STA6800 - Statistical Analysis of Network

LSM for Multiplex Networks

Ick Hoon Jin

Yonsei University, Department of Statistics and Data Science

Variational Inference

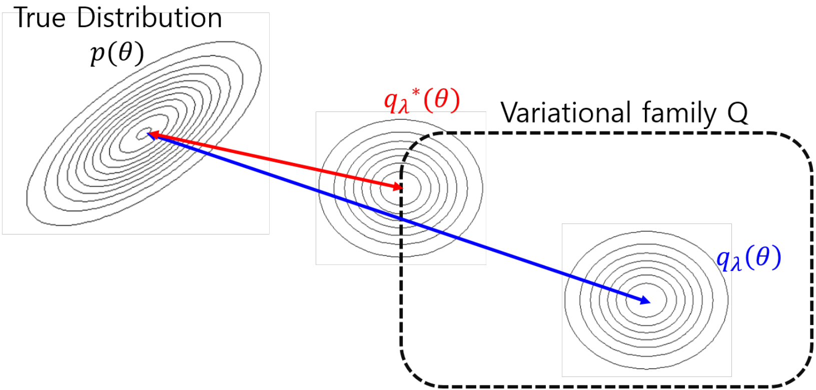


Figure: Variational Inference: Optimization for functions

Variational Inference

- compared to MCMC, biased but fast
- Kullback-Leibler divergence: Measure the divergence between two probability distributions

$$KL\left(q_\lambda(\theta|y)||p(\theta|y)\right) = \int q_\lambda(\theta|y)\ log\frac{q_\lambda(\theta|y)}{p(\theta|y)}d\theta \geq 0$$

- Specify a family of densities for $q_\lambda(\theta)$
- Find

$$q^*_\lambda = \underset{\lambda}{\text{argmin}}\ KL(q_\lambda(\theta|y)||p(\theta|y)) \Leftrightarrow q^*_\lambda = \underset{\lambda}{\text{argmax}}\ ELBO(\lambda)$$

where

$$ELBO(\lambda) = E_q\left[log\ p(\theta, y)\right] - E_q\left[log\ q_\lambda(\theta|y)\right]$$

EM algorithm

- GOAL : Maximize log-likelihood $l(\theta|Y)$ (i.e., Find $\hat{\theta}_{MLE}$)
- Assume that some data is missing
  - Filling the missing data is easy if we know θ
  - Estimating the θ is easy if we have all data

	X_1	X_2	...	X_p
obs_1		N.A.		
obs_2				N.A.
...				
obs_n	N.A.		N.A.	

Table: incomplete data: $Y_{com} = (Y_{obs}, Y_{mis})$

EM algorithm

- Two steps of EM algorithm:
  - Expectation Step (E-step)

$$Q(\theta|\theta^{(t)}) = E[l(\theta|Y_{obs}, Y_{mis})|Y_{obs}, \theta^{(t)}]$$

  - Maximization Step (M-step)

$$\theta^{(t+1)} = \text{argmax}_\theta Q(\theta|\theta^{(t)})$$

- Ascent property

$$l(\theta^{(t+1)}|Y_{obs}) \geq l(\theta^{(t)}|Y_{obs})$$

Variational EM algorithm

- E-step

$$q_\lambda^{(t+1)} = \underset{\lambda}{\operatorname{argmin}}\, KL(q_\lambda(\theta^{(t)}|y)||p(\theta^{(t)}|y))$$
$$Q(\theta|\theta^{(t)}) = E_{q^{(t+1)}}[log\ p(\theta|y)]$$

- M-step

$$\theta^{(t+1)} = \operatorname{argmax}_\theta Q(\theta|\theta^{(t)})$$

- $\mathcal{O}(N^2)$ but it converges in just a few iterations.
- Calculations performed in the estimation procedure are pretty simple.

What is Multiplex Network?

- Many real-world complex systems operate through multiple layers of distinct interactions among constinuents.
  - ex) living organisms, human society, transportation system, ...
- Multiplex network is a class of networks in which the same set of nodes are connected via more than one type of links.
- Multiplex network, Multiple network views, Multidimensional network, Multilayer network, ...

What is multiplex network?

