

# ESTIMATION AND INFERENCE FOR STOCHASTIC BLOCK MODELS

Koen Jochmans\*

Toulouse School of Economics, University of Toulouse Capitole

This version: March 9, 2022

## Abstract

This paper is concerned with nonparametric estimation of the weighted stochastic block model. We first show that the model implies a set of multilinear restrictions on the joint distribution of edge weights of certain subgraphs involving (in its simplest form) triplets and quadruples of nodes. From this system of equations the unknown components of the model can be recovered nonparametrically, up to the usual labeling ambiguity. We introduce a simple and computationally-attractive manner to do this. Estimators then follow from the analogy principle. Limit theory is provided. We find that component distributions and their functionals, as well as their density functions (for the case where edge weights are continuous) are all estimable at the parametric rate. Numerical experiments are reported on.

**Keywords:** heterogeneity, network, mixture model, random graph, stochastic block model

## 1 Introduction

The stochastic block model provides a parsimonious way to incorporate latent heterogeneity in the analysis of network data. The original application of [Holland, Laskey and Leinhardt 1983](#) concerned the binary decision of edge formation between two nodes and generalizes the Erdős-Rényi random-graph model ([Erdős and Rényi 1959](#)). In the latter model, edges are formed independently with a common probability. In the former, the set of nodes is partitioned into a finite set of latent communities, and the link probability between two nodes depends on the communities that they belong to. The latent block structure has

---

\*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: [koen.jochmans@tse-fr.eu](mailto:koen.jochmans@tse-fr.eu).

since been used to study general (discrete or continuous) outcomes generated from pairwise interaction ([Hoff, Raftery and Handcock 2002](#)), thereby extending the applicability of the stochastic block model to weighted graphs.

There is now a large literature on the estimation of the stochastic block model. [Lei \(2016\)](#), [Wang and Bickel \(2017\)](#), [Yan, Sarkar and Cheng \(2018\)](#), and [Le and Levina \(2019\)](#) provide techniques to estimate and test the number of communities. Taking this number as given, [Snijders and Nowicki \(1997\)](#), [Nowicki and Snijders \(2001\)](#), and [Amini, Chen, Bickel and Levina \(2013\)](#) consider likelihood-based estimation of the remaining parameters of the model. Given the well-known computational complexity of this approach variational methods have also been considered ([Daudin, Picard and Robin 2008](#), [Mariadassou, Robin and Vacher 2010](#)). For the binary edge-weight case statistical properties under growth rates on the average degree have been derived by [Celisse, Daudin and Pierre \(2012\)](#) and [Bickel, Choi, Chang and Zhang \(2013\)](#). Related results are also available for approaches based on spectral clustering ([Rohe, Chatterjee and Yu 2011](#), [Sussman, Tang, Fishkind and Priebe 2012](#), [Tang, Cape and Priebe 2022](#)).

This paper is concerned with estimation and inference for the weighted stochastic block model. We present a set of multilinear equations from which all the unknown parameters—the number of communities, their size, and the conditional distributions of the edge weights—can be uniquely recovered. These restrictions involve the probability distributions of edge weights in small subgraphs of certain configurations. As they hold without imposing any parametric structure, our equations can be used as a basis for the construction of a fully nonparametric estimator of the stochastic block model. It would also be possible to complement our approach with (semi-) parametric restrictions, although we do not focus on this here.

The estimation strategy we lay out below is computationally attractive. It is built around a joint (approximate) diagonalization step tailored to our setup. This type of routine has found applicability elsewhere ([Cardoso and Souloumiac 1993](#), [Bonhomme, Jochmans and Robin 2016](#)). Here we use it as an auxiliary first-step estimator in the construction of our main estimators of the components of the model. Moreover, once it has been computed,

our estimators of the stochastic block model are least-squares estimators and are thus immediate to compute. We present an estimator of the distribution of the communities as well as a generic estimator of linear functionals of the conditional distributions. The latter covers (cumulative) distribution functions, their moments, and probability mass functions, for example. We also give results for a kernel estimator for conditional densities for the case where edge weights are continuous.

Limit theory is presented under an asymptotic scheme where the number of nodes in the network,  $n$ , goes to infinity, assuming that the number of communities is known. Under weak regularity conditions they converge in distribution to correctly-centered normal random variables at the rate  $n^{-1/2}$ . Interestingly, this (parametric) rate equally applies to density estimation, as the smoothing bias is small relative to the standard deviation. This is due to the strong dependence induced by the community structure. Undersmoothing is not needed to achieve this result.

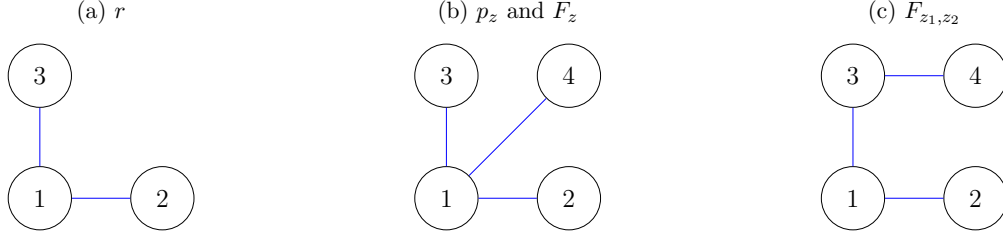
Finally, because our estimators involve averages over only triplets and quadruples of nodes, the conditions underlying our limit results do not impose requirements on network denseness through, for example, (functionals of) the degree distribution. Indeed, they do not attempt to assign nodes to communities. Therefore, our techniques can easily be adapted to a setting where we observe many (possible small) networks generated by the same block structure.

