Random Dot Product Graphs

STAT-S 675

Fall 2021

Recall From Last Time . . .

- Let $G = (V, E)$ be an undirected and hollow graph with $|V| = n$ and adjacency matrix A
  - $A \in \mathbb{R}^{n \times n}$ is symmetric with zero diagonals
- Suppose $G \sim F(\theta)$
  - What kind of $F(\theta)$ make sense here?
  - Given F and observed G, how can we estimate θ?

Recall From Last Time ...

$$A_{ij} = \begin{cases} 1 & \exists \text{ edge between } i \text{ and } j \\ 0 & \text{else} \end{cases}$$

$A_{ji} = A_{ij}$ and $A_{ii} = 0 \ \forall i, j \in [n]$.

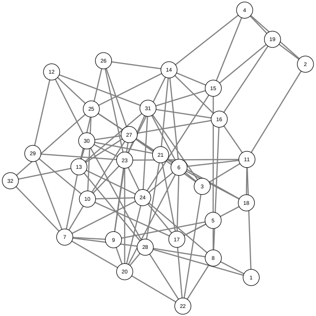
$A \sim \text{BernoulliGraph}(P)$ iff:

1. $P \in [0,1]^{n \times n}$ describes edge probabilities between pairs of vertices.
2. $A_{ij} \overset{\text{ind}}{\sim} \text{Bernoulli}(P_{ij})$ for each $i < j$.

For estimation, we need to impose some structure on P.

Example 1: If every entry $P_{ij} = \theta \in (0,1)$, then $A \sim \text{BernoulliGraph}(P)$ is an Erdos-Renyi graph.

For this model, $A_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$.

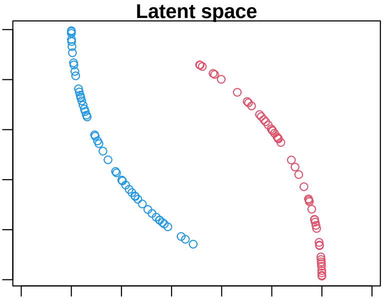


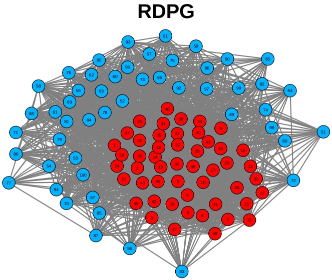
Random Dot Product Graph

Random Dot Product Graph $A \sim \text{RDPG}(X)$
(Young and Scheinerman, 2007)

- Latent vectors $x_1, ..., x_n \in \mathbb{R}^d$ such that $x_i^\top x_j \in [0, 1]$
- $P = XX^\top \in \mathbb{R}^{n \times n}$ where $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^{n \times d}$
- $A \sim \text{BernoulliGraph}(P)$
- Estimation Objectives:
  1. Estimate X from A (assume d is fixed)
  2. If $x_1, ..., x_n \overset{\text{iid}}{\sim} F(\theta)$, estimate θ from A (assuming certain F)
- Non-identifiability: For any $W \in \mathbb{O}(d)$, XW is a latent configuration that produces the same $P = XX^\top = XWW^\top X^\top$

Random Dot Product Graph





Maximum Likelihood Estimation

$$L(X|A) = \prod_{i<j} (x_i^\top x_j)^{A_{ij}} (1 - x_i^\top x_j)^{1-A_{ij}}$$

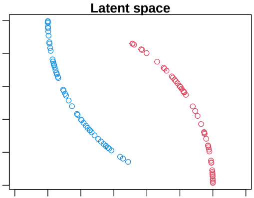
- Intractable
- Not unique

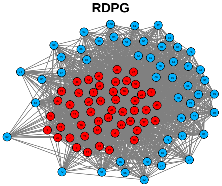
Adjacency Spectral Embedding

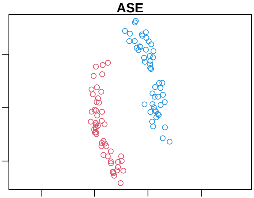
$$\hat{X} = \mathrm{argmin}_{X \in \mathbb{R}^{n \times d}} \|A - XX^\top\|_F^2$$

- Same solution as PCA
- $A = V \Lambda V^\top$
- $\hat{X} = V_d \Lambda_d^{1/2}$
  - $V_d \in \mathbb{R}^{n \times d}$
  - $\Lambda_d \in \mathbb{R}^{d \times d}$
- Not unique
  - If $\hat{X}$ is a solution, then so is $\hat{X}W$
  - Multiplying by W preserves interpoint inner products and distances
- $\max_i \|\hat{x}_i - W x_i\| \stackrel{a.s.}{\to} 0$ (Athreya et al., 2018)

Adjacency Spectral Embedding







Random Dot Product Graph

- For what types of graphs can we justify a latent structure?
- Any $A \sim \text{BernoulliGraph}(P)$ is a RDPG if P is positive semidefinite
- Whether we think of a graph as a RDPG depends on whether the latent structure is useful
- What about P that are not positive semidefinite?

Generalized Random Dot Product Graph

Generlized Random Dot Product Graph $A \sim \text{GRDPG}(X)$ (Rubin-Delanchy et al., 2020)

- Latent vectors $x_1, \ldots, x_n \in \mathbb{R}^{p+q}$ such that $x_i^\top I_{p,q} x_j \in [0, 1]$ and $I_{p,q} = \text{blockdiag}(I_p, -I_q)$
- $P = X I_{p,q} X^\top \in \mathbb{R}^{n \times n}$, $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^{n \times (p+q)}$, $A \sim \text{BernoulliGraph}(P)$
- Non-identifiability: For any $Q \in \mathbb{O}(p, q)$, XQ is a latent configuration that produces the same $P = X I_{p,q} X^\top = X Q I_{p,q} Q^\top X^\top$
    - Multiplication by Q **does not** preserve interpoint inner products or distances
- Any $A \sim \text{BernoulliGraph}(P)$ is a GRDPG

Adjacency Spectral Embedding for GRDPG

$$\hat{X} = \operatorname{argmin}_{X \in \mathbb{R}^{n \times (p+q)}} \|A - X I_{p,q} X^\top\|_F^2$$

- $A = V \Lambda V^\top$
- $\hat{X} = V_{p,q} |\Lambda_{p,q}|^{1/2}$
  - $V_{p,q} = \begin{bmatrix} V_{1:p} & | & V_{n-q+1:n} \end{bmatrix}$
  - $\Lambda_{p,q} = \text{diag}(\lambda_1, ..., \lambda_p, \lambda_{n-q+1}, ..., \lambda_n)$
- Not unique
  - If $\hat{X}$ is a solution, then so is $\hat{X} Q$
- $\max_i \|\hat{x}_i - Q x_i\| \overset{a.s.}{\to} 0$ (Rubin-Delanchy et al., 2020)

Next Time ...

- All $A \sim \text{BernoulliGraph}(P)$ are GRDPGs
    - If P is positive semidefinite, then it is a RDPG
- Includes SBM, DCBM, PABM
- What kind of latent structures correspond to block models?
- Can we use the ASE to estimate block model parameters?