Mixed Membership Stochastic Blockmodels

Journal of Machine Learning Research, 2008

by E.M. Airoldi, D.M. Blei, S.E. Fienberg, E.P. Xing
as interpreted by Ted Westling

STAT 572 Final Talk
May 8, 2014

Overview

1. Notation and motivation
2. The Mixed Membership Stochastic Blockmodel
3. Simulations
4. Applications
5. Conclusions

Overview

1. Notation and motivation
2. The Mixed Membership Stochastic Blockmodel
3. Simulations
4. Applications
5. Conclusions

Network theory: notation

- N nodes or individuals.
- We observe relations/interactions $Y(i,j)$ on pairs of individuals.
- Here we assume $Y(i,j) \in \{0,1\}$, $Y(i,i) = 0$, but do not assume $Y(i,j) = Y(j,i)$ (we deal with directed networks).

Scientific Motivation for Blockmodels

- General blockmodels:
  - Hypothesis: the nodes can be grouped in to non-overlapping blocks such that the probability of observing an edge from i to j is determined by latent block indicators π_i, π_j and the block relation $\pi_i^T B \pi_j$.
  - Goal: recover the blocks and the block relationships.
- MMSB:
  - Hypothesis: nodes exhibit mixed membership over latent blocks such that ... (same as above, except π_i, π_j are now distributions).
  - Goal: recover the mixed memberships and the block relationships.
- Two contrasting examples: **monks** and **rural Indian village**.

Example 1: *Crisis in a Cloister*

- Sampson (1968) collected a directed social network of 18 monks.
- Qualitative analysis indicated three clusters of monks: Young Turks, Loyal Opposition, and Outcasts.