## 2 Stochastic block model

Consider a graph involving  $n$  nodes where the set of nodes is partitioned into  $r$  latent communities, labelled  $1, \dots, r$ . Each node is first assigned to a community independently according to some probability distribution  $\mathbf{p} = (p_1, \dots, p_r)'$ . The community of node  $i$  is recorded in the latent variable  $Z_i$ . Thus,

$$\mathbb{P}(Z_i = z) = p_z > 0 \text{ for each } 1 \leq z \leq r$$

Figure 1: Subgraph configurations used for identification



and is equal to zero otherwise. Next, each unordered pair of nodes  $i \neq j$  draws a real-valued weighted edge  $X_{i,j}$  from some distribution that depends on the communities they belong to,

$$F_{z_1, z_2}(x) := \mathbb{P}(X_{i,j} \leq x | Z_i = z_1, Z_j = z_2).$$

The edge weights  $X_{i,j}$  are independent conditional on the community indicators  $(Z_i, Z_j)$ .

## 2.1 Multilinear restrictions

We first show that it is possible to recover the parameters of the stochastic block model from the distribution of edge weights of subgraphs involving as little as four nodes. To do so we let

$$F_z(x) := \mathbb{P}(X_{i,j} \leq x | Z_i = z) = \sum_{z'=1}^r F_{z,z'}(x) p_{z'}$$

and impose the following condition.

**Assumption 1.** *The functions  $F_1, \dots, F_r$  are linearly independent.*

Rank conditions as this one arise frequently in the analysis of multivariate latent-variable models.

We will prove the following theorem and specialize it further afterwards.

**Theorem 1.** *Suppose that Assumption 1 holds. Then,*

(i) *The number of communities  $r$  is nonparametrically recoverable from the distribution of edge weights in two-star subgraphs;*

(ii) *The distribution of communities,  $\mathbf{p}$ , as well as expectations of the form*

$$\varphi_{z_1, z_2} := \mathbb{E}(\varphi(X_{i,j}) | Z_i = z_1, Z_j = z_2)$$

*are nonparametrically recoverable (up to the labeling of the latent communities) from the distribution of edge weights in three-star subgraphs and the distribution of edge weights in path subgraphs on four nodes.*

Consider, first, the two-star subgraph in Figure 1(a), involving the edges between the three nodes 1, 2, 3. Observe that the edge weights  $X_{1,2}, X_{1,3}$  are independent conditional on  $Z_1$ . Furthermore, their (unconditional) distribution factors as

$$\sum_{z=1}^r p_z F_z \otimes F_z$$

This is a bivariate finite-mixture model. It follows from the work of [Kwon and Mbakop \(2021, Propositions 2.1 and 2.3\)](#) that  $r$  is identified under Assumption 1. See also [Kasahara and Shimotsu \(2014\)](#) for related results.

Next look at the three-star subgraph in Figure 1(b). The edge weights in this three-star graph— $X_{1,2}, X_{1,3}, X_{1,4}$ —are again independent conditional on  $Z_1$ . Hence, their tri-variate distribution again factors as the multivariate finite mixture

$$\sum_{z=1}^r p_z F_z \otimes F_z \otimes F_z.$$

From this, the identification of  $F_1, \dots, F_r$  and  $\mathbf{p} = (p_1, \dots, p_r)'$  (up to an arbitrary but common ordering) follows from [Bonhomme, Jochmans and Robin \(2016, Corollaries 1 and 2\)](#). See also [Allman, Matias and Rhodes \(2009\)](#) for a related result in multivariate mixture models.

Assumption 1 implies that there exists a finite integer  $l$  such that the  $l \times r$  matrix  $\mathbf{G}$ ,

$$(\mathbf{G})_{l', z} := \mathbb{E}(\alpha_{l'}(X_{i,j}) | Z_i = z),$$

for a set of transformation functions  $\alpha_1, \dots, \alpha_l$ , has full column rank. One choice for these transformation functions would be  $\alpha_{l'}(x) = \{x \leq x_{l'}\}$ , where  $x_1, \dots, x_l$  is a grid

of points and  $\{\cdot\}$  denotes the indicator function. In this case,  $(\mathbf{G})_{l',z} = F_z(x_{l'})$ . Other approximating functions such as orthogonal polynomials are also possible. Whatever the choice of functions, the matrix  $\mathbf{G}$  is identified because the  $F_z$  are and the columns of  $\mathbf{G}$  are linear functionals thereof. Moving on, use the joint distribution of the edge weights  $X_{1,2}, X_{1,3}, X_{3,4}$  from the path graph in Figure 1(c) to construct the  $l \times l$  matrix  $\mathbf{M}_\varphi$  with

$$(\mathbf{M}_\varphi)_{l_1, l_2} := \mathbb{E}(\alpha_{l_1}(X_{1,2}) \varphi(X_{1,3}) \alpha_{l_2}(X_{3,4})).$$

Noting that  $X_{1,2}, X_{1,3}, X_{3,4}$  are independent conditional on the pair  $(Z_1, Z_3)$ , we have that

$$\mathbf{M}_\varphi = \mathbf{G} \mathbf{H}_\varphi \mathbf{G}',$$

where

$$(\mathbf{H}_\varphi)_{z_1, z_2} := p_{z_1} p_{z_2} \varphi_{z_1, z_2}.$$

Because  $\mathbf{G}$  has maximal column rank  $\mathbf{G}'\mathbf{G}$  is invertible and, thus, we obtain

$$\mathbf{H}_\varphi = (\mathbf{G}'\mathbf{G})^{-1} \mathbf{G}' \mathbf{M}_\varphi \mathbf{G} (\mathbf{G}'\mathbf{G})^{-1}.$$

With the  $p_z$  already shown to be identified this result suffices to show the theorem. It is nonetheless useful to note that an application of the above argument to the constant function  $\varphi(x) = 1$  gives  $\mathbf{M}_1 = \mathbf{G} \mathbf{H}_1 \mathbf{G}'$  where  $(\mathbf{H}_1)_{z_1, z_2} := p_{z_1} p_{z_2}$ . It therefore follows that

$$\varphi_{z_1, z_2} = (\mathbf{H}_1)_{z_1, z_2}^{-1} (\mathbf{H}_\varphi)_{z_1, z_2},$$

which is a convenient result for the purpose of estimation. The proof of Theorem 1 is complete.

