Network Regression with Graph Laplacians

Yidong Zhou, UC Davis

(joint work with Hans-Georg Müller)

Conference on Statistical Methods for High Dimensional Complex Data

May 23, 2024

- Samples of random objects (non-Euclidean data) that take values in a metric space are becoming increasingly prevalent.
- Due to the absence of a vector space structure, basic statistical tools for scalar/vector data are no longer applicable.

- Examples of random objects:
  - networks,
  - probability measures,
  - covariance/correlation matrices,
  - etc.
- Applications: brain imaging studies, multi-cohort studies, human longevity, etc.

- Networks, a prominent example of random objects, arise in numerous applications, e.g.,

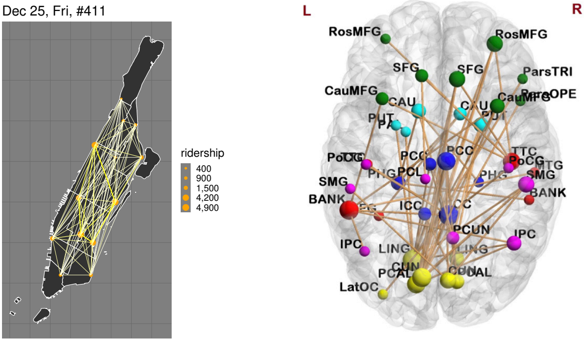


Figure 1: Left: transport network on Dec 25, 2020 in Manhattan. Right: brain functional connectivity network with 40 regions of interest (ROIs).

- Addressing the statistical challenges posed by samples of networks is an emerging frontier, where each network is treated as an individual data point.

- How does a network change as a function of vector covariates?
- This question can be addressed through regression analysis where
    - Response: network;
    - Predictors: $\mathbb{R}^p$.

Related work:

- Severn et al., 2021, 2022: embed the space of graph Laplacians in a Euclidean space, where regression is applied using extrinsic methods.
- Calissano et al., 2022, 2023: implement linear regression in the Euclidean space and then project back to the “graph space” through a quotient map.

Example: Dynamics of Networks in the Aging Brain

- As a person gets older, changes occur in all parts of the body, including the brain.
- Certain parts of the brain shrink, including those important to learning and other complex mental activities.

Example: Dynamics of Networks in the Aging Brain

- The Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).
- Response: functional brain connectivity network of each subject.
- Predictor: the corresponding age.

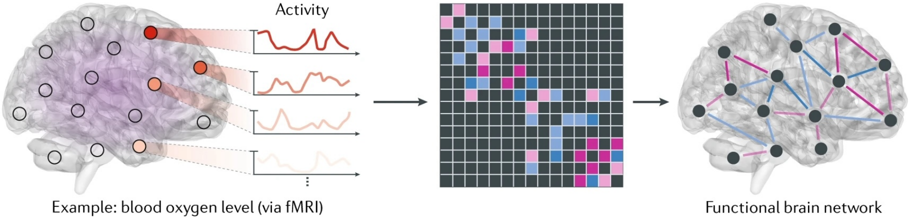


Figure 2: Construction of functional brain connectivity networks (Lynn & Bassett, 2019).

Characterization of the Space of Networks

$G_m = (V, E)$: a network with a set of nodes $V = \{v_1, \ldots, v_m\}$ and a set of edge weights $E = \{w_{ij} : w_{ij} \geq 0, i, j = 1, \ldots, m\}$, where $w_{ij} = 0$ indicates v_i and v_j are unconnected.

(C0) G_m is simple, i.e., there are no self-loops or multi-edges.

(C1) G_m is weighted, undirected, and labeled.

(C2) Edge weights w_{ij} are bounded above by $W \geq 0$, i.e., $0 \leq w_{ij} \leq W$.

Characterization of the Space of Networks

Any network satisfying Conditions (C0)–(C2) can be uniquely associated with its graph Laplacian $L = (l_{ij})$, defined as

$$l_{ij} = \begin{cases} -w_{ij}, & i \neq j \\ \sum_{k \neq i} w_{ik}, & i = j \end{cases}$$

for $i, j = 1, \ldots, m$.

Characterization of the Space of Networks

The space of networks can thus be characterized by the corresponding space of graph Laplacians,

$$\mathcal{L}_m = \{L = (l_{ij}) : \; L = L'; \; L1_m = 0_m; \; -W \leq l_{ij} \leq 0 \text{ for } i \neq j\}, \qquad (1)$$

where 1_m and 0_m are the m-vectors of ones and zeroes, respectively.

Characterization of the Space of Networks

Proposition

The space $\mathcal{L}_m$, defined in (1), is a bounded, closed, and convex subset in $\mathbb{R}^{m^2}$ of dimension $m(m-1)/2$.

- The convexity and closedness of $\mathcal{L}_m$ as shown in Proposition 1 ensures the existence and uniqueness of projections onto $\mathcal{L}_m$ that we will use in the proposed regression approach.

Choice of Metrics

- There are various metrics to choose from for $\mathcal{L}_m$ and one common choice is the Frobenius metric, defined as

$$d_F(L_1, L_2) = [\mathrm{tr}\{(L_1 - L_2)'(L_1 - L_2)\}]^{1/2}.$$

Choice of Metrics

- Let $\mathcal{S}_m^+$ denote the space of real symmetric positive semi-definite $m \times m$ matrices.

