Chapter 6

Degree distributions

6.1 Mean degrees in stochastic block models

In this section we will study a stochastic block model with n nodes and m communities, having density parameter ρ and block interaction matrix K. This is an inhomogeneous Bernoulli random graph G on node set $[n]$ where the link probabilities are of the form

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

where $\rho > 0$ is scalar, $z = (z_1, \ldots, z_n)$ is a list of node attributes with values in $[m]$, and K is a symmetric nonnegative m-by-m matrix. The labelling $i \mapsto z_i$ partitions the node set $[n]$ into m disjoint blocks $C_s = \{i : z_i = s\}$, and the relative size of block s is denoted by

$$\mu_s \;=\; \frac{1}{n} \sum_{i=1}^{n} 1(z_i = s).$$

The vector $(\mu_s)_{s=1}^m$ is a probability distribution on $[m]$ called the empirical block membership distribution, and μ_s can be interpreted as the probability that a randomly selected node belongs to block C_s. The following result describes the expected degrees in the model.

Theorem 6.1. *For a stochastic block model with smallest relative block size $\mu_{\min} = \min_s \mu_s$, the expected degree of any node i in community s satisfies*

$$\mathbb{E} \deg_G(i) \;=\; n\rho\lambda_s - \rho K_{s,s} \;=\; (1 - \epsilon_1) n\rho\lambda_s \tag{6.1}$$

and the expected average degree equals

$$\mathbb{E} \deg_G(U) \;=\; n\rho\lambda - \rho \sum_s \mu_s K_{s,s} \;=\; (1 - \epsilon_2) n\rho\lambda, \tag{6.2}$$

where $\lambda_s = \sum_t K_{s,t}\mu_t$, $\lambda = \sum_{s,t} \mu_s K_{s,t} \mu_t$, and $0 \le \epsilon_1, \epsilon_2 \le (n\mu_{\min})^{-1}$.

Proof. (i) The degree of node i may be written as $\deg_G(i) = \sum_{j\neq i} G_{ij}$ where G_{ij} are independent $\mathrm{Ber}(p_{ij})$-distributed random variables. Hence

$$\mathbb{E}\deg_G(i) \;=\; \sum_{j\neq i} p_{ij} \;=\; \sum_{j\neq i} \rho K_{z_i,z_j} \;=\; \rho\sum_{j=1}^n K_{z_i,z_j} - \rho K_{z_i,z_i}$$

Because $z_i = s$ and the number of nodes in community t equals $n\mu_t$, we find that $\sum_{j=1}^n K_{z_i,z_j} = \sum_{t=1}^m K_{s,t} n\mu_t = n\lambda_s$, and the first equality in (6.1) follows. To verify the second equality in (6.1), note that

$$0 \;\le\; \epsilon_1 \;=\; \frac{K_{s,s}}{n\lambda_s} \;=\; \frac{K_{s,s}\mu_s}{n\lambda_s\mu_s} \;\le\; \frac{\sum_t K_{s,t}\mu_t}{n\lambda_s\mu_s} \;=\; (n\mu_s)^{-1}.$$

(ii) The expected average degree equals

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\deg_G(i) \;=\; \frac{1}{n}\sum_{s=1}^m (n\mu_s)(n\rho\lambda_s - \rho K_{ss}) \;=\; n\rho\sum_{s=1}^m \mu_s\lambda_s - \rho\sum_{s=1}^m \mu_s K_{ss}.$$

The first equality in (6.2) hence follows by noting that $\sum_{s=1}^m \mu_s\lambda_s = \lambda$. Observe next that $\sum_s \mu_s K_{s,s} \le \sum_{s,t}\mu_s K_{s,t} = \sum_{s,t}\mu_s K_{s,t}\mu_t(\mu_t)^{-1} \le (\mu_{\min})^{-1}\lambda$. Hence

$$0 \;\le\; \epsilon_2 \;=\; \frac{\rho\sum_s \mu_s K_{s,s}}{n\rho\lambda} \;\le\; \frac{1}{n\mu_{\min}}.$$

□

Theorem 6.1 shows that the average degree of a random graph generated by a stochastic block model is of the order $n\rho$. When overall link density $\rho = \rho_n$ depends on the scale, we get different limiting regimes corresponding to different levels of sparsity, see Table 6.1.

