STA6800 - Statistical Analysis of Network

LSM for Dynamic Networks

Ick Hoon Jin

Yonsei University, Department of Statistics and Data Science

Introduction

- Network can change as time goes

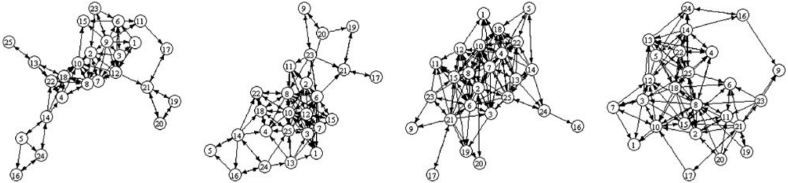


Figure 5. Graphs of Dutch classroom data at, from left to right, times 1, 2, 3, and 4.

- Goal: Modeling dynamic network data into a latent Eucidean space, allowing each actor to have a temporal trajectory in the latent space.

Dynamic Latent Space Model

- Notation
  - n : the number of actors
  - p : dimension of Euclidean latent space
  - X_{it} : p-dimensional vector of the ith actor's latent position at time t
  - X_t : n x p matrix whose ith row is X_{it}
  - $Y_t = \{y_{ijt}\}$: adjacency matrix of the observed network at time t
  - Ψ : model parameters(will be described soon)

Dynamic Latent Space Model

- The latent actor positions are modeled by a Markov process

$$\pi(X_1|\Psi) = \Pi_{i=1}^{n} N(X_{i1}|0, \tau^2 I_p),$$
$$\pi(X_t|X_{t-1}, \Psi) = \Pi_{i=1}^{n} N(X_{it}|X_{i(t-1)}, \sigma^2 I_p).$$

- Hidden Markov Model
  - The system which is modeled by two processes Y(obervable) and X(unobervable, assumed to be a Markov process).
  - Assume that behavior of Y depends on X.
  - $P(Y_n \in A|X_1 = x_1, .., X_n = x_n) = P(Y_n \in A|X_n = x_n)$
  - The goal is to learn about X by observing Y

Dynamic Latent Space Model

- The observed networks at different time points are conditionally independent given the latent positions.

Figure 1. Illustration of the dependence structure for the latent space model. Y_t is the observed graph, $\mathcal{X}_t$ is the unobserved latent actor positions, and ψ is the vector of model parameters.

Dynamic Latent Space Model

- Formulation

$$
\begin{aligned}
P(Y_t|X_t,\Psi) &= \prod_{i\neq j} P(y_{ijt}=1|X_t,\Psi)^{y_{ijt}} * P(y_{ijt}=0|X_t,\Psi)^{1-y_{ijt}} \\
&= \prod_{i\neq j} \frac{exp(y_{ijt}\eta_{ijt})}{1+exp(\eta_{ijt})},
\end{aligned}
$$

where

$$
\eta_{ijt} := \log\left(\frac{P(y_{ijt}=1|X_t,\Psi)}{P(y_{ijt}=0|X_t,\Psi)}\right) = \beta_{IN}\left(1-\frac{d_{ijt}}{r_j}\right) + \beta_{OUT}\left(1-\frac{d_{ijt}}{r_i}\right)
$$

Dynamic Latent Space Model

- $d_{ijt} = \|X_{it} - X_{jt}\|$ and model parameter $\Psi = (\tau^2, \sigma^2, \beta_{IN}, \beta_{OUT}, r_{1:n})$.
- β_{IN} and β_{OUT} are global parameters which reflects the importance of popularity and social activity respectively.
- r_i's are each actor's social reach which reflects the tendency to form and receive edges. Constrained to $\sum_{i=1}^{n} r_i = 1$.

Dynamic Latent Space Model

- Estimation : Repeat steps 1-6
  1. For t = 1,..., T and for i = 1,...,n, draw X_{it} via MH using a normal random walk proposal
  2. Draw τ^2 from its full conditional inverse gamma distribution.
  3. Draw σ^2 from its full conditional inverse gamma distribution.
  4. Draw β_{IN} via MH using a normal random walk proposal.
  5. Draw β_{OUT} via MH using a normal random walk proposal.
  6. Draw $r_{1:n}$ via MH using a Dirichlet proposal.

