

# Community Detection Methods for the Generalized Random Dot Product Graph Model

## Dissertation Proposal

John Koo

June 2021

# Community Detection for Networks

How can we cluster the nodes of a network?

Statistical inference (parametric approach):

1. Define a generative model  $G \mid z_1, \dots, z_n, \vec{\theta} \sim P(\vec{z}, \vec{\theta})$ .
2. Develop a method for obtaining estimators  $f(G) = \hat{z}_1, \dots, \hat{z}_n$ .
3. Identify asymptotic properties of estimators and prove consistency.

# Overview

1. Probability Models for Networks
  - Block Models and Community Structure
  - (Generalized) Random Dot Product Graphs
  - Connecting Block Models to the (G)RDGP
2. Popularity Adjusted Block Model
  - Connecting the PABM to the GRDGP
  - Subspace Clustering for Community Detection
  - Orthogonal Spectral Clustering
3. Community Detection for the (G)RDGP
  - Manifold Clustering
  - Manifolds as (G)RDGP Latent Configurations

# Probability Models for Networks

# Bernoulli Graphs

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

$G$  is described by adjacency matrix  $A$  such

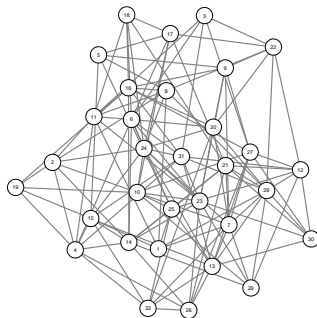
that  $A_{ij} = \begin{cases} 1 & \exists \text{ edge between } i \text{ and } j \\ 0 & \text{else} \end{cases}$

$A_{ji} = A_{ij}$  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.  $A_{ij} \stackrel{\text{ind}}{\sim} \text{Bernoulli}(P_{ij})$  for each  $i < j$ .

Example: If  $G$  is an Erdos-Renyi graph, then  $P_{ij} = \theta$ .



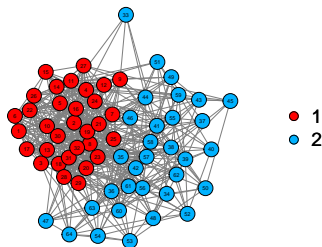
# Block Models

Suppose each vertex  $v_1, \dots, v_n$  has hidden labels  $z_1, \dots, z_n \in [K]$ , and each  $P_{ij}$  depends on labels  $z_i$  and  $z_j$ .

Then  $A \sim \text{BernoulliGraph}(P)$  is a *block model*.

Example 1: Stochastic Block Model with two communities

- $z_1, \dots, z_n \in \{1, 2\}$
- $$P_{ij} = \begin{cases} p & z_i = z_j = 1 \\ q & z_i = z_j = 2 \\ r & z_i \neq z_j \end{cases}$$
- 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

## Stochastic Block Model (Lorrain and White, 1971)

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

## 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  $\theta_{kl}, \omega_i$

## Popularity Adjusted Block Model (Sengupta and Chen, 2017)

- $P_{ij} = \lambda_{iz_j} \lambda_{jz_i}$
- $Kn$  parameters  $\lambda_{ik}$

# Hierarchy of Block Models

PABM  $\rightarrow$  DCBM:  $\lambda_{ik} = \sqrt{\theta_{z_i k}} \omega_i$

DCBM  $\rightarrow$  SBM:  $\omega_i = 1$

SBM  $\rightarrow$  Erdos-Renyi:  $\theta_{kl} = \theta$

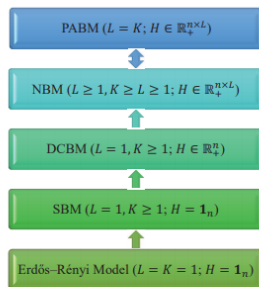


Figure 2: The hierarchy of block models



# (Generalized) Random Dot Product Graph Model

Random Dot Product Graph  $A \sim RDPG(X)$   
(Young and Scheinerman, 2007)

- Latent vectors  $x_1, \dots, x_n \in \mathbb{R}^d$  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  $A \sim GRDPG_{p,q}(X)$   
(Rubin-Delanchy, Cape, Tang, Priebe, 2020)

- Latent vectors  $x_1, \dots, x_n \in \mathbb{R}^{p+q}$  such that  $x_i^\top I_{p,q} x_j \in [0, 1]$   
and  $I_{p,q} = \text{blockdiag}(I_p, -I_q)$
- $A \sim \text{BernoulliGraph}(X I_{p,q} X^\top)$  where  $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$

If latent vectors  $X_1, \dots, X_n \stackrel{iid}{\sim} F$ , then we write  
 $(A, X) \sim RDPG(F, n)$  or  $(A, X) \sim GRDPG_{p,q}(F, n)$ .

