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 Intro Talk
April 22, 2014
Overview

1. Notation and motivation
2. Previous blockmodels
3. The Mixed Membership Stochastic Blockmodel
4. Conclusion and next steps
Overview

1. Notation and motivation

2. Previous blockmodels

3. The Mixed Membership Stochastic Blockmodel

4. Conclusion and next steps
Network theory: notation

- Individuals $p, q \in \{1, \ldots, N\}$.
- We observe relations/interactions $R(p, q)$ on pairs of individuals.
- Here we assume $R(p, q) \in \{0, 1\}$, $R(p, p) = 0$, but do not assume $R(p, q) = R(q, p)$ (we deal with directed networks).
Network theory: data representations

<table>
<thead>
<tr>
<th>p</th>
<th>q</th>
<th>R(p,q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table

Graph

Adjacency matrix, black=1, white=0

$\begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0 \\
\end{pmatrix}$
The Problem: Scientific Motivation

- We believe the relations are a function of unobserved groupings among the individuals.
- We want to recover the groups so we can a) predict new relations or b) interpret the existing network structure.
- Example: Monk network.
The Problem: Pictures

Figure: Two visualizations of the same binary adjacency matrix. Each filled-in square represents a directed edge. Left: ordered randomly. Right: ordered by group membership.
Figure: Two visualizations of the same binary adjacency matrix. Each filled-in square represents a directed edge. Left: ordered randomly. Right: ordered by group membership.
Overview

1. Notation and motivation

2. Previous blockmodels

3. The Mixed Membership Stochastic Blockmodel

4. Conclusion and next steps
Brief blockmodel history

- 1975: CONCOR developed by Harrison White and colleagues
- 1983: Holland, Laskey & Leinhardt introduce *stochastic* blockmodel for blocks known *a priori*.
- 2004: Kemp, Griffiths & Tenenbaum allow unknown and unlimited number of blocks in the Infinite Relational Model.
Infinite Relational Model

- Observe binary relations $R(p, q)$ between nodes $p, q \in \{1, \ldots, N\}$.
- Each node $p$ is a member of exactly one block of $K$ total blocks, $K \leq N$ unknown. Let $z_p$ be an indicator vector of block membership for node $p$, i.e. $z_p = (0, 1, 0)$.
- $B$ is a $K \times K$ matrix of block relationships. If $p$ is in block $g$ and $q$ is in block $h$ then the probability of observing an interaction from node $p$ to node $q$ is $B_{gh}$.
- $R(p, q) \sim \text{Bernoulli}(z^T_p B z_q)$.
- For example, if $p$ is in block 3 and $q$ is in block 2 then $P(R(p, q) = 1) = B_{32}$.
Block structure

\[
\begin{pmatrix}
0.8 & 0.3 & 0 & 0 \\
0.1 & 0.2 & 0.9 & 0 \\
0 & 0.5 & 0.6 & 0 \\
0.1 & 0 & 0.4 & 0.8 \\
\end{pmatrix}
\]
Overview

1. Notation and motivation

2. Previous blockmodels

3. The Mixed Membership Stochastic Blockmodel

4. Conclusion and next steps
Previous models assume each node is assumed to belong to exactly one latent block - e.g. $z_p = (0, 1, 0, 0)$.

Instead, in the MMB we assume each node has a distribution $\pi_p$ over the latent blocks.

For each interaction from $p$ to $q$, both $p$ and $q$ draw a particular block to be a part of for the interaction: $z_{p\rightarrow q} \sim \text{Discrete}(\pi_p)$, $z_{p\leftarrow q} \sim \text{Discrete}(\pi_q)$.

Then $R(p, q) \sim \text{Bernoulli}(z_{p\rightarrow q}^TBz_{p\leftarrow q})$.

$K$ chosen by BIC.
Overview

1. Notation and motivation

2. Previous blockmodels

3. The Mixed Membership Stochastic Blockmodel

4. Conclusion and next steps
Conclusion and next steps

- Blockmodels allow clustering nodes from observed network data.
- MMB extends blockmodels to let nodes be in different groups to different extents, but commit to one group during any given interaction.
- Next steps: estimation...