



Weighted stochastic block model

Tin Lok James Ng¹ · Thomas Brendan Murphy²

Accepted: 23 August 2021 / Published online: 13 September 2021
© The Author(s) 2021

Abstract

We propose a weighted stochastic block model (WSBM) which extends the stochastic block model to the important case in which edges are weighted. We address the parameter estimation of the WSBM by use of maximum likelihood and variational approaches, and establish the consistency of these estimators. The problem of choosing the number of classes in a WSBM is addressed. The proposed model is applied to simulated data and an illustrative data set.

Keywords Weighted stochastic block model · Variational estimators · Maximum likelihood estimators · Consistency · Model selection

1 Introduction

Networks are used in many scientific disciplines to represent interactions among objects of interest. For example, in the social sciences, a network typically represents social ties between actors. In biological sciences, a network can represent interactions between proteins.

The stochastic block model (SBM) (Holland et al. 1983; Snijders and Nowicki 1997) is a popular generative model which partitions vertices into latent classes. Conditional on the latent class allocations, the connection probability between two vertices depends only on the latent classes to which the two vertices belong. Many extensions of the class SBM have been proposed which include the degree correlated SBM (Karrer and Newman 2011; Peng and Carvalho 2016), mixed membership SBM (Airoldi et al. 2008) and overlapping SBM (Latouche et al. 2011).

✉ Tin Lok James Ng
jamesng1105@gmail.com

¹ School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

² School of Mathematics and Statistics, University College Dublin, Dublin, Ireland

The SBM and many of its variants are usually restricted to Bernoulli networks. However, many binary networks are produced after applying a threshold to a weighted relationship (Ghasemian et al. 2016) which results in the loss of potentially valuable information. Although most of the literature has focused on binary networks, there is a growing interest in weighted graphs (Barrat et al. 2004; Newman 2004; Peixoto 2018).

In particular, a number of clustering methods have been proposed for weighted graphs including algorithm based and model based methods. Algorithm based methods for clustering of weighted graphs can be further divided into two classes: algorithms which do not explicitly optimize any criteria (Pons and Latapy 2005; von Luxburg 2007) and those directly optimize a criterion (Clauset et al. 2004; Stouffer and Bascompte 2011). Model based methods (Mariadassou et al. 2010; Aicher et al. 2013, 2015; Ludkin 2020) attempt to take into account the random variability in the data. A recent review of graph clustering methods is given by (Leger et al. 2014).

Mariadassou et al. (2010) presents a Poisson mixture random graph model for integer valued networks and proposes a variational inference approach for parameter estimation. The model can account for covariates via a regression model. In Zanghi et al. (2010), a mixture modelling framework is considered for random graphs with discrete or continuous edges. In particular, the edge distribution is assumed to follow an exponential family distribution. Aicher et al. (2013) proposed a general class of weighted stochastic block model for dense graphs where edge weights are assumed to be generated according to an exponential family distribution. In particular, their construction produces complete graphs, in which every pair of vertices is connected by some real-valued weight. Since most real-world networks are sparse, the constructed model cannot be applied directly. To address this shortcoming, Aicher et al. (2015) extends the work of Aicher et al. (2013) and models the edge existence using a Bernoulli distribution and the edge weights using an exponential family distribution. The contributions of edge-existence distribution and edge-weight distribution in the likelihood function are then combined via a simple tuning parameter. However, their construction does not result in a generative model and it is not obvious how to simulate network observations from the proposed model. More recently, Ludkin (2020) presents a generalization of the SBM which allows arbitrary edge weight distributions and proposes a reversible jump Markov chain Monte Carlo sampler for estimating the parameters and the number of blocks. However, the use of continuous probability distribution to model the edge weights implies that the resulting graph is complete whereby every edge is present. This assumption is unrealistic for many applications whereby a certain proportion of the real-valued edges is 0. Haj et al. (2020) presents a binomial SBM for weighted graphs and proposes a variational expectation maximization algorithm for parameter estimation.

In this paper, we propose a weighted Stochastic Block model (WSBM) with gamma weights which aims to capture the information of weights directly using a generative model. Both maximum likelihood estimation and variational methods are considered for parameter estimation where consistency results are derived. We also address the problem of choosing the number of classes using the Integrated

Completed Likelihood (ICL) criteria (Biernacki et al. 2000). The proposed models and inference methodology are applied to an illustrative data set.

2 Model specification

In this section, we present the weighted stochastic block model in detail and introduce the main notations and assumptions.

We let $\Omega = (\mathcal{V}, \mathcal{X}, \mathcal{Y})$ denote the set of directed weighted random graphs where $\mathcal{V} = \mathbb{N}$ is the set of countable vertices, $\mathcal{X} = \{0, 1\}^{\mathbb{N} \times \mathbb{N}}$ is the set of edge-existence adjacency matrix, and $\mathcal{Y} = \mathbb{R}_+^{\mathbb{N} \times \mathbb{N}}$ is the set of weighted adjacency matrix. Given a random adjacency matrix $X = \{X_{ij}\}_{i,j \in \mathbb{N}}$, $X_{ij} = 1$ if an edge exists from vertex i to vertex j and $X_{ij} = 0$ otherwise. The associated weighted random adjacency matrix is given by: for $i \neq j$, if $X_{ij} = 1$, $Y_{ij} > 0$, and $Y_{ij} = 0$ otherwise. Let \mathbb{P} be a probability measure on Ω .

2.1 Generative model

We now describe the procedure of generating a sample of random graph (V, X, Y) with n vertices from Ω .

- Let $Z_{[n]} = (Z_1, \dots, Z_n)$ be the vector of latent block allocations for the vertices, and set $\theta = (\theta_1, \dots, \theta_Q)$ with $\sum_q \theta_q = 1$. For each vertex v_i , draw its block label $Z_i \in \{1, \dots, Q\}$ from a multinomial distribution

$$Z_i \sim \mathcal{M}(1; \theta_1, \dots, \theta_Q).$$

- Let $\pi = (\pi_{ql})_{q,l=1}^Q$ be a $Q \times Q$ matrix with entries in $[0, 1]$. Conditional on the block allocations $Z_{[n]}$, the entries X_{ij} for $i \neq j$ of the edge-existence adjacency matrix $X_{[n]}$ is generated from independently a Bernoulli distribution

$$X_{ij} | Z_i = q, Z_j = l \sim \mathcal{B}(\pi_{ql}).$$

- Let $\alpha = (\alpha_{ql})_{q,l=1}^Q$ and $\beta = (\beta_{ql})_{q,l=1}^Q$ be $Q \times Q$ matrices with entries taking values in the positive reals. Conditional on the latent block allocations $Z_{[n]}$ and edge-existence adjacency matrix X , the weighted adjacency matrix $Y_{[n]}$ is generated independently from

$$Y_{ij} | X_{ij} = 1, Z_i = q, Z_j = l \sim \text{Ga}(\alpha_{ql}, \beta_{ql}),$$

$$Y_{ij} | X_{ij} = 0, Z_i = q, Z_j = l \sim \delta_{\{0\}},$$

where $\text{Ga}(\cdot, \cdot)$ denotes the gamma distribution and $\delta_{\{\cdot\}}$ is the Dirac delta function.

The generative framework described above is a straightforward extension of the binary stochastic block model whereby a positive weight is generated according to a

gamma distribution for each edge. In particular, (X, Z) is a realization of the binary directed SBM. The gamma distribution is chosen due to its flexibility in the sense that, depending on the value of its shape parameter, it can represent distributions of different shapes.

The log-likelihood of the observations $X_{[n]}$ and $Y_{[n]}$ is given by

$$\mathcal{L}_2(Y_{[n]}, X_{[n]}; \theta, \pi, \alpha, \beta) = \log \left(\sum_{z_{[n]}} e^{\mathcal{L}_1(Y_{[n]}, X_{[n]}; z_{[n]}, \pi, \alpha, \beta)} \mathbb{P}\{Z_{[n]} = z_{[n]}\} \right), \tag{1}$$

where the sum is over all possible latent block allocations, $\mathbb{P}\{Z_{[n]} = z_{[n]}\} = \prod_{i=1}^n \theta_{z_i}$ is the probability of latent block allocation $z_{[n]}$, and

$$\begin{aligned} &\mathcal{L}_1(Y_{[n]}, X_{[n]}; z_{[n]}, \pi, \alpha, \beta) \\ &= \sum_{i \neq j} \left[X_{ij} \{ \log \pi_{z_i, z_j} + \log f(Y_{ij}; \alpha_{z_i, z_j}, \beta_{z_i, z_j}) \} + (1 - X_{ij}) \log(1 - \pi_{z_i, z_j}) \right] \\ &= \sum_{i \neq j} \left\{ X_{ij} \log \pi_{z_i, z_j} + (1 - X_{ij}) \log(1 - \pi_{z_i, z_j}) \right\} \\ &\quad + X_{ij} \left\{ \alpha_{z_i, z_j} \log \beta_{z_i, z_j} + (\alpha_{z_i, z_j} - 1) \log Y_{ij} - \beta_{z_i, z_j} Y_{ij} - \log \Gamma(\alpha_{z_i, z_j}) \right\} \end{aligned} \tag{2}$$

is the complete data log-likelihood, where $f(\cdot; a, b)$ is the gamma probability density function with shape parameter a and rate parameter b ,

2.2 Assumptions

We present several assumptions needed for identifiability and consistency of maximum likelihood estimates. The following four assumptions were presented in Celisse et al. (2012) and are needed in this paper.

Assumption 1 For every $q \neq q'$, there exists $l \in \{1, \dots, Q\}$ such that

$$\pi_{q,l} \neq \pi_{q',l} \text{ or } \pi_{l,q} \neq \pi_{l,q'}.$$

Assumption 2 There exists $\zeta \in (0, 1)$ such that for all $(q, l) \in \{1, \dots, Q\}^2$

$$\pi_{ql} \in [\zeta, 1 - \zeta].$$

Assumption 3 There exists $0 < \gamma < 1/Q$ such that for all $q \in \{1, \dots, Q\}$,

$$\theta_q \in [\gamma, 1 - \gamma].$$

Assumption 4 There exists $0 < \gamma < 1/Q$ and $n_0 \in \mathbb{N}^*$ such that for all $q \in \{1, \dots, Q\}$, for all $n \geq n_0$,

$$\frac{N_q(z_{[n]}^*)}{n} \geq \gamma,$$

where $N_q(z_{[n]}^*) = |\{1 \leq i \leq n : z_i^* = q\}|$ and $z_{[n]}^*$ is any realized block allocation under the WSBM.

(A1) requires that no two classes have the same connectivity probabilities. If this assumption is violated, the resulting model has too many classes and is non-identifiable. (A2) requires that the connectivity probability between any two classes strictly lies within a closed subset of the unit interval. Note that this assumption is slightly more restrictive compared to assumption 2 of Celisse et al. (2012) in that we do not consider the boundary cases where $\pi_{q,l} \in \{0, 1\}$. The boundary cases require special treatment and are not pursued in this paper. (A3) ensures that no class is empty with high probability while (A4) is the empirical version of (A3). We note that (A4) is satisfied asymptotically under the generative framework in Sect. 2.1 since the block allocations are generated according to a multinomial distribution.

In addition to the four assumptions above, we also have the following constraints on the gamma parameters.

Assumption 5 For every $q \neq q'$, there exists $l \in \{1, \dots, Q\}$ such that

$$(\alpha_{q,l}, \beta_{q,l}) \neq (\alpha_{q',l}, \beta_{q',l}) \text{ or } (\alpha_{l,q}, \beta_{l,q}) \neq (\alpha_{l,q'}, \beta_{l,q'})$$

(A5) requires that no two classes have the same weight distribution. This assumption is the exact counterpart of (A1).

The log-likelihood function (1) contains degeneracies that prevent the direct estimation of parameters $\theta, \pi, \alpha, \beta$. To see this, we note that the probability density function of a gamma distribution $\text{Ga}(a, b)$ is given by

$$f(y; a, b) = \frac{y^{a-1} \exp(-by)b^a}{\Gamma(a)}.$$

By Stirling’s formula, we have

$$\Gamma(a) = \sqrt{2\pi}a^{a-1/2} \exp(-a)(1 + \mathcal{O}(a^{-1})).$$

Setting $y = ab$,

$$f(y; a, b) = \frac{y^{a-1} \exp(-a)(a/y)^a}{\Gamma(a)} = \frac{1}{\sqrt{2\pi y}} \frac{\sqrt{a}}{1 + \mathcal{O}(a^{-1})}.$$

Therefore, letting $a \rightarrow \infty$ while keeping $ab = y$, we have $f(y; a, b) \rightarrow \infty$. One can therefore show that the log-likelihood function is unbounded above. To avoid likelihood degeneracy, we compactify the parameter space. That is, we restrict the parameter space to a compact subset which contains the true parameters. Therefore, we have the following assumption.