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

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

## Connecting Block Models to the (G)RDPG Model

All  $G$  with  $A \sim \text{BernoulliGraph}(P)$  are RDPG (if  $P$  is positive semidefinite) or GRDPG (includes all block models).

Example 1 (cont'd): Assortative SBM ( $pq > r^2$ ) with  $K = 2$

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

$$P = XX^\top$$

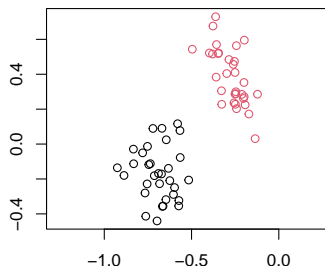
# Connecting Block Models to the (G)RDPG Model

## Example 1 (cont'd)

If we want to perform community detection,

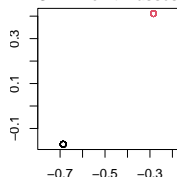
1. Note that  $A \sim \text{BernoulliGraph}(XX^\top)$  is a RDPG
2. Compute the ASE  $A \approx \hat{X}\hat{X}^\top$   
 $\hat{X} = \hat{V}\hat{\Lambda}^{1/2}$
3. Apply clustering algorithm (e.g.,  $K$ -means) to  $\hat{X}$

**ASE of the adjacency matrix drawn from S**

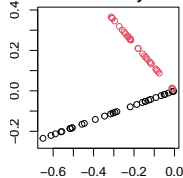


# Connecting Block Models to the (G)RDPG Model

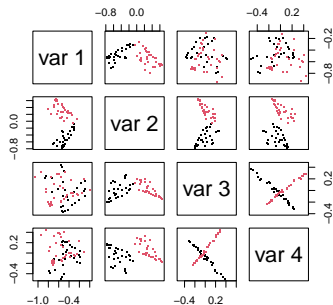
**SBM: Point masses**



**DCBM: Rays**



**PABM: Orthogonal subspaces**



## Popularity Adjusted Block Model

# Popularity Adjusted Block Model

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

Let each vertex  $i \in [n]$  have  $K$  popularity parameters  $\lambda_{i1}, \dots, \lambda_{iK} \in [0, 1]$ . Then  $A \sim PABM(P)$  if each  $P_{ij} = \lambda_{iz_j} \lambda_{jz_i}$ .

**Lemma** (Noroozi, Rimal, and Pensky, 2020):

$A$  is sampled from a PABM if  $A \sim \text{BernoulliGraph}(P)$  such that:

1. Let  $P^{(kl)}$  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  $\lambda_i^{(kl)}$  is the popularity parameter of the  $i^{\text{th}}$  vertex of community  $k$  towards community  $l$ .
3. Each block can be decomposed as  $P^{(kl)} = \lambda^{(kl)} (\lambda^{(lk)})^\top$ .

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

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

**Theorem (KTT):** Given popularity vectors  $\{\lambda^{(kl)}\}_2$ , we can define block diagonal  $X \in \mathbb{R}^{n \times 4}$  and  $U \in \mathbb{O}(4)$  such that  $A \sim PABM(\lambda_2^{(kl)})$  is equivalent to  $A \sim GRDPG(XU)$ .

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

**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  
 $\implies$  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):

If  $A \sim GRDPG_{p,q}(X)$ , then  $A \sim GRDPG_{p,q}(XQ)$  for any  $Q$  in the indefinite orthogonal group with signature  $p, q$ .

# Sparse Subspace Clustering

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

Sparse Subspace Clustering algorithm:

1. Solve  $n$  optimization problems  $c_i = \arg \min_c \|c\|_1$  subject to  $x_i = Xc$  and  $c_i^{(i)} = 0$ .
2. Compile solutions  $C = \begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix}$ .
3. Construct affinity matrix  $B = |C| + |C^\top|$ .

If  $X$  obeys the *Subspace Detection Property*, then  $B$  is sparse such that  $B_{ij} = 0$  if  $i$  and  $j$  belong to different communities and  $\|c_i\| > 0$ .

Step (1) of SSC typically performed via LASSO:

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

# Sparse Subspace Clustering

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.

$$\arg \min_{c_i} \|c_i\|_1 \text{ subject to } A_{:,i} = Ac_i \text{ and } c_i^{(i)} = 0$$

They were able to show that this obeys SDP if we replace  $A$  with  $P$ .

