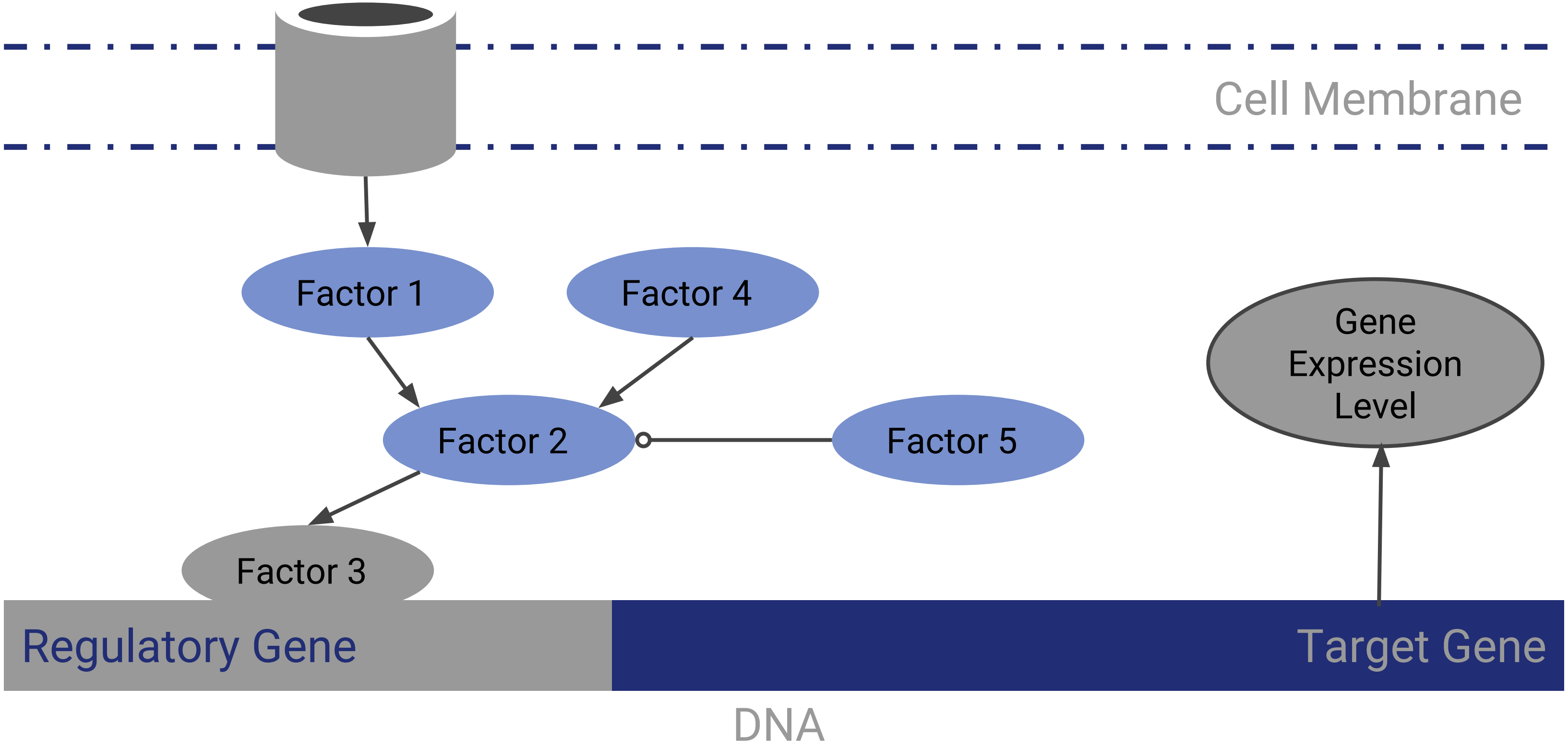
Causal Computational Models for

Gene Regulatory Networks

Sahil Loomba
Parul Jain

Advisors
Dr. Sumeet Agarwal
Dr. Parag Singla

Reintroducing the GRN Problem



Where BTP1 finished…

- Correlation $$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$
- Granger Causality $$x_t = a_0 + \sum_{i=1}^{m} a_i x_{t-i} + \sum_{i=1}^{q} b_i y_{t-i} + \epsilon_t$$
- Mutual Information $$I(X,Y) = H(X) - H(X|Y)$$
- Transfer Entropy $$T(X,Y) = T_{Y\to X} = H(X_t|X_{t-1:t-d}) - H(X_t|X_{t-1:t-d}, Y_{t-1:t-d})$$

Parameters: Size, quantisation, time, lag

Asides: Grid Search, Laplace Smoothing

TE ~ MI > GC > CO

… is where BTP2 picks up

	Linear	Non Linear
Non Predictive	Correlation	Mutual Information
Predictive	Granger Causality	Transfer Entropy

As Random Variable

Convergent Cross Mapping

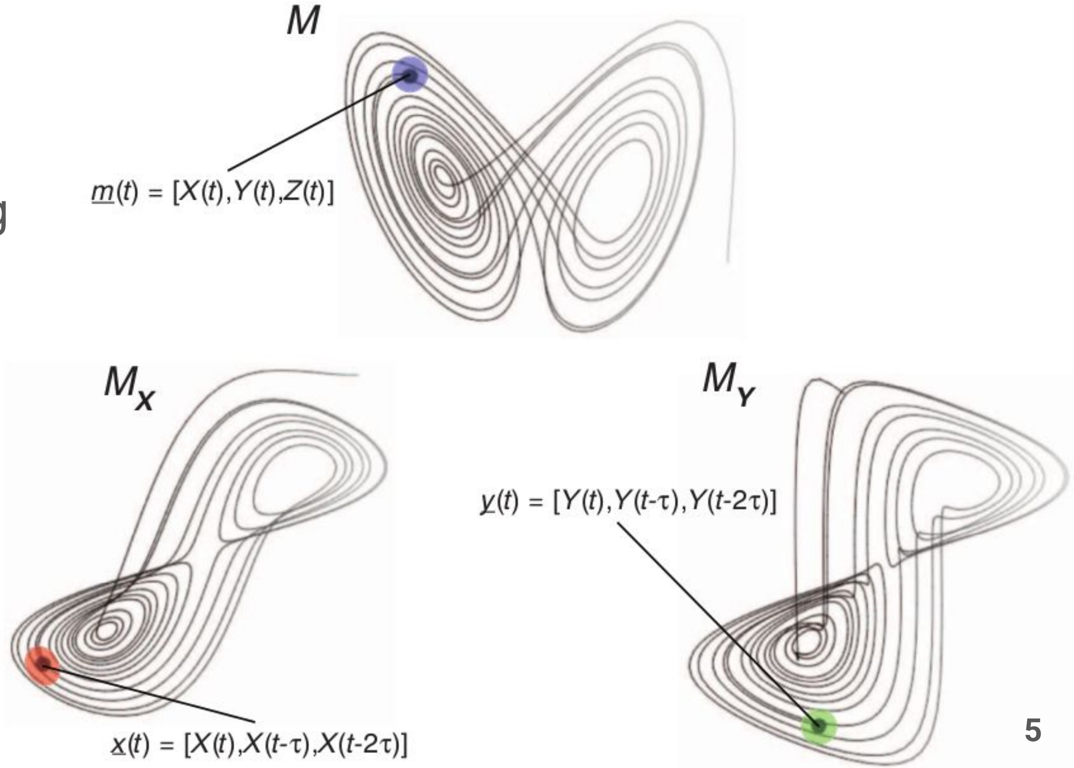
As Dynamical System

Convergent Cross Mapping [Sugihara et al. 2012]