Assumption 6 There exists $0 < \alpha_c < \alpha_C < \infty$ and $0 < \beta_c < \beta_C < \infty$ such that for all $(q, l) \in \{1, \dots, Q\}$,

$$\alpha_c \leq \alpha_{ql} \leq \alpha_C, \beta_c \leq \beta_{ql} \leq \beta_C.$$

With this assumption, it is easy to see that the log-likelihood function is bounded for any sample size.

2.3 Identifiability

Sufficient conditions for identifiability of binary SBM with two classes have been first obtained by Allman et al. (2009). Celisse et al. (2012) show that the SBM parameters are identifiable up to a permutation of class labels under the conditions that $\pi\theta$ has distinct coordinates and $n \geq 2Q$. The condition on $\pi\theta$ is mild since the set of vectors violating this assumption has Lebesgue measure 0. The identifiability of weighted SBM is more challenging where the only known result (Section 4 of Allman et al. (2011)) requires all entries of (π, α, β) to be distinct. We note that the assumptions in the previous section are not necessarily sufficient but are necessary to ensure that the identifiability of the parameters.

3 Asymptotic recovery of class labels

We study the posterior probability distribution of the class labels $Z_{[n]}$ given the random adjacency matrix $X_{[n]}$ and weight matrix $Y_{[n]}$, which is denoted by $\mathbb{P}(Z_{[n]}|X_{[n]}, Y_{[n]})$. Since $X_{[n]}$ and $Y_{[n]}$ are random, $\mathbb{P}(Z_{[n]}|X_{[n]}, Y_{[n]})$ is also random.

Let $P^*(X_{[n]}, Y_{[n]}) := \mathbb{P}(X_{[n]}, Y_{[n]}|Z_{[n]} = z_{[n]}^*)$ be the true conditional distribution of $(X_{[n]}, Y_{[n]})$ which depends on the true parameters $(\theta^*, \pi^*, \alpha^*, \beta^*)$. We study the convergence rate of $\mathbb{P}(Z_{[n]}|X_{[n]}, Y_{[n]})$ towards 1 with respect to P^* .

The matrices π, α, β are permutation-invariant if one permutes both its rows and columns according to some permutation $\sigma : \{1, \dots, Q\} \rightarrow \{1, \dots, Q\}$. Let $\pi^\sigma, \alpha^\sigma, \beta^\sigma$ be the matrices defined by

$$\pi_{ql}^\sigma = \pi_{\sigma(q), \sigma(l)}, \alpha_{ql}^\sigma = \alpha_{\sigma(q), \sigma(l)}, \beta_{ql}^\sigma = \beta_{\sigma(q), \sigma(l)}$$

and define the set

$$\Sigma = \{ \sigma : \{1, \dots, Q\} \rightarrow \{1, \dots, Q\} | \pi^\sigma = \pi, \alpha^\sigma = \alpha, \beta^\sigma = \beta \}.$$

Two vectors of class labels z and z' are equivalent if there exists $\sigma \in \Sigma$ such that $z'_i = \sigma(z_i)$, for all i . We let $[z]$ denote the equivalence class of z and will omit the square-brackets in the equivalence class notation as long as no confusion arises.

The following result extends Theorem 3.1 of Celisse et al. (2012) to the case of WSBM.

Theorem 1 Under assumptions (A1)–(A6), for every $t > 0$,

$$P^* \left[\sum_{[z_{[n]}] \neq [z_{[n]}^*]} \frac{\mathbb{P}([Z_{[n]}] = [z_{[n]}] | X_{[n]}, Y_{[n]})}{\mathbb{P}([Z_{[n]}] = [z_{[n]}^*] | X_{[n]}, Y_{[n]})} > t \right] = \mathcal{O}(ne^{-\kappa n})$$

uniformly with respect to z^* , and for some $\kappa > 0$ depending only on π^*, α^*, β^* but not on z^* . Here $z^* = (z_i^*)_{i=1}^\infty$ with $z_i^* \in \{1, \dots, Q\}$. Furthermore, P^* can be replaced by \mathbb{P} under assumptions (A1)–(A3) and (A5)–(A6).

4 Maximum likelihood estimation of WSBM parameters

For the binary SBM, consistency of parameter estimation have been shown for profile likelihood maximization (Bickel and Chen 2009), spectral clustering method (Rohe et al. 2011), method of moments approach (Bickel et al. 2011), method based on empirical degrees (Channarond et al. 2012), and others (Choi et al. 2012). Consistency of both maximum likelihood estimation and variational approximation method are established in Celisse et al. (2012) and Bickel et al. (2013) where asymptotic normality is also established in Bickel et al. (2013). Abbe (2018) reviews recent development in the stochastic block model and community detections.

Ambroise and Matias (2012) proposes a general class of sparse and weighted SBM where the edge distribution may exhibit any parametric form and studies the consistency and convergence rates of various estimators considered in their paper. However, their model requires the edge existence parameter to be constant across the graph, that is, $\pi_{ql} = \pi$, for all q, l , or π_{ql} can be modelled as $\pi_{ql} = a_1 I_{q=l} + a_2 I_{q \neq l}$ where I is the indicator function and $a_1 \neq a_2$. Furthermore, they also assume that conditional on the block assignments, the edge weight $Y_{ij} | X_i = q, X_j = l$ is modelled using a parametric distribution with a single parameter θ_{ql} . They further impose the restriction that

$$\theta_{ql} = \begin{cases} \theta_{in} & \text{if } q = l, \\ \theta_{out} & \text{if } q \neq l. \end{cases}$$

These assumptions are more restrictive than those imposed in this paper. Jog and Loh (2015) studies the problem of characterizing the boundary between success and failure of MLE when edge weights are drawn from discrete distributions. More recently, Brault et al. (2020) studies the consistency and asymptotic normality of the MLE and variational estimators for the latent block model which is a generalization of the SBM. However, the model considered in Brault et al. (2020) is restricted to the dense setting and requires the observations in the data matrix to be modelled by univariate exponential family distributions.

This section addresses the consistency of the MLE of WSBM. In particular, we extend the results obtained in the pioneering paper of Celisse et al. (2012) to the case of weighted graphs. Our proof closely follows the proof of consistency of the MLE in Celisse et al. (2012). The MLE consistency proof of (π, α, β) and θ require different treatments since there are $n(n - 1)$ edges but only n vertices. The following result established the MLE consistency of (π, α, β) .

Theorem 2 Assume that assumptions (A1), (A2), (A3), (A5), (A6) hold. Let us define the MLE of $(\theta^*, \pi^*, \alpha^*, \beta^*)$ by

$$(\hat{\theta}, \hat{\pi}, \hat{\alpha}, \hat{\beta}) := \arg \max_{\theta, \pi, \alpha, \beta} \mathcal{L}_2(Y_{[n]}, X_{[n]}; \theta, \pi, \alpha, \beta).$$

Then for any metric $d(\cdot, \cdot)$ on (π, α, β) ,

$$d((\hat{\pi}, \hat{\alpha}, \hat{\beta}), (\pi^*, \alpha^*, \beta^*)) \xrightarrow{P} n \rightarrow \infty \mathbb{P}0.$$

Under additional assumption on the rate of convergence of the estimators $(\hat{\pi}, \hat{\alpha}, \hat{\beta})$ of (π, α, β) , consistency of $\hat{\theta}$ can be established.

Theorem 3 Let $(\hat{\theta}, \hat{\pi}, \hat{\alpha}, \hat{\beta})$ denote the MLE of $(\theta^*, \pi^*, \alpha^*, \beta^*)$ and assume that $\|\hat{\pi} - \pi^*\|_\infty = o_P(\sqrt{\log n}/n)$, $\|\hat{\alpha} - \alpha^*\|_\infty = o_P(\sqrt{\log n}/n)$, and $\|\hat{\beta} - \beta^*\|_\infty = o_P(\sqrt{\log n}/n)$, then

$$d(\hat{\theta}, \theta^*) \xrightarrow{P} n \rightarrow \infty \mathbb{P}0$$

for any metric d in \mathbb{R}^Q .

5 Variational estimators

Direct maximization of the log-likelihood function is intractable except for very small graphs since it involves a sum over Q^n terms. In practice, approximate algorithms such as Markov Chain Monte Carlo (MCMC) and variational inference algorithms are often used for parameter inference. For the SBM, both MCMC and variational inference approaches have been proposed (Snijders and Nowicki 1997; Daudin et al. 2008). Variational inference algorithms have also been developed for mixed membership SBM (Airoldi et al. 2008), overlapping SBM (Latouche et al. 2011), and the weighted SBM proposed in Aicher et al. (2013). This section develops a variational inference algorithm for the WSBM which can be considered a natural extension of the algorithm proposed in Daudin et al. (2008) for the SBM.

The variational method consists in approximating $P(Z_{[n]} = \cdot | X_{[n]}, Y_{[n]})$ by a product of n multinomial distributions. Let \mathcal{D}_n denote a set of product multinomial distributions

$$\mathcal{D}_n = \left\{ D_{\tau_{[n]}} = \prod_{i=1}^n \mathcal{M}(1; \tau_{i,1}, \dots, \tau_{i,Q}) \mid \tau_{[n]} \in \mathcal{S}_n \right\}$$

where

$$\mathcal{S}_n = \left\{ \tau_{[n]} = (\tau_1, \dots, \tau_n) \in ([0, 1]^Q)^n \mid \text{for all } i, \tau_i = (\tau_{i,1}, \dots, \tau_{i,Q}), \sum_{q=1}^Q \tau_{i,q} = 1 \right\}.$$

For any $D_{\tau_{[n]}} \in \mathcal{D}_n$, the variational log-likelihood is defined by

$$\mathcal{J}(Y_{[n]}, X_{[n]}; \tau_{[n]}, \theta, \pi, \alpha, \beta) = \mathcal{L}_2(Y_{[n]}, X_{[n]}; \theta, \pi, \alpha, \beta) - \mathbb{K}\mathbb{L}(D_{\tau_{[n]}} \llcorner P(\cdot \mid X_{[n]}, Y_{[n]})).$$

Here $\mathbb{K}\mathbb{L}(\cdot, \cdot)$ denotes the Kullback-Leibler divergence between two probability distributions, which is nonnegative. Therefore, \mathcal{J} provides a lower bound on the log-likelihood function. We have that

$$\begin{aligned} \mathcal{J}(Y_{[n]}, X_{[n]}; \tau_{[n]}, \theta, \pi, \alpha, \beta) &= \sum_{i \neq j} \sum_{q,l} \tau_{i,q} \tau_{j,l} \left(\log \mathbf{b}(X_{ij}; \pi_{ql}) + X_{ij} \log f(Y_{ij}; \alpha_{ql}, \beta_{ql}) \right) \\ &\quad - \sum_i \sum_q \tau_{i,q} (\log \tau_{i,q} - \log \theta_q) \end{aligned}$$

where $\mathbf{b}(\cdot; p)$ denotes the probability mass function of a Bernoulli distribution with parameter p , and recall that $f(\cdot; \alpha_{ql}, \beta_{ql})$ denotes the density function of a gamma distribution $\text{Ga}(\alpha_{ql}, \beta_{ql})$.

