STA6800 - Statistical Analysis of Network

Advanced Topics in LSM

Ick Hoon Jin

Yonsei University, Department of Statistics and Data Science

Euclidean Latent Distance Model

- The most intuitive specification of the latent space model is Euclidean latent distance model.

$$\begin{aligned} \text{log-odds}(y_{ij} = 1) &= \alpha - d(z_i - z_j) \\ d(z_i - z_j) &= \mid z_i - z_j \mid \\ &= \sqrt{(z_i - z_j)'(z_i - z_j)} \end{aligned}$$

where $z_i, z_j \in \mathbb{R}^t$, $z_i \sim f_t(z \mid \phi)$

Euclidean Latent Distance Model

- Add the additive individual random effects act as sociality of popularity effect

$$\text{log-odds}(y_{ij} = 1) = a_i + b_j - d(z_i - z_j)$$

- Pros: easy visualization and interpretation of estimated latent position.
- Limit: It is not sure that this geometry is best to represent complex dependencies.

Projection Model

- Projection model

$$\text{log-odds}(y_{ij} = 1) = \alpha + \frac{v_i^{'} v_j}{\mid v_j \mid},$$

$$v_i \in \mathbf{R}^t$$

- Reparameterizing with alternative latent vectors, $v_i = a_i z_i, v_j = a_j z_j$

$$\text{log-odds}(y_{ij} = 1) = \alpha + a_i z_i^{'} z_j,$$

$$z_i \in \text{t-dimensional hypershpere}$$

where $a_i > 0$ is a sociality parameter for the ith individual.

Hypersphere Latent Space Model

- Then, this projection model is example of latent distance model? $\Rightarrow$ No!!!
- Similarity measure in projection model is the “cosine similarity” between the latent position vectors.
- projection term $(z_i^{'} z_j)$ preserves the relationship between the probability of tie and some measure of “closeness”.

Ultrametric Latent Space

- latent distance model in ultrametric space does not specify Z^t(latent metric space) as a traditional metric space.
- An ultrametric space, or non-Arcimedean space, is a metric space where additionally satisfies the strong triangle inequality.

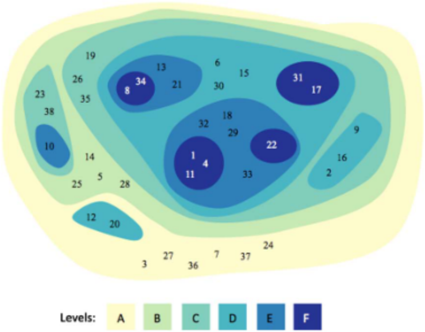
$$d(x,z) \leq \max\{d(x,y), d(y,z)\}$$

- Ultrametric space can be visualized via dendrograms or trees.

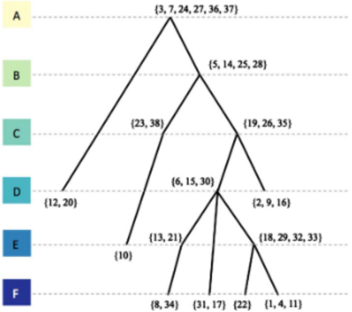
Properties of Ultrametric Latent Space

- All triangles are isosceles with the unequal side being shortest.
- Every point in a dic is a center of that disc, so that two discs intersect only if one disc completely contains the other.
  (where disc: $B_r(x_0) = \{x : d(x, x_0) \leq r\}$)
- Visualizing and understanding the distance can be difficult.

Position in a Latent Ultrametric Space



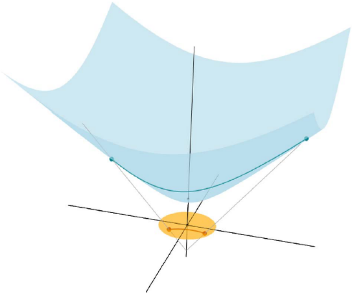
"Topographical" Map



Dendrogram

Hyperbolic Space

- Recent work in latent space models for network data: the use of negatively curved t-dimensional **hyperbolic space,** $\mathbb{H}^t$.



- Parallel postulate: Given any straight line and a point not on it, there are at least two distinct lines passing through that point which do not intersect with the line.

Hyperbolic Space

- 2D hyperbolic space, $\mathbb{H}^2$ satisfy $z^2 - y^2 - x^2 = 1$ with $z > 0$.
- Distance between points: distance along the surface of the hyperboloid
- A generic latent distance model in two-dimensional hyperbolic space:

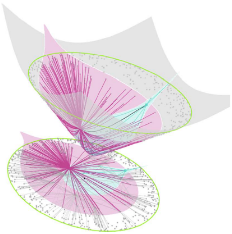
$$Y_{ij} \overset{\text{ind}}{\sim} \text{Bernoulli}(p_{ij}),$$
$$\text{logit}(p_{ij}) = \alpha - d(z_i, z_j),$$
$$z_i \in \mathbb{H}^2, \qquad z_i \overset{\text{ind}}{\sim} f_2(z|\psi)$$

$$\begin{aligned} d(z_i, z_j) &= d((r_i, \phi_i), (r_j, \phi_j)) \\ &= acosh\{cosh(r_i)cosh(r_j) - sinh(r_i)sinh(r_j)cos(\Delta\phi)\} \end{aligned}$$
$$\Delta\phi = \pi - |\pi - |\phi_i - \phi_j||$$

