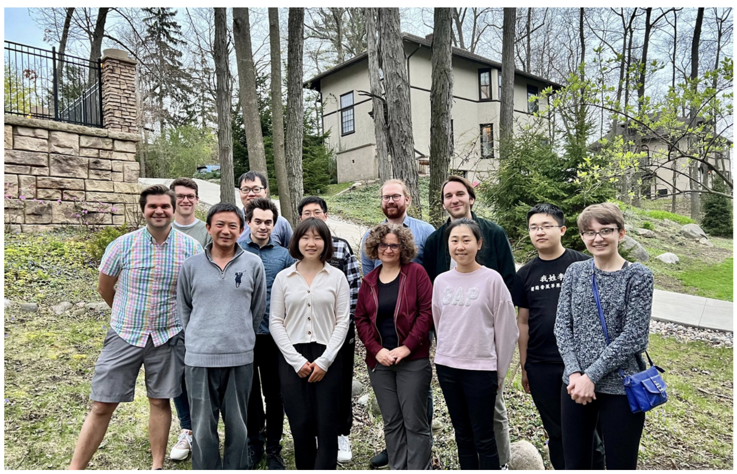
The Ji and Liza Group Super Reading List

2025-07-15

Welcome!



Introduction

Hello! This website was made for the Liza/Ji research group. It will primarily serve as a collection of works to help new students navigate the literature and get up to speed on the workings of the group. In each of the panels to the side, you will see a different topic. For each of these chapters, you will find a selection of works coupled with a brief snippet describing the significance of the paper and how it may pertain to the work of new students. We also try here to supply a narrative describing how each of the works fit into a larger picture.

If you have questions, don't hesitate to reach out to Ji, Liza, or another member of the group. If you find anything on this website broken or lacking for any reason, don't hesitate to contact Chandler (chandle at umich dot edu).

Have fun exploring!

Group Expectations

Group Fun/Memories

Latent Space Modeling

What is it?

In the context of statistical network analysis, latent space modeling refers to the assignment of each node in the network to a latent position vector. Then, the nodes of the network connect with some probability dependent on the positions of the latent position vectors. Each of the following are latent space network models, roughly arranged from least complex to most complex. As a result, we encourage you to read the papers presented *in order*.

The Basics

Hoff et al

You should begin by reading the seminal work *Latent Space Approaches to Social Network Analysis* Hoff, Raftery, and Handcock (2002) here.

This paper views models networks as the ties (edges) between individuals (nodes). In particular, the paper presents a model wherein the probability of a tie between individuals depends on the positions of individuals in some unobserved "social space". This social space corresponds to the latent space described above. In particular, the paper presents the distance model

$$\eta_{ij} = \log \text{odds}(y_{ij} = 1 | z_i, z_j, x_{ij}, \alpha, \beta) = \alpha + \beta' x_{ij} - |z_i - z_j|$$

and a projection model

$$\eta_{ij} = \log \text{odds}(y_{ij} = 1 | z_i, z_j, x_{ij}, \alpha, \beta) = \alpha + \beta' x_{ij} + \frac{z_i' z_j}{|z_j|},$$

where the x_{ij} correspond to covariates and the z_i correspond to the positions of the nodes in the latent space.

These models are then fit using Procrustes analysis and Markov Chain Monte Carlo (don't worry too much about this last part, just try to understand as many details as possible).

The Stochastic Block Model (Holland et al)

Another foundational notion is the **Stochastic Block Model**. The first paper on the subject is *Stochastic Blockmodels: First Steps* Holland, Laskey, and Leinhardt (1983) here.

The fundamental notion of the stochastic block model is very simple. You begin with a graph with n vertices. Partition these vertices into r communities (we might call these $C_1, \ldots, C_r$). Then, for any two vertices $u \in C_i$ and $v \in C_j$, the two are connected by an edge with probability P_{ij}. Notice that the probability of connection is dependent *only* on the community assignment. Doing this for all vertices, we can build a symmetric $r \times r$ matrix P of edge probabilities. When $P_{ij} = p$ for all i, j, then the result is the famous Erdos-Renyi model.

Random Dot Product Graphs (Athreya et al)

Another foundational paper is *Statistical Inference on Random Dot Product Graphs: a Survey* Athreya et al. (2017) here.

As in Hoff, in the Random Dot Product Graph (RDPG), the latent position vectors are drawn from some common distribution F. In particular, we might have $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{iid}}{\sim} F$. That is, we have a graph with n nodes, each of which is associated with a position in the latent space. We then collect these rows and put them into a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\top \in \mathbb{R}^{n \times d}$. Then the adjacency matrix $\mathbf{A}$ is modeled by

$$\text{Prob}[\mathbf{A}|\mathbf{X}] = \prod_{i<j} (\mathbf{x}_i^\top \mathbf{x}_j)^{A_{ij}} (1 - \mathbf{x}_i^\top \mathbf{x}_j)^{1-A_{ij}}$$

In this case, we write $(\mathbf{A}, \mathbf{X}) \sim \text{RDPG}(F, n)$.

Latent Space Modeling with Multiplex Networks

Arroyo et al, COSIE

MacDonald et al, MultiNeSS

Peter MacDonald was another of Liza's students. For one chapter of his dissertation (in fact, his first project for the group), he created the **Multi**ple **Ne**tworks with **S**hared **S**tructure model. As in the case with COSIE, we imagine that we have multiple graphs (a sort of layering of graphs, wherein each layer is one graph) on a common set of n nodes. Let the graphs be labeled by $k = 1, \ldots, K$. As in the previous section, each of the nodes of the graph are associated with a latent position vector. That means that each layer (each graph in the set of graphs) is associated with its *own* latent matrix that we denote by X_k. The key idea is that the latent position matrix for each layer of the multiplex is divided into two pieces:

$$X_k = \begin{bmatrix} x_{1k}^\top \\ \vdots \\ x_{nk}^\top \end{bmatrix} = \begin{bmatrix} V & U_k \end{bmatrix}$$

That is, **all** of the corresponding nodes in the multiplex have latent vectors with some common component (captured by V) and some component individual to that layer (captured by U_k). We can then separate the information that is common to all layers from the information that is unique to individual layers. This paper is very important and very dense; it should be read multiple times!

The paper containing these ideas is *Latent space models for multiplex networks with shared structure* MacDonald, Levina, and Zhu (2021) here.

Community Detection

What is it?

The Basics

Hypergraphs

What is it?

The Basics

Matrix Completion

What is it?

The Basics

George Linderman's Matrix Completion Notes

Candes and Recht

Candes and Tao

Fithian and Mazumder

Connectomics

What is it?

The Basics

Paper Database

Database

In this section, the group has amassed a database of papers that you can search by topic, title, and so on.

Athreya, Avanti, Donniell E. Fishkind, Keith Levin, Vince Lyzinski, Youngser Park, Yichen Qin, Daniel L. Sussman, Minh Tang, Joshua T. Vogelstein, and Carey E. Priebe. 2017. "Statistical Inference on Random Dot Product Graphs: A Survey." https://arxiv.org/abs/1709.05454.

Hoff, Peter D, Adrian E Raftery, and Mark S Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97 (460): 1090–98. https://doi.org/10.1198/016214502388618906.

Holland, Paul W, Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. "Stochastic Blockmodels: First Steps." *Social Networks* 5 (2): 109–37.

MacDonald, Peter W., Elizaveta Levina, and Ji Zhu. 2021. "Latent Space Models for Multiplex Networks with Shared Structure." https://arxiv.org/abs/2012.14409.