The variational algorithm works by iteratively maximizing the lower bound \mathcal{J} with respect to the approximating distribution $D_{\tau_{[n]}}$, and estimating the model parameters. Maximization of \mathcal{J} with respect to $D_{\tau_{[n]}}$ consists of solving

$$\hat{\tau}_{[n]} := \operatorname{argmax}_{\tau_{[n]}} \mathcal{J}(Y_{[n]}, X_{[n]}; \tau_{[n]}, \theta, \pi, \alpha, \beta),$$

where $\theta, \pi, \alpha, \beta$ can be replaced by plug-in estimates. This has a closed-form solution given by

$$\hat{\tau}_{iq} \propto \theta_q \prod_{j \neq i} \prod_l \mathbf{b}(X_{ij}; \pi_{ql})^{\hat{\tau}_{jl}} f(Y_{ij}; \alpha_{ql}, \beta_{ql})^{\hat{\tau}_{ij} X_{ij}}. \tag{3}$$

Conditional on $\hat{\tau}_{[n]}$, the variational estimators of $(\theta, \pi, \alpha, \beta)$ are found by solving,

$$(\tilde{\theta}, \tilde{\pi}, \tilde{\alpha}, \tilde{\beta}) = \operatorname{argmax}_{\theta, \pi, \alpha, \beta} \mathcal{J}(Y_{[n]}, X_{[n]}; \hat{\tau}_{[n]}, \theta, \pi, \alpha, \beta).$$

Closed-form updates for $\tilde{\theta}$ and $\tilde{\pi}$ exist and are given by

$$\tilde{\theta}_q = \frac{1}{n} \sum_i \hat{\tau}_{iq} \tag{4}$$

$$\tilde{\pi}_{ql} = \frac{\sum_{i \neq j} \hat{\tau}_{iq} \hat{\tau}_{jl} X_{ij}}{\sum_{i \neq j} \hat{\tau}_{iq} \hat{\tau}_{jl}}. \tag{5}$$

On the other hand, updates for $\tilde{\alpha}$ and $\tilde{\beta}$ do not have a closed form since the maximum likelihood estimators of the two parameters of a gamma distribution do not have closed forms. However, using the fact that a gamma distribution is a

special case of a generalized gamma distribution, Ye and Chen (2017) derived simple closed-form estimators for the two parameters of gamma distribution. The estimators were shown to be strongly consistent and asymptotically normal. For $q, l = 1, \dots, Q$, let us define the quantities

$$\begin{aligned} \tilde{W}_{ql} &= \sum_{i \neq j, X_{ij}=1} \hat{\tau}_{iq} \hat{\tau}_{jl} \\ \tilde{U}_{ql} &= \sum_{i \neq j, X_{ij}=1} \hat{\tau}_{iq} \hat{\tau}_{jl} Y_{ij} \\ \tilde{V}_{ql} &= \sum_{i \neq j, X_{ij}=1} \hat{\tau}_{iq} \hat{\tau}_{jl} \log Y_{ij} \\ \tilde{S}_{ql} &= \sum_{i \neq j, X_{ij}=1} \hat{\tau}_{iq} \hat{\tau}_{jl} Y_{ij} \log Y_{ij}, \end{aligned}$$

then the updates for α_{ql}, β_{ql} are given by

$$\tilde{\alpha}_{ql} = \frac{\tilde{W}_{ql} \tilde{U}_{ql}}{\tilde{W}_{ql} \tilde{S}_{ql} - \tilde{V}_{ql} \tilde{U}_{ql}} \tag{6}$$

$$\tilde{\beta}_{ql} = \frac{\tilde{W}_{ql}^2}{\tilde{W}_{ql} \tilde{S}_{ql} - \tilde{V}_{ql} \tilde{U}_{ql}}. \tag{7}$$

We obtain the variational estimators $(\tilde{\theta}, \tilde{\pi}, \tilde{\alpha}, \tilde{\beta})$ by computing (3), (4), (5), (6), (7) until convergence.

We now address the consistency of the variational estimators derived above. The following two propositions are the counterpart of Theorem 2 and 3 for variational estimators. We omit the proof since they follow similar arguments as the proof of Corollary 4.3 and Theorem 4.4 of Celisse et al. (2012).

Proposition 1 *Assume that assumptions (A1), (A2), (A3), (A5) and (A6), and let $(\tilde{\theta}, \tilde{\pi}, \tilde{\alpha}, \tilde{\beta})$ be the variational estimators defined above. Then for any distance $d(\cdot, \cdot)$ on (π, α, β) ,*

$$d((\tilde{\pi}, \tilde{\alpha}, \tilde{\beta}), (\pi^*, \alpha^*, \beta^*)) \xrightarrow{[} n \rightarrow \infty] \mathbb{P}0.$$

Proposition 2 *Assume that the variational estimators $(\tilde{\pi}, \tilde{\alpha}, \tilde{\beta})$ converge at rate $1/n$ to $(\pi^*, \alpha^*, \beta^*)$, respectively, and assumptions (A1), (A2), (A3), (A5), (A6) hold. We have*

$$d(\tilde{\theta}, \theta^*) \xrightarrow{[} n \rightarrow \infty] \mathbb{P}0,$$

where d denotes any distance between vectors in \mathbb{R}^Q .

Note a stronger assumption on the convergence rate $1/n$ of $\tilde{\pi}, \tilde{\alpha}, \tilde{\beta}$ is assumed for Proposition 2 compared to $\sqrt{\log n}/n$ in Theorem 3. The same assumption is also used in Theorem 4.4. of Celisse et al. (2012).

6 Choosing the number of classes

In real world applications, the number of classes is typically unknown and needs to be estimated from the data. For the SBM, a number of methods have been developed to determine the number of classes, including log-likelihood ratio statistic (Wang and Bickel 2017), composite likelihood (Saldaña et al. 2017), exact integrated complete data likelihood (Côme and Latouche 2015) and Bayesian framework (Yan 2016) based methods. Model selection for variants of SBM have also been investigated (Latouche et al. 2014).

We apply the integrated classification likelihood (ICL) criterion developed by Biernacki et al. (2000) to choose the number of classes for the WSBM. The ICL is an approximation of the complete-date integrated likelihood. The ICL criterion for SBM have been derived by Daudin et al. (2008) under the assumptions that the prior distribution of (θ, π) factorizes and a non-informative Dirichlet prior on θ . Here we follow the approach of Daudin et al. (2008) to derive an approximate ICL for the WSBM.

Let m_Q denote the model with Q blocks, the ICL criterion is an approximation of the complete-data integrated likelihood:

$$\begin{aligned} &\mathcal{L}(Y_{[n]}, X_{[n]}, Z_{[n]}|m_Q) \\ &= \int_{\theta, \pi, \alpha, \beta} \mathcal{L}(Y_{[n]}, X_{[n]}, Z_{[n]}; \theta, \pi, \alpha, \beta, m_Q)g(\theta, \pi, \alpha, \beta)d\theta d\pi d\alpha d\beta \end{aligned}$$

where $g(\theta, \pi, \alpha, \beta)$ is the prior distribution of the parameters. Assuming a non-informative Jeffreys prior on θ , a Stirling approximation to the gamma function, and finally a BIC approximation to the conditional log-likelihood function, the approximate ICL can be derived. For a model m_Q with Q blocks, and assuming a non-informative Jeffreys prior $Dir(0.5, \dots, 0.5)$ on θ , the approximate ICL criterion is:

$$\begin{aligned} ICL(m_Q) &= \max_{\theta, \pi, \alpha, \beta} \log L(Y_{[n]}, X_{[n]}, \tilde{Z}_{[n]}; \theta, \pi, \alpha, \beta, m_Q) \\ &\quad - \frac{3}{2}(Q(Q + 1)) \log n(n - 1) - \frac{Q - 1}{2} \log n \end{aligned}$$

where $\tilde{Z}_{[n]}$ is the estimate of $Z_{[n]}$. The derivation of above follows exactly the same lines as the proof of Proposition 8 of Daudin et al. (2008).

7 Simulation

We validate the theoretical results developed in previous sections by conducting simulation studies. In particular, we investigate how fast parameter estimates of WSBM converge to their true values and the accuracy of posterior block allocations. Additionally, we investigate the performance of ICL in choosing the number of blocks.

7.1 Experiment 1 (two-class model)

For each fixed number of nodes n , 50 realizations of WSBM are generated based on the fixed parameter setting given in (8 and 9). The variational inference algorithm derived in Sect. 5 is then applied to estimate the model parameters and class allocations.

We can see from Table 1 that the estimated model parameters converge to their true values as the number of nodes increases while the posterior class assignment is accurate across any number of nodes. Table 2 shows the ICL criterion tends to select the correct number of classes, especially when the number of nodes is large.

$$\theta = (0.7, 0.3), \quad (8)$$

$$\pi = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.9 \end{pmatrix}, \alpha = \begin{pmatrix} 10.0 & 0.3 \\ 3.0 & 0.5 \end{pmatrix}, \beta = \begin{pmatrix} 2.0 & 1.0 \\ 0.2 & 1.0 \end{pmatrix}. \quad (9)$$

7.2 Experiment 2 (three-class model)

50 network realizations are obtained under the three-class model with parameter values given in (10 and 11) for a range of values n .

$$\theta = (0.5, 0.3, 0.2) \quad (10)$$

$$\pi = \begin{pmatrix} 0.60 & 0.20 & 0.30 \\ 0.30 & 0.90 & 0.10 \\ 0.60 & 0.50 & 0.20 \end{pmatrix}, \alpha = \begin{pmatrix} 0.50 & 2.00 & 1.00 \\ 0.30 & 0.02 & 6.00 \\ 2.00 & 0.05 & 3.00 \end{pmatrix}, \beta = \begin{pmatrix} 5.00 & 0.40 & 5.00 \\ 3.00 & 12.00 & 0.70 \\ 6.00 & 0.20 & 0.60 \end{pmatrix}. \quad (11)$$

We can see from Table 3 that the estimated parameters converge to their true values quickly as the number of nodes increases. The ICL criterion tends to overestimate the number of classes when the number of nodes is small, but consistently selects the correct model when the number of nodes is large (Table 4).

7.3 Computational complexity

The computational complexity of the variational estimators derived in Sect. 5 scales as $\mathcal{O}(n^2)$, which has the same complexity as the variational algorithm developed by Daudin et al. (2008). Therefore, the algorithm may be prohibitively expensive for

Table 1 Convergence analysis of posterior class allocations and parameter estimates under the two-class model

n	$\sum_{i=1}^n I_{z_i=z_i^*} / n$	$\ \theta - \theta^*\ _2$	$\ \pi - \pi^*\ _F$	$\ \alpha - \alpha^*\ _F$	$\ \beta - \beta^*\ _F$
25	1	0.031	0.057	0.855	0.434
50	1	0.029	0.035	0.654	0.322
100	1	0.027	0.012	0.256	0.070
200	1	0.025	0.006	0.116	0.046
500	1	0.021	0.002	0.053	0.024

Table 2 Frequency of choosing Q blocks by ICL under different number of nodes n under the two-class model

$n \mid Q$	1	2	3	4	5
25	0	44	3	3	0
50	0	48	2	0	0
100	0	50	0	0	0
200	0	50	0	0	0
500	0	50	0	0	0

networks with more than 1000 nodes. The estimated computational time for the two-class model in Sect. 7.1 and for the three-class model in Sect. 7.2 for various values of n are shown in Fig. 1. The estimated computing time at each n is the average running time of the variational algorithm over 20 replications.

8 Application: Washington bike data set

We apply the WSBM to analyse the Washington bike sharing scheme data set.¹ Information with respect to start stations and end stations of trips as well as length (travel time) of trips are available in the data set. We select a time window of one week starting from January 10th, 2016 and construct the adjacency matrix X and weight matrix Y as follows:

- $X_{ij} = 1$ if there is trip starting from station i and finishing at station j .
- Y_{ij} is the total length of trips (in minutes) from station i to station j .

The resulting network consists of 370 nodes with an average out-degree of 36.14. The average total length of trips between any pair of stations is 42.15 minutes. We apply the ICL criterion to select the number of classes for the WSBM. For each number of classes Q , the variational inference algorithm is fitted to the network 20 times with 20 random initializations and the highest value of ICL is recorded, and the six-class model is chosen (Table 5). Each bike station is plotted on the map in

¹ Historical data available at <https://www.capitalbikeshare.com/>.

Table 3 Convergence analysis of posterior class allocations and parameter estimates under the three-class model

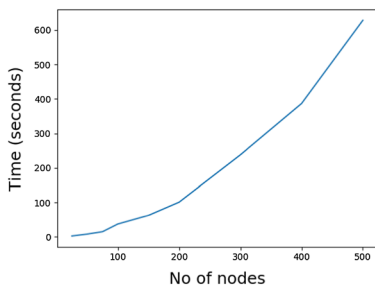
n	$\sum_{i=1}^n I_{z_i=z_i^*} / n$	$\ \theta - \theta^*\ _2$	$\ \pi - \pi^*\ _F$	$\ \alpha - \alpha^*\ _F$	$\ \beta - \beta^*\ _F$
25	0.961	0.116	0.178	7.86	136.72
50	1	0.039	0.039	1.256	7.866
100	1	0.033	0.026	0.481	1.426
200	1	0.014	0.024	0.228	1.011
500	1	0.003	0.006	0.197	0.222

Table 4 Frequency of choosing Q blocks by ICL under different number of nodes n under the three-block model

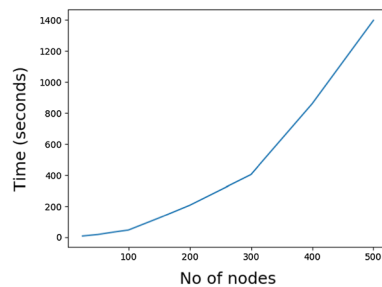
n/Q	1	2	3	4	5
25	0	3	37	8	2
50	0	0	43	7	0
100	0	0	50	0	0
200	0	0	49	1	0
500	0	0	50	0	0

Fig. 2 where its colour represents the estimated class assignment. We observe that bike stations in class 6 (colored in brown) tend to be concentrated in the central area of Washington whereas stations in class 3 (colored in red) tend to be located further from the center. Figure 2 shows some spatial effect in the class assignment of bike stations whereby stations that are close in distance tend to be in the same cluster, with the exception of class 3 (colored in red). One potential extension of the model is to take into account the spatial locations of the bike stations as covariates.

