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

SDSS 2021 Lightning Presentation

John Koo, Indiana University

June 2021

Contributors

John Koo,
PhD Student in
Statistical Science,
Indiana University

Minh Tang,
Assistant Professor of
Statistics,
NC State University

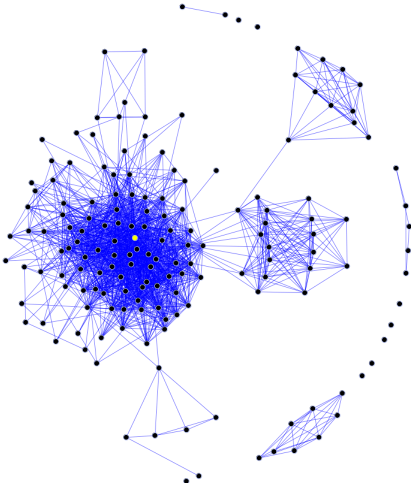
Michael Trosset,
Professor of Statistics,
Indiana University

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

Networks



Screenshot taken by User:DarwinPeacock - Screenshot of free software GUESS, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=6057981

Bernoulli Graphs

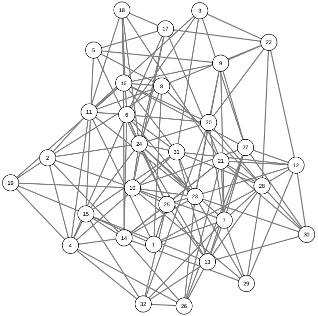
Let $G = (V, E)$ be an undirected and unweighted graph with $|V| = n$.

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

$A \sim BernoulliGraph(P)$ iff:

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

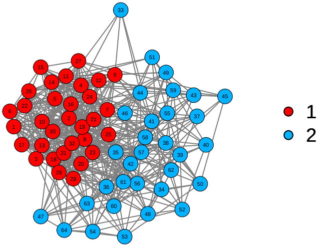


Block Models

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

Example: Stochastic Block Model (Lorrain and White, 1971) with two communities

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



Popularity Adjusted Block Model

Definition based on Noroozi, Rimal, and Pensky (2020); model first proposed by Sengupta and Chen (2017).

$A \sim PABM(\{\lambda^{(kl)}\}_K)$ iff

1. w.l.o.g., organize P such that each block $P^{(kl)} \in [0, 1]^{n_k \times n_l}$ contains edge probabilities between communities k and l.
2. Organize parameters as vectors such that $\lambda^{(kl)} \in \mathbb{R}^{n_k}$ are the popularity parameters of members of community k to community l.
   $\{\lambda^{(kl)}\}_K$ is the set of K^2 popularity vectors.
3. Then we can write each block of P as $P^{(kl)} = \lambda^{(kl)}(\lambda^{(lk)})^\top$.
4. Sample $A \sim BernoulliGraph(P)$.

Example: $K = 2$

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

Community Detection in Block Models

Likelihood

$$L = \prod_{i<j} \prod_{k,l}^{K} \left(p_{k,l,i,j}^{A_{ij}} (1 - p_{k,l,i,j})^{1-A_{ij}} \right)^{z_{ik} z_{jl}}$$

- ML method for community detection: $\hat{\vec{z}} = \arg\max_{\vec{z}} L$
- NP-complete
  - Expectation-Maximization
  - Bayesian methods
  - **Spectral methods**

Generalized Random Dot Product Graphs

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 $(A, X) \sim GRDPG_{p,q}(F, n)$.

(Generalized) Random Dot Product Graph Model

Recovery/Estimation

Want to estimate X given A.

Adjacency Spectral Embedding

To embed in $\mathbb{R}^{p+q}$,

1. Compute $A \approx \hat{V}\hat{\Lambda}\hat{V}^\top$ where $\hat{\Lambda} \in \mathbb{R}^{(p+q)\times(p+q)}$ and $\hat{V} \in \mathbb{R}^{n\times(p+q)}$ by using p most positive and q most negative eigenvalues.
2. Let $\hat{X} = \hat{V}|\hat{\Lambda}|^{1/2}$.