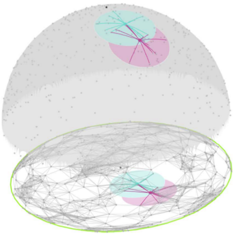
Hyperbolic Space

- Motivation: Many real world networks tend to be tree-like, which the geometry of a tree behaves much like the geometry of hyperbolic space.
- Trees need an exponential amount of space for branching and only hyperbolic space is able to accommodate this since hyperbolic space is ”bigger” than Euclidean space. $\Rightarrow$ help visualize large networks
- 2-dimensional hyperbolic space is the hyperboloid model where points on the hyperbolic plane are modeled as points on the surface of the upper sheet of a hyperboloid in $\mathbb{R}^3$.
- Shortest path between two points must pass through the ”more central” points located in the bottom or bowl of the hyperboloid, those points nearest the origin.

Hyperbolic and Hypersphere Space



$\mathbb{Z}^t$ =disc of radius R in hyperbolic space, $\mathcal{H}^2$

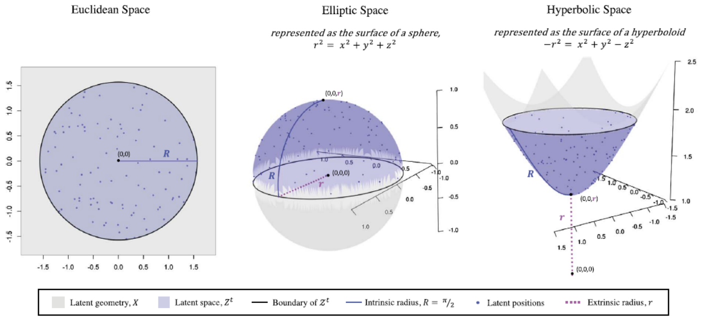


$\mathbb{Z}^t$ = surface of upper half of hypersphere, $\mathcal{S}^3$

- Circles of radius $\gamma = R$ about two particular simulated points
- Hyperbolic space $\mathbb{H}^2$ (negatively-curved space): Connections are radially oriented, centralized(Nodes near the base of the hyperbola are the more central nodes in the network)
- Hypersphere space δ^3 (positively-curved space): Connections are not radially oriented, not much centralization

Function of geometric curvature in LSM

- The key aspect of hyperbolic space which can accommodate complex network structure is its **negative curvature**.



- r (Extrinsic radius): Determines the curvature of elliptic and hyperbolic space

Distribution of Distances in LS Geometries

- Goal: Compare **the role of geometry** in the latent space model for network data
  1. Euclidean Space
  2. Hyperbolic Space
  3. Elliptic Space
- Model used in the simulation

$$Y_{ij} = \Theta\{\gamma - d(z_i, z_j)\}, \quad i < j$$

$$z_i \in \mathcal{Z}^t, z_i \overset{\text{iid}}{\sim} \text{Uniform}(R)$$

Distribution of Distances in LS Geometries

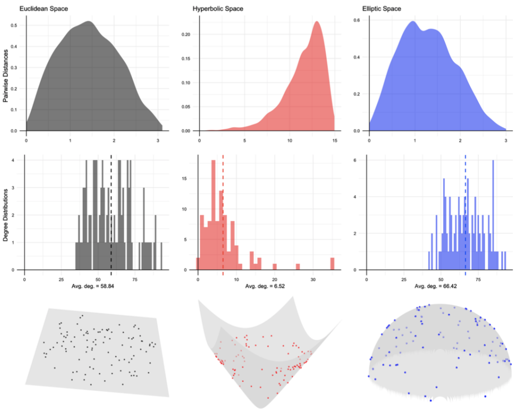
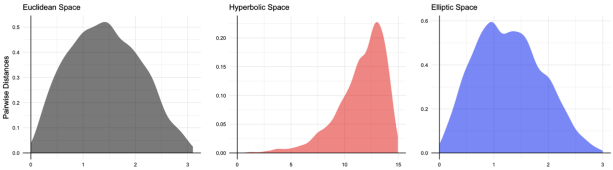


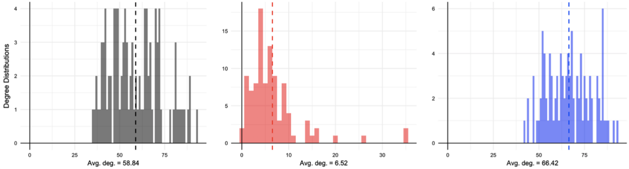
FIG. 8. *Distribution of distances in latent spaces with differing geometries. Panels in the bottom row plot one set of $n = 100$ points simulated in Euclidean, hyperbolic and elliptic space, respectively, while panels in the top row plot histograms of the pairwise distances for these sets of points. The center row plots the corresponding degree distributions across the three latent geometries, where nodes in each network are connected according to model (4.1) with $\gamma = R$.*

Pairwise Distances



- In both Euclidean and Elliptic space, the distribution of distances is much more **symmetric**.
- In Hyperbolic space, there are few points that are very close together and the distribution of distances is **left skewed**.
  - This distribution likely better represents our intuition about the distances between latent positions for nodes in real world social networks.
  - Any randomly selected pair of individuals is not likely to have a tie.

