ST5225: Statistical Analysis of Networks
Lecture 11: Latent Space Model

WANG Wanjie
staww@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

Saturday 14 April, 2018

Review: ERGM

- Model
- Edge prob., MLE
- Example: $p1$ model

Note: the content in Lecture 11 is not covered in the final exam.

Recall our examples for ERGM:

- Random graph model
- Stochastic block model
- many other models...

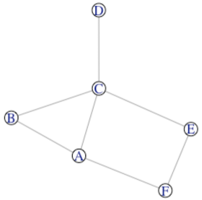
Question. Are there any structure that ERGM cannot catch?

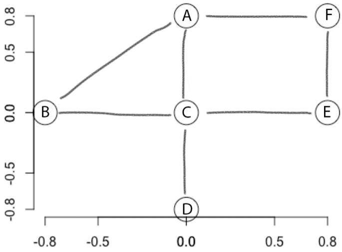
Answer. Yes!

Question: What if the structure in a network is being driven by some 'simpler? form of relationship?

- For stochastic block model, we assume the edges are dependent on the community labels
- Now we generalize the community labels to a broader notion: each node can be represented by a point in the some space, and the edges are dependent on these points
- Given the underlying points, the edges are independent with each other.
- Without the info. of these points, the edges "seems" to be dependent

Latent Space Model





- Left: a network. Right: Place the network in a 2-dim space
- In this 2-dim space, $dist(i, j) < 1$ if $(i, j) \in E$, and $dist(i, j) > 1$ if $(i, j) \neq E$ with one exception.
- The positions of the nodes are the underlying factors that the network is formed
- The positions of each node is unknown

These points are unobserved.

- Recall that for the stochastic block model, it is quite usual that the community labels are unknown, and it is hard to estimate.
- For general models, the underlying points are also unobserved (latent)
- So we call it as a latent space model
- How do we model the network with these latent nodes?
- What is the space? Is it 1-dim? 2-dim? or even higher dimensions?
- How to figure out the representation of each node?

Model assumptions

How do we model the network if we have the latent points?

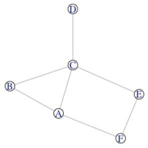
- We are trying to model the edges A_{ij}, which is either 0 or 1
- The probability of A_{ij} is related to the latent representation for nodes i and j, denoted by $Z_i = (z_1^{(i)}, z_2^{(i)}, \cdots, z_d^{(i)})$ and $Z_j = (z_1^{(j)}, z_2^{(j)}, \cdots, z_d^{(j)})$.
- When Z_{ij}'s are not given, the probability for A_{ij} is related to other edges
- When Z_i's are given, the probability is related to the distance between Z_i and Z_j

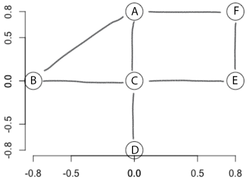
$$\log \left[\frac{P(A_{ij} = 1 | Z_i = z_i, Z_j = z_j)}{P(A_{ij} = 0 | Z_i = z_i, Z_j = z_j)} \right] = \alpha - \|z_i - z_j\|,$$

where α is a parameter to estimate

Remark

- Note that $\log\left[\frac{P(A_{ij}=1|Z_i=z_i,Z_j=z_j)}{P(A_{ij}=0|Z_i=z_i,Z_j=z_j)}\right] z = \alpha - \|z_i - z_j\|$. When $\|z_i - z_j\|$ increases, $\text{logit} P(A_{ij} = 1|Z_i, Z_j)$ decreases, and the probability for connection decreases.
- This model satisfies that *the probability for two nodes to be connected is smaller when the latent distance between them is larger*.
- Recall our example





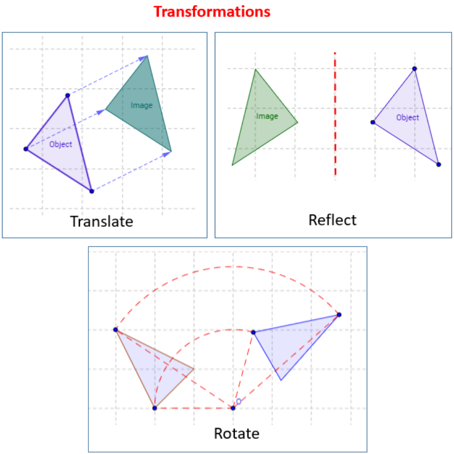
- For this example, when the Euclidean distance between z_i and z_j is smaller than 1, the probability of connection is 1. When $dist(z_i, z_j) > 1$, the probability of connection is small (but not 0).