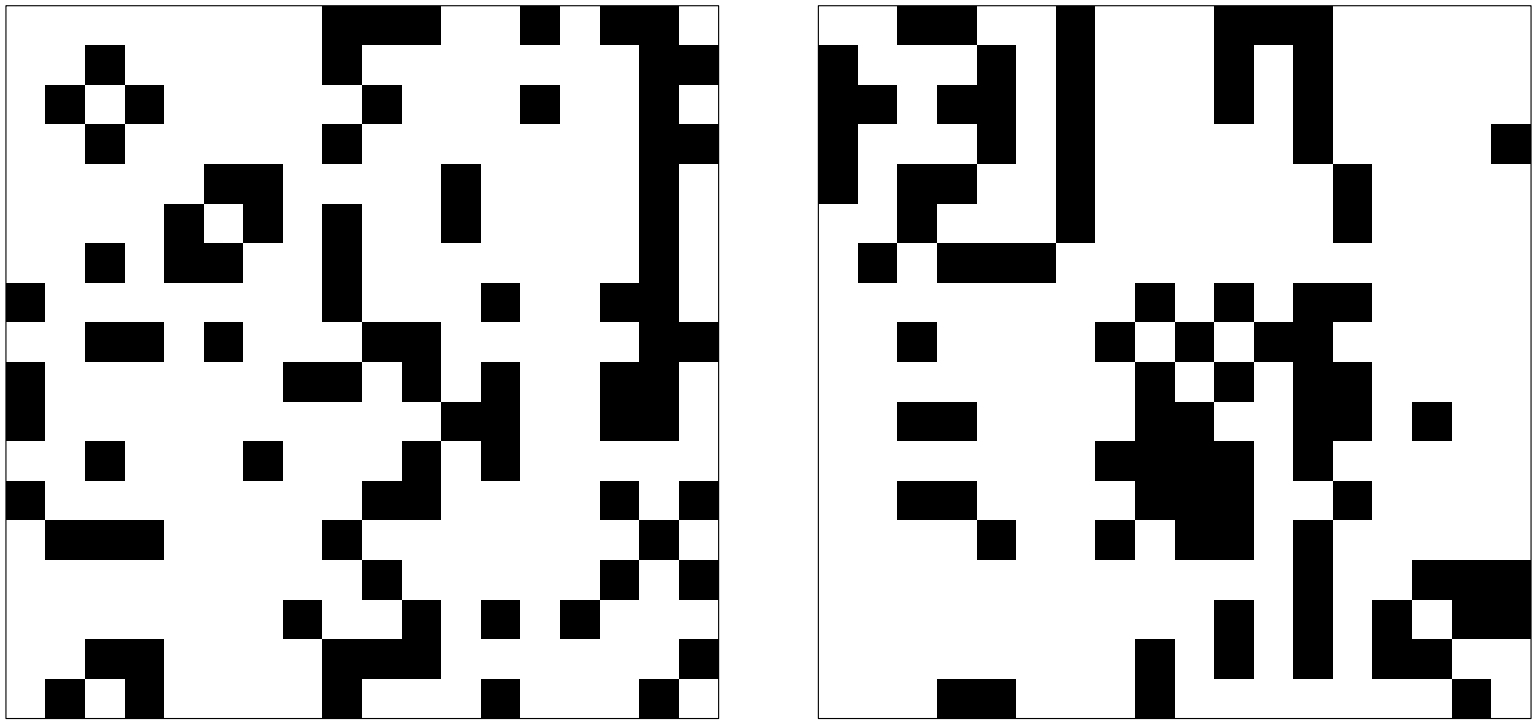


Figure: Monk adjacency matrices, ordered randomly (left) and using the a-priori clusters (right).

Example 1: *Crisis in a Cloister*

- Is this qualitative hypothesis supported by the data?
- What are the relationships between the blocks?
- How strongly defined are the blocks? (MMSB can answer while regular blockmodel cannot.)

Example 2: Indian village network

- Researchers collected individual- and household-level demographic and network data for 75 villages near Bangalore, India.
- I focused on one village of 114 households. Constructed network from household-level visit survey. **No ground truth**.

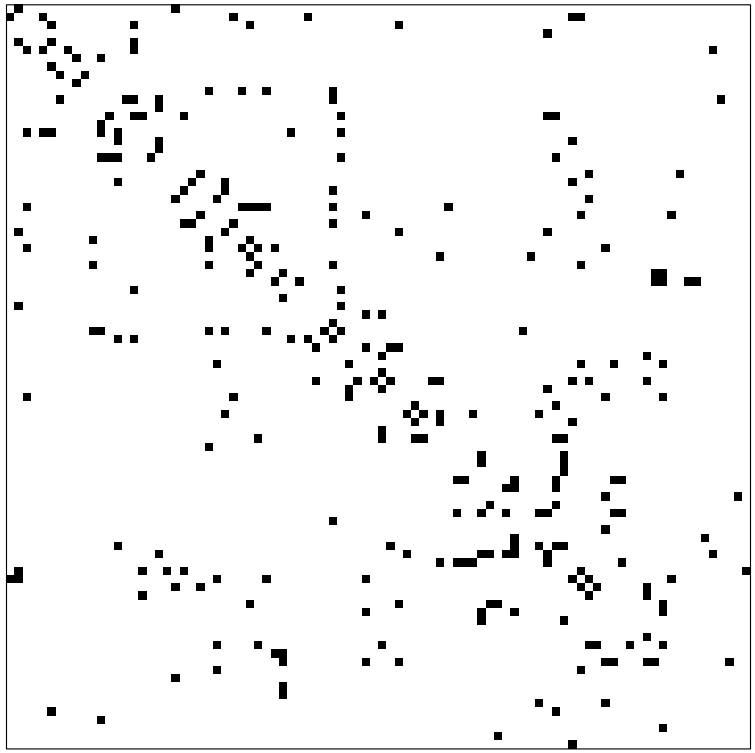


Figure: Indian village adjacency matrix, arbitrarily ordered.

Example 1: Indian village network

- Do the data exhibit block structure?
- Do the estimated blocks correspond to anything in the data?
- How strongly defined are the blocks?
- Are there particularly influential households?