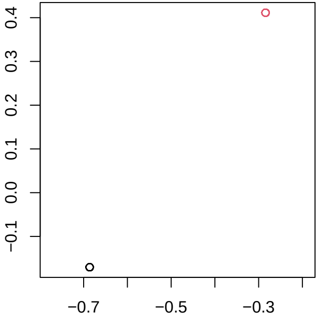
Rubin-Delanchy et al., 2020:

$$\max_i \|\hat{X}_i - Q_n X_i\| \overset{a.s.}{\to} 0$$

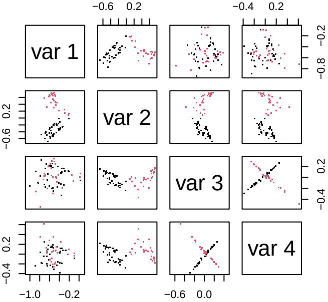
$$Q_n \in \mathbb{O}(p, q)$$

Connecting Block Models to the GRDPG

SBM: Point masses



PABM: Orthogonal subspaces



Connecting the PABM to the GRDPG

Connecting the PABM to the GRDPG ($K = 2$)

Theorem (KTT): $A \sim PABM(\{\lambda^{(kl)}\}_2)$ is equivalent to $A \sim GRDPG_{3,1}(XU)$ for block diagonal X constructed from $\{\lambda^{(kl)}\}_2$ and predetermined $U \in \mathbb{O}(4)$.

Proof: Decompose P as follows

$$X = \begin{bmatrix} \lambda^{(11)} & \lambda^{(12)} & 0 & 0 \\ 0 & 0 & \lambda^{(21)} & \lambda^{(22)} \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$P = (XU) I_{3,1} (XU)^\top = \begin{bmatrix} \lambda^{(11)} (\lambda^{(11)})^\top & \lambda^{(12)} (\lambda^{(21)})^\top \\ \lambda^{(21)} (\lambda^{(12)})^\top & \lambda^{(22)} (\lambda^{(22)})^\top \end{bmatrix}$$

Connecting the PABM to the GRDPG ($K \geq 2$)

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 $Q \in \mathbb{O}(p,q)$.

Community Detection for the PABM

Sparse Subspace Clustering

Sparse Subspace Clustering (Elhamifar and Vidal, 2009):
Let the rows of $X \in \mathbb{R}^{n \times d}$ are vectors that lie on or near a small number of low-dimensional subspaces. SSC constructs a graph from X as follows:

1. For each $i \in [n]$, solve the optimization problem $c_i = \arg\min_c \|c\|_1$ subject to $x_i = Xc$ and $c^{(i)} = 0$.
2. Construct a similarity graph using $B = |C| + |C^\top|$ as its affinity matrix, where $C = \begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix}$.

If X obeys the *Subspace Detection Property*, then it produces a similarity graph B such that $B_{ij} = 0 \; \forall z_i \neq z_j$.

Sparse Subspace Clustering for Community Detection

Noroozi et al. observed that the rank of P is K^2 and the columns of P belonging to each community has rank K to justify SSC for the PABM.

$$c_i = \arg\min_c \|c\|_1 \text{ subject to } A_{\cdot,i} = Ac \text{ and } c^{(i)} = 0$$

GRDPG-based approach: Apply SSC to the ASE of A.

$$c_i = \arg\min_c \|c\|_1 \text{ subject to } \hat{x}_i = \hat{X}c \text{ and } c^{(i)} = 0$$

$$A \approx \hat{X} I_{p,q} \hat{X}^\top$$

Sparse Subspace Clustering

Theorem (KTT):

Let

- P_n describe the edge probability matrix of the PABM with n vertices, and $A_n \sim BernoulliGraph(P_n)$.
- $\hat{V}_n$ 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

- $\exists N < \infty$ such that when $n > N$, $\sqrt{n}\hat{V}_n$ obeys the Subspace Detection Property with probability 1.

Orthogonal Spectral Clustering

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

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

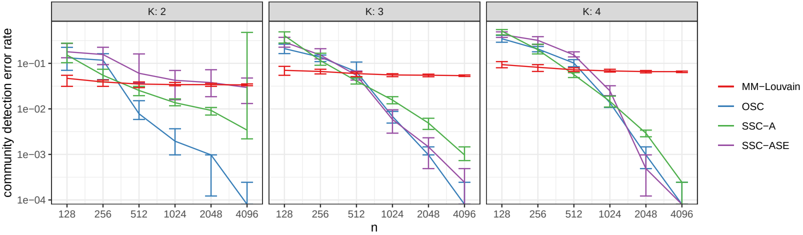
Orthogonal Spectral Clustering (KTT):

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. $A \sim BernoulliGraph(P)$



Conclusion

1. The PABM is a recently developed flexible block model that can be used to describe many graphs with community structure.
2. Likelihood maximization is NP-complete for block models, so we need to take a different approach to community detection.
3. The GRDPG, which describes all block models, motivates a spectral approach to statistical inference on graphs.
4. Under the GRDPG framework, the PABM is induced by a latent configuration in $\mathbb{R}^{K^2}$ consisting of K K-dimensional subspaces that are all orthogonal to each other.
5. The latent configuration of the PABM under the GRDPG framework leads to two methods for community detection, both of which have nice theoretical asymptotic properties.

Thank you

Code and draft of the paper available at
https://github.com/johneverettkoo/pabm-grdpg

R package coming soon at
https://github.com/johneverettkoo/pabm