Chapter 8

Learning SBM parameters

8.1 Statistical network inference

Network data usually consists of relational information between a set of nodes that is represented by an n-by-n matrix (X_{ij}) with binary or numerical entries, and node attribute data represented by an n-vector (Z_i) with numerical or categorical entries, see Figure 8.1. Network inference problems concern *computing estimates, making predictions, and testing hypotheses* of network structure and node attributes based on partial or noisy observations of the network data matrix (X_{ij}), node attribute vectors (Z_i), and possibly some auxiliary data related to temporal dynamics (diffusions, random walks) on the network.

→ intro?

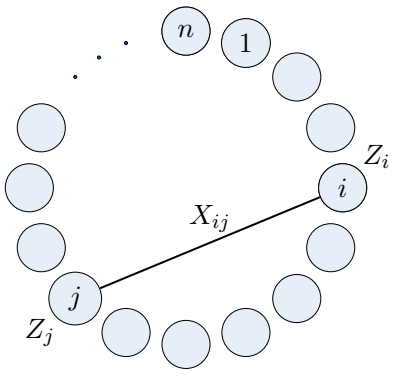

Figure 8.1: Node attributes and relationships.

This framework contains a rich class of applications, for example:

Example 8.1 (Community learning)**.** Estimate node attributes (Z_i) based on fully observed network structure (X_{ij}), up to a permutation of node labels.

This amounts to estimating a partition of the node set generated by the sets $V_s = \{i : Z_i = s\}$ called communities.

Example 8.2 (Phylogenetics)**.** Denote by Z_i a genetic trait of an individual or a group of organisms i. If the values of Z_i have been observed for a set of leaf nodes in an evolutionary tree with fully or partially observed structure (X_{ij}), the task is to infer the value Z_{i_0} of the initial ancestor corresponding to the root node i_0 of the evolutionary tree.

Example 8.3 (Epidemics)**.** Let Z_i be a binary variable indicating whether an individual i falls victim to an infectious disease, and let X_{ij} be a binary variable indicating whether the disease is transmitted through a direct contact between individuals i and j. An important statistical task is to estimate the size of the set $\{i : Z_i = 1\}$ of eventually infected individuals, based on observing values of Z_i for a typically small subset of nodes, and partial observations of the network structure (X_{ij}).

Network data is often given in bipartite form so that we observed relational information (X_{ij}) in the form of an m-by-n matrix between m nodes of a particular type having attributes (Z_i^L), and n nodes of a different type having attributes (Z_j^R), see Figure 8.2. Practical learning tasks involving bipartite data are common in crowdsourcing and collaborative filtering contexts, see the examples below.

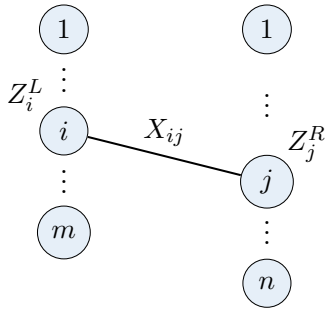

Figure 8.2: Bipartite network data.

Example 8.4 (Crowdsourcing)**.** In microtasking platforms such as Amazon Mechanical Turk, a set of m simple tasks are allocated to n workers who might provide unreliable answers. The unreliability is mitigated by allocating the same task to several workers. Denote by X_{ij} the outcome of task i performed by worker j, by Z_i^L the true outcome of task i, and by Z_j^R the inherent reliability of worker j. The inference problem is to estimate the true outcomes (Z_i^L) based on observed data (X_{ij}).

Example 8.5 (Collaborative filtering)**.** In online recommendation systems a common objective is to infer customer's preferences based on their own and other customers' rankings on a set of items. Let X_{ij} be a number indicating the level of preference of item i by user j. Having observed a partial set of entries of (X_{ij}), the challenge is to complete the matrix by estimating the unobserved remaining values. A famous example of this problem is the Netflix challenge[1]. This problem setup does not involve item attributes (Z_i^L) or customer attributes (Z_j^R), but they could be incorporated as auxiliary model variables.

For notational simplicity, these lecture notes restrict to the unipartite network setting corresponding to Figure 8.1. We will model the joint distribution of the network structure (X_{ij}) and the node attributes (Z_i) using a statistical model where the entries X_{ij} are mutually independent conditionally on the node attributes. This model is described in detail in next section.

