Summary of Dissertation Work Thus Far and Future Steps

John Koo

Introduction and Research Goal/Scope

Statistical analysis on graphs or networks often involves the partitioning of a graph into disconnected subgraphs or clusters, i.e., finding some function $f : [n] \mapsto [K]$. This is often motivated by the assumption that there exist underlying and unobserved communities to which each vertex of the graph belongs, and edges between pairs of vertices are determined by drawing from a probability distribution based on the community relationships between each pair. The goal of the partitioning or clustering then is community detection, or the recovery of the true underlying community labels for each vertex, up to permutation (with some additional parameter estimation being of possible interest), assuming some underlying probability model.

One example of such a model is the Stochastic Block Model (SBM) with two communities. In this case, we may draw edges between communities independently such that the distribution of each edge within community 1 is $Bernoulli(p)$, the distribution of each edge within community 2 is $Bernoulli(q)$, and the distribution of each edge between communities 1 and 2 is $Bernoulli(r)$. Other models have been proposed, such as the Degree-Corrected Block Model (DCBM), Mixed Membership Stochastic Block Model (MMSBM), and Popularity Adjusted Block Model (PABM), which are often considered generalizations of the SBM. All such models involve undirected, unweighted, hollow graphs (although there are some straightforward generalizations to other types of graphs, in particular to weighted graphs). The information in this graph is stored in adjacency $A \in \{0, 1\}^{n \times n}$ where $A_{ij} = 1$ if there is an edge between vertices i and j, and 0 otherwise.

The underlying similarity among all of these Bernoulli edge graphs is the edge probability matrix, $P \in [0, 1]^{n \times n}$. Each element P_{ij} is the probability of the existence of an edge between vertices i and j (P is then symmetric and $P_{ii} = 0 \ \forall i$). The adjacency matrix A is then drawn such that $A_{ij} \overset{indep}{\sim} P_{ij}$ for each $i < j$ and $A_{ji} = A_{ij}$. For instance, in the SBM example from before, $P_{ij} = p$ if vertices i and j both belong to community 1, $P_{ij} = q$ if they both belong to community 2, and $P_{ij} = r$ if one is in community 1 and the other is in community 2.

The Popularity Adjusted Block Model

Definition 1 (Popularity Adjusted Block Model)**.** *Let $P \in [0, 1]^{n \times n}$ be a symmetric edge probability matrix for a set of n vertices, V. Each vertex has a community label $1, ..., K$, and the rows and columns of P are arranged by community label such that $n_k \times n_l$ block $P^{(kl)}$ describes the edge probabilities between vertices in communities k and l ($P^{(lk)} = (P^{(kl)})^\top$). Let graph $G = (V, E)$ be an undirected, unweighted graph such that its corresponding adjacency matrix $A \in \{0, 1\}^{n \times n}$ is a realization of Bernoulli(P), i.e., $A_{ij} \overset{indep}{\sim}$ Bernoulli(P_{ij}) for $i > j$ ($A_{ij} = A_{ji}$ and $A_{ii} = 0$).*

If each block $P^{(kl)}$ can be written as the outer product of two vectors:

$$P^{(kl)} = \lambda^{(kl)} (\lambda^{(lk)})^\top \tag{1}$$

for a set of K^2 fixed vectors $\{\lambda^{(st)}\}_{s,t=1}^K$ where each $\lambda^{(st)}$ is a column vector of dimension n_s, then graph G and its corresponding adjacency matrix A is a realization of a popularity adjusted block model with parameters $\{\lambda^{(st)}\}_{s,t=1}^K$.

We will use the notation $A \sim PABM(\{\lambda^{(kl)}\}_K)$ to denote a random adjacency matrix A drawn from a PABM with parameters $\lambda^{(kl)}$ consisting of K underlying communities.

The Generalized Random Dot Product Graph

The Random Dot Product Graph (RDPG) is another graph model with Bernoulli edge probabilities. Under this model, each vertex of the graph can be represented by a point in some latent space such that the edge probability between any pair of vertices is given by their corresponding dot product in the latent space. It can be shown that any Bernoulli edge graph with positive semidefinite edge probability matrix P can be represented as a RDPG. More specifically, if $P = V \Sigma V^\top$ is the spectral decomposition of P such that Σ is a diagonal matrix of d positive eigenvalues, then $V \Sigma^{1/2}$ is one representation of the latent positions under the RDPG framework.