Part (ii) of Theorem 1 can be applied to  $\varphi(x) = \{x \leq x'\}$  for any chosen value  $x'$ , leading to our first proposition.

**Proposition 1.** *Suppose that Assumption 1 holds. Then the (weighted) stochastic block model is nonparametrically identified up to relabeling of the latent communities.*

Proposition 1 is to be contrasted with the existing identification results to date. For the unweighted model with two communities Allman, Matias and Rhodes (2009, Theorem

7) showed identification from the (joint distribution of) complete subgraphs involving 16 nodes. Assuming the number of communities is known, [Allman, Matias and Rhodes \(2011, Theorems 14 and 15\)](#) obtained results for the general model, but they rely on the complete subgraph on 9 nodes to do so. Their results further require conditions on the support of the edge weights relative to the number of communities being sufficiently large, in addition to  $F_{z_1, z_2}$  (for all  $z_1 < z_2$ ) being linearly independent.

The proof of Theorem 1 is remarkably simple. Yet it provides a constructive approach to estimation and we will consider this below. Note, also, that Theorem 1 is useful beyond as an input to establish Proposition 1 as it can be used to directly show identification—and to construct estimators—of linear functionals of the conditional distributions without the need to first estimate the latter.

## 2.2 A generalization

Assumption 1 cannot be satisfied when the edge weights can take on strictly less values than there are latent communities. In such a case identification can be obtained by looking at subgraphs involving a larger set of nodes. This illustrates the interplay between the richness of the support of the edge weights, the number of latent communities, and the size of the subgraphs that are needed to show identification.

The argument is based on chaining nodes in a particular manner. We let

$$F_z^q(x_1, \dots, x_q) := \mathbb{P}(X_{i, i_1} \leq x_1, X_{i, i_2} \leq x_2, \dots, X_{i, i_q} \leq x_q | Z_i = z),$$

where the indices  $i$  and  $i_1, \dots, i_q$  are all distinct, and impose the following rank requirement.

**Assumption 1’.** *There exists a finite integer  $q$  such that the functions  $F_1^q, \dots, F_r^q$  are linearly independent.*

**Proposition 1’.** *Let  $q_{\min}$  be the smallest integer for which Assumption 1’ holds. If  $q_{\min}$  is positive the (weighted) stochastic block model is nonparametrically identified up to relabeling of the latent communities.*

### 3 Nonparametric estimation

An estimator of the number of latent communities,  $r$ , can be constructed along the lines of [Kwon and Mbakop \(2021\)](#). Here we construct estimators of the distribution of the communities,  $\mathbf{p}$ , and the conditional distributions  $F_{z_1, z_2}$  and functionals thereof, building on the proof of [Theorem 1](#). We consider a setting where we observe data from a single network involving  $n$  nodes.

Our proposal is to proceed in two sequential steps. First, the matrix  $\mathbf{G}$  is estimated by a modification of the diagonalization estimator of [Bonhomme, Jochmans and Robin \(2016\)](#). This estimator,  $\hat{\mathbf{G}}$ , is detailed below. Next, we appeal to the analogy principle to construct our estimators of the components of the stochastic block model. An alternative to the joint diagonalization approach would be to estimate the  $(\mathbf{G})_{l', z} = \mathbb{E}(\alpha_{l'}(X_{i,j})|Z_i = z)$  using estimates of the  $F_z$ . These could be obtained by maximizing a parametric likelihood (using the EM algorithm, see [McLachlan and Peel 2000](#)), or by nonparametric procedures such as those given in [Levine, Hunter and Chauveau \(2011\)](#). A practical advantage of our proposal is that it bypasses estimation of the complete mixture model. A theoretical advantage (relative to other nonparametric estimators) is that distribution theory for the matrix  $\hat{\mathbf{G}}$  can be obtained by adapting the work of [Bonhomme, Jochmans and Robin \(2016\)](#) to deal with the network structure of the data.

If we write  $\mathbf{a} := (a_1, \dots, a_l)'$  for  $a_{l'} := \mathbb{E}(\alpha_{l'}(X_{i,j}))$  we have the univariate mixture representation

$$\mathbf{a} = \mathbf{G}\mathbf{p}.$$

Given  $\hat{\mathbf{G}}$ , a least-squares argument suggests estimating  $\mathbf{p}$  by

$$\hat{\mathbf{p}} := (\hat{\mathbf{G}}'\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}'\hat{\mathbf{a}},$$

for  $\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_l)'$ , with

$$\hat{a}_{l'} := \frac{2}{n(n-1)} \sum_{i < j} \alpha_{l'}(X_{i,j}).$$

This approach is inspired by [Titterton \(1983\)](#), where minimum-distance estimators of mixing proportions were considered.



