance of the algorithm is stable with respect to the tuning parameter λ . Moreover, for any parameter combination of the SBM, we find parameters k and λ which recover the community structure with a non-negligible accuracy. As expected by Theorem 2, increasing the sparsity of the data significantly enough requires also increasing the parameter k . Note that the overall sparsity depends on both parameters θ and ρ : In the case of $\theta = 0.004$ we see on the right-hand side of Figure 4 that the optimal parameter k depends on ρ . Unsurprisingly, we also observe that for a given θ , an increase of the quotient $\frac{\theta}{\rho}$ allows a more accurate recovery of the community structure.

Lastly, we discuss the second question raised above: Can we observe the stated rates of consistency in numerical experiments? To minimize the computation time, we will focus on the case $k = 1$. We have repeated the above experiment for many different parameter combinations (n, θ) with a fixed quotient $\frac{\theta}{\rho} \equiv 4$. For each parameter combination, we computed the maximum Rand index that can be achieved by optimizing the tuning parameter λ . We repeated this experiment ten times for each parameter combination with $n \leq 2000$ and averaged the resulting maximum Rand index. The results are shown in Figure 6. According to Theorem 1, AWCD_1 achieves strong consistency at a rate of $n^{-\frac{2}{3}}$ up to logarithmic factors. Indeed, it seems that the function yielding the minimum θ necessary to achieve a certain level of accuracy for a given community size n can be reasonably well approximated

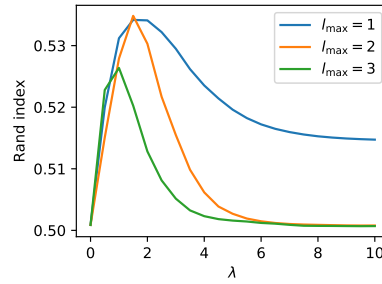


Figure 5: Rand index achieved after the first three iterations of algorithm 1 for a realization of $\text{SBM}(2000, 2, 0.01, 0.0025)$

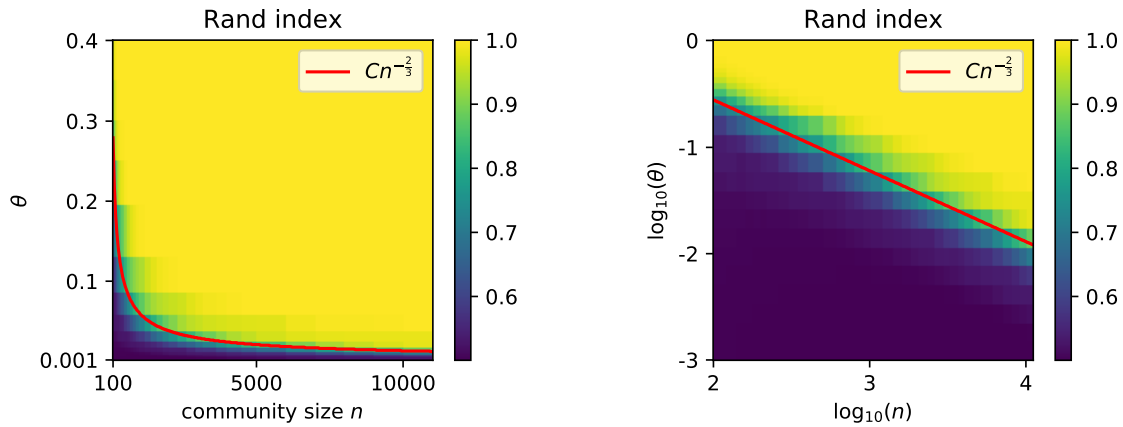


Figure 6: Two plots visualize the Rand index achieved by AWCD_1 for different SBM parameters (n, θ) with $\frac{\theta}{\rho} \equiv 4$ using the optimal tuning parameter λ . The right plot only differs in the logarithmic scale on both axes.

by a function of the form $Cn^{-\frac{2}{3}}$.

4 Proofs

Lemma 1. For $X \sim \text{Binom}(m, q)$ and $a > 0$ we have

$$\mathbb{P}(|X - \mathbb{E}X| \geq am) \leq 2 \exp \left(-\frac{a^2 m}{2 \left(q + \frac{a}{3} \right)} \right).$$

Proof. This is a special case of Bernstein's inequality for bounded variables [38, Theorem 2.8.4]. $\square$

Lemma 2. Suppose X_i are independent Bernoulli variables of mean $\theta_i \leq \theta$ and $a > 0$. Then for $c_i \in \{1, 2\}$ and $m := \sum_{i=1}^N c_i$ we have

$$\mathbb{P} \left(\left| \sum_{i=1}^N c_i X_i - \mathbb{E} c_i X_i \right| \geq am \right) \leq 2 \exp \left(-\frac{a^2 m}{6\theta + \frac{4a}{3}} \right).$$

Proof. This is a special case of Bernstein's inequality for bounded variables [38, Theorem 2.8.4]. $\square$

Proof of Theorem 1. Let us first consider the AWC^D algorithm. Suppose $\epsilon > 0$ and $\alpha \in \{1, 2, \dots, K\}$. We will denote the corresponding true community by $\mathcal{C}_\alpha^*$. According to Lemma 1 we have

$$|\mathcal{C}_i \cap \mathcal{C}_\alpha^*| = \theta_i^\alpha n + \mathcal{O}(\epsilon n + \theta) \quad (1)$$

for

$$\theta_i^\alpha = \begin{cases} \theta & , i \in \mathcal{C}_\alpha^* \\ \rho & , i \notin \mathcal{C}_\alpha^* \end{cases}$$

on an event of probability at least

$$1 - 2 \exp\left(-\frac{\epsilon^2(n-2)}{2\theta + \epsilon}\right).$$

Similarly, we conclude for $i \neq j$

$$|\mathcal{C}_i \cap \mathcal{C}_j| = \mathcal{O}(\theta^2 n + \epsilon n) \quad (2)$$

on an event of the probability of at least the same lower bound. The above identities are still valid if we replace $\mathcal{C}_i$ by $\mathcal{C}_i^j := \mathcal{C}_i \setminus \{j\}$. In the following, we will restrict to an event where (1) and (2) are satisfied for any $i \neq j$. By union bound, this event is of asymptotic probability 1 as long as

$$\min\left\{\frac{\epsilon^2 n}{\theta}, \epsilon n\right\} \gg \log n. \quad (3)$$

Note that according to our assumptions $\theta n \rightarrow \infty$. We conclude from (1) and (2) for $\epsilon < \theta$

$$\begin{aligned} \mathbb{E}[S_{ij} | \mathcal{C}_i^j, \mathcal{C}_j^i] &= \theta \sum_{\alpha} |\mathcal{C}_i^j \cap \mathcal{C}_\alpha^*| |\mathcal{C}_j^i \cap \mathcal{C}_\alpha^*| + p \sum_{\alpha \neq \beta} |\mathcal{C}_i^j \cap \mathcal{C}_\alpha^*| |\mathcal{C}_j^i \cap \mathcal{C}_\beta^*| + \mathcal{O}(\theta |\mathcal{C}_i \cap \mathcal{C}_j|) \\ &= \theta \sum_{\alpha} \theta_i^\alpha \theta_j^\alpha n^2 + \rho \sum_{\alpha \neq \beta} \theta_i^\alpha \theta_j^\beta n^2 + \mathcal{O}(\theta^3 n + \theta^2 \epsilon n^2) \\ &= \begin{cases} a + \mathcal{O}(\theta^3 n + \theta^2 \epsilon n^2) & , \exists \alpha : i, j \in \mathcal{C}_\alpha^* \\ c + \mathcal{O}(\theta^3 n + \theta^2 \epsilon n^2) & , \text{otherwise} \end{cases} \end{aligned} \quad (4)$$

for

$$\begin{aligned} a &:= n^2 [\theta^3 + 3(K-1)\theta\rho^2 + (K-1)(K-2)\rho^3] \text{ and} \\ c &:= n^2 [3\theta^2\rho + 3(K-2)\theta\rho^2 + ((K-1)(K-2) + 1)\rho^3]. \end{aligned}$$

Note that we have to add the term $\mathcal{O}(\theta |\mathcal{C}_i \cap \mathcal{C}_j|)$ because for any point l in the overlap, we have $\mathcal{Y}_{il}\mathcal{Y}_{ll}\mathcal{Y}_{lk} = 0$. To simplify the notation, we will drop the superindex in $\mathcal{C}_i^j$ from now on. Similarly, we have

$$\begin{aligned} \mathbb{E}[S_{ii} | \mathcal{C}_i] &= \theta \sum_{\alpha} |\mathcal{C}_i \cap \mathcal{C}_\alpha^*|^2 + p \sum_{\alpha \neq \beta} |\mathcal{C}_i \cap \mathcal{C}_\alpha^*| |\mathcal{C}_i \cap \mathcal{C}_\beta^*| + \mathcal{O}(\theta |\mathcal{C}_i|) \\ &= \theta \sum_{\alpha} (\theta_i^\alpha)^2 n^2 + p \sum_{\alpha \neq \beta} \theta_i^\alpha \theta_i^\beta n^2 + \mathcal{O}(\theta^2 n + \theta^2 \epsilon n^2) \\ &= a + \mathcal{O}(\theta^2 n + \theta^2 \epsilon n^2) \end{aligned}$$

as well as

$$\begin{aligned}\mathbb{E}[N_{ij}|\mathcal{C}_i, \mathcal{C}_j] &= N_{ij} \\ &= \left(\sum_{\alpha} |\mathcal{C}_i \cap \mathcal{C}_{\alpha}^*| \right) \left(\sum_{\alpha} |\mathcal{C}_j \cap \mathcal{C}_{\alpha}^*| \right) \\ &= d + \mathcal{O}(\theta^2 n + \epsilon \theta n^2)\end{aligned}$$

for

$$d := n^2 [\theta^2 + 2(K-1)\theta\rho + (K-1)^2\rho^2]$$

and

$$\mathbb{E}[N_{ii}|\mathcal{C}_i] = d + \mathcal{O}(\theta n + \epsilon \theta n^2).$$

For

$$R := \theta n + \epsilon \theta n^2 \tag{5}$$

we conclude

$$\mathbb{E}[\tilde{\theta}_{ii}|\mathcal{C}_i] = \frac{a + \mathcal{O}(\theta R)}{d + \mathcal{O}(R)}.$$

For

$$\delta := \max \left\{ \frac{1}{\theta n}, \frac{\epsilon}{\theta} \right\} \tag{6}$$

under the assumption $\delta \ll 1$ we can rewrite

$$\mathbb{E}[\tilde{\theta}_{ii}|\mathcal{C}_i] = \frac{a}{d} (1 + \mathcal{O}(\delta))$$

and analogously

$$\begin{aligned}\mathbb{E}[\tilde{\theta}_{ij}|\mathcal{C}_i, \mathcal{C}_j] &= \begin{cases} \frac{a}{d} (1 + \mathcal{O}(\delta)) & , \exists \alpha : i, j \in \mathcal{C}_{\alpha}^* \\ \frac{c}{d} (1 + \mathcal{O}(\delta)) & , \text{otherwise} \end{cases} \\ \text{and } \mathbb{E}[\tilde{\theta}_{i \vee j}|\mathcal{C}_i, \mathcal{C}_j] &= \begin{cases} \frac{a}{d} (1 + \mathcal{O}(\delta)) & , \exists \alpha : i, j \in \mathcal{C}_{\alpha}^* \\ \frac{a+c}{2d} (1 + \mathcal{O}(\delta)) & , \text{otherwise.} \end{cases}\end{aligned}$$

In the case of i and j belonging to different communities we conclude for large enough n

$$\begin{aligned}\mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{ij}|\mathcal{C}_i, \mathcal{C}_j] &= \frac{a+c}{2d} (1 + \mathcal{O}(\delta)) - \frac{c}{d} (1 + \mathcal{O}(\delta)) \\ &= \frac{a-c}{2d} + \mathcal{O}(\delta\theta) \\ &\propto \frac{(\theta - \rho)^3}{\theta^2} + \mathcal{O}(\delta\theta)\end{aligned} \tag{7}$$

and otherwise

$$\begin{aligned}\mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{ij}|\mathcal{C}_i, \mathcal{C}_j] &= \frac{a}{d} (1 + \mathcal{O}(\delta)) - \frac{a}{d} (1 + \mathcal{O}(\delta)) \\ &= \mathcal{O}(\delta\theta).\end{aligned} \tag{8}$$

Recall that with large probability, $N_{ij} = d(1 + \mathcal{O}(\delta))$. From Lemma 2 we deduce that on an event of large probability,

$$\log \left(\frac{1}{2} \mathbb{P} \left(|S_{ij} - \mathbb{E}[S_{ij} | \mathcal{C}_i^j, \mathcal{C}_j^i]| \geq \frac{\rho d}{3} \middle| \mathcal{C}_i^j, \mathcal{C}_j^i \right) \right) \lesssim -\rho d \\ \lesssim -\theta^3 n^2.$$

As the RHS does not depend on $\mathcal{C}_i^j$ or $\mathcal{C}_j^i$, the bound is valid for the unconditional probability as well. From our assumption $\theta^3 n^2 \gg \log n$ and $\theta < \frac{1}{2}$ we deduce that with large probability for some $x \in \{a, c\} \subseteq [\rho d, \theta d]$

$$\tilde{\theta}_{ij} \in \left[\frac{x - \frac{\rho}{3}}{d} (1 + \mathcal{O}(\delta)), \frac{x + \frac{\rho}{3}}{d} (1 + \mathcal{O}(\delta)) \right] \\ \subseteq \left[\frac{\rho}{2}, \theta + \frac{\rho}{2} \right]. \quad (9)$$

Similarly, with large probability

$$\tilde{\theta}_{i \vee j} \in \left[\frac{\rho}{2}, \theta + \frac{\rho}{2} \right].$$

In particular, the Fisher information of a Bernoulli variable with a mean parameter bounded as above is up to bounded constants given by θ^{-1} .