Identification

Can Z be identified (unique)?

- Why don't we take the function as $\alpha - \beta\|z_i - z_j\|$?
  - If we multiply all the z_i's by the same scalar r, and multiply β by $1/r$, the result does not change
  - Therefore, to make sure that z_i's are unique, we fix $\beta = 1$.
- The Z_i's are still not identified:
  - Nothing changes if rotate all the Z_i's the same way
  - Or if translate all the Z_i's along the same vector
  - Or if reflect all the Z_i's about the same plane
  - Or combine rotations, translations and reflections

Isometry



- Isometry = translations which leave all distances the same
  - For Euclidean space, isometry group built from rotations, translations and reflections
- The Z_i's are "identified up to isometry"

Likelihood function

- Given Z_i's, we have that

$$P(A_{ij} = 1 | Z_i = z_i, Z_j = z_j) = \frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}}$$

So, the joint density function is the likelihood function

$$\begin{aligned} L(z_1, \cdots, z_{|V|}, \alpha) &= \prod_{i,j} [\frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}}]^{A_{ij}} [\frac{1}{1 + \exp\{\alpha - \|z_i - z_j\|\}}]^{1 - A_{ij}} \\ &= \prod_{i,j} [\frac{\exp\{A_{ij}[\alpha - \|z_i - z_j\|]\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}}] \\ &= \frac{\exp\{\sum_{i,j} A_{ij}[\alpha - \|z_i - z_j\|]\}}{\prod_{ij}[1 + \exp\{\alpha - \|z_i - z_j\|\}]} \end{aligned}$$

- Note that it does not belong to ERGM. If we let $\theta_{ij} = \|z_i - z_j\|$, then the statistics are the adjacency matrix A (all the details are required), and θ_{ij}'s are not independent. They have restrictions.

Likelihood function

- Given Z_i's, we have that

$$P(A_{ij} = 1 | Z_i = z_i, Z_j = z_j) = \frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}}$$

So, the joint density function is the likelihood function

$$\begin{aligned} L(z_1, \cdots, z_{|V|}, \alpha) &= \prod_{i,j} [\frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}}]^{A_{ij}} [\frac{1}{1 + \exp\{\alpha - \|z_i - z_j\|\}}]^{1 - A_{ij}} \\ &= \prod_{i,j} [\frac{\exp\{A_{ij}[\alpha - \|z_i - z_j\|]\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}}] \\ &= \frac{\exp\{\sum_{i,j} A_{ij}[\alpha - \|z_i - z_j\|]\}}{\prod_{ij}[1 + \exp\{\alpha - \|z_i - z_j\|\}]} \end{aligned}$$

- Note that it does not belong to ERGM. If we let $\theta_{ij} = \|z_i - z_j\|$, then the statistics are the adjacency matrix A (all the details are required), and θ_{ij}'s are not independent. They have restrictions.

MLE

- Assume that the latent variables are "observed", then the likelihood function is about α only.
- The log-likelihood function is

$$l(\alpha) = \sum_{i,j} A_{ij}[\alpha - \|z_i - z_j\|] - \sum_{i,j} \log[1 + \exp\{\alpha - \|z_i - z_j\|\}]$$

- Take the derivative w.r.t. 0, and let it equals to 0,

$$\frac{dl(\alpha)}{d\alpha} = \sum_{i,j} A_{ij} - \sum_{i,j} \frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}} = 0.$$

  It can be found there is a unique solution for this equation, and that is the MLE.
- However, no explicit formula
- What's more, z_i's are unknown....