Similarly,

$$\hat{\mathbf{H}}_\varphi := (\hat{\mathbf{G}}' \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}' \hat{\mathbf{M}}_\varphi \hat{\mathbf{G}} (\hat{\mathbf{G}}' \hat{\mathbf{G}})^{-1},$$

constructed with

$$(\hat{\mathbf{M}}_\varphi)_{l_1, l_2} := \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \alpha_{l_1}(X_{i_1, i_2}) \varphi(X_{i_2, i_3}) \alpha_{l_2}(X_{i_3, i_4}),$$

yields the estimator

$$\hat{\varphi}_{z_1, z_2} := (\hat{\mathbf{H}}_1)_{z_1, z_2}^{-1} (\hat{\mathbf{H}}_\varphi)_{z_1, z_2}$$

of  $\varphi_{z_1, z_2}$ .

Below we will present the sampling properties of these estimators under asymptotics where the number of nodes,  $n$ , grows large, assuming that the number of communities,  $r$ , is known. An alternative sampling scheme would be to sample  $m$  independent networks, each of size  $n$  and generated from the same stochastic block model. Under asymptotics where  $m \rightarrow \infty$  while  $n$  remains fixed, our estimators achieve the parametric rate of  $m^{-1/2}$  under the same regularity conditions as the ones introduced here. We omit further details for this case for brevity.

### 3.1 Diagonalization step

We construct the matrix  $\hat{\mathbf{G}}$  by relying on an (approximate) simultaneous-diagonalization argument. We summarize the procedure here and refer to [Bonhomme, Jochmans and Robin \(2016\)](#) for additional details on this approach in (stationary) multivariate mixture models. We begin by constructing the  $l \times l$  matrix  $\hat{\mathbf{A}}_0$ ,

$$(\hat{\mathbf{A}}_0)_{l_1, l_2} := \frac{1}{n(n-1)(n-2)} \sum_{i_1 \neq i_2 \neq i_3} \alpha_{l_1}(X_{i_1, i_2}) \alpha_{l_2}(X_{i_1, i_3}),$$

and perform an eigendecomposition on it to construct an  $l \times r$  matrix  $\hat{\mathbf{V}}$  for which  $\hat{\mathbf{V}} \hat{\mathbf{A}}_0 \hat{\mathbf{V}}' = \mathbf{I}_r$ , the  $r \times r$  identity matrix. We next form the  $l \times l$  matrices  $\hat{\mathbf{A}}_{l'}$ ,

$$(\hat{\mathbf{A}}_{l'})_{l_1, l_2} := \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \alpha_{l_1}(X_{i_1, i_2}) \alpha_{l'}(X_{i_1, i_3}) \alpha_{l_2}(X_{i_1, i_4}),$$

where  $l' = 1, \dots, l$ , and transform them using  $\hat{\mathbf{V}}$  to obtain the  $r \times r$  matrices

$$\hat{\mathbf{N}}_{l'} := \hat{\mathbf{V}}' \hat{\mathbf{A}}_{l'} \hat{\mathbf{V}}.$$

We then find the matrix of joint (approximate) eigenvectors of these matrices as

$$\hat{\mathbf{Q}} := \arg \min_{\mathbf{Q} \in \mathcal{Q}} \sum_{l'=1}^l \sum_{z_1 \neq z_2} (\mathbf{Q}' \hat{\mathbf{N}}_{l'} \mathbf{Q})_{z_1, z_2}^2,$$

where we let  $\mathcal{Q}$  be the set of  $r \times r$  orthonormal matrices. With this matrix at hand we construct  $\hat{\mathbf{G}}$  as

$$(\hat{\mathbf{G}})_{l', z} := (\hat{\mathbf{Q}}' \hat{\mathbf{N}}_{l'} \hat{\mathbf{Q}})_{z, z}.$$

The minimization problem that defines  $\hat{\mathbf{Q}}$  can be solved efficiently using the algorithm of [Cardoso and Souloumiac \(1993\)](#).

Before proceeding to the asymptotic properties of this procedure it is useful to comment on its computational complexity. The procedure relies on estimated matrices that take the form of U-statistics up to order four. We note, however, that the kernels of these U-statistics are multiplicatively separable in the indices. Consequently, with some re-arrangement, their computational complexity is of the same order as that of a sample mean. The supplement provides additional details on this.

## 3.2 Regularity conditions

Three regularity conditions will be used. They are collected here. The first two of them impose conventional requirements on second moments.

**Assumption 2.** *The variables  $\alpha_{l'}(X_{i,j})$  have finite variance.*

Note that this assumption can always be satisfied by working with bounded functions.

**Assumption 3.** *The variable  $\varphi(X_{i,j})$  has finite variance.*

The third regularity condition concerns the  $l \times l$  matrix  $\mathbf{A}_0$ , with

$$(\mathbf{A}_0)_{l_1, l_2} := \mathbb{E}(\alpha_{l_1}(X_{i_1, i_2}) \alpha_{l_2}(X_{i_1, i_3})).$$

Observe that  $\mathbf{A}_0 = \mathbf{G} \text{diag}(\mathbf{p}) \mathbf{G}'$ . Its rank is equal to  $r$  by Assumption 1 and so it has  $r$  non-zero eigenvalues. We represent its eigendecomposition as

$$\mathbf{A}_0 = \mathbf{U} \mathbf{L} \mathbf{U}',$$

with  $\mathbf{L}$  the  $r \times r$  diagonal matrix that collects the non-zero eigenvalues and  $\mathbf{U}$  the  $l \times r$  orthonormal matrix whose  $r$  columns contain the associated eigenvectors. Then

$$\mathbf{V} := \mathbf{L}^{-1/2} \mathbf{U}'$$

is the probability limit of  $\hat{\mathbf{V}}$ .

**Assumption 4.** *All non-zero eigenvalues of  $\mathbf{A}_0$  are simple.*

This assumption implies continuity of  $\mathbf{V}$  as a function of  $\mathbf{A}_0$  and is helpful in deriving the properties of  $\hat{\mathbf{V}}$ .