The estimated class proportions $\hat{\theta}$ shown in 12 indicate that class 3 has the largest number of stations whereas class 4 has the smallest number of stations. The estimated $\hat{\pi}$ shows that within class connectivity is generally higher compared to between class connectivity. We further observe that the connection probabilities



(a) Estimated computing time for the two-class model



(b) Estimated computing time for the three-class model

Fig. 1 Estimated computing time for the two-class and three-class models

between bike stations in class 1, 4, and 6 are substantially higher. Interestingly, we observe a near symmetry in the matrix $\hat{\pi}$ indicating that the probability of having a trip from a station in class k to another station in class l is similar with the probability of having a trip from class l to class k .

The estimated densities of travel time between each pair of classes are shown in Fig. 3. We observe that the majority of the estimated densities have mode and mean close to 0, particularly for the estimated densities in the diagonal of Fig. 3. This implies that the total travel times between stations in the same class are quite short. In comparison, the total travel time between stations in different classes tend to be longer. This is reasonable as the distance between bike stations in different classes tend to be longer which in turn requires longer travel time.

$$\hat{\theta} = (0.1985, 0.0975, 0.3256, 0.0270, 0.1377, 0.2136) \tag{12}$$

$$\hat{\pi} = \begin{pmatrix} \mathbf{0.2847} & 0.0539 & 0.0036 & \mathbf{0.4268} & 0.0125 & \mathbf{0.2811} \\ 0.0594 & 0.0630 & 0.0172 & 0.1254 & 0.0156 & 0.0791 \\ 0.0049 & 0.0156 & 0.0185 & 0.0079 & 0.0092 & 0.0029 \\ \mathbf{0.4113} & 0.1140 & 0.0136 & \mathbf{0.8438} & 0.0532 & \mathbf{0.4494} \\ 0.0187 & 0.0318 & 0.0120 & 0.0910 & \mathbf{0.2159} & 0.0241 \\ \mathbf{0.2715} & 0.0514 & 0.0017 & \mathbf{0.4965} & 0.0159 & \mathbf{0.7268} \end{pmatrix}, \tag{13}$$

9 Discussion

This paper proposes a weighted stochastic block model (WSBM) for networks. The proposed model is an extension of the stochastic block model. A variational inference strategy is developed for parameter estimation. Asymptotic properties of maximum likelihood estimators and variational estimators are derived, and the problem of choosing the number of classes is addressed by using an ICL criteria. Simulation studies are conducted to evaluate the performance of variational estimators and the use of ICL to determine the number of classes. The proposed model and inference methods are an illustrative data set.

It is straightforward to extend the WSBM to allow node covariates. Let $w_i \in \mathbb{R}^d$ be the covariates for each node $i = 1, \dots, n$, and let w_{ij} be the covariates for each pair of nodes, $i, j = 1, \dots, n, i \neq j$. The edge probability p_{ij} between a pair of nodes i, j with node i in block q and node j in block l can be modelled as

$$\log \frac{p_{ij}}{1 - p_{ij}} = \xi_{ql,0} + \xi_{ql,1}^T w_{ij} + \xi_{ql,2}^T w_i + \xi_{ql,3}^T w_j,$$

where $\xi_{ql,0} \in \mathbb{R}$ and $\xi_{ql,1}, \xi_{ql,2}, \xi_{ql,3} \in \mathbb{R}^d$. Conditional on block assignments and existence of an edge between a pair of nodes i, j , Y_{ij} can be modelled as a Gamma random variable with mean μ_{ij} and variance σ_{ij} where

Table 5 Model selection for the Washington Bike dataset using ICL criterion

Q	ICL
1	− 107157.24
2	− 93505.61
3	− 91688.33
4	− 90813.17
5	− 90300.90
6	− 89061.38
7	− 89326.35

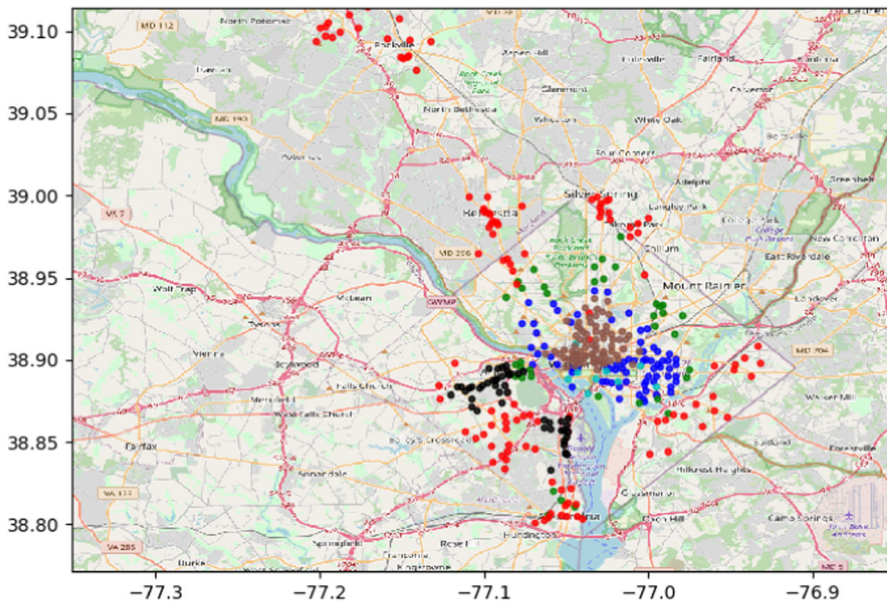


Fig. 2 Bike stations. Class 1: blue. Class 2: green. Class 3: red. Class 4: cyan. Class 5: black. Class 6: brown (color figure online)

$$\mu_{ij} = E(Y_{ij}|Z_i = q, Z_j = l, X_{ij} = 1) = \exp(\phi_{ql,0} + \phi_{ql,1}^T w_{ij} + \phi_{ql,2}^T w_i + \phi_{ql,3}^T w_j),$$

$$\sigma_{ij} = \text{Var}(Z_i = q, Z_j = l, X_{ij} = 1) = v_{ql}$$

where $\phi_{ql,0} \in \mathbb{R}$ and $\phi_{ql,1}, \phi_{ql,2}, \phi_{ql,3} \in \mathbb{R}^d$.

Many possible future extensions are possible. First, it is desirable to investigate further theoretical properties of maximum likelihood and variational estimators of WSBM parameters such as asymptotic normality of the estimators. Furthermore, some of the assumptions imposed in this work in order to ensure consistency of the estimators maybe relaxed. Moreover, the number of blocks is assumed to be fixed in the asymptotic analysis of the estimators. It would be interesting to allow the number of blocks to grows as the number of nodes grows.

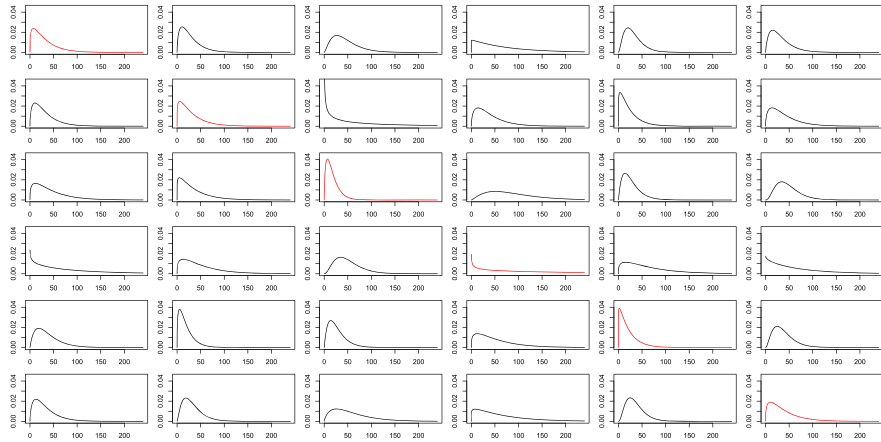


Fig. 3 Estimated densities of total travel time between stations

Auxiliary results

Definition 1 A random variable X with mean $\mu = E(X)$ is sub-exponential if there are non-negative parameters (v, b) such that

$$E(e^{t(X-\mu)}) \leq e^{\frac{v^2 t^2}{2}} \text{ for all } |t| < \frac{1}{b}.$$

The following lemma is a straight forward consequence of the definition.

Lemma 1 *If the independent random variables $\{X_i\}_{i=1}^n$ are sub-exponential with parameters (v_i, b_i) , $i = 1, \dots, n$, then $\sum_{i=1}^n X_i$ is sub-exponential with parameters $(\sqrt{\sum_{i=1}^n v_i^2}, \max_{i=1}^n b_i)$.*

The following results show that for a Gamma random variable Y and a Bernoulli random variable X , both XY and $X \log Y$ are sub-exponential random variables. They are useful for the proof of the main theorems.

Proposition 3 *If $Y \sim Ga(a, b)$ and $X \sim Ber(\pi)$, YX is a sub-exponential random variable.*

Proof The expectation of YX is given by

$$\mu := E(YX) = \pi \frac{a}{b},$$

and thus

$$E(e^{t\mu}) = \exp\left(t\pi \frac{a}{b}\right).$$

The moment generating function of YX is given by

$$E(e^{tYX}) = (1 - \pi) + \pi\left(1 - \frac{t}{b}\right)^{-a}$$

which is defined for $t < b$. Taylor series expansion of $(1 - \frac{t}{b})^{-a}$ around $t = 0$ gives that

$$\left(1 - \frac{t}{b}\right)^{-a} = 1 + \frac{a}{b}t + \frac{a(a+1)}{b^2} \frac{t^2}{2} + \mathcal{O}(t^3).$$

Similarly, Taylor series expansion of $\exp(-t\frac{a}{b}\pi)$ around 0 gives that

$$\exp\left(-t\frac{a}{b}\pi\right) = 1 - \frac{a}{b}\pi t + \frac{a(a+1)}{b^2} \pi \frac{t^2}{2} + \mathcal{O}(t^3).$$

Thus, we have

$$\begin{aligned} E(e^{tYX}) &= 1 - \pi + \pi + \frac{a}{b}\pi t + \frac{a(a+1)}{b^2} \pi \frac{t^2}{2} + \mathcal{O}(t^3) \\ &= 1 + \frac{a}{b}\pi t + \frac{a(a+1)}{b^2} \pi \frac{t^2}{2} + \mathcal{O}(t^3). \end{aligned}$$

This leads to

$$\begin{aligned} E(e^{t(YX-\mu)}) &= \left(1 + \frac{a}{b}\pi t + \frac{a(a+1)}{b^2} \pi \frac{t^2}{2} + \mathcal{O}(t^3)\right) \left(1 - \frac{a}{b}\pi t + \frac{a(a+1)}{b^2} \pi \frac{t^2}{2} + \mathcal{O}(t^3)\right) \\ &= 1 + \frac{(a^2 + 2a)\pi}{b^2} t^2 + \mathcal{O}(t^3). \end{aligned}$$

Hence, we can choose suitable v, b such that

$$E(e^{t(YX-\mu)}) \leq \exp\left(\frac{v^2 t^2}{2}\right) \text{ for all } |t| < \frac{1}{b}.$$

□

Proposition 4 *If $Y \sim Ga(a, b)$ and $X \sim Ber(\pi)$, $X \log Y$ is a sub-exponential random variable.*

Proof The proof is analogous to the proof of Proposition 3. It is straightforward to show that

$$E(e^{tX \log Y}) = (1 - \pi) + \pi \frac{\Gamma(\alpha + t)}{\Gamma(\alpha)\beta^t}$$

and

$$\mu := E(X \log Y) = \pi(\Psi(\alpha) - \log(\beta))$$

where $\Psi(\cdot)$ is the digamma function. The rest of the proof follows the same line as the proof of Proposition 3 by considering the Taylor series expansion of $e^{-\mu t}$ and $E(e^{tX \log Y})$ around $t = 0$. □

The following lemma provides an upper bound for the tail probability of a sub-exponential random variable.

