STA6800 - Statistical Analysis of Network

Latent Network Models

Ick Hoon Jin

Yonsei University, Department of Statistics and Data Science

Introduction

- Unlike ERGMs, this model introduces the notion of **the social space** in which unobserved latent characteristics represent potential transitive tendencies in network relations.
- It is assumed that each actor(or node) i has an unknown position z_i in social space. We call it the latent positions.

Introduction

Main assumption: **Conditional independence** of ties.

- The ties in the network are assumed to be conditionally independent given the latent position.
- The probability of a specific tie between two individuals is modeled as some function of their positions, such as the distance between two two actors in the social space.

Conditional Independence

- Data: *nxn* sociomatrix Y, with entries $y_{i,j}$ denoting the value of the relation from actor i to actor j.
- additional covariate information X.

$$P(Y|Z,X,\theta) = \prod_{i \neq j} P(y_{i,j}|z_i, z_j, x_{i,j}, \theta)$$

- where X and $x_{i,j}$ are observed characteristics and θ and Z are parameters and positions to be estimated.

Distance Models

- A conveniently parameterization of $P(y_{i,j}|z_i, z_j, x_{i,j}, \theta)$ is just the logistic regression model in which the probability of a tie depends on the Euclidean distance between z_i, z_j, where $z_i, z_j \in \mathbb{R}^k$

$$\eta_{i,j} = logodds(y_{i,j} = 1|z_i, z_j, x_{i,j}, \alpha, \beta) = \alpha + \beta^{'} x_{i,j} - |z_i - z_j|.$$

- Note that the $|z_i - z_j|'s$ could be replaced by any metric, satisfying the triangle inequality, $d_{i,j} \leq d_{i,k} + d_{k,j}$.

Distance Models

- The latent position model is inherently reciprocal and transitive.
- If $i \to j$ and $j \to k$, then $d_{i,j}$ and $d_{j,k}$ are probably not too large, making

1. the events $j \to i$ (reciprocity).
2. and $i \to k$. (transitivity)

Projection Models

- The distance model is inherently symmetric in that $p(i \to j) = p(j \to i)$. However, in many (directed) networks such symmetry is not achieved.
- For example, perhaps actor i sends a large number of ties, while j sends ties to a small subset of the actors receiving ties from i.
- So, variable levels of activity have to be modeled in the context of a latent position model which allows for probability transitivity in the relations, as well as individual specific levels of **social activity**.

Projection Models

- Let v_i be the unit-length k-dim. vector of characteristics of actor i.
- We might imagine that i and j are prone to having ties if the angle between them is small, neutral to having ties if the angle is a right angle, and averse to ties if the angle is obtuse.
- Letting $a_i > 0$ be the activity level of actor i, we can model the probability of a tie from i to j as $a_i v_i^{'} v_j$, or $z_i^{'} z_j / |z_j|$, where $z_i = a_i v_i$, then

$$\eta_{i,j} = logodds(y_{i,j} = 1 | z_i, z_j, x_{i,j}, \alpha, \beta) = \alpha + \beta^{'} x_{i,j} + \frac{z_i^{'} z_j}{|z_j|}.$$

Invariant

- Distances between a set points in Euclidean space are **invariant** under rotation, reflection, and translation.
- Therefore, for each *kxn* matrix of latent positions Z, there is an infinite number of other positions giving the same log-likelihood.
- Let $[Z]$ be the class of positions equivalent to Z under rotation, reflection, and translation. For each $[Z]$, there is one set of distances between the nodes. We call this class of positions a *configuration*.
- Similarly, Z at the projection models is invariant under rotation and reflection of positions, but not under translation.

Estimation

- The log-likelihood of a conditional independence model is simple,

$$logP(Y|\eta) = \sum_{i \neq j} \{\eta_{i,j} y_{i,j} - log(1 + e^{\eta_{i,j}})\}$$

Step1

- Use Metropolis-Hastings steps to sample z_j' for each j. Propose z_j' from a proposal distribution $\varphi(\cdot)$ and accept with probability equal to

$$r_z(z_j', .z_j^{(t)}) = \frac{\pi(z_j'|Y, \alpha, \beta)}{\pi(z_j^{(t)}|Y, \alpha, \beta)} \frac{\varphi(z_j' \to z_j^{(t)})}{\varphi(z_j^{(t)} \to z_j')}$$