We remark that, because  $\hat{\mathbf{A}}_0$  is a  $\sqrt{n}$ -consistent estimator of  $\mathbf{A}_0$ , the rank condition on the matrix  $\mathbf{G}$  can be tested by any of a number of standard procedures to test the rank of a matrix.

### 3.3 A linearization

An important step in deriving the large-sample properties of our procedures lies in analyzing the first-step estimator,  $\hat{\mathbf{G}}$ . This estimator is a complicated function of the auxiliary estimators of the matrices  $\mathbf{A}_0$  and  $\mathbf{A} := (\mathbf{A}_1, \dots, \mathbf{A}_l)$ , with the elements of the latter equal to

$$(\mathbf{A}_{l'})_{l_1, l_2} := \mathbb{E}(\alpha_{l_1}(X_{i_1, i_2}) \alpha_{l'}(X_{i_1, i_3}) \alpha_{l_2}(X_{i_1, i_4})).$$

Their respective influence functions are

$$\beta_i(\mathbf{A}_0) := \text{vec}(\mathbf{B}_i(\mathbf{A}_0) - \mathbb{E}(\mathbf{B}_i(\mathbf{A}_0))),$$

where

$$\begin{aligned} (\mathbf{B}_{i_1}(\mathbf{A}_0))_{l_1, l_2} := & \mathbb{E}(\alpha_{l_1}(X_{i_1, i_2}) \alpha_{l_2}(X_{i_1, i_3}) | Z_{i_1}) \\ & + \mathbb{E}(\alpha_{l_1}(X_{i_1, i_2}) \alpha_{l_2}(X_{i_2, i_3}) | Z_{i_1}) \\ & + \mathbb{E}(\alpha_{l_1}(X_{i_2, i_3}) \alpha_{l_2}(X_{i_1, i_2}) | Z_{i_1}), \end{aligned}$$

and

$$\beta_i(\mathbf{A}) := \text{vec}(\mathbf{B}_i(\mathbf{A}) - \mathbb{E}(\mathbf{B}_i(\mathbf{A}))),$$

where  $\mathbf{B}_i(\mathbf{A}) := (\mathbf{B}_i(\mathbf{A}_1), \dots, \mathbf{B}_i(\mathbf{A}_l))$  for

$$\begin{aligned} (\mathbf{B}_{i_1}(\mathbf{A}_{l'}))_{l_1, l_2} &:= \mathbb{E}(\alpha_{l_1}(X_{i_1, i_2}) \alpha_{l_2}(X_{i_1, i_3}) \alpha_{l'}(X_{i_1, i_4}) | Z_{i_1}) \\ &+ \mathbb{E}(\alpha_{l_1}(X_{i_1, i_2}) \alpha_{l_2}(X_{i_2, i_3}) \alpha_{l'}(X_{i_2, i_4}) | Z_{i_1}) \\ &+ \mathbb{E}(\alpha_{l_1}(X_{i_2, i_3}) \alpha_{l_2}(X_{i_1, i_2}) \alpha_{l'}(X_{i_2, i_4}) | Z_{i_1}) \\ &+ \mathbb{E}(\alpha_{l_1}(X_{i_2, i_3}) \alpha_{l_2}(X_{i_2, i_4}) \alpha_{l'}(X_{i_1, i_2}) | Z_{i_1}). \end{aligned}$$

From this a linearization of  $\hat{\mathbf{G}}$  can be derived. Additional notation is needed in order to state the result.

The estimator  $\hat{\mathbf{G}}$  is due to [Cardoso and Souloumiac \(1993\)](#) and is based on the insight that the  $r \times r$  matrices  $\mathbf{N}_{l'} := \mathbf{V} \mathbf{A}_{l'} \mathbf{V}'$  are diagonalizable in the same (orthonormal) basis. We write  $\mathbf{Q}$  for the  $r \times r$  matrix of joint eigenvectors and let

$$\mathbf{D}_{l'} := \mathbf{Q}' \mathbf{N}_{l'} \mathbf{Q}$$

be the  $r \times r$  diagonal matrices that contain their respective eigenvalues. We note that the main diagonal of  $\mathbf{D}_{l'}$  corresponds to the  $l'$ -th row of matrix  $\mathbf{G}$ . If we let  $\mathbf{D} := (\mathbf{D}_1, \dots, \mathbf{D}_l)'$ , then

$$\text{vec}(\mathbf{G}') = (\mathbf{I}_l \otimes \mathbf{S}) \text{vec}(\mathbf{D}'),$$

where  $\mathbf{S} := (\mathbf{s}_1, \dots, \mathbf{s}_r)'$  is the  $r \times r^2$  selection matrix with  $\mathbf{s}_z$  the selection vector whose  $((z-1)(r+1)+1)$ -th entry is equal to one and all other entries are equal to zero. The shorthand  $\mathbf{R} := (\mathbf{D}_1 \ominus \mathbf{D}_1, \dots, \mathbf{D}_l \ominus \mathbf{D}_l)'$ , where  $\mathbf{D}_{l'} \ominus \mathbf{D}_{l'} := (\mathbf{D}_{l'} \otimes \mathbf{I}_r) - (\mathbf{I}_r \otimes \mathbf{D}_{l'})$  is a Kronecker difference, will also be useful.