8.2 Learning stochastic block models

Denote by $\mathcal{G}_n$ the set of undirected graphs on node set $[n] = \{1, \dots, n\}$, or equivalently, the set of all binary arrays (x_{ij}) indexed by $1 \le i < j \le n$. A stochastic block model with n nodes and m blocks generates inhomogeneous Bernoulli random graphs with link probabilities

$$p_{ij} \;=\; K_{z_i z_j},$$

and the model is parameterized by a symmetric m-by-m matrix K with nonnegative entries, called the block interaction matrix, and a vector $z = (z_1, \dots, z_n)$ with entries in $[m] = \{1, \dots, m\}$, called the block membership vector. The law of the random graph is a probability distribution on $\mathcal{G}_n$ with probability mass function

$$f(x \,|\, z) \;=\; \prod_{1 \le i < j \le n} (1 - K_{z_i z_j})^{1 - x_{ij}} K_{z_i z_j}^{x_{ij}}. \tag{8.1}$$

In usual statistical learning tasks, the pairwise node-to-node interactions $x = (x_{ij})$ are observed and the block memberships $z = (z_i)$ are unknown. In a Bayesian inference approach, a standard approach is to assume that the block memberships z_i are independent random samples from a prior distribution α on $[m]$. This leads to a doubly stochastic block model where the joint law of

[1] https://en.wikipedia.org/wiki/Netflix_Prize

the block memberships ("parameter") and the pair interactions ("data") is a probability distribution on $\mathcal{G}_n \times [m]^n$ given by

$$f(x, z) \;=\; \sum_{z \in [m]^n} f(x \,|\, z) \prod_{i=1}^{n} \alpha(z_i). \tag{8.2}$$

Formula (8.2) represents the joint distribution of a random graph (X_{ij}) and a random vector (Z_i) such that the entries of (Z_i) are mutually independent and α-distributed, and conditionally on (Z_i), the entries X_{ij} are independent and Bernoulli distributed with success probability $K(Z_i, Z_j)$. When the number of blocks m is assumed fixed and known, the stochastic block model (8.1) is parameterized by $\theta = (z, K)$, and the doubly stochastic block model (8.2) by $\theta = (\alpha, K)$.

8.3 Learning the block interaction matrix when block memberships are known

The easiest learning problem is to estimate the block interaction matrix K from observed graph sample $x = (x_{ij})$ when the block memberships $z = (z_i)$ are known, or they have been first estimated using some other method. A *maximum likelihood estimate* of K is a symmetric nonnegative m-by-m matrix $\hat{K}$ which maximizes the likelihood $K \mapsto f_K(x \,|\, z)$ corresponding to formula (8.1).

Let us first introduce some helpful notation related to the block structure of the observed graph sample. First, let us represent the attribute vector (z_i) as an n-by-m binary matrix (z_{ij}) with entries $z_{is} = 1(z_i = s)$ indicating whether node i belongs to block s. Then the size of block s can be written as

$$n_s \;=\; \sum_{i=1}^{n} z_{is},$$

and the number of links between blocks s and t as

$$e_{st} \;=\; \begin{cases} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} z_{is} z_{jt}, & s \neq t, \\ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} z_{is} z_{js}, & s = t. \end{cases}$$

As a consequence, the link density between blocks s and t in the observed graph can be written as

$$d_{st} \;=\; \frac{e_{st}}{n_{st}}, \tag{8.3}$$

where

$$n_{st} \;=\; \begin{cases} n_s n_t, & s \neq t, \\ \frac{1}{2} n_s (n_s - 1), & s = t. \end{cases}$$

Theorem 8.6. *The unique maximum likelihood estimate of K is the m-by-m matrix with entries being the observed block densities $\hat{K}_{st} = d_{st}$ defined by (8.3).*

Proof. Recall that maximizing a function is equivalent to maximizing its logarithm. We take logarithm of the likelihood to transform the product in (8.1) into a sum. The log-likelihood can be written as

$$\log f_K(x \,|\, z) \;=\; \sum_{1 \le i < j \le n} \Big\{ (1 - x_{ij}) \log(1 - K_{z_i, z_j}) + x_{ij} \log K_{z_i, z_j} \Big\}.$$

