NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF STOCHASTIC BLOCK MODELS FROM MANY SMALL NETWORKS

Koen Jochmans*

Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France

This version: February 19, 2024

Abstract

This paper concerns the analysis of network data when unobserved node-specific heterogeneity is present. We postulate a weighted version of the classic stochastic block model, where nodes belong to one of a finite number of latent communities and the placement of edges between them and any weight assigned to these depend on the communities to which the nodes belong. A simple rank condition is presented under which we establish that the number of latent communities, their distribution, and the conditional distribution of edges and weights given community membership are all nonparametrically identified from knowledge of the joint (marginal) distribution of edges and weights in graphs of a fixed size. The identification argument is constructive and we present a computationally-attractive nonparametric estimator based on it. Limit theory is derived under asymptotics where we observe a growing number of independent networks of a fixed size. The results of a series of numerical experiments are reported on.

Keywords: heterogeneity, network, random graph, sorting, stochastic block model

*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: koen.jochmans@tse-fr.eu.
I am grateful to the Editor (Elie Tamer), and to the Associate Editor and three referees for constructive comments.
Funded by the European Union (ERC, NETWORK, 101044319) and by the French Government and the French National Research Agency under the Investissements d' Avenir program (ANR-17-EURE-0010). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them

1 Introduction

Pairwise interactions between agents give rise to network data. In a graph where agents are nodes, a (possibly) weighted edge is present between two nodes if they interact; the weight on the edge corresponds to the output of the interaction. In this paper we are interested in the analysis of this type of network data in the presence of agent-specific unobserved heterogeneity.

The setting that we analyse is similar to the ones in Barrat, Bartélemy, Pastor-Satorras and Vespignani (2004), Newman (2004), and Mariadassou, Robin and Vacher (2010) and can be described as follows. Each of n agents first draws his (latent) heterogeneity type independently from a common discrete distribution. Next, $n(n-1)/2$ decisions are made whether or not to place an (undirected) edge between each pair of agents and weights are assigned to the edges that have been placed. Conditional on the types of all n agents edge placement and weight assignment are jointly independent across agent pairs. The probability that an edge is formed between two agents and the distribution from which its weight is generated depend on the latent type of both agents involved. When agents only decide on the placement of edges, and so no weights are present, our setting reduces to the stochastic block model of Holland, Laskey and Leinhardt (1983). Hence, our generalization may be seen as a weighted stochastic block model.

Motivation for considering the weighted case may be found in Newman (2004) and Bonhomme (2021), for example. Borrowing from the latter work, consider a setting where agents pair up in teams to produce an output; examples are researchers producing research papers (Ahmapoor and Jones 2019, Ductor, Fafchamps, Goyal and van der Leij 2014) or scientists creating innovations (Akcigit, Baslandze and Stantcheva 2016, Bell, Chetty, Jaravel, Petkova and Van Reenen 2019). Our model allows for these agents' decision to pair up to depend on their latent types, and thus accommodates different patterns of sorting, and for their joint output to be a function of their latent types, allowing for the presence of complementarity in production.

The weighted stochastic block model has three sets of primitive parameters. These are (i) the number of latent heterogeneity types; (ii) the distribution of these types; and (iii) the conditional distribution of edges and their weights given a pair of types. This paper provides a simple condition under which all these parameters are nonparametrically identified from knowledge of the joint distribution of edges and weights in a graph of fixed size n. The condition in question is a linear-independence requirement on the distributions of (a collection of) edges and weights for a single agent, as a function of his type. We discuss this condition in detail and give examples below. In the simplest case, the (unweighted) stochastic block model with two types, it boils down to the demand that the expected degree is different across the two types. Our findings improve on results in Allman, Matias and Rhodes (2011), where considerably stronger restrictions are used to identify the unknown distributions given knowledge of the number of latent types.

Our identification argument readily suggests an estimation strategy based on solving a collection of multilinear restrictions. We propose a nonparametric estimator of the model parameters that is computationally attractive. It is built around a joint (approximate) diagonalization step tailored to our setup. This type of routine has its origins in blind source separation (Cardoso and Souloumiac 1993) and has found applicability in the estimation of mixtures elswhere (Bonhomme, Jochmans and Robin 2016a,b, 2017). Here we use it as an auxiliary estimator in the construction of our estimators of (ii) and (iii); an estimator of (i) follows readily from adapting arguments in Kwon and Mbakop (2021). Our estimator can be constructed from a single network of size n or from multiple, say m, independent networks, each of size n (say). We provide limit theory for our estimator as m grows large while n is kept fixed. Our estimators of the type distribution and (linear) functionals of the component distributions are estimated at the parametric rate and converge in distribution to a Gaussian random variable. We characterise the variance and provide a consistent plug-in estimator for it. We evaluate our procedures by means of several Monte Carlo experiments.

To the best of our knowledge ours is the first nonparametric estimator of (weighted) stochastic block models for which any theoretical guarantees are available under many small network asymptotics. In the parametric case one can, in principle, proceed by maximizing the likelihood. However, the likelihood can have many local maxima and the EM algorithm is generally intractable (see, e.g., Bickel, Choi, Chang and Zhang 2013, Snijders and Nowicki 1997, and Nowicki and Snijders 2001).Even in very small graphs with only two types of unobserved heterogeneity we found that implementation was cumbersome and convergence to a local optimum was frequent. In contrast, by virtue of the joint-diagonalization step, our estimator is computationally stable and very fast to compute.

Our estimator is also consistent and asymptotically normal when computed from a single network and the number of nodes, n, grows large. Under such single-network asymptotics, however, several alternative estimation strategies could be concocted. Moreover, in large graphs, it is possible to correctly classify units into latent groups even in sparse graphs, where the expected degree shrinks with n; see von Luxburg, Belkin and Bousquet (2008), Rohe, Chatterjee and Yu (2011), Sussman, Tang, Fishkind and Priebe (2012), Lei and Rinaldo (2015), and Yan, Sarkar and Cheng (2018) for consistency results for such spectral clustering techniques under various conditions. Given a consistent assignment of units to types one could, in principle, proceed with estimation of the model parameters; see Channarond, Daudin and Robin (2011) and Tang, Cape and Priebe (2022) for results along these lines in the unweighted case. Our identification argument is for a fixed network size; it does not rely on the ability to correctly classify nodes and the concept of a sparse graph does not arise. Adapting the limit theory for our estimator to the sparse case is left for future work.

The rest of the paper is structured as follows. In Section 2 we formally set up the model of interest. In Section 3 we state and discuss the linear-independence condition that is at the heart of our identification result and provide our constructive proof of identification. In Section 4 we present the implied estimator and provide its limit distribution under many small network asymptotics. In Section 5, finally, we give the results of three simulation studies.

2 Stochastic block model

Consider n units, labeled $1, \ldots, n$. Each unit belongs to one of r latent communities, labeled $1, \ldots, r$. Community membership is recorded in the random variables $Z_1, \ldots, Z_n$, which are not observable in the data. These variables are independent and identically distributed; we let

$$p_z := \mathbb{P}(Z_i = z)$$

for $1 \leq z \leq r$. For each unordered pair of units (i_1, i_2), with $i_1 \neq i_2$, we observe a random variable $X_{i_1,i_2} \in \mathcal{X} \subseteq \mathcal{R}$. Conditional on $Z_1, \ldots, Z_n$, the variables $\{X_{i_1,i_2} : 1 \leq i_1 < i_2 \leq n\}$ are independent and the distribution of X_{i_1,i_2} depends only on Z_{i_1} and Z_{i_2}, with probability distribution

$$P_{z_1 z_2}(x) := \mathbb{P}(X_{i_1,i_2} \leq x | Z_{i_1} = z_1, Z_{i_2} = z_2)$$

for $1 \leq z_1, z_2 \leq r$ and any $x \in \mathcal{X}$. Given that units in a pair are unordered it is natural to work under a symmetry condition on these measures, so that $P_{z_1 z_2} = P_{z_2 z_1}$ for all $1 \leq z_1 \leq z_2 \leq r$.

The distribution of the array $\{X_{i_1,i_2} : 1 \leq i_1 < i_2 \leq n\}$ is invariant to a permutation of the indices $i_1, \ldots, i_n$. It thus follows by the Aldous-Hoover theorem (Hoover 1979, Aldous 1981) that we may equivalently view it as being generated through a structure of the form

$$X_{i_1,i_2} = X(Z_{i_1}, Z_{i_2}, V_{i_1,i_2}), \tag{1}$$

where the variables $\{V_{i_1,i_2} : 1 \leq i_1 < i_2 \leq n\}$ are independent, uniformly distributed on $[0, 1]$, and independent of $Z_1, \ldots, Z_n$, and the function X is symmetric in its first two arguments. This corresponds to a non-separable model for dyadic data with (discrete) two-way heterogeneity.

When $\mathcal{X} = \{0, 1\}$ our setup corresponds to the stochastic block model as in Holland, Laskey and Leinhardt (1983). It presents a parsimonious approach to introduce latent heterogeneity in the Erdős and Rényi (1959) random-graph model. Here, the units can be seen as nodes in a network. The probability of an edge to be formed between any two nodes depends on the types of both nodes. The random graph so formed can then exhibit degree heterogeneity, where some nodes are involved in more edges than others, or assortative matching (homophily), where nodes are more likely to form edges between them when they are of a similar type; see McPherson, Smith-Lovin and Cook (2001) and Newman and Leicht (2007), for example.

More generally, our model can be seen to generate a weighted graph, where nodes do not only decide to place an edge but also to assign a weight to it. This weight can have many interpretations, depending on the problem at hand. It could refer to a type of connection (as in Nowicki and Snijders 2001) or to its intensity (as in Newman 2004), for example. In light of Eq. (1) it could also be interpreted as the result of a generic dyadic interaction between units, i.e., X could be seen as a production function for X_{i_1,i_2}, with its functional form depending on (Z_{i_1}, Z_{i_2}).

In a weighted graph it is useful to code the absence of an edge as the event $X_{i_1,i_2} = 0$, that is, as assigning zero weight to it. Let $p_{z_1 z_2}$ be the probability that an edge is formed between nodes belonging to communities z_1 and z_2. When an edge is placed between two such nodes, in a second step its (non-zero) weight is drawn from some distribution $Q_{z_1 z_2}$ supported on $\mathcal{X} \setminus \{0\}$. The implied density of $P_{z_1 z_2}$ at $x \in \mathcal{X}$ (with respect to the appropriate measure) then factors as

$$f_{z_1 z_2}(x) \coloneqq \{x = 0\}\,(1 - p_{z_1 z_2}) + \{x \neq 0\}\, p_{z_1 z_2}\, q_{z_1 z_2}(x), \tag{2}$$

where $q_{z_1 z_2}$ is the density of $Q_{z_1 z_2}$. From knowledge of $f_{z_1 z_2}$ we can first recover $p_{z_1 z_2}$ as $1 - f_{z_1 z_2}(0)$ and, next, $q_{z_1 z_2}$ as $f_{z_1 z_2}/p_{z_1 z_2}$ on the whole set $\mathcal{X} \setminus \{0\}$. This formulation highlights that the formation of the network and the weights placed on edges may jointly depend on community membership. Furthermore, conditional on $X_{i_1,i_2} \neq 0$, the variables Z_{i_1} and Z_{i_2} are not independent.

As an example, suppose that X_{i_1,i_2} is the pair-level output of two workers, with $X_{i_1,i_2} = 0$ capturing that they did not work together (and thus did not produce any output). This type of setting is discussed in Ahmapoor and Jones (2019) and Bonhomme (2021). The presence of latent communities means the presence of different worker types $1 \leq z \leq r$. The dependence of $p_{z_1 z_2}$ on the types accommodates sorting between workers based on their latent type while the dependence of $q_{z_1 z_2}$ on the types allows for match-quality aspects such as complementarity between the worker types to affect the output that is being produced.