Dynamic Latent Space Model

- Letting $p_{ijt} := P(y_{ijt}|X_t, \Psi)$, the conditional distribution for *Xit* is

$$\pi(X_{it}|Y_{1:T}, \Psi) \propto \begin{cases} \prod_{j:j\neq i} p_{ijt}p_{jit})N(X_{it}|0, \tau^2 I_p)N(X_{i(t+1)}|X_{it}, \sigma^2 I_p), & \text{if } t = 1 \\ \prod_{j:j\neq i} p_{ijt}p_{jit})N(X_{i(t+1)}|X_{it}, \sigma^2 I_p)N(X_{it}|X_{i(t-1)}, \sigma^2 I_p), & \text{if } 1 \leq t \leq T \\ \prod_{j:j\neq i} p_{ijt}p_{jit})N(X_{it}|X_{i(t-1)}, \sigma^2 I_p & \text{if } t = T \end{cases}$$

Missing Data

- This paper focuses on non-responses, that is, missing edge values.
- Let $\mathcal{D}$ denote the sampling pattern.
- $\mathcal{D}$ is the set of $n \times n$ matrices $\{D_1, ..., D_T\}$.
- $D_{ijt} = \begin{cases} 1, & \text{if the dyad is observed} \\ 0, & \text{otherwise} \end{cases}$
- $\mathcal{Y}^{(mis)} = \begin{cases} \text{MCAR}, & \text{if } \mathbb{P}(\mathcal{D} \mid \mathcal{Y}^{(obs)}, \mathcal{Y}^{(mis)}, \xi) = \mathbb{P}(\mathcal{D} \mid \xi) \\ \text{MAR}, & \text{if } \mathbb{P}(\mathcal{D} \mid \mathcal{Y}^{(obs)}, \mathcal{Y}^{(mis)}, \xi) = \mathbb{P}(\mathcal{D} \mid \mathcal{Y}^{(obs)}, \xi) \end{cases}$

Missing Data

- Our posterior distribution is $\pi(\mathcal{X}_{1:T}, \Psi, \mathcal{Y}^{(mis)} \mid \mathcal{Y}^{(obs)}, \mathcal{D})$.
- If the sampling pattern is ignorable, we may make inference based on the posterior distribution like this, $\pi(\mathcal{X}_{1:T}, \Psi, \mathcal{Y}^{(mis)} \mid \mathcal{Y}^{(obs)})$.
- There are two sufficient conditions that must be satisfied in order for the sampling pattern to be ignorable.
  1. $\pi(\mathcal{Y}^{(mis)}, \mathcal{Y}^{(obs)}, \mathcal{X}_{1:T}, \blacksquare, \xi) = \pi(\mathcal{Y}^{(mis)}, \mathcal{Y}^{(obs)}, \mathcal{X}_{1:T}, \blacksquare)\pi(\xi)$
  2. The space of $(\xi, \mathcal{X}_{1:T}, \Psi)$ is a product space, that is, if $\xi \in \Xi, \mathcal{X}_{1:T} \in \boldsymbol{\mathcal{X}}, \Psi \in \Psi$, then, $(\xi, \mathcal{X}_{1:T}, \Psi) \in \Xi \times \boldsymbol{\mathcal{X}} \times \Psi$

Missing Data

- Revise MH within Gibbs sampling scheme for handling the missing data.
- Using the observed data and the current values for the missing data, the full conditionals for $\mathcal{X}_{1:T}$ and Ψ are unchanged.
- The full conditional of $\mathcal{Y}^{(mis)}$ is determined by $\pi(y_{ijt} = 1 \mid \mathcal{X}_{1:T}, \Psi) = 1/(1 + \exp(-\eta_{ijt}))$ for any $y_{ijt} \in \mathcal{Y}^{(mis)}$
- Additional draw for each missing y_{ijt} from a Bernoulli distribution with probability determined previously.

Prediction

- Predicting, for time $T+1$, the adjacency matrix Y_{T+1} and the latent space positions $\mathcal{X}_{T+1}$ is of interest.
- Simple prediction:
  - $\widehat{\mathcal{X}}_{T+1} := \mathbb{E}(\mathcal{X}_{T+1} \mid Y_{1:T}) \approx \frac{1}{L}\sum_{\ell=1}^{L} \mathcal{X}_T^{(\ell)}$
  - A point estimate of $\mathbb{P}(y_{ij(T+1)} = 1)$ can be computed by plugging in $\widehat{\mathcal{X}}_{T+1}$ along with the posterior means of the parameters into the observation equation.

