Popularity Adjusted Block Models are Generalized Random Dot Product Graphs

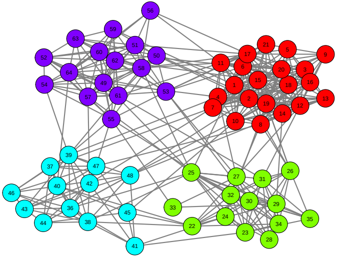
JSM Speed Presentation

August 2022

John Koo,
PhD Student in
Statistical Science,
Indiana University

Minh Tang,
Assistant Professor of
Statistics,
NC State University

Michael Trosset,
Professor of Statistics,
Indiana University



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 $A \sim \text{BernoulliGraph}(P)$ is a PABM if each $P_{ij} = \lambda_{iz_j} \lambda_{jz_i}$

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 $A \sim \text{PABM}(\{\lambda_{ik}\}_K)$ if each $P_{ij} = \lambda_{i z_j} \lambda_{j z_i}$.

Def (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 $P^{(kl)} = \lambda^{(kl)} (\lambda^{(lk)})^\top$.

Generalized Random Dot Product Graph

Def Generalized Random Dot Product Graph
(Rubin-Delanchy, Cape, Tang, Priebe, 2020)

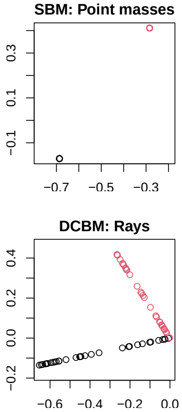
Let $I_{p,q} = \text{blockdiag}(I_p, -I_q)$ and suppose that $x_1, \ldots, x_n \in \mathbb{R}^{p+q}$ are such that $x_i^\top I_{p,q} x_j \in [0, 1]$.
Then $A \sim \text{GRDPG}_{p,q}(X)$ iff $A \sim \text{BernoulliGraph}(X I_{p,q} X^\top)$, where $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$.

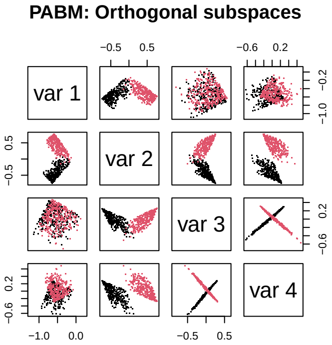
Adjacency Spectral Embedding (Sussman et al., 2012) estimates $x_1, ..., x_n \in \mathbb{R}^{p+q}$ from A:

1. Let $\hat{\Lambda}$ be the diagonal matrix that contains the absolute values of the p most positive and the q most negative eigenvalues.
2. Let $\hat{V}$ be the matrix whose columns are the corresponding eigenvectors.
3. Compute $\hat{X} = \hat{V} \hat{\Lambda}^{1/2}$.

Theorem: $\max_i \|\hat{X}_i - Q_n X_i\| = O_P\left(\frac{(\log n)^c}{n^{1/2}}\right)$ as $n \to \infty$

Connecting Block Models to the (G)RDPG Model





Connecting the PABM to the GRDPG

Theorem (KTT): $A \sim \text{PABM}(\{\lambda_{ik}\}_K)$ is equivalent to $A \sim \text{GRDPG}_{p,q}(XU)$ with

- $p = K(K+1)/2$, $q = K(K-1)/2$;
- U is an orthogonal matrix;
- $X \in \mathbb{R}^{n \times K^2}$ is a block diagonal matrix composed of popularity vectors with each block corresponding to a community.

$$X = \begin{bmatrix} \Lambda^{(1)} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \Lambda^{(K)} \end{bmatrix} \in \mathbb{R}^{n \times K^2}$$

$$\Lambda^{(k)} = \begin{bmatrix} \lambda^{(k1)} & \cdots & \lambda^{(kK)} \end{bmatrix} \in \mathbb{R}^{n_k \times K}$$

$$A \sim \text{PABM}(\{\lambda_{ik}\}_K) \text{ iff } A \sim \text{GRDPG}_{p,q}(XU)$$

Orthogonal Spectral Clustering

Theorem (KTT): If $P = V \Lambda V^\top$ and $B = nVV^\top$, then $B_{ij} = 0$ if $z_i \neq z_j$.

Algorithm: Orthogonal Spectral Clustering:

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 $B = |nVV^\top|$ applying $|\cdot|$ entry-wise.
3. Construct graph G using B as its similarity matrix.
4. Partition G into K disconnected subgraphs.

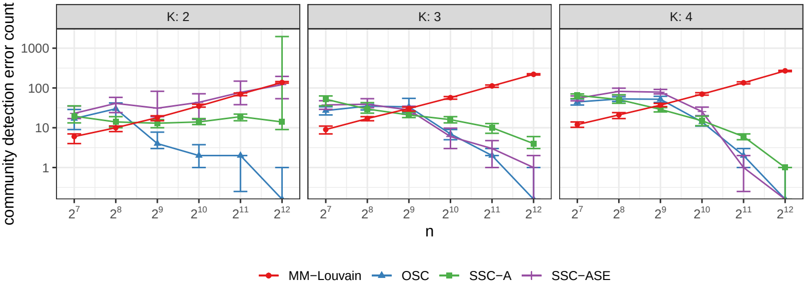
Theorem (KTT): Let $\hat{B}$ with entries $\hat{B}_{ij}$ be the affinity matrix from OSC. Then $\forall$ pairs (i, j) belonging to different communities and sparsity factor satisfying $n\rho_n = \omega((\log n)^{4c})$, $\max_{i,j} \hat{B}_{ij} = O_P\left(\frac{(\log n)^c}{\sqrt{n\rho_n}}\right)$ as $n \to \infty$.

Corollary: OSC results in zero clustering error as $n \to \infty$, with probability 1.

Simulation Results

We compare four algorithms for community detection on randomly generated PABMs:

- Modularity Maximization (Sengupta and Chen) using the Louvain algorithm;
- Orthogonal Spectral Clustering (KTT);
- Sparse Subspace Clustering on the columns of A (Noorozi, Rimal, Pensky);
- Sparse Subspace Clustering on the ASE (KTT).



Popularity Adjusted Block Models are Generalized Random Dot Product Graphs

Published in *Journal of Computational and Graphical Statistics*

arXiv preprint: https://arxiv.org/abs/2109.04010

GitHub repository: https://github.com/johneverettkoo/pabm-grdpg

R package: https://github.com/johneverettkoo/osc