Degree Distributions



- Euclidean and Elliptical space implies that there are some loners and some popular actors but most actors display roughly similar levels of sociality.
- Hyperbolic space implies that most individuals are relatively asocial with only a few very popular actors.
    - It is a natural way of describing **degree heterogeneity**, a long observed trait of many real world networks.

Simulation

- We consider networks of various size (n=20, 50, 100) embedded in Euclidean, Hyperbolic, and Elliptic latent geometries.
- We simulate 5000 networks for each geometry, and increase Heaviside step function cut point from 0 to 1 by increments of 0.20
- We plot the average of common network summary measures
  - clustering, average path length, degree centrality, betweenness centrality, closeness centrality, modularity

Common Network Summary Measures

- Clustering
- Average Path Length

$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j),$$

where n is the number of vertices in G.

- Degree Centrality

- Betweenness Centrality

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

- Closeness Centrality

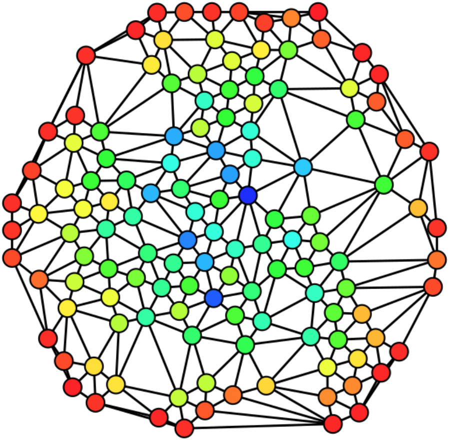
$$C(x) = \frac{1}{\sum_y d(y, x)}$$

where $d(y, x)$ is the distance between vertices x and y.

- Modularity

$$\text{mod}(C) = \sum_{k=1}^{K} \left[f_{kk}(C) - f_{kk}^* \right]^2$$

Betweenness Centrality



Network Statistics in Hyperbolic & Euclidean

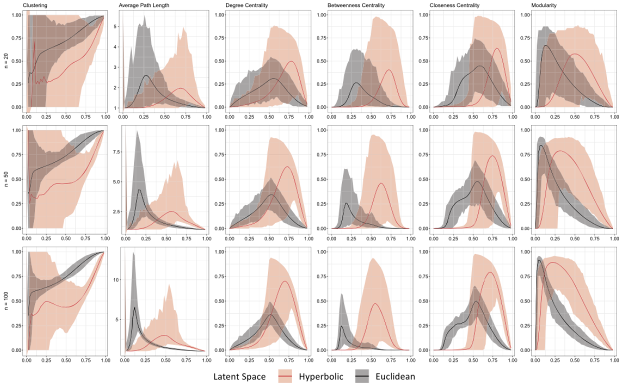


FIG. 9. *Network statistics in hyperbolic space. The x-axis is γ, the Heaviside step function cut point. Distances across the geometries are made comparable by re-scaling each simulation to the* [0, 1] *interval. Each row of plots corresponds to simulated networks of a different size,* $n = 20, 50, 100$ *and each column of plots corresponds to a different network summary measure.*

Network Statistics in Hyperbolic & Euclidean

- Networks in Hyperbolic space can achieve much higher levels of degree centrality, betweenness and closeness, and this effect is more prominent for larger networks.
- This coincides with our intuition for the behavior of the degree distribution of networks in Hyperbolic space.

Network Statistics in Elliptic & Euclidean

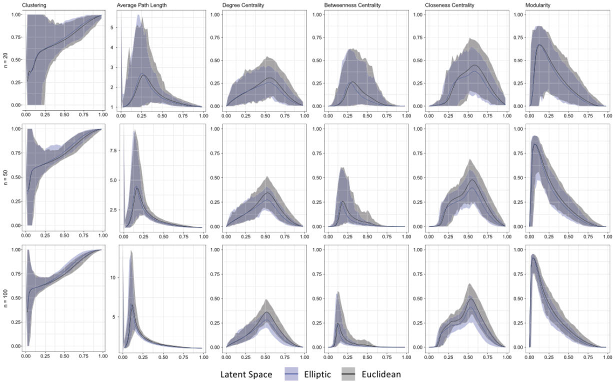


FIG. 10. *Network statistics in elliptic space. The x-axis is γ, the Heaviside step function cut point. Distances across the geometries are made comparable by re-scaling each simulation to the* [0, 1] *interval. Each row of plots corresponds to simulated networks of a different size,* $n = 20, 50, 100$ *and each column of plots corresponds to a different network summary measure.*