Example (Connecting the SBM to the RDPG)**.** *Consider the same SBM as before. Let community 1 have n_1 vertices and community 2 have n_2 vertices such that $n_1 + n_2 = n$. Let the edge probability matrix P be organized by community such that the $P^{(kl)}$ block represents edge probabilities between communities k and l. In this case, $P = \begin{bmatrix} P^{(11)} & P^{(12)} \\ P^{(21)} & P^{(22)} \end{bmatrix}$ where each block is a constant value, e.g., $P^{(11)}_{ij} = p$. Then one RDPG representation of this SBM is:*

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

where the first n_1 rows are $\begin{bmatrix} \sqrt{p} & 0 \end{bmatrix}$ and the next n_2 rows are $\begin{bmatrix} \sqrt{r^2/p} & \sqrt{q - r^2/p} \end{bmatrix}$. Then it can be shown that

$$P = XX^\top$$

If P is not positive semidefinite, then it cannot be the edge probability matrix of a RDPG, but it can be the edge probability matrix of a *Generalized* Random Dot Product Graph (GRDPG).[1]

Definition 2 (Generalized Random Dot Product Graph)**.** *Let $P \in [0,1]^{n \times n}$ be a symmetric edge probability matrix for a set of n vertices, V. If $\exists X \in \mathbb{R}^{n \times d}$ such that*

$$P = X I_{pq} X^\top \tag{2}$$

for some $d, p, q \in \mathbb{N}$ and $p + q = d$, then graph $G = (V, E)$ with adjacency matrix A such that $A_{ij} \overset{indep}{\sim}$ Bernoulli(P_{ij}) for $i > j$ ($A_{ij} = A_{ji}$ and $A_{ii} = 0$) is a draw from the generalized random dot product graph model with latent positions X and signature (p, q). More precisely, if vertices i and j have latent positions x_i and x_j respectively, then the edge probability between the two is $P_{ij} = x_i^\top I_{pq} x_j$, and X contains the latent positions as rows $x_i^\top$.

We will use the notation $A \sim GRDPG_{p,q}(X)$ to denote a random adjacency matrix A drawn from latent positions X and signature (p, q).

Definition 3 (Indefinite Orthogonal Group)**.** *The indefinite orthogonal group with signature (p, q) is the set $\{Q \in \mathbb{R}^{d \times d} : Q I_{pq} Q^\top = I_{pq}\}$, denoted as $\mathbb{O}(p, q)$.*

Remark. Like the RDPG, the latent positions of a GRDPG are not unique. More specifically, if $P_{ij} = x_i^\top I_{pq} x_j$, then we also have for any $Q \in \mathbb{O}(p, q)$, $(Q x_i)^\top I_{pq} (Q x_j) = x_i^\top (Q^\top I_{pq} Q) x_j = x_i^\top I_{pq} x_j = P_{ij}$. Unlike in the RDPG case, transforming the latent positions via multiplication by $Q \in \mathbb{O}(p, q)$ does not necessarily maintain interpoint angles or distances.

[1] It is trivial to show that *all* Bernoulli graphs are GRDPGs.

Literature Review and Previous Work

The PABM was introduced by Sengupta and Chen [4], who intuited that a spectral clustering type of approach would be appropriate for this problem. Their work was furthered by Noroozi et al. [2], who used SSC for community detection. SSC is typically performed on points in Euclidean space, which suggests a spectral embedding, but the authors of this paper instead performed SSC on P and A directly. They were able to show that SSC on P results in perfect community detection.

SSC is described as follows. Given $x_1, ..., x_n \in \mathbb{R}^d$ such that each point lies on one of K subspaces, the goal is to determine which subspace each point belongs to. We might also want to add some noise to the points so that they don't lie exactly on their respective subspaces.

SSC is performed by solving an optimization problem for each observed point. Given $X \in \mathbb{R}^{n \times d}$ with vectors $x_i^\top \in \mathbb{R}^d$ as rows of X, the optimization problem $\min_{c_i} ||c_i||_1$ subject to $x_i = Xc_i$ and $\beta_i = 0$ is solved for each $i = 1, ..., n$. The solutions are collected collected into matrix $C = \begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix}^\top$ to construct an affinity matrix $B = |C| + |C^\top|$. If each x_i lie perfectly on one of K subspaces, B describes an undirected graph consisting of K disjoint subgraphs, i.e., $B_{ij} = 0$ if x_i, x_j are in different subspaces. If X instead represents points near K subspaces with some noise, a final graph partitioning step is performed (e.g., edge thresholding or spectral clustering).