3 Identification

The (weighted) stochastic block model has three sets of primitive parameters. These are (i) the number of communities, r; (ii) the size of each community, p_z for $1 \leq z \leq r$; and (iii) the conditional probability measures $P_{z_1 z_2}$ for $1 \leq z_1 \leq z_2 \leq r$. Our aim is to provide conditions under which these parameters are non-parametrically identified from the (marginal) distribution of $\{X_{i_1,i_2} : 1 \leq i_1 < i_2 \leq n\}$. Of course, as this distribution is invariant to a relabeling of the latent communities, identification here is to be understood as being up to label swapping.

To proceed, for any integer $q \geq 1$, we first introduce the q-dimensional random variable

$$\boldsymbol{X}_{i_1,\boldsymbol{i}'_q} \coloneqq (X_{i_1,i'_1}, X_{i'_1,i'_2}, \ldots, X_{i'_{q-1},i'_q}),$$

where $\boldsymbol{i}'_q \coloneqq (i'_1, \ldots, i'_q)$ and all the indices appearing in i_1 and $i'_1, \ldots, i'_q$ are distinct. We then let

$$G_z(\boldsymbol{x}_q) \coloneqq \mathbb{P}(\boldsymbol{X}_{i_1,\boldsymbol{i}'_q} \leq \boldsymbol{x}_q | Z_{i_1} = z)$$

for $1 \leq z \leq r$ and all $\boldsymbol{x}_q = (x_1, \ldots, x_q) \in \mathcal{X}^q$. This is the distribution of edge weights along a path over q different nodes when starting at a node that belongs to community z. It depends on the path only through its length, q. To conserve on notation we make this dependence explicit only through the dimension of its argument. For the minimal value $q = 1$, $\boldsymbol{X}_{i_1,\boldsymbol{i}'_1} = X_{i_1,i'_1}$ and G_z reduces to the distribution of edge weights for a node from community z. In this case, $G_z(x)$ is a mixture with mixing components $P_{zz'}(x)$ and mixing weights $p_{z'}$ for $1 \leq z' \leq r$. For larger values of q, G_z has an iterative representation as a mixture in the sense that, for any q, $G_z(x_1, \ldots, x_q)$ is a mixture of $P_{zz'}(x_1)\, G_{z'}(x_2, \ldots, x_q)$ with weights $p_{z'}$ for $1 \leq z' \leq r$.

We will work under the following assumption.

Assumption 1. *There exists a finite integer q such that the functions G_z for $1 \leq z \leq r$ are linearly independent.*

Assumption 1 is a rank condition and is reminiscent to conditions that appear in the analysis of multivariate latent-variable models; see, e.g., Hu (2008), Kasahara and Shimotsu (2009), Bonhomme, Jochmans and Robin (2016a,b), and Higgins and Jochmans (2023). It demands that the community to which a unit belongs sufficiently affects how edges involving this unit are weighted. An extreme but simple example where the requirement fails is the Erdős and Rényi (1959) model (or a weighted version thereof). There, $P_{z_1 z_2} = P$ for some P and all $1 \leq z_1 \leq z_2 \leq r$. The latent community structure then has no observable implications, and G_z does not depend on z.

On the other end of the spectrum the $P_{z_1 z_2}$ for $1 \leq z_1 \leq z_2 \leq r$ are all linearly independent. This is the assumption under which Allman, Matias and Rhodes (2011) develop their general results. It implies that Assumption 1 holds for any q and, in particular, for $q = 1$. It can, nevertheless, be a strong assumption to impose. At a minimum, it needs

$$|\mathcal{X}| \geq \binom{r+1}{2};$$

compare this to the corresponding restriction that $|\mathcal{X}| \geq r$ for Assumption 1 to hold for $q = 1$. This rules out the binary-edge case, for example, where the $2 \times \binom{r+1}{2}$ matrix of probability distributions necessarily has reduced column rank, even for the minimal case where $r = 2$. Nevertheless, Assumption 1 goes through for $q = 1$ when $r = 2$ provided that

$$(p_{12} - p_{22}) \neq \frac{p_1}{1 - p_1} (p_{12} - p_{11}), \tag{3}$$

which is rather weak. It allows, notably, that $p_{11} = p_{22}$, provided that $p_1 \neq 1 - p_1$, and thus covers the affiliation structure of Frank and Harary (1982), which falls outside the scope of the linear-independence restrictions of Allman, Matias and Rhodes (2011). For more than two communities Assumption 1 cannot be satisfied for $q = 1$ in the binary-edge case but it can for $q > 1$. Here, the state space of the variable $\boldsymbol{X}_{i_1, \boldsymbol{i}'_q}$ consists of all ordered sequences of zeros and ones of length q, which has cardinality 2^q. Thus, in terms of Assumption 1, working with larger values of q amounts to working with the probability distribution of a random variable with a richer support.

At a minimum, Assumption 1 requires that

$$r \leq |\mathcal{X}|^q.$$

The richer the support of X_{i_1,i_2} the easier this restriction can be satisfied for a given q and r. When $\mathcal{X}$ is an uncountable set $q = 1$ will suffice for Assumption 1 in many cases, but not always. An example where the cardinality of $\mathcal{X}$ is, in fact, irrelevant is a weighted version of the affiliation model. There we have, for (different) probability distributions Q_0 and Q_1,

$$P_{z_1 z_2} = \begin{cases} Q_0 & \text{if } z_1 \neq z_2 \\ Q_1 & \text{if } z_1 = z_2 \end{cases}.$$

For $q = 1$, Assumption 1 concerns the distributions

$$G_z(x) = Q_0(x)\,\mathbb{P}(Z_i \neq z) + Q_1(x)\,\mathbb{P}(Z_i = z) = p_z\,(Q_1(x) - Q_0(x)) + Q_0(x),$$

for $1 \leq z \leq r$. In the two-community case these distributions are linearly-independent if (and only if) $p_1 \neq p_2$; this generalizes our finding below Eq. (3) beyond the binary-edge case. Inspection of the problem, however, reveals that, for larger r, these distributions will always be linearly-dependent, no matter the set $\mathcal{X}$. Nevertheless, looking at larger values of q continues to provide a way forward. When $q = 2$, for example, Assumption 1 involves

$$G_z(x_1, x_2) = Q_1(x_1)\big(Q_1(x_2)\,w_z^{11} + Q_0(x_2)\,w_z^{10}\big) + Q_0(x_1)\big(Q_1(x_2)\,w_z^{01} + Q_0(x_2)\,w_z^{00}\big)$$

for $1 \leq z \leq r$, where we used the shorthands $w_z^{11} := p_z^2$, $w_z^{10} := p_z\,(1 - p_z)$, $w_z^{01} := \sum_{z' \neq z} p_{z'}^2$, and $w_z^{00} := \sum_{z' \neq z} (1 - p_{z'})^2$ to keep the expression light. It can be verified that, in this case, Assumption 1 goes through for all $r \leq 4$, provided that the probabilities p_z are all different.

We note that if Assumption 1 holds for some q, then it also holds for all $q' > q$. From here on out, we will, therefore, take q to indicate the smallest such value for which the linear-independence requirement is satisfied. With this ambiguity settled we can move on to establishing identification.

We begin by showing the following result.

Proposition 1. *Let Assumption 1 hold. Then the number of latent communities, r, is non-parametrically identified from the distribution of edge weights in a graph with $2q+1$ nodes.*

Proposition 1 does not have a counterpart in Allman, Matias and Rhodes (2011). Its proof is simple. It relies on the fact that any two collections of edge weights are dependent only if they have one (or more) nodes in common, and that any such dependence stems from their joint dependence on the latent community-membership indicators. Consider an index i_1 and two index sets $\boldsymbol{i}'_q = (i'_1, \ldots, i'_q)$ and $\boldsymbol{i}''_q = (i''_1, \ldots, i''_q)$, with all indices involved being different. Then the random vectors $\boldsymbol{X}_{i_1,\boldsymbol{i}'_q}$ and $\boldsymbol{X}_{i_1,\boldsymbol{i}''_q}$ are dependent only through their joint dependence on Z_{i_1}. Hence, using the definition of G_z, their joint distribution factors as

$$\mathbb{P}(\boldsymbol{X}_{i_1,\boldsymbol{i}'_q} \leq \boldsymbol{x}'_q, \boldsymbol{X}_{i_1,\boldsymbol{i}''_q} \leq \boldsymbol{x}''_q) = \sum_{z=1}^{r} p_z\, G_z(\boldsymbol{x}'_q)\, G_z(\boldsymbol{x}''_q).$$

This is a bivariate mixture of identically distributed random variables with component distributions G_z and mixture weights p_z for $1 \leq z \leq r$. Further, because the component distributions are linearly independent by Assumption 1, it follows from an application of the results in Kwon and Mbakop (2021, Propositions 2.1 and 2.3) that r is non-parametrically identified.

We now move on to establish identification of the remaining parameters of the stochastic block model. The proof is constructive and will form the basis of our estimation procedure laid out below.

To begin, let $\mathcal{Q}$ be a finite subset of $\mathcal{X}^q$ with $|\mathcal{Q}| = r$ and let $\boldsymbol{G}$ be the $r \times r$ matrix obtained on stacking $G_1(\boldsymbol{x}_q), \ldots, G_r(\boldsymbol{x}_q)$ for all $\boldsymbol{x}_q \in \mathcal{Q}$. Consider an index i_1 and two index sets $\boldsymbol{i}'_q = (i'_1, \ldots, i'_q)$ and $\boldsymbol{i}''_q = (i''_1, \ldots, i''_q)$, with all indices involved being different. Collect the probabilities $\mathbb{P}(\boldsymbol{X}_{i_1,\boldsymbol{i}'_q} \leq \boldsymbol{x}'_q, \boldsymbol{X}_{i_1,\boldsymbol{i}''_q} \leq \boldsymbol{x}''_q)$ for $(\boldsymbol{x}'_q, \boldsymbol{x}''_q) \in \mathcal{Q}^2$ in the $r \times r$ matrix $\boldsymbol{A}$. Then, by the same argument as in the proof of Proposition 1, we have the factorization

$$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{D}\boldsymbol{G}^\top$$

for $\boldsymbol{D} := \operatorname{diag}(p_1, \ldots, p_r)$. By Assumption 1 there exist sets $\mathcal{Q}$ such that $\operatorname{rank}(\boldsymbol{G}) = r$ and we work with such a set henceforth. Note that this rank condition is testable, as rank($\boldsymbol{A}$) $= r$ if (and only if) it holds, and $\boldsymbol{A}$ is identified (and estimable from data; see below). From an eigendecomposition of $\boldsymbol{A}$ we then construct an $r \times r$ matrix $\boldsymbol{V}$ for which

$$\boldsymbol{V}\boldsymbol{A}\boldsymbol{V}^\top = \boldsymbol{I}_r,$$

where $\boldsymbol{I}_r$ denotes the $r \times r$ identity matrix. When combined with the bivariate-mixture decomposition from the previous display this yields

$$\boldsymbol{I}_r = \boldsymbol{V}\boldsymbol{A}\boldsymbol{V}^\top = \boldsymbol{V}(\boldsymbol{G}\boldsymbol{D}\boldsymbol{G}^\top)\boldsymbol{V}^\top = (\boldsymbol{V}\boldsymbol{G}\boldsymbol{D}^{1/2})(\boldsymbol{V}\boldsymbol{G}\boldsymbol{D}^{1/2})^\top = \boldsymbol{Q}_0\boldsymbol{Q}_0^\top,$$

