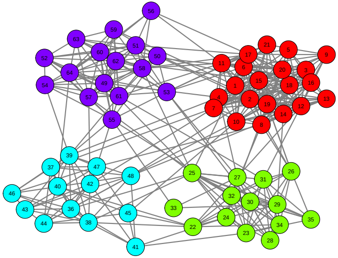
Connecting the Popularity Adjusted Block Model to the Generalized Random Dot Product Graph

SDSS 2021 Lightning Presentation

John Koo, Indiana University

June 2021

The Popularity Adjusted Block Model



Def Popularity Adjusted Block Model (Sengupta and Chen, 2017):

Let each vertex $i \in [n]$ have K popularity parameters $\lambda_{i1}, ..., \lambda_{iK} \in [0, 1]$. Then $A \sim \text{PABM}(P)$ if each $P_{ij} = \lambda_{i z_j} \lambda_{j z_i}$, e.g., if $z_i = k$ and $z_j = l$, $P_{ij} = \lambda_{il} \lambda_{jk}$.

The Popularity Adjusted Block Model

Lemma (Noroozi, Rimal, and Pensky, 2020):

A is sampled from a PABM if P can be described as:

1. Let each $P^{(kl)}$ denote the $n_k \times n_l$ matrix of edge probabilities between communities k and l.
2. Organize popularity parameters as vectors $\lambda^{(kl)} \in \mathbb{R}^{n_k}$ such that $\lambda_i^{(kl)} = \lambda_{k_i l}$ is the popularity parameter of the i^{th} vertex of community k towards community l.
3. Each block can be decomposed as $P^{(kl)} = \lambda^{(kl)} (\lambda^{(lk)})^\top$.

Notation: $A \sim \text{PABM}(\{\lambda^{(kl)}\}_K)$.

The Generalized Random Dot Product Graph

Generalized Random Dot Product Graph $A \sim GRDPG_{p,q}(X)$ (Rubin-Delanchy, Cape, Tang, Priebe, 2020)

- Latent vectors $x_1, ..., x_n \in \mathbb{R}^{p+q}$ such that $x_i^\top I_{p,q} x_j \in [0, 1]$ and $I_{p,q} = blockdiag(I_p, -I_q)$
- $A \sim BernoulliGraph(X I_{p,q} X^\top)$ where $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$

If latent vectors $X_1, ..., X_n \stackrel{iid}{\sim} F$, then we write $(A, X) \sim GRDPG_{p,q}(F, n)$.

Connecting the PABM to the GRDPG

Theorem (KTT): $A \sim PABM(\{\lambda^{(kl)}\}_K)$ is equivalent to $A \sim GRDPG_{p,q}(XU)$ such that

- $p = K(K+1)/2$
- $q = K(K-1)/2$
- U is orthogonal and predetermined for each K
- X is block diagonal and composed of $\{\lambda^{(kl)}\}_K$ $\implies$ if $x_i^\top$ and $x_j^\top$ are two rows of XU corresponding to different communities, then $x_i^\top x_j = 0$.

Remark (non-uniqueness of the latent configuration): $A \sim GRDPG_{p,q}(XU) \implies A \sim GRDPG_{p,q}(XUQ)$ $\forall Q \in \mathbb{O}(p,q)$

Corollary: X is block diagonal by community and U is orthogonal $\implies$ each community corresponds to a subspace in $\mathbb{R}^{K^2}$.

Subspace property holds even with linear transformation $Q \in \mathbb{O}(p,q)$.

Connecting the PABM ot the GRDPG

Theorem (KTT): If $P = V \Lambda V^\top$ is the spectral decomposition of P for the PABM and V has rows $v_i^\top$, then $v_i^\top v_j = 0 \; \forall z_i \neq z_j$.

Theorem (KTT): If $A \approx \hat{V} \hat{\Lambda} \hat{V}^\top$ is the spectral decomposition of A for the PABM using the $K(K+1)/2$ most positive and $K(K-1)/2$ most negative eigenvalues, and we denote $\hat{v}_i^\top$ as the rows of $\hat{V}$, then

$$\max_{i,j : z_i \neq z_j} n \hat{v}_i^\top \hat{v}_j = O_P\Big(\frac{(\log n)^c}{\sqrt{n \rho_n}} \Big)$$

Orthogonal Spectral Clustering (KTT):

1. Let V be the eigenvectors of A corresponding to the $K(K+1)/2$ most positive and $K(K-1)/2$ most negative eigenvalues.
2. Compute $B = |n V V^\top|$ applying $|\cdot|$ entry-wise.
3. Construct graph $\hat{G}$ using B as its similarity matrix.
4. Partition $\hat{G}$ into K disconnected subgraphs.

Simulation Study

Simulation setup:

1. $Z_1, ..., Z_n \overset{iid}{\sim} Categorical(1/K, ..., 1/K)$
2. $\lambda_{ik} \overset{iid}{\sim} Beta(a_{ik}, b_{ik})$
   $$a_{ik} = \begin{cases} 2 & z_i = k \\ 1 & z_i \neq k \end{cases}, \; b_{ik} = \begin{cases} 1 & z_i = k \\ 2 & z_i \neq k \end{cases}$$
3. $P_{ij} = \lambda_{iz_j} \lambda_{jz_i}$
4. $A \sim BernoulliGraph(P)$