Towards Contagion Driven Homophily in Large Networks

November 13, 2017

Motivation for Community Detection

- Subnetwork-level measures provide better understanding for processes than network-level ones
  - A hub at community level may not be so much of a hub at network level.
  - Sub-networks can provide better meaning and context for behavioral patterns.

How We Do It

- (Correlated)Mixed Membership Stochastic Blockmodel(allowing for large directed networks)
    - Individuals within and between communities connect according to a probability matrix(higher probabilities within than between)
    - Each person has a probability vector determining its propensity to belong to different communities, which could be correlated with other individuals
    - Each person determines its community according to their potential interaction to other individuals in a network
    - Two individuals connect according to the probability of the communities they have determined in potential interaction

Big Picture

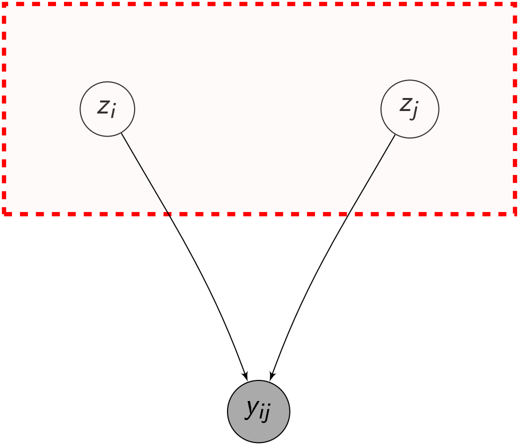
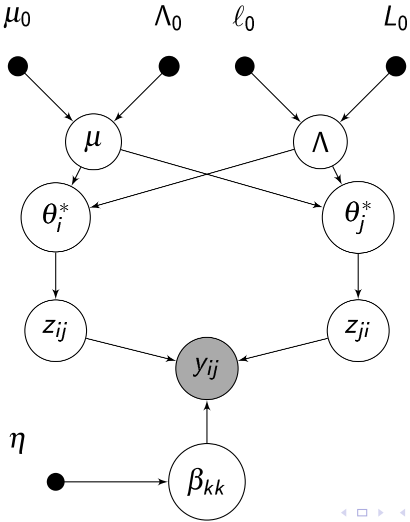


Figure: Graphical representation of network structure

Generative Algorithm

- $\forall k \in [1,..,K]$
  - draw the diagonal elements of the block matrix B via$\beta_{k,k} \sim Beta(\eta_0, \eta_1)$
- $\forall i \in \mathcal{N}$
  - draw the mean of the logit mixed membership vector through $\mu \sim Normal(\mu_0, \mathbf{\Lambda}_0)$
  - draw the precision of the logit mixed membership vector through $\mathbf{\Lambda} \sim Wishart(\ell_0, \mathbf{L}_0)$
  - draw a K-dimensional vector,$\theta_i^* \sim Normal(\mu, \mathbf{\Lambda})$
  - construct the simplical mixed membership via logistic transformation , $\theta_{i,k} = \frac{exp(\theta^*_{\mathbf{i},\mathbf{k}})}{\sum_l exp(\theta^*_{\mathbf{i},\mathbf{l}})}$
- $\forall (i,j) \in \mathcal{E}$
  - draw one-hot membership indicator vector for i when contacting j, $\mathbf{z}_{i\rightarrow j} \sim Categorical(\theta_i)$
  - draw one-hot membership indicator vector for j when contacted by i, $\mathbf{z}_{i\leftarrow j} \sim Categorical(\theta_j)$
  - sample a link between $i \rightarrow j$ with probability $\mathbf{z}_{i\rightarrow j}\mathbf{B}\mathbf{z}_{i\leftarrow j}$, $Y(i,j) \sim Bernoulli(\mathbf{z}_{i\rightarrow j}\mathbf{B}\mathbf{z}_{i\leftarrow j})$

MMSB model


μ_0
Λ_0
ℓ_0
L_0
μ
Λ
θ_i^*
θ_j^*
z_{ij}
z_{ji}
y_{ij}
η
β_{kk}

Inference Engine

- For all models we use Stochastic Variational Inference to get back paramaters
- Variational Inference(VI)
  - Approximating posterior through an easier tractable proposal distribution by minimizing the KL divergence between true posterior and proposal
- Stochastic Variational Inference(SVI)
  - Using Minibatch sampling to have a noisy estimate of the gradient with decreasing step size to optimize the variational objective
  - allows handling large datasets