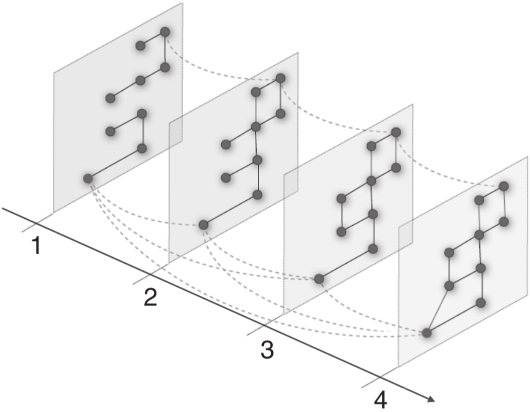


Figure: Multiplex network of nine nodes with four layers

What is multiplex network?

- Multiplex network is not just a simple stack of many network layers.
- Two important facets of multiplexity
  - The pattern of multiplexity: How the layers are coupled "structurally"
  - The type of multiplexity: How the layers are coupled "functionally"

Latent Space Joint Model

- Latent Space Model(LSM)
  - The posterior distribution of the LSM cannot be computed analytically
  - Variational EM algorithm
- Latent Space Joint Model(LSJM)
  - Multiple Network Views
  - $p(y_{ij} = 1|\eta_{ij})$ is explained by a unique latent variable z
  - Estimate using Variational EM algorithm

Latent Space Joint Model

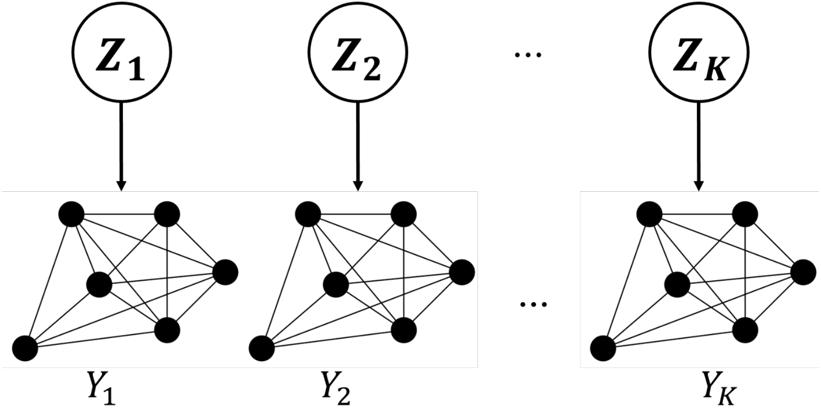


Figure: Latent Space Model

Latent Space Joint Model

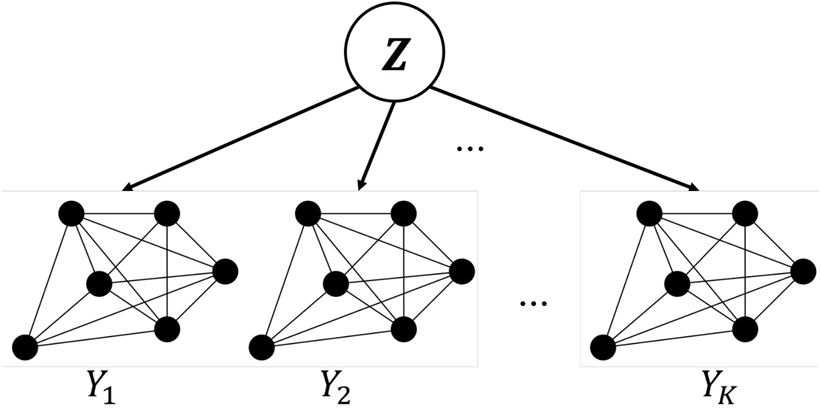


Figure: Latent Space Joint Model

Latent Space Model

- Latent Space Model

$$p(Y|Z,\alpha) = \prod_{i \neq j}^{N} p(y_{ij}|z_i, z_j, \alpha) = \prod_{i \neq j}^{N} \frac{\exp(\alpha - |z_i - z_j|^2)^{y_{ij}}}{1 + \exp(\alpha - |z_i - z_j|^2)}$$

  - $logodds(y_{ij} = 1|z_i, z_j, \alpha) = \alpha - |z_i - z_j|^2$
  - $p(\alpha) = \mathcal{N}(\xi, \psi^2)$
  - $p(z_i) \stackrel{iid}{=} \mathcal{N}(0, \sigma^2 I_D)$
  - $|z_i - z_j|^2 = \sum_{d=1}^{D} (z_{id} - z_{jd})^2$
- Advantages of using squared Euclidean distance measure
  - It allows one to visualize the data more clearly
  - It requires fewer approximation steps to be made in the estimation procedure