We have

$$\text{vec}(\hat{\mathbf{G}}' - \mathbf{G}') = \frac{1}{n} \sum_{i=1}^n \beta_i(\mathbf{G}') + o_p(n^{-1/2}),$$

for

$$\beta_i(\mathbf{G}') := (\mathbf{I}_l \otimes \mathbf{S}) ((\mathbf{I}_l \otimes \mathbf{I}_{r^2}) + \mathbf{P}_{\mathbf{R}}) ((\mathbf{I}_l \otimes \mathbf{K}_{r^2}) (\mathbf{T}_1 + \mathbf{T}_2) \beta_i(\mathbf{A}_0) + \mathbf{T}_3 \beta_i(\mathbf{A})),$$

where  $\mathbf{P}_R := \mathbf{R}(\mathbf{R}'\mathbf{R})^*\mathbf{R}'$  with a  $*$  superscript denoting the Moore-Penrose pseudo inverse of a matrix,  $\mathbf{K}_{r^2} := \mathbf{I}_{r^2} + \mathbf{C}_{r^2}$  with  $\mathbf{C}_{r^2}$  denoting the  $r^2 \times r^2$  commutation matrix, and

$$\mathbf{T}_1 := -(\mathbf{D} \otimes \mathbf{I}_r)(\mathbf{Q}' \otimes \mathbf{Q}')(\mathbf{L} \ominus \mathbf{L})^*(\mathbf{L} \otimes \mathbf{I}_r)(\mathbf{V} \otimes \mathbf{V}),$$

and

$$\mathbf{T}_2 := -\frac{1}{2}(\mathbf{D} \otimes \mathbf{I}_r)(\mathbf{Q}' \overset{c}{\otimes} \mathbf{Q}')(\mathbf{V} \overset{r}{\otimes} \mathbf{V}), \quad \mathbf{T}_3 := \mathbf{I}_l \otimes (\mathbf{Q}' \otimes \mathbf{Q}')(\mathbf{V} \otimes \mathbf{V}).$$

Here, we use  $\overset{c}{\otimes}$  and  $\overset{r}{\otimes}$  to denote columnwise and rowwise Kronecker products, respectively.

### 3.4 Limit behavior

Given the large-sample behavior of the first-step estimator it is readily established that

$$\text{vec}(\hat{\mathbf{G}}^* - \mathbf{G}^*) = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\beta}_i(\mathbf{G}^*) + o_p(n^{-1/2}),$$

for

$$\boldsymbol{\beta}_i(\mathbf{G}^*) := (((\mathbf{I}_l - \mathbf{G}\mathbf{G}^*) \otimes (\mathbf{G}'\mathbf{G})^{-1}) - (\mathbf{G}' \otimes \mathbf{G})^* \mathbf{C}_{rl}) \boldsymbol{\beta}_i(\mathbf{G}').$$

We can now present the asymptotic behavior of our estimators of the main components of the stochastic block model. We state these in the form of two theorems.

We first provide the limit distribution of the estimator  $\hat{\mathbf{p}}$ . The asymptotic variance of this estimator is equal to

$$\mathbf{V}_p := \mathbb{E}(\boldsymbol{\theta}_i \boldsymbol{\theta}_i'),$$

where

$$\boldsymbol{\theta}_i := \mathbf{G}^* \boldsymbol{\beta}_i(\mathbf{a}) + (\mathbf{a}' \otimes \mathbf{I}_r) \boldsymbol{\beta}_i(\mathbf{G}^*),$$

with  $\boldsymbol{\beta}_i(\mathbf{a}) := (\beta_i(a_1), \dots, \beta_i(a_l))'$  for

$$\beta_i(a_{l'}) := 2(\mathbb{E}(\alpha_{l'}(X_{i,j})|Z_i) - a_{l'}).$$

Theorem 2 follows.

**Theorem 2.** Suppose that Assumptions 1, 2, and 4 hold. Then

$$\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}) \rightsquigarrow N(\mathbf{0}, \mathbf{V}_p),$$

as  $n \rightarrow \infty$ .

We next state the limit distribution of our estimator of  $\varphi_{z_1, z_2}$ . We will need

$$\boldsymbol{\beta}_i(\mathbf{M}_\varphi) := \text{vec}(\mathbf{B}_i(\mathbf{M}_\varphi) - \mathbb{E}(\mathbf{B}_i(\mathbf{M}_\varphi))),$$

with

$$\begin{aligned} (\mathbf{B}_{i_1}(\mathbf{M}_\varphi))_{l_1, l_2} &:= \mathbb{E}(\alpha_{l_1}(X_{i_1, i_3}) \varphi(X_{i_1, i_2}) \alpha_{l_2}(X_{i_2, i_4}) | Z_{i_1}) \\ &\quad + \mathbb{E}(\alpha_{l_1}(X_{i_2, i_3}) \varphi(X_{i_1, i_2}) \alpha_{l_2}(X_{i_1, i_4}) | Z_{i_1}) \\ &\quad + \mathbb{E}(\alpha_{l_1}(X_{i_1, i_2}) \varphi(X_{i_2, i_3}) \alpha_{l_2}(X_{i_3, i_4}) | Z_{i_1}) \\ &\quad + \mathbb{E}(\alpha_{l_1}(X_{i_3, i_4}) \varphi(X_{i_2, i_3}) \alpha_{l_2}(X_{i_1, i_2}) | Z_{i_1}). \end{aligned}$$

We then have that

$$\boldsymbol{\beta}_i(\mathbf{H}_\varphi) := \mathbf{K}_{r^2}(\mathbf{G}^* \mathbf{M}_\varphi \otimes \mathbf{I}_r) \boldsymbol{\beta}_i(\mathbf{G}^*) + (\mathbf{G}^* \otimes \mathbf{G}^*) \boldsymbol{\beta}_i(\mathbf{M}_\varphi)$$

is the influence function of  $\hat{\mathbf{H}}_\varphi$ . From this, the asymptotic behavior of  $\hat{\varphi}_{z_1, z_2}$  will follow after a linearization. Writing  $\mathbf{I}_r = (\mathbf{e}_1, \dots, \mathbf{e}_r)$ , the influence function of  $\hat{\varphi}_{z_1, z_2}$  is

$$\vartheta_i(z_1, z_2) := (p_{z_1} p_{z_2})^{-1} (\mathbf{e}'_{z_2} \otimes \mathbf{e}'_{z_1}) (\boldsymbol{\beta}_i(\mathbf{H}_\varphi) - \varphi_{z_1, z_2} \boldsymbol{\beta}_i(\mathbf{H}_1)).$$

We let

$$v_\varphi(z_1, z_2) := \mathbb{E}(\vartheta_i(z_1, z_2)^2)$$

in the next theorem.

