Stochastic Block Models

COS 424/524, SML 302: Fundamentals of Machine Learning

Professor Engelhardt

COS424/524, SML302

Lecture 19

Mixed membership models

In previous lectures, we learned three ways to perform exploratory data analysis using dimension reduction:

- principal component analysis (PCA);
- factor analysis (FA);
- latent Dirichlet allocation (LDA), a mixed membership model.

Today we will learn a second mixed membership model, this time for the goal of *community detection in networks*: the **mixed membership stochastic block model**.

Community detection: problem definition

Community detection is the problem of labeling communities of nodes in a network, where community membership drives the probability of a direct edge between the two nodes.

- in networks, global properties of a network are a function of local network structure
- our goal is to describe a latent set of communities among nodes based on their global network structure

Networks: two differences from previous approaches

The data we discuss today have two main differences from previous data analytic problems:

- the samples are not independent and identically distributed (IID)
- the structure on the samples is a network, which is often difficult to work with

Network analysis: challenges

- Network representation:
    - missing edges
    - smooth structure
    - network motifs
    - network semantics?
- Population statistics:
    - how does sample size (non-IID) influence analysis?
    - variability among features? edges?
- Diffusion of information in network:
    - what is background? what is network-based interaction?
    - are nodes only activating? or can they repress?

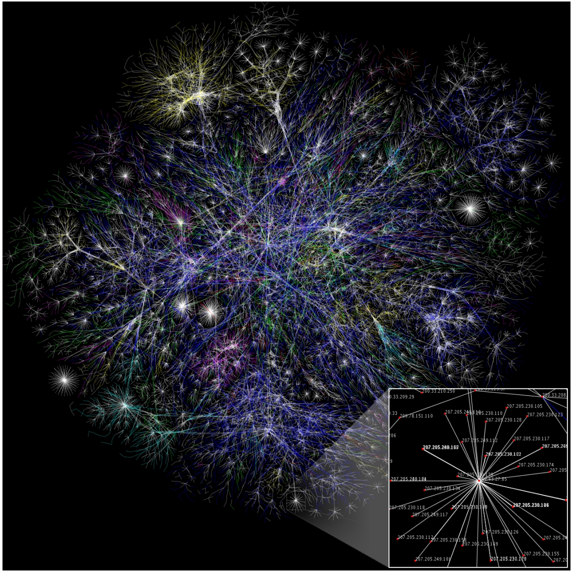
What types of networks can we consider?

Networks are:

- a collection of $\mathcal{D} = \{x_1, \ldots, x_n\}$ samples of p features;
- a binary matrix $Z \in \{0, 1\}^{n \times n}$ representing edges between each sample (nodes).

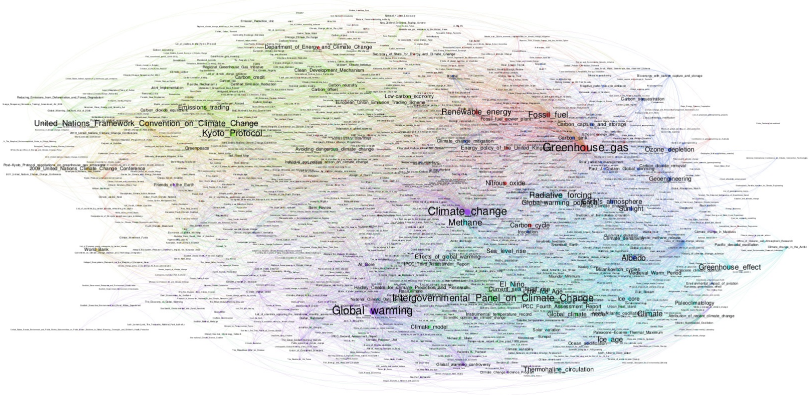
- Current networks may be directed or undirected;
- often $n = O(millions)$;
- in some networks, edges are given (e.g., "friends" in a social network);
- in other networks, edges are inferred (e.g., protein-protein interactions);
- no edge may mean edge does not exist between nodes, or edge is missing.

