Popularity Adjusted Block Models are Generalized Random Dot Product Graphs

Future Leaders Summit 2022 Lightning Presentation

John Koo, Indiana University

April 2022

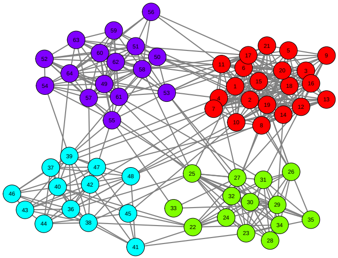
Contributors

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

Community Detection for Networks



How can we cluster the nodes of a network?

Statistical inference (parametric approach):

1. Define a generative model for graph: $G \mid z_1, ..., z_n, \vec{\theta} \sim P(\vec{z}, \vec{\theta})$.
2. Develop a method for obtaining estimators: $f(G) = (\hat{\vec{z}}, \hat{\vec{\theta}})$.
3. Describe asymptotic properties of estimators: $(\hat{\vec{z}}, \hat{\vec{\theta}}) \rightarrow (\vec{z}, \vec{\theta})$.

Overview

1. Block Models and the Popularity Adjusted Block Model
2. Generalized Random Dot Product Graphs
3. Connecting the PABM to the GRDPG
4. Community Detection for the PABM

Block Models

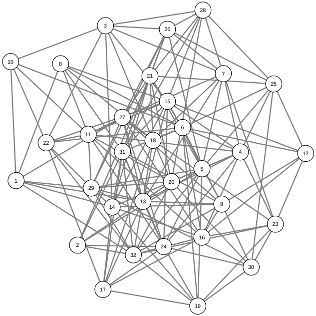
Bernoulli Graphs

Let $G = (V, E)$ be an undirected and unweighted graph with $|V| = n$.
G is described by adjacency matrix A such that $A_{ij} = \begin{cases} 1 & \exists \text{ edge between } i \text{ and } j \\ 0 & \text{else} \end{cases}$
$A_{ji} = A_{ij}$ 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 $i < j$.

Example 1: If every entry $P_{ij} = \theta$, then $A \sim \text{BernoulliGraph}(P)$ is an Erdos-Renyi graph. For this model, $A_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$.

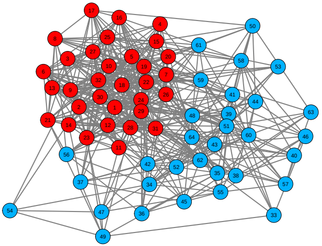


Block Models

Suppose each vertex $v_1, \ldots, v_n$ has labels $z_1, \ldots, z_n \in \{1, \ldots, K\}$,
and each P_{ij} depends on labels z_i and z_j.
Then $A \sim \text{BernoulliGraph}(P)$ is a *block model*.

Example 2: Stochastic Block Model with two communities

- $z_1, \ldots, z_n \in \{1, 2\}$
- $P_{ij} = \begin{cases} p & z_i = z_j = 1 \\ q & z_i = z_j = 2 \\ r & z_i \neq z_j \end{cases}$
- To make this an assortative SBM, set $pq > r^2$.
- In this example, $p = 1/2$, $q = 1/4$, and $r = 1/8$.



Block Models

Erdos-Renyi Model (1959)

- $P_{ij} = \theta$ (not a block model)
- 1 parameter θ

Stochastic Block Model (Lorrain and White, 1971)

- $P_{ij} = \theta_{z_i z_j}$
- $K(K+1)/2$ parameters θ_{kl}

Degree Corrected Block Model (Karrer and Newman, 2011)

- $P_{ij} = \theta_{z_i z_j} \omega_i \omega_j$
- $K(K+1)/2 + n$ parameters θ_{kl}, ω_i

Popularity Adjusted Block Model (Sengupta and Chen, 2017)

- $P_{ij} = \lambda_{i z_j} \lambda_{j z_i}$
- Kn parameters λ_{ik}

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 $z_j = l$, $P_{ij} = \lambda_{il} \lambda_{jk}$.