where we introduced the shorthand $\boldsymbol{Q}_0 \coloneqq \boldsymbol{V}\boldsymbol{G}\boldsymbol{D}^{1/2}$, which is a full-rank $r \times r$ orthonormal matrix. Now, for some $s \leq q$, consider a third index set $\boldsymbol{i}_s''' = (i_1''', \ldots, i_s''')$, which again does not involve i_1 and does not share any indices with $\boldsymbol{i}_q'$ and $\boldsymbol{i}_q''$. For each $\boldsymbol{x}_s \in \mathcal{S} \subset \mathcal{X}^s$, collect the probabilities $\mathbb{P}(\boldsymbol{X}_{i_1,\boldsymbol{i}_q'} \leq \boldsymbol{x}_q', \boldsymbol{X}_{i_1,\boldsymbol{i}_s'''} \leq \boldsymbol{x}_s, \boldsymbol{X}_{i_1,\boldsymbol{i}_q''} \leq \boldsymbol{x}_q'')$ for $(\boldsymbol{x}_q', \boldsymbol{x}_q'') \in \mathcal{Q}^2$ in the $r \times r$ matrix $\boldsymbol{A}_{\boldsymbol{x}_s}$. Then, using the shorthand notation $\boldsymbol{D}_{\boldsymbol{x}_s} \coloneqq \operatorname{diag}(\boldsymbol{G}_1(\boldsymbol{x}_s), \ldots, \boldsymbol{G}_r(\boldsymbol{x}_s))$, we have

$$\boldsymbol{A}_{\boldsymbol{x}_s} = \boldsymbol{G}\boldsymbol{D}^{1/2}\boldsymbol{D}_{\boldsymbol{x}_s}\boldsymbol{D}^{1/2}\boldsymbol{G},$$

because $(\boldsymbol{X}_{i_1,\boldsymbol{i}_q'}, \boldsymbol{X}_{i_1,\boldsymbol{i}_s'''}, \boldsymbol{X}_{i_1,\boldsymbol{i}_q''})$ are dependent only because they share the index i_1 and so are independent conditional on Z_{i_1}. By using the matrix $\boldsymbol{V}$, we have that, for each $\boldsymbol{x}_s \in \mathcal{S}$,

$$\boldsymbol{V}\boldsymbol{A}_{\boldsymbol{x}_s}\boldsymbol{V}^\top = \boldsymbol{Q}_0\,\boldsymbol{D}_{\boldsymbol{x}_s}\boldsymbol{Q}_0^\top.$$

Thus the collection of matrices $\boldsymbol{V}\boldsymbol{A}_{\boldsymbol{x}_s}\boldsymbol{V}^\top$ for $\boldsymbol{x}_s \in \mathcal{S}$ share the same eigenvectors, $\boldsymbol{Q}_0$. Furthermore, by Assumption 1 there exists a set $\mathcal{S}$ on which the distributions G_z for $1 \leq z \leq r$ are all different, and so it follows from Belouchrani, Abed-Meraim, Cardoso and Moulines (1997, Theorem 3) that the matrix $\boldsymbol{Q}_0$ is identified up to permutation and sign of its columns. Moreover, for an $r \times r$ diagonal matrix $\boldsymbol{\Delta}$ containing only 1 or -1 and an $r \times r$ permutation matrix $\boldsymbol{\Sigma}$, we have recovered $\boldsymbol{Q} \coloneqq \boldsymbol{Q}_0\boldsymbol{\Delta}\boldsymbol{\Sigma}$.

With $\boldsymbol{W} \coloneqq \boldsymbol{V}^\top\boldsymbol{Q}$ at hand we can now proceed to show that the p_z for $1 \leq z \leq r$ and the $P_{z_1 z_2}$ for $1 \leq z_1 \leq z_2 \leq r$ are identified up to the same labelling of the latent communities. We will first establish that, for any chosen scalar function φ, the expectation

$$\varphi_{z_1 z_2} \coloneqq \mathbb{E}(\varphi(X_{i_1,i_2})|Z_{i_1} = z_1, Z_{i_2} = z_2)$$

(assuming that it exists) is identified. For $\varphi(X_{i_1,i_2}) = \{X_{i_1,i_2} \leq x\}$ this yields $P_{z_1 z_2}(x)$, which suffices for the current proof, but allowing for other functions will lead to a convenient way to estimate linear functionals of $P_{z_1 z_2}$ without the need to first estimate the distribution itself. For a chosen φ, collect the $\varphi_{z_1 z_2}$ in the $r \times r$ matrix $\boldsymbol{H}_\varphi$. Next collect the expectations $\mathbb{E}(\{\boldsymbol{X}_{i_1,\boldsymbol{i}'_q} \leq \boldsymbol{x}'_q\}\, \varphi(X_{i_1,i_2})\, \{\boldsymbol{X}_{i_2,\boldsymbol{i}''_q} \leq \boldsymbol{x}''_q\})$ for $(\boldsymbol{x}'_q, \boldsymbol{x}''_q) \in \mathcal{Q}^2$, where i_2 is any of the indices in $\boldsymbol{i}'''_s$, in the $r \times r$ matrix $\boldsymbol{A}_\varphi$. Because $(\boldsymbol{X}_{i_1,\boldsymbol{i}'_q}, X_{i_1,i_2}, \boldsymbol{X}_{i_2,\boldsymbol{i}''_q})$ are independent conditional on (Z_{i_1}, Z_{i_2}) we have

$$\boldsymbol{A}_\varphi = \boldsymbol{G}\boldsymbol{D}\boldsymbol{H}_\varphi \boldsymbol{D}\boldsymbol{G}^\top.$$

Therefore,

$$\boldsymbol{V}\,\boldsymbol{A}_\varphi \boldsymbol{V}^\top = (\boldsymbol{V}\boldsymbol{G}\boldsymbol{D}^{1/2})\boldsymbol{D}^{1/2}\boldsymbol{H}_\varphi \boldsymbol{D}^{1/2}(\boldsymbol{V}\boldsymbol{G}\boldsymbol{D}^{1/2})^\top = \boldsymbol{Q}_0 \boldsymbol{D}^{1/2}\boldsymbol{H}_\varphi \boldsymbol{D}^{1/2}\boldsymbol{Q}_0^\top,$$

from which we can recover

$$\boldsymbol{W}^\top \boldsymbol{A}_\varphi \boldsymbol{W} = \boldsymbol{\Sigma}^\top \boldsymbol{\Delta}\boldsymbol{D}^{1/2}\boldsymbol{H}_\varphi\, \boldsymbol{D}^{1/2}\boldsymbol{\Delta}\boldsymbol{\Sigma}$$

for any function φ. Here, the permutation matrix $\boldsymbol{\Sigma}$ reshuffles the rows and columns of $\boldsymbol{\Delta}\,\boldsymbol{D}^{1/2}\boldsymbol{H}_\varphi \boldsymbol{D}^{1/2}\boldsymbol{\Delta}$, whose entries are $\varphi_{z_1 z_2}\, \delta_{z_1}\delta_{z_2} \sqrt{p_{z_1}\, p_{z_2}}$ for $1 \leq z_1 \leq z_2 \leq r$, using the notation $\boldsymbol{\Delta} = \operatorname{diag}(\delta_1, \ldots, \delta_r)$. The reshuffling amounts to a relabelling of the latent communities, and the presence of $\boldsymbol{\Sigma}$ is merely a manifestation of the fact that identification can only be achieved up to an arbitrary labelling. What matters for our purposes is that this labelling is independent of the function φ as, then, with $\boldsymbol{W} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_r)$, we have that, for $1 \leq z_1 \leq z_2 \leq r$, the scalars $\boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_\varphi\, \boldsymbol{\omega}_{z_2}$ identify the $\varphi_{z_1 z_2}\, \delta_{z_1}\delta_{z_2} \sqrt{p_{z_1}\, p_{z_2}}$ for any function φ up to an arbitrary but common (and, thus, irrelevant) labelling of the latent communities $1 \leq z \leq r$.

For the constant function $\varphi(x) = 1$ we have $\boldsymbol{H}_1 = \boldsymbol{1}_r \boldsymbol{1}_r^\top$ for $\boldsymbol{1}_r$ the r vector of ones, and that the matrix $\boldsymbol{A}_1$ collects the products $\mathbb{P}(\boldsymbol{X}_{i_1,\boldsymbol{i}'_q} \leq \boldsymbol{x}'_q)\, \mathbb{P}(\boldsymbol{X}_{i_2,\boldsymbol{i}''_q} \leq \boldsymbol{x}''_q)$ for $(\boldsymbol{x}'_q, \boldsymbol{x}''_q) \in \mathcal{Q}^2$. In this case the scalars $\boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_1\, \boldsymbol{\omega}_{z_2}$ for $1 \leq z_1 \leq z_2 \leq r$ identify the $\delta_{z_1}\delta_{z_2} \sqrt{p_{z_1}\, p_{z_2}}$ up to the same common labelling as before. Putting everything obtained so far together shows that

$$\varphi_{z_1 z_2} = (\boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_1\, \boldsymbol{\omega}_{z_2})^{-1} (\boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_\varphi\, \boldsymbol{\omega}_{z_2}) \tag{4}$$

for $1 \leq z_1 \leq z_2 \leq r$ and

$$p_z = \boldsymbol{\omega}_z^\top \boldsymbol{A}_1 \, \boldsymbol{\omega}_z \tag{5}$$

for $1 \leq z \leq r$ are all identified up to an arbitrary but common labelling of the latent communities. Here, to arrive at the last equality, we have made use of the fact that $\delta_z^2 = 1$ for all z.

Finally, note that, from Eq. (4), we immediately obtain the identification of ratios of conditional expectations, i.e., ratios of the form $\varphi'_{z_1 z_2}/\varphi''_{z_1 z_2}$ for two functions φ' and φ'' as

$$(\boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_{\varphi''} \, \boldsymbol{\omega}_{z_2})^{-1} (\boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_{\varphi'} \, \boldsymbol{\omega}_{z_2}) = (\varphi''_{z_1 z_2})^{-1} (\varphi'_{z_1 z_2}).$$

A prime application of this result has $\varphi''(x) = \{\underline{x} < x \leq \overline{x}\}$ and $\varphi'(x) = \varphi(x)\, \varphi''(x)$ for chosen φ, which returns conditional expectations of $\varphi(X_{i_1,i_2})$ given that $\underline{x} < X_{i_1,i_2} \leq \overline{x}$ and latent-community membership. In the context of Eq. (2) this allows to immediately recover functionals of the density $q_{z_1 z_2}$ that conditions on a link being present, as well as the match probability $p_{z_1 z_2}$.

We have established our next result.

Proposition 2. *Let Assumption 1 hold. Then the probabilities p_z for $1 \leq z \leq r$ and the distributions $P_{z_1 z_2}$ for $1 \leq z_1 \leq z_2 \leq r$ are non-parametrically identified (up to an arbitrary but common labelling of the latent communities) from the distribution of edge weights in a graph with $2q + s + 1$ nodes.*

Propositions 1 and 2 dictate values for the size of the graph, n, as a function of q. Moreover, Proposition 1 requires that $n \geq 2q + 1$ while Proposition 1 goes through for $n \geq 2q + s + 1$ or, more conservatively but as a function of q only (by recalling that we have $s \leq q$), that $n \geq 3q + 1$. The (weighted) stochastic block model may thus be fully identifiable from the distribution of (certain subsets of) edge weights in networks involving as little as four agents. This finding is to be contrasted with the results of Allman, Matias and Rhodes (2011, Theorems 14 and 15) where, assuming that r is known, graphs of size nine are needed to arrive at the conclusion of Proposition 2 while, at the same time, demanding the more stringent condition that the conditional distributions $P_{z_1 z_2}$ for $1 \leq z_1 \leq z_2 \leq r$ are all linearly independent.

