Popularity Adjusted Block Models are Generalized Random Dot Product Graphs

JSM Speed Presentation

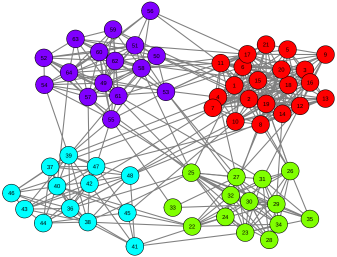
August 2022

John Koo,
PhD Student in Statistics,
Indiana University

Minh Tang,
Assistant Professor of Statistics,
NC State University

Michael W. Trosset,
Professor of Statistics,
Indiana University

Community Detection for Networks

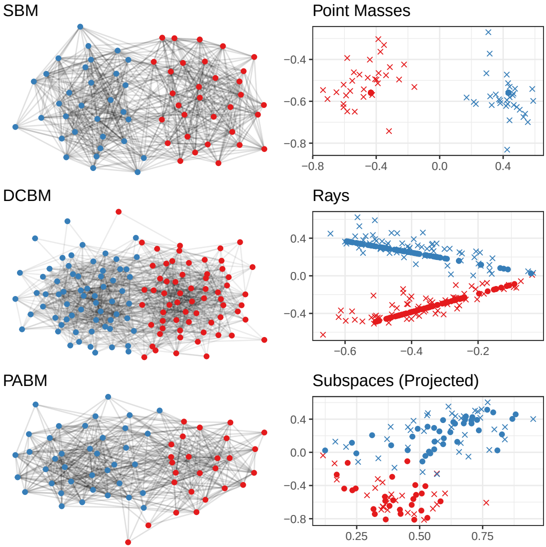


Def Popularity Adjusted Block Model (Sengupta and Chen, 2017):

Let each vertex $i \in [n]$ have K popularity parameters $\lambda_{i1}, ..., \lambda_{iK} \in [0, 1]$.

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

Connecting Block Models to the GRDPG



- K-means clustering
- Gaussian mixture models

- K-means with cosine similarity
- GMM on angles

- ???

Orthogonal Spectral Clustering

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

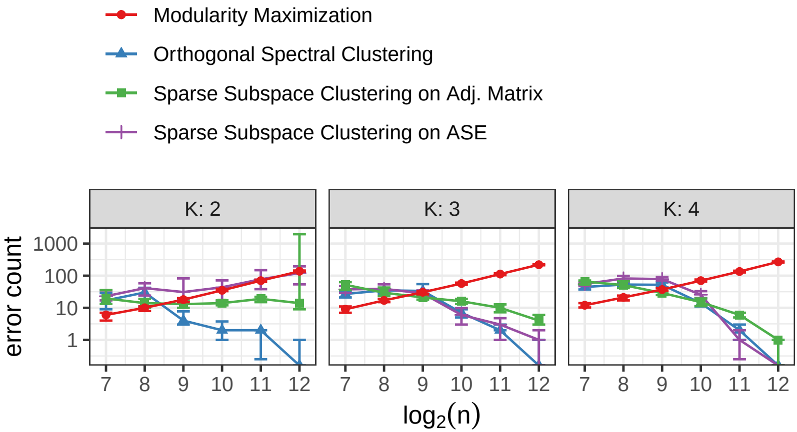
Algorithm: Orthogonal Spectral Clustering:

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): For all pairs (i, j) belonging to different communities, $\max\limits_{i,j} B_{ij} = O_P\left(\frac{(\log n)^c}{\sqrt{n \rho_n}}\right)$.

Corollary: OSC results in zero clustering error almost surely as $n \to \infty$.

Simulation Study



Thank you

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 (WIP): https://github.com/johneverettkoo/osc