Latent Space Model: Variational Inference

- Posterior probability of the unknown (Z,α)

$$p(Z, \alpha | Y) = Cp(Y|Z, \alpha)p(\alpha) \prod_{i=1}^{N} p(z_i)$$

- Variational posterior

$$q(Z, \alpha | Y) = q(\alpha) \prod_{i=1}^{N} q(z_i)$$

  - $q(\alpha) = \mathcal{N}(\tilde{\xi}, \tilde{\psi}^2)$
  - $q(z_i) = \mathcal{N}(\tilde{z}_i, \tilde{\Sigma})$
- Kullback-Leibler

$$KL[q(\alpha)||p(\alpha)] + \sum_{i=1}^{N} KL[q(z_i)||p(z_i)] - E_{q(Z,\alpha|Y)}[log(p(Y|Z, \alpha)]$$

Latent Space Model: Variational Inference

- E-step: Estimate the parameters of the latent posterior distributions $\tilde{z}_i{}^{i+1}$ and $\tilde{\Sigma}^{i+1}$ evaluating

$$\tilde{\Sigma}^{(i+1)} = \frac{N}{2}\left[\left(\frac{N}{2\sigma^2} + 2\sum_{i=1}^{N}\sum_{j\neq i} y_{ij}\right) I + J(\tilde{\Sigma}^{(i)})\right]^{-1}$$

$$\tilde{z}_i{}^{(i+1)} = \left[\left(\frac{1}{2\sigma^2} + \sum_{j\neq i}(y_{ji} + y_{ij})\right) I + H\left(\tilde{z}_i{}^{(i)}\right)\right]^{-1} \times \left[\sum_{j\neq i}(y_{ji} + y_{ij})\tilde{z}_j - G(\tilde{z}_i{}^{(i)}) + H(\tilde{z}_i{}^{(i)})\tilde{z}_i{}^{(i)}\right]$$

Latent Space Model: Variational Inference

- M-step: Estimate the parameters of the posterior distribution of α evaluating

$$\tilde{\xi}^{(i+1)} = \frac{\xi + \psi^2 \left(\Sigma_{i=1}^{N} \Sigma_{j \neq i} y_{ij} - f'(\tilde{\xi}^{(i)}) + \tilde{\xi}^{(i)} f''(\tilde{\xi}^{(i)}) \right)}{1 + \psi^2 f''(\tilde{\xi}^{(i)})}$$

$$\tilde{\psi}^{2(i+1)} = \left(\frac{1}{\psi^2} + 2f'(\tilde{\psi}^{2(i)}) \right)^{-1}$$

Latent Space Joint Model

- Latent Space Joint Model

$$
\begin{aligned}
p(Y_1, \cdots, Y_K | Z, \alpha_1, \cdots, \alpha_K) &= \prod_{k=1}^{K} p(Y_k | Z; \alpha_k) \\
&= \prod_{k=1}^{K} \prod_{i \neq j}^{N} \frac{\exp(\alpha_k - |z_i - z_j|^2)^{y_{ijk}}}{1 + \exp(\alpha_k - |z_i - z_j|^2)}
\end{aligned}
$$

- $\log \text{odds}(y_{ijk} = 1 | z_i, z_j, \alpha_k) = \alpha_k - |z_i - z_j|^2$
- $p(\alpha_k) = \mathcal{N}(\xi_k, \psi_k^2)$
- $p(z_i) \overset{iid}{=} \mathcal{N}(0, \sigma^2 I_D)$
- $|z_i - z_j|^2 = \sum_{d=1}^{D} (z_{id} - z_{jd})^2$

Latent Space Joint Model: Variational Inference

- Posterior distribution of the z_i given K models

$$
\begin{aligned}
p(z_i | Y_1, \cdots, Y_k; \alpha_1, \cdots, \alpha_K) &\propto \frac{\prod_{k=1}^{K} p(z_i | Y_k; \alpha_k)}{p(z_i)^{K-1}} \\
&\propto \mathcal{N}(\bar{z}_i, \bar{\Sigma})
\end{aligned}
$$

- Approximate $p(z_i | Y_k; \alpha_k)$ with $q(zi) \sim \mathcal{N}(\tilde{z}_{ik}, \tilde{\Sigma}_k)$
- $z_i \sim \mathcal{N}(0, \sigma^2 I_D)$
- $\bar{\Sigma} = \left[\sum_{k=1}^{K} \tilde{\Sigma}_k^{-1} - \frac{K-1}{\sigma^2} I_D \right]^{-1}$
- $\bar{z}_i = \bar{\Sigma} \left[\sum_{k=1}^{K} \tilde{\Sigma}_k^{-1} \tilde{z}_{ik} \right]$