Assumption 1 demands that $|\mathcal{X}|^q \geq r$ or, equivalently, $q \geq \lceil \log(r)/\log(|\mathcal{X}|) \rceil$, where $\lceil a \rceil$ is the smallest integer greater than or equal to a. For Proposition 2, this translates to

$$n \geq \left\lceil 2 \frac{\log(r)}{\log(|\mathcal{X}|)} + s + 1 \right\rceil,$$

which, as a function of the number of communities, grows at a logarithmic rate. Recalling that $s \leq q$ a lower bound on n that does not depend on s is $\lceil 3 \log(r)/\log(|\mathcal{X}|) + 1 \rceil$. For the case of binary edge weights and probabilities $p_{z_1 z_2}$ that are different for all $1 \leq z_1 \leq z_2 \leq r$, Allman, Matias and Rhodes (2011, Theorem 2) give another lower bound on n. Theirs, however, grows like r^4.

4 Estimation

We now turn to the construction of estimators of $\varphi_{z_1 z_2}$ for $1 \leq z_1 \leq z_2 \leq r$ and of p_z for $1 \leq z \leq r$; an estimator of r can be constructed as in Kwon and Mbakop (2021). The proof of Proposition 2, and Equations (4) and (5) in particular, suggests estimators of the form

$$\hat{\varphi}_{z_1 z_2} := (\hat{\boldsymbol{\omega}}_{z_1}^\top \hat{\boldsymbol{A}}_1 \hat{\boldsymbol{\omega}}_{z_2})^{-1} (\hat{\boldsymbol{\omega}}_{z_1}^\top \hat{\boldsymbol{A}}_\varphi \hat{\boldsymbol{\omega}}_{z_2}) \tag{6}$$

and

$$\hat{p}_z = \hat{\boldsymbol{\omega}}_z^\top \hat{\boldsymbol{A}}_1 \hat{\boldsymbol{\omega}}_z, \tag{7}$$

where $\hat{\boldsymbol{\omega}}_z$ is an estimator of $\boldsymbol{\omega}_z$ for each $1 \leq z \leq r$ and $\hat{\boldsymbol{A}}_\varphi$ is an estimator of $\boldsymbol{A}_\varphi$ for any function φ.

The proof equally provides a routine to construct the matrix $(\hat{\boldsymbol{\omega}}_1, \ldots, \hat{\boldsymbol{\omega}}_r) = \hat{\boldsymbol{W}} = \hat{\boldsymbol{V}}^\top \hat{\boldsymbol{Q}}$ from estimators $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{A}}_{\boldsymbol{x}_s}$ for $\boldsymbol{x}_s \in \mathcal{S}$ of the matrices $\boldsymbol{A}$ and $\boldsymbol{A}_{\boldsymbol{x}_s}$ for $\boldsymbol{x}_s \in \mathcal{S}$. First, matrix

$$\hat{\boldsymbol{V}} := \hat{\boldsymbol{S}}^{-1/2} \hat{\boldsymbol{U}}^\top$$

is constructed from the eigendecomposition $\hat{\boldsymbol{A}} = \hat{\boldsymbol{U}} \hat{\boldsymbol{S}} \hat{\boldsymbol{U}}^\top$, where $\hat{\boldsymbol{S}}$ is an $r \times r$ diagonal matrix containing the eigenvalues of $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{U}}$ is an $r \times r$ orthonormal matrix collecting the associated eigenvectors. Next, the $r \times r$ matrix $\hat{\boldsymbol{Q}}$ is computed as the (approximate) joint diagonalizer of the collection of matrices $\hat{\boldsymbol{V}}^\top \hat{\boldsymbol{A}}_{\boldsymbol{x}_s} \hat{\boldsymbol{V}}$ for $\boldsymbol{x}_s \in \mathcal{S}$. More precisely, we let

$$\hat{\boldsymbol{Q}} := \underset{\bar{\boldsymbol{Q}}}{\operatorname{argmin}} \sum_{\boldsymbol{x}_s \in \mathcal{S}} \sum_{z_1=1}^{r} \sum_{z_2 \neq z_1} \left(\bar{\boldsymbol{Q}}^\top (\hat{\boldsymbol{V}} \hat{\boldsymbol{A}}_{\boldsymbol{x}_s} \hat{\boldsymbol{V}}^\top) \bar{\boldsymbol{Q}} \right)^2_{z_1, z_2},$$

where the minimization is over the set of all $r \times r$ orthonormal matrices. This can be achieved efficiently using the JADE algorithm developed by Cardoso and Souloumiac (1993).

From a single network of size n, the estimators of the matrices $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{A}}_{\boldsymbol{x}_s}$ for $\boldsymbol{x}_s \in \mathcal{S}$ take the form of U-statistics. More precisely, the entries of the $r \times r$ matrix $\hat{\boldsymbol{A}}$ are equal to

$$\frac{1}{n(n-1)\cdots(n-2q)} \sum_{i_1 \neq i_1' \neq \cdots \neq i_q''} \{\boldsymbol{X}_{i_1, \boldsymbol{i}_q'} \leq \boldsymbol{x}_q'\} \{\boldsymbol{X}_{i_1, \boldsymbol{i}_q''} \leq \boldsymbol{x}_q''\}$$

for $(\boldsymbol{x}_q', \boldsymbol{x}_q'') \in \mathcal{Q}^2$. The entries of each of the $|\mathcal{S}|$ matrices $\hat{\boldsymbol{A}}_{\boldsymbol{x}_s}$, each again of size $r \times r$, are similarly given by

$$\frac{1}{n(n-1)\cdots(n-2q-s)} \sum_{i_1 \neq i_1' \neq \cdots \neq i_q''} \{\boldsymbol{X}_{i_1, \boldsymbol{i}_q'} \leq \boldsymbol{x}_q'\} \{\boldsymbol{X}_{i_1, \boldsymbol{i}_s'''} \leq \boldsymbol{x}_s\} \{\boldsymbol{X}_{i_1, \boldsymbol{i}_q''} \leq \boldsymbol{x}_q''\}$$

for $(\boldsymbol{x}_q', \boldsymbol{x}_q'') \in \mathcal{Q}^2$ and all $\boldsymbol{x}_s \in \mathcal{S}$. In the same way, for any function φ, matrix $\hat{\boldsymbol{A}}_\varphi$ has entries

$$\frac{1}{n(n-1)\cdots(n-2q-1)} \sum_{i_1 \neq i_1' \neq \cdots \neq i_q''} \{\boldsymbol{X}_{i_1, \boldsymbol{i}_q'} \leq \boldsymbol{x}_q'\} \, \varphi(X_{i_1, i_2}) \, \{\boldsymbol{X}_{i_2, \boldsymbol{i}_q''} \leq \boldsymbol{x}_q''\}$$

for $(\boldsymbol{x}_q', \boldsymbol{x}_q'') \in \mathcal{Q}^2$. In the above estimators, the summation ranges over all ordered subsets of, respectively, $2q+1$, $2q+s+1$, and $2q+2$ distinct indices from the set of n nodes. Each estimator is unbiased for their respective estimand. In fact, from Janson and Nowicki (1991),

$$\|\hat{\boldsymbol{A}} - \boldsymbol{A}\|_{\max} = O_p(n^{-1/2}), \qquad \|\hat{\boldsymbol{A}}_{\boldsymbol{x}_s} - \boldsymbol{A}_{\boldsymbol{x}_s}\|_{\max} = O_p(n^{-1/2}) \text{ for all } \boldsymbol{x}_s \in \mathcal{S},$$

and

$$\|\hat{\boldsymbol{A}}_\varphi - \boldsymbol{A}_\varphi\|_{\max} = O_p(n^{-1/2})$$

for any function φ for which $\varphi(X_{i_1,i_2})$ has finite variance. From this we can deduce that, as the size of the network grows, the estimators in Eq. (6) and Eq. (7) will be $\sqrt{n}$-consistent and asymptotically normally distributed. The limit distributions follow from our results below on plugging the appropriate influence functions for the estimators of the input matrices just given into Eq. (8). These influence functions equal the projection of the (symmetrized) kernel of the relevant U-statistic on Z_{i_1}.

Here we pursue the case where multiple independent networks, each of size n (say), are available. Let $\boldsymbol{X}_i \coloneqq \{X_{i_1,i_2;i} : 1 \leq i_1 \leq i_2 \leq n\}$ be the data from network $1 \leq i \leq m$. Then estimators of the different input matrices can be constructed as the sample averages

$$\hat{\boldsymbol{A}} \coloneqq {}^{1}\!/\!{}_{m} \sum_{i=1}^{m} \hat{\boldsymbol{A}}(\boldsymbol{X}_i), \quad \hat{\boldsymbol{A}}_{\boldsymbol{x}_s} \coloneqq {}^{1}\!/\!{}_{m} \sum_{i=1}^{m} \hat{\boldsymbol{A}}_{\boldsymbol{x}_s}(\boldsymbol{X}_i) \text{ for } \boldsymbol{x}_s \in \mathcal{S}, \quad \hat{\boldsymbol{A}}_{\varphi} \coloneqq {}^{1}\!/\!{}_{m} \sum_{i=1}^{m} \hat{\boldsymbol{A}}_{\varphi}(\boldsymbol{X}_i),$$

where $\hat{\boldsymbol{A}}(\boldsymbol{X}_i)$, $\hat{\boldsymbol{A}}_{\boldsymbol{x}_s}(\boldsymbol{X}_i)$, and $\hat{\boldsymbol{A}}_{\varphi}(\boldsymbol{X}_i)$ are the U-statistic estimators above computed from network $\boldsymbol{X}_i$. We now characterize the limit distribution of the estimators in Equations (6) and (7) as $m \to \infty$ while n is being kept fixed when the number of latent communities, r, is treated as known.

We will work under the following regularity condition.

Assumption 2. *The eigenvalues of $\boldsymbol{A}$ are all simple.*

This assumption simplifies the analysis of the estimator $\hat{\boldsymbol{V}}$ and could, in principle, be relaxed. As it holds generically, we see it as mild.

Moving on requires the introduction of some additional notation. With $\otimes$ the Kronecker product, let $\boldsymbol{S} \oplus \boldsymbol{S} \coloneqq (\boldsymbol{S} \otimes \boldsymbol{I}_r) + (\boldsymbol{I}_r \otimes \boldsymbol{S})$ and $\boldsymbol{S} \ominus \boldsymbol{S} \coloneqq (\boldsymbol{S} \otimes \boldsymbol{I}_r) - (\boldsymbol{I}_r \otimes \boldsymbol{S})$ denote, respectively, the Kronecker sum and difference of $\boldsymbol{S}$ with itself. Write $\overset{c}{\otimes}$ for the columnwise Kronecker product and $\overset{r}{\otimes}$ for the row-wise Kronecker product and let a $*$ superscript on a matrix denote its Moore-Penrose generalized inverse. We then introduce the two matrices

$$\boldsymbol{M}_1 \coloneqq (\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top) \left\{ (\boldsymbol{S} \ominus \boldsymbol{S})^* (\boldsymbol{I}_r \otimes \boldsymbol{S}) - {}^{1}\!/\!{}_{2} (\boldsymbol{I}_r \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{I}_r \overset{r}{\otimes} \boldsymbol{I}_r) \right\} (\boldsymbol{V} \otimes \boldsymbol{V})$$

and

$$\boldsymbol{M}_2 \coloneqq (\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{S} \ominus \boldsymbol{S})^* (\boldsymbol{S} \oplus \boldsymbol{S})(\boldsymbol{V} \otimes \boldsymbol{V}).$$