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

Generalized Random Dot Product Graphs

(Generalized) Random Dot Product Graph Model

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 $x_i^\top x_j \in [0,1]$
- $A \sim \text{BernoulliGraph}(XX^\top)$ where $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$

Generalized Random Dot Product Graph $A \sim \text{GRDPG}_{p,q}(X)$
(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 $I_{p,q} = \text{blockdiag}(I_p, -I_q)$
- $A \sim \text{BernoulliGraph}(XI_{p,q}X^\top)$ where $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$

If latent vectors $X_1, ..., X_n \overset{\text{iid}}{\sim} F$, then we write
$(A, X) \sim \text{RDPG}(F, n)$ or $(A, X) \sim \text{GRDPG}_{p,q}(F, n)$.

(Generalized) Random Dot Product Graph Model

Recovery/Estimation

Want to estimate X from A, or alternatively, interpoint distances, inner products, or angles.

Adjacency Spectral Embedding

To embed in $\mathbb{R}^d$,

1. Compute $A \approx \hat{V}\hat{\Lambda}\hat{V}^\top$ where $\hat{\Lambda} \in \mathbb{R}^{d \times d}$ and $\hat{V} \in \mathbb{R}^{n \times d}$.
   For RDPG, use d greatest eigenvalues; for GRDPG, use p most positive and q most negative eigenvalues.
2. For RDPG, let $\hat{X} = \hat{V}\hat{\Lambda}^{1/2}$; for GRDPG, let $\hat{X} = \hat{V}|\hat{\Lambda}|^{1/2}$.

RDPG: $\max_i \|\hat{X}_i - W_n X_i\| \stackrel{a.s.}{\to} 0$ (Athreya et al., 2018)

GRDPG: $\max_i \|\hat{X}_i - Q_n X_i\| \stackrel{a.s.}{\to} 0$ (Rubin-Delanchy et al., 2020)

Connecting the PABM to the GRDPG

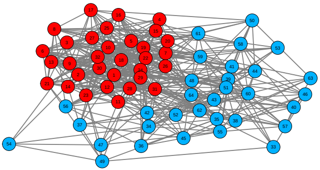
Connecting Block Models to the (G)RDPG Model

All Bernoulli Graphs are RDPG (if P is positive semidefinite) or GRDPG (in general).

Example 2 (cont'd): Assortative SBM ($pq > r^2$) with $K = 2$

$$P_{ij} = \begin{cases} p & z_i = z_j = 1 \\ q & z_i = z_j = 2 \\ r & z_i \neq z_j \end{cases}$$

$$P = \begin{bmatrix} P^{(11)} & P^{(12)} \\ P^{(21)} & P^{(22)} \end{bmatrix} = XX^\top$$

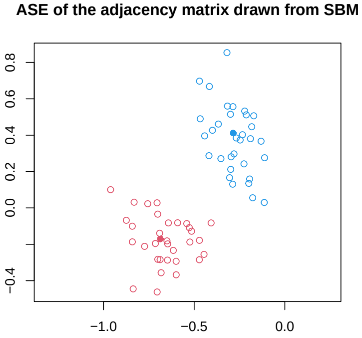


$$X = \begin{bmatrix} \sqrt{p} & 0 \\ \vdots & \vdots \\ \sqrt{p} & 0 \\ \sqrt{r^2/p} & \sqrt{q - r^2/p} \\ \vdots & \vdots \\ \sqrt{r^2/p} & \sqrt{q - r^2/p} \end{bmatrix}$$

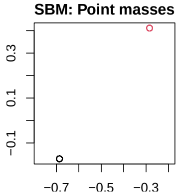
Connecting Block Models to the (G)RDPG Model

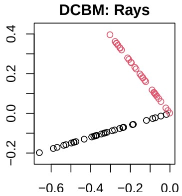
Example 2 (cont'd): If we want to perform community detection,

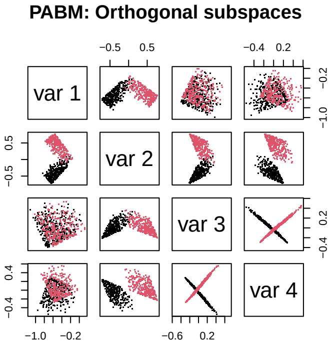
1. Note that A is a RDPG because $P = XX^\top$.
2. Compute the ASE $A \approx \hat{X}\hat{X}^\top$ with $\hat{X} = \hat{V}\hat{\Lambda}^{1/2}$.
3. Apply clustering algorithm (e.g., K-means) to $\hat{X}$,
   noting that as $n \to \infty$, the ASE approaches point masses.