Latent Space Joint Model: Variational Inference

- E-step: Estimate the parameters $\bar{\Sigma}^{(i+1)}$ and $\bar{z}_i^{(i+1)}$ of the posterior distribution

$$\tilde{\Sigma}_k^{(i+1)} = \frac{N}{2}\left[\left(\frac{N}{2\sigma^2} + 2\sum_{i=1}^{N}\sum_{j\neq i} y_{ijk}\right) I + J(\bar{\Sigma}^{(i)})\right]^{-1}$$

$$\tilde{z}_{ik}^{(i+1)} = \left[\left(\frac{1}{2\sigma^2} + \sum_{j\neq i}(y_{jik} + y_{ijk})\right) I + H_k\left(\bar{z}_i{}^{(i)}\right)\right]^{-1}$$

$$\times \left[\sum_{j\neq i}(y_{jik} + y_{ijk})\bar{z}_j{}^{(i)} - G_k(\bar{z}_i{}^{(i)}) + H_k(\bar{z}_i{}^{(i)})\bar{z}_i{}^{(i)}\right]$$

:Can derive $\bar{\Sigma}^{(i+1)}$ and $\bar{z}_i^{(i+1)}$

Latent Space Joint Model: Variational Inference

- M-step: Update the model parameters evaluating

$$\tilde{\xi}_k^{(i+1)} = \frac{\xi_k + \psi_k^2\left(\Sigma_{i=1}^N \Sigma_{j\neq i} y_{ijk} - f_k'(\tilde{\xi}_k^{(i)}) + \tilde{\xi}_k^{(i)} f_k''(\tilde{\xi}_k^{(i)})\right)}{1 + \psi_k^2 f_k''(\tilde{\xi}_k^{(i)})}$$

$$\tilde{\psi}_k^{2(i+1)} = \left(\frac{1}{\psi_k^2} + 2f_k'(\tilde{\psi}_k^{2(i)})\right)^{-1}$$

Missing Link Data

- Missing (unobserved) links can be easily managed by the LSJM using the information given by all the network views.
- To estimate the probability of presence($y_{ijk} = 1$) of an edge, we employ the posterior mean of the α_k and of the latent positions.

$$y^*_{ijk} = p(y_{ijk} = 1 | \bar{\mathbf{z}}_i, \bar{\mathbf{z}}_j, \tilde{\xi}_k) = \frac{\exp\left(\tilde{\xi}_k - |\bar{\mathbf{z}}_i - \bar{\mathbf{z}}_j|^2\right)}{1 + \exp\left(\tilde{\xi}_k - |\bar{\mathbf{z}}_i - \bar{\mathbf{z}}_j|^2\right)}$$

- If we want to infer whether to assign $y_{ijk} = 1$ or $y_{ijk} = 0$, we need to introduce a threshold τ_k.

Evaluation

- To evaluate link prediction, we use 10-fold CV procedure in each network.
- We predict the links of each subset given the others fitting an LSM to each network independently and fitting the LSJM.
- In LSJM, a missing node in one network can be located by employing the information provided by the other network views.
- In LSM, otherwise, a missing node depends on only prior distribution of the nodes $p(\boldsymbol{z}_i)$.

Computational Aspects

- Prior:

$$p(\alpha) = N(0, 2)$$
$$p(\boldsymbol{z}_i) = N(0, I_2)$$

- Initialize the variational parameter

$$\tilde{\xi}_k = 0, \tilde{\psi}_k^2 = 2, \bar{\boldsymbol{z}}_i \text{ from } N(0, I_2), \bar{\Sigma} = I_2$$

Computational Aspects

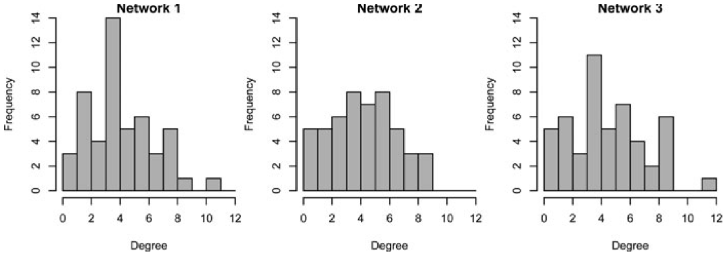
- The EM algorithm was stopped after at least 10 iterations when

$$||E_q[\log(p(Y|Z,\alpha))]^{(i+1)} - E_q[\log(p(Y|Z,\alpha))]^{(i)}|| < \text{tol}$$

- To evaluate the fit of the model, We produced ROC curve, AUC and the boxplots of the estimated link probabilities for both the true positive and the true negative links.