Next, write $\boldsymbol{O}_-$ for the matrix obtained on (vertically) concatenating $(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})$, $\boldsymbol{O}_+$ for the matrix obtained on concatenating $(\boldsymbol{D}_{\boldsymbol{x}_s} \oplus \boldsymbol{D}_{\boldsymbol{x}_s})$, and $\boldsymbol{O}_\times$ for the matrix obtained on concatenating $(\boldsymbol{D}_{\boldsymbol{x}_s} \otimes \boldsymbol{I}_r)$. Each of these matrices is of dimension $|\mathcal{S}|\, r^2 \times r^2$. Also write $\boldsymbol{R}$ for the $|\mathcal{S}|\, r^2 \times r^2\, |\mathcal{S}|$ block-diagonal matrix $(\boldsymbol{I}_{|\mathcal{S}|} \otimes \boldsymbol{W}^\top \otimes \boldsymbol{W}^\top)$. With these matrices in hand, we introduce

$$\boldsymbol{N}_1 \coloneqq (\boldsymbol{I}_r \otimes \boldsymbol{W}) \left\{ \left(\boldsymbol{I}_{r^2} + (\boldsymbol{O}_-^\top \boldsymbol{O}_-)^* (\boldsymbol{O}_-^\top \boldsymbol{O}_+) \right) \boldsymbol{M}_1 - (\boldsymbol{O}_-^\top \boldsymbol{O}_-)^* (\boldsymbol{O}_-^\top \boldsymbol{O}_\times) \boldsymbol{M}_2 \right\}$$

and $\boldsymbol{N}_2 \coloneqq (\boldsymbol{I}_r \otimes \boldsymbol{W})(\boldsymbol{O}_-^\top \boldsymbol{O}_-)^* (\boldsymbol{O}_-^\top \boldsymbol{R})$. These two matrices have the interpretation of Jacobians, translating estimation noise in, respectively, $\hat{\boldsymbol{A}} - \boldsymbol{A}$ and $\hat{\bar{\boldsymbol{A}}} - \bar{\boldsymbol{A}}$ into randomness in $\hat{\boldsymbol{W}} - \boldsymbol{W}$, where $\hat{\bar{\boldsymbol{A}}}$ and $\bar{\boldsymbol{A}}$ are the horizontal concatenations of the matrices $\hat{\boldsymbol{A}}_{\boldsymbol{x}_s}$ and $\boldsymbol{A}_{\boldsymbol{x}_s}$ for $\boldsymbol{x}_s \in \mathcal{S}$.

Let $\boldsymbol{\Xi}_\varphi(\boldsymbol{X}_i)$ for $1 \leq i \leq m$ be the $r \times r$ matrix with entries

$$\begin{aligned}
(\boldsymbol{\Xi}_\varphi(\boldsymbol{X}_i))_{z_1,z_2} \coloneqq\ & (\boldsymbol{\omega}_{z_2}^\top \otimes \boldsymbol{\omega}_{z_1}^\top) \operatorname{vec}(\hat{\boldsymbol{A}}_\varphi(\boldsymbol{X}_i) - \boldsymbol{A}_\varphi) \\
& + \left\{ (\boldsymbol{e}_{z_2}^\top \otimes \boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_\varphi) + (\boldsymbol{e}_{z_1}^\top \otimes \boldsymbol{\omega}_{z_2}^\top \boldsymbol{A}_\varphi) \right\} \boldsymbol{N}_1 \operatorname{vec}(\hat{\boldsymbol{A}}(\boldsymbol{X}_i) - \boldsymbol{A}) \\
& + \left\{ (\boldsymbol{e}_{z_2}^\top \otimes \boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_\varphi) + (\boldsymbol{e}_{z_1}^\top \otimes \boldsymbol{\omega}_{z_2}^\top \boldsymbol{A}_\varphi) \right\} \boldsymbol{N}_2 \operatorname{vec}(\hat{\bar{\boldsymbol{A}}}(\boldsymbol{X}_i) - \bar{\boldsymbol{A}}),
\end{aligned} \tag{8}$$

where the r-vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r$ denote the standard basis of the r-dimensional Euclidean space (i.e., they are the columns of $\boldsymbol{I}_r$). For any pair of functions φ' and φ'' (with finite variance), we let

$$\boldsymbol{\Sigma}_{\varphi',\varphi''} \coloneqq \mathbb{E}(\operatorname{vec}(\boldsymbol{\Xi}_{\varphi'}(\boldsymbol{X}_i)) \operatorname{vec}(\boldsymbol{\Xi}_{\varphi''}(\boldsymbol{X}_i))^\top),$$

which is a matrix of dimension $r^2 \times r^2$ containing covariances.

Let

$$\vartheta_\varphi \coloneqq \boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_\varphi \boldsymbol{\omega}_{z_2}$$

and write $\hat{\vartheta}_\varphi$ for the corresponding estimator. The notation leaves the dependence of these objects on (z_1, z_2) implicit.

Proposition 3. *Let Assumptions 1 and 2 hold. Consider functions φ' and φ'' so that* $\operatorname{var}(\varphi'(X_{i_1,i_2})) < \infty$ *and* $\operatorname{var}(\varphi''(X_{i_1,i_2})) < \infty$. *Then*

$$\sqrt{m} \begin{pmatrix} \hat{\vartheta}_{\varphi'} - \vartheta_{\varphi'} \\ \hat{\vartheta}_{\varphi''} - \vartheta_{\varphi''} \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\varphi',\varphi'} & \sigma_{\varphi',\varphi''} \\ \sigma_{\varphi'',\varphi'} & \sigma_{\varphi'',\varphi''} \end{pmatrix} \right),$$

where $\sigma_{\varphi',\varphi''} := (\boldsymbol{e}_{z_2} \otimes \boldsymbol{e}_{z_1})^\top \boldsymbol{\Sigma}_{\varphi',\varphi''} (\boldsymbol{e}_{z_2} \otimes \boldsymbol{e}_{z_1})$*, as* m *grows large while* n *is being held fixed.*

This result yields two corollaries. The first deals with estimators $\hat{\theta}$ of ratios of the form $\theta := {}^{\vartheta_{\varphi'}}/_{\vartheta_{\varphi''}}$; an example is the estimator $\hat{\varphi}_{z_1 z_2}$ given in Eq. (6), which has $\varphi'(x) = \varphi(x)$ and $\varphi''(x) = 1$.

Corollary 1. *Let Assumptions 1 and 2 hold. Consider functions* φ' *and* φ'' *so that* $\mathrm{var}(\varphi'(X_{i_1,i_2})) < \infty$ *and* $\mathrm{var}(\varphi''(X_{i_1,i_2})) < \infty$*. Then, provided that* $\vartheta_{\varphi''}$ *is not equal to zero,*

$$\sqrt{m}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \frac{(\boldsymbol{e}_{z_2} \otimes \boldsymbol{e}_{z_1})^\top (\boldsymbol{\Sigma}_{\varphi',\varphi'} - 2\theta \boldsymbol{\Sigma}_{\varphi',\varphi''} + \theta^2 \boldsymbol{\Sigma}_{\varphi'',\varphi''})(\boldsymbol{e}_{z_2} \otimes \boldsymbol{e}_{z_1})}{\vartheta^2_{\varphi''}}\right)$$

as m *grows large while* n *is being held fixed.*

This result covers $\theta = \int \varphi(x)\, f_{z_1 z_2}(x)\, dx$ and $\theta = \int \varphi(x)\, q_{z_1 z_2}(x)\, dx$, for example, which includes the conditional distributions at a given point as well as their moments. The corollary could equally be used to establish the asymptotic distribution of (generalized) method-of-moment estimators of a parameter δ defined through moment conditions of the form $\int \varphi(x, \delta)\, f_{z_1 z_2}(x)\, dx = 0$ under standard identification and regularity conditions for such estimators.

The second corollary concerns the size of each latent community.

Corollary 2. *Let Assumptions 1 and 2. Then,*

$$\sqrt{m}(\hat{p}_z - p_z) \xrightarrow{d} N(0, (\boldsymbol{e}_z \otimes \boldsymbol{e}_z)^\top \boldsymbol{\Sigma}_{1,1} (\boldsymbol{e}_z \otimes \boldsymbol{e}_z))$$

as m *grows large while* n *is being held fixed.*

Inference based on the result of the corollaries can be done using the variance estimator

$$\hat{\boldsymbol{\Sigma}}_{\varphi',\varphi''} := {}^1/_m \sum_{i=1}^{m} \mathrm{vec}(\hat{\boldsymbol{\Xi}}_{\varphi'}(\boldsymbol{X}_i))\, \mathrm{vec}(\hat{\boldsymbol{\Xi}}_{\varphi''}(\boldsymbol{X}_i))^\top,$$

where $\hat{\boldsymbol{\Xi}}_{\varphi}(\boldsymbol{X}_i)$ is the plug-in estimator of $\boldsymbol{\Xi}_{\varphi}(\boldsymbol{X}_i)$ obtained on replacing all unknown quantities in Eq. (8) by their respective sample counterparts. Because those estimators are all consistent it is easy to see that, for any function φ for which $\text{var}(\varphi(X_{i_1,i_2})) < \infty$, we have

$$^1\!/_m \sum_{i=1}^{m} \|\text{vec}(\hat{\boldsymbol{\Xi}}_\varphi(\boldsymbol{X}_i)) - \text{vec}(\boldsymbol{\Xi}_\varphi(\boldsymbol{X}_i))\|^2 = o_p(1),$$

from which we obtain $\hat{\boldsymbol{\Sigma}}_{\varphi',\varphi''} \xrightarrow{p} \boldsymbol{\Sigma}_{\varphi',\varphi''}$ under the conditions given in Proposition 3 as m grows large while n is being held fixed.

5 Numerical illustrations

We next present the results of three Monte Carlo experiments. Our first experiment is built around Eq. (2). There are two latent communities, with $p_1 = .3$ and $p_2 = .7$. Edges are placed between units of the different communities with probabilities $p_{11} = .50$, $p_{12} = .20$, and $p_{11} = .80$. Next, weights are assigned to edges that have been placed. These weights are drawn from $Q_{z_1 z_2}$, which are taken to be Beta distributions with (shape) parameters $\boldsymbol{\varpi}_{11} = (2,6)^\top$, $\boldsymbol{\varpi}_{12} = (2,2)^\top$, and $\boldsymbol{\varpi}_{22} = (6,2)^\top$. To motivate, one can think about units as workers of either low type or high type that are paired up into (multiple) teams to produce an output. Because Beta distributions are supported on $(0,1)$ we can think about the weight on an edge as a quality measure of the output produced by the pair of units involved. With $p_{12} < p_{11} < p_{22}$ there is assortative matching in link formation, so units of the same type are more likely to be sorted into a team. Also, with $\mu_{z_1 z_2} \coloneqq \int x\, q_{z_1 z_2}(x)\, dx$, we have $\mu_{11} = .25$, $\mu_{12} = .50$, and $\mu_{22} = .75$, and so $\mu_{11} < \mu_{12} < \mu_{22}$. This means that average quality increases when more workers are of the high type.