- Similary, use Metropolis-Hasting steps to sample α, β

Estimation

Step2

- post-process the MCMC samples using Procrustes matching

1. To find a reference set of latent positions, pick out the latent positions Z_0 in MCMC samples that achieved the highest value of the full log posterior density
2. apply Procrustes matching to each of the MCMC samples, using the Z_0

$$Z^* = \text{argmin}_{TZ} tr(Z_0 - TZ)'(Z_0 - TZ)$$

Monk Data

- interpersonal relations among 18 monks divided into 4 groups
- directed relation
- stochastic block model
  - SBM assumes that probabilities of tie within a group are same, and probabilities of tie between groups are also same.
- use distance model in $\mathbb{R}^2$

Monk Data

Table 1. Model Fitting Results for the Monk Data

Model	*Maximized log-likelihood*	*Number of parameters*
Distance model ($\Re^3$)	−34.04	50
Distance model ($\Re^2$)	−66.02	34
Stochastic blockmodel	−82.12	37

- distance model takes advantage of transitivity and can achieve a better fit than stochastic block model, using fewer parameters and not presuming the a priori existence of distinct groups
- the fit cannot be improved by going into much higher dimensions

Monk Data

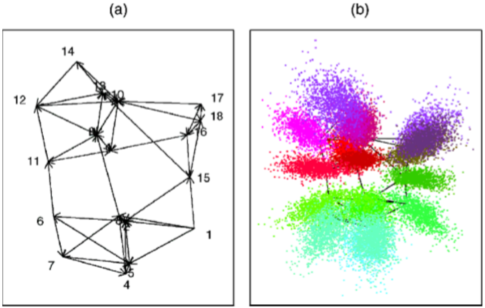

Figure 1. Maximum Likelihood Estimates (a) and Bayesian Marginal Posterior Distributions (b) for Monk Positions. The direction of a relation is indicated by an arrow.

Classroom Data

- friendship ties between 13 boys and 14 girls in a sixth-grade classroom
- directed relation
- levels of social activity of actors vary considerably
- projection model
  - no covariate : $logit(p_{i,j}) = \alpha + z_i' z_j / |z_j|$
  - one covariate : $logit(p_{i,j}) = \alpha + \beta x_{i,j} + z_i' z_j / |z_j|$

    where $x_{i,j}$ is homophily statistics about sex.

Classroom Data

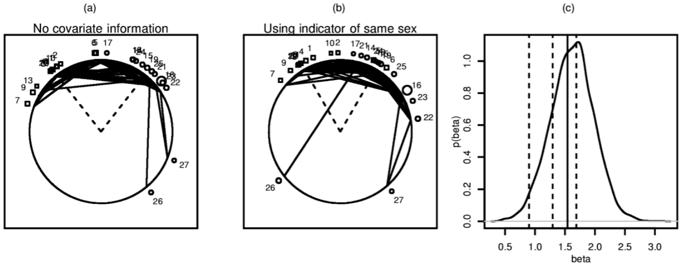

Figure 5. Maximum Likelihood Estimates of Student Positions (a) No covariate information; (b) using indicator of same sex, and (c) the posterior of β.

- single covariate $x_{i,j}$ does not fully explain the different rates of within-sex and between-sex friendship ties
- ”full” model would have different baseline rates for the four different types of ties

Advantages

- provide a visual and interpretable model-based spatial representation of network relationships
- It is flexible and can be easily generalized to allow for multiple relationships, ties with varying strengths, and time-varying relations
- deal easily with missing data, at least if information on ties is missing at random
- the model is inherently transitive, and so we can expect an improved fit over models lacking such structure when the relations are transitive in nature

Latent Position Cluster Model

- Latent Position Cluster Model

$$log \frac{P(y_{i,j} = 1|z_i, z_j, x_{i,j}, \beta)}{1 - P(y_{i,j} = 1|z_i, z_j, x_{i,j}, \beta)} = \beta_0^T x_{i,j} - \beta_1 |z_i - z_j|$$

$$z_i \sim \sum_{g=1}^{G} \lambda_g MVN_d(\mu_g, \sigma_g^2 I_d), \quad \text{where} \quad \sqrt{\frac{1}{n} \sum_i |z_i|^2} = 1$$

and λ_g is the proportion of an individual distribution.