In practice, SSC is often performed by solving the LASSO problems

$$c_i = \arg\min_c \frac{1}{2} ||x_i - X_{-i}c||_2^2 + \lambda ||c||_1 \tag{3}$$

for some sparsity parameter $\lambda > 0$. The c_i vectors are then collected into C and B as before.

Definition 4 (Subspace Detection Property)**.** *Let $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$ be noisy points sampled from K subspaces. Let C and B be constructed from the solutions of LASSO problems as described in (3). If each column of C has nonzero norm and $B_{ij} = 0 \; \forall \; x_i$ and x_j sampled from different subspaces, then X obeys the subspace detection property.*

Wang and Xu [6] showed that in the case where the subspaces are orthogonal (or at least the cosine of their angles is very small) and the points are close enough to their respective subspaces, the subspace detection property will hold given an appropriate choice of λ. It turns out that this is precisely the case for a particular embedding choice for the PABM, of which the properties have been described by Rubin-Delanchy et al. [3].

Completed Work Thus Far

The PABM can be expressed as a GRDPG. In particular, there is one expression of this that results in a very simple structure such that each community corresponds to a subspace and the subspaces are mutually orthogonal.

Theorem 1 (Connecting the PABM to the GRDPG for $K = 2$)**.** *Let*

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

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

where each $\lambda^{(kl)}$ is a vector as in Definition 1. Then $A \sim GRDPG_{3,1}(XU)$ and $A \sim PABM(\{(\lambda^{(kl)}\}_2)$ are equivalent.

Theorem 2 (Generalization to $K > 2$)**.** *There exists a block diagonal matrix $X \in \mathbb{R}^{n \times K^2}$ defined by PABM parameters $\{\lambda^{(kl)}\}_K$ and orthonormal matrix $U \in \mathbb{R}^{K^2 \times K^2}$ that is known for each K such that $A \sim GRDPG_{K(K+1)/2, K(K-1)/2}(XU)$ and $A \sim PABM(\{(\lambda^{(kl)}\})_K)$ are equivalent.*

Proof. Define the following matrices from $\{\lambda^{(kl)}\}_K$:

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

$$X = \text{blockdiag}(\Lambda^{(1)}, ..., \Lambda^{(K)}) \in \mathbb{R}^{n \times K^2} \tag{4}$$

$$L^{(k)} = \text{blockdiag}(\lambda^{(1k)}, ..., \lambda^{(Kk)}) \in \mathbb{R}^{n \times K}$$

$$Y = \begin{bmatrix} L^{(1)} & \cdots & L^{(K)} \end{bmatrix} \in \mathbb{R}^{n \times K^2}$$

Then $P = XY^\top$.

Similar to the $K = 2$ case, we have $Y = X\Pi$ for a permutation matrix Π, resulting in $P = X\Pi X^\top$. The permutation described by Π has K fixed points, which correspond to K eigenvalues equal to 1 with corresponding eigenvectors e_k where $k = r(K+1)+1$ for $r = 0, ..., K-1$. It also has $\binom{K}{2} = K(K-1)/2$ cycles of order 2. Each cycle corresponds to a pair of eigenvalues $+1$ and -1 and a pair of eigenvectors $(e_s + e_t)/\sqrt{2}$ and $(e_s - e_t)/\sqrt{2}$.

Then Π has $K(K+1)/2$ eigenvalues equal to 1 and $K(K-1)/2$ eigenvalues equal to -1. Π has the decomposed form

$$\Pi = U I_{K(K+1)/2, K(K-1)/2} U^\top \tag{5}$$

The edge probability matrix then can be written as:

$$P = XUI_{p,q}(XU)^\top \tag{6}$$

$$p = K(K+1)/2 \tag{7}$$

$$q = K(K-1)/2 \tag{8}$$

and we can describe the PABM with K communities as a GRDPG with latent positions XU with signature $\big(K(K+1)/2, K(K-1)/2\big)$. □