Lemma 2 (Sub-exponential tail bound) *Suppose that X with mean μ is sub-exponential with parameters (v, b) . Then*

$$\mathbb{P}(X \geq \mu + t) \leq \exp \left[- \min \left(\frac{t^2}{2v^2}, \frac{t}{2b} \right) \right].$$

The following inequality for suprema of random processes is needed for the proof of Theorem 2.

Proposition 5 (Baraud 2010, Theorem 2.1) *Let $(S(g))_{g \in \mathcal{G}}$ be a family of real valued and centered random variables. Fix some g_0 in \mathcal{G} . Suppose the following two conditions hold:*

1. *There exist two arbitrary norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$ on \mathcal{G} and a nonnegative constant c such that for all $g_1, g_2 \in \mathcal{G} (g_1 \neq g_2)$,*

$$\mathbb{E} \left[e^{t(S(g_1) - S(g_2))} \right] \leq \exp \left(\frac{t^2 \|g_1 - g_2\|_2^2}{2(1 - tc \|g_1 - g_2\|_\infty)} \right)$$

for all $t \in \left[0, \frac{1}{c \|g_1 - g_2\|_\infty} \right)$.

2. *Let \mathcal{S} be a linear space with finite dimension D endowed with the two norms $\|\cdot\|_\infty, \|\cdot\|_2$ defined above, we assume that for constants $u > 0$ and $b \geq 0$,*

$$\mathcal{G} \subset \{g \in \mathcal{S} : \|g - g_0\|_\infty \leq u_1, c \|g - g_0\|_2 \leq u_2\}.$$

we have for all $x > 0$,

$$P \left[\sup_{g \in \mathcal{G}} |S(g) - S(g_0)| \geq \kappa(\sqrt{u^2(D+x)} + b(D+x)) \right] \leq 2e^{-x}.$$

where $\kappa = 18$.

Proof

In this section, we prove Theorems 1, 2, 3 for the special case $\alpha_{q,l} = \alpha_{11}$ for all $(q, l) \in \{1, \dots, Q\}$. The more general case can be proved similarly by using the fact that $X_{ij} \log Y_{ij}$ is a sub-exponential random variable.

Proof of Theorem 1

Proof Our proof is adapted from Celisse et al. (2012). Using the fact that for every $z'_{[n]} \in [z_{[n]}]$, $\mathbb{P}(z'_{[n]}|X_{[n]}, Y_{[n]}) = \mathbb{P}(z_{[n]}|X_{[n]}, Y_{[n]})$, we have that

$$\begin{aligned} \mathbb{P}(Z_{[n]} = [z_{[n]})|X_{[n]}, Y_{[n]}) &= \sum_{z'_{[n]} \in [z_{[n]}} \mathbb{P}(z'_{[n]}|X_{[n]}, Y_{[n]}) \\ &= |[z_{[n]}]| \mathbb{P}(z_{[n]}|X_{[n]}, Y_{[n]}) \end{aligned}$$

where $|[z_{[n]}]|$ is the cardinality of the equivalence class $[z_{[n]}]$.

By applying a similar approach as in (Celisse et al. (2012); ‘‘Appendix B2’’), one

can derive an upper bound for $P^* \left[\sum_{[z_{[n]}] \neq [z'_{[n]}]} \frac{\mathbb{P}([Z_{[n]}] = [z_{[n]}]|X_{[n]}, Y_{[n]})}{\mathbb{P}([Z_{[n]}] = [z'_{[n]}]|X_{[n]}, Y_{[n]})} > t \right]$:

$$\begin{aligned} P^* \left[\sum_{[z_{[n]}] \neq [z^*_{[n]}]} \frac{\mathbb{P}([z_{[n]}]|X_{[n]}, Y_{[n]})}{\mathbb{P}([z^*_{[n]}]|X_{[n]}, Y_{[n]})} > t \right] \\ \leq \sum_{r=1}^n \sum_{z_{[n]} \notin [z^*_{[n]}], \|z_{[n]} - z^*_{[n]}\|_0 = r} P^* \left[\frac{\mathbb{P}(z_{[n]}|X_{[n]}, Y_{[n]})}{\mathbb{P}(z^*_{[n]}|X_{[n]}, Y_{[n]})} > \frac{t}{n^{r+1}(Q-1)^r} \right] \end{aligned}$$

where $\|z_{[n]} - z^*_{[n]}\|_0$ is the number of difference between $z_{[n]}$ and $z^*_{[n]}$.

To simplify notation, we write $z := z_{[n]}$, $X := X_{[n]}$, and $Y = Y_{[n]}$. We have that

$$\begin{aligned}
 & P^* \left[\frac{\mathbb{P}(z|X, Y)}{\mathbb{P}(z^*|X, Y)} > \frac{t}{n^{r+1}(Q-1)^r} \right] \\
 &= P^* \left[\frac{\mathbb{P}(Y|X, z)\mathbb{P}(X|z)\mathbb{P}(z)}{\mathbb{P}(Y|X, z^*)\mathbb{P}(X|z^*)\mathbb{P}(z^*)} > \frac{t}{n^{r+1}(Q-1)^r} \right] \\
 &= P^* \left[\log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} + \log \frac{\mathbb{P}(X|z)\mathbb{P}(z)}{\mathbb{P}(X|z^*)\mathbb{P}(z^*)} > \log \frac{t}{n^{r+1}(Q-1)^r} \right] \\
 &\leq P^* \left[\left| \log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} \right) \right| \right. \\
 &> \frac{1}{2} \left(\log \frac{t}{n^{r+1}(Q-1)^r} - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} \right) - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(X|z)\mathbb{P}(z)}{\mathbb{P}(X|z^*)\mathbb{P}(z^*)} \right) \right) \Bigg] \\
 &\quad + P^* \left[\left| \log \frac{\mathbb{P}(X|z)\mathbb{P}(z)}{\mathbb{P}(X|z^*)\mathbb{P}(z^*)} - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(X|z)\mathbb{P}(z)}{\mathbb{P}(X|z^*)\mathbb{P}(z^*)} \right) \right| \right] \\
 &> \frac{1}{2} \left(\log \frac{t}{n^{r+1}(Q-1)^r} - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} \right) - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(X|z)\mathbb{P}(z)}{\mathbb{P}(X|z^*)\mathbb{P}(z^*)} \right) \right) \Bigg] \\
 &=: I_1 + I_2
 \end{aligned}$$

Consider the first term on the RHS of the last inequality,

$$\begin{aligned}
 & \log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} \right) \\
 &= \sum_{i \neq j} \left(\alpha_{11}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} \right) (X_{ij} - \pi_{z_i^*, z_j^*}^*) + (\beta_{z_i^*, z_j^*}^* - \beta_{z_i, z_j}^*) \left(Y_{ij} X_{ij} - \frac{\alpha_{11}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* \right)
 \end{aligned}$$

The summand in the summation above vanishes when $\beta_{z_i, z_j}^* = \beta_{z_i^*, z_j^*}^*$. For two vectors z and z' , we define

$$D(z, z') = \{(i, j) | i \neq j, \beta_{z_i, z_j}^* \neq \beta_{z_i', z_j'}^*\}$$

and let $N_r(z) = |D(z, z^*)|$ denote the number of terms in the summation. Set

$$\begin{aligned}
 s &= \frac{1}{N_r(z)} \frac{1}{2} \left\{ \log \frac{t}{n^{r+1}(Q-1)^r} - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} \right) \right. \\
 &\quad \left. - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(X|z)\mathbb{P}(z)}{\mathbb{P}(X|z^*)\mathbb{P}(z^*)} \right) \right\},
 \end{aligned}$$

one can show by Lemma B.3 of Celisse et al. (2012) that there exists some positive constant c^* such that $s \geq c^* > 0$. We have that

$$\begin{aligned}
 & P^* \left[\frac{1}{N_r(z)} \left| \log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} - \mathbb{E}^{Z=z^*} \left(\log \frac{\mathbb{P}(Y|X, z)}{\mathbb{P}(Y|X, z^*)} \right) \right| > 2s \right] \\
 & \leq P^* \left[\frac{1}{N_r(z)} \left| \sum_{i \neq j} \alpha_{11}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} (X_{ij} - \pi_{z_i^*, z_j^*}^*) \right| > s \right] \\
 & \quad + P^* \left[\frac{1}{N_r(z)} \left| \sum_{i \neq j} (\beta_{z_i^*, z_j^*}^* - \beta_{z_i, z_j}^*) \left(Y_{ij} X_{ij} - \frac{\alpha_{11}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* \right) \right| > s \right]
 \end{aligned}$$

Assumption (A6) implies that the random variable $\alpha_{11}^* (\log \beta_{z_i, z_j}^* / \beta_{z_i^*, z_j^*}^*) X_{ij}$ is bounded for all i, j . An application of Hoeffding’s inequality yields that

$$P^* \left[\frac{1}{N_r(z)} \left| \sum_{i \neq j} \alpha_{11}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} (X_{ij} - \pi_{z_i^*, z_j^*}^*) \right| > s \right] \leq \exp \left(-\frac{N_r(z)s^2}{L_1} \right) \tag{14}$$

for some constant $L_1 > 0$.

By Proposition 3, the random variable $Y_{ij} X_{ij}$ is sub-exponential and the tail bound for sub-exponential random variables (Lemma 2) implies that

$$\begin{aligned}
 & P^* \left[\frac{1}{N_r(z)} \left| \sum_{i \neq j} (\beta_{z_i^*, z_j^*}^* - \beta_{z_i, z_j}^*) \left(Y_{ij} X_{ij} - \frac{\alpha_{11}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* \right) \right| > s \right] \\
 & \leq \max \left\{ \exp \left(-\frac{N_r(z)s^2}{L_2} \right), \exp \left(-\frac{N_r(z)s}{L_3} \right) \right\}
 \end{aligned} \tag{15}$$

for some constants $L_2, L_3 > 0$. Proposition B.4 of Celisse et al. (2012) shows that $N_r(z)$ is bounded below by

$$N_r(z) \geq \frac{\gamma^2}{2} n \|z_{[n]} - z_{[n]}^*\|_0. \tag{16}$$

Combining inequalities (14), (15), (16), it is straight forward to show that

$$I_1 = \mathcal{O}(\exp(-A_1 n)).$$

By Theorem 3.1 of Celisse et al. (2012), we have

$$I_2 = \mathcal{O}(\exp(-A_2 n))$$

for some constant $A_2 > 0$. Therefore, with $A := \min\{A_1, A_2\}$

$$\begin{aligned}
 P^* & \left[\sum_{[z_{[n]}] \neq [z_{[n]}^*]} \frac{\mathbb{P}([z_{[n]}] | X_{[n]}, Y_{[n]})}{\mathbb{P}([z_{[n]}^*] | X_{[n]}, Y_{[n]})} > t \right] \\
 & \leq \sum_{r=1}^n \binom{n}{r} (Q-1)^r \mathcal{O}(\exp(-An)) \\
 & \rightarrow 0
 \end{aligned}$$

as $n \rightarrow \infty$. Since the upper bound does not depend on z^* , P^* can be replaced by \mathbb{P} . □

Proof of Theorem 2

Proof As the first step of the proof, we define the normalized complete data log-likelihood function

$$\begin{aligned}
 \phi_n(z_{[n]}, \pi, \alpha, \beta) & = \frac{1}{n(n-1)} \mathcal{L}_1(X_{[n]}, Y_{[n]}; z_{[n]}, \pi, \alpha, \beta) \\
 & = \frac{1}{n(n-1)} \left[\sum_{i \neq j} \left\{ X_{ij} \log \pi_{z_i, z_j} + (1 - X_{ij}) \log(1 - \pi_{z_i, z_j}) \right\} \right. \\
 & \quad \left. + X_{ij} \left\{ \alpha_{z_i, z_j} \log \beta_{z_i, z_j} + (\alpha_{z_i, z_j} - 1) \log Y_{ij} - \beta_{z_i, z_j} Y_{ij} - \log \Gamma(\alpha_{z_i, z_j}) \right\} \right],
 \end{aligned}$$

and its expectation

$$\begin{aligned}
 \Phi_n(z_{[n]}, \pi, \alpha, \beta) & = \mathbb{E} \left[\phi_n(z_{[n]}, \pi, \alpha, \beta) \mid Z_{[n]} = z_{[n]}^* \right] \\
 & = \frac{1}{n(n-1)} \left[\sum_{i \neq j} \left\{ \pi_{z_i^*, z_j^*}^* \log \pi_{z_i, z_j} + (1 - \pi_{z_i^*, z_j^*}^*) \log(1 - \pi_{z_i, z_j}) \right\} \right. \\
 & \quad \left. + \pi_{z_i^*, z_j^*}^* \left\{ \alpha_{z_i, z_j} \log \beta_{z_i, z_j} + (\alpha_{z_i, z_j} - 1) (\log \beta_{z_i^*, z_j^*}^* + \psi(\alpha_{z_i^*, z_j^*}^*)) \right. \right. \\
 & \quad \left. \left. - \beta_{z_i, z_j} \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} - \log \Gamma(\alpha_{z_i, z_j}) \right\} \right].
 \end{aligned}$$

The following proposition shows that $\phi_n(z_{[n]}, \pi, \alpha, \beta)$ uniformly converges to $\Phi_n(z_{[n]}, \pi, \alpha, \beta)$. This result is an extension of Proposition 3.5 of Celisse et al. (2012).