Challenge

- The behavior of these measures across latent spaces of different geometries can provide us with some indication of the role of geometry in these models.
- However, it is difficult to formally relate these statistics to properties of the latent space.
- This issue is rooted in the fact that **it is difficult to find a common language describing both networks and geometric spaces**
    - such a common language is a prerequisite for comparing properties between the two sorts of structures

Solution

- Both the geometry of a space and the combinatorics of a graph or network can be described by linear operators whose eigenvalues are comparable.
- The eigenvalues of graph Laplacian describe many combinational features of a graph G.
- For example,
  - the algebraic multiplicity of the eigenvalue 0 reveals how many connected components G has.
  - The second smallest eigenvalue of graph Laplacian quantifies how interconnected the nodes in G are.

Simulated Real Networks Across Diff. Geometries

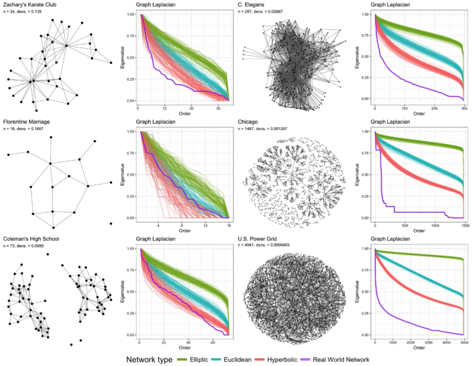


FIG. 11. *A small collection of real world networks. In these plots, we compare a set of real world networks to networks simulated from our general latent distance model (Equation* (2.4)) *across different geometries. Each row includes a network plot (using the default settings of the* `ggnet` *function in the* `GGally` *package for* R) *and a plot of the eigenvalue curves, where the color corresponds to the geometry of the latent space. In the eigenvalue plots, each network is represented by a curve which plots the network's eigenvalues against their order. Networks on the left side of the plot are examples of social networks, that range in size from* $n = 16$ *to* $n = 73$*, while networks on the right side of the plot are nonsocial networks, including a brain network, road networks and a power grid network.*

Summary

- **Using a negatively curved latent space** allows us to smoothly grow the complexity of the latent space model without adding any additional structure to the model.
- Models in Hyperbolic latent space **preserve the use of distance to model dependence**, a common feature of many popular statistical models, and is able to draw upon the long success of latent space models for network data.

Introduction

- Goal: To model **the population distribution** of networks under the assumption that replicated networks share common underlying structure
  - Example: brain connectivity networks for multiple subjects
- Consider the following $V \times V$ symmetric adjacency matrices $A_1, \ldots, A_n$ as realizations from a common network-valued random variable $\boldsymbol{A}$:

$$\begin{bmatrix} \quad A_1 \quad \end{bmatrix}, \begin{bmatrix} \quad A_2 \quad \end{bmatrix}, \cdots, \begin{bmatrix} \quad A_n \quad \end{bmatrix} \sim \begin{bmatrix} \quad \boldsymbol{A} \quad \end{bmatrix}$$

Introduction

Previous Models

1. Separately analyzing each A_i (Hagmann et al., 2008)
2. Averaging $A_1, \ldots, A_n$ (Scheinerman and Tucker, 2010)

Proposed Model

- **Latent space models** using a **Bayesian Nonparametric** approach to induce a prior on the unknown population distribution
- Bayesian Nonparametric model deals with infinite dimensional parameter space using finitely chosen subset of available parameter dimension.

Nonparametric Model

- Let $\mathcal{L}(A_1), \ldots, \mathcal{L}(A_n)$ be realizations from a multivariate R.V.: $\mathcal{L}(\boldsymbol{A})$, where

$$\mathcal{L}(A_i) = (A_{i[21]}, A_{i[31]}, \ldots, A_{i[V1]}, A_{i[32]}, \ldots, A_{i[V2]}, \ldots, A_{i[V(V-1)]})'$$

denotes lower triangluar elements of A_i for each $i = 1, 2, \ldots, n$.

- $\mathcal{L}(\boldsymbol{A})$ has binary entries $\mathcal{L}(\boldsymbol{A})_l \in \{0, 1\}, \; l = 1, \ldots, V(V-1)/2$,

$$\text{e.g., } \boldsymbol{A}_{V \times V} = \begin{bmatrix} & & & & \\ 0 & & & & \\ 1 & 0 & & & \\ 0 & 0 & 1 & & \\ \vdots & \vdots & \vdots & & \\ 1 & 1 & 0 & \cdots & 1 \end{bmatrix}$$