Example ($K = 3$)**.** *Using the same notation as in Theorem 2:*

$$X = \begin{bmatrix} \lambda^{(11)} & \lambda^{(12)} & \lambda^{(13)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda^{(21)} & \lambda^{(22)} & \lambda^{(23)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda^{(31)} & \lambda^{(32)} & \lambda^{(33)} \end{bmatrix}$$

$$Y = \begin{bmatrix} \lambda^{(11)} & 0 & 0 & \lambda^{(12)} & 0 & 0 & \lambda^{(13)} & 0 & 0 \\ 0 & \lambda^{(21)} & 0 & 0 & \lambda^{(22)} & 0 & 0 & \lambda^{(23)} & 0 \\ 0 & 0 & \lambda^{(31)} & 0 & 0 & \lambda^{(32)} & 0 & 0 & \lambda^{(33)} \end{bmatrix}$$

Algorithm 1: Orthogonal Spectral Clustering.

```
Data: Adjacency matrix $A$, number of communities $K$
Result: Community assignments $1, ..., K$
Compute the eigenvectors of $A$ that correspond to the $K(K+1)/2$ most positive eigenvalues and
  $K(K-1)/2$ most negative eigenvalues. Construct $V$ using these eigenvectors as its columns.
Compute $B = |nVV^\top|$, applying $|\cdot|$ entry-wise.
Construct graph $G$ using $B$ as its similarity matrix.
Partition $G$ into $K$ disconnected subgraphs (e.g., using edge thresholding or spectral clustering).
Map each partition to the community labels $1, ..., K$.
```

Then $P = XY^\top$ and $Y = X\Pi$ where Π is a permutation matrix consisting of 3 fixed points and 3 cycles of order 2:

$$\Pi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

** Positions 1, 5, 9 are fixed.*

** The cycles of order 2 are $(2,4)$, $(3,7)$, and $(6,8)$.*

Therefore, we can decompose $\Pi = UI_{6,3}U^\top$ where the first three columns of U consist of e_1, e_5, and e_9 corresponding to the fixed positions 1, 5, and 9, the next three columns consist of eigenvectors $(e_k + e_l)/\sqrt{2}$, and the last three columns consist of eigenvectors $(e_k - e_l)/\sqrt{2}$, where pairs (k,l) correspond to the cycles of order 2 described above.

The latent positions are the rows of

$$XU = \begin{bmatrix} \lambda^{(11)} & 0 & 0 & \lambda^{(12)}/\sqrt{2} & \lambda^{(13)}/\sqrt{2} & 0 & \lambda^{(12)}/\sqrt{2} & \lambda^{(13)}/\sqrt{2} & 0 \\ 0 & \lambda^{(22)} & 0 & \lambda^{(21)}/\sqrt{2} & 0 & \lambda^{(23)}/\sqrt{2} & -\lambda^{(21)}/\sqrt{2} & 0 & \lambda^{(23)}/\sqrt{2} \\ 0 & 0 & \lambda^{(33)} & 0 & \lambda^{(31)}/\sqrt{2} & \lambda^{(32)}/\sqrt{2} & 0 & -\lambda^{(31)}/\sqrt{2} & -\lambda^{(32)}/\sqrt{2} \end{bmatrix}$$

This connection leads to the following property:

Theorem 3. *Let $P = VDV^\top$ be the spectral decomposition of the edge probability matrix. Let $B = nVV^\top$. Then $B_{ij} = 0$ if vertices i and j are from different communities.*

In practice, however, we observe A rather than P. Using the asymptotic properties of the Adjacency Spectral Embedding (ASE) of the GRDPG [3], it can be shown that constructing $\hat{B}_n$ from the eigenvectors of A will result in entries of $\hat{B}_n$ corresponding to between-community pairings going to 0 with probability 1.

Theorem 4. *Let $\hat{B}_n$ with entries $\hat{B}_n^{(ij)}$ be the affinity matrix from OSC (Alg. 1). Then $\forall$ pairs (i,j) belonging to different communities and sparsity factor satisfying $n\rho_n = \omega\{(\log n)^{4c}\}$,*

$$\max_{i,j} |n(\hat{v}_n^{(i)})^\top \hat{v}_n^{(j)}| = O_P\Big(\frac{(\log n)^c}{\sqrt{n\rho_n}}\Big) \tag{9}$$

This provides the result that for i,j in different communities, $\hat{B}_n^{(ij)} \overset{a.s.}{\to} 0$.