The internet



- relationships between IP addresses
- undirected network; edges are given; no-edge may be missing
- built for redundancy

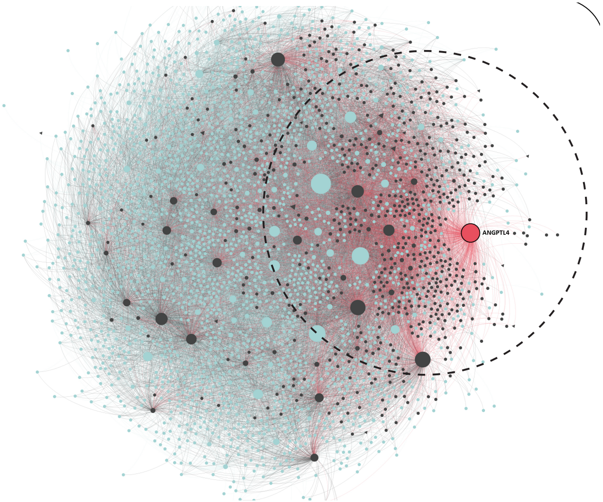
Wikipedia hyperlinks



emapsproject.com: Hyperlinks b/t wikipedia articles on climate change

- edges are hyperlinks between articles on climate change
- directed network; edges are given; no-edge may be missing

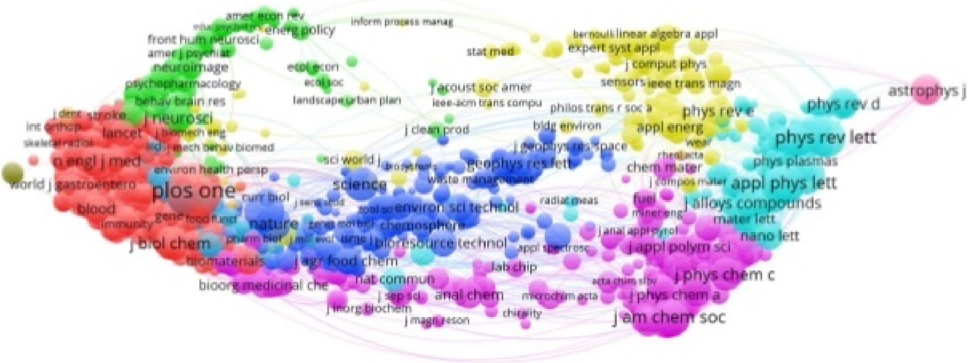
Gene interaction networks



Bianca Dumitrascu and Jonathan Lu

- pairwise interactions between genes
- directed network; edges are inferred; no-edge may be missing
- inferred from observational or time series data

Journal citation network



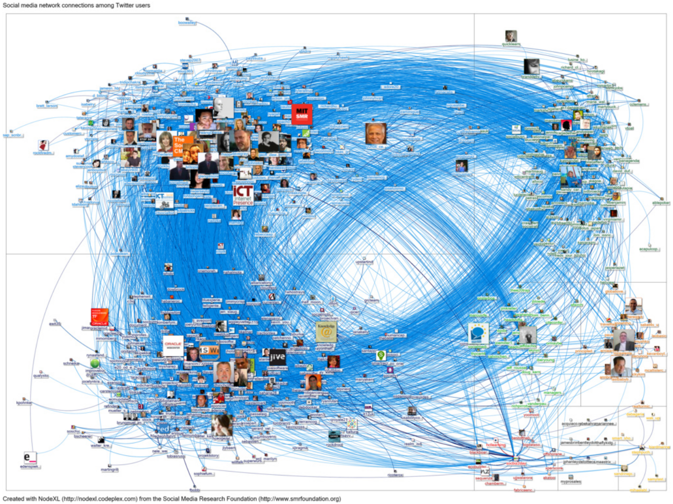
Nees Jan van Eck (Leiden University)

- edges are references between articles in scientific journals
- directed network; edges are given; no-edge is not missing

Social networks

- edges are *friend* relationships on `facebook.com`
- undirected network; edges are given; no-edge may be missing