Nonparametric Model

- Let $\boldsymbol{a} \in \mathbb{A}_V = \{0,1\}^{V(V-1)/2}$ be one of the possible network configurations.
- Goal: To estimate the probability mass function

$$P_{\mathcal{L}(\boldsymbol{A})}(\boldsymbol{a}) = \Pr\{\mathcal{L}(\boldsymbol{A}) = \boldsymbol{a}\} = \sum_{h=1}^{H} \nu_h \prod_{l=1}^{V(V-1)/2} \left(\boldsymbol{\pi}_l^{(h)}\right)^{a_l} \left(\mathbf{1} - \boldsymbol{\pi}_l^{(h)}\right)^{1-a_l}$$

where $\nu_h \in (0,1)$ is the probability assigned to mixture component h and $\boldsymbol{\pi}_l^{(h)} \in (\mathbf{0}, \mathbf{1})$ **is edge probability** for the lth pair of nodes in mixture component h for every $h = 1, \ldots, H$, $l = 1, \ldots, V(V-1)/2$.

$\Rightarrow$ Mixture of Conditionally independent Bernoulli random variables

Nonparametric Model

Latent Space Characterizations for

1. **Borrowing network information** from replicated networks
2. **Reducing dimensionality** within each mixture component

- Set $\pi^{(h)} = (\pi_1^{(h)}, \ldots, \pi_{V(V-1)/2}^{(h)})' \in (0,1)^{V(V-1)/2}$ as following :

$$\pi^{(h)} = \{1 + \exp(-Z - D^{(h)})\}^{-1},$$
$$D^{(h)} = \mathcal{L}(X^{(h)} \Lambda^{(h)} X^{(h)'})$$

i.e., the component-specific edge probability $\pi^{(h)}$ is a function of

1. **Shared Similarity Vector** $Z \in \mathbb{R}^{V(V-1)/2}$
2. **Component-specific Deviation** $D^{(h)} \in \mathbb{R}^{V(V-1)/2}$ via matrix factorization

Nonparametric Model

Reminder: Latent Distance Model

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

- In our model, the logistic mapping is applied component-wise as following :

$$\pi^{(h)} = \{1 + \exp\left(-Z - D^{(h)}\right)\}^{-1}$$

$$\log \text{odds} = \log\left(\frac{\pi^{(h)}}{1 - \pi^{(h)}}\right) = \log\left(\frac{1}{\exp(-Z - D^{(h)})}\right) = Z + D^{(h)},$$

  where $D^{(h)}$ and Z correspond to $\alpha + \beta' x_{ij}$ and $-|z_i - z_j|$, respectively.
- Note that Z denotes the shared **similarity** vector that incorporates shared network information from replicated networks.

Nonparametric Model

- Why dimension reduction?

$$D^{(h)} = \mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)'}),$$

where $X^{(h)} \in \mathbb{R}^{V\times R}$ whose rows $X_v^{(h)'} = (X_{v1}^{(h)}, \ldots, X_{vR}^{(h)}), v = 1, \ldots, V$ denote the R **latent coordinates** of each node $v = 1, \ldots, V$ and $\Lambda^{(h)} \in \mathbb{R}^{R\times R}$ is a diagonal matrix whose diagonal elements $\lambda_r^{(h)} \geq 0$ measures the **importance of each dimension** $r = 1, \ldots, R$.

1. We have to estimate H number of $\pi^{(h)}$ for $h = 1, \ldots, H$.
2. Among V number of pathways, let's focus on R ($\ll V$) number of them.

$$D_l^{(h)} = \mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)'})_l = \sum_{r=1}^{R} \lambda_r^{(h)} X_{vr}^{(h)} X_{ur}^{(h)},$$

where l denotes the pair of nodes v and u, with $v > u$.

Prior Specification

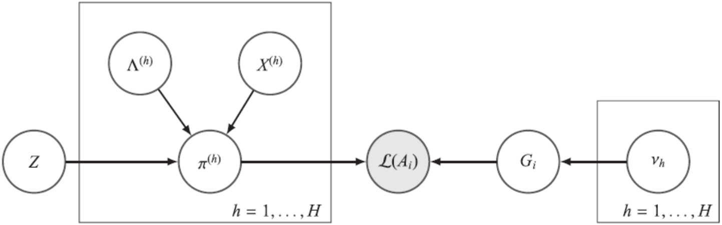


Figure: Probabilistic mechanism generating data $\mathcal{L}(A_i)$

- Component indicator variable $G_i \in \{1, \ldots, H\}$ with pmf $\Pr(G_i = h) = \nu_h$
- Prior specification for $\nu_h, Z, \lambda^{(h)}, X^{(h)}$

Prior Specification

- How to choose H and R?

  Choose H and R sufficiently large for satisfying the **full prior support** that ensures consistent posterior concentration around true population distribution.