**Theorem 3.** Suppose that Assumptions 1, 2, 3, and 4 hold. Then,

$$\sqrt{n}(\hat{\varphi}_{z_1, z_2} - \varphi_{z_1, z_2}) \rightsquigarrow N(0, v_\varphi(z_1, z_2)),$$

as  $n \rightarrow \infty$ .

This theorem covers distribution functions, probability mass functions when edge weights are discrete, and moments, for example.

More generally, the result can be used in combination with standard asymptotic theory to construct estimators of a parameter defined as

$$\arg \max_{\boldsymbol{\delta}} \mathbb{E}(\varphi(X_{i,j}; \boldsymbol{\delta}) | Z_i = z_1, Z_j = z_2).$$

Under regularity conditions the implied estimator will be  $\sqrt{n}$ -consistent and asymptotically normal.

### 3.5 Density estimation

Next consider the case where the edge weights are continuous and  $F_{z_1, z_2}$  admits a density function,  $f_{z_1, z_2}$ , say. The estimation of  $f_{z_1, z_2}$  may be of interest. Theorem 3 does not immediately cover this as such nonparametric estimators involve smoothing- or truncation bias. However, the dependence between edge weights reduces these issues to second-order problems. This is in line with the conclusion reached in recent related work by [Graham, Niu and Powell \(2019\)](#), who considered estimation of the marginal density of the edge weights in a setting that encompasses ours.

We consider a standard kernel estimator at a point  $x$ . Similar results to the one to follow may be established for locally-linear (or polynomial) versions of the kernel estimator, as well as for estimators based on series expansions. The kernel estimator can be cast into our generic formula for  $\hat{\varphi}_{z_1, z_2}$  by setting

$$\varphi(X_{i,j}) = \frac{1}{h_n} k\left(\frac{X_{i,j} - x}{h_n}\right),$$

where  $k$  is a kernel function and  $h_n$  is a non-negative bandwidth.

The following conditions are standard in density estimation.

**Assumption 3'.** *The kernel function  $k$  is symmetric, bounded, and integrates to one. The density functions  $f_{z_1, z_2}$  are bounded and are twice differentiable with bounded derivatives.*

In the sequel, Assumption 3' replaces Assumption 3.

The dependence that the stochastic block model induces between the edge weights means that the variance of the kernel estimator will be of the order

$$n^{-1} + (n^2 h_n)^{-1} + (n^3 h_n)^{-1} + (n^4 h_n)^{-1}.$$

These terms arise from the covariances between the (symmetrized) kernel of

$$\frac{1}{n(n-1)(n-2)(n-3)} \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \alpha_{l_1}(X_{i_1, i_2}) \frac{1}{h_n} k\left(\frac{X_{i_2, i_3} - x}{h_n}\right) \alpha_{l_2}(X_{i_3, i_4}),$$

evaluated at two different quadruples of nodes that have exactly one, two, three, or all four indices in common. Hence, if  $nh_n \rightarrow \infty$ , the variance is of order  $n^{-1}$ , and only terms involving quadruples of nodes that have one index in common contribute to the asymptotic variance. By the usual arguments for U-statistics on graphs (Janson and Nowicki, 1991) their contribution is equal to the variance of the sample mean of the projection of the kernel onto the  $Z_i$ .

Further, exploiting the fact that the three terms in the kernel above are independent conditional on  $(Z_{i_2}, Z_{i_3})$  it is readily confirmed that a standard argument, as validated by Assumption 3', implies that the expectation of the statistic given in the previous display is

$$\sum_{z_1, z_2} (\mathbf{G})_{l_1, z_1} f_{z_1, z_2}(x) p_{z_1} p_{z_2} (\mathbf{G})_{l_2, z_2} + O(h_n^2).$$

Thus, the smoothing bias of the kernel estimator is of the order  $h_n^2$ , which is the conventional result for such a procedure. In light of the variance being of the order  $n^{-1}$ , this means that asymptotic bias will be absent provided that  $nh_n^4 \rightarrow 0$ . Taking these observations together leads to the following conclusion.

**Proposition 2.** *Suppose that Assumptions 1, 2, 3', and 4 hold. Then, if the bandwidth satisfies  $nh_n \rightarrow \infty$  and  $nh_n^4 \rightarrow 0$  as  $n \rightarrow \infty$ , Theorem 3 applies to the kernel density estimator.*

Note that, from above, the mean-squared error of the density estimator is of the order

$$h_n^4 + n^{-1} + (n^2 h_n)^{-1}.$$



Equating the rate of the first term to the rate of the second term gives us the optimal-rate requirement  $h_n \propto n^{-2/5}$ . Observe that this requirement is compatible with the condition  $nh_n^4 \rightarrow 0$  in Proposition 2. Hence, undersmoothing is not needed to prevent asymptotic bias.

