Mixed Membership Stochastic Blockmodels

(2008) Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg and Eric P. Xing

Herrissa Lamothe

Princeton University

Outline

Motivating question

Are there certain "rules" dictating how individual vertices (or nodes) make "decisions" to connect/not to connect to other vertices?

- Let's suppose we can draw a network graph, G from a **generative model**, so that G comes from a probability distribution $Pr(G|\theta)$, governed by parameter θ
- so that if θ is partly known, it can act as a constraint in generating synthetic graphs (similar to G)
- if G is partly known, we can use it to infer a θ that make G likely

How do we obtain θ?

$$Pr(G|\theta) = \prod_{ij} Pr(A_{ij}|\theta)$$

- **Vertex-level attributes:** too chaotic (over-fit)
- **Global-network attributes:** too general (over-generalize)

Introducing structure

- Differentiates data from noise
- Captures relevant patterns
- Describing patterns and predicting them

Vertices $\rightarrow$ **Communities**

Let's Rephrase the Motivating Question

How do **groups** of vertices make decisions to connect/not to connect to other groups?

The stochastic block model

Class of models of which Mixed Membership Stochastic Blockmodel is a variant.

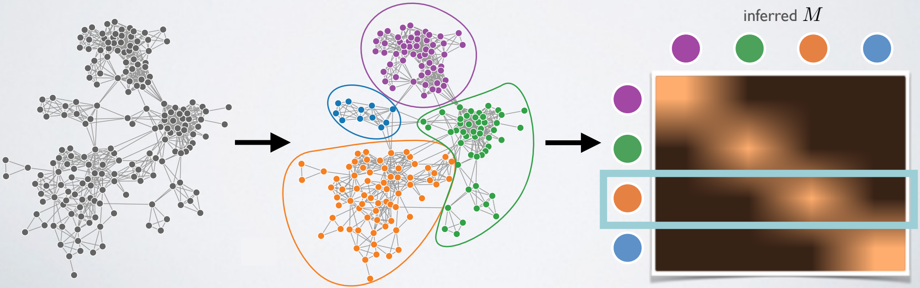
Aaron Clauset (UColorado at Boulder, Computer Science)

has amazing slides...

the stochastic block model

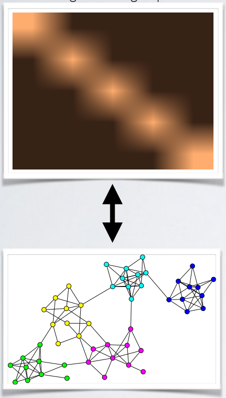
- each vertex i has type $z_i \in \{1, \ldots, k\}$ (k vertex types or groups)
- stochastic block matrix M of group-level connection probabilities
- probability that i, j are connected = M_{z_i, z_j}

community = vertices with same pattern of inter-community connections

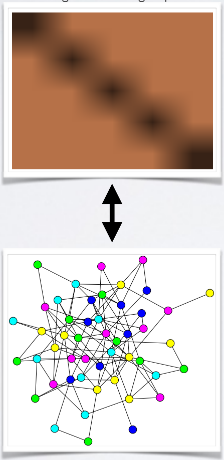


the stochastic block model

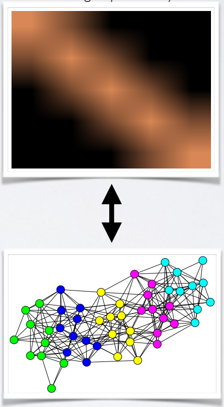
assortative
edges within groups



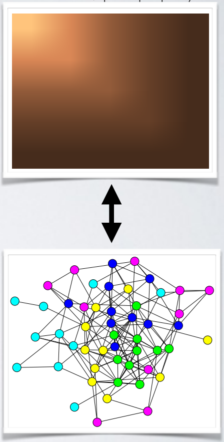
disassortative
edges between groups



ordered
linear group hierarchy



core-periphery
dense core, sparse periphery



the stochastic block model

the most general SBM

$$\Pr(A \mid z, \theta) = \prod_{i,j} f\left(A_{ij} \mid \theta_{\mathcal{R}(z_i, z_j)}\right)$$

A_{ij} : value of adjacency

$\mathcal{R}$: partition of adjacencies

f : probability function

$\theta_{a,*}$: pattern for a-type adjacencies

Binomial = simple graphs
Poisson = multi-graphs
Normal = weighted graphs
etc.

θ_{11}	θ_{12}	θ_{13}	θ_{14}
θ_{21}	θ_{22}	θ_{23}	θ_{24}
θ_{31}	θ_{32}	θ_{33}	θ_{34}
θ_{41}	θ_{42}	θ_{43}	θ_{44}

Interactions across blocks

- Classes of SBMs are looking at interactions **across** "blocks"
- They are **not** directly looking for patches of "connectedness" **among** nodes.
- Assume that individual nodes' behavior can be explained **entirely** by group membership.

Sociological contributions

If this way of operationalizing the problem (in terms of group membership) seems familiar, that is because sociologists have made many contributions to this line of research.

Implications for interactions within blocks

Given this framework, what are we implying about connections between vertices i and j if they belong to the same group k?

Outline

Mixed Membership

Mixed membership stochastic block model (MMSB) (f = Bernoulli)

Similar to SBM, but with an extra layer of parameters to estimate.

Key assumptions remain: $Pr(i \to j) = M_{zi,zj}$

M = Stochastic Block Matrix

But, z_i and z_j must be **estimated** for each *dyadic interaction* between all i and j vertices, based on a latent **mixed membership vector** for each i.