Analogously, the condition $\theta \gg (\log n)^{\frac{1}{3}} n^{-\frac{2}{3}}$ and Lemma 2 imply that with large probability

$$|S_{ij} - \mathbb{E}[S_{ij} | \mathcal{C}_i, \mathcal{C}_j]| \lesssim (\log n)^{\frac{1}{2}} \theta^{\frac{3}{2}} n$$

and

$$|S_{i \vee j} - \mathbb{E}[S_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j]| \lesssim (\log n)^{\frac{1}{2}} \theta^{\frac{3}{2}} n.$$

In particular,

$$|\tilde{\theta}_{ij} - \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j]| \lesssim (\log n)^{\frac{1}{2}} \theta^{-\frac{1}{2}} n^{-1} \quad (10)$$

and

$$|\tilde{\theta}_{i \vee j} - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j]| \lesssim (\log n)^{\frac{1}{2}} \theta^{-\frac{1}{2}} n^{-1}.$$

Using the quadratic Taylor expansion of the Kullback-Leibler and (7), we conclude in the case where i and j belong to different communities

$$\theta^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\tilde{\theta}_{ij}, \tilde{\theta}_{i \vee j}) \gtrsim |\tilde{\theta}_{ij} - \tilde{\theta}_{i \vee j}| \\ \geq \left| \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j] - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j] \right| - \left| \tilde{\theta}_{ij} - \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j] \right| - \left| \tilde{\theta}_{i \vee j} - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j] \right| \\ \gtrsim (\theta - \rho)^3 \theta^{-2} + \mathcal{O} \left(\delta \theta + (\log n)^{\frac{1}{2}} \theta^{-\frac{1}{2}} n^{-1} \right), \quad (11)$$

whereas in the other case we conclude analogously from (8)

$$\theta^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\theta_{ij}, \theta_{i \vee j}) \lesssim \delta \theta + (\log n)^{\frac{1}{2}} \theta^{-\frac{1}{2}} n^{-1}. \quad (12)$$

The same inequalities hold for $\mathcal{K}^{\frac{1}{2}}(\tilde{\theta}_{ii}, \tilde{\theta}_{i \vee j})$ and $\mathcal{K}^{\frac{1}{2}}(\tilde{\theta}_{jj}, \tilde{\theta}_{i \vee j})$. To ensure that the Kullback-Leibler divergence in the second case (12) is significantly smaller than in the first case (11), it will suffice to check

$$\delta \theta \ll \frac{(\theta - \rho)^3}{\theta^2} \\ \Leftrightarrow \max \left\{ \frac{1}{\theta n}, \frac{\epsilon}{\theta} \right\} \ll \left(\frac{\theta - \rho}{\theta} \right)^3 \quad (13)$$

and

$$\sqrt{\frac{\log n}{\theta n^2}} \ll \frac{(\theta - \rho)^3}{\theta^2}. \quad (14)$$

With $\cdot \ll \cdot$ we denote that inequality is satisfied for small enough yet not specified constant. These are all conditions in the proof that remain to be checked except the large probability assumption

$$\min \left\{ \frac{\epsilon^2 n}{\theta \log n}, \frac{\epsilon n}{\log n} \right\} \gg 1 \quad (15)$$

given in (3). An $\epsilon > 0$ satisfying (13) and (15) exists as long as

$$\frac{\log n}{n} + \sqrt{\frac{\theta \log n}{n}} \ll \frac{(\theta - \rho)^3}{\theta^2}, \quad (16)$$

whereas the ϵ -independent part of conditions (13) and (14) can be summarized by

$$\frac{1}{n} + \sqrt{\frac{\log n}{\theta n^2}} \ll \frac{(\theta - \rho)^3}{\theta^2}. \quad (17)$$

By simple calculus, we can verify that our assumption

$$\theta - \rho \gg \max \{ n^{-\frac{1}{3}} \theta^{\frac{1}{2}} (\log n)^{\frac{1}{6}}, n^{-\frac{1}{6}} \theta^{\frac{5}{6}} (\log n)^{\frac{1}{6}} \}$$

implies conditions (16) and (17). In view of the fact that test is scaled with $N_{ij} = d(1 + \mathcal{O}(\delta))$ and $\delta \ll 1$ this also ensures consistency of the test (provided a proper threshold is given).

Next, let us consider the original AWCD algorithm. The proof deviates because we need to add an additional summand $\mathcal{O}(|\mathcal{C}_i| + |\mathcal{C}_j|) = \mathcal{O}(\theta n)$ to the conditional expectation of S_{ij} in (4). Therefore we also need to modify the definitions $R := n + \epsilon \theta n^2$ and $\delta := \max \{ \frac{1}{\theta^2 n}, \frac{\epsilon}{n} \}$, cf. (5) and (6). Otherwise, we can follow the above proof, although S_{ij} is, after conditioning on $\mathcal{C}_i$ and $\mathcal{C}_j$, not any longer a sum of independent Bernoulli variables. However, it can be split into a deterministic part and the same sum of independent Bernoulli variables as above. By bounding the deterministic part, we can still establish inequality (9). Similarly to (16) and (17) we end up with the sufficient condition

$$\frac{1}{\theta n} + \sqrt{\frac{\theta \log n}{n}} + \sqrt{\frac{\log n}{\theta n^2}} \ll \frac{(\theta - \rho)^3}{\theta^2}$$

which is satisfied by our condition

$$\theta - \rho \gg \max \{ n^{-\frac{1}{3}} \theta^{\frac{1}{3}}, n^{-\frac{1}{6}} \theta^{\frac{5}{6}} (\log n)^{\frac{1}{6}} \}.$$

□

Proof of Observation 1. We can follow the proof of theorem 1 by modifying (1) to

$$|\mathcal{C}_i \cap \mathcal{C}_\alpha^*| = \mathbb{1}(i \in \mathcal{C}_\alpha) \theta n$$

in the first case and in the second case to

$$|\mathcal{C}_i \cap \mathcal{C}_\alpha^*| = \mathbb{1}(i \in \mathcal{C}_\alpha) \theta^{\frac{1}{2}} n.$$

In both cases, we end up with

$$\mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j] \propto \theta - \rho + \mathcal{O}(n^{-1}).$$

However, using the larger starting guess, the stochastic bound (10) can be improved to

$$|\tilde{\theta}_{ij} - \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j]| \lesssim (\log n)^{\frac{1}{2}} n^{-1}.$$

Finally, we end up with the sufficient condition

$$\theta - \rho \gg (\log n)^{\frac{1}{2}} \theta^{-\frac{1}{2}} n^{-1}$$

in the case of the smaller starting guess and in the other case with

$$\theta - \rho \gg (\log n)^{\frac{1}{2}} n^{-1}.$$

□

Lemma 3. *Suppose K, k and $\epsilon > 0$ are fixed. We assume*

- $\min \left\{ \frac{\epsilon^2 n}{\theta}, \epsilon n \right\} \gg \log n,$
- $\epsilon \ll \theta,$
- $\theta n \gg 1,$ and
- $\theta^k n^{k-1} \ll 1.$

Then with large probability, we have

$$|\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| = a_k^{i,\alpha} n^k \left(1 + \mathcal{O} \left(\frac{\epsilon}{\rho} + \theta^k n^{k-1} \right) \right),$$

where we define recursively

$$\begin{aligned} a_1 &= \theta, \\ b_1 &= \rho, \\ a_k &= \theta a_{k-1} + (K-1) \rho b_{k-1}, \\ b_k &= \rho a_{k-1} + \theta b_{k-1} + (K-2) \rho b_{k-1} \text{ and} \\ a_k^{i,\alpha} &= \begin{cases} a_k & , i \in \mathcal{C}_\alpha^* \\ b_k & , i \notin \mathcal{C}_\alpha^*. \end{cases} \end{aligned}$$

The upper bound can be improved to

$$|\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| \leq a_k^{i,\alpha} n^k \left(1 + \mathcal{O} \left(\frac{\epsilon}{\rho} \right) \right).$$

Remark 6. *As $|\mathcal{C}_i^{k-1} \cap \mathcal{C}_\alpha^*|$ is significantly smaller than $|\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*|$, exactly the same concentration result holds for $|\mathcal{C}_i^{\leq k} \cap \mathcal{C}_\alpha^*|$. The increase corresponds to an additional factor $(1 + \mathcal{O}(\frac{1}{\theta n}))$ - however in view of $\frac{\epsilon}{\rho} \gg \frac{1}{\theta n}$ this factor can be omitted.*

Remark 7. By induction, we can compute an explicit formula for the difference

$$\begin{aligned} a_k - b_k &= (\theta - \rho)(a_{k-1} - b_{k-1}) \\ &= (\theta - \rho)^k. \end{aligned}$$

Proof of Lemma 3. Applying Lemma 1 and the union bound, a simple induction yields the following upper bound

$$|\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| \leq a_k^{i,\alpha} n^k \left(1 + \frac{\epsilon}{\rho}\right)^k \quad (18)$$

as long as

$$\min \left\{ \frac{\epsilon^2 n}{\theta}, \epsilon n \right\} \geq c(k) \log n. \quad (19)$$

The lower bound is not as simple because of the potential overlap of the involved 1-neighborhoods. First of all, our assumptions $\epsilon \ll \theta$, $\theta n \gg 1$ as well as $\theta^k n^{k-1} \ll 1$ are designed to ensure the upper bound

$$\begin{aligned} |\mathcal{C}_i^{\leq k} \cap \mathcal{C}_\alpha^*| &\stackrel{(18)}{\leq} \sum_{l=1}^k K^{l-1} \theta^l n^l \left(1 + \frac{\epsilon}{\rho}\right)^l \\ &\leq \frac{n}{2}. \end{aligned}$$

This allows us to apply Lemma 1 to the set of members not contained in $\mathcal{C}_{i,j}^{\leq k}$ while the large probability of the bound is still ensured by condition (19). The obtained lower bounds on the size of the 1-neighborhoods around points in $\mathcal{C}_i^{k-1}$ take into account the potential overlaps and can thus be summarized. To be precise, let us condition on the event that the $(k-1)$ -neighborhood is equal to $\{v_1, v_2, \dots, v_m\}$. Let us fix $v \in \mathcal{C}^{k-1}$ and denote by $\mathcal{S}(\mathcal{C}_i^{\leq l})$ the set of edges contained by a path of length at most l starting from i . We condition additionally on $\mathcal{S}(\mathcal{C}_i^{\leq k-1}) \cup \left(\bigcup_{v_i \neq v} \mathcal{S}(\mathcal{C}_{v_i})\right)$. Now we apply Lemma 1 to get a lower bound of the number of members in $\mathcal{C}_\alpha^*$ that are connected to v and not contained in $\mathcal{C}_i^{\leq k-1} \cup \left(\bigcup_{v_i \neq v} \mathcal{C}_{v_i}\right)$. This is in fact a lower bound on the number of members in $\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*$ that are only connected to i via a k -path containing v . This lower bound is given by

$$\begin{aligned} &(\theta_v^\alpha - \epsilon) \left| \mathcal{C}_\alpha^* \setminus \left(\{i\} \cup \mathcal{C}_i^{\leq k} \cap \mathcal{C}_\alpha^* \right) \right| \\ &\geq (\theta_v^\alpha - \epsilon) \left(n - 1 - a_1^{i,\alpha} n^1 \left(1 + \frac{\epsilon}{\rho}\right)^1 - \dots - a_k^{i,\alpha} n^k \left(1 + \frac{\epsilon}{\rho}\right)^k \right) \\ &= (\theta_v^\alpha - \epsilon) \left(n - 1 - \sum_{l=1}^k a_l^{i,\alpha} n^l \left(1 + \frac{\epsilon}{\rho}\right)^l \right) \\ &\geq (\theta_v^\alpha - \epsilon) \left(n - 1 - \sum_{l=1}^k K^{l-1} \theta^l n^l \left(1 + \frac{\epsilon}{\rho}\right)^l \right) \\ &= (\theta_v^\alpha - \epsilon) \left(n - \frac{1}{K} \frac{K^{k+1} \theta^{k+1} n^{k+1} \left(1 + \frac{\epsilon}{\rho}\right)^{k+1} - 1}{K \theta n \left(1 + \frac{\epsilon}{\rho}\right) - 1} \right) \\ &= (\theta_v^\alpha - \epsilon) n (1 - \delta_k) \end{aligned}$$

for

$$\delta_k = \frac{1}{n} \left(\frac{1}{K} \frac{K^{k+1} \theta^{k+1} n^{k+1} \left(1 + \frac{\epsilon}{\rho}\right)^{k+1} - 1}{K \theta n \left(1 + \frac{\epsilon}{\rho}\right) - 1} \right).$$

This bound is of course valid in the unconditional form as well. By induction, we conclude

$$\begin{aligned} |\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| &\geq \sum_{\beta} |\mathcal{C}_i^{k-1} \cap \mathcal{C}_\beta^*| [\theta \mathbb{1}(\alpha = \beta) + \rho \mathbb{1}(\alpha \neq \beta) - \epsilon] n(1 - \delta_k) \\ &\geq \alpha_i^k n^k \left(1 - \frac{\epsilon}{\rho}\right)^k \prod_{l=1}^k (1 - \delta_l). \end{aligned}$$

So we get in view of $\delta_{k-1} < \delta_k$

$$|\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| \in \left[a_k^{i,\alpha} n^k \left(1 - \frac{\epsilon}{\rho}\right)^k (1 - \delta_k)^k, a_k^{i,\alpha} n^k \left(1 + \frac{\epsilon}{\rho}\right)^k \right].$$

