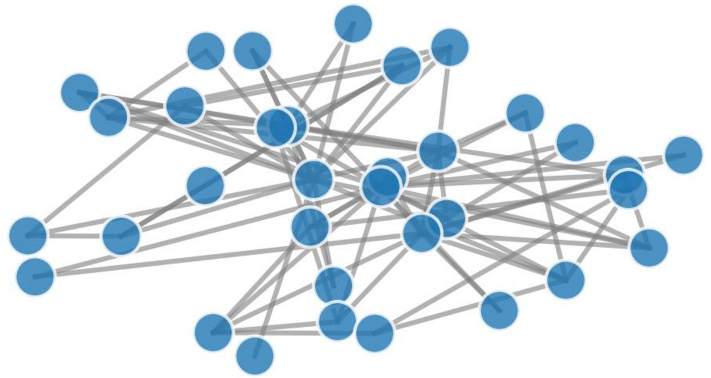
Data-Driven Reconstruction of Infrastructure Networks

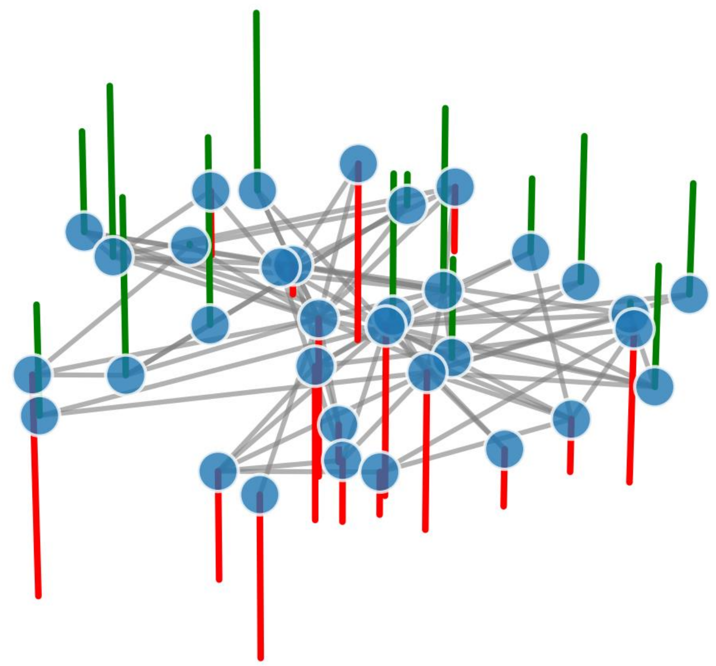
Estéban Nocet-Binois

Supervised by Jürgen Hackl

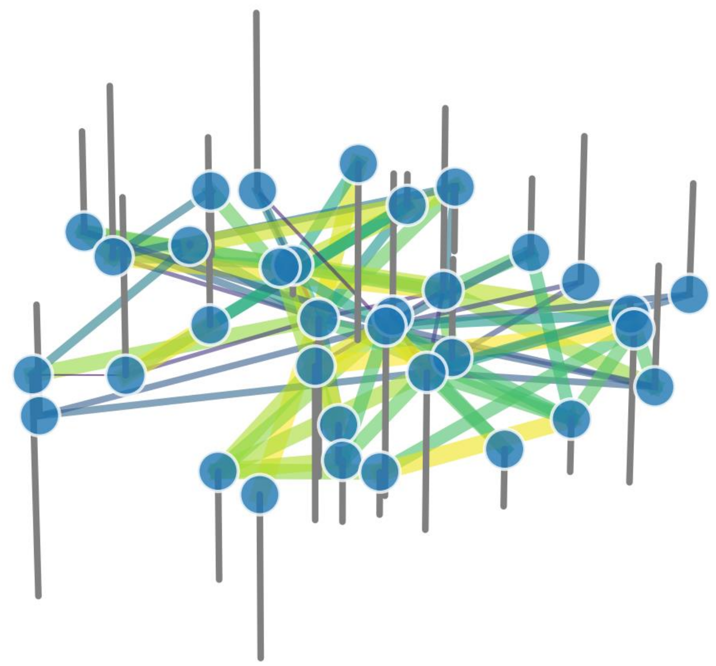
Infrastructures as Networks