- Another popular metric is the power metric,

$$d_{F,\alpha}(L_1, L_2) = d_F\{F_\alpha(L_1), F_\alpha(L_2)\},$$

  where $F_\alpha(S) = U\Lambda^\alpha U'$, and $U\Lambda U'$ represents the spectral decomposition of $S \in \mathcal{S}_m^+$.
- For $\alpha = 1$, $d_{F,\alpha}$ reduces to the Frobenius metric d_F.

Fréchet Mean and Conditional Fréchet Mean

- Consider the random pair $(X, L) \sim F$, where X takes values in $\mathbb{R}^p$, $L \in \mathcal{L}_m$ is a graph Laplacian.
- The Fréchet mean and conditional Fréchet mean are generalizations of the mean and conditional mean from the real line to general metric spaces.

- Let $Y \in \mathbb{R}$ denote a random variable on the real line.

$$E(Y) = \underset{y \in \mathbb{R}}{\arg\min}\, E\{(Y - y)^2\}, \quad E(Y|X) = \underset{y \in \mathbb{R}}{\arg\min}\, E\{(Y - y)^2|X\}.$$

Fréchet Mean and Conditional Fréchet Mean

$$E(Y) = \underset{y \in \mathbb{R}}{\arg\min}\, E\{(Y - y)^2\}, \quad E(Y|X) = \underset{y \in \mathbb{R}}{\arg\min}\, E\{(Y - y)^2|X\}.$$

- Mean ⇝ Fréchet mean (Fréchet, 1948):

$$E(Y) \rightsquigarrow \underset{\omega \in \mathcal{L}_m}{\arg\min}\, E\{d^2(L, \omega)\}.$$

- Conditional mean ⇝ conditional Fréchet mean (Petersen & Müller, 2019):

$$E(Y|X) \rightsquigarrow \underset{\omega \in \mathcal{L}_m}{\arg\min}\, E\{d^2(L, \omega)|X\}.$$

- d can be either the Frobenius or power metric.

Global Network Regression

- To model the relationship between networks and vector predictors, a natural target is the conditional Fréchet mean (Petersen & Müller, 2019),

$$m(x) = \operatorname*{arg\,min}_{\omega \in \mathcal{L}_m} E\{d^2(L, \omega) | X = x\}. \tag{2}$$

- Recall that for scalar responses, linear regression assumes a linear relationship between X and the conditional mean of Y given X, i.e.,

$$E(Y|X) = \beta_0 + \beta_1' X.$$

Global Network Regression

Using ordinary least squares, the regression function can be alternatively characterized by

$$E(Y|X = x) = \underset{y \in \mathbb{R}}{\arg\min}\, E\{s_G(x)(Y - y)^2\},$$

where the weight function $s_G(x) = 1 + (X - \mu)'\Sigma^{-1}(x - \mu)$ with $\mu = E(X)$ and $\Sigma = \mathrm{Var}(X)$.

Global Network Regression

The global network regression, extending linear regression to network responses, is defined as

$$m_G(x) = \operatorname*{arg\,min}_{\omega \in \mathcal{L}_m} E\{s_G(x) d^2(L, \omega)\}, \tag{3}$$

where d can be either the Frobenius or power metric.

Global Network Regression

- Suppose that $(X_k, L_k) \sim F, k = 1, \ldots, n$ are independent and define

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad \hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})(X_k - \bar{X})'.$$

- The regression function in (3) can be estimated by

$$\hat{m}_G(x) = \operatorname*{arg\,min}_{\omega \in \mathcal{L}_m} \frac{1}{n}\sum_{k=1}^{n} s_{kG}(x) d^2(L_k, \omega), \tag{4}$$

where $s_{kG}(x) = 1 + (X_k - \bar{X})'\hat{\Sigma}^{-1}(x - \bar{X})$.

Local Network Regression

The local network regression, a generalization of local linear regression to network responses, follows a similar form but employs a different weight function.

Rates of Convergence

Pointwise and uniform rates of convergence:

	Global network regression	Local network regression
Frobenius metric		
Power metric		

Remark

The power metric necessitates specific considerations when investigating asymptotic properties.

Rates of Convergence Using Frobenius Metric

Theorem

Let the space of graph Laplacian matrices $\mathcal{L}_m$ *be endowed with the Frobenius metric* d_F*. Then for a fixed* $x \in \mathbb{R}^p$*, it holds for* $m_G(x)$ *and* $\hat{m}_G(x)$ *as per* (3) *and* (4) *that,*

$$d_F\{m_G(x), \hat{m}_G(x)\} = O_p(n^{-1/2}). \tag{5}$$

Furthermore, for a given $B > 0$ *and any* $\varepsilon > 0$*,*

$$\sup_{\|x\|_E \leq B} d_F\{m_G(x), \hat{m}_G(x)\} = O_p(n^{-1/\{2(1+\varepsilon)\}}). \tag{6}$$