We implemented our estimators with $q = 1$ and used $\mathcal{Q} = \{.3, .7\}$ and $\mathcal{S} = \{.2, .4, .6, .8\}$. Alternative sets were also experimented with; they led to essentially the same conclusions as those reported on here. In Table 1 we report the median, interquartile range, and (actual) coverage rate of 95% confidence intervals of our estimators of (p_1, p_2), of (p_{11}, p_{12}, p_{22}), and of $(\mu_{11}, \mu_{12}, \mu_{22})$ (as computed over $10,000$ Monte Carlo replications) for various sample sizes. There is some bias and undercoverage in the smaller sample sizes for some of the parameters, but the results show that our method is able to accurately recover all parameters and yield

Table 1: Experiment 1

	p_1	p_2	p_{11}	p_{12}	p_{22}	μ_{11}	μ_{12}	μ_{22}
truth	.3000	.7000	.5000	.2000	.8000	.2500	.5000	.7500
				$(m, n) = (500, 5)$				
median	.3533	.6735	.4752	.2043	.8090	.2940	.5731	.7516
iqr	.2114	.1780	.2445	.2022	.1976	.2779	.4038	.0767
coverage	.8060	.8250	.8390	.9476	.9507	.8635	.9624	.9676
				$(m, n) = (500, 10)$				
median	.2992	.7003	.5002	.2006	.7990	.2503	.5008	.7501
iqr	.0597	.0324	.0711	.0537	.0419	.0859	.1436	.0254
coverage	.9451	.9463	.9486	.9479	.9509	.9490	.9547	.9496
				$(m, n) = (1000, 5)$				
median	.3256	.6881	.4939	.2003	.8051	.2616	.5301	.7494
iqr	.1456	.1171	.1894	.1468	.1270	.2101	.3154	.0527
coverage	.8573	.8661	.8591	.9535	.9512	.9073	.9556	.9538
				$(m, n) = (1000, 10)$				
median	.2991	.7007	.5007	.2000	.7993	.2504	.4991	.7502
iqr	.0412	.0224	.0500	.0373	.0294	.0626	.1009	.0179
coverage	.9523	.9513	.9532	.9514	.9472	.9514	.9586	.9552

reliable inference.

Our second experiment serves to assess the impact of increasing the number of latent communities. We use a variation of the affiliation model to construct a design that scales naturally as a function of r. Moreover, we set $p_z = {}^1\!/_r$ for $1 \leq z \leq r$ and draw weights from the (continuous) uniform distribution on $(0, r+1)$ when units belong to different communities and from the (symmetric) triangular distribution on $(z-1, z+1)$ when both units belong to community z. We implement our approach with $\mathcal{Q}$ and $\mathcal{S}$ set to equidistant points on $(0, r+1)$ and consider $r \in \{2, 3, 4, 5\}$ for $(m, n) = (1000, 25)$ in Table 2. We

Table 2: Experiment 2

	p_1	p_2	p_3	p_4	p_5	μ_{11}	μ_{22}	μ_{33}	μ_{44}	μ_{55}
					$r = 2$					
truth	.5000	.5000	—	—	—	1.0000	2.0000	—	—	—
median	.4999	.5000	—	—	—	0.9998	2.0000	—	—	—
iqr	.0117	.0101	—	—	—	.0115	.0191	—	—	—
coverage	.9564	.9474	—	—	—	.9504	.9504	—	—	—
					$r = 3$					
truth	.3333	.3333	.3333	—	—	1.0000	2.0000	3.0000	—	—
median	.3330	.3332	.3331	—	—	0.9991	1.9999	3.0013	—	—
iqr	.0119	.0205	.0207	—	—	.0350	.0527	.0830	—	—
coverage	.9544	.9520	.9506	—	—	.9502	.9638	.9488	—	—
					$r = 4$					
truth	.2500	.2500	.2500	.2500	—	1.0000	2.0000	3.0000	4.0000	—
median	.2496	.2508	.2515	.2499	—	1.0012	2.0085	2.9936	3.9874	—
iqr	.0264	.0315	.0577	.0609	—	.1832	.1526	.1950	.5182	—
coverage	.9464	.9520	.9294	.9362	—	.9464	.9656	.9492	.9098	—
					$r = 5$					
truth	.2000	.2000	.2000	.2000	.2000	1.0000	2.0000	3.0000	4.0000	5.0000
median	.1972	.2017	.2047	.2209	.1939	0.9787	2.0039	3.0007	3.9155	4.9447
iqr	.0284	.0423	.0481	.0799	.0971	.3198	.4303	.2107	.5627	1.1861
coverage	.9366	.9512	.9388	.8760	.8774	.9546	.8846	.9444	.8598	.8574

Table 3: Experiment 3

	p_1	p_2	μ_{11}	μ_{12}	μ_{22}
			$\gamma = .00$		
truth	.5000	.5000	1.0000	1.5000	2.0000
median	.4999	.5000	0.9998	1.5000	1.9998
iqr	.0116	.0101	.0116	.0107	.0191
coverage	. 9528	.9514	.9534	.9532	.9488
			$\gamma = .25$		
truth	.5000	.5000	1.0000	1.3750	1.7500
median	.4999	.4999	1.0001	1.3749	1.7500
iqr	.0118	.0096	.0097	.0078	.0111
coverage	. 9500	.9494	.9508	.9558	.9504
			$\gamma = .50$		
truth	.5000	.5000	1.0000	1.250	1.5000
median	.5000	.4999	1.0000	1.2501	1.5000
iqr	.0168	.0135	.0095	.0083	.0103
coverage	. 9450	.9484	.9496	.9486	.9484
			$\gamma = .75$		
truth	.5000	.5000	1.0000	1.1250	1.2500
median	.4999	.4992	1.0004	1.1251	1.2497
iqr	.0500	.0490	.0167	.0158	.0170
coverage	. 9454	.9468	.9474	.9502	.9482

provide the same summary statistics as before for our estimators of p_z and μ_{zz} for $1 \leq z \leq r$. To conserve on space we do not report results for $\mu_{z_1 z_2}$ when $z_1 \neq z_2$, which are all equal to $^{(r+1)}/_2$ and were very well estimated.

The results show that in all cases our estimators perform well. As r grows the probability of drawing from any of the latent communities shrinks and it becomes more likely to draw weights from the uniform distribution. This explains why the interquartile range of all estimators increases as r grows. The volatility also tends to be most pronounced for the larger z, which is in line with the asymptotic variance formula from Corollary 1 depending on the estimand. When r is at its largest the increased sampling noise is starting to affect the quality of inference. The additional model complexity from increasing r should then be offset by an increase in the sample size.

Our third experiment maintains the mixture structure from the previous experiment but focusses on the impact of moving the distributions closer to each other. We maintain a two-community structure, with $p_1 = p_2$ as before. Let $\gamma \in \{.00, .25, .50, .75\}$. When $z_1 \neq z_2$ we draw weights from the uniform distribution on $(0, 3-\gamma)$. When $z_1 = z_2 = 1$ we draw weights from the Triangular distribution on $(0, 2)$ and when $z_1 = z_2 = 1$ we draw weights from the Triangular distribution on $(1-\gamma, 3-\gamma)$. In Table 3 we observe that shifting the distributions toward each other does not have any major impact on the performance of the estimators for the range over which γ is being moved. If the shift parameter γ is moved down further the label-swapping issue starts making reliable Monte Carlo evaluation complicated. Of course, as $\gamma \to 1$ we have that $G_2 \to G_1$, at which point we lose identification. Our techniques will no longer provide reliable inference in the weakly-identified setting where γ is very close to unity. In practice the estimator $\hat{\boldsymbol{A}}$ is often close to being a singular matrix in such cases, making the estimator $\hat{\boldsymbol{\omega}}_z$ for $1 \leq z \leq r$ unstable.

Appendix

Proof of Proposition 3. Consider $\hat{\vartheta}_\varphi = \hat{\boldsymbol{\omega}}_{z_1}^\top \hat{\boldsymbol{A}}_\varphi \, \hat{\boldsymbol{\omega}}_{z_2}$ and $\vartheta_\varphi = \boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_\varphi \, \boldsymbol{\omega}_{z_2}$. We first derive the limit distribution of the estimators $\hat{\boldsymbol{A}}_\varphi$ and $\hat{\boldsymbol{W}} = (\hat{\boldsymbol{\omega}}_1, \ldots, \hat{\boldsymbol{\omega}}_r)$. We then combine the results to obtain the result of the proposition.

Input matrix $\hat{\boldsymbol{A}}_\varphi$. Observe that

$$\hat{\boldsymbol{A}}_\varphi = {}^1\!/\!_m \sum_{i=1}^m \hat{\boldsymbol{A}}_\varphi(\boldsymbol{X}_i)$$

is a sample average of i.i.d. random variables. The entries of the $r \times r$ matrix $\hat{\boldsymbol{A}}_\varphi(\boldsymbol{X}_i)$ are

$$\frac{1}{n(n-1)\cdots(n-2q-1)} \sum_{i_1 \neq i_1' \neq \cdots \neq i_q''} \{\boldsymbol{X}_{i_1, \boldsymbol{i}_q'; i} \leq \boldsymbol{x}_q'\} \, \varphi(X_{i_1, i_2; i}) \, \{\boldsymbol{X}_{i_2, \boldsymbol{i}_q''; i} \leq \boldsymbol{x}_q''\}$$

for $(\boldsymbol{x}_q', \boldsymbol{x}_q''), \in \mathcal{Q}^2$. They are unbiased for the corresponding entries of $\boldsymbol{A}$. Furthermore, they have finite second moment. To see this, we observe that the expected squared summand is equal to $\mathbb{E}(\{\boldsymbol{X}_{i_1, \boldsymbol{i}_q'; i} \leq \boldsymbol{x}_q'\} \, \varphi(X_{i_1, i_2; i})^2 \, \{\boldsymbol{X}_{i_2, \boldsymbol{i}_q''; i} \leq \boldsymbol{x}_q''\})$ and factors as

$$\sum_{z_1=1}^r \sum_{z_2=1}^r G_{z_1}(\boldsymbol{x}_q') \, p_{z_1} \, \mathbb{E}(\varphi(X_{i_1,i_2})^2 | Z_{i_1} = z_1, Z_{i_2} = z_2) \, p_{z_2} \, G_{z_2}(\boldsymbol{x}_q''),$$

using the conditional-independence restrictions of the stochastic block model. This is bounded by $\sum_{z_1=1}^r \sum_{z_2=1}^r \mathbb{E}(\varphi(X_{i_1,i_2})^2 | Z_{i_1} = z_1, Z_{i_2} = z_2) \, p_{z_1} \, p_{z_2} = \mathbb{E}(\varphi(X_{i_1,i_2})^2)$ because $\max_z \sup_{\boldsymbol{x}_q} |G_z(\boldsymbol{x}_q)| = 1$. The latter moment is finite by assumption. We therefore have that

$$\operatorname{vec}(\hat{\boldsymbol{A}}_\varphi - \boldsymbol{A}_\varphi) = {}^1\!/\!_m \sum_{i=1}^m \operatorname{vec}(\hat{\boldsymbol{A}}_\varphi(\boldsymbol{X}_i) - \boldsymbol{A}_\varphi) + o_p(m^{-1/2}),$$

and a standard central limit theorem applied to the sample average on the right-hand side yields asymptotic normality.

Transformation matrix $\hat{\boldsymbol{W}}$. The estimator $\hat{\boldsymbol{W}} = \hat{\boldsymbol{V}}^\top \hat{\boldsymbol{Q}}$ of $\boldsymbol{W} = \boldsymbol{V}^\top \boldsymbol{Q}$ is a nonlinear function of the input matrices $\hat{\boldsymbol{A}}$ and $\hat{\bar{\boldsymbol{A}}}$. By the same argument as the one just used for $\hat{\boldsymbol{A}}_\varphi$, each of these input matrices is $m^{-1/2}$-consistent and asymptotically normal for their respective population counterpart. To arrive at an influence-function representation for our estimator of the transformation matrix $\hat{\boldsymbol{W}}$ we begin with deriving such representations for $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{Q}}$.