MLE

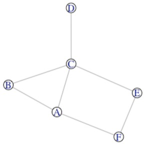
- Assume that the latent variables are "observed", then the likelihood function is about α only.
- The log-likelihood function is

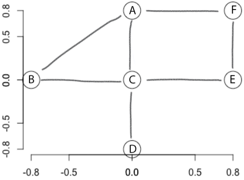
$$l(\alpha) = \sum_{i,j} A_{ij}[\alpha - \|z_i - z_j\|] - \sum_{i,j} \log[1 + \exp\{\alpha - \|z_i - z_j\|\}]$$

- Take the derivative w.r.t. 0, and let it equals to 0,

$$\frac{dl(\alpha)}{d\alpha} = \sum_{i,j} A_{ij} - \sum_{i,j} \frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}} = 0.$$

  It can be found there is a unique solution for this equation, and that is the MLE.
- However, no explicit formula
- What's more, z_i's are unknown....



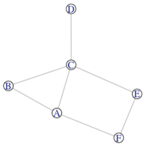


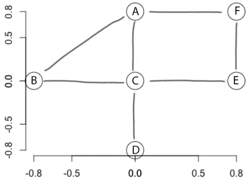
Assume the network on the left follows latent variable model, where the link function is

$$P(A_{ij} = 1 | Z_i = z_i, Z_j = z_j) = \frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}}.$$

Assume it is known that the latent points for the nodes are $z_A = (0, 0.8)$, $z_B = (-0.8, 0)$, $z_C = (0, 0)$, $z_D = (0, -0.8)$, $z_E = (0.8, 0)$, $z_F = (0.8, 0.8)$. What is the MLE for α?

Example, II





There are 6 nodes, so 15 pairs of nodes in this network. According to the formula for likelihood function, we have

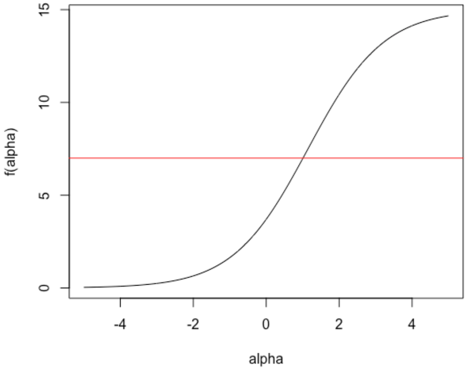
$$L(\alpha) = \frac{\exp\{\sum_{i,j} A_{ij}(\alpha - \|z_i - z_j\|)\}}{\prod_{i,j}[1 + \exp\{\alpha - \|z_i - z_j\|\}]},$$

and we should solve the equation

$$\sum_{i,j} \frac{\exp\{\alpha - \|z_i - z_j\|\}}{1 + \exp\{\alpha - \|z_i - z_j\|\}} = \sum_{i,j} A_{ij} = 7.$$

Example, III

Define $f(\alpha) = \sum_{i,j} \frac{\exp\{\alpha - \|z_i - z_j\|\}}{1+\exp\{\alpha - \|z_i - z_j\|\}}$, we want to find the intersection of it and the horizontal line $y = 7$.



- The numerical solution is $\hat{\alpha} = 1.01$.
- The prob. given is 0.0002050468

Estimate

- For random graph model, SBM and ERGM, we calculate MLE
- For this latent space model, MLE is hard to calculate
- We further assume the latent points Z_i's also follow some distribution, and include that in the likelihood. In this way, z_i's are random variables, not the parameters to estimate
- With data, we are interested in the corrected distribution (posterior) for Z_i's, and also the estimate of Z_i's and α

- Now we assume a prior distribution for the latent points $Z_i \in \mathcal{R}^d$

$$Z_i \overset{i.i.d}{\sim} N(0, I_d), \qquad i = 1, 2, \cdots, |V|,$$

where I_d is the $d \times d$ identity matrix, and $N(0, I_d)$ denotes the multivariate normal distribution.

- Multivariate Normal distribution: We say $\mathbf{X} \sim N(\mu, \Sigma)$, where $\mathbf{X} = (X_1, X_2, \cdots, X_d)'$, $\mu = (\mu_1, \mu_2, \cdots, \mu_d)$ is the mean vector, and Σ is a $d \times d$ covariance matrix.

  With these parameters, the joint density is

$$f(x_1, x_2, \cdots, x_d) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\},$$

  where $\mathbf{x} = (x_1, x_2, \cdots, x_d)$.