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

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

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

# Sparse Subspace Clustering

Wang and Xu (2016): If data matrix  $X$  consists of points lying close to low-dimensional subspaces such that:

1. Each point's distance to its subspace is sufficiently small.
2. The cosine of the angles between pairs of subspaces is sufficiently small.
3. The points corresponding to each subspace cover a sufficient amount of that subspace.

Then SDP holds with probability 1.

**Theorem** (KTT): If  $P = V\Lambda V^\top$  and  $B = nVV^\top$ , then  $B_{ij} = 0$   $\forall i, j$  in different communities.

# 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

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

Remarks:

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

# Orthogonal Spectral Clustering

**Theorem (KTT):** If  $P = V\Lambda V^\top$  and  $B = nVV^\top$ , then  $B_{ij} = 0$   $\forall i, j$  in different communities.

Orthogonal Spectral Clustering algorithm:

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  $n\rho_n = \omega\{(\log n)^{4c}\}$ ,

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

## General Community Detection for the (G)RDPG

# Generative Model

Let  $(A, X) \sim RDPG(F, n)$  such that

1. Define functions  $f_1, \dots, f_K$  such that  $f_k : [0, 1]^r \mapsto \mathbb{R}^d$  and  $f_k(t) \neq f_l(t) \ \forall k, l \in [K]$ .
2. Sample labels  $Z_1, \dots, Z_n \stackrel{iid}{\sim} \text{Categorical}(\pi_1, \dots, \pi_K)$ .
3. Sample  $T_1, \dots, T_n \stackrel{iid}{\sim} D$  with support  $[0, 1]^r$ .
4. Set latent positions  $X_i = f_{Z_i}(T_i)$  and  $X = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix}^\top$ .
5.  $A \sim \text{BernoulliGraph}(XX^\top)$



# Community Detection

Athreya et al. and Rubin-Delanchy et al.: we can approximate properties of the latent configurations via ASE.

General community detection method:

Given  $A$ ,  $K$ , and  $d$  (or  $p$  and  $q$ ),

1. Use ASE to approximate the latent configuration.
2. Use the appropriate clustering algorithm for the form of the latent configuration (manifolds).

## Parallel Segments

**Example:** Let  $U_1, \dots, U_{n_1}, U_{n_1+1}, \dots, U_n \stackrel{iid}{\sim} \text{Uniform}(0, \cos \frac{\pi}{2} a)$ ,  $f_1(t) = (t, 0)$ , and  $f_2(t) = (t, a)$ .  $X_i = f_1(U_i)$  for  $i \leq n_1$  and  $X_j = f_2(U_j)$  for  $n_1 + 1 \leq j \leq n$ . If we observe  $X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n$ , what approach will allow us to group the observations by segment?

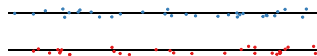


$\forall a \in (0, 1)$ ,  $\delta \in (0, 1)$ , and  $K \geq 2$ ,  $\exists N(a, \delta, K) < \infty$  such that when  $\min_k n_k \geq N$ , with probability at least  $1 - \delta$ ,

1. Single linkage clustering will produce perfect community detection.
2. Any  $\epsilon$ -neighborhood graph with  $\epsilon \leq a$  will consist of at least  $K$  disjoint subgraphs such that no subgraph contains members of two different communities.

# Noisy Parallel Segments and One-Dimensional Manifolds

**Example:** Starting with the parallel segments as before, suppose instead of observing  $X_1, \dots, X_n$ , we have noisy observations  $X_1 + \xi_1, \dots, X_n + \xi_n$  such that  $\max_i \|\xi_i\| = \xi \leq a/3$ .

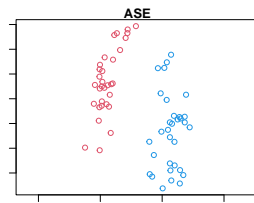
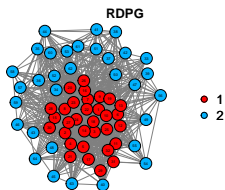
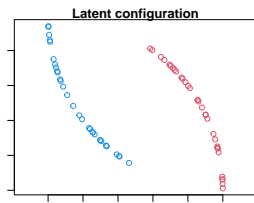


Then  $\forall a \in (0, 1)$ ,  $\delta \in (0, 1)$ ,  $K \geq 2$ ,  $\xi \leq a/3$ ,  $\exists N(a, \delta, K, \xi) < \infty$  such that when  $\min_k n_k \geq N$ , with probability at least  $1 - \delta$ ,

1. Single linkage will produce perfect community detection.
2. Any  $\epsilon$ -neighborhood graph will consist of at least  $K$  sub-graphs with no subgraph containing vertices from multiple communities.