Diffeomorphism across Shadow Manifolds

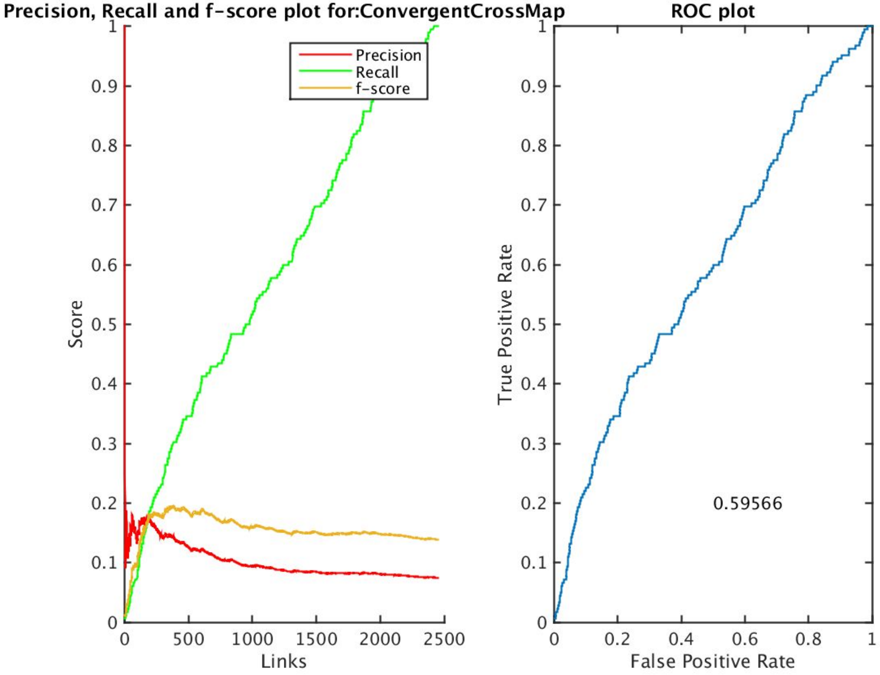
- For weakly coupled dynamical systems
- New notion of causality: belonging to same dynamical system
- Library size ∝ Time series length
- Parameters:
  - Dimensions: M, E where E ≥ M
  - Lag: τ