In the above sum there is a lot of redundancy in the sense that the only possible values of the terms are $\log(1 - K_{s,t})$ and $\log K_{s,t}$ for some $1 \le s \le t \le m$. By counting how many times these values occur in the sum, we see that the log-likelihood can be written as

$$\begin{aligned} \log f_K(x \,|\, z) \;&=\; \sum_{1 \le s \le t \le m} \Big\{ (N_{st} - e_{st}) \log(1 - K_{st}) + e_{st} \log(K_{st}) \Big\} \\ &=\; \sum_{1 \le s \le t \le m} N_{st} \Big\{ (1 - d_{st}) \log(1 - K_{st}) + d_{st} \log(K_{st}) \Big\}. \end{aligned}$$

After brief algebraic manipulations, one can also verify that

$$\log f_K(x \,|\, z) \;=\; \sum_{1 \le s \le t \le m} N_{st} \Big\{ - H(\mathrm{Ber}(d_{st})) - d_{\mathrm{KL}}(\mathrm{Ber}(K_{st}) || \,\mathrm{Ber}(d_{st})) \Big\},$$

where $H(f) = -\sum_x f(x) \log f(x)$ denotes the Shannon entropy of probability distribution f, and $d_{\mathrm{KL}}(f||g) = \sum_x f(x) \log \frac{f(x)}{g(x)}$ the *Kullback–Leibler divergence* of f with respect to g. Because $d_{\mathrm{KL}}(f||g) \ge 0$ always, with equality holding if and only if $f = g$, it follows that the above quantity is maximized when $\mathrm{Ber}(K_{st}) = \mathrm{Ber}(d_{st})$ for all s and t, that is, when $K_{st} = d_{st}$. □

8.4 Learning block frequencies and block interaction parameters

We will discuss the article [BCL11]. A large random graph is modeled as a sequence of doubly stochastic block models $\mathrm{SBM}(\alpha, K^{(n)})$ on node set $[n]$

indexed by $n = 1, 2, \ldots$, where the prior block membership distribution α is a probability distribution on a set $S = [m]$, and the connectivity matrix is given by

$$K^{(n)}(s,t) \;=\; \rho_n K(s,t) \wedge 1, \tag{8.4}$$

where the *link density* ρ_n is a scalar such that $\rho_n \to 0$ as $n \to \infty$, and the *normalized kernel* $K : S \times S \to [0, \infty)$ is a symmetric function[2] normalized according to[3]

$$\sum_s \sum_t K(s,t)\,\alpha(s)\alpha(t) \;=\; 1.$$

As $n \to \infty$, one can verify (exercise) that any particular node pair is linked with probability $(1 + o(1))\rho_n$, and the expected degree of any node equals $(1+o(1))n\rho_n$. The statistical learning problem is now to determine the prior block membership distribution α and the normalized block interaction matrix K from a graph sample $X^{(n)}$ obtained from the SBM$(\alpha, K^{(n)})$ distribution.

A moment-based estimation approach for learning the model parameters is to compute the R-matching (or R-covering) densities defined in (7.4) of the observed graph sample $X^{(n)}$ for a suitable collection of small graphs R, and try to match the so-obtained empirical densities to the corresponding theoretical densities of the model. Because in the sparse setting with $\rho_n \to 0$, the empirical and model densities converge to zero, we need to work with normalized densities. For a graph R on node set $[r]$, the *normalized R-covering density* of the model is defined by

$$Q^*(R) \;=\; \sum_{z_1} \cdots \sum_{z_r} \alpha(z_1) \cdots \alpha(z_r) \prod_{ij \in E(R)} K(z_i, z_j),$$

and the normalized empirical R-covering density of the graph sample $X^{(n)}$ is defined by $\rho_n^{-|E(R)|}\hat{Q}_{X^{(n)}}(R)$ where $\hat{Q}_{X^{(n)}}(R)$ is defined in (7.5). The following result provides a sufficient condition for the normalized empirical R-covering density to be a consistent estimator of $Q^*_{\alpha,K}(R)$.

Theorem 8.7. *Assume that $cn^{-1} \le \rho_n \ll 1$, and that*

$$\sum_s \sum_t K(s,t)^{2r}\alpha(s)\alpha(t) < \infty.$$

Then for any acyclic graph R with r nodes,

$$\rho_n^{-|E(R)|}\hat{Q}_{X^{(n)}}(R) \;\xrightarrow{\mathbb{P}}\; Q^*(R).$$

[2] In the paper [BCL11] a different truncation $w1(w \le 1)$ was used in place of $w \wedge 1$, and $S = (0,1)$, but this should not make a difference.

[3] If S is an uncountable measurable space, then all sums over S involving $\alpha(u)$ should be replaced by integrals involving $\alpha(du)$.