Exercise 6.2. Verify that under the assumptions of Theorem 6.1, for a node in community s, the mean number of neighbors in community t is approximately $n\rho K_{s,t}\mu_t$.

Density	**Average degree**	**Regime**
$\rho \ll n^{-1}$	$d_{\mathrm{ave}} \ll 1$	Very sparse
$\rho \approx cn^{-1}$	$d_{\mathrm{ave}} \approx c$	Sparse with bounded degree
$n^{-1} \ll \rho \ll 1$	$1 \ll d_{\mathrm{ave}} \ll n$	Sparse with diverging degree
$\rho \approx c$	$d_{\mathrm{ave}} \approx cn$	Dense

Table 6.1: Different regimes of large graph models.

6.2 Poisson approximation

The following result, sometimes called *Le Cam's inequality* after a famous Berkeley statistician Lucien Le Cam, illustrates how to apply the stochastic coupling method to get an upper bound on the distance between a sum of independent $\{0,1\}$-valued random variables and a Poisson distribution.

Theorem 6.3. *Let A_i be independent $\{0,1\}$-valued random variables such that $\mathbb{E}A_i = a_i$ and $\sum_i a_i < \infty$. Then*

$$d_{\mathrm{tv}}\Big(\operatorname{Law}(\sum_i A_i),\, \operatorname{Poi}(\sum_i a_i)\Big) \;\le\; \sum_i a_i^2.$$

Proof. By applying (5.3) and Theorem 5.1, we see that for every i there exists a coupling $(\hat{A}_i, \hat{B}_i)$ of X_i and a $\operatorname{Poi}(a_i)$-distributed random integer B_i, so that

$$\mathbb{P}(\hat{A}_i \ne \hat{B}_i) \;\le\; a_i(1 - e^{-a_i}). \tag{6.3}$$

By a standard technique of probability theory, it is possible to construct all of the bivariate random variables $(\hat{A}_i, \hat{B}_i)$, $i \in I$, on a common probability space and in such a way that these bivariate random variables are mutually independent (nevertheless, $\hat{A}_i$ and $\hat{B}_i$ are dependent for each i). Then define $\hat{A} = \sum_i \hat{A}_i$ and $\hat{B} = \sum_i \hat{B}_i$. Then $\operatorname{Law}(\hat{A}) = \operatorname{Law}(\sum_i A_i)$. Moreover, because the sum of independent Poisson-distributed random integers is Poisson-distributed, it follows that $(\hat{A}, \hat{B})$ is a coupling of $\sum_i A_i$ and a $\operatorname{Poi}(\sum_i a_i)$-distributed random integer B. By applying (6.3) and the union bound, this coupling satisfies

$$\mathbb{P}(\hat{A} \ne \hat{B}) \;=\; \mathbb{P}(\cup_{i\in I}\{\hat{A}_i \ne \hat{B}_i\}) \;\le\; \sum_{i\in I} \mathbb{P}(\hat{A}_i \ne \hat{B}_i) \;\le\; \sum_{i\in I} a_i(1 - e^{-a_i}).$$

By applying Theorem 5.1, it now follows that

$$d_{\mathrm{tv}}(\sum_i A_i, \operatorname{Poi}(\sum_i a_i)) \;\le\; \mathbb{P}(\hat{A} \ne \hat{B}) \;\le\; \sum_{i\in I} a_i(1 - e^{-a_i}).$$

This implies the claim after noting that $1 - e^{-a_i} \le a_i$. $\square$

Exercise 6.4. For a sequence of probability distributions we denote $\mu_n \xrightarrow{tv} \mu$ when $d_{\mathrm{tv}}(\mu_n, \mu) \to 0$.