Applications for Community Detection

- Summarizing data in a lower dimensional space
- Identifying new roles at the sub-network level(i.e local hubs/bridges)
  - domain specific influentials may not be influential/central in other domains
- Extensions:Simplifying the study of social influence when latent homophily is present
- Extensions:Allowing for monitoring the growth of a network in dynamic setting

Our Contribution

- Although many network-level measures have been useful in predictive ability, but networks are in fact mixture of multiple subnetworks
- Other models such as latent space models aim at answering similar questions but fails to give meaning and context to the euclidean latent space
  - Latent space models do not provide measures for subnetworks.
- MMSB model estimations for larger networks make unnecessary simplifying assumptions
  - For inference convenience they deal with undirected graphs
  - They do not allow for correlated membership by assuming conjugate priors
  - They do not allow for potential evolution of networks by assuming conjugate priors
- We address all of these problems by assuming
  - directed links
  - correlated membership through logistic normal prior
  - allowing for evolution in later dynamic model by introducing logistic normal priors
  - Allow large networks by incorporarting our scalable SVI with these new constraints

Using Behavioral Data

- We exploit behavioral data in better detecting communities, and labeling them
- We can assess improved prediction of signal penetration rate
- We can redefine opinion leadership as a function of community roles

Underlying Latent Structure

- Individuals connect according the similarity in preferences they have for tie formation
- People may choose their ties according to both their tie and behavioral preferences
- Link and Behavioral latent space may overlap
  - Solution:joint modeling of behavior and link formation