Rates of Convergence Using Power Metric

Theorem

Let the space of graph Laplacian matrices $\mathcal{L}_m$ *be endowed with the power metric* $d_{F,\alpha}$*. Then for a fixed* $x \in \mathbb{R}^p$*, it holds for* $m_G(x)$ *and* $\hat{m}_G(x)$ *that,*

$$d_F\{m_G(x), \hat{m}_G(x)\} = \begin{cases} O_p(n^{-1/2}) & 0 < \alpha \leq 1 \\ O_p(n^{-1/(2\alpha)}) & \alpha > 1 \end{cases}. \tag{7}$$

Furthermore, for a given $B > 0$ *and any* $\varepsilon > 0$*,*

$$\sup_{\|x\|_E \leq B} d_F\{m_G(x), \hat{m}_G(x)\} = \begin{cases} O_p(n^{-1/\{2(1+\varepsilon)\}}) & 0 < \alpha \leq 1 \\ O_p(n^{-1/\{2(1+\varepsilon)\alpha\}}) & \alpha > 1 \end{cases}. \tag{8}$$

Global v.s. Local Network Regression

- While global network regression relies on stronger model assumptions, it does not require a tuning parameter and is applicable for categorical predictors.

- Local network regression, by contrast, is more flexible and may be preferable as long as the regression relation is smooth, the covariate dimension is low and the covariates are continuous.

Dynamics of Networks in the Aging Brain

- Data: $n = 404$ cognitively normal elderly subjects with age ranging from 55.61 to 95.39 years.

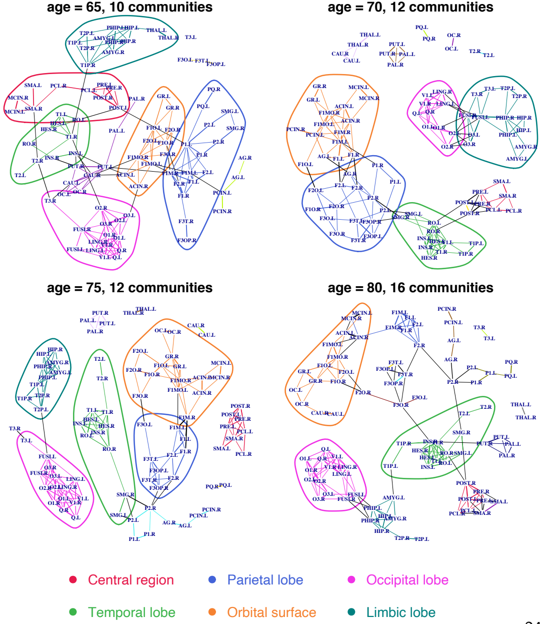
- Response: functional brain connectivity network of each subject.

- Predictor: the corresponding age.

- Model: local network regression.

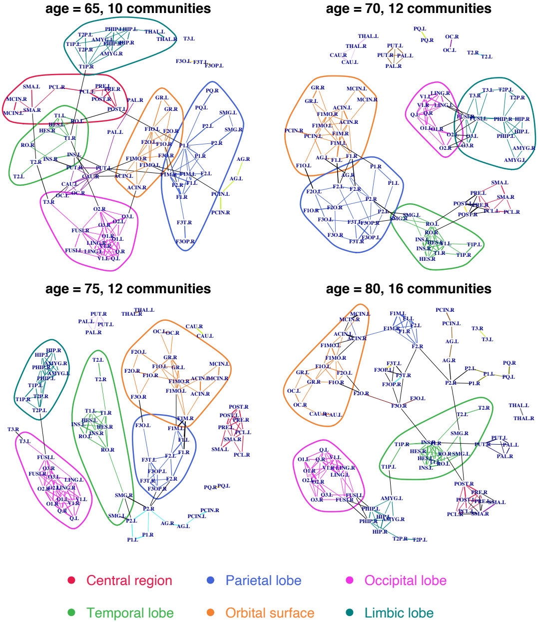
Dynamics of Networks in the Aging Brain

- The number of communities for ages 65, 70, 75, and 80 is 10, 12, 12, and 16, respectively.



Dynamics of Networks in the Aging Brain

- The number of communities for ages 65, 70, 75, and 80 is 10, 12, 12, and 16, respectively.
- These communities are found to be associated with different anatomical regions of the brain.



age = 65, 10 communities

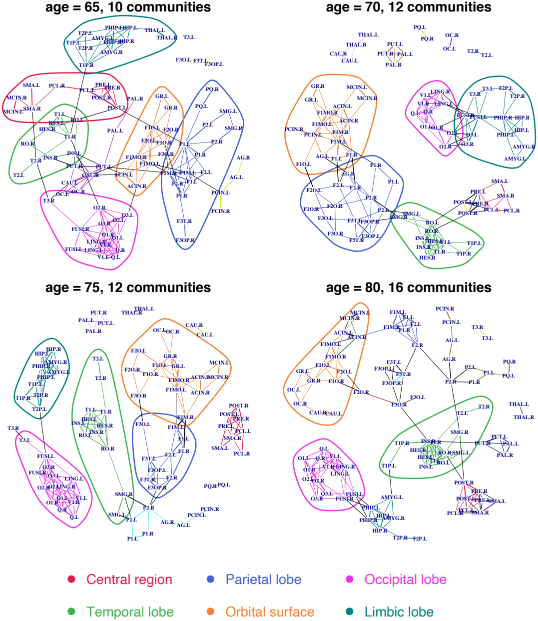
age = 70, 12 communities

age = 75, 12 communities

age = 80, 16 communities

Dynamics of Networks in the Aging Brain

- The number of communities for ages 65, 70, 75, and 80 is 10, 12, 12, and 16, respectively.
- These communities are found to be associated with different anatomical regions of the brain.
- Higher age is associated with increased local interconnectivity and cliquishness.