Proposition 6 *Under assumptions (A1), (A2), (A5), (A6), we have*

$$\sup_{\mathcal{P}} |\phi_n(z_{[n]}, \pi, \alpha, \beta) - \Phi_n(z_{[n]}, \pi, \alpha, \beta)| \xrightarrow{[n]} n \rightarrow \infty \mathbb{P}0$$

where $\mathcal{P} := \{(z_{[n]}, \pi, \alpha, \beta) : (A1), (A2), (A5), (A6)\}$.

Proof We have that

$$\begin{aligned} & |\phi_n(z_{[n]}, \pi, \alpha, \beta) - \Phi_n(z_{[n]}, \pi, \alpha, \beta)| \\ & \leq \rho_n \left| \sum_{i \neq j} (X_{ij} - \pi_{z_i^*, z_j^*}^*) \left(\log \frac{\pi_{z_i, z_j}}{1 - \pi_{z_i, z_j}} + \alpha_{z_i, z_j} \log \beta_{z_i, z_j} - \log \Gamma(\alpha_{11}) \right) \right| \\ & \quad + \rho_n \left| \sum_{i \neq j} \left(X_{ij} Y_{ij} - \pi_{z_i^*, z_j^*}^* \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \right) \beta_{z_i, z_j} \right| \\ & \quad + \rho_n \left| \sum_{i \neq j} \left(X_{ij} \log Y_{ij} - \pi_{z_i^*, z_j^*}^* \log \beta_{z_i^*, z_j^*}^* - \pi_{z_i, z_j}^* \psi(\alpha_{z_i^*, z_j^*}^*) \right) (\alpha_{z_i, z_j} - 1) \right| \end{aligned} \tag{17}$$

where $\rho_n = 1/n(n - 1)$. Notice that under Assumption (A2) and (A6),

$$\Phi_n(z_{[n]}, \pi, \alpha, \beta) < +\infty.$$

Therefore, by Proposition 3.5 of Celisse et al. (2012),

$$\sup_{\mathcal{P}} \rho_n \left| \sum_{i \neq j} (X_{ij} - \pi_{z_i^*, z_j^*}^*) \left(\log \frac{\pi_{z_i, z_j}}{1 - \pi_{z_i, z_j}} + \alpha_{z_i, z_j} \log \beta_{z_i, z_j} - \log \Gamma(\alpha_{z_i, z_j}) \right) \right| \xrightarrow{[n]} n \rightarrow \infty \mathbb{P}0$$

To bound the second term on the RHS of the inequality 17, we apply Proposition 5 (Theorem 2.1 of Baraud (2010)). We first define

$$S_n(\beta) = \sum_{i \neq j} \left(X_{ij} Y_{ij} - \pi_{z_i^*, z_j^*}^* \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \right) \beta_{z_i, z_j}.$$

Now, for two parameters $\beta^{(1)}$ and $\beta^{(2)}$,

$$\begin{aligned} S_n(\beta^{(1)}) - S_n(\beta^{(2)}) &= \sum_{i \neq j} \left(X_{ij} Y_{ij} - \pi_{z_i^*, z_j^*}^* \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \right) (\beta_{z_i, z_j}^{(1)} - \beta_{z_i, z_j}^{(2)}) \\ &= \sum_{q, l} \sum_{i \neq j, z_i = q, z_j = l} \left(X_{ij} Y_{ij} - \pi_{z_i^*, z_j^*}^* \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \right) (\beta_{ql}^{(1)} - \beta_{ql}^{(2)}). \end{aligned}$$

Since $X_{ij} Y_{ij}$ is sub-exponential with parameters (v, b) , $\sum_{i \neq j, z_i = q, z_j = l} X_i X_j$ is sub-exponential with parameters $(\sqrt{n_{ql}}v, b)$ by Lemma 3, where $n_{ql} = |\{(i, j) : z_i = q, z_j = l\}|$. Therefore, define the norms

$$\begin{aligned} \|\beta^{(1)} - \beta^{(2)}\|_2^2 &= n^2 v^2 \sum_{q,l} (\beta_{ql}^{(1)} - \beta_{ql}^{(2)})^2 \\ \|\beta^{(1)} - \beta^{(2)}\|_\infty^2 &= \sum_{q,l} (\beta_{ql}^{(1)} - \beta_{ql}^{(2)})^2, \end{aligned}$$

we have

$$\begin{aligned} &\mathbb{E} \left[\exp(t(S_n(\beta^{(1)}) - S_n(\beta^{(2)}))) \right] \\ &= \prod_{q,l} E \left[\exp \left(t(\beta_{ql}^{(1)} - \beta_{ql}^{(2)}) \sum_{i \neq j, z_i = q, z_j = l} \left(X_{ij} Y_{ij} - \pi_{z_i^*, z_j^*} \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \right) \right) \right] \\ &\leq \prod_{q,l} \exp \left\{ \frac{n_{ql} v^2 t^2 (\beta_{ql}^{(1)} - \beta_{ql}^{(2)})^2}{2} \right\} \\ &\leq \exp \left\{ \frac{n^2 v^2 t^2 (\beta_{ql}^{(1)} - \beta_{ql}^{(2)})^2}{2} \right\} \\ &= \exp \left\{ \frac{t^2 \|\beta^{(1)} - \beta^{(2)}\|_2^2}{2} \right\} \\ &\leq \exp \left\{ \frac{t^2 \|\beta^{(1)} - \beta^{(2)}\|_2^2}{2(1 - tc \|\beta^{(1)} - \beta^{(2)}\|_\infty)} \right\} \end{aligned}$$

for all

$$t \in \left[0, \frac{1}{c \|\beta^{(1)} - \beta^{(2)}\|} \right),$$

where c is some negative constant. Therefore, the first condition of Proposition 5 is satisfied.

Fix some $\beta^{(0)}$, (A6) implies that there exist $u_1 = \mathcal{O}(n)$ and $u_2 = \mathcal{O}(1)$ such that $\|\beta - \beta^{(0)}\|_2 \leq u_1$, and $\|\beta - \beta^{(0)}\|_2 \leq u_2$. Therefore the second condition of Proposition 5 is also satisfied with $D = Q^2$. Now, we have

$$\begin{aligned} P^*(\rho_n \sup_{\beta} |S_n(\beta)| > \eta) &\leq P^*(\rho_n \sup_{\beta} |S_n(\beta) - S_n(\beta^{(0)})| > \frac{\eta}{2}) \\ &\quad + P^*(\rho_n |S_n(\beta^{(0)})| > \frac{\eta}{2}). \end{aligned} \tag{18}$$

Since $\sum_{i \neq j} X_{ij} Y_{ij} \beta_{z_i, z_j}^{(0)}$ is a sub-exponential random variable with parameters (nv, b) , the second term on the RHS of inequality 18 can be bounded by

$$P^*(\rho_n |S_n(\beta^{(0)})| > \frac{\eta}{2}) = \mathcal{O}(\exp(-n^2)). \tag{19}$$

To bound the first term on the RHS of inequality 18, we introduce the set $\mathcal{P}^{z_{[n]}} = \{(\pi, \alpha, \beta) : (z_{[n]}, \pi, \alpha, \beta) \in \mathcal{P}\}$ for every $z_{[n]}$, and define the event

$$\Omega_n(z_{[n]}) = \left\{ \sup_{\mathcal{P}(z_{[n]})} \rho_n |S_n(\beta) - S_n(\beta^{(0)})| \leq \rho_n \kappa \sqrt{u^2(Q^2 + x_n)} + \rho_n b(Q^2 + x_n) \right\}.$$

We have

$$P^*(\Omega_n(z_{[n]})^c) = 2e^{-x_n} \tag{20}$$

by Proposition 5. Combining 18, 19, and 20,

$$\begin{aligned} & P^* \left[\rho_n \left| \sum_{i \neq j} \left(X_{ij} Y_{ij} - \pi_{z_i^*, z_j^*}^* \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \right) \beta_{z_i, z_j} \right| > \eta \right] \\ & \leq \sum_{z_{[n]}} P^* \left[\left\{ \sup_{\mathcal{P}(z_{[n]})} \rho_n |S_n(\beta) - S_n(\beta^{(0)})| > \frac{\eta}{2} \right\} \right] + P^* \left[\rho_n |S_n(\beta_0)| > \frac{\eta}{2} \right] \\ & \leq \sum_{z_{[n]}} P^* \left[\left\{ \sup_{\mathcal{P}(z_{[n]})} \rho_n |S_n(\beta) - S_n(\beta^{(0)})| > \frac{\eta}{2} \right\} \cap \Omega_n(z_{[n]}) \right] + \sum_{z_{[n]}} 2e^{-x_n} + \sum_{z_{[n]}} \mathcal{O}(e^{-n^2}) \\ & \leq \sum_{z_{[n]}} P^* \left[\rho_n \kappa \sqrt{u^2(D + x_n)} + \rho_n b(D + x_n) > \frac{\eta}{2} \right] + \sum_{z_{[n]}} 2e^{-x_n} + \sum_{z_{[n]}} \mathcal{O}(e^{-n^2}) \end{aligned}$$

Since $z_{[n]}$ belongs to a set of cardinality at most Q^n , by choosing $x_n = n \log(n)$, the three sums converge to 0.

By using the fact that $X_{ij} \log Y_{ij}$ is a sub-exponential random variable, we can similarly show that

$$P^* \left[\sup_{\mathcal{P}} \rho_n \left| \sum_{i \neq j} \left(X_{ij} \log Y_{ij} - \pi_{z_i^*, z_j^*}^* \log \beta_{z_i^*, z_j^*}^* - \pi_{z_i^*, z_j^*}^* \psi(\alpha_{z_i^*, z_j^*}^*) \right) (\alpha_{z_i, z_j} - 1) \right| > \eta \right] \xrightarrow{I} n \rightarrow \infty] \mathbb{P}0.$$

Since the convergence is uniform with respect to $z_{[n]}^*$, P^* can be replaced by \mathbb{P} and the proof is completed. □

Proposition 6 allows us to establish the following result concerning the convergence of the normalized log-likelihood function. Proposition 7 is an extension of Theorem 3.6 of Celisse et al. (2012) and allows us to establish the consistency of MLE of (π, α, β) .