Sketch of proof. Because the distribution of the random graph $X = X^{(n)}$ is invariant with respect to node relabeling, we may relabel the nodes of R so that $V(R) = [r]$. Moreover, $\mathbb{P}(X \supset R') = \mathbb{P}(X \supset R)$ whenever R' is isomorphic to R. Hence the expected R-covering density of X equals

$$\mathbb{E}\hat{Q}_{X^{(n)}}(R) \;=\; \mathbb{E}\frac{\sum_{R' \in \mathcal{G}_n(R)} 1(X^{(n)} \supset R')}{|\mathcal{G}_n(R)|} \;=\; \mathbb{P}(X^{(n)} \supset R).$$

Because the the entries X_{ij} are conditionally independent given the node labeling Z, it follows that

$$\mathbb{P}(X^{(n)} \supset R \,|\, Z = z) \;=\; \prod_{ij \in E(R)} (\rho_n K(z_i, z_j) \wedge 1),$$

and

$$\rho_n^{-|E(R)|}\mathbb{P}(X^{(n)} \supset R \,|\, Z = z) \;=\; \prod_{ij \in E(R)} (K(z_i, z_j) \wedge \rho_n^{-1}) \;\to\; \prod_{ij \in E(R)} K(z_i, z_j).$$

After multiplying the left side above by $\alpha(z_1) \cdots \alpha(z_r)$ and summing over $z_1, \ldots, z_r$, it follows (by Lebesgue's monotone convergence if there are infinitely many labels) that

$$\rho_n^{-|E(R)|}\mathbb{P}(X^{(n)} \supset R) \;\to\; Q^*(R),$$

and we conclude that

$$\mathbb{E}\,\rho_n^{-|E(R)|}\hat{Q}_{X^{(n)}}(R) \;\to\; Q^*(R).$$

To finish the proof by Chebyshev's inequality (i.e. the second moment method), it suffices to show that the fact that R is acyclic helps here

$$\mathrm{Var}\left(\rho_n^{-|E(R)|}\hat{Q}_{X^{(n)}}(R)\right) \;\to\; 0.$$

This is done in [BCL11, Proof of Theorem 1] (see also [Bol01, Sec 4.1]). □

Using matching densities instead of covering densities

This is nice to know, but not crucially important. For sparse doubly stochastic block models, the empirical matching and covering densities behave roughly similarly. By similar arguments as for the R-covering density, it follows that the expected R-matching density of $X = X^{(n)}$ equals

$$\mathbb{E}\hat{P}_X(R) \;=\; \mathbb{P}\Big(X_{ij} = R_{ij} \text{ for all } 1 \le i < j \le r\Big).$$

Observe that the difference between the covering and the matching densities is bounded by $\hat{Q}_X(R) - \hat{P}_X(R) \geq 0$ and

$$\begin{aligned}
\hat{Q}_X(R) - \hat{P}_X(R) &= \frac{1}{|\mathcal{G}_n(R)|} \sum_{R' \in \mathcal{G}_n(R)} 1(X \supset R') 1(X_{k\ell} = 1 \text{ for some } k\ell \notin E(R')) \\
&= \frac{1}{|\mathcal{G}_n(R)|} \sum_{R' \in \mathcal{G}_n(R)} \prod_{ij \in E(R')} X_{ij} 1(X_{k\ell} = 1 \text{ for some } k\ell \notin E(R')) \\
&\leq \frac{1}{|\mathcal{G}_n(R)|} \sum_{R' \in \mathcal{G}_n(R)} \left(\prod_{ij \in E(R')} X_{ij} \right) \left(\sum_{k\ell \notin E(R')} X_{k\ell} \right),
\end{aligned}$$

where $k\ell \notin E(R')$ refers to the $k\ell \in \binom{[r]}{2} \setminus E(R')$, so that

$$\begin{aligned}
\mathbb{E}|\hat{Q}_X(R) - \hat{P}_X(R)| &\leq \mathbb{E} \left(\prod_{ij \in E(R)} X_{ij} \right) \left(\sum_{k\ell \in \binom{[r]}{2} \setminus E(R)} X_{k\ell} \right) \\
&\leq \rho_n^{|E(R)|+1} \mathbb{E} \left(\prod_{ij \in E(R)} K(Z_i, Z_j) \right) \left(\sum_{k\ell \in \binom{[r]}{2} \setminus E(R)} K(Z_k, Z_\ell) \right) \\
&\leq c \rho_n^{|E(R)|+1}
\end{aligned}$$

under sufficient moment conditions on w. Hence by Markov's inequality,

$$\rho_n^{-|E(R)|} \hat{Q}_X(R) - \rho_n^{-|E(R)|} \hat{P}_X(R) \xrightarrow{\mathbb{P}} 0.$$