Note that $\delta_k = \mathcal{O}(\theta^k n^{k-1})$, so our assumptions ensure $\delta_k \ll 1$. Also taking into account $\epsilon \ll \theta$, we conclude

$$|\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| = a_k^{i,\alpha} n^k \left(1 + \mathcal{O}\left(\frac{\epsilon}{\rho} + \theta^k n^{k-1}\right)\right).$$

□

Lemma 4. *Under the same assumptions as in Lemma 3 it holds on an event of large probability*

$$|\mathcal{C}_i^k \cap \mathcal{C}_j^k \cap \mathcal{C}_\alpha^*| \lesssim \theta^{2k} n^{2k-1} + \epsilon \theta^{k-1} n^k.$$

If we further assume that there exists no path of length at most $k - l$ between i and j for some $1 \leq l < k$, we have

$$|\mathcal{C}_i^l \cap \mathcal{C}_j^k \cap \mathcal{C}_\alpha^*| \lesssim \theta^{l+k} n^{l+k-1} + \epsilon \theta^{l-1} n^l.$$

Proof. According to Lemma 3

$$|\mathcal{C}_i^{\leq k} \cap \mathcal{C}_\alpha^*| \leq a_k^{i,\alpha} n^k \left(1 + \mathcal{O}\left(\frac{\epsilon}{\theta} + \theta^k n^{k-1}\right)\right). \quad (20)$$

From our assumptions, we conclude further for any β

$$|\mathcal{C}_i^{\leq k} \cap \mathcal{C}_\beta^*| < 2K^{k-1} \theta^k n^k.$$

Let us introduce the notation

$$m := n - 2K^{k-1} \theta^k n^k.$$

Note that our assumptions imply $m > \frac{n}{2}$. After conditioning on $\mathcal{C}_j^1$, we can apply Lemma 1 to get a lower bound on the number of edges between i and the remaining points in $\mathcal{C}_\beta^*$:

$$\begin{aligned} |\mathcal{C}_\beta^* \cap \mathcal{C}_i^1 \setminus \mathcal{C}_j^1| &\geq a_1^{i,\beta} m - \mathcal{O}(\epsilon n) \\ &= a_1^{i,\beta} m \left(1 - \mathcal{O}\left(\frac{\epsilon}{\theta}\right)\right) \end{aligned} \quad (21)$$

Note that this holds for any β with large probability due to the lower bound on ϵ and $m > \frac{n}{2}$. Let us denote by $\mathcal{S}(\mathcal{C}_i^{\leq k})$ the set of all edges contained by a path starting at i of length at most k . After additionally conditioning on $\mathcal{C}_i^1 = \{v_1, \dots, v_{k'}\}$ and $\mathcal{S}(\mathcal{C}_j^{\leq 2})$ as well as $\mathcal{C}_{v_2}^1, \dots$ and $\mathcal{C}_{v_{k'}}^1$, the same argument yields for any α

$$\left| \mathcal{C}_\alpha^* \cap \mathcal{C}_{v_1}^1 \setminus \left(\mathcal{C}_j^{\leq 2} \cup (\mathcal{C}_{v_2}^1 \cup \dots \cup \mathcal{C}_{v_{k'}}^1) \right) \right| \geq a_1^{v_1, \alpha} m \left(1 - \mathcal{O} \left(\frac{\epsilon}{\theta} \right) \right). \quad (22)$$

Summarizing the lower bound (22) yields in view of (21) and $\frac{\epsilon}{\theta} \ll 1$

$$|\mathcal{C}_\alpha^* \cap \mathcal{C}_i^2 \setminus \mathcal{C}_j^{\leq 2}| \geq a_2^{i, \alpha} m^2 \left(1 - \mathcal{O} \left(\frac{\epsilon}{\theta} \right) \right).$$

Iterating the argument further, we end up with

$$|\mathcal{C}_\alpha^* \cap \mathcal{C}_i^k \setminus \mathcal{C}_j^{\leq k}| \geq a_k^{i, \alpha} m^k \left(1 - \mathcal{O} \left(\frac{\epsilon}{\theta} \right) \right). \quad (23)$$

From (20), (23) and $\theta^k n^{k-1} \ll 1$ we conclude

$$\begin{aligned} |\mathcal{C}_\alpha^* \cap \mathcal{C}_i^k \cap \mathcal{C}_j^{\leq k}| &\leq a_k^{i, \alpha} (n^k - m^k) + \mathcal{O}(\epsilon \theta^{k-1} n^k) \\ &= a_k^{i, \alpha} n^k (1 - (1 - \mathcal{O}(\theta^k n^{k-1}))^k) + \mathcal{O}(\epsilon \theta^{k-1} n^k) \\ &\lesssim \theta^{2k} n^{2k-1} + \epsilon \theta^{k-1} n^k. \end{aligned}$$

This is of course also an upper bound for $|\mathcal{C}_i^k \cap \mathcal{C}_j^k \cap \mathcal{C}_\alpha^*|$. Next, we discuss the overlap of two neighborhoods of different sizes $|\mathcal{C}_i^l \cap \mathcal{C}_j^k \cap \mathcal{C}_\alpha^*|$ with $l < k$ in case there exists no path of length at most $k - l$ between i and j . We can then argue very similar (after conditioning on $\mathcal{S}(\mathcal{C}_j^{\leq k-l+1})$) as in the case of $l = k$ and start the induction with a lower lower bound analogous to (21):

$$|\mathcal{C}_\beta^* \cap \mathcal{C}_i^1 \setminus \mathcal{C}_j^{\leq k-l+1}| \geq a_1^{i, \beta} m \left(1 - \mathcal{O} \left(\frac{\epsilon}{\theta} \right) \right)$$

Analogously to (22) and (23) we iterate this argument further and end up with

$$|\mathcal{C}_\alpha^* \cap \mathcal{C}_i^l \setminus \mathcal{C}_j^{\leq k}| \geq a_l^{i, \alpha} m^l \left(1 - \mathcal{O} \left(\frac{\epsilon}{\theta} \right) \right).$$

Together with the upper bound from Lemma 3

$$|\mathcal{C}_i^l \cap \mathcal{C}_\alpha^*| \leq \alpha_l^{i, \alpha} n^l \left(1 + \mathcal{O} \left(\frac{\epsilon}{\theta} + \theta^l n^{l-1} \right) \right)$$

we conclude

$$\begin{aligned} |\mathcal{C}_i^l \cap \mathcal{C}_j^k \cap \mathcal{C}_\alpha^*| &\leq \alpha_l^{i, \alpha} (n^l - m^l) + \mathcal{O}(\epsilon \theta^{l-1} n^l) \\ &\lesssim \theta^{l+k} n^{l+k-1} + \epsilon \theta^{l-1} n^l. \end{aligned}$$

□

Proof of Theorem 2. We start discussing the algorithm AWCD_k and consider $i \neq j$. Let us condition on $\mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})$ and write

$$\begin{aligned} \mathbb{E}[S_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] &= \sum_{v_1 \in \mathcal{C}_i^k, v_2 \in \mathcal{C}_j^k} \mathcal{Y}_{v_1 v_2} \\ &= D + S, \end{aligned}$$

where D denotes the deterministic part of the sum and S denotes the stochastic part (i.e. an independent sum of Bernoulli variables). Note that a summand $\mathcal{Y}_{v_1 v_2}$ is deterministic and $\equiv 1$ if and only if $v_1 \in \mathcal{C}_j^{k-1}$ or $v_2 \in \mathcal{C}_i^{k-1}$. First of all, let us consider the case $\mathcal{Y}_{ij} = 0$. According to Lemma 4 we have w.h.p.

$$|\mathcal{C}_i^{k-1} \cap \mathcal{C}_j^k \cap \mathcal{C}_\alpha^*| \lesssim \theta^{2k-1} n^{2k-2} + \epsilon \theta^{k-2} n^{k-1}. \quad (24)$$

Moreover, the 1-neighborhood around any member is of size at most $\mathcal{O}(\theta n(1 + \frac{\epsilon}{\theta}))$, so

$$D \lesssim \theta^{2k} n^{2k-1} + \epsilon \theta^{k-1} n^k.$$

Next, let us discuss the case where $\mathcal{Y}_{ij} = 1$. We start studying the case when we do not correct the bias at all, i.e. with the algorithm AWCD_k^0 . Then the upper bound (24) is no longer valid, instead we get

$$\begin{aligned} |\mathcal{C}_i^{k-1} \cap \mathcal{C}_j^k \cap \mathcal{C}_\beta^*| &\leq |\mathcal{C}_i^{k-1} \cap \mathcal{C}_\beta^*| \\ &\leq a_{k-1}^{i,\beta} n^{k-1} \left(1 + \mathcal{O}\left(\frac{\epsilon}{\theta}\right)\right). \end{aligned} \quad (25)$$

Again, using the upper bound $\theta_v^\alpha n(1 + \mathcal{O}(\frac{\epsilon}{\theta}))$ on size of the intersection of $\mathcal{C}_\alpha^*$ with the 1-neighborhood around any member v we conclude from (25)

$$D \leq \sum_{\alpha} (a_k^{i,\alpha} + a_k^{j,\alpha}) n^k \left(1 + \mathcal{O}\left(\frac{\epsilon}{\theta}\right)\right).$$

Next, we want to compute a lower bound on D . Note that if $v_2 \in \mathcal{C}_j^k \cap \mathcal{C}_i^{k-1}$, then any neighbor v of v_2 that is not contained by $\mathcal{C}_i^{\leq k-1}$ contributes to a deterministic summand $\mathcal{Y}_{vv_2} \equiv 1$ of D . From the proof of Lemma 4 we already have the lower bound (23)

$$\begin{aligned} |\mathcal{C}_j^k \cap \mathcal{C}_i^{k-1} \cap \mathcal{C}_\beta^*| &= |\mathcal{C}_\beta^* \cap \mathcal{C}_i^{k-1} \setminus \mathcal{C}_j^{\leq k-1}| \\ &\geq a_{k-1}^{i,\beta} (n - 2K^{k-2} \theta^{k-1} n^{k-1})^{k-1} \left(1 - \mathcal{O}\left(\frac{\epsilon}{\theta}\right)\right) \\ &= a_{k-1}^{i,\beta} n^{k-1} (1 - \mathcal{O}(\theta^{k-1} n^{k-2})) \left(1 - \mathcal{O}\left(\frac{\epsilon}{\theta}\right)\right). \end{aligned} \quad (26)$$

At the same time, any member v has at least $\theta_v^\alpha n (1 - \mathcal{O}(\theta^{k-1} n^{k-2})) (1 - \mathcal{O}(\frac{\epsilon}{\theta}))$ neighbors inside $\mathcal{C}_\alpha^* \setminus \mathcal{C}_i^{\leq k-1}$. Combining this with (26) yields

$$D \geq \sum_{\alpha} (a_k^{i,\alpha} + a_k^{j,\alpha}) n^k (1 - \mathcal{O}(\theta^{k-1} n^{k-2})) \left(1 - \mathcal{O}\left(\frac{\epsilon}{\theta}\right)\right).$$

Considering that according to Lemma 3

$$|\mathcal{C}_i^k| + |\mathcal{C}_j^k| = \sum_{\alpha} (a_k^{i,\alpha} + a_k^{j,\alpha}) n^k \left(1 + \mathcal{O}\left(\frac{\epsilon}{\theta} + \theta^k n^{k-1}\right)\right),$$

the deterministic part D is significantly smaller after we adjust the definition of S_{ij} to

$$\sum_{v_1 \in \mathcal{C}_i^k, v_2 \in \mathcal{C}_j^k} \mathcal{Y}_{v_1 v_2} - \mathbb{1}(\mathcal{Y}_{ij} = 1)(|\mathcal{C}_i^k| + |\mathcal{C}_j^k|).$$

To be precise, considering again AWCD_k , we now have the same bound as in the case $\mathcal{Y}_{ij} = 0$

$$\begin{aligned} D &= \mathcal{O} \left(\left(\frac{\epsilon}{\theta} + \theta^k n^{k-1} \right) \theta^k n^k \right) \\ &= \mathcal{O} \left(\theta^{2k} n^{2k-1} + \epsilon \theta^{k-1} n^k \right). \end{aligned} \quad (27)$$