Teenage Friends and Lifestyle Study

- Data: Three directed networks about friendship relations between students.
  - Year : 1995, 1996, 1997
  - 50 girls that were present at all measurement points.
  - # links : 113(0.046), 116(0.047), 122(0.049)



Teenage Friends and Lifestyle Study

- $q(\alpha_1) = N(-0.63, 0.01), q(\alpha_2) = N(-0.66, 0.01), q(\alpha_3) = N(-0.48, 0.01)$
- Estimated latent posterior positions $\tilde{\boldsymbol{z}}_i$

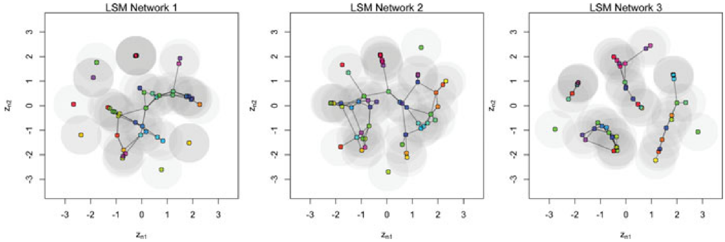


Figure: Latent positions in LSM

Teenage Friends and Lifestyle Study

- AUC : 89%, 97%, 98%

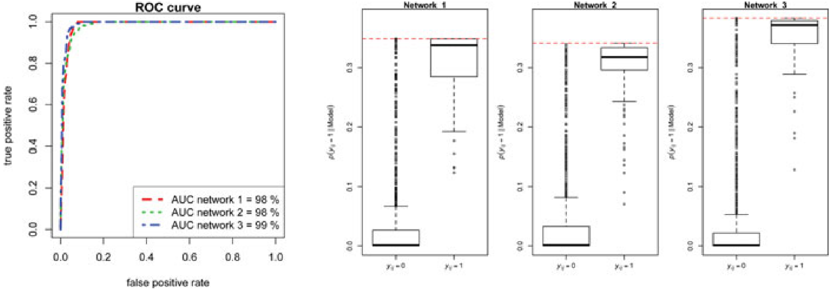


Figure: ROC, AUC, boxplot in LSM

Teenage Friends and Lifestyle Study

- $q(\alpha_1) = N(-0.42, 0.01), q(\alpha_2) = N(-0.39, 0.01), q(\alpha_3) = N(-0.32, 0.01)$

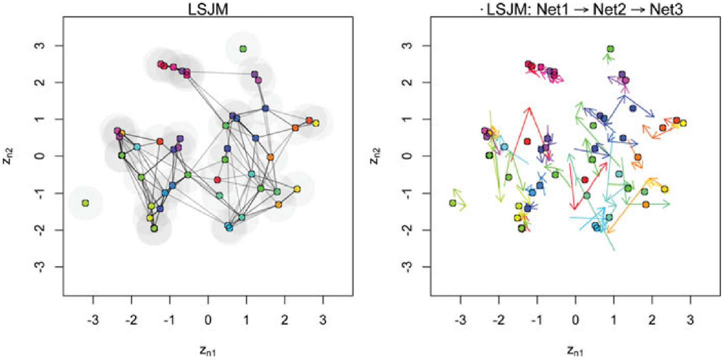


Figure: Latent positions in LSJM

Teenage Friends and Lifestyle Study

- The proximity between the disconnected components in network 3 depends on their previous relations.

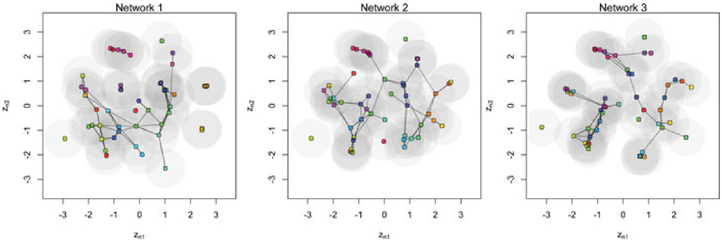


Figure: Latent positions in LSJM

Teenage Friends and Lifestyle Study

- AUC : 97%, 96%, 99%

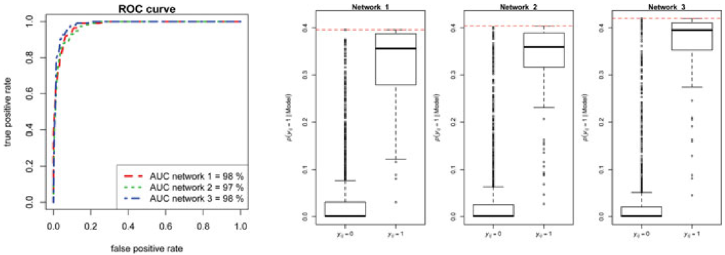


Figure: ROC, AUC, boxplot in LSJM

Teenage Friends and Lifestyle Study

- estimated # links : 109(0.044), 115(0.047), 118(0.048)