$$C(X, Y) = \rho(X, \hat{X}_Y)$$

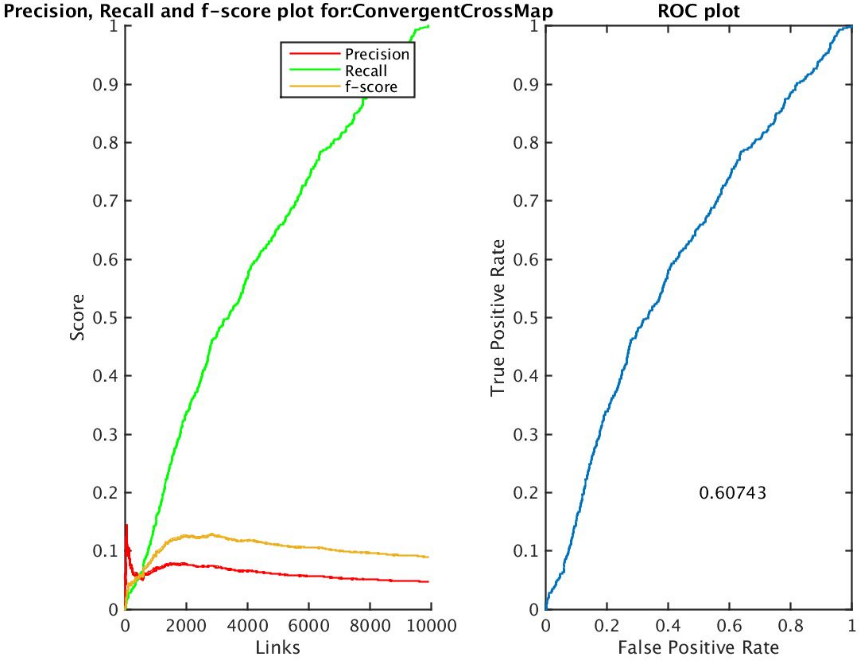


Convergent Cross Mapping

Some Results

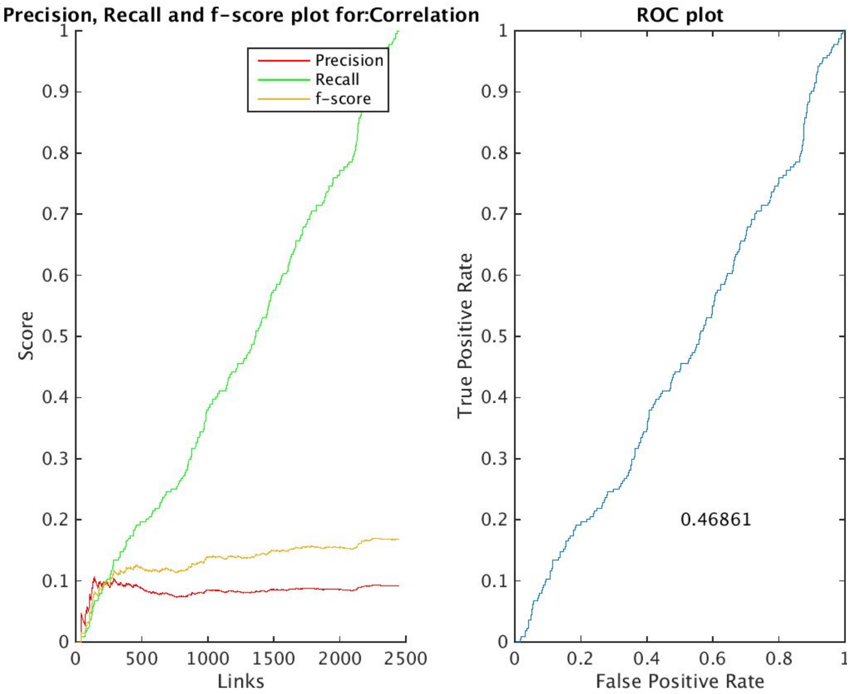


Network Size 50, Time Series Length = 500

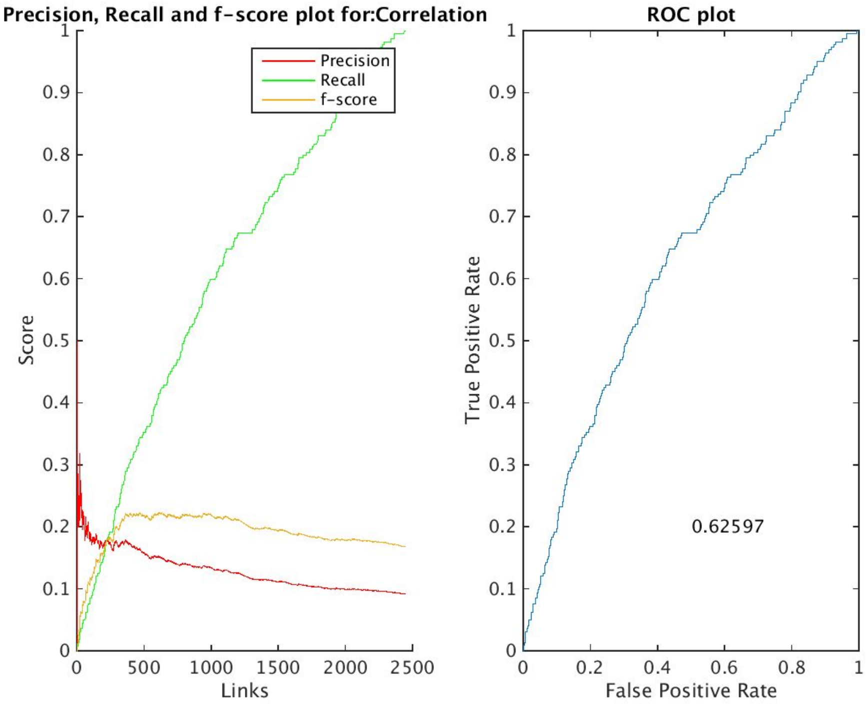


Network Size 100, Time Series Length = 500

Normalisation should be the norm!



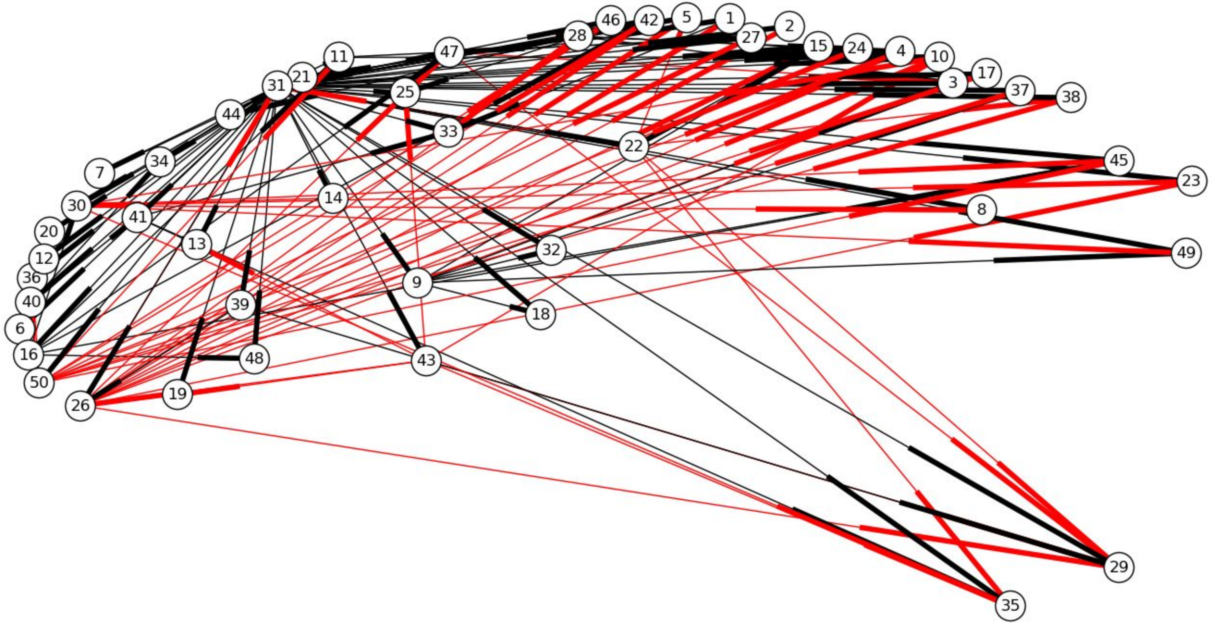
Network Size 50, Time Series Length = 1000



Normalised Time Series data

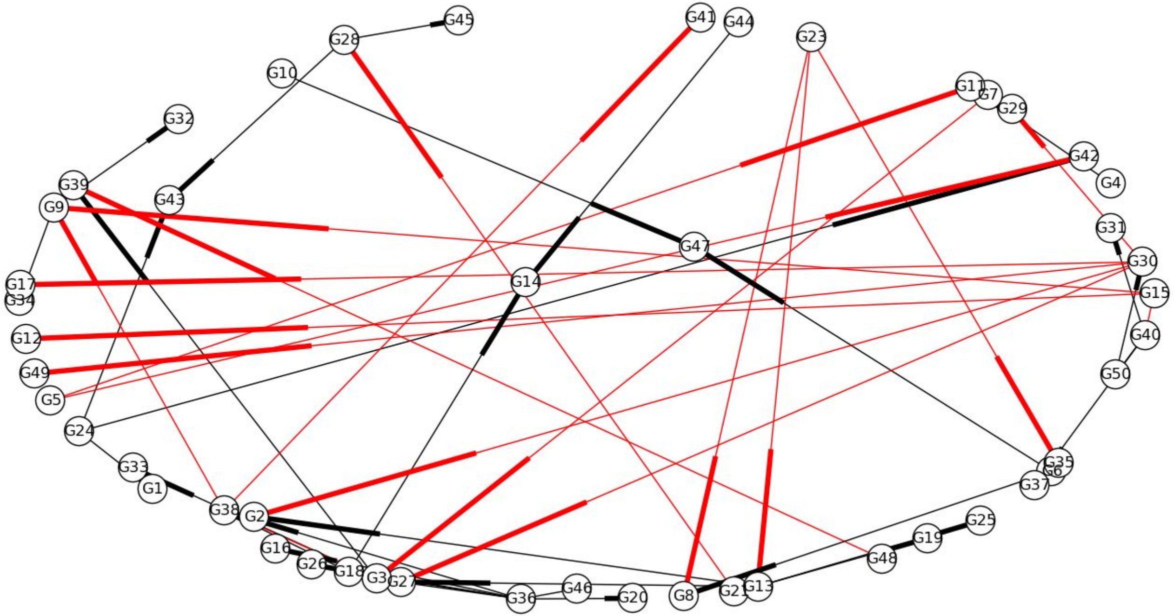
High degrees are degrading!