Next, let us consider the stochastic part S . Note that if $v_1 \in \mathcal{C}_i^k \setminus \mathcal{C}_j^{\leq k-1}$ and $v_2 \in \mathcal{C}_j^k \setminus \mathcal{C}_i^{\leq k-1}$, then even after conditioning on $\mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})$, the term $\mathcal{Y}_{ij}$ is still a Bernoulli variable and not deterministic. In view of

$$\begin{aligned} |\mathcal{C}_\alpha^* \cap \mathcal{C}_i^k \setminus \mathcal{C}_j^{\leq k-1}| &= |\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| \left(1 + \mathcal{O} \left(\frac{|\mathcal{C}_j^{\leq k-1} \cap \mathcal{C}_\alpha^*|}{|\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*|} \right) \right) \\ &= |\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| \left(1 + \mathcal{O} \left(\frac{1}{\theta n} \right) \right) \end{aligned}$$

we conclude similar to (4) from Lemma 3 and Lemma 4

$$\begin{aligned} S &= \theta \sum_{\alpha} |\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| |\mathcal{C}_j^k \cap \mathcal{C}_\alpha^*| \left(1 + \mathcal{O} \left(\frac{1}{\theta n} \right) \right) \\ &\quad + \rho \sum_{\alpha \neq \beta} |\mathcal{C}_i^k \cap \mathcal{C}_\alpha^*| |\mathcal{C}_j^k \cap \mathcal{C}_\beta^*| \left(1 + \mathcal{O} \left(\frac{1}{\theta n} \right) \right) + \mathcal{O}(\theta |\mathcal{C}_i^k \cap \mathcal{C}_j^k|) \\ &= \begin{cases} an^{2k} \left(1 + \mathcal{O} \left(\frac{\epsilon}{\theta} + \theta^k n^{k-1} + \frac{1}{\theta n} \right) \right) + \mathcal{O}(\theta^{2k+1} n^{2k-1} + \epsilon \theta^k n^k + D) & , \exists \alpha : i, j \in \mathcal{C}_\alpha^* \\ cn^{2k} \left(1 + \mathcal{O} \left(\frac{\epsilon}{\theta} + \theta^k n^{k-1} + \frac{1}{\theta n} \right) \right) + \mathcal{O}(\theta^{2k+1} n^{2k-1} + \epsilon \theta^k n^k + D) & , \text{otherwise} \end{cases} \\ &= \begin{cases} an^{2k} + \mathcal{O}(\theta R) & , \exists \alpha : i, j \in \mathcal{C}_\alpha^* \\ cn^{2k} + \mathcal{O}(\theta R) & , \text{otherwise} \end{cases} \end{aligned} \quad (28)$$

for

$$\begin{aligned} a &= \theta [a_k^2 + (K-1)b_k^2] + \rho [2(K-1)a_k b_k + (K-1)(K-2)b_k^2], \\ c &= \theta [2a_k b_k + (K-2)b_k^2] + \rho [a_k^2 + 2(K-2)a_k b_k + ((K-1)(K-2) + 1)b_k^2] \text{ and} \\ \theta R &= \mathcal{O} \left(\theta^{2k+1} n^{2k} \left(\frac{\epsilon}{\theta} + \theta^k n^{k-1} + \frac{1}{\theta n} \right) + \theta^{2k+1} n^{2k-1} + \epsilon \theta^k n^k + \theta^{2k} n^{2k-1} + \epsilon \theta^{k-1} n^k \right) \\ &= \mathcal{O} \left(\theta^{3k+1} n^{3k-1} + \epsilon \theta^{k-1} n^k + \epsilon \theta^{2k} n^{2k} \right). \end{aligned}$$

Note that according to Remark 7 we have

$$\begin{aligned} a - c &= \theta a_k^2 + \theta b_k^2 - 2\theta a_k b_k + 2\rho a_k b_k - \rho a_k^2 - \rho b_k^2 \\ &= (\theta - \rho)(a_k - b_k)^2 \\ &= (\theta - \rho)^{2k+1}. \end{aligned}$$

In case $i = j$ the concentration result (28) on S as well as the upper bound (27) for D are still valid: After conditioning on $\mathcal{S}(\mathcal{C}_i^{\leq k})$ we have $D \equiv 0$ as the only deterministic edges are $\mathcal{Y}_{vv} = 0$ for any $v \in \mathcal{C}_i^k$. This leads to an additional term $\mathcal{O}(\theta^{k+1} n^k)$ that needs to be considered in (28), however this is much smaller than $\mathcal{O}(\theta R)$.

Using again Lemma 3, we compute for the denominator in case $i \neq j$

$$\begin{aligned} N_{ij}^k &= \left(\sum_{\alpha} a_k^{i,\alpha} \right)^2 n^{2k} \left(1 + \mathcal{O} \left(\frac{\epsilon}{\theta} + \theta^k n^{k-1} \right) \right) \\ &= dn^{2k} + \mathcal{O}(\epsilon \theta^{2k-1} n^k + \theta^{3k} n^{3k-1}) \\ &= dn^{2k} + \mathcal{O}(R) \end{aligned}$$

for

$$d = a_k + (K - 1)b_k.$$

In case $i = j$ the above is still valid: We only need to add a term $\mathcal{O}(C_i^k) = \mathcal{O}(\theta^k n^k)$ which is in view of $\epsilon \theta^{2k-1} n^{2k} \gg \theta^{2k-1} n^{2k-1} \gg \theta^k n^k$ already contained in $\mathcal{O}(R)$.

The rest of the proof is very similar to the proof of Theorem 1: From the above, we conclude

$$\begin{aligned} \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] &= \begin{cases} \frac{a}{d} (1 + \mathcal{O}(\delta)) & , \exists \alpha : i, j \in \mathcal{C}_{\alpha}^* \\ \frac{c}{d} (1 + \mathcal{O}(\delta)) & , \text{otherwise} \end{cases} \\ \text{and } \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] &= \begin{cases} \frac{a}{d} (1 + \mathcal{O}(\delta)) & , \exists \alpha : i, j \in \mathcal{C}_{\alpha}^* \\ \frac{a+c}{2d} (1 + \mathcal{O}(\delta)) & , \text{otherwise} \end{cases} \end{aligned}$$

for

$$\delta = \frac{R}{\theta^{2k} n^{2k}}$$

and under the assumption $\delta \ll 1$, which we discuss later. For large enough n in the case of i and j belonging to different communities we have

$$\begin{aligned} \mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] &= \frac{a+c}{2d} (1 + \mathcal{O}(\delta)) - \frac{c}{d} (1 + \mathcal{O}(\delta)) \\ &= \frac{a-c}{2d} + \mathcal{O}(\delta \theta) \\ &\propto \frac{(\theta - p)^{2k+1}}{\theta^{2k}} + \mathcal{O}(\delta \theta) \end{aligned} \tag{29}$$

and otherwise

$$\begin{aligned} \mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] &= \frac{a}{d} (1 + \mathcal{O}(\delta)) - \frac{a}{d} (1 + \mathcal{O}(\delta)) \\ &= \mathcal{O}(\delta \theta). \end{aligned} \tag{30}$$

Next we apply Lemma 2: Considering that the sum S_{ij} consists of $\mathcal{O}(\theta^{2k} n^{2k})$ summands, we have with large probability

$$|S_{ij} - \mathbb{E}[S_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})]| \lesssim \log n + (\log n)^{\frac{1}{2}} \theta^{k+\frac{1}{2}} n^k.$$

The upper bound simplifies under the condition $\theta^{2k+1} n^{2k} \gtrsim \log n$ to

$$|S_{ij} - \mathbb{E}[S_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})]| \lesssim (\log n)^{\frac{1}{2}} \theta^{k+\frac{1}{2}} n^k,$$

implying

$$|\tilde{\theta}_{ij} - \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})]| \lesssim \frac{(\log n)^{\frac{1}{2}}}{\theta^{k-\frac{1}{2}} n^k}. \quad (31)$$

As $\theta^{2k+1} n^{2k} \gg \log n$, the upper bound is $\ll \theta$, implying furthermore

$$\tilde{\theta}_{ij}, \tilde{\theta}_{i \vee j} \in \left[\frac{\rho}{2}, \theta + \frac{\rho}{2} \right].$$

Consequently, the corresponding Fisher information is up to bounded constants given by θ^{-1} . Using the quadratic Taylor expansion of the Kullback-Leibler, we conclude in the case where i and j belong to different communities from (29) and (31)

$$\begin{aligned} \theta^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\tilde{\theta}_{ij}, \tilde{\theta}_{i \vee j}) &\gtrsim |\tilde{\theta}_{ij} - \tilde{\theta}_{i \vee j}| \\ &\geq \left| \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] \right| \\ &\quad - \left| \tilde{\theta}_{ij} - \mathbb{E}[\tilde{\theta}_{ij} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] \right| \\ &\quad - \left| \tilde{\theta}_{i \vee j} - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{S}(\mathcal{C}_i^{\leq k}) \cup \mathcal{S}(\mathcal{C}_j^{\leq k})] \right| \\ &\gtrsim \frac{(\theta - \rho)^{2k+1}}{\theta^{2k}} + \mathcal{O} \left(\delta\theta + \frac{(\log n)^{\frac{1}{2}}}{\theta^{k-\frac{1}{2}} n^k} \right), \end{aligned} \quad (32)$$

whereas in the other case we conclude analogously from (30)

$$\theta^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\tilde{\theta}_{ij}, \tilde{\theta}_{i \vee j}) \lesssim \delta\theta + \frac{(\log n)^{\frac{1}{2}}}{\theta^{k-\frac{1}{2}} n^k}. \quad (33)$$

From (33) and (32) we conclude: A sufficient condition (together with the lower bound from Lemma 3 on ϵ that guarantees the large probability of the concentration results) for consistency of the algorithm is

$$\begin{aligned} \frac{(\theta - \rho)^{2k+1}}{\theta^{2k}} &\gg \delta\theta + \frac{(\log n)^{\frac{1}{2}}}{\theta^{k-\frac{1}{2}} n^k} \\ \Leftrightarrow (\theta - \rho)^{2k+1} &\gg \theta R n^{-2k} + (\log n)^{\frac{1}{2}} \theta^{k+\frac{1}{2}} n^{-k} \\ \Leftrightarrow (\theta - \rho)^{2k+1} &\gg \theta^{3k+1} n^{k-1} + \epsilon \theta^{k-1} n^{-k} + \epsilon \theta^{2k} + (\log n)^{\frac{1}{2}} \theta^{k+\frac{1}{2}} n^{-k}. \end{aligned} \quad (34)$$

At the same time, considering $\theta \gg \frac{\log n}{n}$, all of the above concentration results above only hold with large probability as long as ϵ satisfies the lower bound (c.f. Lemma 2)

$$\begin{aligned} \min \left\{ \frac{\epsilon^2 n}{\theta \log n}, \frac{\epsilon n}{\log n} \right\} &\gg 1 \\ \Leftrightarrow \epsilon &\gg \frac{\log n}{n} + \sqrt{\frac{\theta \log n}{n}} \\ \Leftrightarrow \epsilon &\gg \sqrt{\frac{\theta \log n}{n}}, \end{aligned} \quad (35)$$

whereas the ϵ -dependent part of (34) is equivalent to the upper bound

$$\epsilon \ll \min \left\{ \frac{(\theta - \rho)^{2k+1}}{\theta^{2k}}, \frac{(\theta - \rho)^{2k+1} n^k}{\theta^{k-1}} \right\}. \quad (36)$$

An $\epsilon > 0$ satisfying (35) as well as (36) exists if and only if

$$\min \left\{ \frac{(\theta - \rho)^{2k+1}}{\theta^{2k}}, \frac{(\theta - \rho)^{2k+1} n^k}{\theta^{k-1}} \right\} \gg \sqrt{\frac{\theta \log n}{n}} \\ \Leftrightarrow (\theta - \rho)^{2k+1} \gg \theta^{2k+\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}} + \theta^{k-\frac{1}{2}} n^{-k-\frac{1}{2}} (\log n)^{\frac{1}{2}}. \quad (37)$$

Because we only discuss the case $k \geq 2$, our assumptions ensure $\theta \ll n^{-\frac{k-1}{k}} \ll n^{-\frac{1}{2}}$ and in particular $\theta^{k-\frac{1}{2}} n^{-k-\frac{1}{2}} (\log n)^{\frac{1}{2}} \ll \theta^{k+\frac{1}{2}} n^{-k} (\log n)^{\frac{1}{2}}$. Thus we can simplify (34) and (37) into the following sufficient condition for consistency of the algorithm

$$\theta - \rho \gg \max\{A, B, C\}$$

for

$$A = \theta^{\frac{3k+1}{2k+1}} n^{\frac{k-1}{2k+1}}, \\ B = \theta^{\frac{4k+1}{4k+2}} n^{-\frac{1}{4k+2}} (\log n)^{\frac{1}{4k+2}} \text{ and} \\ C = \theta^{\frac{2k-1}{4k+2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{4k+2}}.$$

Next, let us consider the algorithm AWCD^+ . The proof is almost identical - we only need to modify $D = 0$, leading to

$$\theta R = \mathcal{O}(\theta^{3k+1} n^{3k-1} + \epsilon \theta^{2k} n^{2k}).$$

Consequently, in the lower bound (34) we can drop the term $\epsilon \theta^{k-1} n^{-k}$ leading to the following analogous condition

$$(\theta - \rho)^{2k+1} \gg \theta^{3k+1} n^{k-1} + \epsilon \theta^{2k} + (\log n)^{\frac{1}{2}} \theta^{k+\frac{1}{2}} n^{-k}$$

and (37) simplifies to

$$(\theta - \rho)^{2k+1} \gg \theta^{2k+\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}.$$

We end up with the final sufficient condition

$$\theta - \rho \gg \max\{A, B, C^+\}$$

for

$$C^+ = \theta^{\frac{1}{2}} n^{-\frac{k}{2k+1}} (\log n)^{\frac{1}{4k+2}}.$$

Finally, let us consider the algorithm AWCD_k° without any correction of the bias term. The deterministic part of the conditional expectation of S_{ij} increases to $D = \mathcal{O}(\theta^k n^k)$ implying

$$\theta R = \mathcal{O}(\theta^{3k+1} n^{3k-1} + \epsilon \theta^{2k} n^{2k} + \theta^k n^k).$$

We end up with the following lower bound corresponding to (34)

$$(\theta - \rho)^{2k+1} \gg \theta^{3k+1} n^{k-1} + \epsilon \theta^{2k} + \theta^k n^{-k} + (\log n)^{\frac{1}{2}} \theta^{k+\frac{1}{2}} n^{-k} \\ \Leftrightarrow (\theta - \rho)^{2k+1} \gg \theta^{3k+1} n^{k-1} + \epsilon \theta^{2k} + \theta^k n^{-k}$$

and instead of (37) we have again

$$(\theta - \rho)^{2k+1} \gg \theta^{2k+\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}.$$