Prediction

- Better prediction
  - Eliminate unnecessary uncertainty by not conditioning on the posterior means of the model parameters.
  - Derive and use the marginal distribution,

$$\mathbb{P}(Y_{ij(T+1)} \mid Y_{1:T}, \widehat{\mathcal{X}}_{T+1}) \approx \sum_{\ell=1}^{L} w_\ell \pi(y_{ij(T+1)} \mid \widehat{\mathbf{X}}_{i(T+1)}, \widehat{\mathbf{X}}_{j(T+1)}, \blacksquare^{(\ell)})$$

- This method outperforms the simpler method because of using fewer estimated parameters to make predictions, hence introducing less uncertainty into the predictions estimates.

Edge Attraction

- Edge Attraction
  - How one actor affects the edges of another actor.
  - This attraction is manifested in an increased tendency for the influenced actor to move in the direction of the influencing actor in the social space.

Edge Attraction

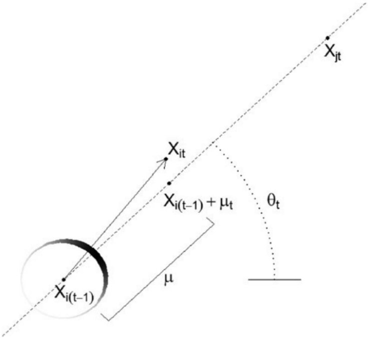


Figure 3. The extension of the transition equation to allow for actor j's influence on actor i. Actor i is more likely to move *toward* actor j. The circle around $\mathbf{X}_{i(t-1)}$ represents a von Mises distribution for the angle component of ϵ_{it}'s polar coordinates, where dark values indicate high probability regions and light values indicate low probability regions.

Detection of Edge Attraction

- Consider extension of the transition equation.
  - $\mathbf{X}_{it} = \mathbf{X}_{i(t-1)} + \epsilon_{it}$ where $\epsilon_{it} \sim N(\boldsymbol{\mu}_t, \sigma^2 I_p), p = 2$
- $\theta_t = \text{atan2}(\mathbf{X}_{jt} - \mathbf{X}_{i(t-1)})$ from "2-argument arctangent".
- $\boldsymbol{\mu}_t = \mathcal{R}_t \begin{pmatrix} \mu \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\theta_t) & -\sin(\theta_t) \\ \sin(\theta_t) & \cos(\theta_t) \end{pmatrix} \begin{pmatrix} \mu \\ 0 \end{pmatrix}$,

  where μ is some unknown parameter taking nonnegative values.
- $\mu = \begin{cases} 0 & \text{if no edge attraction} \\ > 0 & \text{if exist some edge attraction} \end{cases}$
- Under the extended transition eqation, the Markov property still holds for the latent positions, $\pi(\mathcal{X}_t \mid \mathcal{X}_{1:(t-1)}, \Psi, \mu) = \pi(\mathcal{X}_t \mid \mathcal{X}_{t-1}, \Psi, \mu)$.

Detection of Edge Attraction

- The prior distribution of μ is

$$\pi(\mu) = \begin{cases} p_0 & \text{if } \mu = 0 \\ (1 - p_0) f(\mu) & \text{for } \mu > 0 \end{cases} \text{, where } f \sim \exp(\lambda)$$

- For notation, let

$$\begin{aligned} \pi_0(\mu = 0 \mid Y_{1:T}) &= \pi(Y_{1:T} \mid \mu = 0) p_0 / \pi(Y_{1:T}) \\ \pi_+(\mu \mid Y_{1:T}) &= \pi(Y_{1:T} \mid \mu)(1 - p_0) f(\mu) / \pi(Y_{1:T}) \end{aligned}$$

Detection of Edge Attraction

- Since

$$1 = \pi_0(\mu = 0 \mid Y_{1:T}) + \int_0^\infty \pi_+(\mu \mid Y_{1:T}) d\mu$$

$$\pi_0(\mu = 0 \mid Y_{1:T}) = \frac{1}{1 + \int_0^\infty \kappa(\nu) d\nu}$$

where $\kappa(\nu) = \pi_+(\mu = \nu \mid Y_{1:T}) / \pi_0(\mu = 0 \mid Y_{1:T})$