Algorithm 2: Sparse Subspace Clustering for the PABM using LASSO.

```
Data: Adjacency matrix $A$, number of communities $K$, hyperparameter $\lambda$
Result: Community assignments $1, ..., K$
Find $V$, the matrix of eigenvectors of $A$ corresponding to the $K(K+1)/2$ most positive and the $K(K-1)/2$ most negative eigenvalues.
for $i = 1, ..., n$ do
    Assign $v_i^\top$ as the $i^{th}$ row of $V$. Assign $V_{-i} = \begin{bmatrix} v_1 & \cdots & v_{i-1} & v_{i+1} & \cdots & v_n \end{bmatrix}^\top$.
    Solve the LASSO problem $c_i = \arg\min_\beta \frac{1}{2}||\sqrt{n} v_i - \sqrt{n} V_{-i} \beta||_2^2 + \lambda ||\beta||_1$.
    Assign $\tilde{c}_i = \begin{bmatrix} c_i^{(1)} & \cdots & c_i^{(i-1)} & 0 & c_i^{(i)} & \cdots & c_i^{(n-1)} \end{bmatrix}^\top$ such that the superscript is the index of $\tilde{c}_i$.
end
Assign $C = \begin{bmatrix} \tilde{c}_1 & \cdots & \tilde{c}_n \end{bmatrix}$.
Compute the affinity matrix $B = |C| + |C^\top|$.
Construct graph $G$ using $B$ as its similarity matrix.
Partition $G$ into $K$ disconnected subgraphs (e.g., using edge thresholding or spectral clustering).
Map each partition to the community labels $1, ..., K$.
```

Theorems 2, 3, and 4 also provide a very natural path toward using SSC for community detection for the PABM. We established in Theorem 2 that an ASE of the edge probability matrix P can be constructed such that the communities lie on mutually orthogonal subspaces, and this property can be recovered from the eigenvectors of P. Then Theorems 3 and 4 show that this property holds for the unscaled ASE of A drawn from P as $n \to \infty$.

Theorem 5. *Let P_n describe the edge probability matrix of the PABM with n vertices, and let $A_n \sim$ Bernoulli(P_n). Let $\hat{V}_n$ 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 $\exists \lambda > 0$ such that as $n \to \infty$, $\hat{V}_n$ obeys the subspace detection property with probability 1.*

We are also able to show convergence properties for parameter estimation using the ASE, the details of which have been omitted for brevity but can be found in the main paper.

Future Work

So far, we've shown that there is a very convenient GRDPG representation of the PABM that makes community detection almost trivially simple. More specifically, we showed that an embedding of the PABM results in orthogonal subspaces and noise that converges to 0 with probability 1, which makes for a very straightforward application of SSC. It can also be shown that this type of embedding has the same properties for the SBM, DCBM, and Nested Block Model (NBM), which are really just special cases of the PABM (see Noroozi and Pensky [1], in particular Fig 2.).

Furthermore, all Bernoulli graphs can be represented as a GRDPG, and it may be the case that other, more complicated graphs have nice ASE representations that lead to community detection methods. In particular, we are currently exploring the nature of RDPGs and GRDPGs that result from latent positions for which each community lies on a manifold rather than on a subspace. Much of this work is based on work by Trosset et al. [5] who considered the RDPG with latent positions that lie on a manifold and developed methods for recovering the manifold from an adjacency matrix sampled from it.

References

[1] Majid Noroozi and Marianna Pensky. The hierarchy of block models, 2021.

[2] Majid Noroozi, Ramchandra Rimal, and Marianna Pensky. Estimation and clustering in popularity adjusted block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, n/a(n/a). doi: https://doi.org/10.1111/rssb.12410. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12410.

[3] Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey E. Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph, 2017.

[4] Srijan Sengupta and Yuguo Chen. A block model for node popularity in networks with community structure. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 80(2):365–386, March 2018. ISSN 1369-7412. doi: 10.1111/rssb.12245.

[5] Michael W. Trosset, Mingyue Gao, Minh Tang, and Carey E. Priebe. Learning 1-dimensional submanifolds for subsequent inference on random dot product graphs, 2020.

[6] Yu-Xiang Wang and Huan Xu. Noisy sparse subspace clustering. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 89–97, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL http://proceedings.mlr.press/v28/wang13.html.