- For this example, $f(x_1, x_2, \cdots, x_d) = (2\pi)^{-d/2} \exp\{-\frac{1}{2}(\mathbf{x}'\mathbf{x})\}$.

Joint distribution

- To figure out the joint dist., usually we assume the latent space is known, say, $d = 2$.
- Therefore, we have the joint dist.,

$$P(A) = [\frac{\exp\{\sum_{i,j} A_{ij}[\alpha - \|\mathbf{z_i} - \mathbf{z_j}\|]\}}{\prod_{i,j}[1 + \exp\{\alpha - \|\mathbf{z_i} - \mathbf{z_j}\|\}]}] \prod_i [(2\pi)^{-d/2} \exp\{-\frac{1}{2}(\mathbf{z_i}'\mathbf{z_i})\}]$$

- The only unknown parameter is α, and obviously it does not belong to ERGM.
- To solve the problem, we may
  - For given α, find the posterior and the estimate for Z_i's
  - With the estimation of Z_i's, find the MLE of α
  - Repeat until convergence

Generalization

Generalization on the link function:

- Recall that we link the probability with the function

$$\text{logit}P(A_{ij} = 1|z_i, z_j, \alpha) = \alpha - \|z_i - z_j\|.$$

  This is the most basic one. According to the data, we may consider some other choices.

- For some data sets, such as our toy example, the nodes can be embedded in a way that the relationship between $dist(Z_i, Z_j)$ and 1 largely impacts the probability of connection. For this case, a proper function is

$$\text{logit}P(A_{ij} = 1|z_i, z_j, \alpha) = \alpha(1 - \|z_i - z_j\|).$$

- Projection methods. For some data sets, it can be found that node i is connected to many nodes, and node j is connected to a small subset of the neighbors of node i. In this case, we model that both i and j are "similar" but i is more "socially active".

$$\text{logit}P(A_{ij} = 1|z_i, z_j, \alpha) = \alpha + \frac{z_i' z_j}{\|z_j\|}.$$

Generalization

Additional covariate information:

- Some networks contain the covariate information matrix X. The covariate information may also impact the probability of connection.
- In this case, we model it as

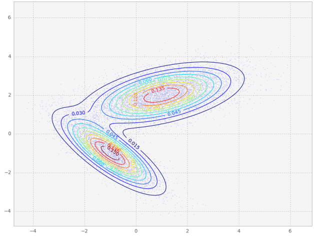
$$\text{logit}P(A_{ij} = 1 | z_i, z_j, \alpha) = \alpha + \beta' x_{i,j} - \|z_i - z_j\|,$$

  where $x_{i,j}$ is the covariate information.
- Of course, the covariate information can be added to other linkage functions:

$$\text{logit}P(A_{ij} = 1 | z_i, z_j, \alpha) = \alpha + \beta x_{i,j} + \frac{z_i' z_j}{\|z_j\|}.$$

Prior distribution:

- In our model, we assume the prior dist. for Z_i's are multivariate normal dist.
- Combine it with the community recovery problem, then we assume the latent points also have community structure

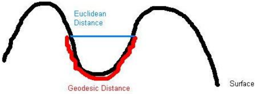


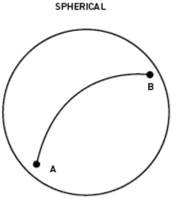
- The latents points follow a mixture normal dist., where

$$Z_i \sim \sum_{k=1}^{K} \lambda_k N(\mu_k, \sigma_k^2 I_d), \qquad \sum_{k=1}^{K} \lambda_k = 1.$$

Generalization on the distance measurement:

- In the model, we consider the Euclidean distance between two latent points.
- With the projection method, we consider the projection of one latent vector on the other latent vector
- There are more possibilities:
  - Geodesic distance: for points on some surfaces, other than the Euclidean distance, there is also the geodesic distance, which measures the distance on the surface, e.g., the airlines





- The latent space is a manifold

Summary

- Latent space model assume latent points for every node. The prob. of connection depends on the latent points only.
- The model based on the latent points varies. We mainly discuss the distance model
- MLE is hard to calculate; latent points are hard to get
- Bayesian approach
- Generalizations