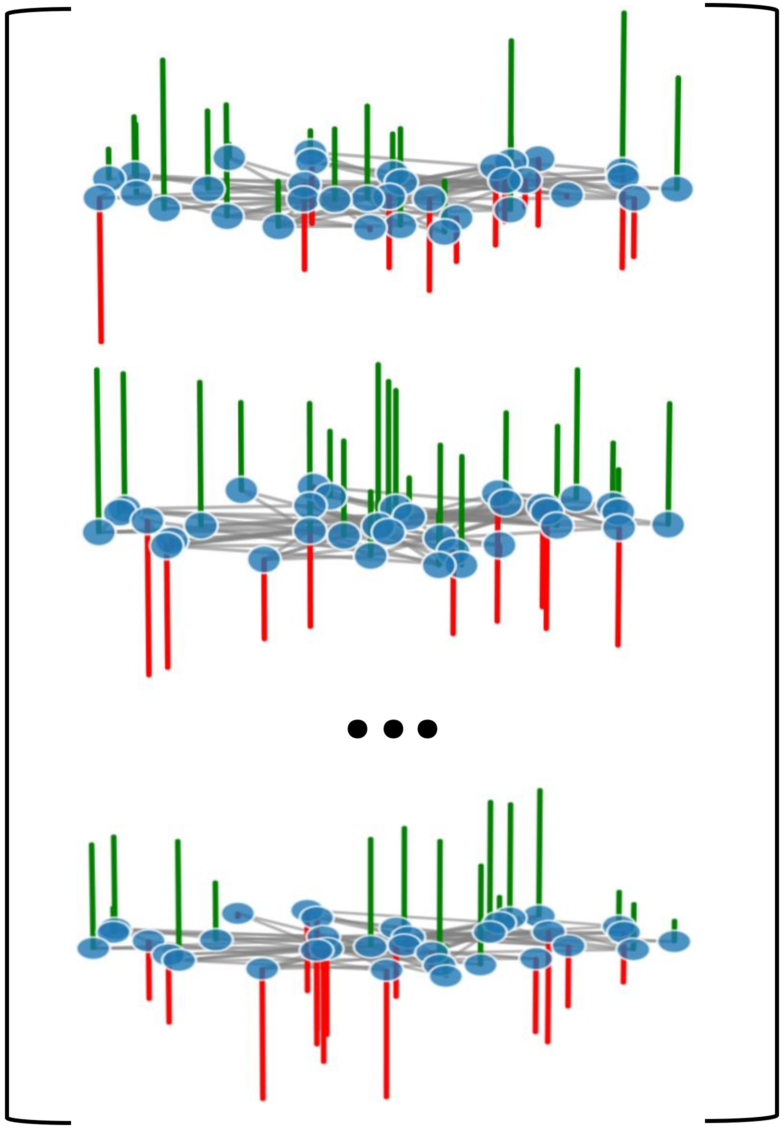
Signals on Graphs



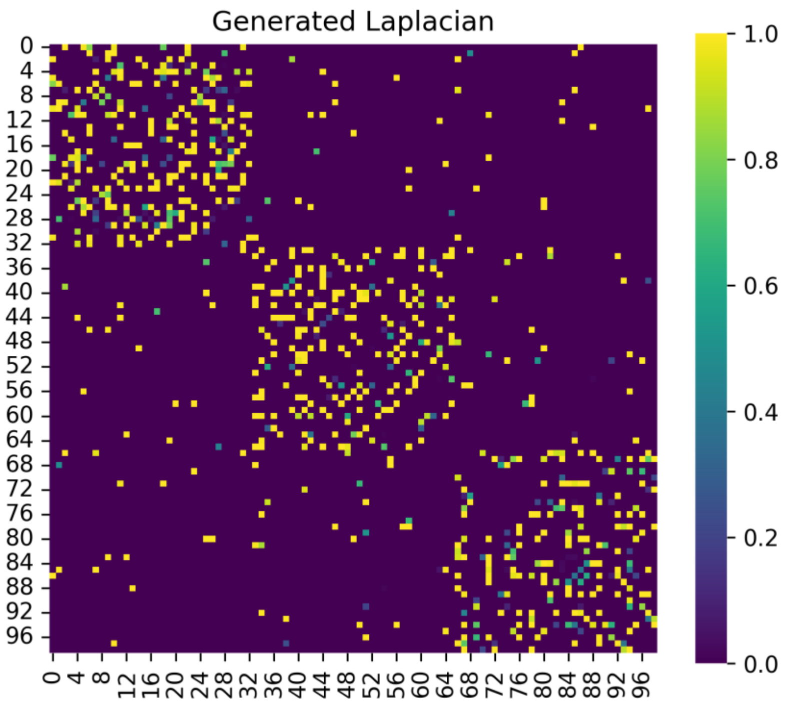
Data-driven Flows

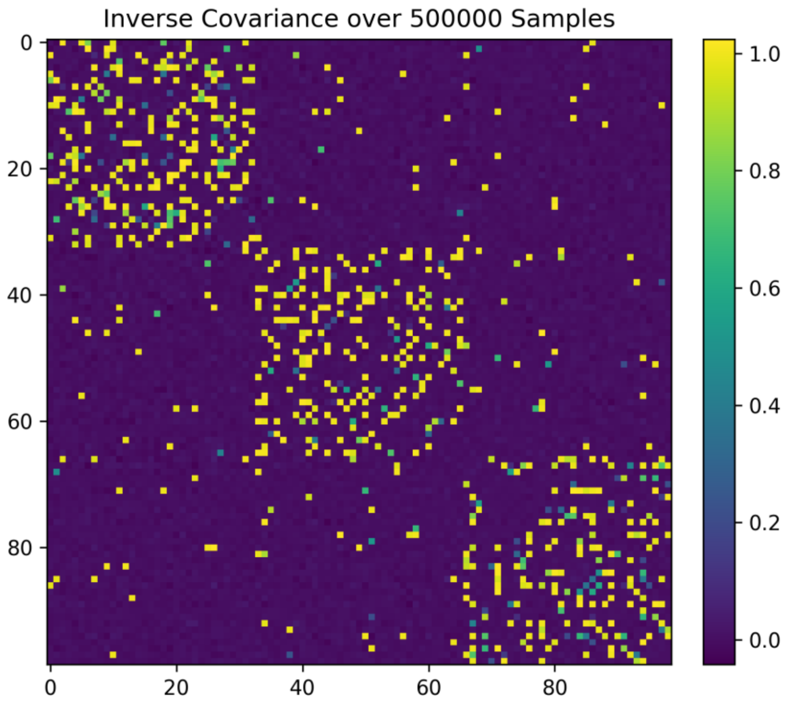


Signals Matrix



$$X = \begin{pmatrix} (x_{1,1} & x_{1,2} & \dots & x_{1,n}) \\ (x_{2,1} & x_{2,2} & \dots & x_{2,n}) \\ & \vdots & & \\ (x_{k,1} & x_{k,2} & \dots & x_{k,n}) \end{pmatrix}$$





Laplacian as Precision Matrix

$\mathbf{x}_i \sim \mathcal{N}(0, \mathbf{L}^\dagger)$

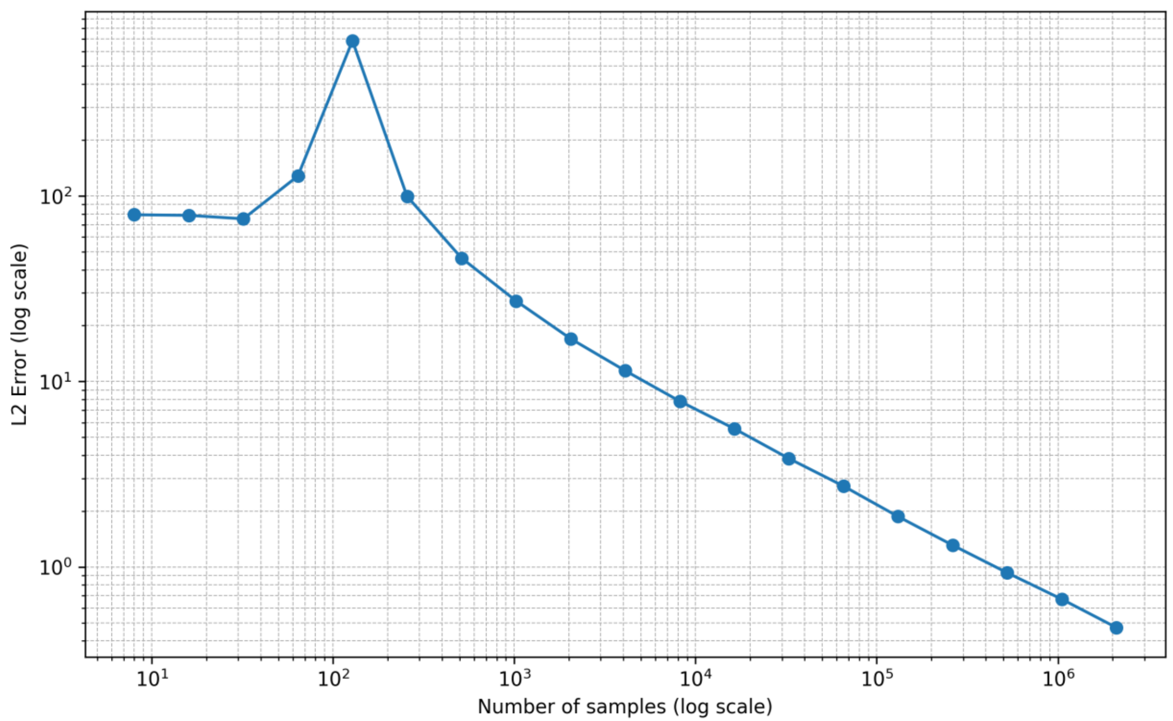
$\mathbf{L}^\dagger$ *encodes graph topology via* $\mathbb{E}[\Sigma]$.