References I

Calissano, A., Feragen, A., & Vantini, S. (2022).Graph-valued regression: Prediction of unlabelled networks in a non-Euclidean graph space. *Journal of Multivariate Analysis*, *190*, 104950.

Calissano, A., Feragen, A., & Vantini, S. (2023).Populations of unlabelled networks: Graph space geometry and generalized geodesic principal components. *Biometrika*, *111*, 147–170.

Deutsch, F. (2001). *Best approximation in inner product spaces* (Vol. 7). Springer New York.

Dryden, I. L., Koloydenko, A., & Zhou, D. (2009).Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, *3*, 1102–1123.

Fréchet, M. (1948).Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, *10*(4), 215–310.

References II

- Lynn, C. W., & Bassett, D. S. (2019).The physics of brain network structure, function and control. *Nature Reviews Physics, 1*(5), 318–332.
- Petersen, A., & Müller, H.-G. (2019).Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics, 47*(2), 691–719.
- Severn, K. E., Dryden, I. L., & Preston, S. P. (2021).Non-parametric regression for networks. *Stat, 10*(1), e373.
- Severn, K. E., Dryden, I. L., & Preston, S. P. (2022).Manifold valued data analysis of samples of networks, with applications in corpus linguistics. *Annals of Applied Statistics, 16*(1), 368–390.

Questions?

Characterization of the Space of Networks

$$L = D - A$$

Degree matrix D

$$\begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.4 \end{pmatrix}$$

$$-$$

Adjacency matrix A

$$\begin{pmatrix} 0 & 0.2 & 0.3 \\ 0.2 & 0 & 0.1 \\ 0.3 & 0.1 & 0 \end{pmatrix}$$

$$\downarrow$$

Graph Laplacian L

$$\begin{pmatrix} 0.5 & -0.2 & -0.3 \\ -0.2 & 0.3 & -0.1 \\ -0.3 & -0.1 & 0.4 \end{pmatrix}$$

Choice of Metrics

- While d_F is the simplest among the possible metrics on $\mathcal{L}_m$, it suffers from the swelling effect. $\Longrightarrow$ The Euclidean average preserves the trace, while the determinant typically inflates.
- For $\alpha = 1/2$, the square root metric $d_{F,1/2}$ is a canonical choice that has been widely studied.

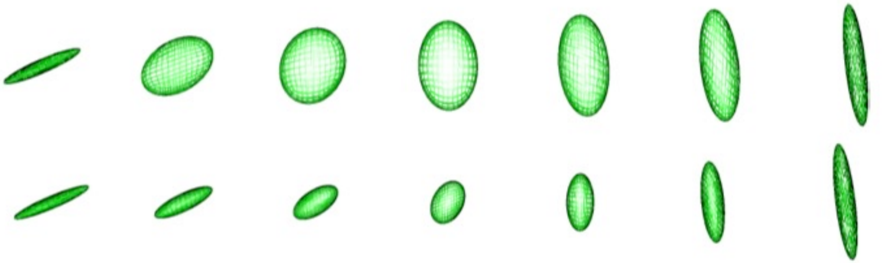


Figure 3: Interpolation of two matrices, where the top row is obtained using d_F and the bottom row is obtained using $d_{F,1/2}$ (Dryden et al., 2009).

Global Network Regression

The global network regression estimate in (4) using the Frobenius metric d_F can be simplified as a projection $P_{\mathcal{L}_m}$ onto $\mathcal{L}_m$,

$$\hat{m}_G(x) = \underset{\omega \in \mathcal{L}_m}{\arg\min}\, d_F^2\{B_G(x), \omega\} = P_{\mathcal{L}_m}\{B_G(x)\}, \tag{9}$$

where

$$B_G(x) = \frac{1}{n} \sum_{k=1}^{n} s_{kG}(x) L_k.$$

Network Regression Using Power Metric

- The largest eigenvalue of L is bounded, say by D, a nonnegative constant depending on m and W.
- Define the embedding space $\mathcal{M}_m$ to be a subset of $\mathcal{S}_m^+$,

$$\mathcal{M}_m = \{S \in \mathcal{S}_m^+ : \lambda_1(S) \leq C^\alpha\}, \tag{10}$$

  where $\lambda_1(S)$ denotes the largest eigenvalue of S.

The image of $\mathcal{L}_m$ under the matrix power map F_α, i.e., $F_\alpha(\mathcal{L}_m)$, is a subset of $\mathcal{M}_m$.