Twitter networks



`forbes.com` article on Twitter influence networks

- pairwise interactions between users
- directed network; edges are given; no-edge may be missing
- edges represent *following*, *retweeting*, *liking*, *mentioning*

Network representation: Random graphs

First, let's describe a specific model of networks, the *random graph (Erdös-Renyi-Gilbert)*

- Represent relational data as $G = (X, Y)$
- $X \in \Re^{n \times p}$ represents n observations of p features
- $Y \in \Re^{n \times n}$ represents edge weights between n nodes
- Here, consider only binary graphs: $Y \in \{0, 1\}^{n \times n}$

Example: consider Facebook, where edges represent "friends" relationship

Network representation: ERG networks

Erdös-Renyi-Gilbert (ERG networks)

- Let $Y_{i,j}$ represent the edge between nodes i, j
- Binary edges are all sampled independently: $Y_{i,j} \sim Bernoulli(\theta)$
- Then, likelihood is written as:

$$p(G|\theta) = \prod_{i,j \in n \times n; i \neq j} \theta_{i,j}^{Y_{i,j}} (1 - \theta_{i,j})^{1 - Y_{i,j}}.$$

Much theoretical work on the value of θ and subsequent network properties

ERG networks: limitations

- Probability of edge is uniform across pairs of nodes
- cannot encode known pairwise relationships

Extend these models by adding node specific parameters

Mixed membership stochastic block model

Generative model for the mixed membership stochastic block model

- For every network node $i \in n$:
  - Draw a K dimensional vector on the simplex, given α:

$$\pi_i \sim Dirichlet(\alpha)$$

- For every pair of network nodes, $i, j \in N$:
  - Draw a membership indicator for initiator, $z_{i,j} \sim Multinomial(\pi_i)$
  - Draw a membership indicator for receiver, $z_{j,i} \sim Multinomial(\pi_j)$
  - Draw an indicator of their edge, $Y_{i,j} \sim Bernoulli(z_{i,j}^T B z_{j,i})$

Here, B is a $K \times K$ matrix capturing the probability that pairs of nodes from each of the K communities have an edge

Properties of MM stochastic block model

- Edges are *context dependent*: node may play different roles with different neighbors
- *Admixture*: each node is a combination of different communities
- $z_{i,j}$ latent variables label each possible interaction for every node with one community
- *Asymmetry*: an edge may have different underlying community labels from the two nodes
- When π, B given, can draw networks trivially from generative model
- B, and, correspondingly, Y may be free to take on arbitrary values and distributions

Mixed membership stochastic block model

Generative model for the mixed membership stochastic block model

- For every network node $i \in n$:
  - Draw a K dimensional vector on the simplex, given α:

$$\pi_i \sim Dirichlet(\alpha)$$

- For every pair of network nodes, $i, j \in n$:
  - Draw a membership indicator for initiator, $z_{i,j} \sim Multinomial(\pi_i)$
  - Draw a membership indicator for receiver, $z_{j,i} \sim Multinomial(\pi_j)$
  - Draw an indicator of their edge, $Y_{i,j} \sim Bernoulli(z_{i,j}^T B z_{j,i})$

What are the latent variables? What are the parameters?