(a) Apply Le Cam's inequality to show that when $p_n \ll n^{-1/2}$,

$$d_{\mathrm{tv}}\Big(\operatorname{Bin}(n, p_n),\, \operatorname{Poi}(np_n)\Big) \;\to\; 0.$$

(b) As a consequence, derive Poisson's law of small numbers:

$$\operatorname{Bin}\left(n, \frac{\lambda}{n}\right) \xrightarrow{tv} \operatorname{Poi}(\lambda).$$

6.3 Degree distributions in sparse SBMs

6.3.1 Blockwise degree distribution

In a stochastic block model, all nodes in the same block are statistically identical, and the common degree distribution of nodes in block s is here denoted by f_s. We denote by entrywise max norm of a matrix by $\|K\|_{\max} = \max_{s,t} |K_{s,t}|$. As corollary of the following theorem, it follows that f_s is accurately approximated by a Poisson distribution when $\rho = \rho_n$ satisfies $\rho_n \ll n^{-1/2}$.

Theorem 6.5. *In a stochastic block model, the degree distribution* $f_s = \mathrm{Law}(\deg_G(i))$ *of any node* i *in block* s *is approximated by a Poisson distribution* $g_s = \mathrm{Poi}(n\rho\lambda_s)$ *according to*

$$d_{\mathrm{tv}}(f_s, g_s) \ \le\ \|K\|_{\max}^2 n\rho^2 + \|K\|_{\max}\rho,$$

where $\lambda_s = \sum_t K_{s,t}\mu_t$.

Proof. The degree of node i in block s may be written as $\deg_G(i) = \sum_{j \ne i} G_{ij}$ where G_{ij} are mutually independent $\mathrm{Ber}(p_{ij})$-distributed random variables. By Theorem 6.1, we know that the mean of $\deg_G(i)$ equals $\ell_s = n\rho\lambda_s - \rho K_{s,s}$. By Le Cam's inequality (Theorem 6.3), it follows that

$$d_{\mathrm{tv}}\Big(f_s,\, \mathrm{Poi}(\ell_s)\Big) \ \le\ \sum_{j\ne i} p_{ij}^2 \ =\ \rho^2 \sum_{j \ne i} (K_{z_i,z_j})^2 \ \le\ n\rho^2 \|K\|_{\max}^2,$$

where $\|K\|_{\max} = \max_{s,t} |K_{s,t}|$ denotes the entrywise max norm of K. Now by Lemma 5.4 and the inequality $1 - t \le e^{-t}$, it follows that it follows that

$$d_{\mathrm{tv}}\Big(\mathrm{Poi}(\ell_s),\, \mathrm{Poi}(n\rho_n\lambda_s)\Big) \ \le\ |\ell_s - n\rho_n\lambda_s| \ =\ \rho K_{s,s} \ \le\ \|K\|_{\max}\rho.$$

We conclude using the triangle inequality of the total variation metric that

$$d_{\mathrm{tv}}\Big(f_s,\, \mathrm{Poi}(n\rho\lambda_s)\Big) \ \le\ n\rho^2\|K\|_{\max}^2 + \|K\|_{\max}\rho.$$

□

6.3.2 Typical degree distribution

By a *typical node* of a graph G we mean a node U sampled uniformly at random from the node set of the graph. When the graph is random, the degree of a typical node $\deg_G(U)$ involves two sources of randomness: the randomness associated with the graph G, and the randomness associated with the sampling of U.

A *mixed Poisson distribution* with mixing distribution ϕ is the probability distribution $\mathrm{MPoi}(\phi)$ on the nonnegative integers with probability density

$$\mathbb{E}e^{-\Lambda}\frac{\Lambda^x}{x!}, \quad x = 0, 1, \ldots,$$

where Λ is a random variable distributed according to ϕ, a probability distribution on $\mathbb{R}_+$. Samples from $\mathrm{MPoi}(\phi)$ can be generated by first sampling a random variable Λ from ϕ, and conditionally on $\Lambda = \lambda$, sampling from a Poisson distribution with mean λ.

The distribution $g = \sum_{s=1}^m \mu_s g_s$ in Theorem 6.6 below is a mixed Poisson with mixing distribution ϕ being a probability distribution on a finite set $\{a_1, \ldots, a_m\}$ with $a_s = n\rho_n\lambda_s$ such that $\phi(a_s) = \mu_s$ for all $s = 1, \ldots, m$.