Network Regression Using Power Metric

$$\begin{tikzcd}
 & F_\alpha(\mathcal{L}_m) \arrow[dr, bend left, "\mathrm{Id}"] & \\
\mathcal{L}_m \arrow[ur, bend left, "F_\alpha"] & & \mathcal{M}_m \arrow[dl, bend left, "F_{1/\alpha}"] \\
 & F_{1/\alpha}(\mathcal{M}_m) \arrow[ul, bend left, "P_{\mathcal{L}_m}"] &
\end{tikzcd}$$

Figure 4: Schematic diagram for network regression with power metric $d_{F,\alpha}$.

- Think about data transformation in linear regression: $Y \longrightarrow Y^\alpha$.
- Global and local network regression are carried out in the embedding space $\mathcal{M}_m$ using d_F, where the existence and uniqueness can be ensured.

Optimal Pointwise Rates of Convergence

- Both pointwise and uniform rates of convergence have been established for global and local network regression, utilizing both the Frobenius and power metrics.

- Pointwise rates of convergence are optimal for both global and local network regression when employing the Frobenius and power metrics with $0 < \alpha < 1$.

New York Yellow Taxi System After COVID-19 Outbreak

- Responses: traffic networks from Apr 12, 2020 to Sep 30, 2020.
- Predictors: COVID-19 new cases per day, weekend indicator.
- Model: global network regression.

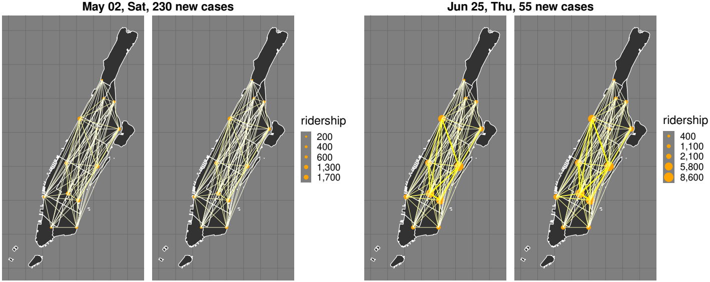
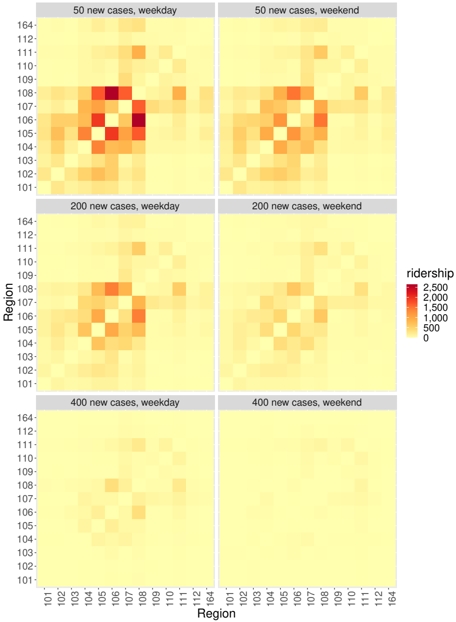


Figure 5: True (left) and fitted (right) networks on May 2, 2020 and Jun 25, 2020. The corresponding number of COVID-19 new cases are in the headline.

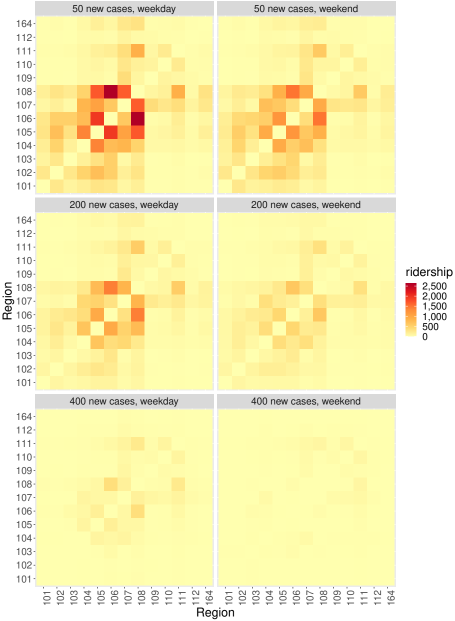
New York Yellow Taxi System After COVID-19 Outbreak

- Predicted networks represented as heatmaps at different number of COVID-19 new cases on weekdays or weekends.



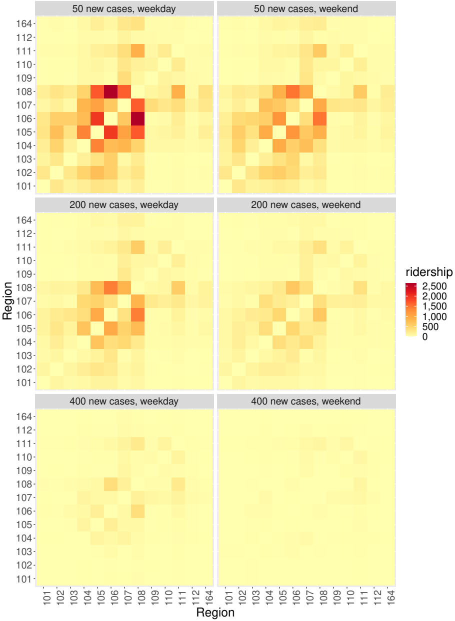
New York Yellow Taxi System After COVID-19 Outbreak

- Predicted networks represented as heatmaps at different number of COVID-19 new cases on weekdays or weekends.
- Edge weights are seen to decrease for increasing COVID-19 new cases.