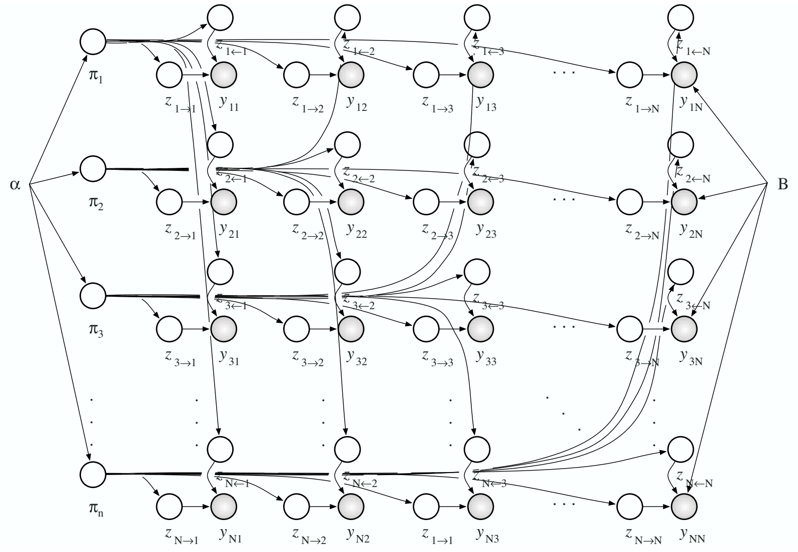
MMB: Likelihood

Mixed membership stochastic block model likelihood

$$p(Y, \pi, Z \mid \alpha, B) = \prod_{i,j \in n \times n; i \neq j} p(Y_{i,j} \mid z_{i,j}, z_{j,i}, B) p(z_{i,j} \mid \pi_j) \prod_{i \in n} p(\pi_i \mid \alpha)$$

Exercise: write out the expected complete log likelihood

Graphical model of MMB with *Z [Airoldi et al. 2008]*



Marginal mixed membership stochastic block model

Marginalize out latent variables Z:

Generative model for the marginal mixed membership SBM

- For every network node $i \in n$:
  - Draw a K dimensional vector on the simplex, given α:
$$\pi_i \sim Dirichlet(\alpha)$$
- For every pair of network nodes, $i, j \in n$:
  - Draw an indicator of their edge, $Y_{i,j} \sim Bernoulli(\pi_i^T B \pi_j)$

What are the latent variables? What are the parameters?

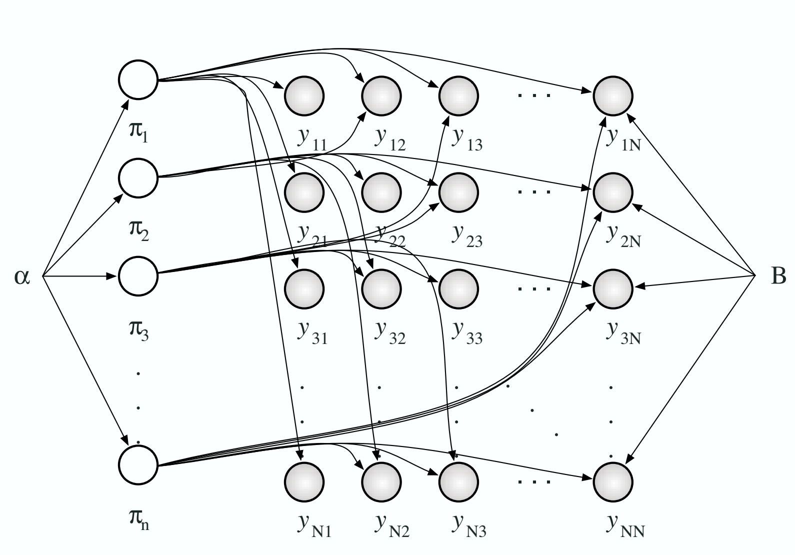
MMB: Likelihood, marginalizing out Z

Marginal mixed membership stochastic block model likelihood

$$p(Y, \pi \mid \alpha, B) \quad = \quad \prod_{i,j \in n \times n; i \neq j} p(Y_{i,j} \mid \pi_i, \pi_j, B) \prod_{i \in n} p(\pi_i \mid \alpha)$$

Exercise: write out the expected complete log likelihood

Graphical model of MMB *[Airoldi et al. 2008]*



Additional problem: sparsity

Often communities are large, graph structures are sparse.

No-edge may indicate community structure or missing data.

Extend to mixture model: data may be missing or observed; if not missing, what is the probability of an edge?

Sparse MMB likelihood

$$p(Y_{i,j} \mid Z, B, \rho) \quad = \quad \rho\delta(0) + (1-\rho) z_{i,j}^T B z_{j,i}$$

where ρ indicates probability of missing data between any pair of nodes.

Includes indicator for each pair of nodes with posterior prob of missingness

How can variables be interpreted for a MMB?