Overview

1. Notation and motivation
2. The Mixed Membership Stochastic Blockmodel
3. Simulations
4. Applications
5. Conclusions

Generative MMSB

1. Fix the number of latent blocks K.
2. For each node $i = 1, \ldots, N$:
   1. Have a distribution π_i over the latent blocks, distributed as $\pi_i \sim \text{Dirichlet}(\alpha)$.
3. For each ordered pair of nodes (i, j):
   1. Assume i and j are in particular blocks for the interaction $i \to j$ indicated by $z_{i \to j}, z_{i \leftarrow j}$, where:
      - $z_{i \to j} \sim \text{Categorical}(\pi_i)$.
      - $z_{i \leftarrow j} \sim \text{Categorical}(\pi_j)$.
   2. Draw the binary relation $Y(i, j) \sim \text{Bernoulli}(b)$ where $b = z_{i \to j}^T B z_{i \leftarrow j}$. This is $B(g, h)$ when $z_{i \to j} = e_g$ and $z_{i \leftarrow j} = e_h$.
4. Choose K by estimating the model for $K = 1, 2, 3, \ldots$ and picking the model with the highest approximate BIC.

MMSB vs. Previous methods

1. Central advancement over previous methods: each node i given a distribution π_i over blocks rather than being a member of exactly one block.
2. Allows us to determine how strong someone's allegiance to each block is.
3. Learn about the block-level relationships through B.
4. Limitations: unlike previous methods, requires fixing the number of blocks K. Also does not incorporate other possible network mechanisms, e.g. reciprocity or triangle-closing.

Model Estimation: Overview

- Strategy: treat $\{\pi, Z_{\rightarrow}, Z_{\leftarrow}\} \equiv \theta$ as random latent variables and obtain posterior distribution. Treat $\{\alpha, B\} \equiv \beta$ as fixed parameters to estimate via Empirical Bayes.
- Cannot use EM algorithm, because there is no closed form for $p(\theta|Y, \beta)$. Sampling from exact posterior does not scale well, so instead use **Variational Bayes**.

Variational Bayes

- Main idea: write down a simple parametric approximation $q(\pi, Z|\Delta)$ for the posterior distribution that depends on free variational parameters Δ.
- **Minimize the KL divergence** between q and the true posterior in terms of Δ, which is equivalent to **maximizing the evidence lower bound** or ELBO

$$\mathcal{L}(\Delta|Y, \alpha, B) = \mathbb{E}_q \left[\log p(\pi, Z, Y|\alpha, B)\right] - \mathbb{E}_q \left[\log q(\pi, Z|\Delta)\right].$$

Empirical Bayes hyperparameter estimation

- Said we would update α and B with Empirical Bayes.
- This usually means maximizing $p(Y|\alpha, B)$, which we can't get in closed form.
- However, $\mathcal{L}$ is a lower bound:

$$\log p(Y|\alpha, B) = K(p, q) + \mathcal{L}(\Delta|Y, \alpha, B)$$

  so we maximize this instead, hoping that $K(p, q)$ is relatively constant in α, B.
- Really approximate Empirical Bayes. Different sort of approximation than variational inference.

"Nested" algorithm for the MMSB

1. Initialize $B^{(0)}, \alpha^{(0)}, \gamma_{1:N}^{(0)}, \Phi_{\rightarrow}^{(0)}, \Phi_{\leftarrow}^{(0)}$.
2. E-step:
   (a) for each i, j:
      (i) Update $\phi_{i \rightarrow j}, \phi_{i \leftarrow j}$
      (ii) Update γ_i, γ_j
      (iii) Update B.
   (b) Until convergence
3. M-step: Update α.
4. Until convergence.