New York Yellow Taxi System After COVID-19 Outbreak

- Predicted networks represented as heatmaps at different number of COVID-19 new cases on weekdays or weekends.
- Edge weights are seen to decrease for increasing COVID-19 new cases.
- Weekend taxi traffic, with lighter and less essential traffic compared to weekdays, is more severely affected by COVID-19.



Assumptions for Local Network Regression (cont'd)

In the following, $f_X(\cdot)$ and $f_{X|L}(\cdot, \omega)$ stand for the marginal density of X and the conditional density of X given $L = \omega$, respectively. $\mathcal{T}$ is a closed interval in $\mathbb{R}$ with interior $\mathcal{T}^o$.

(A1) The kernel $K(\cdot)$ is a probability density function, symmetric around zero. Furthermore, defining $K_{kj} = \int_{\mathbb{R}} K^k(u) u^j du$, $|K_{14}|$ and $|K_{26}|$ are both finite.

(A2) $f_X(\cdot)$ and $f_{X|L}(\cdot, \omega)$ both exist and are twice continuously differentiable, the latter for all $\omega \in \mathcal{L}_m$, and $\sup_{x,\omega} |(\partial^2 f_{X|L}/\partial x^2)(x, \omega)| < \infty$. Additionally, for any open set $U \subset \mathcal{L}_m$, $\int_U dF_{L|X}(x, \omega)$ is continuous as a function of x.

(A3) The kernel $K(\cdot)$ is a probability density function, symmetric around zero, and uniformly continuous on $\mathbb{R}$. Furthermore, defining $K_{jk} = \int_{\mathbb{R}} K(u)^j u^k du$ for $j, k \in \mathbb{N}$, $|K_{14}|$ and $|K_{26}|$ are both finite. The derivative K' exists and is bounded on the support of K, i.e., $\sup_{K(x)>0} |K'(x)| < \infty$; additionally, $\int_{\mathbb{R}} x^2 |K'(x)| (|x \log |x||)^{1/2} dx < \infty$.

(A4) $f_X(\cdot)$ and $f_{X|L}(\cdot, \omega)$ both exist and are continuous on $\mathcal{T}$ and twice continuously differentiable on $\mathcal{T}^o$, the latter for all $\omega \in \mathcal{L}_m$. The marginal density $f_X(\cdot)$ is bounded away from zero on $\mathcal{T}$, $\inf_{x \in \mathcal{T}} f_X(x) > 0$. The second-order derivative f_X'' is bounded, $\sup_{x \in \mathcal{T}^o} |f_X''(x)| < \infty$. The second-order partial derivatives $(\partial^2 f_{X|L}/\partial x^2)(\cdot, \omega)$ are uniformly bounded, $\sup_{x \in \mathcal{T}^o, \omega \in \mathcal{L}_m} |(\partial^2 f_{X|L}/\partial x^2)(x, \omega)| < \infty$. Additionally, for any open set $U \subset \mathcal{L}_m$, $\int_U dF_{L|X}(x, \omega)$ is continuous as a function of x; $M(\cdot, x)$ is equicontinuous, i.e., for all $x \in \mathcal{T}$,

$$\limsup_{z \to x} \sup_{\omega \in \mathcal{L}_m} |M(\omega, z) - M(\omega, x)| = 0.$$

Local Network Regression

For $X \in \mathbb{R}$, consider a smoothing kernel $K(\cdot)$ corresponding to a probability density and $K_h(\cdot) = h^{-1}K(\cdot/h)$ with h a bandwidth.

$$m_{L,h}(x) = \underset{\omega \in \mathcal{L}_m}{\arg\min}\, E[s_L(x,h) d^2(L,\omega)], \tag{11}$$

where $s_L(x,h) = K_h(X-x)[\mu_2 - \mu_1(X-x)]/\sigma_0^2$ with $\mu_j = E[K_h(X-x)(X-x)^j]$ for $j = 0,1,2$ and $\sigma_0^2 = \mu_0\mu_2 - \mu_1^2$.

$$\hat{m}_{L,n}(x) = \underset{\omega \in \mathcal{L}_m}{\arg\min}\, \frac{1}{n}\sum_{k=1}^{n} s_{kL}(x,h) d^2(L_k,\omega). \tag{12}$$

Here $s_{kL}(x,h) = K_h(X_k-x)[\hat{\mu}_2 - \hat{\mu}_1(X_k-x)]/\hat{\sigma}_0^2$, where $\hat{\mu}_j = n^{-1}\sum_{k=1}^n K_h(X_k-x)(X_k-x)^j$ for $j = 0,1,2$ and $\hat{\sigma}_0^2 = \hat{\mu}_0\hat{\mu}_2 - \hat{\mu}_1^2$. The dependency on n is through the bandwidth sequence $h = h_n$.