SysGenSIM



Network Size 50, Average Degree = 7

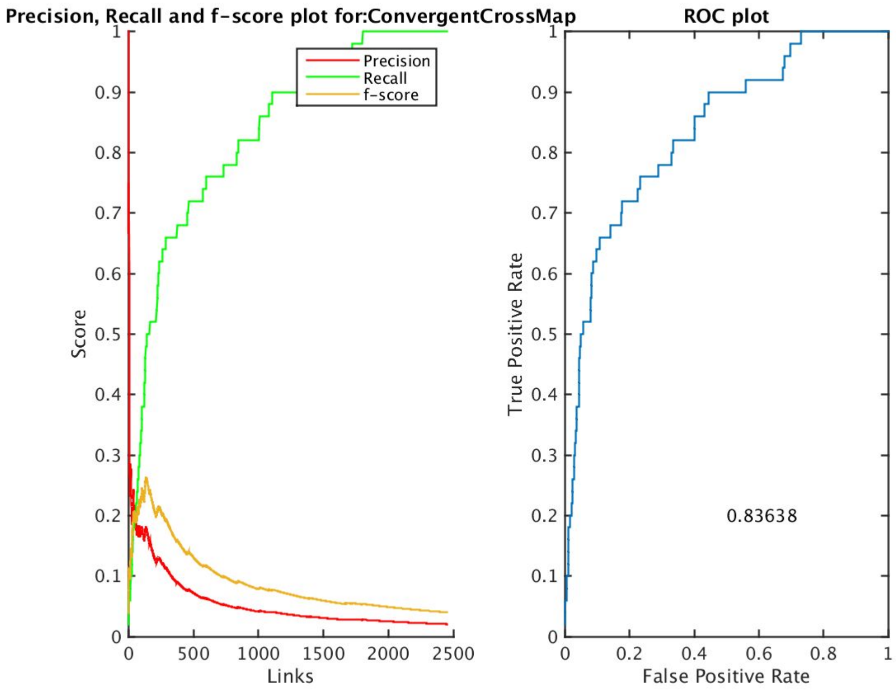
High degrees are degrading!



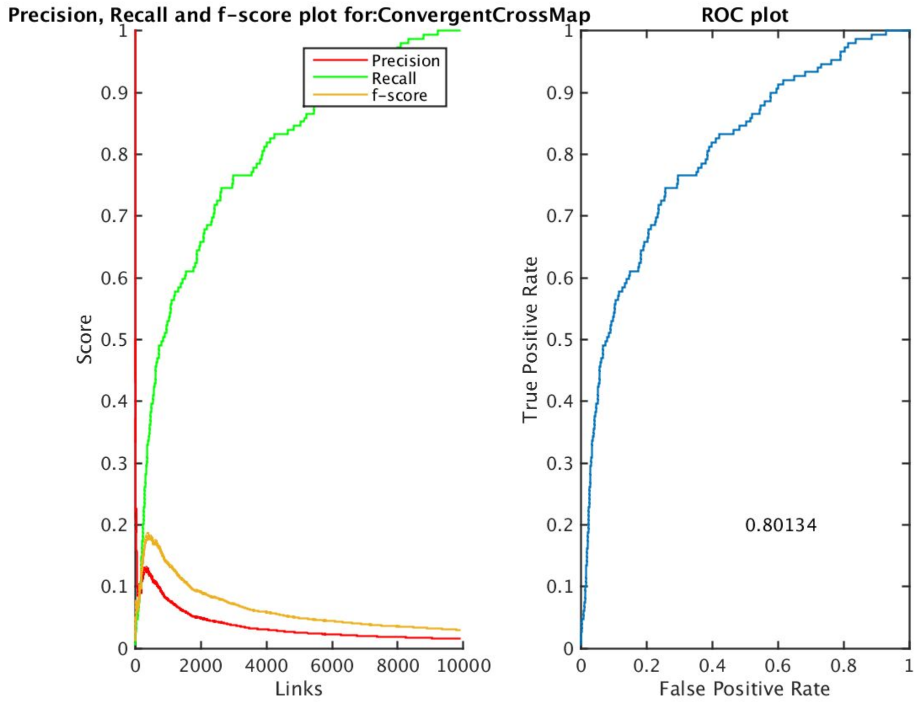
Network Size 50, Average Degree = 2

Convergent Cross Mapping

Some Results



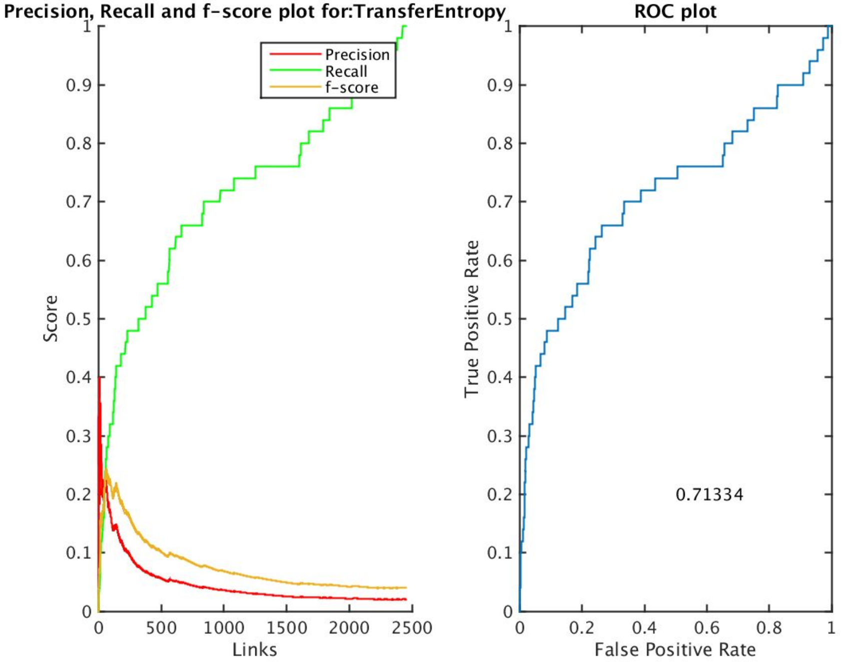
Network Size 50, Time Series Length = 1000