Edge Interpretation

- $\boldsymbol{\Omega} = \mathbf{L}$: Precision matrix
- $\boldsymbol{\Omega}_{i,j} = -W_{i,j}$ (partial correlations)
- $\boldsymbol{\Omega}_{i,i} = D_i$ (degree)

Sample Complexity

- The main question is: How few signals would we need to estimate the graph Laplacian at best
- Limitations:
  - Rank deficiency of the covariance matrix
  - Gaussian noise
  - Non-linear effects



When the sample covariance matrix $\Sigma = X^T X/m$ is ill-conditioned or rank-deficient (i.e. when $m < n$), its pseudo-inverse $(L^*)^\dagger$ becomes numerically unstable.

Maximum Likelihood Estimation

For m i.i.d. samples $X = [\mathbf{x}_1, \ldots, \mathbf{x}_m]$, the log-likelihood becomes:

$$\ln p(X|L) = \sum_{i=1}^{m} p(\mathbf{x}_i|L) = -\frac{mn}{2}\ln(2\pi) + \frac{m}{2}\ln|L| - \frac{1}{2}\underbrace{\sum_{i=1}^{m}\mathbf{x}_i L \mathbf{x}_i^T}_{m\mathrm{tr}(L\Sigma)}$$

Dong et al. (2016)

$$\hat{L}_{\text{MLE}} = \arg\max_{L\in\mathcal{L}} \left\{ \log \det^{\dagger}(L) - \mathrm{tr}(L\Sigma) \right\}$$

where $\mathcal{L} = \{L \succeq 0 \mid L\mathbf{1} = 0,\ L_{ij} \le 0\, \forall i \neq j\}$ and $\det^{\dagger}(L)$ is the pseudo-determinant.

Egilmez et al. (2017)

$$\hat{L} = \arg\min_{L\in\mathcal{L}} \left\{ \mathrm{tr}(L\Sigma) + \alpha\|L\|_F^2 \right\}$$

where $\alpha > 0$ and $\mathcal{L}$ as before.

Purpose:

- **Convex**: Strongly convex due to Frobenius norm.
- **Stabilizes** estimation when Σ is rank-deficient ($m < n$).

Medvedovsky et al. (2024)

$$\hat{L} = \arg\min_{L\in\mathcal{L}} \left\{ \mathrm{tr}(L\Sigma) + \alpha\psi(L) + \beta\sum_{i\neq j}\rho(L_{ij}) \right\}$$

where $\psi(L)$ can be $-\log\det^{\dagger}(L)$, and $\rho(\cdot)$ is a sparsity penalty (e.g., MCP, ℓ_1).

Purpose:

- Promotes **sparsity** in the learned graph (edge selection).
- Non-convex for MCP/SCAD, but tractable under restricted convexity.