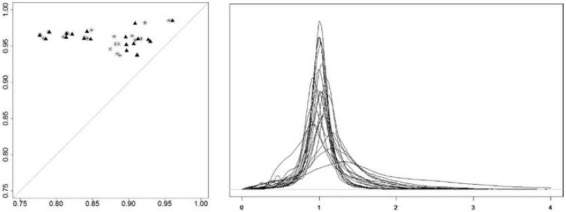
Simulations

- 20 data set
- For each data set: $n = 100$, $T(timepoints) = 10$
- Setting: $\beta_{in} = 1$, $\beta_{out} = 2$,
- $r_{1:n}$: randomly drawn from Dirichlet distribution.
- For 10 of the 20 simulations, 25 actors were randomly selected to be influenced, each of which was accompanied by another randomly selected actor to do the influencing
- For remaining 10 simulations No edge attraction.

Simulations

	Mean(sd) over 20 simulation
$\widehat{\beta_{in}}$	0.9172(0.06207)
$\widehat{\beta_{out}}$	2.045(0.1438)
$corr(\widehat{r_{1:n}}, r_{1:n}^{true})$	0.9298(0.06402)

Simulations



(a) AUC (b) Distances

Figure 4. Results for 20 simulations. (a) AUC using Sarkar and Moore's method (horizontal axis) and our method (vertical axis) on both undirected (triangles) and directed (asterisks) networks; (b) Distribution of pairwise distance ratios, comparing estimated latent positions with true latent positions.

Simulations

- Bayesian estimation does a very good job at detecting edge attraction without giving many false positives when no such influence exists.
  - Mean of specificity: 0.082(with edge attraction) vs 0.868(without edge attraction)
- The estimation is robust to the hyperparameters for the prior of σ^2.
  - $\pi \sim U(3, 15), \quad \sigma^2 \sim U(0, 01, 2)$
  - Averaging AUC = 0.9621
- By using the approximations, there is a drastic decrease in computational time with very little loss in model fit.
  - Mean(sd) decrease in computational time of 68.6%(0.835)

Real Data Analysis: Dutch Classroom Data

- A longitudinal study in which 26 students aged 11 to 13 years in a Dutch class were surveyed over **four time points**
- There are four asymmetric adjacency matrices where the $(i, j)th$ entry denotes whether student i claims student j as a friend.
- n=25 ($\because$ one student failed to complete the study)
- Missing edges exist.

Real Data Analysis: Dutch Classroom Data

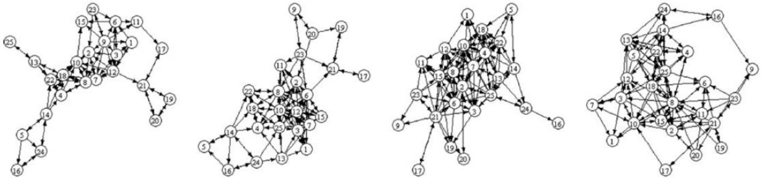
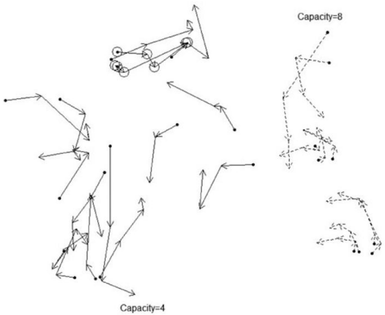


Figure 5. Graphs of Dutch classroom data at, from left to right, times 1, 2, 3, and 4.

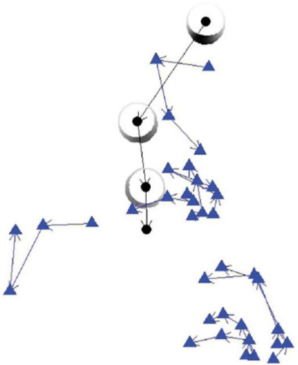
Real Data Analysis: Dutch Classroom Data

- AUC = 0.917
- The posterior means of β_{in} and β_{out} : $1.29 > 1$



Real Data Analysis: Dutch Classroom Data

- AUC = 0.917
- The posterior means of β_{in} and β_{out} : $1.29 > 1$



Conclusion

- Rich visualization of the dynamics of the network
    - insight into the characteristics of the actors
    - the overall groupings
    - communities that exist within the network
- Handling directed edges, missing data
- Predict future latent positions and future edges,
- Detect and visualize edge attraction
- Approximation method
- This model can be easily generalized to dyadic data types other than binary (by changing the link function).