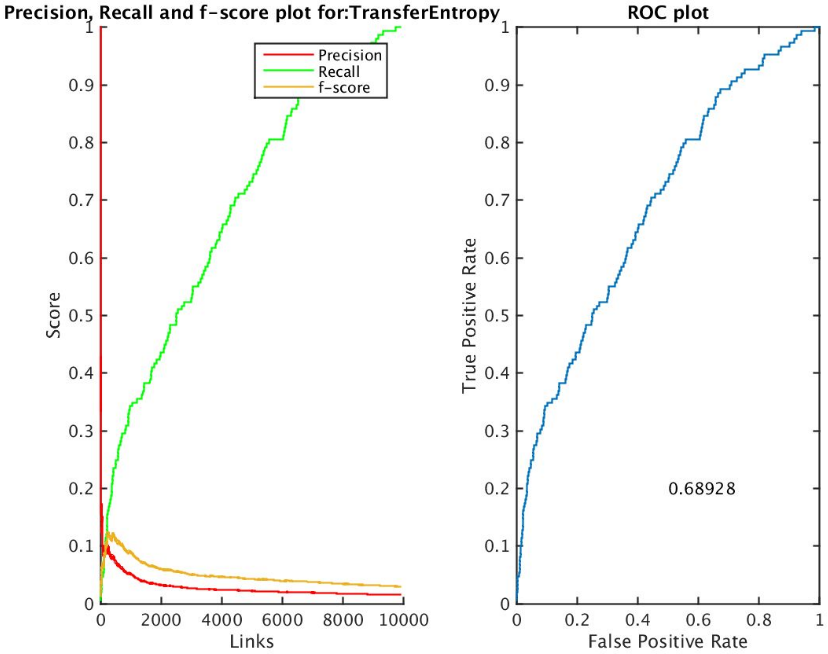
Network Size 100, Time Series Length = 1000

Transfer Entropy

Some Results



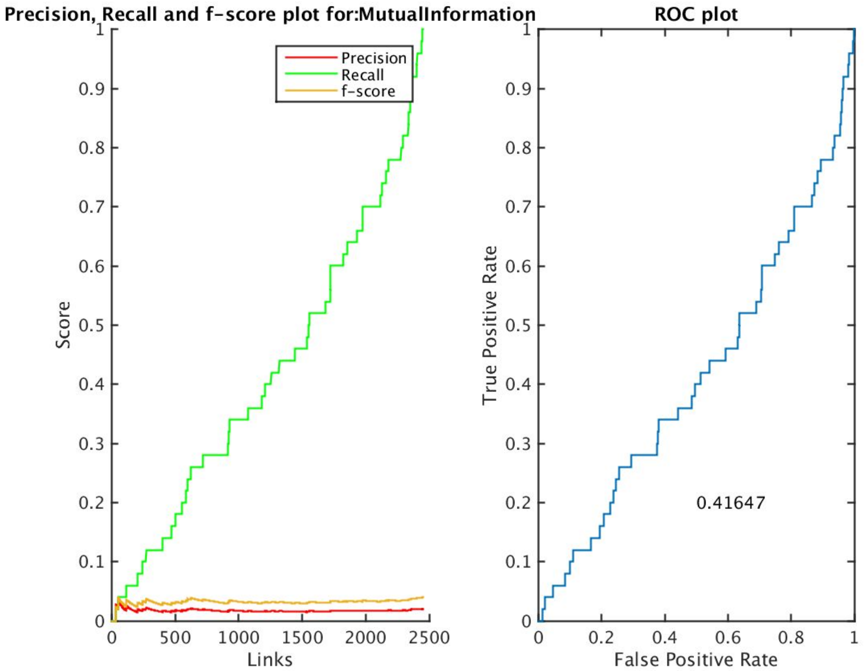
Network Size 50, Time Series Length = 1000
Quantization Level = 10



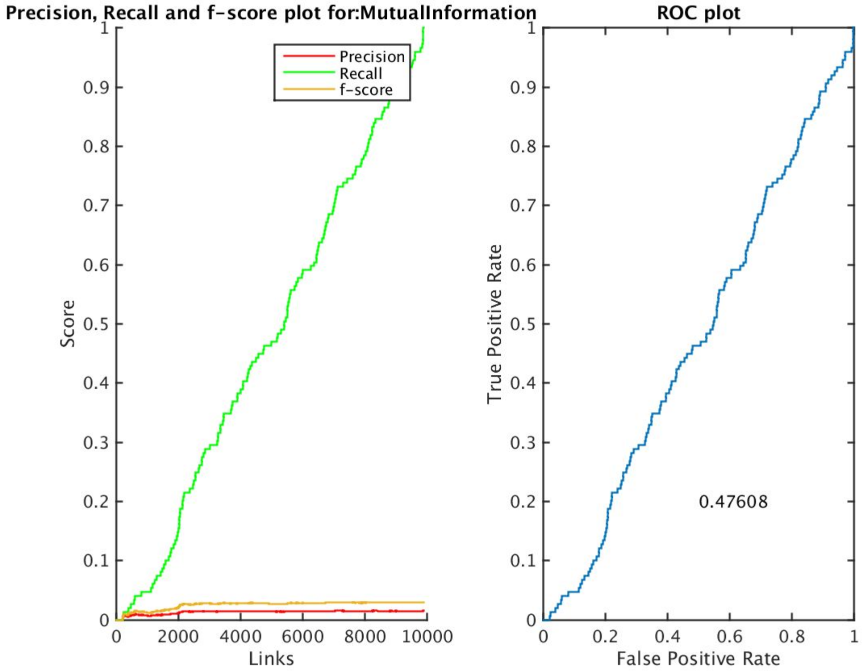
Network Size 100, Time Series Length = 1000
Quantization Level = 10