This leads to the final sufficient condition

$$\theta - \rho \gg \max\{A, B, C^\circ\}$$

with

$$C^\circ = \theta^{\frac{k}{2k+1}} n^{-\frac{k}{2k+1}}.$$

□

Proof of Proposition 1. We proceed analogously as in the case of identical block size. Suppose $\epsilon > 0$ and $\alpha \in \{1, 2, \dots, K\}$. Moreover, we use the notation α_i for a member i to denote the corresponding community index such that $i \in \mathcal{C}_{\alpha_i}^*$. According to Lemma 1 we have

$$|\mathcal{C}_i^j \cap \mathcal{C}_\alpha^*| = \theta_i^\alpha n_\alpha + \mathcal{O}(\epsilon n_{\max} + \theta) \quad (38)$$

for

$$\theta_i^\alpha = \begin{cases} \theta & , i \in \mathcal{C}_\alpha^* \\ \rho & , i \notin \mathcal{C}_\alpha^* \end{cases}$$

on an event of probability at least

$$1 - 2 \exp\left(-\frac{\epsilon^2 n_{\min}}{2\theta + \epsilon}\right).$$

Similarly, we conclude for $i \neq j$

$$|\mathcal{C}_i^j \cap \mathcal{C}_j^j| = \mathcal{O}(\theta^2 n_{\max} + \epsilon n_{\max}) \quad (39)$$

on an event of probability of at least the same lower bound. In the following, we will restrict to an event where (38) and (39) are satisfied for any $i \neq j$. By union bound, this event of asymptotic probability 1 as long as

$$\min\left\{\frac{\epsilon^2 n_{\min}}{\theta}, \epsilon n_{\min}\right\} \gg \log n_{\min}.$$

Before moving on, we introduce the following notation:

$$\begin{aligned} T &= \sum_{\alpha \notin \{\alpha_i, \alpha_j\}} n_\alpha \\ a_i^\circ &= \theta^3 n_i^2 + \theta \rho^2 (n_j^2 + 2n_i n_j) \\ c_{ij}^\circ &= \theta^2 \rho (n_i^2 + n_j^2 + n_i n_j) + \rho^3 n_i n_j \\ r_i &= \theta \rho^2 \left(\sum_{\alpha \notin \{\alpha_i, \alpha_j\}} n_\alpha^2 + 2n_i T \right) + 2\rho^3 n_j T + \rho^3 \sum_{\alpha \neq \beta, \alpha, \beta \notin \{\alpha_i, \alpha_j\}} n_\alpha n_\beta \\ r_{ij} &= \theta \rho^2 \left(\sum_{\alpha \notin \{\alpha_i, \alpha_j\}} n_\alpha^2 + (n_i + n_j) T \right) + \rho^3 (n_i + n_j) T + \rho^3 \sum_{\alpha \neq \beta, \alpha, \beta \notin \{\alpha_i, \alpha_j\}} n_\alpha n_\beta \end{aligned}$$

Note that according to our assumptions $\theta n_{\min} \rightarrow \infty$. We conclude from (38) and (39) for $\epsilon < \theta$

$$\begin{aligned} \mathbb{E}[S_{ij}|\mathcal{C}_i^j, \mathcal{C}_j^i] &= \theta \sum_{\alpha} |\mathcal{C}_i^j \cap \mathcal{C}_{\alpha}^*| |\mathcal{C}_j^i \cap \mathcal{C}_{\alpha}^*| + \rho \sum_{\alpha \neq \beta} |\mathcal{C}_i^j \cap \mathcal{C}_{\alpha}^*| |\mathcal{C}_j^i \cap \mathcal{C}_{\alpha}^*| + \mathcal{O}(\theta |\mathcal{C}_i^j \cap \mathcal{C}_j^i|) \\ &= \theta \sum_{\alpha} \theta^{\alpha} \theta_j^{\alpha} n_{\alpha}^2 + \rho \sum_{\alpha \neq \beta} \theta_i^{\alpha} \theta_j^{\beta} n_{\alpha} n_{\beta} + \mathcal{O}(\theta^3 n_{\max} + \theta^2 \epsilon n_{\max}^2) \\ &= \begin{cases} a_i + \mathcal{O}(\theta^3 n_{\max} + \theta^2 \epsilon n_{\max}^2) & , \exists \alpha : i, j \in \mathcal{C}_{\alpha}^* \\ c_{ij} + \mathcal{O}(\theta^3 n_{\max} + \theta^2 \epsilon n_{\max}^2) & , \text{otherwise} \end{cases} \end{aligned}$$

for

$$\begin{aligned} a_i &:= a_i^{\circ} + r_i \text{ and} \\ c_{ij} &:= c_{ij}^{\circ} + r_{ij}. \end{aligned}$$

Similarly

$$\mathbb{E}[S_{ii}|\mathcal{C}_i] = a_i + \mathcal{O}(\theta^2 n_{\max} + \theta^2 \epsilon n_{\max}^2)$$

as well as

$$\begin{aligned} \mathbb{E}[N_{ij}|\mathcal{C}_i, \mathcal{C}_j] &= N_{ij} \\ &= \left(\sum_{\alpha} |\mathcal{C}_i \cap \mathcal{C}_{\alpha}^*| \right) \left(\sum_{\alpha} |\mathcal{C}_j \cap \mathcal{C}_{\alpha}^*| \right) \\ &= d_{ij} + \mathcal{O}(\theta^2 n_{\max} + \epsilon \theta n_{\max}^2) \end{aligned}$$

for

$$\begin{aligned} d_{ij} &= (N_i^{\circ} + \rho T)(N_j^{\circ} + \rho T), \\ N_i^{\circ} &= \theta n_i + \rho n_j \text{ and} \\ N_j^{\circ} &= \rho n_i + \theta n_j. \end{aligned}$$

Analogously as before, we introduce

$$R := \theta n_{\max} + \epsilon \theta n_{\max}^2$$

and get

$$\mathbb{E}[\tilde{\theta}_{ii}|\mathcal{C}_i] = \frac{a_i + \mathcal{O}(\theta R)}{d_{ii} + \mathcal{O}(R)}.$$

For

$$\delta := \max \left\{ \frac{1}{\theta n_{\max}}, \frac{\epsilon}{\theta} \right\}$$

and under the assumption $\delta \ll 1$ we can rewrite

$$\mathbb{E}[\tilde{\theta}_{ii}|\mathcal{C}_i] = \frac{a_i}{d_{ii}} (1 + \mathcal{O}(\delta))$$

and analogously

$$\mathbb{E}[\tilde{\theta}_{ij}|\mathcal{C}_i, \mathcal{C}_j] = \begin{cases} \frac{a_i}{d_{ii}} (1 + \mathcal{O}(\delta)) & , \exists \alpha : i, j \in \mathcal{C}_\alpha^* \\ \frac{c_{ij}}{d_{ij}} (1 + \mathcal{O}(\delta)) & , \text{otherwise.} \end{cases}$$

Let us consider the case $i \in \alpha_i$. In this case, we have

$$\begin{aligned} \mathbb{E}[\tilde{\theta}_{ii} - \tilde{\theta}_{ij}|\mathcal{C}_i, \mathcal{C}_j] &= \frac{a_i}{d_{ii}} (1 + \mathcal{O}(\delta)) - \frac{a_i}{d_{ii}} (1 + \mathcal{O}(\delta)) \\ &= \mathcal{O}(\delta\theta). \end{aligned}$$

Since $\tilde{\theta}_{i \vee j}$ is the weighted mean of $\tilde{\theta}_{ii}$, $\tilde{\theta}_{jj}$ and $\tilde{\theta}_{ij}$, we conclude also

$$\mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{ij}|\mathcal{C}_i, \mathcal{C}_j] = \mathcal{O}(\delta\theta). \quad (40)$$

Next, we consider the other case $i \notin \alpha_i$. First, we need to calculate a lower bound for the difference $\frac{a_i}{d_{ii}} - \frac{c_i}{d_{ij}}$. We have

$$\begin{aligned} \frac{a_i}{d_{ii}} - \frac{c_i}{d_{ij}} &= \frac{a_i^\circ + r_i}{(N_i^\circ + \rho T)^2} - \frac{c_{ij}^\circ + r_{ij}}{(N_i^\circ + \rho T)(N_j^\circ + \rho T)} \\ &= \frac{1}{(N_i^\circ + \rho T)^2(N_j^\circ + \rho T)} (A_{ij} + B_{ij} + C_{ij}) \end{aligned}$$

for

$$\begin{aligned} A_{ij} &= a_i^\circ N_j^\circ - c_{ij}^\circ N_i^\circ, \\ B_{ij} &= \rho T(a_i^\circ - c_{ij}^\circ) \text{ and} \\ C_{ij} &= r_i(N_j^\circ - N_i^\circ) + (r_i - r_{ij})(N_i^\circ + \rho T). \end{aligned}$$

We have

$$\begin{aligned} r_i - r_{ij} &= \theta \rho^2 (n_i - n_j) T + \rho^3 (n_j - n_i) T \\ &= \rho^2 T (\theta n_i + \rho n_j - (\theta n_j + \rho n_i)) \\ &= \rho^2 T (N_i^\circ - N_j^\circ). \end{aligned}$$

Consequently,

$$\begin{aligned} C_{ij} &= (N_i^\circ - N_j^\circ)(\rho^2 T N_i^\circ + \rho^3 T^2 - r_i) \\ &= -(N_i^\circ - N_j^\circ) \left((\theta - \rho) \rho^2 \sum_{\alpha \notin \{\alpha_i, \alpha_j\}} n_\alpha^2 + \theta \rho^2 n_i T + \rho^3 n_j T \right) \\ &= -(N_i^\circ - N_j^\circ) \left((\theta - \rho) \rho^2 \sum_{\alpha \notin \{\alpha_i, \alpha_j\}} n_\alpha^2 + \rho^2 N_i^\circ T \right) \end{aligned}$$

and

$$\begin{aligned} C_{ij} + C_{ji} &= (N_i^\circ - N_j^\circ)(-\rho^2 N_i^\circ T + \rho^2 N_j^\circ T) \\ &= -(N_i^\circ - N_j^\circ)^2 \rho^2 T \\ &= -(\theta - \rho)^2 (n_i - n_j)^2 \rho^2 T. \end{aligned}$$

Moreover,

$$\begin{aligned} B_{ij} + B_{ji} &= \rho T (a_i^\circ + a_j^\circ - 2c_{ij}^\circ) \\ &= \rho T [\theta^3(n_i^2 + n_j^2) + \theta\rho^2(n_i^2 + n_j^2 + 4n_in_j) - 2\theta^2\rho(n_i^2 + n_j^2 + n_in_j) - 2\rho^3n_in_j] \\ &= \rho T [\rho(\theta - \rho)^2(n_i - n_j)^2 + (\theta - \rho)^3(n_i^2 + n_j^2)]. \end{aligned}$$

The first term A_{ij} simplifies to

$$\begin{aligned} A_{ij} &= a_i^\circ N_j^\circ - c_{ij}^\circ N_i^\circ \\ &= [\theta^3n_i^2 + \theta\rho^2(n_j^2 + 2n_in_j)](\rho n_i + \theta n_j) - [\theta^2\rho(n_i^2 + n_j^2 + n_in_j) + \rho^3n_in_j](\theta n_i + \rho n_j) \\ &= \theta^4n_i^2n_j + \theta^2\rho^2n_in_j^2 + \theta\rho^3n_in_j^2 + \theta\rho^3n_i^2n_j - \theta^3\rho n_in_j^2 - \theta^3\rho n_i^2n_j - \theta^2\rho^2n_i^2n_j - \rho^4n_in_j^2 \\ &= n_in_j^2 \underbrace{(\theta^4 - 2\theta^3\rho + 2\theta\rho^3 - \rho^4)}_{=(\theta+\rho)(\theta-\rho)^3} + n_in_j(n_i - n_j) \underbrace{(\theta^4 - \theta^3\rho - \theta^2\rho^2 + \theta\rho^3)}_{=-\theta(\theta+\rho)(\theta-\rho)^2} \\ &= n_in_j^2(\theta + \rho)(\theta - \rho)^3 + D_{ij} \end{aligned}$$

for

$$D_{ij} = (n_i - n_j)n_in_j\theta(\theta + \rho)(\theta - \rho)^2.$$

Note that

$$\underbrace{D_{ij} + D_{ji}}_{=0} + B_{ij} + B_{ji} + C_{ij} + C_{ji} = \rho(\theta - \rho)^3(n_i^2 + n_j^2)T.$$

So w.l.o.g. we can assume

$$D_{ij} + B_{ij} + C_{ij} \geq \frac{1}{2}\rho(\theta - \rho)^3(n_i^2 + n_j^2)T,$$

implying

$$A_{ij} + B_{ij} + C_{ij} > \frac{1}{2}\rho(\theta - \rho)^3(n_i^2 + n_j^2)T.$$

We conclude w.l.o.g. (possibly after exchanging the indices i and j)

$$\frac{a_i}{d_{ii}} - \frac{c_i}{d_{ij}} \gtrsim \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2}$$

and thus

$$\begin{aligned} \mathbb{E}[\tilde{\theta}_{ii} - \tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j] &= \frac{a_i}{d_{ii}} (1 + \mathcal{O}(\delta)) - \frac{c_{ij}}{d_{ij}} (1 + \mathcal{O}(\delta)) \\ &\gtrsim \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2} + \mathcal{O}(\delta\theta). \end{aligned}$$

Since $\tilde{\theta}_{i \vee j}$ is the weighted mean of $\tilde{\theta}_{ii}$, $\tilde{\theta}_{jj}$ and $\tilde{\theta}_{ij}$, there exists indices $(i', j') \in \{(i, i), (j, j), (i, j)\}$ such that

$$\mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{i'j'} | \mathcal{C}_i, \mathcal{C}_j] \gtrsim \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2} + \mathcal{O}(\delta\theta). \quad (41)$$