Limitations

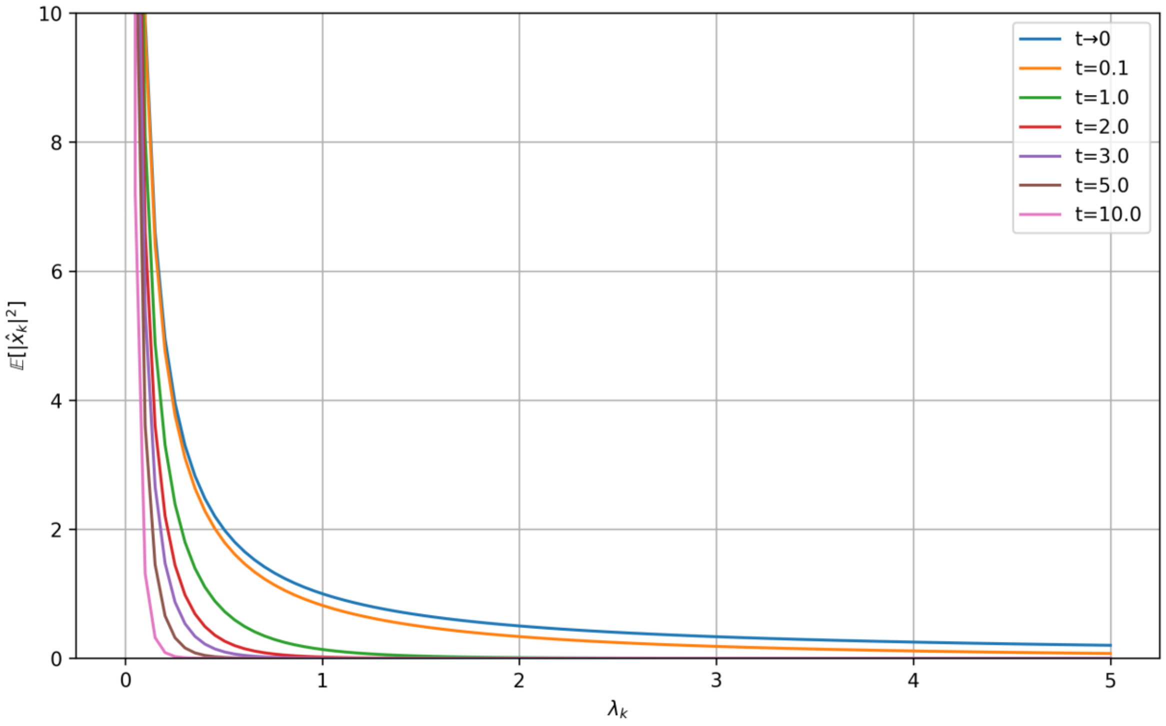
All these methods assume, *in expectation*, the covariance structure to reflect edge weights. Certain non-linear effects can break this assumption.

- Data corruption, and incomplete data
- **Phase shifts**, or time lags
- Interference leading to **scale asymmetry**
- Sampling rates increase *autocorrelation*

Two nodes with identical time series patterns but shifted in time, or with different overall magnitudes, will yield a larger Euclidean distance (hence smaller similarity).

Diffusion Map

- Construct $\mathbf{S}_{\text{DM}} = \boldsymbol{\Psi}_t \boldsymbol{\Psi}_t^\top$, where $\boldsymbol{\Psi}_t$ are diffusion map embeddings at time t:
$$\boldsymbol{\Psi}_t = \mathbf{U}\boldsymbol{\Gamma}_t,\ \boldsymbol{\Gamma}_t = \mathrm{diag}(e^{-t\lambda_1}, \ldots, e^{-t\lambda_{n-1}}, 0)$$
- $\boldsymbol{\Psi}_t$ encode multiscale geometry via heat kernel $e^{-t\mathbf{L}}$. Each row encodes a node's position in a diffusion geometry, where distances approximate connectivity.
- The heat kernel approaches the identity matrix as $t \to 0$: $e^{-t\lambda_k} \approx 1 - t\lambda_k$.



Beyond i.i.d.

Hierarchical Model

- Kronecker product:

$$\boldsymbol{\Sigma} = \underbrace{\mathbf{L}}_{\text{space}} \otimes \underbrace{\mathbf{K}}_{\text{time}}$$

- Matrix variate Gaussian density:

$$\mathbf{X} \sim \mathcal{MN}(\mathbf{0}, \mathbf{L}, \mathbf{K})$$

Kronecker PCA

- Approximates the covariance matrix using a sum of Kronecker products

Challenges

- Identifiability: Separating space/time correlations
- Scalability: Kronecker decomposition scales as $O(n^3 + m^3)$

Open Problems

- Can we generalize diffusion-based approaches to distortion-based ones?
- Can we define non-separable kernels?
- Learning $\mathbf{L}$ from non-i.i.d. time series
- Scalable, $O(n^3)$ eigendecomposition-free, learning
- Theoretical guarantees for nonlinear dynamics

Thanks for listening !

- Estéban Nocet-Binois
- en4624@princeton.edu
- https://cis.princeton.edu/