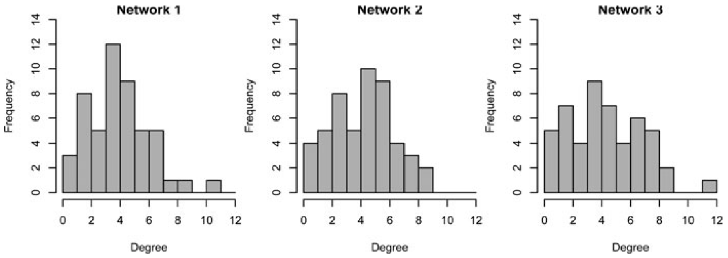


Figure: Estimated degree distribution in LSJM

Conclusion

- LSJM for multiple network views can describe their complex connectivity structure.
- LSJM merges the information given by multiple network views and can be explained by a unique latent position.
- Variational EM algorithm allows us to apply the LSJM to larger networks.

Introduction

- Goal: To capture the heterogeneity
  (heterogeneity : Global sparsity and Local density)
- Existing statistical models for social networks
  1. Carefully representing structure among actors that have a relatively high likelihood of interaction
  2. Clearly differentiating between groups of actors, that is, communities, within the graph that have high within-group connectivity and low between-group connectivity

$\Rightarrow$ Multiresolution Network Models

Existing Statistical Models for Networks

- LPCM (Latent Position Cluster Model): The probability of network ties is a function of the distance between actor positions in a latent space.
    - These distance calculations are computationally expensive and the propensity for actors in different groups to interact is often very small
    - The latent position for each node is heavily influenced by the numerous other nodes with which it has no relation

Existing Statistical Models for Networks

- Locally dependent ERGM (Exponential Random Graph Model): Local dependence on a graph as a decoupling of the dependence structure such that dependence exists only among ties within the same community and among ties between the same two communities
  - When partial communities are observed (i.e., some of the actors in various communities are not included in the sample), the parameter estimates from locally depen- dent ERGMs are difficult to interpret. Further, since the ties within each community are modeled using an ERGM
  - It is not possible to compare parameters across communities within the same graph unless the communities happen to be the same size

Multiresolution Network Model

The Class of Distribution over $\mathbf{Y}_N$

- $P_{\gamma,\alpha,\omega,N}(\mathbf{Y}_N) = \prod_{k=1}^{K_N} W_\alpha(\mathbf{Y}_{N,kk}) \prod_{k=1}^{K_N-1} \prod_{l=k+1}^{K_N} B_\omega(\mathbf{Y}_{N,kl})$

- γ = The community of actor i.
- W = The distribution depending on parameter α w.r.t within-community model
- B = The distribution depending on parameter ω w.r.t between-community model
- N = The number of actors

Multiresolution Network Model

Within- and Between-community Distributions as Mixture Distributions

- $W_\alpha(\mathbf{Y}_{N,kk}) = \int W(\mathbf{Y}_{N,kk}|\eta_k)dR_\alpha(\eta_k)$
- $B_\omega(\mathbf{Y}_{N,kl}) = \int B(\mathbf{Y}_{N,kl}|\tau_{kl})dS_\omega(\tau_{kl})$

- η_k = Within-community random effect for community k under random effect distribution R_α
- τ_{kl} = Between community random effect for community k and ℓ under random effect distribution S_ω
- $R_\alpha(.)$, $S_\omega(.)$ = Random effect for block-level probabilities

Latent Space Stochastic Block Model

- In the LS-SBM,
  - Within-block edges is modeled with a latent space model.
  - Between-block edges are modeled with a stochastic block model.

Description for model variables

- $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{K_N})$: vector of membership probabilities.
  - π_k : probability an actor belongs to block k.
- $\eta_k \equiv (\beta_k, \log \sigma_k)$: within-community random effect in community k.
  - β_k : maximum logit-probability of a relation in block k.
  - $\log \sigma_k$: measure of heterogeneity in block k.