Connecting Block Models to the (G)RDPG Model







Connecting the PABM to the GRDPG

Theorem (KTT): $A \sim$ PABM($\{\lambda^{(kl)}\}_K$) is equivalent to $A \sim$ GRDPG$_{p,q}(XU)$ with

- $p = K(K+1)/2$, $q = K(K-1)/2$
- $U \in \mathbb{O}(K^2)$
- $X \in \mathbb{R}^{n \times K^2}$ is block diagonal and composed of $\{\lambda^{(kl)}\}_K$ with each block corresponding to a community.

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

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

Connecting the PABM to the GRDPG

$$X = \begin{bmatrix} \Lambda^{(1)} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \Lambda^{(K)} \end{bmatrix} \qquad U \in \mathbb{O}(K^2)$$

$$A \sim \text{PABM}(\{\lambda^{(kl)}\}_K) \iff A \sim \text{GRDPG}_{p,q}(XU)$$

Remark 1 (orthogonality of subspaces): If $y_i^\top$ and $y_j^\top$ are two rows of XU corresponding to different communities, then $y_i^\top y_j = 0$.

Remark 2 (non-uniqueness of the latent configuration): If $A \sim \text{GRDPG}_{p,q}(Y)$, then $A \sim \text{GRDPG}_{p,q}(YQ)$ for any Q in the indefinite orthogonal group with signature p, q.

Remark 3: Communities correspond to subspaces even with linear transformation $Q \in \mathbb{O}(p, q)$, but this may break the orthogonality property.

Community Detection for the PABM

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 $\hat{B}_n$ with entries $\hat{B}_n^{(ij)}$ be the affinity matrix from OSC. Then $\forall$ pairs (i, j) belonging to different communities and sparsity factor satisfying $n\rho_n = \omega((\log n)^{4c})$,

$$\max_{i,j} \hat{B}_n^{(ij)} = O_P\Big(\frac{(\log n)^c}{\sqrt{n\rho_n}}\Big)$$

Sparse Subspace Clustering

Corollary: The ASE of $A \sim \text{PABM}(\{\lambda^{(kl)}\}_K)$ lies near a collection of K-dimensional subspaces in K^2 dimensions.

Algorithm: Sparse Subspace Clustering (Elhamifar & Vidal, 2009):

1. Solve n optimization problems $c_i = \arg\min_c \|c\|_1$ subject to $x_i = X^\top c$ and $c^{(i)} = 0$.
   This is typically performed via LASSO:
   $c_i = \arg\min \frac{1}{2}\|x_i - X_{-i}^\top c\|_2^2 + \lambda\|c\|_1$
2. Compile solutions $C = \begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix}$.
3. Construct affinity matrix $B = |C| + |C^\top|$.

Sparse Subspace Clustering

Theorem (KTT):

Let

- P_n describe the edge probability matrix of the PABM with n vertices, and $A_n \sim \text{BernoulliGraph}(P_n)$;
- $\hat{V}_n$ be the matrix of eigenvectors of A_n corresponding to the $K(K+1)/2$ most positive and $K(K-1)/2$ most negative eigenvalues.

Then

- For some $\lambda > 0$ and $N < \infty$, $\sqrt{n}\hat{V}_n$ obeys the Subspace Detection Property with probability 1 when $n > N$.

Remarks:

- For large n, we can identify λ for SDP (Wang and Xu, 2016).
- SDP does not guarantee community detection.

Simulation Results