Local Network Regression

Note that for the case of $X \in \mathbb{R}^p$ with $p > 1$, we still have (11) and (12). However, the weight function takes a slightly different form,

$$s_L(x, h) = \frac{1}{\mu_0 - \mu_1' \mu_2^{-1} \mu_1} K_h(X - x)[1 - \mu_1' \mu_2^{-1}(X - x)],$$

where $\mu_0 = E[K_h(X - x)], \mu_1 = E[K_h(X - x)(X - x)]$, and $\mu_2 = E[K_h(X - x)(X - x)(X - x)']$ is non-degenerate. The sample version $s_{kL}(x, h)$ can be defined similarly.

Rates of Convergence Using Frobenius Metric

Theorem

Let the space of graph Laplacian matrices $\mathcal{L}_m$ be endowed with the Frobenius metric d_F. Suppose (A1), (A2) hold, then for a fixed $x \in \mathbb{R}$, it holds for $m(x), m_{L,h}(x)$, and $\hat{m}_{L,n}(x)$ as per (2), (11), and (12), respectively that

$$d_F\{m(x), m_{L,h}(x)\} = O(h^2),$$

$$d_F\{m_{L,h}(x), \hat{m}_{L,n}(x)\} = O_P\{(nh)^{-\frac{1}{2}}\}.$$

With $h \sim n^{-1/5}$, it holds that

$$d_F\{m(x), \hat{m}_{L,n}(x)\} = O_P(n^{-\frac{2}{5}}).$$

Rates of Convergence Using Frobenius Metric

Theorem

Furthermore, suppose (A3), (A4) hold, for a given closed interval $\mathcal{T}$, *if* $h \to 0, nh^2(-\log h)^{-1} \to \infty$ *as* $n \to \infty$, *then for any* $\varepsilon > 0$,

$$\sup_{x\in\mathcal{T}} d_F\{m(x), m_{L,h}(x)\} = O(h^2),$$

$$\sup_{x\in\mathcal{T}} d_F\{m_{L,h}(x), \hat{m}_{L,n}(x)\} = O_P(\max[(nh^2)^{-\frac{1}{2+\varepsilon}}, \{nh^2(-\log h)^{-1}\}^{-\frac{1}{2}}]).$$

With $h \sim n^{-1/(6+2\varepsilon)}$, *it holds that*

$$\sup_{x\in\mathcal{T}} d_F\{m(x), \hat{m}_{L,n}(x)\} = O_P(n^{-\frac{1}{3+\varepsilon}}).$$

Rates of Convergence Using Power Metric

Theorem

Let the space of graph Laplacian matrices $\mathcal{L}_m$ be endowed with the power metric $d_{F,\alpha}$. Suppose (A1), (A2) hold, then for a fixed $x \in \mathbb{R}$, it holds for $m(x), m_{L,h}(x)$, and $\hat{m}_{L,n}(x)$, respectively that

$$d_F\{m(x), m_{L,h}(x)\} = \begin{cases} O(h^2) & 0 < \alpha \leq 1 \\ O(h^{\frac{2}{\alpha}}) & \alpha > 1 \end{cases},$$

$$d_F\{m_{L,h}(x), \hat{m}_{L,n}(x)\} = \begin{cases} O_P\{(nh)^{-\frac{1}{2}}\} & 0 < \alpha \leq 1 \\ O_P\{(nh)^{-\frac{1}{2\alpha}}\} & \alpha > 1 \end{cases}.$$

With $h \sim n^{-1/5}$, it holds that

$$d_F\{m(x), \hat{m}_{L,n}(x)\} = \begin{cases} O_P(n^{-\frac{2}{5}}) & 0 < \alpha \leq 1 \\ O_P(n^{-\frac{2}{5\alpha}}) & \alpha > 1 \end{cases}.$$

Rates of Convergence Using Power Metric

Theorem

Furthermore, suppose (A3), (A4) hold, for a given closed interval $\mathcal{T}$, if $h \to 0, nh^2(-\log h)^{-1} \to \infty$ as $n \to \infty$, then for any $\varepsilon > 0$,

$$\sup_{x\in\mathcal{T}} d_F\{m(x), m_{L,h}(x)\} = \begin{cases} O(h^2) & 0 < \alpha \leq 1 \\ O(h^{\frac{2}{\alpha}}) & \alpha > 1 \end{cases},$$

$$\sup_{x\in\mathcal{T}} d_F\{m_{L,h}(x), \hat{m}_{L,n}(x)\} =$$

$$\begin{cases} O_P(\max\{(nh^2)^{-\frac{1}{2+\varepsilon}}, (nh^2(-\log h)^{-1})^{-\frac{1}{2}}\}) & 0 < \alpha \leq 1 \\ O_P(\max\{(nh^2)^{-\frac{1}{\alpha(2+\varepsilon)}}, (nh^2(-\log h)^{-1})^{-\frac{1}{2\alpha}}\}) & \alpha > 1 \end{cases}.$$