Theorem 6.6. *In a stochastic block model, the degree distribution $f = \mathrm{Law}(\deg_G(U))$ of a typical node is approximated by a mixed Poisson distribution $g = \sum_s \mu_s g_s$ with $g_s = \mathrm{Poi}(n\rho_n\lambda_s)$ according to*

$$d_{\mathrm{tv}}(f, g) \ \le \ \|K\|_{\max}^2 n\rho^2 + \|K\|_{\max}\rho,$$

where $\lambda_s = \sum_t K_{s,t}\mu_t$. Especially, for $\rho_n = n^{-1}$, the expected relative frequency $f(x) = f^{(n)}(x)$ of nodes of degree x satisfies

$$f^{(n)}(x) \ \to \ \sum_{s=1}^m \mu_s e^{-\lambda_s}\frac{\lambda_s^x}{x!}.$$

Proof. Denote by $f(x) = \mathbb{P}(\deg_G(U) = x)$ the typical node degree distribution. Because $\mathbb{P}(z_U = s) = \mu_s$, we find that

$$f(x) \ = \ \sum_{s=1}^m \mu_s f_s(x),$$

where f_s is the common degree distribution of nodes in block s. By Theorem 6.5

$$d_{\mathrm{tv}}(f_s, g_s) \ \le \ \|K\|_{\max}^2 n\rho^2 + \|K\|_{\max}\rho.$$

Then (use the L_1-distance representation of the total variation distance and triangle inequalities for the total variation distance),

$$d_{\mathrm{tv}}(f, g) \ = \ d_{\mathrm{tv}}\Big(\sum_s \mu_s f_s, \sum_s \mu_s g_s\Big) \ \le \ \sum_s \mu_s d_{\mathrm{tv}}(f_s, g_s).$$

□

6.4 Joint degree distribution

Many results related to large random graph rely on the fact that several local characteristics of the graph are approximately independent for large n. In statistics it is important to quantify how close certain observables are to being fully independent. Here we discuss the case of degrees.

Let $G(p)$ be a Bernoulli random graph on node set $[n]$ where each unordered node pair $\{i, j\}$ is connected by a link with probability p_{ij}, independently of other node pairs. Denote by

$$\operatorname{Law}(D_i : i \in I)$$

the joint distribution of the degrees $D_i = \deg_G(i)$ for a set of nodes $I \subset [n]$. The degrees D_i are not independent, but the dependence is not strong in large sparse random graphs. We may quantify this by measuring how much the joint degree distribution deviates from the product distribution

$$\prod_{i \in I} \operatorname{Law}(D_i)$$

which represents the joint distribution of *independently* sampled random integers from the distributions $\operatorname{Law}(D_i)$.

A collection of random variables $(X_i :\in I)$ whose joint distribution depends on a scale parameter n, is called *asymptotically independent* if

$$d_{\mathrm{tv}}\left(\operatorname{Law}(X_i : i \in I), \prod_{i \in I} \operatorname{Law}(X_i)\right) \ \to \ 0 \quad \text{as } n \to \infty.$$

Theorem 6.7. *The joint degree distribution of an arbitrary set of nodes I in a Bernoulli random graph with link probabilities p_{ij} satisfies*

$$d_{\mathrm{tv}}\left(\operatorname{Law}(D_i : i \in I), \prod_{i \in I} \operatorname{Law}(D_i)\right) \ \le \ 4 \sum_{i,j \in I : i < j} (1 - p_{ij}) p_{ij}.$$

As an immediate application of the above theory, we obtain the following result for sparse SBMs.

Proposition 6.8. *For a sparse stochastic block model with density parameter $\rho_n \ll 1$ and community link matrix K, the degrees of any set of $n_0 \ll \rho_n^{-1/2}$ nodes are asymptotically independent.*

Proof. For any node set I of size n_0,

$$4 \sum_{i,j \in I : i < j} (1 - p_{ij}) p_{ij} \ \le \ 4 \sum_{i,j \in I : i < j} p_{ij} \ \le \ 4 \rho_n \|K\|_{\max} n_0^2$$

The right side tends to zero when $\rho_n n_0^2 \to 0$ (and the community link matrix K does not depend on the scale parameter, which we implicitly assume here throughout). $\square$