Mixed Membership

- each vertex i has a **mixed membership vector** $\theta_i \sim \text{Dirichlet}(\alpha)$
- vertex i takes on a *single* group membership with probability θ_i in the **context** of a **directed** dyadic interaction with vertex j

For each pair of vertices $(i, j) \in [1, N]x[1, N]$ (adjacency matrix),

Sample group membership of i and j **independently**

Sample group $z_{i \rightarrow j} \sim \text{Multinomial}(\theta_i, 1)$

Sample group $z_{i \leftarrow j} \sim \text{Multinomial}(\theta_i, 1)$

Note: This model specification can be adapted to *undirected* interactions easily, $z_{i \leftrightarrow j}$

Mixed Membership

Key Innovations:

- nodes belong to **more than one** group
- nodes belong to groups with **different strengths** of membership
- nodes take on a **specific group membership** for the duration of an *interaction*
- include sparsity parameter, control model's sensitivity to zeros in adjacency matrix due to noise.

Final sampling of $A_{ij} \sim \text{Bernoulli}\left(\rho \; z_{i \to j} \; M \; z_{i \leftarrow j} + (1 - \rho) \; \delta_0\right)$

Mixed Membership

joint probability distribution:

$$p(A, \theta, z_i, z_j|\alpha, M) = \prod_{i=1}^{N} p(\theta_i|\alpha) \prod_{j=1}^{N} p(z_{i\rightarrow j}|\theta_i) p(z_{i\leftarrow j}|\theta_j) p(A_{ij}|z_{i\rightarrow j}, z_{i\leftarrow j}, M)$$

Where,

- A is the observed adjacency matrix
- M is the block matrix
- θ_i and θ_j are the mixed membership vectors for i and j
- z_i and z_j are the group membership indicators for i and j during their interaction

The only input to this model is the **number of groups.**

Outline

MMSB Estimation

We won't focus on this too much...

- Only to note that the actual marginal probability (likelihood) of $p(A|\alpha, M)$ is **not tractable to compute** (i.e. we cannot integrate out z and α).
- Airoldi et al. carry out an *approximate* inference and parameter estimation.

In order to compute the posterior degrees of membership for all i given hyperparameters (θ and α):

$$p(\theta|A, \alpha, M) = \frac{p(\theta, A|\alpha, M)}{p(A|\alpha, M)}$$

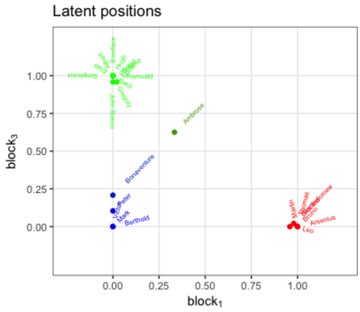
They use **variational methods**: "find a lower bound of the likelihood and approximate posterior distributions for each objects membership vector." (Airoldi et al, 2015 p. 7)

MMSB Estimation

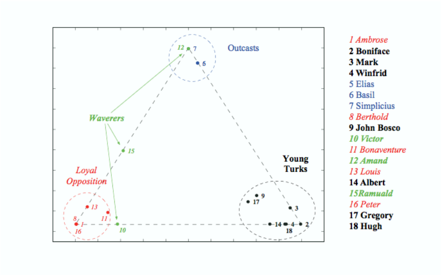
What are the implications of this estimation strategy for the model?

Outline

Sampson monk factions



(a) MMSB with LDA



(b) Airoldi et al.: Variational Methods

Synthetic data

How to create "good" synthetic data?

- various levels of difficulty of detection? ($c_{in} - c_{out}$)
- specific block patterns?
- always the problem of linking recovered partitions to actual theorized groups

Outline

Potential Issues

1. Estimation procedure
2. Selecting "K" number of groups

Two Goals: Prediction and Interpretation

1. **Prediction**: We want to identify the number of group affiliations, K, most predictive of observed patterns of interaction, G.
2. **Interpretation (Hypothesis-Driven)**: We want to determine how predictive a *specific* set of hypothesized group affiliations, K, is of observed patterns of interaction, G.

Selecting K

Our goal affects how we deal with two key issues in community detection:

1. How do we determine the right number of groups in analyzing the interactions in G?

   prediction: Find the most predictive K (i.e. BIC, or cross-validation).
   interpretation: ?

2. How do we know that the way our algorithm ("heuristic") partitioned the data is always the same for K groups?

   prediction: permutation of components of θ_i to interpret $[E|\theta_i(k)|A]$, **if** we know components of each functional group.
   interpretation: ?

Outline

There are also times when we, as researchers, have **prior (if partial) information** that we want to test and inject in our model...

- partial information on K
- partial information on M (block, or "mixing" matrix)
- partial information on θ (mixed membership vectors
- partial information on **decision rule** for selecting $z_{i\rightarrow j}$ or $z_{i\leftarrow j}$. (i.e. we might want to break the independence assumption).
- or any combination of the above

Questions to the room

- Are there tools available to allow researchers to take advantage of this potential flexibility in the MMSB and SBMs in general?
- What types of research question can we address using the MMSB model?
- What types of research question can we address if we can take full advantage of the model's potential flexibility?

and as always, there are questions about the **assumptions** made by this model.

- Under what circumstances (for what research questions) are we willing to make them?
- Under what circumstances would these assumptions not hold?