Mutual Information

Some Results

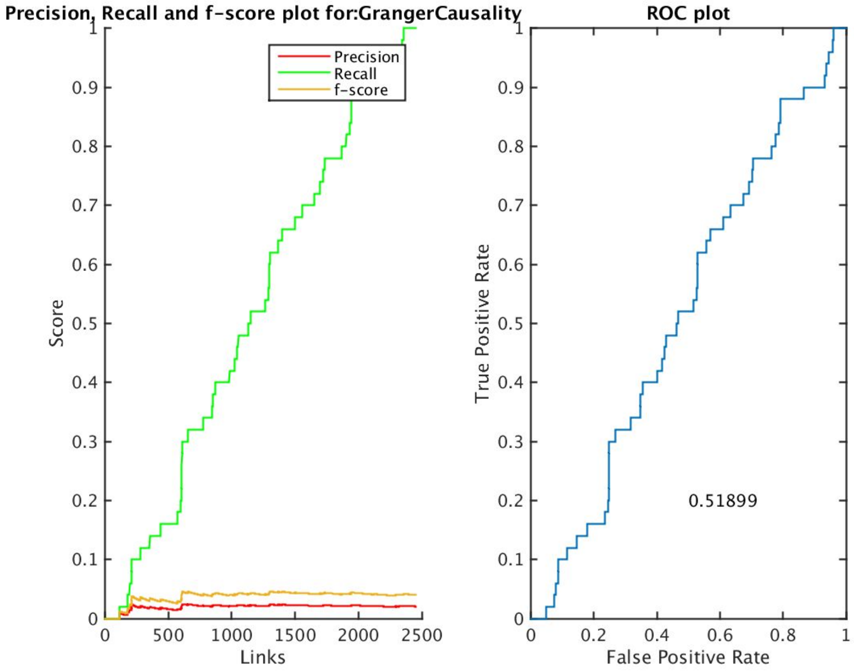


Network Size 50, Time Series Length = 1000
Quantization Level = 10

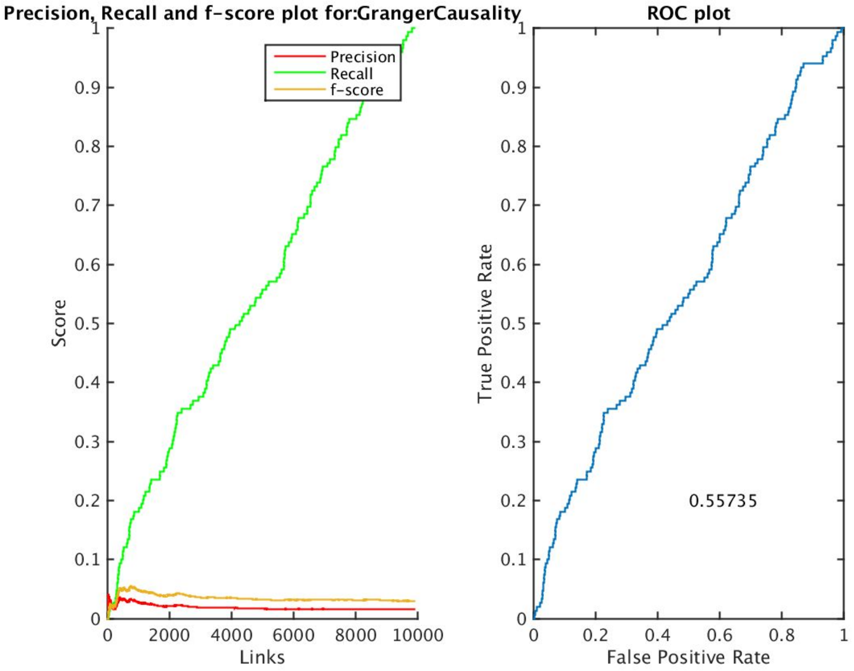


Network Size 100, Time Series Length = 1000
Quantization Level = 10

Granger Causality
Some Results



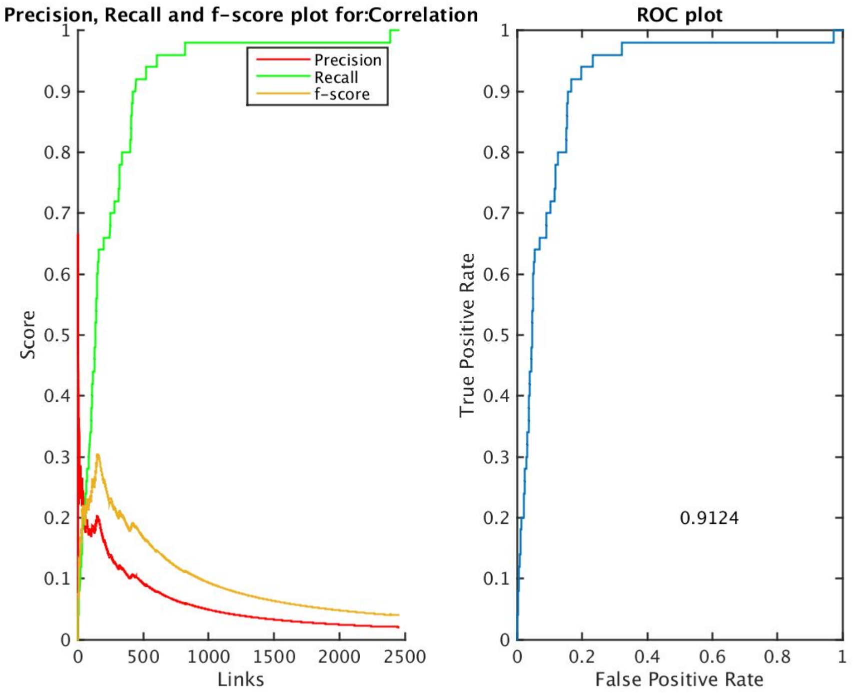
Network Size 50, Time Series Length = 1000



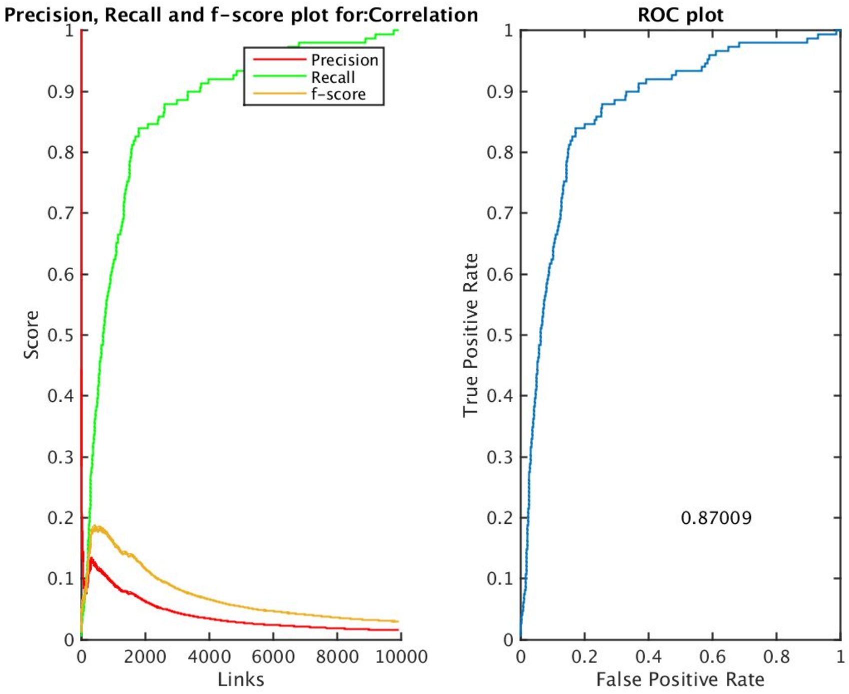
Network Size 100, Time Series Length = 1000