- Estimated values of π: node-specific community membership
- Estimated value of B: connectivity within and across communities
- Estimated value of $z_{i,j}$: which community gave rise to (possible) connection

What if I do not have an explicit graph to begin with?

Gaussian field or Markov random field or Gaussian graphical model

A Markov random field (MRF) is an undirected graphical model with marginal distributions on single nodes and pairs of nodes:

$$
\begin{aligned}
p(x \mid \theta) &\propto \prod_{i \sim j} \psi_{i,j}(x_i, x_j) \prod_i \psi_i(x_i) \\
\psi_i(x_i) &= \exp\left(-\frac{1}{2}\Lambda_{i,i} x_i^2 + \eta_i x_i\right) \\
\psi_{i,j}(x_i, x_j) &= \exp\left(-\frac{1}{2} x_i \Lambda_{i,j} x_j\right).
\end{aligned}
$$

Gaussian graphical models

Gaussian field or Markov random field or Gaussian graphical model

The joint distribution of a GGM then can be written as:

$$p(x|\theta) \propto \exp\left(\eta^T x - \frac{1}{2} x^T \Lambda x\right),$$

which is an unnormalized multivariate Gaussian, with $\Lambda = \Sigma^{-1}$ and $\eta = \Lambda\mu$.

Σ^{-1} is the precision matrix, and an entry is the covariance of $x_i, x_j | x_{\neg i,j}$.

When $\Lambda_{i,j} = 0$, then x_i, x_j are conditionally independent and have no edge.

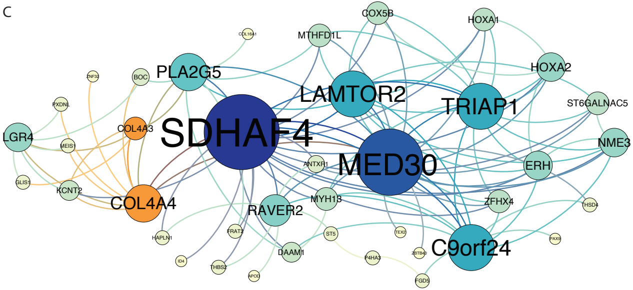
Gaussian graphical models in practice

From data $X \in \Re^{n \times p}$, estimate a sparse, $n \times n$ precision matrix

This can practically be done in three ways (probably more)

- When $n \gg p$ this can be done analytically
- Graphical lasso: $x_i = \beta^T x_{\neg i} + \epsilon$
- Sparse, low dimensional precision matrix estimates.

c



What analyses can I perform with MMB?

Dimension reduction

From network Y $(n \times n)$ to node-specific communities π $(n \times K)$ & interactions B $(K \times K)$

Prediction

Expected posterior of $Y(i, j)$ (marginalize over Z):

$$E[Y(i, j) = 1 | \hat{\pi}_i, \hat{B}, \hat{\pi}_j] \approx \hat{\pi}_i^T \hat{B} \hat{\pi}_j.$$

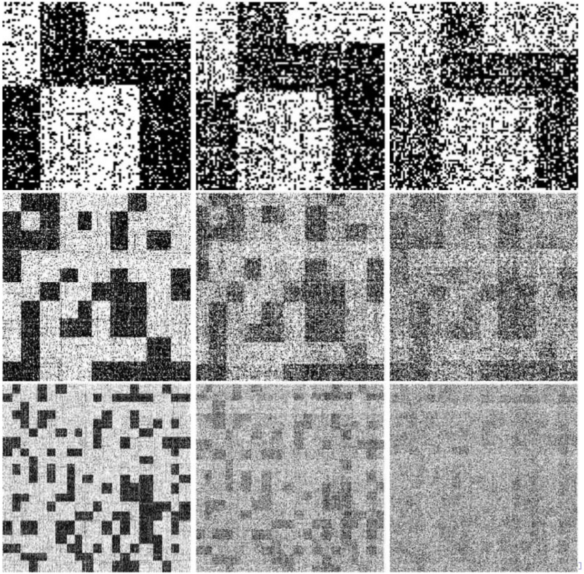
Network painting

Color each node in a network with $\hat{\pi}$; color each edge with $\hat{Z}$

Network generation

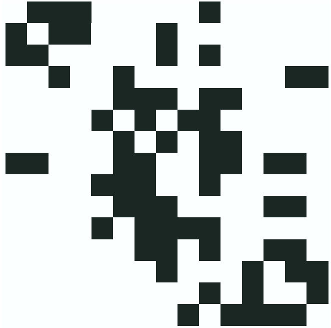
Generate networks with same latent structure as data using $\hat{\pi}, \hat{B}, \hat{Z}$

