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

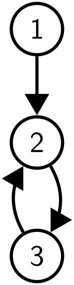
Network theory: notation

- 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

p	q	R(p,q)
1	2	1
2	3	1
3	2	1

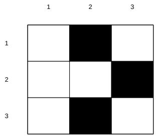
Table



Graph

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Adjacency matrix
i, j element is
$R(i, j)$

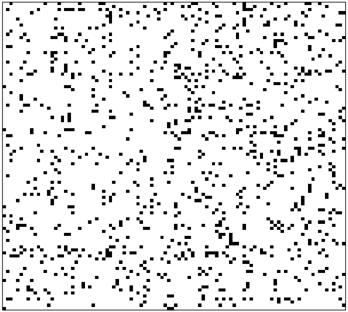


Adjacency matrix,
black=1, white=0

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



The Problem: Pictures

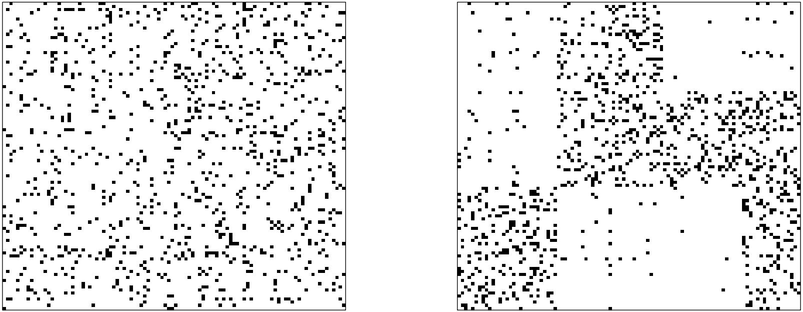


Figure: Two visualizations of the same binary adjacency matrix. Each filled-in square represents a directed edge. Left: ordered randomly. Right: ordered by group membership.

Overview

Brief blockmodel history

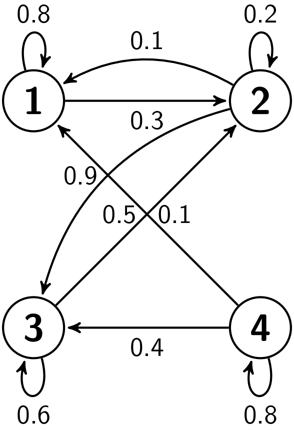
- 1975: CONCOR developed by Harrison White and colleagues
- 1983: Holland, Laskey & Leinhardt introduce *stochastic* blockmodel for blocks known *a priori*.
- 1987: Wasserman & Anderson extend to *a posteriori* estimation.
- 2004: Kemp, Griffiths & Tenenbaum allow unknown and unlimited number of blocks in the Infinite Relational Model.

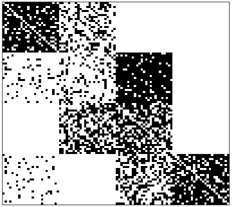
Infinite Relational Model

- 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}.
- $R(p, q) \sim \text{Bernoulli}(z_p^T B z_q)$.
- For example, if p is in block 3 and q is in block 2 then $P(R(p, q) = 1) = B_{32}$.

Block structure

$$\begin{pmatrix} .8 & .3 & 0 & 0 \\ .1 & .2 & .9 & 0 \\ 0 & .5 & .6 & 0 \\ .1 & 0 & .4 & .8 \end{pmatrix}$$



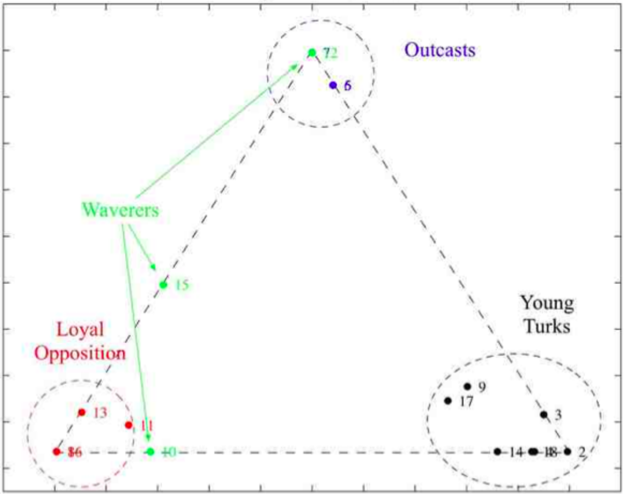


Overview

The Mixed Membership Stochastic Blockmodel

- 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 π_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: $z_{p \to q} \sim \text{Discrete}(\pi_p)$, $z_{p \leftarrow q} \sim \text{Discrete}(\pi_q)$.
- Then $R(p, q) \sim \text{Bernoulli}(z_{p \to q}^T B z_{p \leftarrow q})$.
- K chosen by BIC.

Breiger et al. (1975)



1	Ambrose
2	Boniface
3	Mark
4	Winfrid
5	Elias
6	Basil
7	Simplicius
8	Berthold
9	John Bosco
10	Victor
11	Bonaventure
12	Amand
13	Louis
14	Albert
15	Ramuald
16	Peter
17	Gregory
18	Hugh

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...