Correlation

Some Results



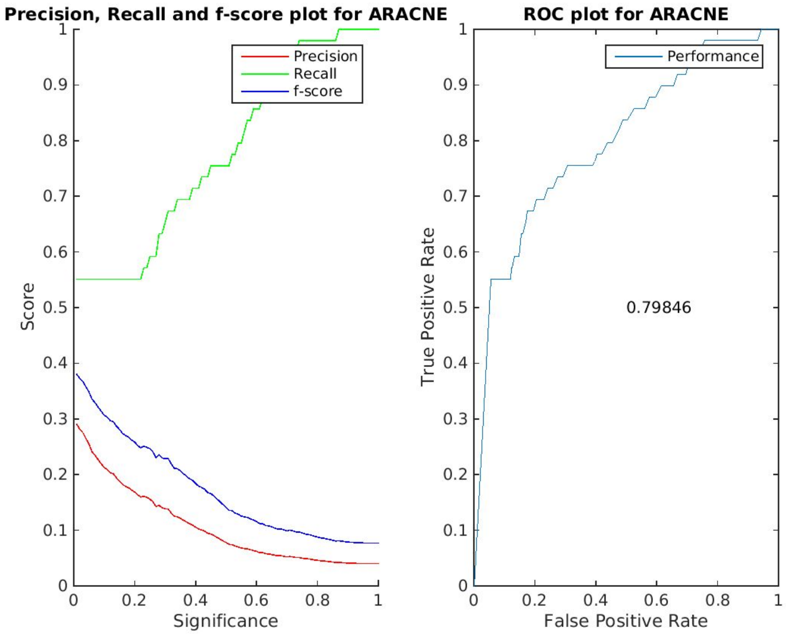
Network Size 50, Time Series Length = 1000



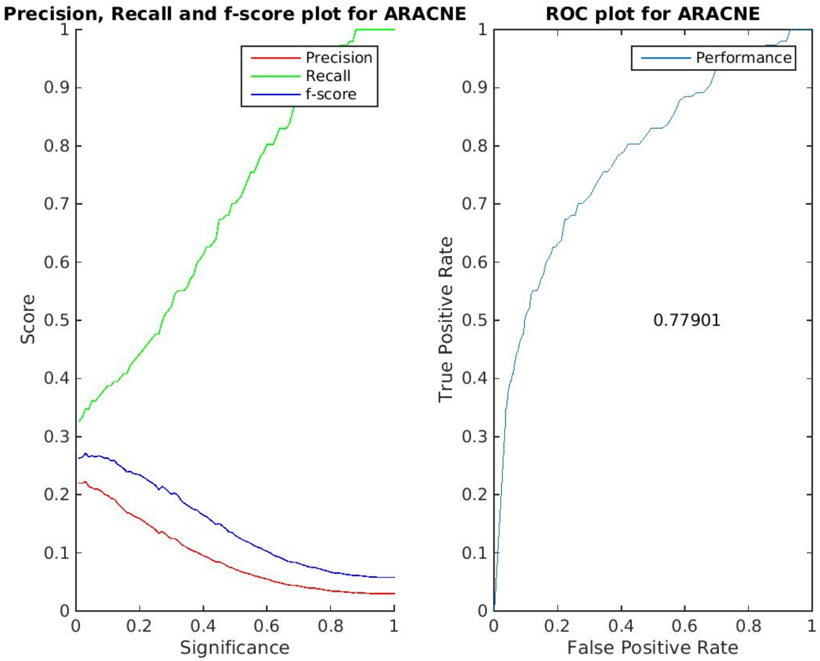
Network Size 100, Time Series Length = 1000

ARACNE [Margilin et al. 2006]

Graph Structure Estimation



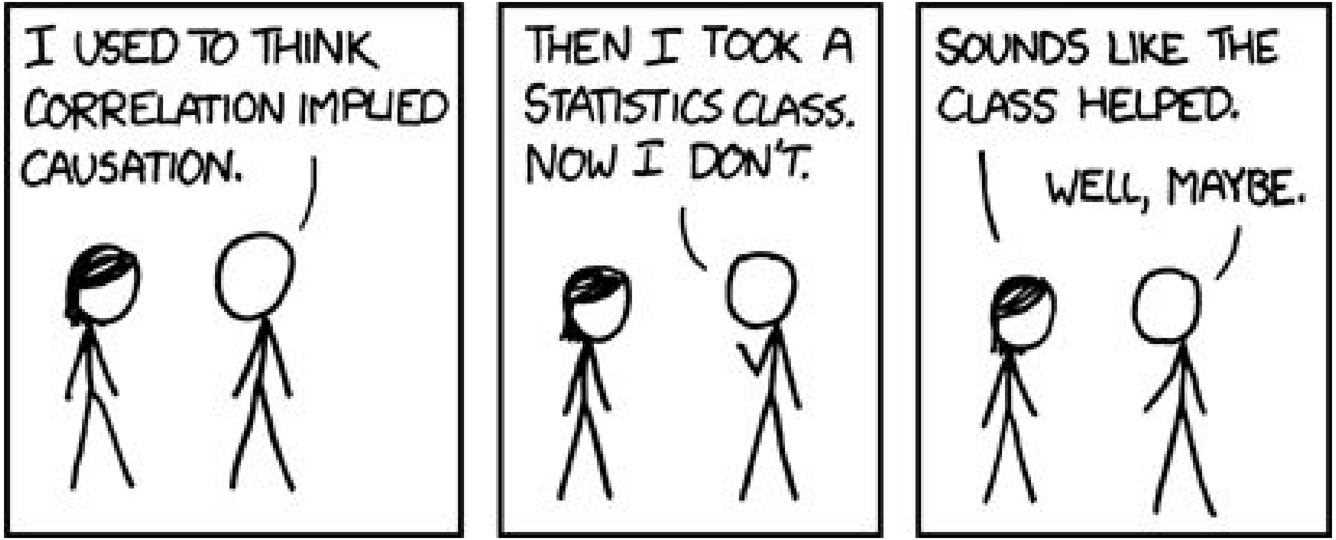
Network Size 50, Time Series Length = 1000



Network Size 100, Time Series Length = 1000

Pairwise Metrics of Causality

A Summary



CCM and Correlation *seem* to work the best at a pairwise level

Naive Edge Selection

1. (Since all pairwise metrics are directly proportional to the strength of causality,) sort ${}^{n}C_{2}$ edges by metric value in decreasing order.
2. Choose top-k edges and output as graph G.