Proposition 7 *We assume that assumptions (A1), (A2), (A3), (A5), (A6) hold. For every $(\theta, \pi, \alpha, \beta)$, set*

$$M_n(\theta, \pi, \alpha, \beta) = (n(n - 1))^{-1} \mathcal{L}_2(Y_{[n]}, X_{[n]}; \theta, \pi, \alpha, \beta),$$

and

$$\begin{aligned} \mathbb{M}(\pi, \alpha, \beta, \mathcal{A}) &= \sum_{q,l} \theta_q^* \theta_l^* \sum_{q',l'} a_{q,q'} a_{l,l'} \left[\pi_{q,l}^* \log \pi_{q',l'} + (1 - \pi_{q,l}^*) \log(1 - \pi_{q',l'}) \right. \\ &\quad \left. + \pi_{q,l}^* E_{\alpha_{q',l'}, \beta_{q',l'}} \left(\log f(\cdot; \alpha_{q',l'}, \beta_{q',l'}) \right) \right] \end{aligned}$$

where

$$\mathcal{A} = \{A = (a_{q,l})_{1 \leq q,l \leq Q} | a_{q,l} \geq 0, \sum_{l=1}^Q a_{kl} = 1\}.$$

Then for any $\eta > 0$,

$$\begin{aligned} \sup_{d((\pi, \alpha, \beta), (\pi^*, \alpha^*, \beta^*)) \geq \eta} \mathbb{M}(\pi, \alpha, \beta) &< \mathbb{M}(\pi^*, \alpha^*, \beta^*), \\ \sup_{\theta, \pi, \alpha, \beta} |M_n(\theta, \pi, \alpha, \beta) - \mathbb{M}(\pi, \alpha, \beta)| &\xrightarrow{[n \rightarrow \infty]} \mathbb{P}0. \end{aligned}$$

Proof We define the following:

$$\begin{aligned} \hat{z}_{[n]}(\pi, \alpha, \beta) &= \operatorname{argmax}_z \phi_n(z_{[n]}, \pi, \alpha, \beta), \\ \tilde{z}_{[n]}(\pi, \alpha, \beta) &= \operatorname{argmax}_z \Phi_n(z_{[n]}, \pi, \alpha, \beta), \\ \bar{A}_{\pi, \alpha, \beta} &= \operatorname{argmax}_{A \in \mathcal{A}} \mathbb{M}(\pi, \alpha, \beta, A), \\ \mathbb{M}(\pi, \alpha, \beta) &= \mathbb{M}(\pi, \alpha, \beta, \bar{A}_{\pi, \alpha, \beta}). \end{aligned}$$

By a similar reasoning as in the proof of Theorem 3.6 of Celisse et al. (2012), we can show that $\bar{A}_{\pi^*, \alpha^*, \beta^*} = I_Q$ and is unique. To show that for all $\eta > 0$, $\sup_{d((\pi, \alpha, \beta), (\pi^*, \alpha^*, \beta^*)) \geq \eta} \mathbb{M}(\pi, \alpha, \beta) < \mathbb{M}(\pi^*, \alpha^*, \beta^*)$, we let $(\bar{a}_{q,l})_{q,l}$ denote coefficients of $\bar{A}_{\pi, \alpha, \beta}$. We have that

$$\begin{aligned} \mathbb{M}(\pi, \alpha, \beta) - \mathbb{M}(\pi^*, \alpha^*, \beta^*) &= - \sum_{q,l} \theta_q^* \theta_l^* \sum_{q',l'} \bar{a}_{q,q'} \bar{a}_{l,l'} \{ \mathbb{K}\mathbb{L}_B(\pi_{q,l}^*, \pi_{q',l'}) + \pi_{q,l}^* \mathbb{K}\mathbb{L}_G((\alpha_{q,l}^*, \beta_{q,l}^*), (\alpha_{q',l'} \beta_{q',l'})) \}, \end{aligned}$$

where $\mathbb{K}\mathbb{L}_B(p, q)$ denotes the Kullback-Leibler divergence between two bernoulli distributions $\operatorname{Ber}(p)$ and $\operatorname{Ber}(q)$, and $\mathbb{K}\mathbb{L}_G((a, b_1), (a, b_2))$ denotes the Kullback-Leibler divergence between two gamma distributions $\operatorname{Ga}(a, b_1)$ and $\operatorname{Ga}(a, b_2)$. Since the set $\{(\pi, \alpha, \beta) | d((\pi, \alpha, \beta), (\pi^*, \alpha^*, \beta^*)) \geq \eta\}$ is compact by assumptions, there exists

$(\pi^0, \alpha^0, \beta^0) \neq (\pi^*, \alpha^*, \beta^*)$ such that

$$\begin{aligned} & \sup_{d((\pi, \alpha, \beta), (\pi^*, \alpha^*, \beta^*)) \geq \eta} \mathbb{M}(\pi, \alpha, \beta) - \mathbb{M}(\pi^*, \alpha^*, \beta^*) \\ &= \mathbb{M}(\pi^0, \alpha^0, \beta^0) - \mathbb{M}(\pi^*, \alpha^*, \beta^*) < 0. \end{aligned}$$

Next we show that

$$\sup_{\theta, \pi, \alpha, \beta} |M_n(\theta, \pi, \alpha, \beta) - \mathbb{M}(\pi, \alpha, \beta)| \xrightarrow{1} n \rightarrow \infty \mathbb{P}0.$$

We first have the following bound:

$$\begin{aligned} |M_n(\theta, \pi, \alpha, \beta) - \mathbb{M}(\pi, \alpha, \beta)| &\leq |M_n(\theta, \pi, \alpha, \beta) - \phi_n(\hat{z}, \pi, \alpha, \beta)| \\ &\quad + |\phi_n(\hat{z}, \pi, \alpha, \beta) - \Phi_n(\tilde{z}, \pi, \alpha, \beta)| \\ &\quad + |\Phi_n(\tilde{z}, \pi, \alpha, \beta) - \mathbb{M}(\pi, \alpha, \beta)|. \end{aligned} \tag{21}$$

We first consider the first term on the RHS of inequality (21):

$$\begin{aligned} & \sup_{\theta, \pi, \alpha, \beta} |M_n(\theta, \pi, \alpha, \beta) - \phi_n(\hat{z}, \pi, \alpha, \beta)| \\ &= \sup_{\theta, \pi, \alpha, \beta} \frac{|\mathcal{L}_2(Y_{[n]}, X_{[n]}; \theta, \pi, \alpha, \beta) - \mathcal{L}_1(Y_{[n]}, X_{[n]}; \hat{z}_{[n]}, \pi, \alpha, \beta)|}{n(n-1)} \\ &\leq \frac{\log(1/\gamma)}{n-1} \xrightarrow{1} n \rightarrow \infty 0. \end{aligned} \tag{22}$$

Consider the second term on the RHS of inequality (21), if $\phi_n(\hat{z}, \pi, \alpha, \beta) < \Phi_n(\tilde{z}, \pi, \alpha, \beta)$, we have

$$\begin{aligned} |\phi_n(\hat{z}, \pi, \alpha, \beta) - \Phi_n(\tilde{z}, \pi, \alpha, \beta)| &= \Phi_n(\tilde{z}, \pi, \alpha, \beta) - \phi_n(\hat{z}, \pi, \alpha, \beta) \\ &\leq \Phi_n(\tilde{z}, \pi, \alpha, \beta) - \phi_n(\tilde{z}, \pi, \alpha, \beta). \end{aligned}$$

On the other hand, if $\phi_n(\hat{z}, \pi, \alpha, \beta) \geq \Phi_n(\tilde{z}, \pi, \alpha, \beta)$,

$$\begin{aligned} |\phi_n(\hat{z}, \pi, \alpha, \beta) - \Phi_n(\tilde{z}, \pi, \alpha, \beta)| &= \phi_n(\tilde{z}, \pi, \alpha, \beta) - \Phi_n(\hat{z}, \pi, \alpha, \beta) \\ &\leq \phi_n(\hat{z}, \pi, \alpha, \beta) - \Phi_n(\hat{z}, \pi, \alpha, \beta). \end{aligned}$$

Therefore, Proposition 6 implies that

$$\sup_{\pi, \alpha, \beta} |\phi_n(\hat{z}, \pi, \alpha, \beta) - \Phi_n(\tilde{z}, \pi, \alpha, \beta)| \xrightarrow{1} n \rightarrow \infty \mathbb{P}0. \tag{23}$$

Last, we notice that under our assumptions (A2) and (A6), and using the strong law of large number,

$$\sup_{\pi, \alpha, \beta} |\Phi_n(\tilde{z}, \pi, \alpha, \beta) - \mathbb{M}(\pi, \alpha, \beta)| \xrightarrow{1} n \rightarrow \infty \mathbb{P}0. \tag{24}$$

Combining (22), (23) and (24), we have the desired result. □

The consistency of $(\hat{\pi}, \hat{\alpha}, \hat{\beta})$ follows from Proposition 7 and Theorem 3.4 of Celisse et al. (2012). \square

Proof of Theorem 3

Proof We denote $\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]}, Y_{[n]})$ as the conditional distribution of $Z_{[n]}$ under the parameters $(\hat{\theta}, \hat{\pi}, \hat{\alpha}, \hat{\beta})$. The following result is an extension of Proposition 3.8 of Celisse et al. (2012) and is needed to establish the consistency of $\hat{\theta}$.

Proposition 8 Assume that assumptions (A1)-(A6) hold, and there exists estimators $\hat{\pi}, \hat{\alpha}, \hat{\beta}$ such that $\|\hat{\pi} - \pi\|_{\infty} = o_P(v_n)$, $\|\hat{\alpha} - \alpha\|_{\infty} = o_P(v_n)$, $\|\hat{\beta} - \beta\|_{\infty} = o_P(v_n)$, with $v_n = o(\sqrt{\log n}/n)$. Let also $\hat{\theta}$ denote any estimator of θ^* . Then for every $\epsilon > 0$,

$$P^* \left[\sum_{z_{[n]} \neq z_{[n]}^*} \frac{\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]}, Y_{[n]})}{\hat{P}(Z_{[n]} = z_{[n]}^* | X_{[n]}, Y_{[n]})} > \epsilon \right] \leq \kappa_1 n e^{-\kappa_2 \frac{(\log n)^2}{m v_n^2}} + \mathbb{P}[\|\hat{\pi} - \pi^*\|_{\infty} > v_n] + \mathbb{P}[\|\hat{\alpha} - \alpha^*\|_{\infty} > v_n] + \mathbb{P}[\|\hat{\beta} - \beta^*\|_{\infty} > v_n]$$

for n large enough, and for some constants $\kappa_1, \kappa_2 > 0$ and

$$\begin{aligned} & \log \left(\frac{\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]}, Y_{[n]})}{\hat{P}(Z_{[n]} = z_{[n]}^* | X_{[n]}, Y_{[n]})} \right) \\ &= \log \left(\frac{\hat{P}(X_{[n]} | Z_{[n]} = z_{[n]}) P(Z_{[n]} = z_{[n]})}{\hat{P}(X_{[n]} | Z_{[n]} = z_{[n]}^*) P(Z_{[n]} = z_{[n]}^*)} \right) + \log \left(\frac{\hat{P}(Y_{[n]} | X_{[n]}, Z_{[n]} = z_{[n]})}{\hat{P}(Y_{[n]} | X_{[n]}, Z_{[n]} = z_{[n]}^*)} \right) \\ &= \sum_{i \neq j} \left\{ X_{ij} \log \left(\frac{\hat{\pi}_{z_i, z_j}}{\hat{\pi}_{z_i^*, z_j^*}} \right) + (1 - X_{ij}) \log \left(\frac{1 - \hat{\pi}_{z_i, z_j}}{1 - \hat{\pi}_{z_i^*, z_j^*}} \right) \right\} + \sum_i \log \frac{\hat{\theta}_{z_i}}{\hat{\theta}_{z_i^*}} \\ & \quad + \sum_{i \neq j} \left\{ X_{ij} (\hat{\alpha}_{z_i, z_j} \log \hat{\beta}_{z_i, z_j} - \hat{\alpha}_{z_i^*, z_j^*} \log \hat{\beta}_{z_i^*, z_j^*}) + (\hat{\alpha}_{z_i, z_j} - \hat{\alpha}_{z_i^*, z_j^*}) X_{ij} \log Y_{ij} \right. \\ & \quad \left. + X_{ij} Y_{ij} (\hat{\beta}_{z_i^*, z_j^*} - \hat{\beta}_{z_i, z_j}) + X_{ij} (\log \Gamma(\hat{\alpha}_{z_i^*, z_j^*}) - \log \Gamma(\hat{\alpha}_{z_i, z_j})) \right\} \\ & := T_0 + T_1 \end{aligned}$$