Simulated networks *[Airoldi et al. 2008]*



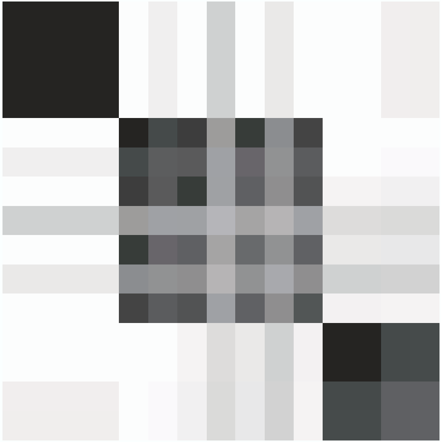
Example: Analysis of a monastery *[Airoldi et al. 2008]*

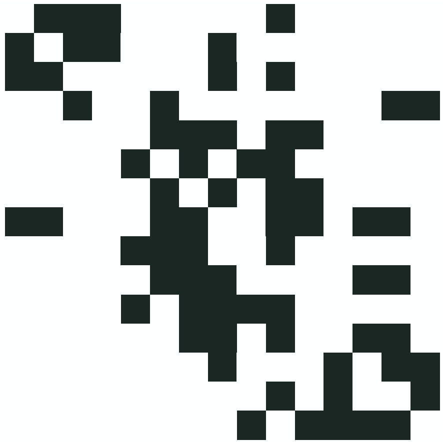
"Sampson (1968) surveyed 18 novice monks in a monastery [over two years] and asked them to rank the other novices in terms of four sociometric relations: like/dislike, esteem, personal influence, and alignment with the monastic credo."



- pairwise interactions between monks
- Sampson suggested four communities:
  - *loyal opposition*: members joined monastery first
  - *young turks*: members joined later
  - *outcasts*: members were not accepted into either main faction
  - *waverers*: members did not take sides

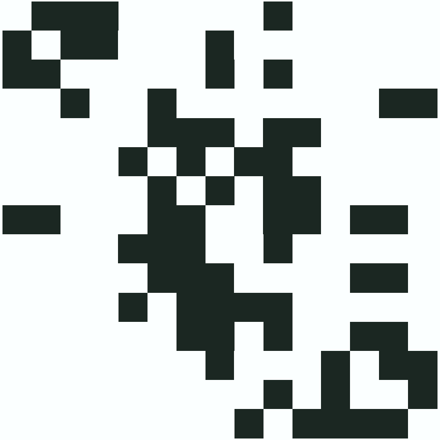
Example: Predicted interactions *[Airoldi et al. 2008]*





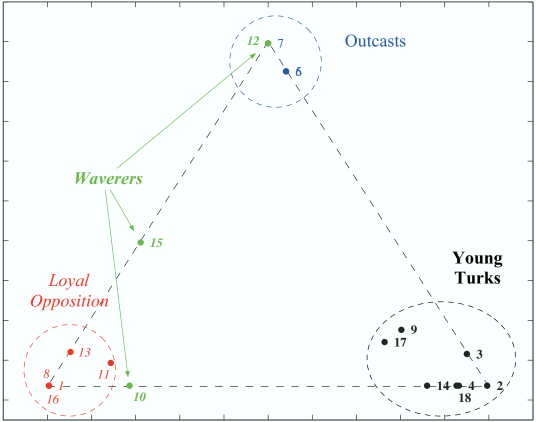
- predicted interactions between monks
- predictions are symmetric, *soft*: not binary