Smart Edge Selection

Future Work

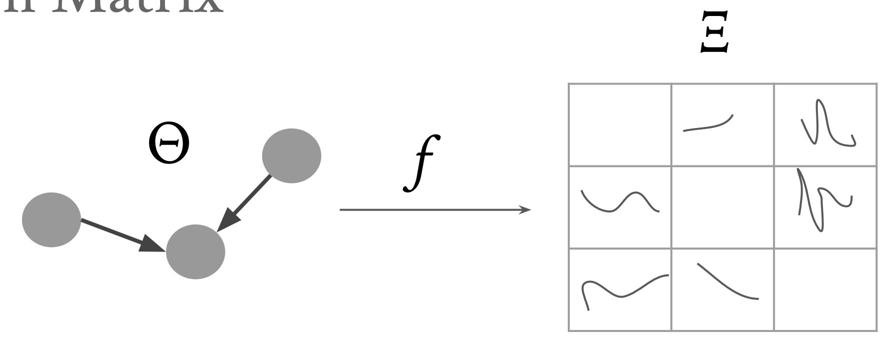
Use a “sophisticated” algorithm which selects top-k edges, by making use of graph connectivity and other constraints information.

Intrinsic Graph Structure Estimation [Hino et al. 2015]

Adjacency Matrix to Observation Matrix

$$f : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$$

$$\Theta \mapsto f(\Theta) = \Xi$$



(Pairwise Metrics used here)

$$\xi_{ij} = c_i + c_{ij}\theta_{ij} + \sum_{k \in V} c_{ij}^{k} \theta_{ik}\theta_{kj} + \sum_{k,l \in V} c_{ij}^{kl} \theta_{ik}\theta_{kl}\theta_{lj} + \ldots$$

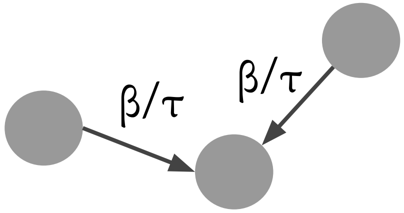
$$t_{ij} = \xi_{ij} + \epsilon$$

Intrinsic Graph Structure Estimation [Hino et al. 2015]

The Random Walk Model

$$L(\Theta) = \begin{bmatrix} \sum \theta_{1k} & -\theta_{12} & \cdots & -\theta_{1n} \\ -\theta_{21} & \sum \theta_{2k} & \cdots & -\theta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -\theta_{n1} & -\theta_{n2} & \cdots & \sum \theta_{nk} \end{bmatrix}$$

Digraph Laplacian



$$f(\Theta) = \alpha e^{\beta L(\Theta)}$$

$$\rho = \{\alpha, \beta\}$$

Intrinsic Graph Structure Estimation [Hino et al. 2015]

Parameter Estimation Algorithm

$$J(\rho, \Theta) = \sum_{i,j \in V, i \neq j} \left(t_{ij} - [f(\Theta)]_{ij} \right)^2$$

In an EM style algorithm, iterating over k (number of edges in graph):

$$\Theta^k = f^{-1}(\Xi)$$

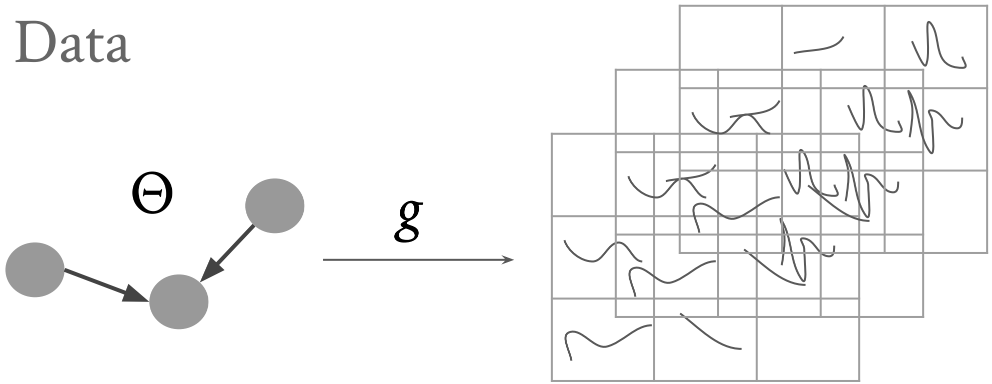
$$\rho_k = \arg\min_{\rho} J(\rho, \Theta^k)$$

Intrinsic Graph Structure Estimation

Moving towards Multi-attribute Data

$$g : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n \times r}$$

$$\Theta \mapsto g(\Theta) = \Xi$$



(Multiple Pairwise Metrics used here)

$$^{q}\xi_{ij} = {}^{q}c_i + {}^{q}c_{ij}\theta_{ij} + \sum_{k \in V} {}^{q}c_{ij}^{k}\theta_{ik}\theta_{kj} + \sum_{k,l \in V} {}^{q}c_{ij}^{kl}\theta_{ik}\theta_{kl}\theta_{lj} + \ldots$$

$$^{q}t_{ij} = {}^{q}\xi_{ij} + \epsilon$$

Intrinsic Graph Structure Estimation

Moving towards Multi-attribute Data

$$g(\Theta) = cat\left({}^{1}f(\Theta), {}^{2}f(\Theta), \ldots, {}^{r}f(\Theta)\right)$$

$$J(\rho, \Theta) = \sum_{1 \leq q \leq r} \left(\sum_{i,j \in V, i \neq j} \left({}^{q}t_{ij} - {}^{q}[g(\Theta)]_{ij}\right)^2 \right)$$

$$\Theta^k = \frac{\sum_{1 \leq q \leq r} {}^{q}f^{-1}(\Xi)}{r}$$

$${}^{q}\rho_k = \arg\min_{\rho} J(\rho, \Theta^k)$$

Ideally, Θ should be exactly mapped by every ${}^{q}f^{-1}(\Xi)$

Intrinsic Graph Structure Estimation

More Considerations & Postprocessing

- Imposing extra constraints on Θ
- Explore different ways to map g to Θ (hard versus soft constraints)
- For reduced computational complexity, use a small window of k
- Further postprocessing: Data Processing Inequality

$$g_1 \leftrightarrow \ldots \leftrightarrow g_2 \leftrightarrow \ldots \leftrightarrow g_3$$

$$I(g_1, g_3) \leq min[I(g_1, g_2), I(g_2, g_3)]$$

Thank You

Questions and Feedback