Proof of Theorem 6.7. The proof is based on a coupling argument described in [vdH17, Theorem 6.7(b)]. After relabeling the node set if necessary, we may and will assume that $I = \{1, 2, \ldots, m\}$. Let G the adjacency matrix of the random graph, and let $\hat{G}$ be an *independent copy* of G. That is, we sample $\hat{G}$ from the same distribution as G, independently. Then we define

$$\tilde{D}_i \;=\; \sum_{j:j<i} \hat{G}_{ij} + \sum_{j:j>i} G_{ij}. \tag{6.4}$$

The random integers $\tilde{D}_i$ are *not* degrees of G nor $\hat{G}$. Nevertheless, we see that $\mathrm{Law}(\tilde{D}_i) = \mathrm{Law}(D_i)$ because all random variables on the right side above are independent. Note that

$$\begin{aligned}
\tilde{D}_1 &= G_{12} + G_{13} + G_{14} + \cdots \\
\tilde{D}_2 &= \hat{G}_{12} + G_{23} + G_{24} + \cdots \\
\tilde{D}_3 &= \hat{G}_{13} + \hat{G}_{23} + G_{34} + \cdots
\end{aligned}$$

and so on. Because all terms in the three above sums are independent, it follows that $\tilde{D}_1, \tilde{D}_2, \tilde{D}_3$ are independent. In fact, one may verify by induction that $\tilde{D}_1, \ldots, \tilde{D}_n$ are all mutually independent. Hence so is the sublist $\tilde{D}_I = (\hat{D}_i : i \in I)$, and the distribution of the list $\tilde{D}_I$ equals $\prod_{i \in I} \mathrm{Law}(D_i)$. Now the pair $(D_I, \tilde{D}_I)$ constitutes a coupling of $\mathrm{Law}(D_I)$ and $\prod_{i \in I} \mathrm{Law}(D_i)$. Hence

$$\begin{aligned}
d_{\mathrm{tv}}\left(\mathrm{Law}(D_i : i \in I), \prod_{i \in I} \mathrm{Law}(D_i) \right) &\le \mathbb{P}(D_I \neq \tilde{D}_I) \\
&= \mathbb{P}\Big(\bigcup_{i \in I} \{D_i \neq \tilde{D}_i\} \Big) \\
&\le \sum_{i \in I} \mathbb{P}(D_i \neq \tilde{D}_i).
\end{aligned}$$

From (6.4) we see that $\tilde{D}_i - D_i = \sum_{j:j<i} (\hat{G}_{ij} - G_{ij})$. Hence $D_i = \tilde{D}_i$ unless $G_{ij} \neq \hat{G}_{ij}$ for one or more indices $j < i$. Therefore

$$\mathbb{P}(D_i \neq \tilde{D}_i) \;\le\; \mathbb{P}(\cup_{j:j<i} \{G_{ij} \neq \hat{G}_{ij}\}) \;\le\; \sum_{j:j<i} \mathbb{P}(G_{ij} \neq \hat{G}_{ij}).$$

Because

$$\begin{aligned}\mathbb{P}(G_{ij} \neq \hat{G}_{ij}) &= \mathbb{P}(G_{ij} = 0, \hat{G}_{ij} = 1) + \mathbb{P}(G_{ij} = 1, \hat{G}_{ij} = 0) \\ &= 2(1 - p_{ij})p_{ij},\end{aligned}$$

we conclude that

$$d_{\mathrm{tv}}\left(\mathrm{Law}(D_i : i \in I), \prod_{i \in I} \mathrm{Law}(D_i)\right) \leq 2 \sum_{i \in I} \sum_{j \in I : j \neq i} (1 - p_{ij}) p_{ij},$$

and the claim follows. □

Exercise 6.9. If D_1 and D_2 are asymptotically independent, show that $\mathrm{cov}(\phi(D_1, \phi(D_2)) \to 0$ for any bounded function ϕ.