Example: Reconstructed interactions *[Airoldi et al. 2008]*



- reconstructed interactions between monks
- reconstruction is asymmetric, *soft*: not binary

Selecting K [Airoldi et al. 2008]

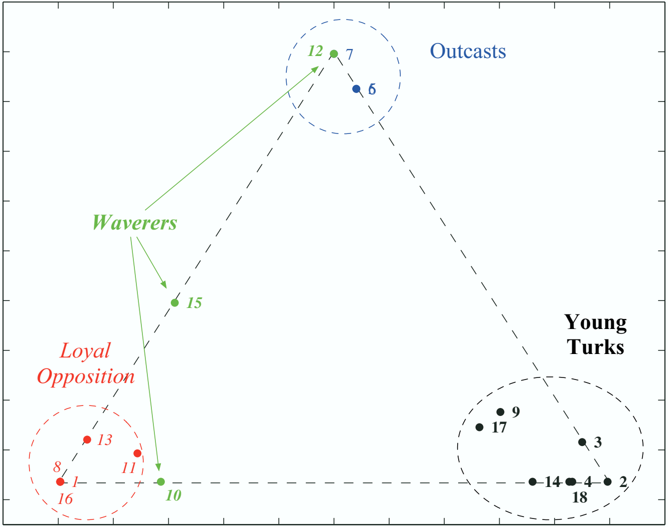


1 Ambrose
2 Boniface
3 Mark
4 Winfrid
5 Elias
6 Basil
7 Simplicius
8 Berthold
9 John Bosco
10 Victor
11 Bonaventure
12 Amand
13 Louis
14 Albert
15Ramuald
16 Peter
17 Gregory
18 Hugh

Used Bayesian information criterion (BIC) to select $K = 3$:

$$BIC = -2\log(p(Y|B,\alpha)) + \log(n)K$$

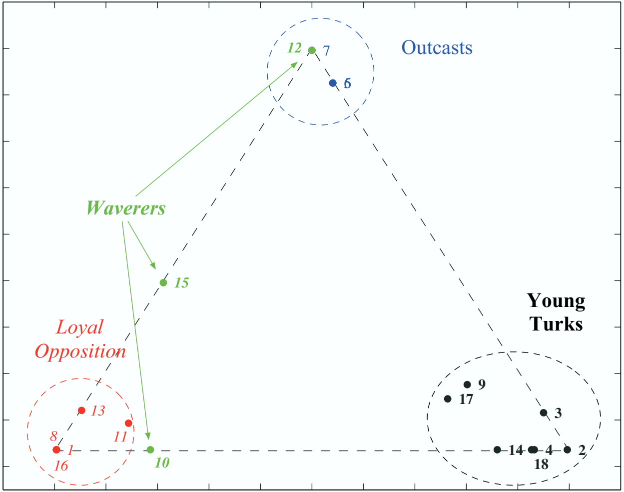
Node-specific community proportions *[Airoldi et al. 2008]*



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
15Ramuald
16 Peter
17 Gregory
18 Hugh

Is this type of community structure what you expected ($K = 3$)?

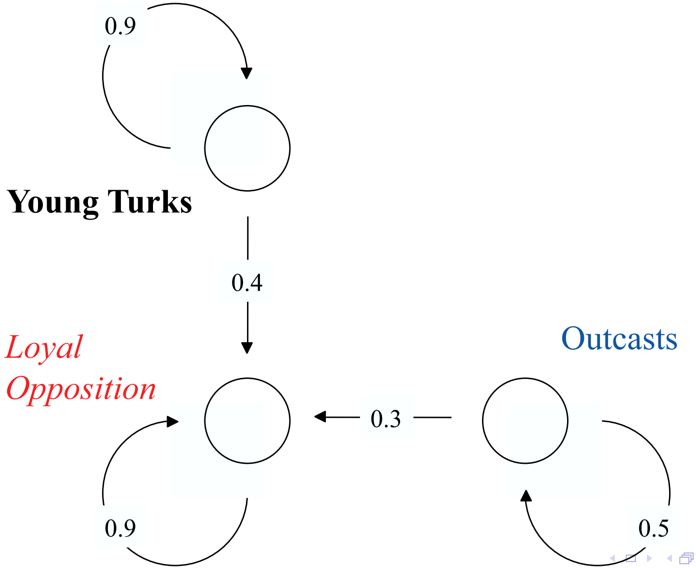
Node-specific community proportions *[Airoldi et al. 2008]*



