1 Random Graph Review

\( G(V, E), E \subset V \times V \), random graph. \( A_{ij} \sim \text{Bernoulli}(p) \) (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>
<thead>
<tr>
<th>Random graph</th>
<th>Block model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edges are independent</td>
<td>Edges are independent given ( Z )</td>
</tr>
<tr>
<td>All edges have equal probability</td>
<td>All edges between 2 blocks have equal probability (Within a block model it looks like a random graph)</td>
</tr>
<tr>
<td>All nodes have equal binomial degree distribution</td>
<td>All nodes in the same block have the same distribution ( \sum \text{Binomial}(n_s - \delta_{rs}, b_{rs}) )</td>
</tr>
<tr>
<td>Proportion of triangles ( p^3 ), no-transitivity</td>
<td>Overall density ( \sum_{r,s} b_{rs} \frac{n_r n_s}{n^2} ). ( P(A_{ik} = 1</td>
</tr>
<tr>
<td>Exponential family: ( \hat{p} = \sum_{i,j} \frac{A_{ij}}{\binom{n}{2}} )</td>
<td>Exponential family: ( b_{rs} = \frac{e_{rs}}{n_r n_s} )</td>
</tr>
</tbody>
</table>
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 \( \sum_{s=1}^{k} \text{Binomial}(n_s - \delta_{rs}, b_{rs}) \).

Baseline probability of an edge in a model:

\[
P_{\text{ef}} = Pr(A_{ij} = 1) = \sum_{r=1}^{k} \sum_{s=1}^{k} P(A_{ij}, Z_i = r, Z_j = s) = \sum_{r=1}^{k} \sum_{s=1}^{k} P(A_{ij}|Z_i = r, Z_j = s)P(Z_i = r, Z_j = s)
\]

\[
= \sum_{r=1}^{k} \sum_{s=1}^{k} b_{rs}Pr(Z_i = r, Z_j = s) = \sum_{r=1}^{k} \sum_{s=1}^{k} b_{rs} \frac{n_r n_s}{n^2}
\]

Probability of completing a triangle:

\[
P(A_{ik} = 1|A_{ij} = 1, A_{jk} = 1) = \sum_{r,s,q} \Pr(Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1)
\]

\[
= \sum_{r,s,q} \Pr(A_{ik} = 1, Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1)
\]

\[
= \sum_{r,s,q} \Pr(A_{ik} = 1|Z_i = r, Z_j = s, Z_k = q, A_{ij} = 1, A_{jk} = 1)Pr(Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1)
\]

\[
= \sum_{r,s,q} \Pr(A_{ik} = 1|Z_i = r, Z_k = q)Pr(Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1)
\]

\[
= \sum_{r,s,q} b_{rq}Pr(Z_i = r, Z_j = s, Z_k = q|A_{ij} = 1, A_{jk} = 1)
\]

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 \( s \)

\[
l = \log(L) = \sum_{ij} A_{ij} \log b_{Z_i Z_j} + (1 - A_{ij}) \log (1 - b_{Z_i Z_j})
\]

\[
= \sum_{r=1}^{k} \sum_{s=1}^{k} e_{rs} \log(b_{rs}) + (n_r n_s - e_{rs} \logit(b_{rs}))
\]

\[
= \sum_{r,s} n_r n_s \log(1 - b_{rs}) + e_{rs} \logit(b_{rs})
\]

We can show \( b_{rs}^* = \frac{e_{rs}}{n_r n_s}. \)

3 Routes to the Block Model

- Give a partition into blocks, most random graphs with observed edges and densities.
Exponential families at MLE. $\iff$ 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 $\epsilon_j$, we can divide the nodes into $s(n, \epsilon)$ blocks, and the edge counts come within $\epsilon$ 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.

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 $\theta$ set to MLE.