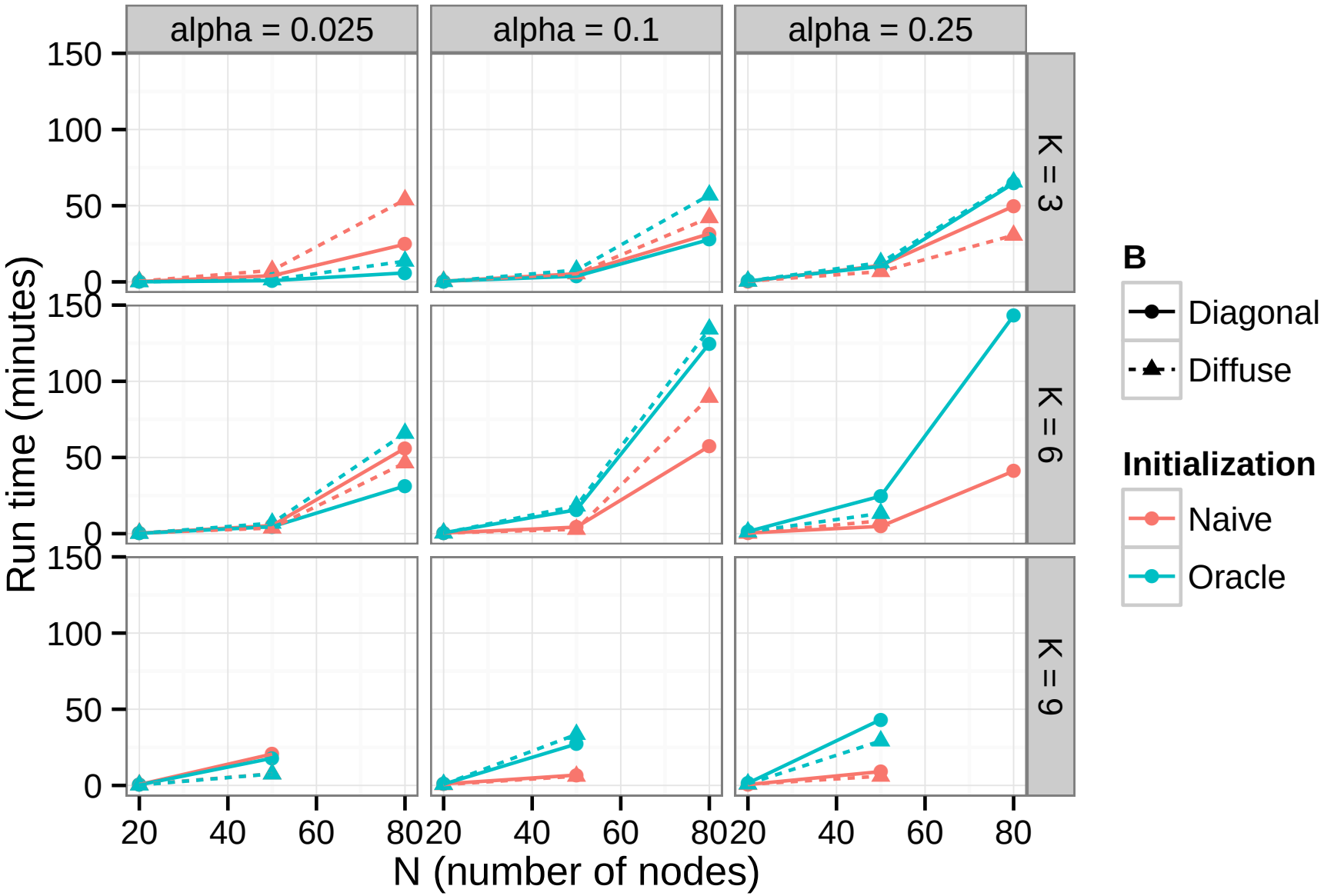
Overview

1. Notation and motivation
2. The Mixed Membership Stochastic Blockmodel
3. Simulations
4. Applications
5. Conclusions

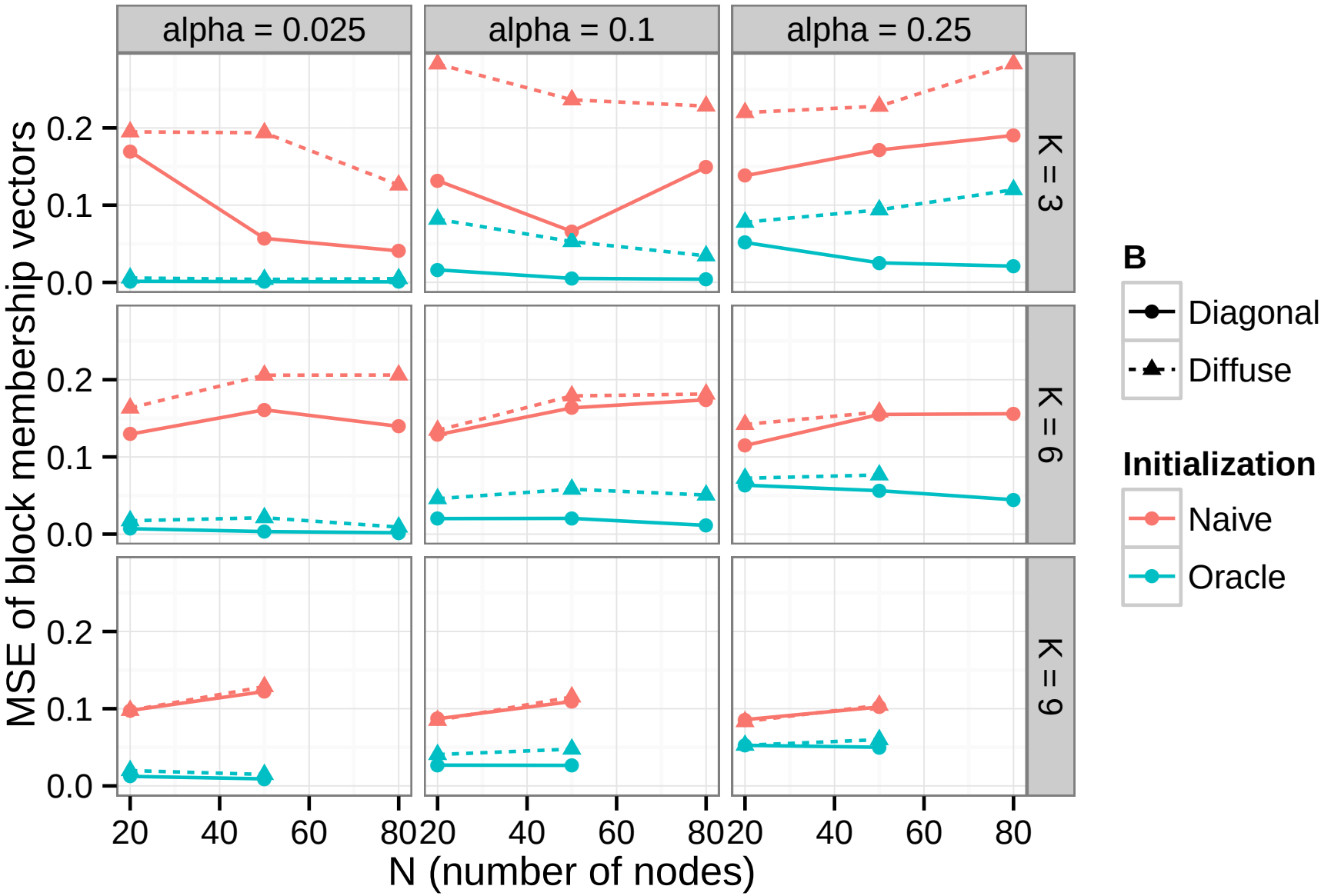
Model Simulations

- Simulated data from the **assumed model** under various parameter values.
  - $K = 3, 6, 9$.
  - $N = 20, 50, 80$.
  - $\alpha = 0.025 \cdot \mathbf{1}, 0.1 \cdot \mathbf{1}, 0.25 \cdot \mathbf{1}$.
  - B with .8 on the diagonal and .1 off-diagonal ("Diagonal"), with Uniform(.5, 1) on the diagonal and Uniform(.2, .5) off-diagonal ("Diffuse").
- For each combination of parameter values, ran **five simulations**.
- For each simulation, initialized at the true parameter values ("Oracle") and at uninformed values ("Naive").