Latent Space Stochastic Block Model

- Uses a latent Euclidean distance model.
- $Z_i \overset{\text{iid}}{\sim} N_D(0, \sigma_k^2 I_D)$; Latent positions in block k.
- $G(Y_{ij}|\beta_k, \mathbf{Z}_i, \mathbf{Z}_j) \sim Bernoulli(logit^{-1}(\beta_k - \| \mathbf{Z}_i - \mathbf{Z}_j \|))$

$\Rightarrow$ Edges in $\mathbb{Y}_{N,kk}$ are conditionally independent given $\mathbf{Z}_i$ and $\mathbf{Z}_j$ where $i, j \in S_k$.

1. Within-community distribution, W

$$W(\mathbb{Y}_{N,kk}|\eta_k) = \int \left(\prod_{i,j} G(Y_{ij}; \beta_k, \mathbf{Z}_i, \mathbf{Z}_j) \right) \times \left(\prod_i dN_D(\mathbf{Z}_i; 0, \sigma_k^2 I_D) \right)$$

where the products are taken with respect to all nodes in block k.

Latent Space Stochastic Block Model

- Use an Erdos–Renyi model.
- τ_{kl} : between-community random effect for k and l under random effect distribution S_ω

$\Rightarrow$ τ_{kl} is modeled as Beta distributed with parameters $\omega = (a_0, b_0)$.

$\Rightarrow$ $S_\omega(\tau_{kl}) \sim Beta(a_0, b_0)$

2. Between-community distribution, B

- All edges between communities k and l are IID with probability τ_{kl}.

$$B(\mathbb{Y}_{N,kl}|\tau_{kl}) = \prod_{i \in S_k} \prod_{j \in S_l} \tau_{kl}^{Y_{ij}} (1 - \tau_{kl})^{1 - Y_{ij}}$$

Prior Specification for α and ω

1. $\alpha = \{\boldsymbol{\mu}, \Sigma\}$: parameter of bivariate normal for within-block random effect η_k. $\Rightarrow R_\alpha(\eta_k) = R_\alpha((\beta_k, \log \sigma_k)) \sim N_2(\boldsymbol{\mu}, \Sigma)$
2. $\omega = \{a_0, b_0\}$: parameter of Beta distribution for between-block random effect τ_{kl}. $\Rightarrow S_\omega(\tau_{kl}) \sim Beta(a_0, b_0)$

- Set the prior on $\boldsymbol{\mu}$ and Σ to be a conjugate Normal-Inverse-Wishart distribution, with parameters $\{\mathbf{m}_0, s_0, \Psi_0, \nu_0\}$, subject to an additional assortativity restriction.

Prior Specification

- Given a_0 and b_0,

$$P(\boldsymbol{\alpha}|a_0, b_0, \mathbf{m}_0, s_0, \Psi_0, \nu_0) \propto N_2(\boldsymbol{\mu}; \mathbf{m}_0, \Sigma_0/s_0) \times Inv.Wish(\Sigma; \Psi_0, \nu_0)\mathbf{1}(a_0, b_0, \boldsymbol{\mu})$$

where $\mathbf{1}(a_0, b_0, \boldsymbol{\mu})$ is the indicator f_n enforcing the assortativity condition.

Assortativity Condition

The marginal probability of a within-community tie for the average block, induced by $\boldsymbol{\mu}$, be larger than the average between-block probability of a tie, $\frac{a_0}{(a_0+b_0)}$:

$$E\left[logit\left(Pr(Y_{ij}=1)\right)|\gamma_i=\gamma_j\right] \geq E\left[logit\left(Pr(Y_{ij}=1)\right)|\gamma_i \neq \gamma_j\right]$$

$\Rightarrow$ proceed with estimation using this global assortativity restriction.

Estimation Procedure

- Unlike the general estimation procedure using the MCMC algorithm, our model estimates using a two-stage procedure.
- Stage1. To estimate block membership K,

  $\Rightarrow$ use an assortative graph clustering algorithm.
- Stage2. To estimate parameters within each block,

  $\Rightarrow$ use variational inference.