Whitening matrix $\hat{\boldsymbol{V}}$. Recall the eigendecompositions $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^\top$ and $\hat{\boldsymbol{A}} = \hat{\boldsymbol{U}}\hat{\boldsymbol{S}}\hat{\boldsymbol{U}}^\top$, which define the matrices $\boldsymbol{V} = \boldsymbol{S}^{-1/2}\boldsymbol{U}^\top$ and $\hat{\boldsymbol{V}} = \hat{\boldsymbol{S}}^{-1/2}\hat{\boldsymbol{U}}^\top$. Under Assumption 2 we can appeal to the results of Anderson (1963) to obtain

$$\operatorname{vec}(\hat{\boldsymbol{U}} - \boldsymbol{U}) = (\boldsymbol{I}_r \otimes \boldsymbol{U})\,(\boldsymbol{S} \ominus \boldsymbol{S})^*(\boldsymbol{U}^\top \otimes \boldsymbol{U}^\top)\operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}) + o_p(m^{-1/2}), \tag{A.1}$$

where $\boldsymbol{S} \ominus \boldsymbol{S} = (\boldsymbol{S} \ominus \boldsymbol{I}_r) - (\boldsymbol{I}_r \ominus \boldsymbol{S})$ and the superscript $*$ denotes the Moore-Penrose pseudoinverse. Similarly, from Magnus (1985), using the delta method and the definition of $\boldsymbol{V}$,

$$\operatorname{vec}(\hat{\boldsymbol{S}}^{-1/2} - \boldsymbol{S}^{-1/2}) = -1/2(\boldsymbol{S}^{-1/2} \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{V} \overset{r}{\otimes} \boldsymbol{V})\operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}) + o_p(m^{-1/2}), \tag{A.2}$$

where $\overset{c}{\otimes}$ and $\overset{r}{\otimes}$ denote, respectively, the columnwise and rowwise Kronecker product. Now, by a linearization, and letting $\boldsymbol{K}_{r^2}$ be the $r^2 \times r^2$ commutation matrix (Magnus and Neudecker, 1979),

$$\operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V}) = (\boldsymbol{I}_r \otimes \boldsymbol{S}^{-1/2})\boldsymbol{K}_{r^2}\operatorname{vec}(\hat{\boldsymbol{U}} - \boldsymbol{U}) + (\boldsymbol{U} \otimes \boldsymbol{I}_r)\operatorname{vec}(\hat{\boldsymbol{S}}^{-1/2} - \boldsymbol{S}^{-1/2}) + o_p(m^{-1/2}).$$

Plugging in Equations (A.1) and (A.2), using that $(\boldsymbol{I}_r \otimes \boldsymbol{S}^{-1/2})\boldsymbol{K}_{r^2} = \boldsymbol{K}_{r^2}(\boldsymbol{S}^{-1/2} \otimes \boldsymbol{I}_r)$ and that $\boldsymbol{V}^\top = \boldsymbol{U}\boldsymbol{S}^{-1/2}$ we obtain that, up to $o_p(m^{-1/2})$,

$$\operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V}) = \Big\{\boldsymbol{K}_{r^2}(\boldsymbol{I}_r \otimes \boldsymbol{U})\,(\boldsymbol{S} \ominus \boldsymbol{S})^*(\boldsymbol{V} \otimes \boldsymbol{U}^\top) - 1/2(\boldsymbol{V}^\top \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{V} \overset{r}{\otimes} \boldsymbol{V})\Big\}\operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}).$$

The first term in the expression in the curly braces can be beautified by inserting the diagonal matrix $(\boldsymbol{I}_r \otimes \boldsymbol{S}^{-1/2})(\boldsymbol{I}_r \otimes \boldsymbol{S})(\boldsymbol{I}_r \otimes \boldsymbol{S}^{-1/2})$ and re-arranging, and subsequently using the definition of $\boldsymbol{V}$ to arrive at, again up to $o_p(m^{-1/2})$,

$$\begin{aligned}\operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V}) = \Big\{&\boldsymbol{K}_{r^2}(\boldsymbol{I}_r \otimes \boldsymbol{V}^\top)\,(\boldsymbol{S} \ominus \boldsymbol{S})^*(\boldsymbol{I}_r \otimes \boldsymbol{S})(\boldsymbol{V} \otimes \boldsymbol{V}) - 1/2(\boldsymbol{V}^\top \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{V} \overset{r}{\otimes} \boldsymbol{V})\Big\}\\ &\times \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A})\end{aligned}$$

as influence function for $\hat{\boldsymbol{V}}$ that we will bring forward.

Joint (approximate) diagonalizer $\hat{\boldsymbol{Q}}$. Let $\boldsymbol{C}_{\boldsymbol{x}_s} := \boldsymbol{V}\boldsymbol{A}_{\boldsymbol{x}_s}\boldsymbol{V}^\top$ and $\hat{\boldsymbol{C}}_{\boldsymbol{x}_s} := \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}_{\boldsymbol{x}_s}\hat{\boldsymbol{V}}^\top$ for all $\boldsymbol{x}_s \in \mathcal{S}$. From our results so far we know that these matrices are $m^{-1/2}$-consistent and asymptotically normal. Indeed,

$$\operatorname{vec}(\hat{\boldsymbol{C}}_{\boldsymbol{x}_s} - \boldsymbol{C}_{\boldsymbol{x}_s}) = (\boldsymbol{V} \otimes \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}}_{\boldsymbol{x}_s} - \boldsymbol{A}_{\boldsymbol{x}_s}) + (\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r^2})(\boldsymbol{V}\boldsymbol{A}_{\boldsymbol{x}_s} \otimes \boldsymbol{I}_r) \operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V}) + o_p(m^{-1/2})$$

for all $\boldsymbol{x}_s \in \mathcal{S}$. From Bonhomme and Robin (2009), but making use of the Moore-Penrose pseudo inverse to obtain a more convenient expression for our purposes, we further deduce that

$$\operatorname{vec}(\hat{\boldsymbol{Q}} - \boldsymbol{Q}) = (\boldsymbol{I}_r \otimes \boldsymbol{Q}) \left(\sum_{\boldsymbol{x}_s \in \mathcal{S}} (\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})^2 \right)^* \\ \times \left(\sum_{\boldsymbol{x}_s \in \mathcal{S}} (\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top) \operatorname{vec}(\hat{\boldsymbol{C}}_{\boldsymbol{x}_s} - \boldsymbol{C}_{\boldsymbol{x}_s}) \right) + o_p(m^{-1/2}).$$

We now aim to work out this expression in terms of the influence functions of the input matrices. First, we exploit the definitions $\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{G}\boldsymbol{D}^{1/2}$ and $\boldsymbol{V} = \boldsymbol{S}^{-1/2}\boldsymbol{U}^\top$, and the fact that

$$\boldsymbol{C}_{\boldsymbol{x}_s} = \boldsymbol{V}\boldsymbol{A}_{\boldsymbol{x}_s}\boldsymbol{V}^\top = \boldsymbol{Q}\boldsymbol{D}_{\boldsymbol{x}_s}\boldsymbol{Q}^\top$$

for all $\boldsymbol{x}_s \in \mathcal{S}$ to write $(\boldsymbol{V}\boldsymbol{A}_{\boldsymbol{x}_s} \otimes \boldsymbol{I}_r) \operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V})$ as the sum of two terms; the first of these two terms is

$$\boldsymbol{K}_{r^2}(\boldsymbol{I}_r \otimes \boldsymbol{Q})(\boldsymbol{I}_r \otimes \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{I}_r \otimes \boldsymbol{Q}^\top)(\boldsymbol{S} \ominus \boldsymbol{S})^*(\boldsymbol{I}_r \otimes \boldsymbol{S})(\boldsymbol{V} \otimes \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A})$$

and comes from the estimation noise in the eigenvectors, i.e., $\operatorname{vec}(\hat{\boldsymbol{U}} - \boldsymbol{U})$, while the second of the two terms is

$$-1/2(\boldsymbol{Q} \otimes \boldsymbol{I}_r)(\boldsymbol{D}_{\boldsymbol{x}_s} \otimes \boldsymbol{I}_r)(\boldsymbol{Q}^\top \otimes \boldsymbol{I}_r)(\boldsymbol{I}_r \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{V} \overset{r}{\otimes} \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}),$$

which comes from the noise in the eigenvalues, i.e., $\operatorname{vec}(\hat{\boldsymbol{S}}^{-1/2} - \boldsymbol{S}^{-1/2})$. Next, premultiplying each of these two terms with $(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r^2})$ and re-arranging results in

$$(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r^2})(\boldsymbol{I}_r \otimes \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{S} \ominus \boldsymbol{S})^*(\boldsymbol{I}_r \otimes \boldsymbol{S})(\boldsymbol{V} \otimes \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}), \quad \text{(A.3)}$$

and

$$-{}^1\!/{}_2(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{I}_{r^2} + \boldsymbol{K}_{r^2})(\boldsymbol{D}_{\boldsymbol{x}_s} \otimes \boldsymbol{I}_r)(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{I}_r \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{V} \overset{r}{\otimes} \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}), \quad \text{(A.4)}$$

respectively. Both expressions depend on $\boldsymbol{x}_s$ only through the matrix $\boldsymbol{D}_{\boldsymbol{x}_s}$. Now note that $\boldsymbol{K}_{r^2}$ is a symmetric and orthonormal matrix; therefore,

$$\boldsymbol{K}_{r^2}(\boldsymbol{S} \ominus \boldsymbol{S})^* = \boldsymbol{K}^*_{r^2}(\boldsymbol{S} \ominus \boldsymbol{S})^* = ((\boldsymbol{S} \ominus \boldsymbol{S})\, \boldsymbol{K}_{r^2})^* = (-\boldsymbol{K}_{r^2}(\boldsymbol{S} \ominus \boldsymbol{S}))^* = -(\boldsymbol{S} \ominus \boldsymbol{S})^* \boldsymbol{K}_{r^2}.$$

Also, $\boldsymbol{K}_{r^2}(\boldsymbol{I}_r \overset{c}{\otimes} \boldsymbol{I}_r) = (\boldsymbol{I}_r \overset{c}{\otimes} \boldsymbol{I}_r)$ and $\boldsymbol{K}_{r^2} \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}) = \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A})$ because $\hat{\boldsymbol{A}}$ and $\boldsymbol{A}$ are symmetric. Therefore, (A.3) can be written as

$$\begin{aligned} &(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{D}_{\boldsymbol{x}_s} \oplus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{S} \ominus \boldsymbol{S})^*(\boldsymbol{I}_r \otimes \boldsymbol{S})(\boldsymbol{V} \otimes \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}) \\ -&(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{D}_{\boldsymbol{x}_s} \otimes \ \boldsymbol{I}_r\ )(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{S} \ominus \boldsymbol{S})^*(\ \boldsymbol{S} \oplus \boldsymbol{S})(\boldsymbol{V} \otimes \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}), \end{aligned} \quad \text{(A.5)}$$

and (A.4) can be written as

$$-{}^1\!/{}_2(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{D}_{\boldsymbol{x}_s} \oplus \boldsymbol{D}_{\boldsymbol{x}_s})(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{I}_r \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{I}_r \overset{r}{\otimes} \boldsymbol{I}_r)(\boldsymbol{V} \otimes \boldsymbol{V}) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}), \quad \text{(A.6)}$$

where we have used that $(\boldsymbol{V} \overset{r}{\otimes} \boldsymbol{V}) = (\boldsymbol{I}_r \overset{r}{\otimes} \boldsymbol{I}_r)\,(\boldsymbol{V} \otimes \boldsymbol{V})$. If we then use the matrices defined in the main text,

$$\boldsymbol{M}_1 = (\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top) \left\{ (\boldsymbol{S} \ominus \boldsymbol{S})^*(\,\boldsymbol{I}_r \otimes \boldsymbol{S}) - {}^1\!/{}_2(\boldsymbol{I}_r \overset{c}{\otimes} \boldsymbol{I}_r)(\boldsymbol{I}_r \overset{r}{\otimes} \boldsymbol{I}_r) \right\} (\boldsymbol{V} \otimes \boldsymbol{V}),$$

and

$$\boldsymbol{M}_2 = (\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{S} \ominus \boldsymbol{S})^*(\ \boldsymbol{S} \oplus \boldsymbol{S})(\boldsymbol{V} \otimes \boldsymbol{V})$$