This also holds for noisy points sampled from one-dimensional manifolds such that the manifolds are distance at least  $a$  apart.

# Recovery from the Adjacency Matrix



## Future Work

1. Show that the ASE of a random graph generated by these latent vectors produces the correct conditions for sufficiently large  $n$ .
2. Extend results to non-uniform distributions.
3. Extend results to multidimensional manifolds.
4. Relax condition for the minimum distance between manifolds.
5. Explore more robust clustering techniques for these latent configurations.
6. Extend results to the GRDPG.

Additional Slides

# Parameter Estimation

# Indefinite Orthogonal Group

$$\mathbb{O}(p, q) = \{Q : QI_{p,q}Q^\top = I_{p,q}\}$$

- $Q^\top Q \neq I$
- If  $A \sim GRDPG_{p,q}(X)$ , then  $A \sim GRDPG_{p,q}(XQ)$
- $(Qx)^\top(Qy) = x^\top Q^\top Qy \neq x^\top y$
- $\|Q\| \neq 1 \implies \|Qx - Qy\| \neq \|x - y\|$



# Community Detection in Block Models

## Likelihood

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

Example: DCBM ( $p_{k, l, i, j} = \theta_{kl} \omega_i \omega_j$ )

$$L = \prod_{i < j} \prod_{k, l}^K ((\theta_{kl} \omega_i \omega_j)^{A_{ij}} (1 - \theta_{kl} \omega_i \omega_j)^{1 - A_{ij}})^{z_{ik} z_{jl}}$$

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

# Expectation Maximization for the PABM

Full data log-likelihood

$$\begin{aligned}\log L &= \sum_{i < j} \sum_{k, l} z_{ik} z_{jl} (A_{ij} \log \lambda_{il} \lambda_{jk} + (1 - A_{ij}) \log(1 - \lambda_{il} \lambda_{jk})) \\ &\quad + \sum_i \sum_k z_{ik} \log \pi_k\end{aligned}$$

E-step

- $\gamma_{ik} = P(Z_i = k \mid \{\pi_l\}, \{\lambda_{jl}\})$
- $\log \gamma_{ik} \propto$   
 $\log \pi_k + \sum_{j \neq i} \sum_l \pi_{jl} (A_{ij} \log \lambda_{il} \lambda_{jk} + (1 - A_{ij}) \log(1 - \lambda_{il} \lambda_{jk}))$

M-step

- $\pi_k = \frac{1}{n} \sum_i \gamma_{ik}$
- $\{\lambda_{ik}\} = \arg \max_{\{\lambda_{ik}\}} E_Z[\log L]$

# MCMC Sampling for the PABM

Priors:

- $Z_i \stackrel{iid}{\sim} \text{Categorical}(\pi_1, \dots, \pi_K)$
- $\lambda_{ik} \stackrel{ind}{\sim} \text{Beta}(a_{ik}, b_{ik})$

Full joint distribution:

$$\log p = \text{constant}$$

$$\begin{aligned} &+ \sum_{i < j} \sum_k \sum_l z_{ik} z_{jl} (A_{ij} \log \lambda_{il} \lambda_{jk} + (1 - A_{ij}) \log(1 - \lambda_{il} \lambda_{jk})) \\ &+ \sum_k \sum_i z_{ik} \log \pi_k \\ &+ \sum_i \sum_k (a_{ik} - 1) \log \lambda_{ik} + (b_{ik} - 1) \log(1 - \lambda_{ik}) \end{aligned}$$

# Variational Inference for the PABM

## Mean Field Variational Inference

- Minimize  $d_{KL}(p||q)$ 
  - $p$  is the joint distribution
  - $q$  is a density of some form
- Restrict  $q(\vec{z}, \{\lambda_{ik}\}) = \left( \prod_i q_{z_i}(z_i) \right) \left( \prod_{i,k} q_{\lambda_{ik}}(\lambda_{ik}) \right)$
- Iterative solution:  $q_{\theta_i}^{(t+1)} \propto \exp(E_{\theta_{-i}^{(t)}}[\log p])$
- Approximate solution for the PABM
  - $Z_i \mid \{a'_{ik}\}, \{b'_{ik}\} \sim \text{Categorical}(\pi'_1, \dots, \pi'_K)$
  - $\lambda_{ik} \mid \{a'_{-i,-k}\}, \{b'_{-i,-k}\}, \{\pi'_k\} \sim \text{Beta}(a'_{ik}, b'_{ik})$
  - Iteratively update  $\{\pi'_K\}, \{a'_{ik}\}, \{b'_{ik}\}$  until convergence