Latent Space

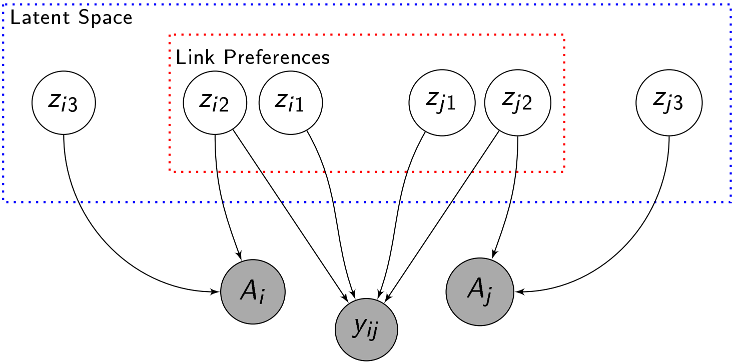


Figure: Graphical representation of network structure and behavior

When Influence is Present

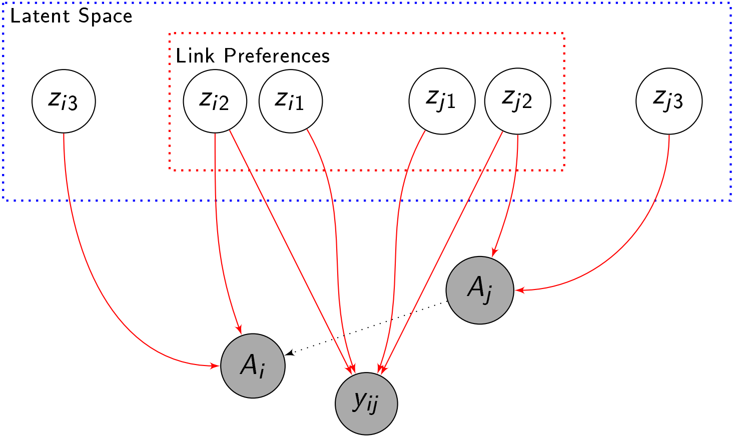
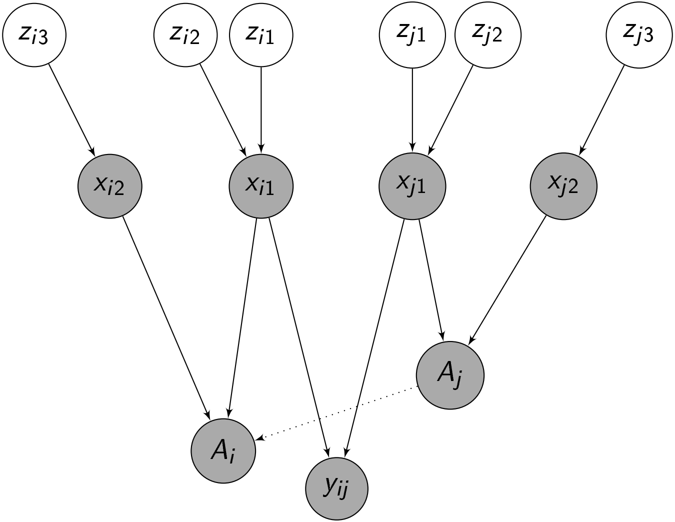
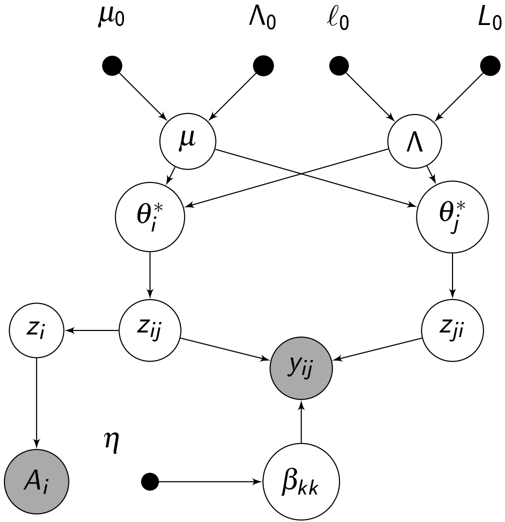


Figure: Graphical representation of network structure and behavior with influence

What We Do



What We Do in Practice



Implications of Joint modeling

- Improved community discovery with meaningful labels
- Improved predcition of quality of contagion(signal level vs. individual level)
- Relevance of social data in predicting behavior
- Identifying contagion from homophily

Our Contribution

- Scalable inference for large networked behavioral data
- Joint modeling explains the network processes better
- Other models that try to jointly model these two
    - either make strong parametric assumptions without accounting for latent homophily(Stochastic Actor Based model)
    - Rely only on observed characteristics and ignore latent homophily(Mathced sample estimation)
- Extensions:Allowing for dynamic modeling of the diffusion process(next project)

Dynamic Networks:Contagion-Driven Homophily

- Dynamic model of the trajectory of the latent space
  - there exists evidence that networks dynamically evolve
  - links can be formed or severed(evolution of bursts)
- Finding evidence of behavior affecting network positions
  - there exists evidence of change in preferences for tie formation when behavior changes(evolution of bursts)

Dynamic Networks:Contagion-Driven Homophily

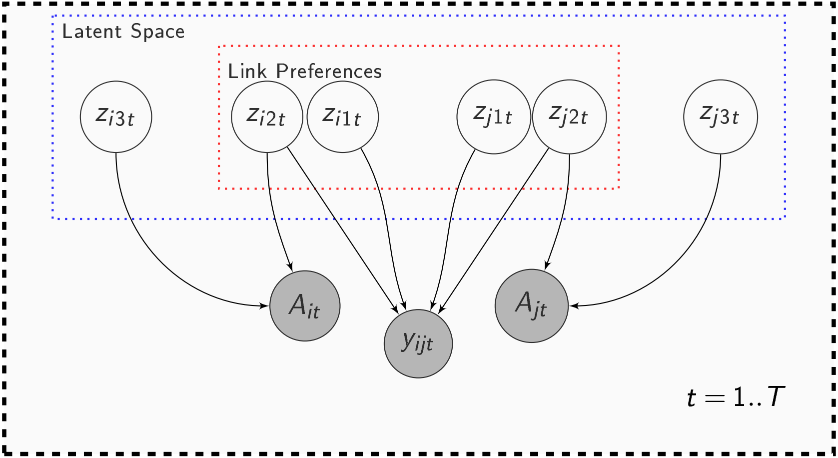


Figure: Dynamic representation of co-evolution

Our Contribution

- Finding evidence and possibly measuring how behavior affects tie preferences
- Better redefinition of opinion leadership at subnetwork level
- Better seeding strategies