Simulations: run time



Simulations: MSE of γ



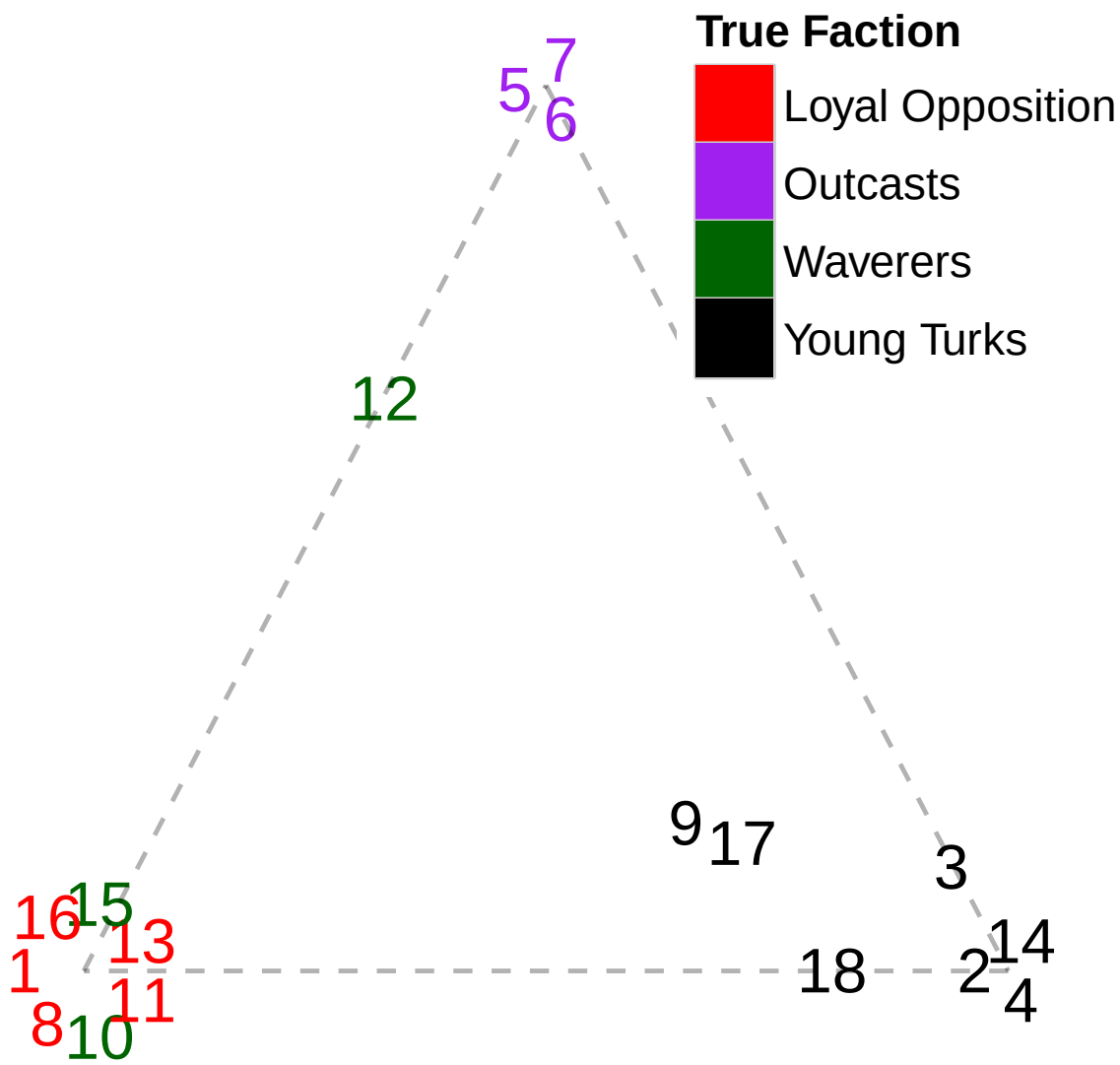
Simulations: conclusions

- Runtime is order N^2.
- Initialization matters, especially when α is small.
- Better performance when there is less mixing (small α, B close to diagonal).