6.5 Empirical degree distributions

6.5.1 Empirical distributions of large data sets

To obtain a tractable sparse graph model, we need to impose some regularity assumptions on the behavior of node attributes. We will denote the *empirical distribution* of the list $x^{(n)}$ by

$$\mu_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1(x_i^{(n)} \in B)$$

returns the relative frequency of node attributes with values in $B \subset \mathbb{R}$. Alternatively, μ_n is the probability distribution of random variable X_n obtained by picking an element of the list uniformly at random. We assume that for large graphs, the distribution of attributes can be approximated by a limiting probability distribution μ on $(0, \infty)$. More precisely, we assume that $\mu_n \to \mu$ *weakly*, that is, example

$$\mathbb{E}\phi(X_n) \to \mathbb{E}\phi(X)$$

for any continuous and bounded function $\phi : (0, \infty) \to \mathbb{R}$ and random variables X_n distributed according to μ_n and X distributed according to μ. We also say that $\mu_n \to \mu$ *weakly with k-th moments* if in addition $\mathbb{E}X_n^k \to \mathbb{E}X^k$ and $\mathbb{E}X_n^k$, $\mathbb{E}X^k$ are finite[1]. For a thorough treatment of the aspects of weak convergence of probability measures, see for example [Kal02, Section 4]. The main fact is that when the limiting distribution has a continuous cumulative distribution F, then $\mu_n \to \mu$ weakly if and only if $F_n(t) \to F(t)$ for all t, where $F_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1(x_i^{(n)} \leq t)$ is the *empirical cumulative distribution* of the list $x^{(n)}$.

[1] This corresponds to convergence of probability measure in the Wasserstein-k metric.

Example 6.10 (Random attribute lists)**.** A fundamental example is the following setting. Assume that $X_1, X_2, \ldots$ are independent random numbers sampled from a probability distribution μ which has a finite k-th moment. Then the empirical distribution μ_n of the list $X^{(n)} = (X_1, \ldots, X_n)$ is a *random* probability distribution. As consequence of the strong law of large numbers and the Glivenko–Cantelli theorem it follows that with probability one, $\mu_n \to \mu$ weakly with k-th moments.

Here, as elsewhere, we denote $f_n \ll g_n$ or $f_n = o(g_n)$ when $f_n/g_n \to 0$.

Lemma 6.11. *Assume the empirical distribution of $x^{(n)}$ converges weakly and with first moments to a probability distribution μ. Then $\max_{i \in [n]} x_i^{(n)} \ll n$.*

Proof. Let X_n be a μ_n-distributed random number for each n. Then by Lemma A.5, the sequence (X_n) is uniformly integrable, and for any $\epsilon > 0$, it follows that

$$\begin{aligned} n\mathbb{P}(X_n > \epsilon n) \;=\; \epsilon^{-1}\mathbb{E}\epsilon n 1(X_n > \epsilon n) \;&\le\; \epsilon^{-1}\mathbb{E}X_n 1(X_n > \epsilon n) \\ &\le\; \epsilon^{-1} \sup_m \mathbb{E}X_m 1(X_m > \epsilon n) \;\to\; 0. \end{aligned}$$

But this means that

$$\sum_{i=1}^n 1(x_i^{(n)} > \epsilon n) \;=\; n \frac{1}{n} \sum_{i=1}^n 1(x_i^{(n)} > \epsilon n) \;=\; n\,\mathbb{P}(X_n > \epsilon n) \;\to\; 0.$$

Because the left side above is integer-valued, we conclude that exists n_0 such that $\sum_{i=1}^n 1(x_i^{(n)} > \epsilon n) = 0$ for all $n > n_0$. This implies that $n^{-1}x_i^{(n)} \le \epsilon$ for all $i \in [n]$, or equivalently, $n^{-1} \max_{i \in [n]} x_i^{(n)} \le \epsilon$ for all $n > n_0$, and the claim follows.

□

6.6 Product-form kernels (not part of Fall 2018 course)

Recall from Section 1.4 the definition of inhomogeneous Bernoulli graphs and latent position graphs. Many real-world data sets have highly varying degree distributions, where most nodes have a relatively small degree and a few hub nodes have an extremely high degree. Such data sets can be modeled as large inhomogeneous random graphs where the attribute space is $\mathcal{S} = [0, \infty)$ and the attributes are considered weights. A natural idea is the multiplicative