1. $(\nu_1, \ldots, \nu_H) \sim \text{Dirichlet}(1/H, \ldots, 1/H)$
2. $\lambda^{(h)} \sim \text{MIG}(a_1, a_2)$ i.e., $\lambda_r^{(h)} = \prod_{m=1}^{r} \frac{1}{\delta_m^{(h)}}, \delta_1^{(h)} \sim \text{Ga}(a_1, 1),\ \delta_{m\geq 2}^{(h)} \sim \text{Ga}(a_2, 1)$
3. $Z \sim \text{N}_{V(V-1)/2}(\mu, \Sigma)$ where $\mu \in \mathbb{R}^{V(V-1)/2},\ \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_{V(V-1)/2}^2)$
4. $X_{vr}^{(h)} \sim \text{N}(0,1),\ v = 1, \ldots, V,\ r = 1, \ldots, R,\ h = 1, \ldots, H$

Comments from T. Broderick

- Nonparametric Bayes : **Density estimation** vs Clustering

$\Rightarrow$ Recovering a particular number of components H may not be a priority.

- Alternative : Dirichlet Process (DP) prior for flexible choice of H

  Letting $H = \infty$, consider a stick-breaking prior for $(\nu_1, \ldots, \nu_H)$ with appropriately Beta-distributed stick-breaking weights, i.e., DP prior.

$$\nu_1, \ldots, \nu_H \mid F \sim F$$
$$F \sim \text{DP}(\alpha, F_0)$$

$\Rightarrow$ $f(\nu) = \sum_{k=1}^{\infty} \beta_k \delta_{\nu_k}(\nu)$ where $\beta \sim \text{beta}(1, \alpha)$

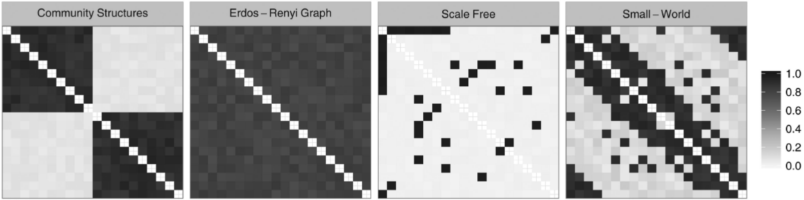
The smaller α is, the less of the stick H will be left.

Simulation Study

- 4 mixture component, and simulate 25 networks for each component by sampling edge probability from conditionally independent Bernoulli random variables.
- $V = 20$ nodes
- Gibbs iteration 5000 with $H = 30$, $R = 10$, and set burn-in of 1000.
- $\mu_1 = \cdots = \mu_{V(V-1)/2} = 0$
- $\sigma_1^2 = \cdots = \sigma_{V(V-1)/2}^2 = 10$
- $a_1 = 2.5$, $a_2 = 3.5$ in the $MIG(a_1, a_2)$ prior.

True Component Specific Edge Probability Vector

- Each component specific edge probability vector is constructed to assign high probability to a subset of network configurations characterized by a specific property.



Comparison

- Comparing this approach with hierarchical Bayesian formulations (Handcock et al. 2007) and bilinear models (Hoff, 2005).
- Authors claim that their approach is superior due to flexibility.

Network Measures: Eigen-centrality

- Eigen-centrality: For a given network $G := (V, E)$ and adjacency matrix A, the relative centrality score of vertex v can be defined as:

$$x_v = \frac{1}{\lambda} \sum_{t \in G} A_{i,j} x_t$$

where λ is the largest eigen value of A.

- In neuroscience, the eigenvector centrality of a neuron in a model neural network has been found to correlate with its relative firing rate.

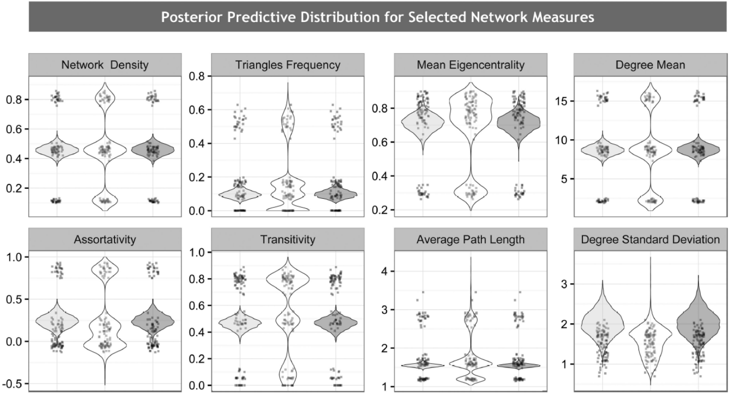
Network Measures: Density and Assortativity

- Density: For a given network $G := (V, E)$ and adjacency matrix A the density of a graph G is

$$den(G) = \frac{|E|}{|V|(|V-1|)/2}$$

- Network density describes the portion of the potential connections in a network that are actual connections.
- Assortativity: The assortativity is based on a characteristics, how much vertices tends to connect other nodes with same characteristics.