Karnataka Villages

- ”Visit” relation data from Karnataka Villages (Banerjee et al., 2013)
  - undirected relationships between individuals and households in 75 villages
- Case1: from village 59, N = 293 households, K = 6 (cross-validation selection procedure)
- Case2: from all 75 villages, N = 13,009, K = 534
- Goal: estimate LS-SBM

Karnataka Villages: Case1

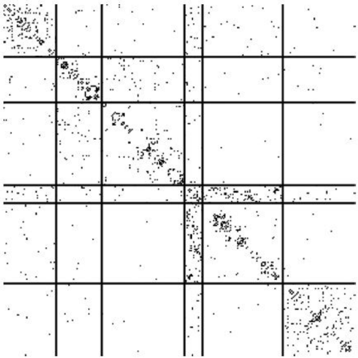


Figure: ”Visit” relation adjacency matrix of village 59, nodes are grouped by marginal posterior mode block membership

Karnataka Villages: Case1

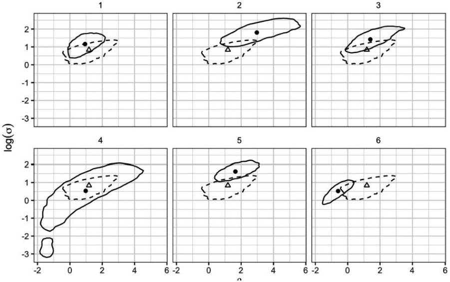


Figure: posterior mean and 95% highest posterior density region for a single block parameter η_k (β_k, $\log \sigma_k$) and global parameter (μ_1, μ_2)

- B2 : heterogeneity in tie probabilities are larger in B2
- B6 : smaller heterogeneity in tie probabilities

Karnataka Villages: Case1

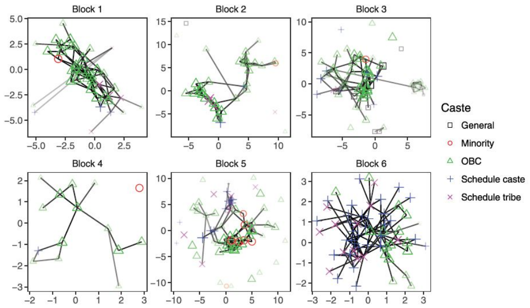
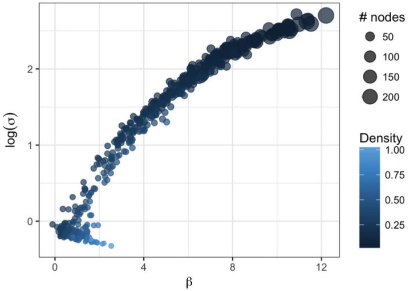


Figure: latent positions within each block

- Strong sorting by caste, almost all the members of schedule castes and schedule tribes (the lowest castes) are grouped into B6

Karnataka Villages: Case2

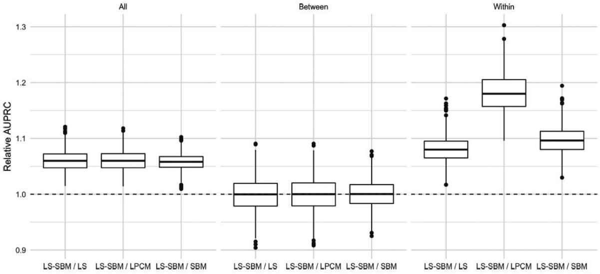


- estimate for all 75 villages, N = 13,009, K = 534
- no blocks containing households from multiple villages
- the number of nodes per block varied with median of 14, mean of 24.4, maximum of 233
- larger blocks tend to be sparser and more heterogeneous

Simulation Study

- LS-SBM model: binary, undirected network data, N = 300, K = 5 with equally sized
  - between-block tie probability = 0.2 or 0.02
  - within-block probabilities stemmed from heterogeneous set of two-dimensional block specific latent space
- 1000 simulations

Simulation Study



- LS-SBM outperforms LS, SBM and LPCM as relative the AUPRCs are greater than one
  - LS-SBM predicts edges between nodes in different blocks as well as other models
  - LS-SBM notably predicts edges between nodes within the same block

Projectivity of Multiresolution Network Models

Projectivity of Model Family

$$P_{\theta,N}(\mathbb{Y}_N) = P_{\theta,M}(\mathbb{Y}_N, \mathbb{Y}_{M\setminus N} \in \mathcal{Y}_{M\setminus N}),$$

where $\pi^{-1}_{M\mapsto N}(\mathbb{Y}_N)$ is the set of graphs on $\{1,2,\ldots,M\}$

- Projectivity: the sample can explain the characteristics of the population
- W_α and B_ω are projective: W are B are projective, R_α and S_ω do not depend on N.
- Multiresolution network models are projective.
- Latent space network model within-block tie in LS-SBM are also projective.

Conclusion

- Present multiresolution model for social network that are overall very sparse, but locally dense
- Use mixture of projective models to separately characterize tie structure within and between dense pockets in the graph
- Introduce LS-SBM as one example of multiresolution model
  - Latent space model representing with-in community relations can be replaced with LPCMs and so on
- projectivity allows to compare across communities within a graph, even if the communities are different size.