Proof We can write

$$\begin{aligned}
 T_1 &= \sum_{D^*} \left\{ \pi_{z_i^*, z_j^*}^* \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} (\beta_{z_i^*, z_j^*}^* - \beta_{z_i, z_j}^*) + \pi_{z_i^*, z_j^*}^* \alpha_{z_i, z_j}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} \right\} \\
 &+ \sum_{D^*} \left\{ \left(X_{ij} Y_{ij} - \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* \right) (\beta_{z_i^*, z_j^*}^* - \hat{\beta}_{z_i, z_j}) + (X_{ij} - \pi_{z_i^*, z_j^*}^*) \alpha_{z_i, z_j}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} \right\} \\
 &+ \sum_{\bar{D} \cup D^*} \left\{ \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* (\hat{\beta}_{z_i^*, z_j^*} - \beta_{z_i^*, z_j^*}^*) + \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* (\hat{\beta}_{z_i, z_j} - \beta_{z_i, z_j}^*) \right. \\
 &\left. + \pi_{z_i^*, z_j^*}^* \left(\hat{\alpha}_{z_i, z_j} \log \frac{\hat{\beta}_{z_i, z_j}}{\hat{\beta}_{z_i^*, z_j^*}} - \alpha_{z_i, z_j}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} \right) \right\} \\
 &+ \sum_{\bar{D} \cup D^*} \left\{ (X_{ij} - \pi_{z_i^*, z_j^*}^*) \left(\hat{\alpha}_{z_i, z_j} \log \frac{\hat{\beta}_{z_i, z_j}}{\hat{\beta}_{z_i^*, z_j^*}} - \alpha_{z_i, z_j}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} \right) \right\} \\
 &+ \sum_{\bar{D} \cup D^*} \left\{ \left(X_{ij} Y_{ij} - \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* \right) (\hat{\beta}_{z_i^*, z_j^*} - \beta_{z_i^*, z_j^*}^*) \right. \\
 &\left. + \left(X_{ij} Y_{ij} - \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} \pi_{z_i^*, z_j^*}^* \right) (\beta_{z_i, z_j}^* - \hat{\beta}_{z_i, z_j}) \right\} \\
 &=: T_{1,1} + T_{1,2} + T_{1,3} + T_{1,4} + T_{1,5}
 \end{aligned}$$

where

$$\begin{aligned}
 D^* &:= \{(i, j) : i \neq j, \pi_{z_i, z_j}^* \neq \pi_{z_i^*, z_j^*}^*\}, \\
 \hat{D} &:= \{(i, j) : i \neq j, \hat{\pi}_{z_i, z_j} \neq \hat{\pi}_{z_i^*, z_j^*}\}.
 \end{aligned}$$

By the proof of Proposition 3.8 of Celisse et al. (2012), we have

$$\begin{aligned}
 P^* &\left[\left\{ \sum_{[z_{[n]}] \neq [z_{[n]}^*]} \frac{\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]}, Y_{[n]})}{\hat{P}(Z_{[n]} = z_{[n]}^* | X_{[n]}, Y_{[n]})} > \epsilon \right\} \cap \Omega_n \right] \\
 &\leq \sum_{r=1}^n \sum_{z_{[n]} \notin [z_{[n]}^*], \|z_{[n]} - z_{[n]}^*\|_0 = r} P^* \left[\left\{ \log \frac{\hat{P}(Z_{[n]} = z_{[n]} | X_{[n]}, Y_{[n]})}{\hat{P}(Z_{[n]} = z_{[n]}^* | X_{[n]}, Y_{[n]})} > -5r \log n \right\} \cap \Omega_n \right] \\
 &= \sum_{r=1}^n \sum_{z_{[n]} \notin [z_{[n]}^*], \|z_{[n]} - z_{[n]}^*\|_0 = r} P^* [\{T_0 + T_1 > -5r \log n\} \cap \Omega_n].
 \end{aligned}$$

We have that

$$P^*[\{T_0 + T_1 > -5r \log n\} \cap \Omega_n] \leq P^*[\{T_0 > -10r \log n\} \cap \Omega_n] + P^*[\{T_1 > 5r \log n\} \cap \Omega_n].$$

The proof of Proposition 3.8 of Celisse et al. (2012) shows that

$$P^*[\{T_0 > -10r \log n\} \cap \Omega_n] \leq C_1 \left\{ \exp \left(8n \log n - C_2 \frac{(\log n)^2}{nv_n^2} \right) \right\}^r. \tag{25}$$

We have that

$$\begin{aligned} T_{1,1} &= \sum_{D^*} \pi_{z_i^*, z_j^*}^* \frac{\alpha_{z_i^*, z_j^*}^*}{\beta_{z_i^*, z_j^*}^*} (\beta_{z_i^*, z_j^*}^* - \beta_{z_i, z_j}^*) + \pi_{z_i^*, z_j^*}^* \alpha_{11}^* \log \frac{\beta_{z_i, z_j}^*}{\beta_{z_i^*, z_j^*}^*} \\ &\leq |D^*| \max_{(q,l) \neq (q',l')} -\mathbb{K} \mathbb{L}_G((\alpha_{ql}^*, \beta_{ql}^*) || (\alpha_{q'l'}^*, \beta_{q'l'}^*)) \\ &= -|D^*| K^* \end{aligned}$$

Since $X_{ij}Y_{ij}$ is a sub-exponential random variable and X_{ij} is bounded for all i, j , one can show that

$$P^*[T_{1,2} + \frac{1}{2}T_{1,1} > t] \leq \max \left\{ \exp \left\{ -C_1 \frac{(t + |D^*|K^*)^2}{|D^*|} \right\}, \exp \left\{ -C_2(t + |D^*|K^*) \right\} \right\}. \tag{26}$$

Using similar techniques as in the proof of Proposition 3.8 of Celisse et al. (2012), one can show that

$$P^*[\{|T_{1,3}| > t\} \cap \Omega_n] \leq P^* \left(v_n > \frac{C_3 t}{nr} \right), \tag{27}$$

$$P^*[\{T_{1,4} > t\} \cap \Omega_n] \leq \sum_k \sum_{D, |D|=k} Q^2 \exp \left(-\frac{C_4 t^2}{v_n^2(k + |D^*|)} \right), \tag{28}$$

$$\begin{aligned} &P^*[T_{1,5} + \frac{1}{2}T_{1,1} > t] \\ &\leq \sum_k \sum_{D, |D|=k} Q^2 \max \left\{ \exp \left[-\frac{C_5(t + (|D^*| + k)K^*)}{v_n} \right], \right. \\ &\quad \left. \exp \left[-\frac{C_6(t + (|D^*| + k)K^*)^2}{(|D^*| + k)v_n^2} \right] \right\} \end{aligned} \tag{29}$$

Combining inequalities 25, 26, 27, 28, 29, and letting $v_n = o(\sqrt{\log n}/n)$, we have for some $B_1, B_2, B_3 > 0$,

$$\begin{aligned}
P^* & \left[\left\{ \sum_{[z_{[n]}] \neq [z_{[n]}^*]} \frac{\hat{P}(z_{[n]} | X_{[n]}, Y_{[n]})}{\hat{P}(z_{[n]}^* | X_{[n]}, Y_{[n]})} > \epsilon \right\} \cap \Omega_n \right] \\
& \leq \sum_{r=1}^n \binom{n}{r} (Q-1)^r B_1 \left(\exp \left[B_2 n \log n - B_3 \omega(n \log n) \right] \right)^r \\
& = B_1 \left[(1 + (Q-1)u'_n)^n - 1 \right]
\end{aligned}$$

where $u'_n = \exp \left[B_2 n \log n - B_3 \frac{(\log n)^2}{n} \right]$. Since $(1 + (Q-1)u'_n)^n \rightarrow 1$ as $n \rightarrow \infty$, and the proof is completed. \square

The proof of the consistency of $\hat{\theta}$ is a consequence of the proposition above and follows the same lines as the proof of Theorem 3.9 of Celisse et al. (2012). \square

Funding Open Access funding provided by the IReL Consortium. The funding was provided by Science Foundation Ireland (Grant No. SFI/12/RC/2289-2).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbe E (2018) Community detection and stochastic block models: recent developments. *J Mach Learn Res* 18:1–86
- Airoldi EM, Blei DM, Fienberg SE, Xing EP (2008) Mixed membership stochastic blockmodels. *J Mach Learn Res* 9:1981–2014
- Aicher C, Jacobs AZ, Clauset A (2013) Adapting the stochastic block model to edge-weighted networks. *ICML workshop on structured learning*
- Aicher C, Jacobs AZ, Clauset A (2015) Learning latent block structure in weighted networks. *J Compl Netw* 3:221–248
- Allman ES, Matias C, Rhodes JA (2009) Identifiability of parameters in latent structure models with many observed variables. *Ann Stat* 37:3099–3132
- Allman ES, Matias C, Rhodes JA (2011) Parameter identifiability in a class of random graph mixture models. *J Stat Plan Inference* 141:1719–1736
- Ambroise C, Matias C (2012) New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J R Stat Soc Ser B Stat Methodol* 74:3–35
- Baraud Y (2010) A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression. *Bernoulli* 16:1064–1085
- Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci* 101:3747–3752
- Bickel PJ, Chen A (2009) A nonparametric view of network models and Newman–Girvan and other modularities. *Proc Natl Acad Sci USA* 106:21068–21073
- Bickel PJ, Chen A, Levina E (2011) The method of moments and degree distributions for network models. *Ann Stat* 39:2280–2301
- Bickel P, Choi D, Chang X, Zhang H (2013) Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann Stat* 41:1922–1943

- Biernacki C, Celeux G, Govaert G (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans Pattern Anal Mach Intell* 22:719–725
- Brault V, Keribin C, Mariadassou M (2020) Consistency and asymptotic normality of Latent Block Model estimators. *Electron J Stat* 14:1234–1268
- Celisse A, Daudin J-J, Pierre L (2012) Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron J Stat* 6:1847–1899
- Channaronand A, Daudin J-J, Robin S (2012) Classification and estimation in the stochastic blockmodel based on the empirical degrees. *Electron J Stat* 6:2574–2601
- Choi DS, Wolfe PJ, Airolidi EM (2012) Stochastic blockmodels with a growing number of classes. *Biometrika* 99:273–284
- Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. *Phys Rev E* 70:066111
- Côme E, Latouche P (2015) Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. *Stat Model* 15:564–589
- Daudin J-J, Picard F, Robin S (2008) A mixture model for random graphs. *Stat Comput* 18:173–183
- Ghasebian A, Zhang P, Clauset A, Moore C, Peel L (2016) Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Phys Rev X* 6:031005
- Haj AE, Slaoui Y, Louis P-Y, Khraibani Z (2020) Estimation in a binomial stochastic blockmodel for a weighted graph by a variational expectation maximization algorithm. *Commun Stat Simul Comput*:1–20
- Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. *Soc Netw* 5:109–137
- Jog V, Loh P-L (2015) Information-theoretic bounds for exact recovery in weighted stochastic block models using the Rényi divergence. *CoRR*, arXiv:abs/1509.06418
- Karrer B, Newman MEJ (2011) Stochastic blockmodels and community structure in networks. *Phys Rev E* 83:016107
- Latouche P, Birmelé E, Ambroise C (2011) Overlapping stochastic block models with application to the French political blogosphere. *Ann Appl Stat* 5:309–336
- Latouche P, Birmelé E, Ambroise C (2014) Model selection in overlapping stochastic block models. *Electron J Stat* 8:762–794
- Leger J-B, Vacher C, Daudin J-J (2014) Detection of structurally homogeneous subsets in graphs. *Stat Comput* 24:675–692
- Ludkin M (2020) Inference for a generalised stochastic block model with unknown number of blocks and non-conjugate edge models. *Comput Stat Data Anal* 152:107051
- Mariadassou M, Robin S, Vacher C (2010) Uncovering latent structure in valued graphs: a variational approach. *Ann Appl Stat* 4:715–742
- Newman MEJ (2004) Analysis of weighted networks. *Phys Rev E* 70:056131
- Peixoto TP (2018) Nonparametric weighted stochastic block models. *Phys Rev E* 97:012306
- Peng L, Carvalho L (2016) Bayesian degree-corrected stochastic blockmodels for community detection. *Electron J Stat* 10:2746–2779
- Pons P, Latapy M (2005) Computing communities in large networks using random walks. In: *Proceedings of the 20th international conference on computer and information sciences*. Springer, Berlin, ISCIS'05, pp 284–293
- Rohe K, Chatterjee S, Yu B (2011) Spectral clustering and the high-dimensional stochastic blockmodel. *Ann Stat* 39:1878–1915
- Saldaña DF, Yu Y, Feng Y (2017) How many communities are there? *J Comput Graph Stat* 26:171–181
- Snijders TAB, Nowicki K (1997) Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J Classif* 14:75–100
- Stouffer DB, Bascombe J (2011) Compartmentalization increases food-web persistence. *Proc Natl Acad Sci USA* 108:3648–3652
- von Luxburg U (2007) A tutorial on spectral clustering. *Stat Comput* 17:395–416
- Wang YXR, Bickel PJ (2017) Likelihood-based model selection for stochastic block models. *Ann Stat* 45:500–528
- Yan X (2016) Bayesian model selection of stochastic block models. In: *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pp 323–328
- Ye Z-S, Chen N (2017) Closed-form estimators for the gamma distribution derived from likelihood equations. *Am Stat* 71:177–181

Zanghi H, Picard F, Miele V, Ambroise C (2010) Strategies for online inference of model-based clustering in large and growing networks. *Ann Appl Stat* 4:687–714

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.