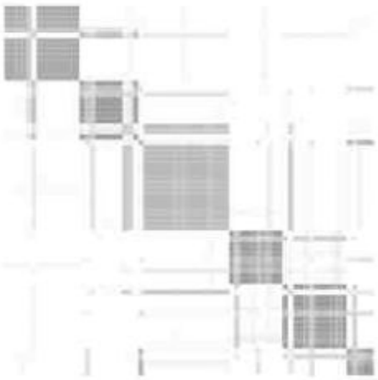
1 Ambrose
2 Boniface
3 Mark
4 Winfrid
5 Elias
6 Basil
7 Simplicius
8 Berthold
9 John Bosco
10 Victor
11 Bonaventure
12 Amand
13 Louis
14 Albert
15Ramuald
16 Peter
17 Gregory
18 Hugh

"After two years, John Bosco and Gregory get expelled, most young turks leave, and the order dissolves"

Monk block structure *[Airoldi et al. 2008]*

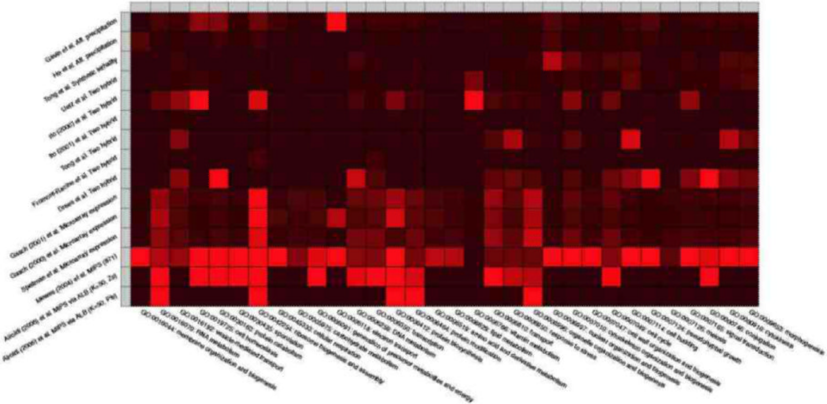


Middle+high school friendship network *[Airoldi et al. 2008]*



True (L) and estimated (R) friendships among people at a school.

Yeast protein interaction network *[Airoldi et al. 2008]*



Red is area under precision-recall curve for functions (columns) of genes in a group (rows)

Relationship with LDA model

- Both are mixed membership models, where each node/sample has proportions of membership
- Both are linear models, with expected complete log likelihood being a linear function of variables
- Latent representation, however, differs: LDA has no $K \times K$ matrix;

A note on inference

- As with LDA models, posterior inference is often intractable
- Variational expectation-maximization (and variants) has been used
- Sampling approaches have also been suggested, with various success

MMB: Extensions

- Sparsity
- Informative priors
- Fully Bayesian treatment
- Node attributes included
- Dynamic MMB
- Hierarchical relational models
- Non-parametric MMBs

MMBs: Summary & Assumptions

- Mixed membership stochastic block models assign each edge in a network to a pair of communities according to the node-specific community proportions, and draw from the blockmodel of that pair of communities
- MMBs do not use observations; instead, only use network connectivity.
- Samples and populations have alternative meaning from standard statistical framework: only one observation of network edges; samples not IID

Network analysis has a rich history

- Sociology and friendship networks (1930s)
- Sociology, mathematics, and psychology (1950s)
- Psychology (1960s)
- Statistics (1970s)
- Computer science and physics (1990s)

Additional Resources

- Many details of this lecture were borrowed from Airoldi's slides on MMBs (Guest lecture for EE380L)
- MMB paper: *[Airoldi, Blei, Fienberg, Xing 2008]*
- R code: iGraph, LDA