## 4 Numerical experiments

We now provide simulation evidence for the standard block model with binary outcomes and two latent communities. Here, the conditional distributions are fully characterized by their success probabilities and so we consider estimation of

$$\varphi_{z_1, z_2} = \mathbb{P}(X_{i,j} = 1 | Z_i = z_1, Z_j = z_2),$$

along with the relative sizes of the two latent communities,  $p_1$  and  $p_2$ . We report results for several combinations of these probabilities. For each, we simulated 10,000 networks of size  $n = 100$ , with  $p_2 = .70$ , and report the mean, median, standard deviation, and interquartile range across the Monte Carlo replications. To give a sense of the numerical complexity, estimation of the model for a single replication takes just under  $1/3$  of a second on my desktop computer.

Table 1 contains results for three designs that feature complementarity, i.e., success is more likely if agents are from the same community. The three designs vary in how much  $\varphi_{1,1}$  is separated from  $\varphi_{2,2}$ . The specification is peculiar in that agents from different communities never generate successes. We do this to highlight that such a degeneracy does not cause problems for our procedure.

The table shows good performance of our procedure. The conditional distributions are accurately recovered. As  $\varphi_{2,2}$  moves further away from  $\varphi_{1,1}$  the standard deviation of the estimated success probabilities goes down, as expected. The estimator of the population shares of the communities equally does well across the designs. Its performance is essentially unaffected by the design changes.

Table 1: Simulation results

	$\varphi_{1,1}$	$\varphi_{1,2}$	$\varphi_{2,2}$	$p_1$	$p_2$
Design 1					
true value	0.200	0.000	0.400	0.300	0.700
mean	0.220	0.000	0.392	0.285	0.715
median	0.219	0.000	0.392	0.284	0.716
std. dev.	0.029	0.006	0.014	0.044	0.044
iqr	0.039	0.008	0.018	0.059	0.059
Design 2					
true value	0.200	0.000	0.600	0.300	0.700
mean	0.209	0.000	0.590	0.287	0.713
median	0.209	0.000	0.591	0.287	0.714
std. dev.	0.024	0.003	0.012	0.043	0.043
iqr	0.032	0.004	0.016	0.059	0.059
Design 3					
true value	0.200	0.000	0.800	0.300	0.700
mean	0.202	0.000	0.789	0.287	0.713
median	0.202	0.000	0.789	0.287	0.713
std. dev.	0.023	0.002	0.009	0.044	0.044
iqr	0.030	0.002	0.012	0.058	0.058

# References

- Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics* 37, 3099–3132.
- Allman, E. S., C. Matias, and J. A. Rhodes (2011). Parameter identifiability in a class of random graph mixture models. *Journal of Statistical Planning and Inference* 141, 1719–1736.
- Amini, A. A., A. Chen, P. J. Bickel, and E. Levina (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics* 41, 2097–2122.
- Bickel, P. J., D. Choi, X. Chang, and H. Zhang (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Annals of Statistics* 41, 1922–1943.
- Bonhomme, S., K. Jochmans, and J.-M. Robin (2016). Nonparametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society, Series B* 78, 211–229.
- Cardoso, J.-F. and A. Souloumiac (1993). Blind beamforming for non-Gaussian signals. *IEEE-Proceedings, F* 140, 362–370.
- Celisse, A., J. J. Daudin, and L. Pierre (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics* 6, 1847–1899.
- Daudin, J. J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistical Computing* 18, 173–183.
- Erdős, P. and A. Rényi (1959). On random graphs. *Publicationes Mathematicae* 6, 290–297.
- Graham, B. S., F. Niu, and J. L. Powell (2019). Kernel density estimation for undirected dyadic data. Mimeo.
- Hoff, P., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97, 1090–1098.
- Holland, P., K. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social Networks* 5, 109–137.
- Janson, S. and K. Nowicki (1991). The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probability Theory and Related Fields* 90, 341–375.
- Kasahara, H. and K. Shimotsu (2014). Nonparametric identification and estimation of the number of components in multivariate mixtures. *Journal of the Royal Statistical Society, Series B* 76,

97–111.

- Kwon, C. and E. Mbakop (2021). Estimation of the number of components of non-parametric multivariate finite mixture models. *Annals of Statistics* 49, 2178–2205.
- Le, C. M. and E. Levina (2019). Estimating the number of communities by spectral methods. Mimeo.
- Lei, J. (2016). A goodness-of-fit test for stochastic block models. *Annals of Statistics* 44, 401–424.
- Levine, M., D. R. Hunter, and D. Chauveau (2011). Maximum smoothed likelihood for multivariate mixtures. *Biometrika* 98, 403–416.
- Mariadassou, M., S. Robin, and C. Vacher (2010). Uncovering latent structure in valued graphs: A variational approach. *Annals of Applied Statistics* 4, 715–742.
- McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models*. Wiley-Blackwell.
- Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96, 1077–1087.
- Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics* 39, 1878–1915.
- Snijders, T. A. B. and K. Nowicki (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification* 14, 75–100.
- Sussman, D. L., M. Tang, D. E. Fishkind, and C. E. Priebe (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association* 107, 1119–1128.
- Tang, M., J. Cape, and C. E. Priebe (2022). Asymptotically efficient estimators for stochastic blockmodels: The naive MLE, the rank-constrained MLE, and the spectral estimator. *Bernoulli* 28, 1049–1073.
- Titterton, D. M. (1983). Minimum distance non-parametric estimation of mixture proportions. *Journal of the Royal Statistical Society, Series B* 45, 37–46.
- Wang, Y. X. R. and P. J. Bickel (2017). Likelihood-based model selection for stochastic block models. *Annals of Statistics* 45, 500–528.
- Yan, B., P. Sarkar, and X. Cheng (2018). Provable estimation of the number of blocks in block models. *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*.