Hence by Theorem 8.7 it follows that also the normalized empirical matching density converges to $Q^*(R)$, according to

$$\rho_n^{-|E(R)|} \hat{P}_X(R) \xrightarrow{\mathbb{P}} Q^*(R).$$

8.5 Identifiability of the doubly stochastic block model from covering densities

The prior block membership distribution α can be viewed as a column vector of m numbers $\alpha_s \in [0, 1]$ normalized according to

$$\sum_{s=1}^{m} \alpha_s = 1. \tag{8.5}$$

In a finite label space $S = [m]$ we can ignore the truncation term in the kernel definition (8.4), and we can write $K^{(n)}(s,t) = \rho_n K_{s,t}$, where $\rho_n \in (0,1)$ is the overall link density and the limiting kernel K is now a symmetric m-by-m matrix with entries in $K_{st} \in [0,1]$ normalized by

$$\sum_{s=1}^{m}\sum_{t=1}^{m} K_{st}\alpha_s\alpha_t \;=\; 1. \tag{8.6}$$

To learn the model it is then sufficient to determine the m real numbers α_s and the $m(m+1)/2$ real numbers $K_{st}, 1 \le s \le t \le m$. Actually, a bit less is sufficient. Namely, (8.5) and (8.6) imply that we can omit learning one entry of α and one entry of K. Therefore, the number of free parameters in the model equals $m(m+3)/2 - 2$.

The limiting normalized R-covering density of a doubly stochastic block model with label distribution α and kernel K was found to be

$$Q^*(R) \;=\; \sum_{z_1}\cdots\sum_{z_r}\prod_{ij\in E(R)} K_{z_i,z_j}\,\alpha_{z_1}\cdots\alpha_{z_r}.$$

We have seen that the above model covering densities can be consistently estimated by the empirical covering densities computed from the observed graph. After observing a graph sample (X_{ij}) and then estimating the covering densities for a collection $R_1, \ldots, R_M$ of small graphs, we obtain M equations

$$\sum_{z_1}\cdots\sum_{z_r}\prod_{ij\in E(R_k)} K_{z_i,z_j}\,\alpha_{z_1}\cdots\alpha_{z_r} \;=\; Q^*(R_k), \qquad k = 1,\ldots,M,$$

involving the unknown parameters α_s and K_{st}. The identifiability problem then asks: *Do the above moment equations admit a unique solution?* This is a problem of algebraic statistics. To get a first feeling about whether or not the problem is easy to solve, let us compute the theoretical normalized R-covering density for some simple graphs first.

When R is a single link, we get

$$Q^*(\text{link}) \;=\; \sum_{s}\sum_{t} K_{st}\,\alpha_s\alpha_t \;=\; 1$$

due to the normalization constraint (8.6). For the triangle we obtain

$$Q^*(\text{triangle}) \;=\; \sum_{s}\sum_{t}\sum_{u} K_{st}K_{tu}K_{su}\,\alpha_s\alpha_t\alpha_u,$$

but this appears a complicated formula to analyze. To obtain simpler algebraic expressions, we will try computing covering densities for some acyclic

graphs. For the 3-path with $V(R) = \{1,2,3,4\}$ and $E(R) = \{\{1,2\},\{2,3\},\{3,4\}\}$, we find that

$$\begin{aligned} Q^*(\text{3-path}) &= \sum_{u_1,u_2,u_3,u_4} K_{u_1u_2}K_{u_2u_3}K_{u_3u_4}\,\alpha_{u_1}\alpha_{u_2}\alpha_{u_3}\alpha_{u_4} \\ &= \sum_{u_1,u_2,u_3,u_4} \alpha_{u_1} L_{u_1u_2}L_{u_2u_3}L_{u_3u_4} \\ &= \sum_{u_1}\sum_{u_4} \alpha_{u_1} L^3_{u_1u_4} \\ &= \sum_{u} \alpha_u (L^3 e)_u, \end{aligned}$$