Similar as in the proof of Theorem 1, we deduce from our assumptions $\theta^3 n_{\min}^2 \gg \log n_{\min}$ and $\theta \leq \frac{1}{2}$ as well as Lemma 2 that

- the estimates $\tilde{\theta}$. (including $\tilde{\theta}_{ii}$, $\tilde{\theta}_{jj}$, $\tilde{\theta}_{ij}$ and $\tilde{\theta}_{i \vee j}$) are with large probability inside an interval of the form $[\frac{\rho}{2}, \theta + \frac{\rho}{2}]$, implying that the corresponding Fisher information of a Bernoulli variable with such a mean is up to bounded constants given by θ^{-1} and
- with large probability,

$$|\tilde{\theta} - \mathbb{E}[\tilde{\theta} | \mathcal{C}_i, \mathcal{C}_j]| \lesssim (\log n_{\min})^{\frac{1}{2}} \theta^{-\frac{1}{2}} n_{\min}^{-1}.$$

Combining the above with the quadratic Taylor expansion of the Kullback-Leibler and (41), we conclude in the case where i and j belong to different communities

$$\begin{aligned} \theta^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\tilde{\theta}_{i'j'}, \tilde{\theta}_{i \vee j}) &\gtrsim |\tilde{\theta}_{i'j'} - \tilde{\theta}_{i \vee j}| \\ &\geq \left| \mathbb{E}[\tilde{\theta}_{i'j'} | \mathcal{C}_i, \mathcal{C}_j] - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j] \right| - \left| \tilde{\theta}_{i'j'} - \mathbb{E}[\tilde{\theta}_{i'j'} | \mathcal{C}_i, \mathcal{C}_j] \right| - \left| \tilde{\theta}_{i \vee j} - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j] \right| \\ &\gtrsim \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2} + \mathcal{O}\left(\delta\theta + (\log n_{\min})^{\frac{1}{2}} \theta^{-\frac{1}{2}} n_{\min}^{-1}\right), \end{aligned}$$

whereas in the other case we conclude analogously from (40)

$$\theta^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\tilde{\theta}, \tilde{\theta}_{i \vee j}) \lesssim \delta\theta + (\log n_{\min})^{\frac{1}{2}} \theta^{-\frac{1}{2}} n_{\min}^{-1}.$$

To ensure that the Kullback-Leibler divergence in the second case is significantly smaller than in the first case, it will suffice to check

$$\begin{aligned} \delta\theta &\ll \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2} \\ \Leftrightarrow \quad \max\left\{\frac{1}{\theta n_{\max}}, \frac{\epsilon}{\theta}\right\} &\ll \left(\frac{\theta - \rho}{\theta}\right)^3 \frac{n_{\min}^2}{n_{\max}^2} \end{aligned} \quad (42)$$

and

$$\sqrt{\frac{\log n_{\min}}{\theta n_{\min}^2}} \ll \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2}. \quad (43)$$

Those are all conditions in the proof that remain to be checked except the large probability assumption

$$\min\left\{\frac{\epsilon^2 n_{\min}}{\theta \log n_{\min}}, \frac{\epsilon n_{\min}}{\log n_{\min}}\right\} \gg 1. \quad (44)$$

An $\epsilon > 0$ satisfying (42) and (44) exists as long as

$$\frac{\log n_{\min}}{n_{\min}} + \sqrt{\frac{\theta \log n_{\min}}{n_{\min}}} \ll \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2}, \quad (45)$$

whereas the ϵ -independent part of conditions (42) and (43) can be summarized by

$$\frac{1}{n_{\max}} + \sqrt{\frac{\log n_{\min}}{\theta n_{\min}^2}} \ll \frac{(\theta - \rho)^3}{\theta^2} \frac{n_{\min}^2}{n_{\max}^2}. \quad (46)$$

By simple calculus, we can verify that our assumption

$$\theta - \rho \gg \left(\frac{n_{\max}}{n_{\min}}\right)^{\frac{2}{3}} \max\{n_{\min}^{-\frac{1}{3}} \theta^{\frac{1}{2}} (\log n_{\min})^{\frac{1}{6}}, n_{\min}^{-\frac{1}{6}} \theta^{\frac{5}{6}} (\log n_{\min})^{\frac{1}{6}}\}$$

implies conditions (45) and (46). Note that d_{ii} , d_{jj} and d_{ij} only differ by bounded factors. Taking into account the concentration result $N = d(1 + \mathcal{O}(\delta))$ with $\delta \ll 1$, this also ensures consistency of the test - provided a proper threshold is given. $\square$

Proof of Proposition 2. We proceed analogously as previously. Suppose $\epsilon > 0$ and $\alpha \in \{1, 2\}$. According to Lemma 1 we have

$$|\mathcal{C}_i^j \cap \mathcal{C}_\alpha^*| = \theta_i^\alpha n_\alpha + \mathcal{O}(\epsilon n_{\max} + \theta_{\max}) \quad (47)$$

for

$$\theta_i^\alpha = \begin{cases} \theta_\alpha & , i \in \mathcal{C}_\alpha^* \\ \rho & , i \notin \mathcal{C}_\alpha^* \end{cases}$$

on an event of probability at least

$$1 - 2 \exp\left(-\frac{\epsilon^2(n_{\min} - 2)}{2\theta_{\max} + \epsilon}\right).$$

Similarly, we conclude for $i \neq j$

$$|\mathcal{C}_i^j \cap \mathcal{C}_j^j| = \mathcal{O}(\theta_{\max}^2 n_{\max} + \epsilon n_{\max}) \quad (48)$$

on an event of probability of at least the same lower bound. In the following, we will restrict to an event where (47) and (48) are satisfied for any $i \neq j$. By union bound, this event of asymptotic probability 1 as long as

$$\min\left\{\frac{\epsilon^2 n_{\min}}{\theta_{\max}}, \epsilon n_{\min}\right\} \gg \log n_{\min}.$$

Note that according to our assumptions $\theta_{\min} n_{\min} \rightarrow \infty$. We conclude from (47) and (48) for $\epsilon < \theta_{\min}$

$$\begin{aligned} \mathbb{E}[S_{ij} | \mathcal{C}_i^j, \mathcal{C}_j^i] &= \sum_{\alpha} \theta_{\alpha} |\mathcal{C}_i^j \cap \mathcal{C}_{\alpha}^*| |\mathcal{C}_j^i \cap \mathcal{C}_{\alpha}^*| + \rho \sum_{\alpha \neq \beta} |\mathcal{C}_i^j \cap \mathcal{C}_{\alpha}^*| |\mathcal{C}_j^i \cap \mathcal{C}_{\alpha}^*| + \mathcal{O}(\theta_{\max} |\mathcal{C}_i^j \cap \mathcal{C}_j^i|) \\ &= \sum_{\alpha} \theta_{\alpha} \theta_i^{\alpha} \theta_j^{\alpha} n_{\alpha}^2 + \rho \sum_{\alpha \neq \beta} \theta_i^{\alpha} \theta_j^{\beta} n_{\alpha} n_{\beta} + \mathcal{O}(\theta_{\max}^3 n_{\max} + \theta_{\max}^2 \epsilon n_{\max}^2) \\ &= \begin{cases} a_{11} + \mathcal{O}(\theta^3 n_{\max} + \theta^2 \epsilon n_{\max}^2) & , i, j \in \mathcal{C}_1^* \\ a_{22} + \mathcal{O}(\theta^3 n_{\max} + \theta^2 \epsilon n_{\max}^2) & , i, j \in \mathcal{C}_2^* \\ a_{12} + \mathcal{O}(\theta^3 n_{\max} + \theta^2 \epsilon n_{\max}^2) & , \text{otherwise} \end{cases} \end{aligned}$$

for

$$\begin{aligned} a_{11} &:= \theta_1^3 n_1^2 + 2\theta_1 \rho^2 n_1 n_2 + \theta_2 \rho^2 n_2^2, \\ a_{22} &:= \theta_2^3 n_2^2 + 2\theta_2 \rho^2 n_1 n_2 + \theta_1 \rho^2 n_1^2 \text{ and} \\ a_{12} &:= \theta_1^2 \rho n_1^2 + \theta_2^2 \rho n_2^2 + \theta_1 \theta_2 \rho n_1 n_2 + \rho^3 n_1 n_2. \end{aligned}$$

Similarly

$$\mathbb{E}[S_{ii} | \mathcal{C}_i] = a_i + \mathcal{O}(\theta_{\max}^2 n_{\max} + \theta_{\max}^2 \epsilon n_{\max}^2)$$

as well as

$$\begin{aligned} \mathbb{E}[N_{ij} | \mathcal{C}_i, \mathcal{C}_j] &= N_{ij} \\ &= \left(\sum_{\alpha} |\mathcal{C}_i \cap \mathcal{C}_{\alpha}^*| \right) \left(\sum_{\alpha} |\mathcal{C}_j \cap \mathcal{C}_{\alpha}^*| \right) \\ &= \begin{cases} d_{11} + \mathcal{O}(\theta_{\max}^2 n_{\max} + \epsilon \theta_{\max} n_{\max}^2) & , i, j \in \mathcal{C}_1^* \\ d_{22} + \mathcal{O}(\theta_{\max}^2 n_{\max} + \epsilon \theta_{\max} n_{\max}^2) & , i, j \in \mathcal{C}_2^* \\ d_{12} + \mathcal{O}(\theta_{\max}^2 n_{\max} + \epsilon \theta_{\max} n_{\max}^2) & , \text{otherwise} \end{cases} \end{aligned}$$

for

$$\begin{aligned} d_{11} &:= (\theta_1 n_1 + \rho n_2)^2, \\ d_{22} &:= (\theta_2 n_2 + \rho n_1)^2 \text{ and} \\ d_{12} &:= (\theta_1 n_1 + \rho n_2)(\theta_2 n_2 + \rho n_1). \end{aligned}$$

For

$$R := \theta_{\max} n_{\max} + \epsilon \theta_{\max} n_{\max}^2$$

we conclude

$$\mathbb{E}[\tilde{\theta}_{ii} | \mathcal{C}_i] = \frac{a_i + \mathcal{O}(\theta_{\max} R)}{d_{ii} + \mathcal{O}(R)}.$$

For

$$\delta := \left(\frac{\theta_{\max}}{\theta_{\min}} \right)^2 \left(\frac{n_{\max}}{n_{\min}} \right)^2 \max \left\{ \frac{1}{\theta_{\min} n_{\max}}, \frac{\epsilon}{\theta_{\min}} \right\}$$

and under the assumption $\delta \ll 1$ we can rewrite

$$\mathbb{E}[\tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j] = \begin{cases} \frac{a_{11}}{d_{11}} (1 + \mathcal{O}(\delta)) & , i, j \in \mathcal{C}_1^* \\ \frac{a_{22}}{d_{22}} (1 + \mathcal{O}(\delta)) & , i, j \in \mathcal{C}_2^* \\ \frac{a_{12}}{d_{12}} (1 + \mathcal{O}(\delta)) & , \text{otherwise.} \end{cases}$$

Let us consider the case $i \in \alpha_j$. W.l.o.g. let us assume $\alpha_i = 1$. Then

$$\begin{aligned} \mathbb{E}[\tilde{\theta}_{ii} - \tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j] &= \frac{a_{11}}{d_{11}} (1 + \mathcal{O}(\delta)) - \frac{a_{11}}{d_{11}} (1 + \mathcal{O}(\delta)) \\ &= \mathcal{O}(\delta \theta_{\max}). \end{aligned}$$

Since $\tilde{\theta}_{i \vee j}$ is the weighted mean of $\tilde{\theta}_{ii}$, $\tilde{\theta}_{jj}$ and $\tilde{\theta}_{ij}$, we also conclude

$$\mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{ij} | \mathcal{C}_i, \mathcal{C}_j] = \mathcal{O}(\delta \theta_{\max}). \quad (49)$$

Next, we consider the other case $i \notin \alpha_i$. W.l.o.g. let us assume $\alpha_i = 1$ and $\alpha_j = 2$. We compute

$$\left(\frac{a_{11}}{d_{11}} - \frac{a_{22}}{d_{22}} \right) + \left(\frac{a_{11}}{d_{11}} - \frac{a_{12}}{d_{12}} \right) = \frac{x}{(\theta_1 n_1 + \rho n_2)^2 (\theta_2 n_2 + \rho n_1)^2}$$

with

$$\begin{aligned} x &= 2a_{11}(\theta_2 n_2 + \rho n_1)^2 - a_{22}(\theta_1 n_1 + \rho n_2)^2 - a_{12}(\theta_1 n_1 + \rho n_2)(\theta_2 n_2 + \rho n_1) \\ &= n_1 n_2 (\theta_1 \theta_2 - \rho^2) [(\theta_2 - \rho)(\theta_1^2 - \rho^2) n_1 n_2 + R_{ij}] \end{aligned}$$

for

$$R_{ij} = 3\rho(\theta_2 - \rho)(\theta_1 n_1^2 - \theta_2 n_2^2) + (\theta_1 - \theta_2) [n_1 n_2 (\theta_1 \theta_2 + \theta_1 \rho + \rho^2) + 3\rho \theta_1 n_1^2].$$

Note that

$$\begin{aligned} R_{ij} + R_{ji} &= 3\rho(\theta_2 - \theta_1)(\theta_1 n_1^2 - \theta_2 n_2^2) + (\theta_1 - \theta_2) [\rho n_1 n_2 (\theta_1 - \theta_2) + 3\rho(\theta_1 n_1^2 - \theta_2 n_2^2)] \\ &= \rho n_1 n_2 (\theta_1 - \theta_2)^2 \\ &\geq 0. \end{aligned}$$