Rates of Convergence Using Power Metric

Theorem

With $h \sim n^{-1/(6+2\varepsilon)}$*, it holds that*

$$\sup_{x \in \mathcal{T}} d_F\{m(x), \hat{m}_{L,n}(x)\} = \begin{cases} O_P(n^{-\frac{1}{3+\varepsilon}}) & 0 < \alpha \leq 1 \\ O_P(n^{-\frac{1}{\alpha(3+\varepsilon)}}) & \alpha > 1 \end{cases}.$$

Computational Details

To implement $P_{\mathcal{L}_m}(B)$ where $B = (b_{ij})$ is a constant $m \times m$ matrix, we reformulate it into the following convex optimization problem.

$$\begin{array}{lrl}
\text{minimize} & f(L) = d_F^2(B, L) = \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m}(b_{ij} - l_{ij})^2 & \\
\text{subject to} & l_{ij} - l_{ji} = 0, & 1 \leq i, j \leq m, \\
 & \displaystyle\sum_{j=1}^{m} l_{ij} = 0, & 1 \leq i \leq m, \\
 & -W \leq l_{ij} \leq 0, & 1 \leq i \neq j \leq m,
\end{array}$$

where $L = (l_{ij})$ is a graph Laplacian.

Model Inference

The Fréchet R^2 coefficient of determination is defined as

$$R^2_\oplus = 1 - \frac{E[d^2\{L, m_G(x)\}]}{V_\oplus}.$$

The corresponding sample version is

$$\hat{R}^2_\oplus = 1 - \frac{\sum_{k=1}^n d^2\{L_k, \hat{m}_G(X_k)\}}{\sum_{k=1}^n d^2(L_k, \hat{\omega}_\oplus)},$$

where

$$\hat{\omega}_\oplus = \underset{\omega \in \Omega}{\arg\min} \frac{1}{n} \sum_{k=1}^n d^2(L_k, \omega)$$

is the sample Fréchet mean.

Model Selection

The adjusted Fréchet R^2 for a fitted submodel $\mathcal{M}$ using $q \leq p$ predictors is defined as

$$R_{\oplus,\mathrm{adj}}(\mathcal{M}) = R^2_{\oplus} - (1 - R^2_{\oplus})\frac{q}{n-q-1}. \tag{13}$$

where the sample version $\hat{R}_{\oplus,\mathrm{adj}}(\mathcal{M})$ can be obtained by plugging in $\hat{R}^2_{\oplus}$. Let $\mathcal{C}_q$ be the class of submodels using q predictors. Computing

$$q^* = \operatorname*{arg\,max}_{1\leq q\leq p} \max_{\mathcal{M}\in\mathcal{C}_q} \hat{R}^2_{\oplus,\mathrm{adj}}(\mathcal{M}), \tag{14}$$

the final model can then be taken as $\mathcal{M}^* = \operatorname{arg\,max}_{\mathcal{M}\in\mathcal{C}_{q^*}} \hat{R}^2_{\oplus,\mathrm{adj}}(\mathcal{M})$.

Another alternative for model selection is to minimize prediction error, which can be estimated by k-fold cross-validation.

Power Metric

The convexity of the target space is crucial in the proof of existence and uniqueness for the minimizers in (2)–(4) and (11)–(12). Indeed, as stated in Deutsch (2001, Chapter 12), every Chebyshev subset of a finite-dimensional Hilbert space is convex. Let K be a nonempty subset of the inner product space X, then K is called a Chebyshev subset if each $x \in X$ has exactly one best approximation in K. It can be shown that $F_\alpha(\mathcal{L}_m)$ as a subset of $\mathbb{R}^{m^2}$ is not convex, suggesting that it cannot be a Chebyshev subset. Hence uniqueness for the minimizers in (2)–(4) and (11)–(12) cannot be guaranteed. For this reason, we consider embedding $F_\alpha(\mathcal{L}_m)$ in $\mathcal{M}_m$ as defined in (10), where uniqueness for the minimizers in (2)–(4) and (11)–(12) can be ensured.

Hölder Continuity for F_α

Suppose that K is a set in $\mathbb{R}^{n_1}$, E is a non-empty subset of K, and $0 < \beta \leq 1$. A function $g : K \mapsto \mathbb{R}^{n_2}$ is uniformly Hölder continuous with exponent β and Hölder coefficient H in the set E, shortly (β, H)-Hölder continuous when there exists $H \geq 0$ such that

$$\|g(x) - g(y)\|_F \leq H\|x - y\|_F^\beta, \quad \text{for all } x, y \in E.$$

For $\beta = 1$ the function g is said to be Lipschitz continuous in E with Lipschitz constant H, shortly H-Lipschitz continuous.

Hölder Continuity for F_α

Proposition

Define $\mathcal{E}_m$ as $\{S \in \mathcal{S}_m^+ : \lambda_1(S) \leq C\}$, where $\lambda_1(S)$ is the largest eigenvalue of S and $C \geq 0$ is a constant. Then the matrix power map F_α is

- $(\alpha, m^{(1-\alpha)/2})$*-Hölder continuous in* $\mathcal{S}_m^+$ *for* $0 < \alpha < 1$,
- *and* $\alpha C^{\alpha-1}$*-Lipschitz continuous in* $\mathcal{E}_m$ *for* $\alpha \geq 1$.