Goodness-of-Fit



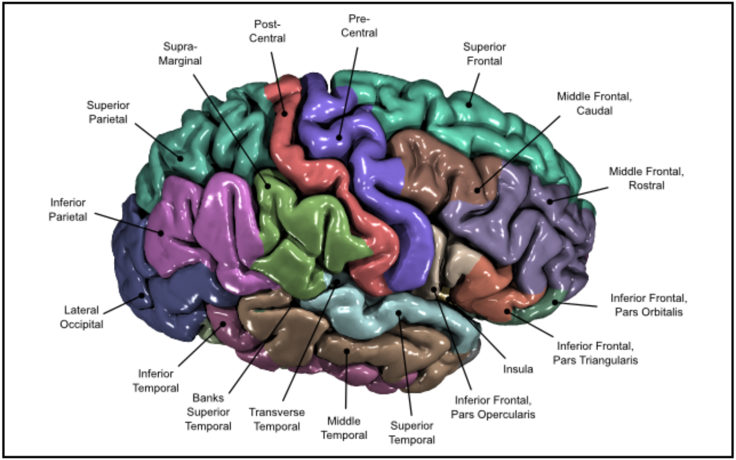
Comments from Adrian E. Raftery

- Is it a fair comparison?
- Durant and co-authors report that their model does better than the Handcock model in their simulation model.
- However, all of this difference is likely due to the simple fact that they allow the latent space to vary between replicates.
- A fairer comparison would be between their model and an extension of Handcock model in which the latent space is allowed to vary between replicates.

Brain Networks

- 21 healthy subjects with no history of neurological diseases, for each subject two brain network observations are available, for a total of $n = 42$.
- Brain regions are constructed according to the Desikan atlas, for a total of $V = 68$.
- If at least one connected white matter fiber exists, there exists a connection between two regions.
- Others are similar to above simulation settings.

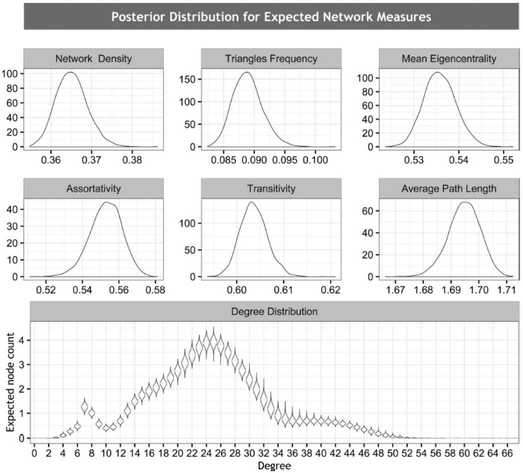
Desikan Atlas



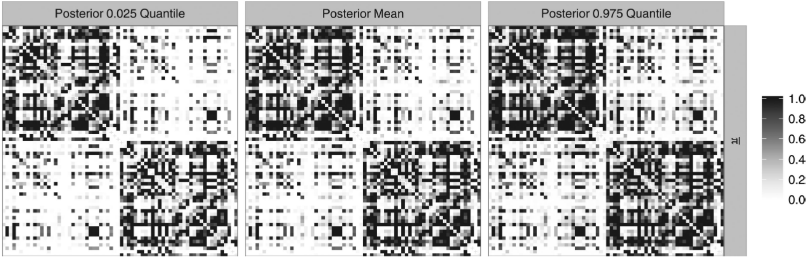
Purpose of Brain Connectome Analysis

- The connectome will significantly increase our understanding of how functional brain states emerge from their underlying structural substrate, and will provide new mechanistic insights into how brain function is affected if this structural substrate is disrupted.
- That is, we can utilize it to classify neurological disease, such as autism, by investigating absence of crucial connection in the estimated connectome.

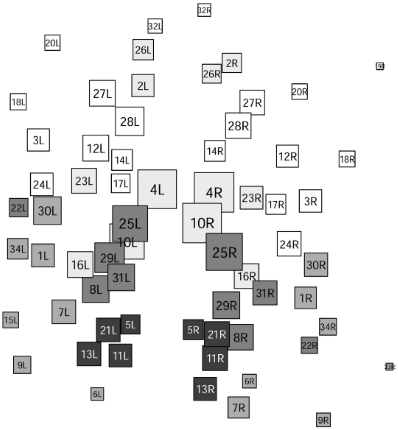
Posterior for Expected Network Measures



Summary of Posterior for the Elements $\bar{\pi}_l$



Weighted Network Visualization



Discussion

- One important topic is developing supervised approaches that include predictor variables along with network-valued responses.
- For example, behavioral phenotypes are available in some studies along with brain structural connection network data, with interest focusing on testing for variations in connectivity patterns across such phenotypes.

