Lecture 5

Zongge Liu

1 Random Graph Review

$G(V, E)$, $E \subset V \times V$, random graph. $A_{ij} \sim Bernoulli(p)$ (Undirected).

- Phase transition to giant component at $\lambda = p(n-1) = 1$
- Binomial (Poisson) Degree of Distribution.
- Diameter $O(\log n)$ above $\lambda = 1$.
- Rare p^3 triangles, little transitivity. $P(A_{ik} = 1 | A_{ij} = 1, A_{jk} = 1) = p$

2 Block Model

2.1 Definition

All nodes are divided into k blocks. Z_i is the block to which node i belongs.

$Pr(A_{ij} = 1 | Z_i = r, Z_j = s) = b_{rs}$ where b is a $b \times b$ matrix (affinity matrix). All the edges are independent given Z ("block assignment").

2.2 Comparison Between Random Graph and Block Model

Table 1: Comparison between random graph and block model

Random graph	Block model
Edges are independent	Edges are independent given Z
All edges have equal probability	All edges between 2 blocks have equal probability (Within a block model it looks like a random graph)
All nodes have equal binomial degree distribution	All nodes in the same block have the same distribution $\sum Binomial(n_s - \delta_{rs}, b_{rs})$
Proportion of triangles p^3, no-transitivity $P(A_{ik} = 1 \| A_{ij} = 1, A_{jk} = 1) = p$	Overall density $\sum_{r,s} b_{r,s} \frac{n_r n_s}{n^2}$. $P(A_{ik} = 1 \| A_{ij} = 1, A_{jk} = 1) = \sum_{r,s,q} b_{rq} Pr(Zi = r, Z_j = s, Z_k = q \| A_{ij} = 1, A_{jk} = 1)$
Exponential family: $\hat{p} = \sum_{i,j} \frac{A_{ij}}{\binom{n}{2}}$	Exponential family: $\hat{b_{rs}} = \frac{e_{rs}}{n_r n_s}$

2.3 Some Derivations for Table 1

Degree Distribution: say n_r nodes in block r$(n = \sum_{r=1}^{k} n_r)$, degree distribution of a node in block r is $\sum_{s=1}^{k} Binomial(n_s - \delta_{rs}, b_{rs})$.
Baseline probability of an edge in a model:

$$\begin{aligned} P_{eff} = Pr(A_{ij} = 1) &= \sum_{r=1}^{k}\sum_{s=1}^{k} P(A_{ij}, Z_i = r, Z_j = s) = \sum_{r=1}^{k}\sum_{s=1}^{k} P(A_{ij}|Z_i = r, Z_j = s)P(Z_i = r, Z_j = s) \\ &= \sum_{r=1}^{k}\sum_{s=1}^{k} b_r s Pr(Z_i = r, Z_j = s) = \sum_{r=1}^{k}\sum_{s=1}^{k} b_{rs}\frac{n_r n_s}{n^2} \end{aligned}$$

Probability of completing a triangle:

$$\begin{aligned} &P(A_{ik} = 1|A_{ij} = 1, A_{jk} = 1) \\ &= \sum_{r,s,q} Pr(Zi = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1) \\ &= \sum_{r,s,q} Pr(A_{ik} = 1, Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1) \\ &= \sum_{r,s,q} Pr(A_{ik} = 1|Z_i = r, Z_j = s, Z_k = q, A_{ij} = 1, A_{jk} = 1)Pr(Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1) \\ &= \sum_{r,s,q} Pr(A_{ik} = 1|Z_i = r, Z_k = q)Pr(Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1) \\ &= \sum_{r,s,q} b_{rq} Pr(Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1) \end{aligned}$$

Likelihood: firstly notice that $L = \prod_{i,j} b_{Z_i,Z_j}^{A_{ij}}(1 - b_{Z_i,Z_j})^{1-A_{ij}}$. We assume we have n_r nodes in block r and e_{rs} edges between block r and block s

$$\begin{aligned} l = log(L) &= \sum_{ij} A_{ij} \log b_{Z_i} Z_j + (1 - A_{ij}) \log(1 - b_{Z_i} Z_j) \\ &= \sum_{r=1}^{k}\sum_{s=1}^{k} e_{rs} \log(b_{rs}) + (n_r n_s - e_{rs} logit(b_{rs})) \\ &= \sum_{r,s} n_r n_s \log(1 - b_{rs}) + e_{rs} logit(b_{rs}) \end{aligned}$$

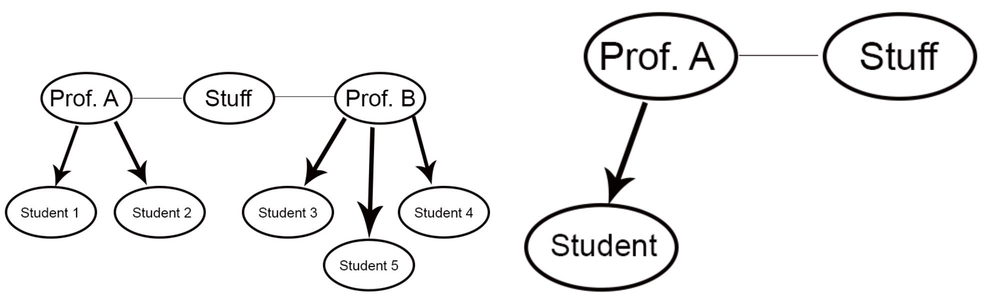
We can show $\hat{b_{rs}} = \frac{e_{rs}}{n_r n_s}$.

3 Routes to the Block Model

- Give a partition into blocks, most random graphs with observed edges and densities.

Exponential families at MLE. $\Longleftrightarrow$ Most random distributions where
observed statistics match expectations.
(maximum entropy distribution)
(distributions closest to homogeneous random graphs)

- Graph theory: Szemeredi regularity lemma. For any graph G of n nodes and any ϵ_j, we can divide the nodes into $s(n, \epsilon)$ blocks, and the edge counts come within ϵ of expectations for a k-block model.
- Third source for block model: *role model* in sociology, *regular equivalence*. Two models are structurally equivalent when they have the same neighbors. This can be illustrated in Fig. 1.



(a) a

(b) b

Figure 1: The left figure a can be represented by the right figure b under the regular equivalence principle.

4 Discussion

For what value of parameters are all $E(e_{rs})$ equal to observed e_{rs}? Maximum Entropy Problem: Given observed statistics $t_1, t_2 ... t_p$, find the distribution where $E(T_i) = t_i$ and entropy is maximized. The solution is always $p \propto e^{\sum_{i=1}^{n} \theta_i t_i(x)}$ with θ set to MLE.