and write $\boldsymbol{O}_-$ for the $r^2\,|\mathcal{S}| \times r^2$ matrix obtained on concatenating $(\boldsymbol{D}_{\boldsymbol{x}_s} \ominus \boldsymbol{D}_{\boldsymbol{x}_s})$, $\boldsymbol{O}_+$ for the $r^2\,|\mathcal{S}| \times r^2$ matrix obtained on concatenating $(\boldsymbol{D}_{\boldsymbol{x}_s} \oplus \boldsymbol{D}_{\boldsymbol{x}_s})$, and $\boldsymbol{O}_\times$ for the $r^2\,|\mathcal{S}| \times r^2$matrix obtained on concatenating $(\boldsymbol{D}_{\boldsymbol{x}_s} \otimes \boldsymbol{I}_r)$, and $\boldsymbol{R}$ for the $r^2\,|\mathcal{S}| \times r^2|\mathcal{S}|$ block-diagonal matrix $(\boldsymbol{I}_{|\mathcal{S}|} \otimes \boldsymbol{W}^\top \otimes \boldsymbol{W}^\top)$, we can combine both equations compactly to finally arrive at the expression

$$\begin{aligned} \operatorname{vec}(\hat{\boldsymbol{Q}} - \boldsymbol{Q}) = (\boldsymbol{I}_r \otimes \boldsymbol{Q})\left(\boldsymbol{O}_-^\top \boldsymbol{O}_-\right)^* &\left((\boldsymbol{O}_-^\top \boldsymbol{O}_+)\boldsymbol{M}_1 - (\boldsymbol{O}_-^\top \boldsymbol{O}_\times)\boldsymbol{M}_2\right) \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}) \\ &+ (\boldsymbol{I}_r \otimes \boldsymbol{Q})\left(\boldsymbol{O}_-^\top \boldsymbol{O}_-\right)^* (\boldsymbol{O}_-^\top \boldsymbol{R}) \operatorname{vec}(\hat{\bar{\boldsymbol{A}}} - \bar{\boldsymbol{A}}) + o_p(m^{-1/2}), \end{aligned}$$

which we carry forward.

Transformation matrix $\hat{\boldsymbol{W}}$. Because $\boldsymbol{W} = \boldsymbol{V}^\top \boldsymbol{Q}$ and the estimators of the individual matrices making up $\boldsymbol{W}$ are $m^{-1/2}$-consistent and asymptotically normal we have, by a linearization

$$\operatorname{vec}(\hat{\boldsymbol{W}} - \boldsymbol{W}) = (\boldsymbol{Q}^\top \otimes \boldsymbol{I}_r)\, \boldsymbol{K}_{r^2} \operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V}) + (\boldsymbol{I}_r \otimes \boldsymbol{V}^\top) \operatorname{vec}(\hat{\boldsymbol{Q}} - \boldsymbol{Q}) + o_p(m^{-1/2}).$$

Now, using the expression for $\operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V})$ obtained above together with the definition of $\boldsymbol{M}_1$ and the fact that $\boldsymbol{Q}$ is orthonormal—so that $(\boldsymbol{Q}^\top \otimes \boldsymbol{Q}^\top)(\boldsymbol{Q} \otimes \boldsymbol{Q}) = (\boldsymbol{I}_r \otimes \boldsymbol{I}_r) = \boldsymbol{I}_{r^2}$, we see that

$$(\boldsymbol{Q}^\top \otimes \boldsymbol{I}_r)\, \boldsymbol{K}_{r^2} \operatorname{vec}(\hat{\boldsymbol{V}} - \boldsymbol{V}) = (\boldsymbol{I}_r \otimes \boldsymbol{W})\, \boldsymbol{M}_1 \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}) + o_p(m^{-1/2}),$$

while, using the definition of $\boldsymbol{W}$, $(\boldsymbol{I}_r \otimes \boldsymbol{V}^\top) \operatorname{vec}(\hat{\boldsymbol{Q}} - \boldsymbol{Q})$ can be written, up to $o_p(m^{-1/2})$, as

$$(\boldsymbol{I}_r \otimes \boldsymbol{W})\ \left(\boldsymbol{O}_-^\top \boldsymbol{O}_-\right)^* \left\{ \left((\boldsymbol{O}_-^\top \boldsymbol{O}_+)\boldsymbol{M}_1 - (\boldsymbol{O}_-^\top \boldsymbol{O}_\times)\boldsymbol{M}_2\right)\ \operatorname{vec}(\hat{\boldsymbol{A}} - \boldsymbol{A}) + \boldsymbol{O}_-^\top \boldsymbol{R} \operatorname{vec}(\hat{\bar{\boldsymbol{A}}} - \bar{\boldsymbol{A}}) \right\}.$$

Thus, with

$$\boldsymbol{N}_1 = (\boldsymbol{I}_r \otimes \boldsymbol{W}) \left\{ \left(\boldsymbol{I}_{r^2} + (\boldsymbol{O}_-^\top \boldsymbol{O}_-)^* (\boldsymbol{O}_-^\top \boldsymbol{O}_+)\right) \boldsymbol{M}_1 - (\boldsymbol{O}_-^\top \boldsymbol{O}_-)^* (\boldsymbol{O}_-^\top \boldsymbol{O}_\times) \boldsymbol{M}_2 \right\}$$

and $\boldsymbol{N}_2 = (\boldsymbol{I}_r \otimes \boldsymbol{W}) \left\{ (\boldsymbol{O}_-^\top \boldsymbol{O}_-)^* (\boldsymbol{O}_-^\top \boldsymbol{R}) \right\}$ as defined in the main text we arrive at the influence-function representation

$$\operatorname{vec}(\hat{\boldsymbol{W}} - \boldsymbol{W}) = \boldsymbol{N}_1\, {}^1\!/\!_m \sum_{i=1}^m \operatorname{vec}(\hat{\boldsymbol{A}}(\boldsymbol{X}_i) - \boldsymbol{A}) + \boldsymbol{N}_2\, {}^1\!/\!_m \sum_{i=1}^m \operatorname{vec}(\hat{\bar{\boldsymbol{A}}}(\boldsymbol{X}_i) - \bar{\boldsymbol{A}}) + o_p(m^{-1/2}).$$

Quadratic form $\hat{\vartheta}_\varphi$. A linearization readily gives, up to $o_p(m^{-1/2})$,

$$\hat{\vartheta}_\varphi - \vartheta_\varphi = (\boldsymbol{\omega}_{z_2}^\top \otimes \boldsymbol{\omega}_{z_1}^\top) \operatorname{vec}(\hat{\boldsymbol{A}}_\varphi - \boldsymbol{A}_\varphi) + \left\{ (\boldsymbol{e}_{z_2}^\top \otimes \boldsymbol{\omega}_{z_1}^\top \boldsymbol{A}_\varphi) + (\boldsymbol{e}_{z_1}^\top \otimes \boldsymbol{\omega}_{z_2}^\top \boldsymbol{A}_\varphi) \right\} \operatorname{vec}(\hat{\boldsymbol{W}} - \boldsymbol{W}),$$

where we have used that $(\boldsymbol{\omega}_{z_2}^\top \boldsymbol{A}_\varphi \otimes \boldsymbol{e}_{z_1}^\top)\, \boldsymbol{K}_{r^2} = (\boldsymbol{e}_{z_1}^\top \otimes \boldsymbol{\omega}_{z_2}^\top \boldsymbol{A}_\varphi)$. This holds for every function φ (whose variance is finite) and for all $1 \leq z_1 \leq z_2 \leq r$. From above we know that $\sqrt{m} \operatorname{vec}(\hat{\boldsymbol{A}}_\varphi - \boldsymbol{A}_\varphi)$ and $\sqrt{m} \operatorname{vec}(\hat{\boldsymbol{W}} - \boldsymbol{W})$ are both asymptotically normally distribution as $m \to \infty$. Asymptotic normality of the linear combination in the influence function then follows readily. □

Proof of Corollary 1. A delta-method argument together with Proposition 3 yields that

$$\hat{\vartheta}_{\varphi'}/\hat{\vartheta}_{\varphi''} - \vartheta_{\varphi'}/\vartheta_{\varphi''} = \frac{(\hat{\vartheta}_{\varphi'} - \vartheta_{\varphi'}) - (\vartheta_{\varphi'}/\vartheta_{\varphi''})(\hat{\vartheta}_{\varphi''} - \vartheta_{\varphi''})}{\vartheta_{\varphi''}} + o_p(m^{-1/2})$$

and that the dominant term on the right-hand side is asymptotically normal with the variance given in the corollary. □

Proof of Corollary 2. This is an immediate consequence of Proposition 3. □

References

Ahmapoor, M. and B. F. Jones (2019). Decoding team and individual impact in science and invention. *Proceedings of the National Academy of Sciences 116*, 13885–13890.

Akcigit, U., S. Baslandze, and S. Stantcheva (2016). Taxation and the international mobility of investors. *American Economic Review 106*, 2930–2981.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis 11*, 581–598.

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.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics 34*, 122–148.

Barrat, A., M. Bartélemy, R. Pastor-Satorras, and A. Vespignani (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences 101*, 3747–3752.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019). Who becomes an inventor in America? The importance of exposure to innovation. *Quarterly Journal of Economics 134*, 647–713.

Belouchrani, A., K. Abed-Meraim, J.-F. Cardoso, and E. Moulines (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing 45*, 434–444.

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. (2021). Teams: Heterogeneity, sorting, and complementarity. Proceedings of the 15th World Congress of the Econometric Society, to appear.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016a). Estimating multivariate latent-structure models. *Annals of Statistics 44*, 540–563.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016b). Nonparametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society, Series B 78*, 211–229.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2017). Nonparametric estimation of non-exchangeable latent-variable models. *Journal of Econometrics 201*, 237–248.

Bonhomme, S. and J.-M. Robin (2009). Consistent noisy independent component analysis. *Journal of Econometrics 149*, 12–25.

Cardoso, J.-F. and A. Souloumiac (1993). Blind beamforming for non-Gaussian signals. *IEEE-Proceedings, F 140*, 362–370.

Channarond, A., J.-J. Daudin, and S. Robin (2011). Classification and estimation in the stochastic block model based on the empirical degrees. Mimeo.

Ductor, L., M. Fafchamps, S. Goyal, and M. J. van der Leij (2014). Social networks and research output. *Review of Economics and Statistics 96*, 936–948.

Erdős, P. and A. Rényi (1959). On random graphs. *Publicationes Mathematicae 6*, 290–297.

Frank, O. and F. Harary (1982). Cluster inference by using transitivity indices in empirical graphs. *Journal of the American Statistical Association 77*, 835–840.

Higgins, A. and K. Jochmans (2023). Identification of mixtures of dynamic discrete choices. *Journal of Econometrics 237*, 105462.

Holland, P., K. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social Networks 5*, 109–137.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. Technical report, Institute of Advanced Study, Princeton.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics 144*, 27–61.

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 (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica 77*, 135–175.

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.

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic blockmodels. *Annals of Statistics 43*, 215–237.

Magnus, J. R. (1985). On differentiating eigenvalues and eigenvectors. *Econometric Theory 1*, 179–191.

Magnus, J. R. and H. Neudecker (1979). The commutation matrix: Some properties and applications. *Annals of Statistics 7*, 381–394.

Mariadassou, M., S. Robin, and C. Vacher (2010). Uncovering latent structure in valued graphs: A variational approach. *Annals of Applied Statistics 4*, 715–742.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*, 415–444.

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E 70*, 056131.

Newman, M. E. J. and E. A. Leicht (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences 104*, 9564–9569.

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.

von Luxburg, U., M. Belkin, and O. Bousquet (2008). Consistency of spectral clustering. *Annals of Statistics 36*, 555–586.

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 84*, 1185–1194.