where $L_{uv} = K_{uv}\alpha_v$ is the matrix product of K and the diagonal matrix with entries $\alpha_1,\ldots,\alpha_m$, and e is the column vectors of m ones. For the 3-star with $V(R) = \{1,2,3,4\}$ and $E(R) = \{\{1,2\},\{1,3\},\{1,4\}\}$, we get

$$\begin{aligned} Q^*(\text{3-star}) &= \sum_{u_1,u_2,u_3,u_4} K_{u_1u_2}K_{u_1u_3}K_{u_1u_4}\,\alpha_{u_1}\alpha_{u_2}\alpha_{u_3}\alpha_{u_4} \\ &= \sum_{u_1} \alpha_{u_1} \left(\sum_{u_2}\sum_{u_3}\sum_{u_4} K_{u_1u_2}K_{u_1u_3}K_{u_1u_4}\alpha_{u_2}\alpha_{u_3}\alpha_{u_4} \right) \\ &= \sum_{u} \alpha_u \left(\sum_{v} K_{uv}\alpha_v \right)^3 \\ &= \sum_{u} \alpha_u \left((Le)_u \right)^3. \end{aligned}$$

The above computations can be generalized to (exercise)

$$\begin{aligned} Q^*(k\text{-path}) &= \sum_{u} \alpha_u (L^k e)_u, \\ Q^*(\ell\text{-star}) &= \sum_{u} \alpha_u \left((Le)_u \right)^\ell. \end{aligned}$$

Even more generally, one can verify (exercise) that

$$Q^*(k\ell\text{-star}) = \sum_{u} \alpha_u (L^k e)_u^\ell, \tag{8.7}$$

where a *$k\ell$-star* refers to a graph of radius k obtained by joining the endpoints of ℓ paths of length k at a common hub node in the center. Hence an $(1,\ell)$-star is the usual ℓ-star.

Can we identify α and K from the covering densities of paths and stars? The first claim is that we can identify $(\alpha_1, \ldots, \alpha_m)$ from the ℓ-star covering densities with $\ell = 1, \ldots, 2m-1$. Why? Let X_1 be a random variable which takes on value $(Le)_u$ with probability α_u for all $u = 1, \ldots, m$. Then

$$\mathbb{E}X_1^\ell \;=\; \sum_u \alpha_u \left((Le)_u\right)^\ell \;=\; Q^*(\ell\text{-star}).$$

Hence the knowledge of normalized ℓ-star covering densities amounts to the knowledge of the moments $\mathbb{E}X_1^\ell$ for $\ell = 1, \ldots, 2m-1$. Then a classical theorem about the method of moments [Fel71] tells that the distribution (support and probabilities) of X_1 can be recovered from sufficiently many moments $\mathbb{E}X_1, \mathbb{E}X_1^2, \ldots$ Hence, we may obtain the label distribution α and the rows sums $(Le)_1, \ldots (Le)_m$ from the star covering densities.

Next, let X_k be a random variable which takes on value $(L^k e)_u$ with probability α_u. Then by (8.7),

$$\mathbb{E}X_k^\ell \;=\; \sum_u \alpha_u (L^k e)_u^\ell \;=\; Q^*(k\ell\text{-star}),$$

and hence we may recover the rows sums $(L^k e)_1, \ldots, (L^k e)_m$ from the $k\ell$-star covering densities. Let us now define the m-by-m square matrices

$$V^{(1)} \;=\; \begin{bmatrix} e & Le & \cdots & L^{m-1}e \end{bmatrix}$$

and

$$V^{(2)} \;=\; \begin{bmatrix} Le & L^2 e & \cdots & L^m e \end{bmatrix}.$$

Then

$$LV^{(1)} \;=\; V^{(2)},$$

and if the columns of $V^{(1)}$ are linearly independent, we obtain the matrix L from

$$L \;=\; V^{(2)}(V^{(1)})^{-1},$$

and thereafter the matrix K by $K_{uv} = L_{uv}\alpha_v^{-1}$. Hence we have proved the following result.

Theorem 8.8. *Assume that the vectors $e, Le, \ldots, L^{m-1}e$ are linearly independent and $\alpha_u > 0$ for all u. Then the label distribution α and the normalized kernel K can be identified from the normalized covering densities $Q^*(k\ell\text{-star})$ with $k, \ell \geq 1$.*

Combining Theorems 8.7 and 8.8 yields a consistent way to estimate the parameters α and K of a large and sparse doubly stochastic block model from the normalized empirical covering densities of $k\ell$-stars computed from a single large sample $X^{(n)}$.