Overview

1. Notation and motivation
2. The Mixed Membership Stochastic Blockmodel
3. Simulations
4. Applications
5. Conclusions

Crisis in a Cloister: posterior expected memberships



Crisis in a Cloister: adjacency matrices

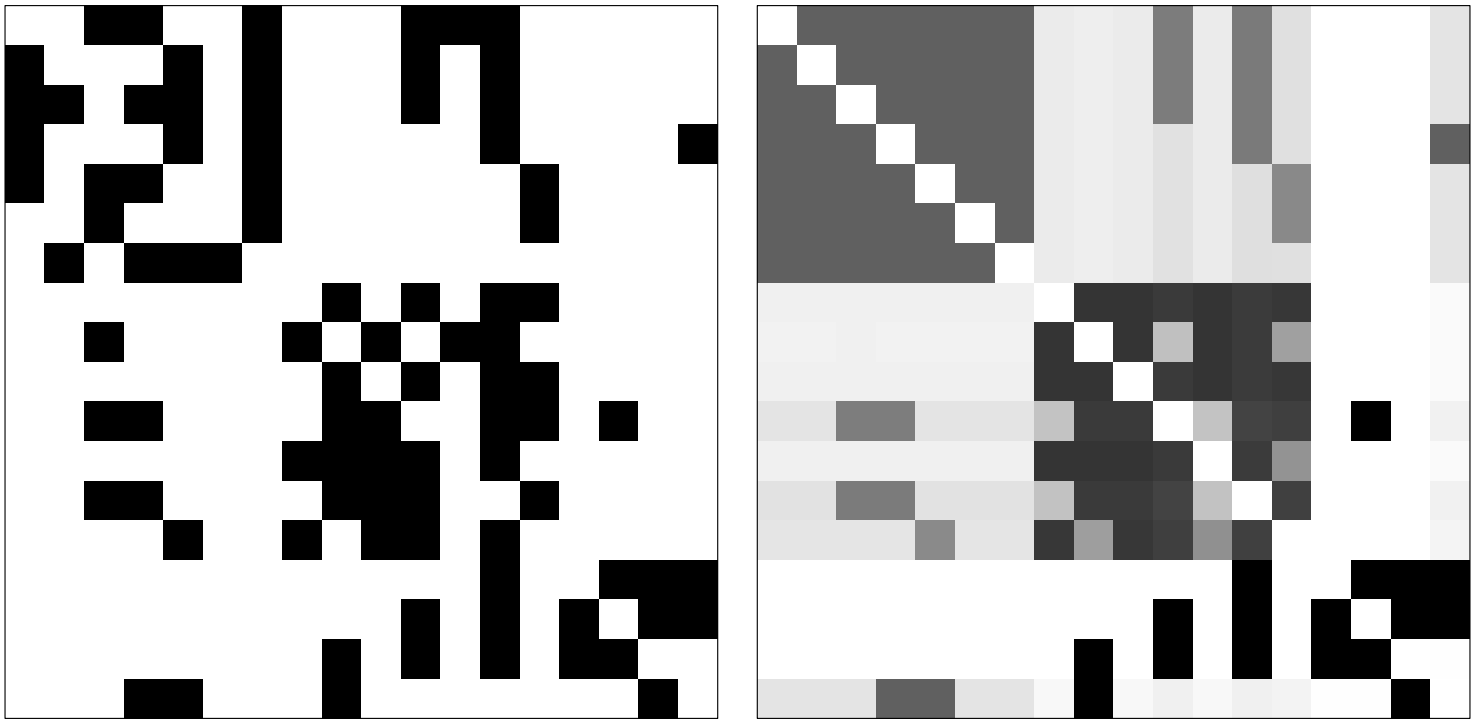


Figure: Left: Original adjacency matrix, ordered according to posterior group memberships. Right: smoothed adjacency matrix using posterior Z.