- Likelihood

$$P(Y|Z, X, \beta) = \prod_{i \neq j} P(y_{i,j}|z_i, z_j, x_{i,j}, \beta)$$

Bayesian Estimation

Fully Bayesian Estimation Procedure

1. Specify prior distributions of β, λ_g, μ_g, and σ_g^2 (model parameters).
2. Specify full conditional posterior distributions of $\mathbf{z}_i$, β, λ, μ_g, σ_g^2, and K_i.
3. MCMC Algorithm

Bayesian Estimation

1. Prior distributions of β, λ_g, μ_g, and σ_g^2 (model parameters)

$$
\begin{aligned}
\beta &\sim \mathrm{MVN}_p\Big(\xi, \Psi\Big), \\
\lambda &\sim \mathrm{Dirichlet}\Big(\nu\Big), \\
\sigma_g^2 &\sim \sigma_0^2 \mathrm{Inv}\text{-}\chi^2_\alpha, \quad g = 1, \cdots, G \\
\mu_g &\sim \mathrm{MVN}_d\Big(0, \omega^2 I_d\Big), \quad g = 1, \cdots, G.
\end{aligned}
$$

Bayesian Estimation

2. Full conditional posterior distribution of $\mathbf{z}_i$, β, λ, μ_g, σ_g^2, and K_i.

$$\mathbf{z}_i \mid K_i = g, \text{others} \sim \phi_d\left(\mathbf{z}_i; \mu_g, \sigma_g^2 I_d\right) P\left(\mathbf{Y} \mid \mathbf{Z}, \mathbf{X}, \beta\right), \quad i = 1, \cdots, n,$$

$$\beta \mid \mathbf{Z}, \text{others} \sim \phi_p\left(\beta; \xi, \Psi\right) P\left(\mathbf{Y} \mid \mathbf{Z}, \mathbf{X}, \beta\right),$$

$$\lambda \mid \text{others} \sim \text{Dirichlet}\left(m + \nu\right),$$

$$\mu_g \mid \text{others} \sim \text{MVN}_d\left(\frac{m_g \mathbf{z}_g}{m_g + \sigma_g^2/\omega^2}, \frac{\sigma_g^2}{m_g + \sigma_g^2/\omega^2}\mathbf{I}\right), \quad g = 1, \cdots, G,$$

$$\sigma_g^2 \mid \text{others} \sim \left(\sigma_0^2 + d s_g^2\right) \text{Inv-}\chi^2_{\alpha + m_s d}, \quad g = 1, \cdots, G,$$

$$P\left(K_i = g \mid \text{others}\right) = \frac{\lambda_g \phi_d\left(z_i; \mu_g, \sigma_g^2 \mathbf{I}_d\right)}{\sum_{r=1}^{G} \lambda_r \phi_d\left(z_i; \mu_r, \sigma_r^2 \mathbf{I}_d\right)}, \quad g = 1, \cdots, G.$$

Bayesian Estimation

Step 1. Update Z_{t+1} using Metropolis-Hastings steps

1. Proposal $\mathbf{z}_i^* \sim MVN_d(\mathbf{z}_{it}, \delta_Z^2 \mathbf{I}_d)$.
2. Accept $\mathbf{z}_i^*$ as the i-th element of $\mathbf{z}_{t+1}$ with probability

$$\frac{P\left(\mathbf{Y}|\mathbf{Z}^*, \mathbf{X}, \beta_t\right)\phi_d\left(\mathbf{Z}_i^*; \mu K_i, \sigma_K^2 \mathbf{I}_d\right)}{P\left(\mathbf{Y}|\mathbf{Z}_t, \mathbf{X}, \beta_t\right)\phi_d\left(\mathbf{Z}_{it}; \mu K_i, \sigma_K^2 \mathbf{I}_d\right)}$$

Bayesian Estimation

Step 2. Update β_{t+1} using Metropolis-Hastings steps

1. Propose $\beta^* \sim MVN_d(\beta_t, \delta_\beta^2 I_p)$
2. Accept β^* as β_{t+1} with probability

$$\frac{P\left(\mathbf{Y}|\mathbf{Z}_{t+1}, \mathbf{X}, \beta^*\right)\phi_p\left(\beta^*; \xi, \Psi\right)}{P\left(\mathbf{Y}|\mathbf{Z}_{t+1}, \mathbf{X}, \beta_t)\phi_p(\beta_t; \xi, \Psi\right)}.$$

Bayesian Estimation

Step 3. Update λ, μ_g, σ_g^2, K_i

Update λ, μ_g, σ_g^2, K_i using full conditional posterior distributions

Pros and Cons

- Advantage: Perform better
- Disadvantage: More Complicated

Identifiability of Positions and Cluster Labels

Problem: Non-identifiabilities of positions and cluster labels

Likelihood is invariant to

1. Reflections, rotations, translations of the latent positions
2. Relabelling of the clusters $\rightarrow$ Label switching problem

Label switching problem

Permuting the cluster labels does not change the likelihood but we have a problem in clustering the observations into groups

Identifiability of Positions and Cluster Labels

How to resolve? $\rightarrow$ Minimizing Bayes risk

1. Find the positions of the actors that minimize the estimated Bayes risk
2. Procrustes transform the posterior draws of latent positions and using the same transformation matrix, transform the cluster means and covariances
3. Find the cluster membership probabilities of the actors that minimize the estimated Bayes risk