Appendix: Posterior Computation

1. Allocate each network $\mathcal{L}(A_i)$ to one of the mixture components

$$G_i \sim \Pr(G_i = h \mid \cdot) = \frac{\nu_h \prod_{l=1}^{V(V-1)/2} (\pi_l^{(h)})^{\mathcal{L}(A_i)_l} (1 - \pi_l^{(h)})^{1-\mathcal{L}(A_i)_l}}{\sum_{q=1}^{H} \nu_q \prod_{l=1}^{V(V-1)/2} (\pi_l^{(q)})^{\mathcal{L}(A_i)_l} (1 - \pi_l^{(q)})^{1-\mathcal{L}(A_i)_l}}$$

2. Update the **mixing probabilities**

$$(\nu_1, \ldots, \nu_H) \mid \cdot \sim \text{Dirichlet}\{1/H + \sum_{i=1}^{n} I(G_i = 1), \ldots, 1/H + \sum_{i=1}^{n} I(G_i = H)\}$$

- For updating $Z, X^{(h)}$ and $\lambda^{(h)}$, let's use **Polya-gamma data augmentation**
  to circumvent inconvenience in Bayesian logistic regression problem. (Polson and Scott, 2013)

Appendix: Posterior Computation

- Polya-gamma data augmentation for aggregated networks $Y^{(1)}, \ldots, Y^{(H)}$

  with $Y^{(h)} = \sum_{i:G_i=h} \mathcal{L}(A_i)$, for $h = 1, \ldots, H$,

$$\left(Y_l^{(h)} \mid Z, X^{h)}, \lambda^{(h)}\right) \sim \text{Binom}(n_h, \pi_l^{(h)}),$$

$$\left(\pi_l^{(h)} = [1 + \exp\{-Z_l - \mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)'})_l\}]^{-1}\right),$$

  independently for $l = 1, \ldots, V(V-1)/2$ and $h = 1, \ldots, H$ with n_h the number of networks in component h.

- Now, we can use simple (Polya-gamma) Gibbs-sampling algorithms for the Bayesian logistic regression rather than designing complex MH-samplers.

Appendix: Posterior Computation

3 Update the Polya-gamma augmented variable

$$\omega_l^{(h)} \mid \cdot \sim \mathrm{PG}\{n_h, Z_l + \mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)'})\}$$

4 Update the **shared similarity vector**

$$Z \mid \cdot \sim \mathrm{N}_{V(V-1)/2}(\mu_Z, \Sigma_Z)$$

with Σ_Z diagonal having elements $\sigma_{Z_l}^2 = 1/(\sigma_l^{-1} + \sum_{h=1}^H \omega_l^{(h)})$, whereas $\mu_Z = \sigma_{Z_l}^2[\sigma_l^{-2}\mu_l + \sum_{h=1}^H \{Y_l^{(h)} - n_h/2 - \omega_l^{(h)}\mathcal{L}(X^{(h)}\Lambda^{(h)}X^{(h)'})_l\}]$

Appendix: Posterior Computation

5 Update the **component-specific weighted latent coordinates**

Let's block-sample each row of $\bar{X}^{(h)}$ conditionally on the other parameters :

$$\bar{X}_{\nu}^{(h)} \mid \cdot \sim \mathrm{N}_R\left\{\left(\bar{X}_{(-\nu)}^{(h)\prime}\Omega_{(\nu)}^{(h)}\bar{X}_{(-\nu)}^{(h)} + \Lambda^{(h)^{-1}}\right)^{-1}\eta_{\nu}^{(h)}, \left(\bar{X}_{(-\nu)}^{(h)\prime}\Omega_{(\nu)}^{(h)}\bar{X}_{(-\nu)}^{(h)} + \Lambda^{(h)^{-1}}\right)^{-1}\right\}$$

with $\eta_{\nu}^{(h)} = \bar{X}_{(-\nu)}^{(h)\prime}(Y_{(\nu)}^{(h)} - 1_{V-1}n_h/2 - \Omega_{(\nu)}^{(h)}Z_{(h)})$,

where $Y_{(\nu)}^{(h)} \sim \mathrm{Binom}\left(n_h, \pi_{(\nu)}^{(h)}\right)$, $\mathrm{logit}\left(\pi_{(\nu)}^{(h)}\right) = Z_{(\nu)} + \bar{X}_{(-\nu)}^{(h)}\bar{X}_{\nu}^{(h)}$

Appendix - Posterior Computation

6 Update the **component-specific weight parameters**

$$\delta_1^{(h)} \mid \cdot \sim \text{Ga}\left\{a_1 + \frac{VR}{2}, 1 + \frac{1}{2}\sum_{m=1}^{R}\theta_m^{(-1)}\sum_{v=1}^{V}(\bar{X}_{vm}^{(h)})^2\right\},$$

$$\delta_{r\geq 2}^{(h)} \mid \cdot \sim \text{Ga}\left\{a_2 + \frac{V(R-r+1)}{2}, 1 + \frac{1}{2}\sum_{m=r}^{R}\theta_m^{(-1)}\sum_{v=1}^{V}(\bar{X}_{vm}^{(h)})^2\right\}$$

where $\theta_m^{(-r)} = \prod_{t=1, t\neq r}^{m} \delta_t^{(h)}$, for $r = 1, \ldots, R$

7 Update the **component-specific edge probability vectors**

Compute $\pi^{(h)} = [1 + \exp\{-Z - \mathcal{L}(\bar{X}^{(h)}\bar{X}^{(h)'})\}]^{-1}$