Indian Village: adjacency matrices

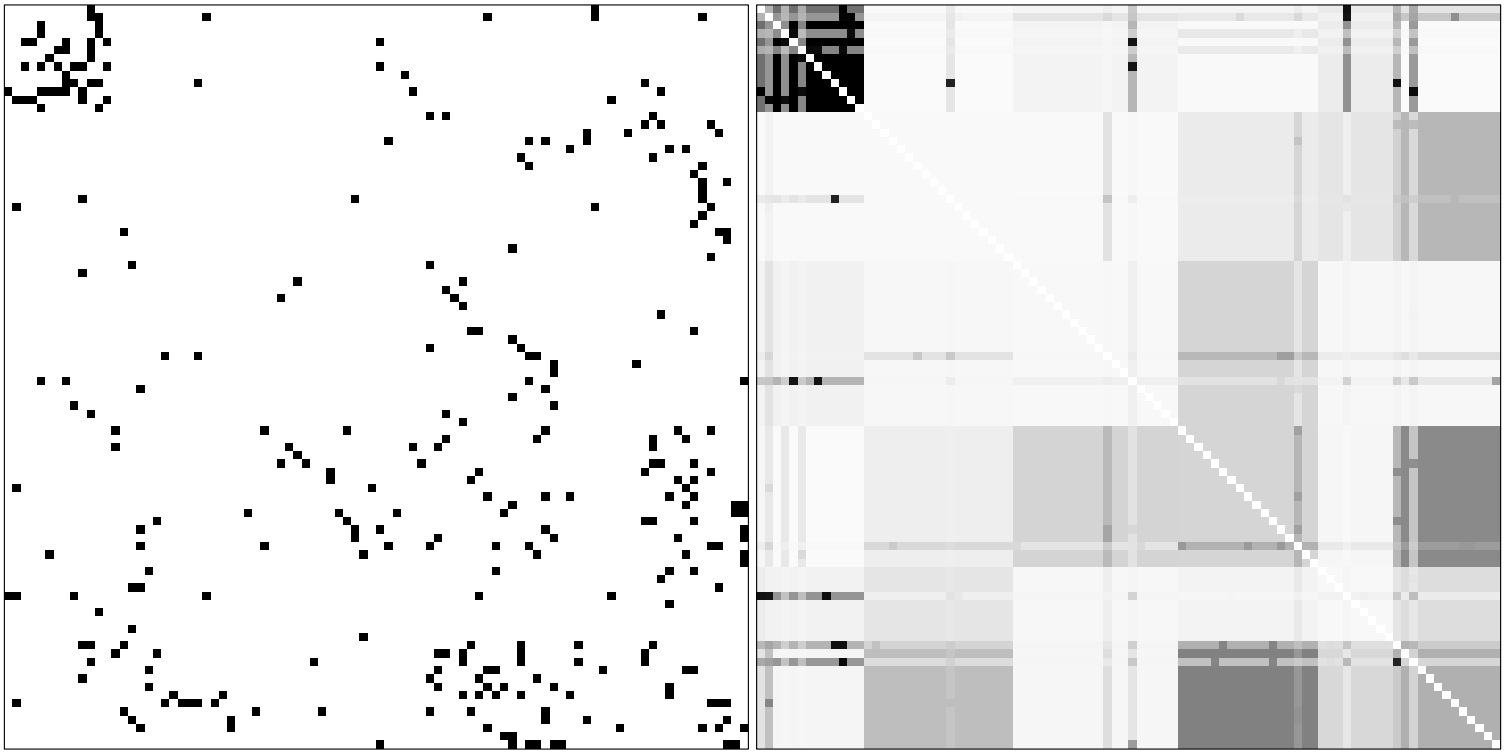


Figure: Left: Original adjacency matrix, ordered according to posterior group memberships. Right: smoothed adjacency matrix using posterior Z.

Overview

1. Notation and motivation
2. The Mixed Membership Stochastic Blockmodel
3. Simulations
4. Applications
5. Conclusions

Conclusions

- MMSB is a useful extension to blockmodels when there is a ground truth or the goal is prediction.
- However, when lacking ground truth model lacks interpretability.
- Could be improved by incorporating other network components.
- Variational algorithm also makes interpretation difficult since posterior distribution and emprical bayes are both approximations.

Thank you!