So w.l.o.g. (after possibly exchanging the two indices) we can assume

$$\begin{aligned} \left(\frac{a_{11}}{d_{11}} - \frac{a_{22}}{d_{22}} \right) + \left(\frac{a_{11}}{d_{11}} - \frac{a_{12}}{d_{12}} \right) &\geq \frac{(\theta_1 \theta_2 - \rho^2)(\theta_1 - \rho)(\theta_2 - \rho)(\theta_1 + \rho)n_1^2 n_2^2}{(\theta_1 n_1 + \rho n_2)^2 (\theta_2 n_2 + \rho n_1)^2} \\ &\gtrsim \left(\frac{\theta_{\min}}{\theta_{\max}} \right)^2 \left(\frac{n_{\min}}{n_{\max}} \right)^2 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2} \end{aligned}$$

and thus

$$\begin{aligned} &\mathbb{E}[(\tilde{\theta}_{ii} - \tilde{\theta}_{jj}) + (\tilde{\theta}_{ii} - \tilde{\theta}_{ij}) | \mathcal{C}_i, \mathcal{C}_j] \\ &= \left(\frac{a_{11}}{d_{11}} (1 + \mathcal{O}(\delta)) - \frac{a_{22}}{d_{22}} (1 + \mathcal{O}(\delta)) \right) + \left(\frac{a_{11}}{d_{11}} (1 + \mathcal{O}(\delta)) - \frac{a_{12}}{d_{12}} (1 + \mathcal{O}(\delta)) \right) \\ &\gtrsim \left(\frac{\theta_{\min}}{\theta_{\max}} \right)^2 \left(\frac{n_{\min}}{n_{\max}} \right)^2 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2} + \mathcal{O}(\delta \theta_{\max}). \end{aligned}$$

Since $\tilde{\theta}_{i \vee j}$ is the weighted mean of $\tilde{\theta}_{ii}$, $\tilde{\theta}_{jj}$ and $\tilde{\theta}_{ij}$, there exists indices $(i', j') \in \{(i, i), (j, j), (i, j)\}$ such that

$$\mathbb{E}[\tilde{\theta}_{i \vee j} - \tilde{\theta}_{i' j'} | \mathcal{C}_i, \mathcal{C}_j] \gtrsim \left(\frac{\theta_{\min}}{\theta_{\max}} \right)^2 \left(\frac{n_{\min}}{n_{\max}} \right)^2 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2} + \mathcal{O}(\delta \theta_{\max}). \quad (50)$$

Similar as in the proof of Theorem 1, we deduce from our assumptions $\theta_{\min}^3 n_{\min}^2 \gg \log n_{\min}$ and $\theta_{\max} \leq \frac{1}{2}$ as well as Lemma 2 that

- the estimates $\tilde{\theta}$. (including $\tilde{\theta}_{ii}$, $\tilde{\theta}_{jj}$, $\tilde{\theta}_{ij}$ and $\tilde{\theta}_{i \vee j}$) are with large probability inside an interval of the form $[\frac{\rho}{2}, \theta_{\max} + \frac{\rho}{2}]$, implying that the corresponding Fisher information of a Bernoulli variable with such a mean is up to bounded constants bounded from below by $\theta_{\max}^{-1}$ and
- with large probability,

$$|\tilde{\theta} - \mathbb{E}[\tilde{\theta} | \mathcal{C}_i, \mathcal{C}_j]| \lesssim (\log n_{\min})^{\frac{1}{2}} \theta_{\min}^{-\frac{1}{2}} n_{\min}^{-1}.$$

Combining the above with the quadratic Taylor expansion of the Kullback-Leibler and (50), we conclude in the case where i and j belong to different communities

$$\begin{aligned} \theta_{\max}^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\tilde{\theta}_{i' j'}, \tilde{\theta}_{i \vee j}) &\gtrsim |\tilde{\theta}_{i' j'} - \tilde{\theta}_{i \vee j}| \\ &\geq \left| \mathbb{E}[\tilde{\theta}_{i' j'} | \mathcal{C}_i, \mathcal{C}_j] - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j] \right| - \left| \tilde{\theta}_{i' j'} - \mathbb{E}[\tilde{\theta}_{i' j'} | \mathcal{C}_i, \mathcal{C}_j] \right| - \left| \tilde{\theta}_{i \vee j} - \mathbb{E}[\tilde{\theta}_{i \vee j} | \mathcal{C}_i, \mathcal{C}_j] \right| \\ &\gtrsim \left(\frac{\theta_{\min}}{\theta_{\max}} \right)^2 \left(\frac{n_{\min}}{n_{\max}} \right)^2 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2} + \mathcal{O} \left(\delta \theta_{\max} + (\log n_{\min})^{\frac{1}{2}} \theta_{\min}^{-\frac{1}{2}} n_{\min}^{-1} \right), \end{aligned}$$

whereas in the other case we conclude analogously from (49)

$$\theta_{\max}^{\frac{1}{2}} \mathcal{K}^{\frac{1}{2}}(\tilde{\theta}, \tilde{\theta}_{i \vee j}) \lesssim \delta \theta_{\max} + (\log n_{\min})^{\frac{1}{2}} \theta_{\min}^{-\frac{1}{2}} n_{\min}^{-1}.$$

Recall that the concentration result $N = d \cdot (1 + \mathcal{O}(\delta))$ and with $\delta \ll 1$ and note that d_{ii} , d_{jj} and d_{ij} differ only by a factor $\lesssim \left(\frac{\theta_{\max}}{\theta_{\min}} \right)^2 \left(\frac{n_{\max}}{n_{\min}} \right)^2$. We conclude that to show the consistency of the test under a proper choice of the threshold, it will suffice to check

$$\begin{aligned} \delta \theta_{\max} &\ll \left(\frac{\theta_{\min}}{\theta_{\max}} \right)^4 \left(\frac{n_{\min}}{n_{\max}} \right)^4 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2} \\ \Leftrightarrow \max \left\{ \frac{1}{\theta_{\min} n_{\max}}, \frac{\epsilon}{\theta_{\min}} \right\} &\ll \left(\frac{\theta_{\min}}{\theta_{\max}} \right)^6 \left(\frac{n_{\min}}{n_{\max}} \right)^6 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^3} \end{aligned} \quad (51)$$

and

$$\sqrt{\frac{\log n_{\min}}{\theta_{\min} n_{\min}^2}} \ll \left(\frac{\theta_{\min}}{\theta_{\max}}\right)^4 \left(\frac{n_{\min}}{n_{\max}}\right)^4 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2}, \quad (52)$$

while also considering the large probability assumption

$$\min \left\{ \frac{\epsilon^2 n_{\min}}{\theta_{\max} \log n_{\min}}, \frac{\epsilon n_{\min}}{\log n_{\min}} \right\} \gg 1. \quad (53)$$

An $\epsilon > 0$ satisfying (42) and (44) exists as long as

$$\frac{\log n_{\min}}{n_{\min}} + \sqrt{\frac{\theta_{\max} \log n_{\min}}{n_{\min}}} \ll \left(\frac{\theta_{\min}}{\theta_{\max}}\right)^7 \left(\frac{n_{\min}}{n_{\max}}\right)^6 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2}, \quad (54)$$

whereas the ϵ -independent part of conditions (42) and (43) is implied by

$$\left(\frac{\theta_{\max}}{\theta_{\min}}\right) \frac{1}{n_{\max}} + \sqrt{\frac{\log n_{\min}}{\theta_{\min} n_{\min}^2}} \ll \left(\frac{\theta_{\min}}{\theta_{\max}}\right)^6 \left(\frac{n_{\min}}{n_{\max}}\right)^6 \frac{(\theta_{\max} - \rho)(\theta_{\min} - \rho)^2}{\theta_{\max}^2}. \quad (55)$$

By simple calculus we can verify that our assumption

$$\theta_{\min} - \rho \gg \left(\frac{\theta_{\max}}{\theta_{\min}}\right)^{\frac{7}{3}} \left(\frac{n_{\max}}{n_{\min}}\right)^2 \max \left\{ n_{\min}^{-\frac{1}{3}} \theta_{\max}^{\frac{1}{2}} (\log n_{\min})^{\frac{1}{6}}, n_{\min}^{-\frac{1}{6}} \theta_{\max}^{\frac{5}{6}} (\log n_{\min})^{\frac{1}{6}} \right\}$$

implies conditions (54) and (55). $\square$

A Equivalent formulations and $\log_n$ - $\log_n$ plot for rates of consistency

Recall that for $k = 1$, Theorem 1 guarantees strong consistency as long as

$$\theta - \rho \gg \max \{ n^{-\frac{1}{3}} \theta^{\frac{1}{2}} (\log n)^{\frac{1}{6}}, n^{-\frac{1}{6}} \theta^{\frac{5}{6}} (\log n)^{\frac{1}{6}} \}. \quad (56)$$

Considering $\theta - \rho < \theta$, condition (56) implies $\theta \gg n^{-\frac{2}{3}} (\log n)^{\frac{1}{3}}$. This is equivalent to (56) in case $\frac{\theta}{\rho} \equiv C$. Furthermore, in view of

$$\begin{aligned} n^{-\frac{1}{3}} \theta^{\frac{1}{2}} (\log n)^{\frac{1}{6}} &> n^{-\frac{1}{6}} \theta^{\frac{5}{6}} (\log n)^{\frac{1}{6}} \\ \Leftrightarrow \theta &< n - \frac{1}{2} \end{aligned}$$

we can rewrite the consistency condition (56) as

$$\theta - \rho \gg \begin{cases} n^{-\frac{1}{6}} \theta^{\frac{5}{6}} (\log n)^{\frac{1}{6}} & , n^{-\frac{1}{2}} \lesssim \theta < \frac{1}{2} \\ n^{-\frac{1}{3}} \theta^{\frac{1}{2}} (\log n)^{\frac{1}{6}} & , n^{-\frac{2}{3}} (\log n)^{\frac{1}{3}} \lesssim \theta \lesssim n^{-\frac{1}{2}}, \end{cases}$$

or equivalently, using the notation $\theta = n^\gamma$

$$\begin{aligned} \log_n(\theta - \rho) - \frac{c}{\log n} &\geq \\ \begin{cases} \frac{5\gamma-1}{6} + \frac{1}{6} \log_n(\log n) & , -\frac{1}{2} + \mathcal{O}((\log n)^{-1}) \leq \gamma < 0 + \log_n\left(\frac{1}{2}\right) \\ \frac{3\gamma-2}{6} + \frac{1}{6} \log_n(\log n) & , -\frac{2}{3} + \frac{1}{3} \log_n(\log n) \leq \gamma + \mathcal{O}((\log n)^{-1}) \leq -\frac{1}{2} \end{cases} \end{aligned}$$

for a large enough constant $c \in \mathbb{R}$. Up to the logarithmic terms (which are vanishing for $n \rightarrow \infty$), this can be visualized in a plot of $\log_n \left(\frac{\theta - \rho}{\theta} \right)$ versus $\log_n(\theta)$ by the area of a polygon consisting of the following points:

$$\begin{aligned} & (0, 0) \\ & \left(0, -\frac{1}{6} \right) \\ & \left(-\frac{1}{2}, -\frac{1}{12} \right) \\ & \left(-\frac{2}{3}, 0 \right) \end{aligned}$$

Using the original algorithm AWCD^o instead, we have the sufficient condition for strong consistency

$$\theta - \rho \gg \max\{n^{-\frac{1}{3}}\theta^{\frac{1}{3}}, n^{-\frac{1}{6}}\theta^{\frac{5}{6}}(\log n)^{\frac{1}{6}}\}. \quad (57)$$

Considering $\theta - \rho < \theta$, condition (57) implies $\theta \gg n^{-\frac{1}{2}}$. This is equivalent to (57) in case $\frac{\theta}{\rho} \equiv C$. Furthermore, because of

$$\begin{aligned} n^{-\frac{1}{3}}\theta^{\frac{1}{3}} &> n^{-\frac{1}{6}}\theta^{\frac{5}{6}}(\log n)^{\frac{1}{6}} \\ \Leftrightarrow \theta &< n^{-\frac{1}{3}}(\log n)^{-\frac{1}{3}} \end{aligned}$$

we can rewrite the consistency condition (57) as

$$\theta - \rho \gg \begin{cases} n^{-\frac{1}{6}}\theta^{\frac{5}{6}}(\log n)^{\frac{1}{6}} & , n^{-\frac{1}{3}}(\log n)^{-\frac{1}{3}} \lesssim \theta < \frac{1}{2} \\ n^{-\frac{1}{3}}\theta^{\frac{1}{3}} & , n^{-\frac{1}{2}} \lesssim \theta \lesssim n^{-\frac{1}{3}}(\log n)^{-\frac{1}{3}} \end{cases}$$

or equivalently, using the notation $\theta = n^\gamma$

$$\begin{aligned} \log_n(\theta - \rho) - \frac{c}{\log n} &\geq \\ \begin{cases} \frac{5\gamma-1}{6} + \frac{1}{6} \log_n(\log n) & , -\frac{1}{3} - \frac{1}{3} \log_n(\log n) + \mathcal{O}((\log n)^{-1}) \leq \gamma < 0 + \log_n\left(\frac{1}{2}\right) \\ \frac{\gamma-1}{3} & , -\frac{1}{2} \leq \gamma + \mathcal{O}((\log n)^{-1}) \leq -\frac{1}{3} - \frac{1}{3} \log_n(\log n). \end{cases} \end{aligned}$$

Again ignoring the logarithmic terms, this can be visualized in a plot of $\log_n \left(\frac{\theta - \rho}{\theta} \right)$ versus $\log_n(\theta)$ by the area of a polygon consisting of the following points:

$$\begin{aligned} & (0, 0) \\ & \left(0, -\frac{1}{6} \right) \\ & \left(-\frac{1}{3}, -\frac{1}{9} \right) \\ & \left(-\frac{1}{2}, 0 \right) \end{aligned}$$