- Bayes risk = Expecation of loss function
- Use a Kullback-Leibler loss as a loss function

Choosing the Number of Clusters

Bayesian estimation

Select the model with the highest posterior probability

$$P(Y, \hat{Z}|G) = \int P(Y|\hat{Z}, X, \beta)p(\beta)d\beta \int P(\hat{Z}|\theta)p(\theta)d\theta$$

$$\int P(Y|\hat{Z}, X, \beta)p(\beta)d\beta = \textit{integrated likelihood for the logistic regression}$$

$$\int P(\hat{Z}|\theta)p(\theta)d\theta = \textit{integrated likelihood for the mixture model}$$

$$\int P(Y|\hat{Z}, X, \beta)p(\beta)d\beta \approx BIC_{lr}(\textit{logistic regression})$$

$$\int P(\hat{Z}|\theta)p(\theta)d\theta \approx BIC_{mbc}(\textit{mixture model})$$

$$BIC = BIC_{lr} + BIC_{mbc}$$

Choosing the Number of Clusters

BIC for the logistic regression

$$BIC_{lr} = 2log[P\{Y|\hat{Z}, X, \hat{\beta}(\hat{Z})\}] - d_{logit} log(n_{logit})$$

d_{logit}=number of parameters in the logitic regression

n_{logit}=number of ties in data

BIC for the mixture model

$$BIC_{mbc} = 2log[P\{\hat{Z}|\hat{\theta}(\hat{Z})\}] - d_{mbc} log(n)$$

d_{mbc}=number of parameters in the clustering model

Example 1. Linking between Monks

- Social relations between 18 monks in an isolated American monastery
- A monk has the social relation of ‘like’ to another monk if he ranked that monk in the top three monks for positive affect in any of three interviews given over a 12-month period.
- 3 main groups: the Young Turks (7), the Loyal Opposition (5), the Outcasts(3) + Waverers(3)
- MCMC sampling with 5000 burn-in iterations and a further 30000 iterations, of which we kept every 30th value.
- Clear choice of three clusters

Example 1. Linking between Monks

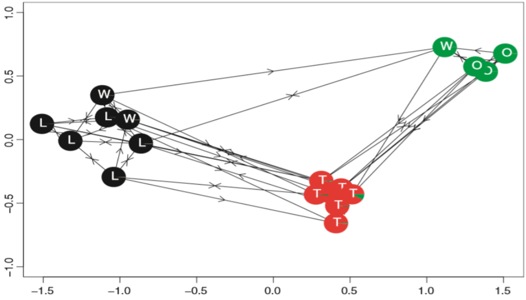

Figure: Latent position cluster model

- Bayesian estimate of the latent position cluster model produces greater distinctions between the groups and firmly identifies the grouping of the Waverers.

Example 2. Adolescent Health

- Social network for the adolescent health data
- 69 adolescents
- Each respondent was asked to nominate up to five boys and five girls within the school whom they regarded as their best friends
- Groups : grades 7–12
- MCMC sampling with 50000 burn-in iterations and a further 2 million iterations, of which we kept every 1000th value
- Clear choice of 6 clusters

Example 2. Adolescent Health

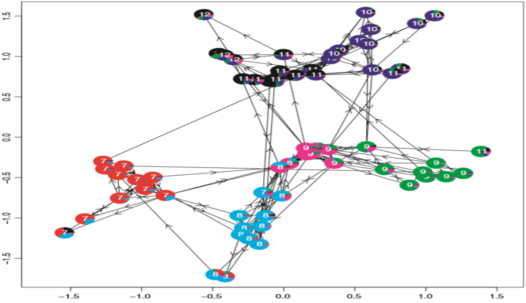

Figure: Latent position cluster model

- The clusters based on the bayesian method are well defined in terms of both their positions in space and their correspondence to the grades

Example 2. Adolescent Health

Grade	*Results for the following clusters:*					
	1	*2*	*3*	*4*	*5*	*6*
7	13	1	0	0	0	0
8	0	11	1	0	0	0
9	0	0	7	6	3	0
10	0	0	0	0	3	7
11	0	0	0	0	3	10
12	0	0	0	0	0	4

Figure: Latent position cluster model

- Within-grade cohesion gets weak as students move up in the school
- Students form links increasingly based on common interests and personal affinity and less on the grade as they gain seniority.

Summary

- Takes account of transitivity, homophily on attributes and clustering simultaneously
- Simultaneous estimation
- Good at estimating the latent positions and the number of groups
- Slow & Complicate