Next, we discuss the case $k \geq 2$, starting with main version AWCD. Recall from Theorem 2 the sufficient consistency condition

$$\theta - \rho \gg \max\{A, B, C\}$$

for

$$\begin{aligned} A &= \theta^{\frac{3k+1}{2k+1}} n^{\frac{k-1}{2k+1}}, \\ B &= \theta^{\frac{4k+1}{4k+2}} n^{-\frac{1}{4k+2}} (\log n)^{\frac{1}{4k+2}} \text{ and} \\ C &= \theta^{\frac{2k-1}{4k+2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{4k+2}}. \end{aligned}$$

Since $\frac{B}{A} = \theta^{-\frac{1}{2}} n^{-\frac{2k-1}{4k+2}} (\log n)^{\frac{1}{4k+2}}$, we have

$$B > A \Leftrightarrow \theta < n^{-\frac{2k-1}{2k+1}} (\log n)^{\frac{1}{2k+1}}.$$

Since $\frac{C}{B} = \theta^{-\frac{k+1}{2k+1}} n^{-\frac{k}{2k+1}}$, we have

$$C > B \Leftrightarrow \theta < n^{-\frac{k}{k+1}}.$$

Furthermore, $\theta \gg A$ implies the upper bound

$$\theta \ll n^{-\frac{k-1}{k}}$$

and $\theta \gg C$ implies the lower bound

$$\theta \gg n^{-\frac{2k+1}{2k+3}} (\log n)^{\frac{1}{2k+3}}.$$

This allows us to rewrite the consistency condition as

$$\theta - \rho \gg \begin{cases} A & , n^{-\frac{2k-1}{2k+1}} (\log n)^{\frac{1}{2k+1}} \lesssim \theta \lesssim n^{-\frac{k-1}{k}} \\ B & , n^{-\frac{k}{k+1}} \lesssim \theta \lesssim n^{-\frac{2k-1}{2k+1}} (\log n)^{\frac{1}{2k+1}} \\ C & , n^{-\frac{2k+1}{2k+3}} (\log n)^{\frac{1}{2k+3}} \lesssim \theta \lesssim n^{-\frac{k}{k+1}} \end{cases}$$

for

$$n^{-\frac{2k+1}{2k+3}} (\log n)^{\frac{1}{2k+3}} \ll \theta \ll n^{-\frac{k-1}{k}}.$$

Note that the latter is in fact equivalent to the consistency condition in case $\frac{\theta}{\rho} \equiv C$. Using the notation

$$\theta = n^\gamma,$$

we can write equivalently

$$\log_n(\theta - \rho) - \frac{c}{\log n} \geq \begin{cases} \frac{\gamma(3k+1)+k-1}{2k+1} & , \gamma + \mathcal{O}((\log n)^{-1}) \in [-\frac{2k-1}{2k+1} + \frac{1}{2k+1} \log_n(\log n), -\frac{k-1}{k}] \\ \frac{\gamma(4k+1)-1}{4k+2} + \frac{1}{4k+2} \log_n(\log n) & , \gamma + \mathcal{O}((\log n)^{-1}) \in [-\frac{k}{k+1}, -\frac{2k-1}{2k+1} + \frac{1}{2k+1} \log_n(\log n)] \\ \frac{\gamma(2k-1)}{4k+2} - \frac{1}{2} + \frac{1}{4k+2} \log_n(\log n) & , \gamma + \mathcal{O}((\log n)^{-1}) \in [-\frac{2k+1}{2k+3} + \frac{1}{2k+3} \log_n(\log n), -\frac{k}{k+1}] \end{cases}$$

for a large enough constant $c \in \mathbb{R}$. Up to the logarithmic terms (which are vanishing for $n \rightarrow \infty$), we can visualize the consistency regime in a plot of $\log_n(\frac{\theta-\rho}{\theta})$ versus $\log_n(\theta)$ by the area of a polygon

consisting of the following points:

$$\begin{aligned} & \left(-\left(\frac{k-1}{k} \right), 0 \right) \\ & \left(-\left(\frac{2k-1}{2k+1} \right), -\left(\frac{1}{(2k+1)^2} \right) \right) \\ & \left(-\left(\frac{k}{k+1} \right), -\frac{1}{(4k+2)(k+1)} \right) \\ & \left(-\left(\frac{2k+1}{2k+3} \right), 0 \right) \end{aligned}$$

Next, we consider AWCD^+ . Recall the consistency condition

$$\theta - \rho \gg \max\{A, B, C^+\}$$

for

$$C^+ = \theta^{\frac{1}{2}} n^{-\frac{k}{2k+1}} (\log n)^{\frac{1}{4k+2}}.$$

Since $\frac{C^+}{B} = \theta^{-\frac{k}{2k+1}} n^{-\frac{2k-1}{4k+2}}$ we have

$$C^+ > B \Leftrightarrow \theta < n^{-\frac{2k-1}{2k}}.$$

Moreover, $\theta \gg C^+$ implies the lower bound

$$\theta \gg n^{-\frac{2k}{2k+1}} (\log n)^{\frac{1}{2k+1}}.$$

The condition

$$n^{-\frac{k}{2k+1}} \ll \theta \ll n^{-\frac{k-1}{k}}$$

is in fact equivalent to the final sufficient condition above in case of $\frac{\theta}{\rho} \equiv C$. Moreover, according to the above, the consistency condition can be rewritten as

$$\theta - \rho \gg \begin{cases} A & , n^{-\frac{2k-1}{2k+1}} (\log n)^{\frac{1}{2k+1}} \lesssim \theta \lesssim n^{-\frac{k-1}{k}} \\ B & , n^{-\frac{2k-1}{2k}} \lesssim \theta \lesssim n^{-\frac{2k-1}{2k+1}} (\log n)^{\frac{1}{2k+1}} \\ C^+ & , n^{-\frac{2k}{2k+1}} (\log n)^{\frac{1}{2k+1}} \lesssim \theta \lesssim n^{-\frac{2k-1}{2k}}. \end{cases}$$

Using the notation

$$\theta = n^\gamma,$$

we can write equivalently

$$\log_n(\theta - \rho) - \frac{c}{\log n} \geq \begin{cases} \frac{\gamma(3k+1)+k-1}{2k+1} & , \gamma + \mathcal{O}((\log n)^{-1}) \in \left[-\frac{2k-1}{2k+1} + \frac{1}{2k+1} \log_n(\log n), -\frac{k-1}{k} \right] \\ \frac{\gamma(4k+1)-1}{4k+2} + \frac{1}{4k+2} \log_n(\log n) & , \gamma + \mathcal{O}((\log n)^{-1}) \in \left[-\frac{2k-1}{2k}, -\frac{2k-1}{2k+1} + \frac{1}{2k+1} \log_n(\log n) \right] \\ \frac{\gamma}{2} - \frac{k}{2k+1} + \frac{1}{4k+2} \log_n(\log n) & , \gamma + \mathcal{O}((\log n)^{-1}) \in \left[-\frac{2k}{2k+1} + \frac{1}{2k+1} \log_n(\log n), -\frac{2k-1}{2k} \right]. \end{cases}$$

The corresponding polygon in a plot of $\log_n \left(\frac{\theta - \rho}{\theta} \right)$ versus $\log_n(\theta)$ is defined by the following points:

$$\begin{aligned} & \left(-\left(\frac{k-1}{k} \right), 0 \right) \\ & \left(-\left(\frac{2k-1}{2k+1} \right), -\left(\frac{1}{(2k+1)^2} \right) \right) \\ & \left(-\left(\frac{2k-1}{2k} \right), -\frac{1}{2k(4k+2)} \right) \\ & \left(-\left(\frac{2k}{2k+1} \right), 0 \right) \end{aligned}$$

Lastly, we also discuss the version AWCD° without bias correction. Recall the consistency condition

$$\theta - \rho \gg \max\{A, B, C^\circ\}$$

for

$$C^\circ = \theta^{\frac{k}{2k+1}} n^{-\frac{k}{2k+1}}.$$

Since $\frac{C^\circ}{B} = \theta^{-\frac{1}{2}} n^{-\frac{2k-1}{4k+2}} (\log n)^{-\frac{1}{4k+2}}$, we have

$$C^\circ > B \Leftrightarrow \theta < n^{-\frac{2k-1}{2k+1}} (\log n)^{-\frac{1}{2k+1}}.$$

Moreover, $\theta \gg C^\circ$ implies the lower bound

$$\theta \gg n^{-\frac{k}{k+1}}.$$

The condition

$$n^{-\frac{k}{2k+1}} \ll \theta \ll n^{-\frac{k-1}{k}}$$

is equivalent to the final sufficient condition above in case of $\frac{\theta}{\rho} \equiv C$. Moreover, we can rewrite the consistency condition as

$$\theta - \rho \gg \begin{cases} A & , n^{-\frac{2k-1}{2k+1}} (\log n)^{\frac{1}{2k+1}} \lesssim \theta \lesssim n^{-\frac{k-1}{k}} \\ B & , n^{-\frac{2k-1}{2k+1}} (\log n)^{-\frac{1}{2k+1}} \lesssim \theta \lesssim n^{-\frac{2k-1}{2k+1}} (\log n)^{\frac{1}{2k+1}} \\ C^\circ & , n^{-\frac{k}{k+1}} \lesssim \theta \lesssim n^{-\frac{2k-1}{2k+1}} (\log n)^{-\frac{1}{2k+1}}. \end{cases}$$

The corresponding polygon in a plot of $\log_n \left(\frac{\theta - \rho}{\theta} \right)$ versus $\log_n(\theta)$ is defined by the following points:

$$\begin{aligned} & \left(-\left(\frac{k-1}{k} \right), 0 \right) \\ & \left(-\left(\frac{2k-1}{2k+1} \right), -\left(\frac{1}{(2k+1)^2} \right) \right) \\ & \left(-\left(\frac{k}{k+1} \right), 0 \right) \end{aligned}$$

References

- [1] E. Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.
- [2] L. Adamyan, K. Efimov, and V. Spokoiny. Adaptive nonparametric community detection. *Discussion Papers by IRTG 1792*, 2019-006, 2019.
- [3] A. Arenas, J. Duch, A. Fernández, and S. Gómez. Size reduction of complex networks preserving modularity. *New Journal of Physics*, 9(6):176, 2007.
- [4] J. P. Bagrow. Evaluating local community methods in networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(05):P05001, may 2008.
- [5] E. R. Barnes. An algorithm for partitioning the nodes of a graph. *SIAM Journal on Algebraic Discrete Methods*, 3(4):541–550, 1982.
- [6] P. J. Bickel and A. 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.
- [7] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008, 04 2008.
- [8] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70:066111, Dec 2004.
- [9] V. Cohen-Addad, A. Kosowski, F. Mallmann-Trenn, and D. Saulpic. On the power of louvain in the stochastic block model. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4055–4066. Curran Associates, Inc., 2020.
- [10] Y. Dourisboure, F. Geraci, and M. Pellegrini. Extraction and classification of dense implicit communities in the web graph. *ACM Trans. Web*, 3(2), apr 2009.
- [11] K. Efimov, L. Adamyan, and V. Spokoiny. Adaptive nonparametric clustering. *IEEE Trans. Inform. Theory*, 65(8):4875–4892, 2019.
- [12] E. Ferrara. A large-scale community structure analysis in facebook. *EPJ Data Science*, 1:1–30, 11 2012.
- [13] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305, 1973.
- [14] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.
- [15] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Community detection in degree-corrected block models. *The Annals of Statistics*, 46(5):2153–2185, 2018.
- [16] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.
- [17] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airolidi. A survey of statistical network models. *Found. Trends Mach. Learn.*, 2(2):129–233, feb 2010.

- [18] R. Guimera and L. A. Nunes Amaral. Functional cartography of complex metabolic networks. *nature*, 433(7028):895–900, 2005.
- [19] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
- [20] B. H. Junker and F. Schreiber. Analysis of biological networks. 2008.
- [21] B. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *AT&T Technical Journal*, 49(2):291–307, 1970.
- [22] O. Klopp, A. B. Tsybakov, and N. Verzelen. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316–354, 2017.
- [23] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.
- [24] M. E. Newman. Analysis of weighted networks. *Physical review E*, 70(5):056131, 2004.
- [25] M. E. Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133, 2004.
- [26] M. E. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.
- [27] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- [28] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.
- [29] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *nature*, 435(7043):814–818, 2005.
- [30] M. Pensky. Dynamic network models and graphon estimation. *The Annals of Statistics*, 47(4):2378–2403, 2019.
- [31] J. Polzehl and V. Spokoiny. Propagation-separation approach for local likelihood estimation. *Probab. Theory Related Fields*, 135(3):335–362, 2006.
- [32] A. Pothen, H. D. Simon, and K.-P. Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM journal on matrix analysis and applications*, 11(3):430–452, 1990.
- [33] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.
- [34] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [35] H. Shen, X. Cheng, K. Cai, and M.-B. Hu. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 388(8):1706–1712, 2009.
- [36] X. Su, S. Xue, F. Liu, J. Wu, J. Yang, C. Zhou, W. Hu, C. Paris, S. Nepal, D. Jin, et al. A comprehensive survey on community detection with deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

- [37] A. L. Traud, P. J. Mucha, and M. A. Porter. Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180, 2012.
- [38] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer.
- [39] A. Wickramasinghe and S. Muthukumarana. Social network analysis and community detection on spread of covid-19. *Model Assisted Statistics and Applications*, 16:37–52, 03 2021.
- [40] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.
- [41] A. Y. Zhang and H. H. Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.