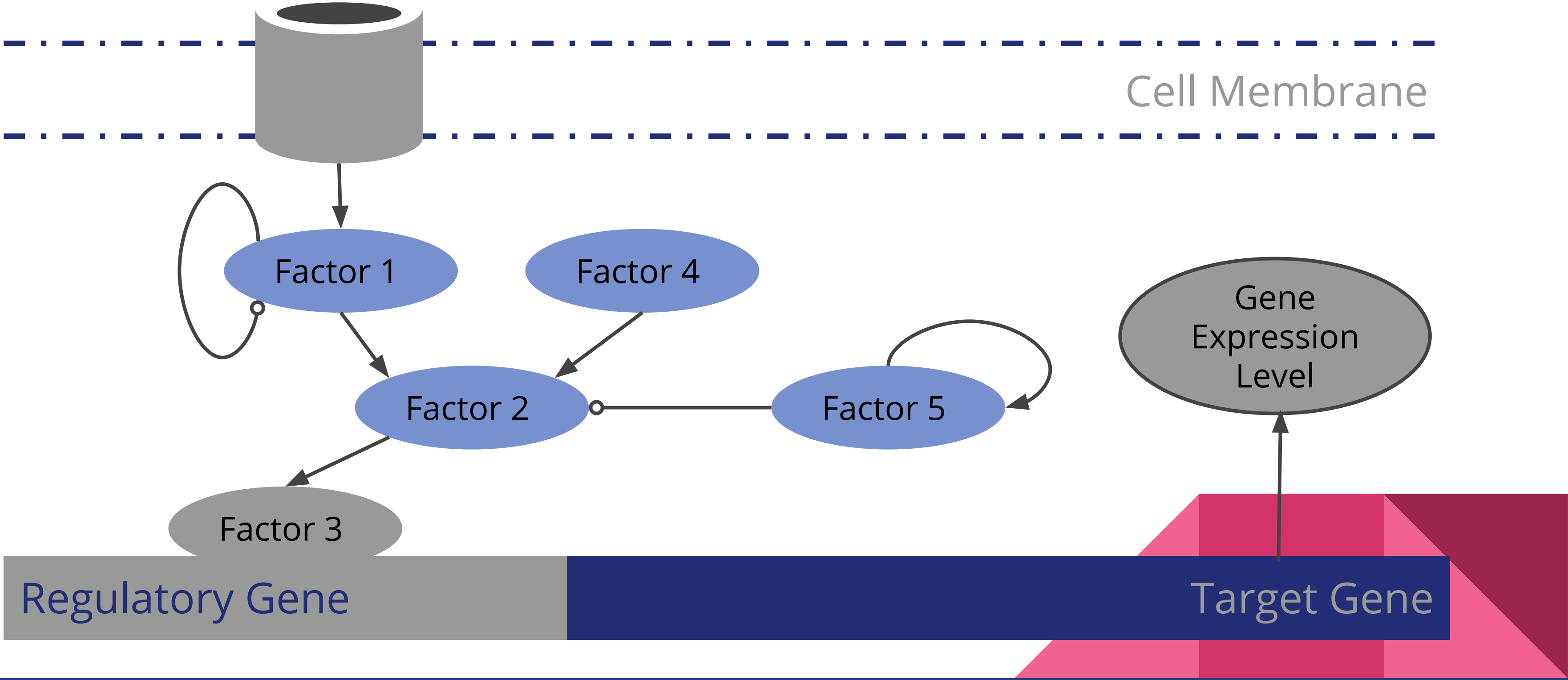
Causal Computational Models for Gene Regulatory Networks

Parul Jain
Sahil Loomba

Advisors
Dr. Sumeet Agarwal
Dr. Parag Singla

Gene Regulatory Networks



Problem Statement

To create computational models of GRNs,

which capture causal interactions between the genes ,

with emphasis on reducing dimensionality of the problem,

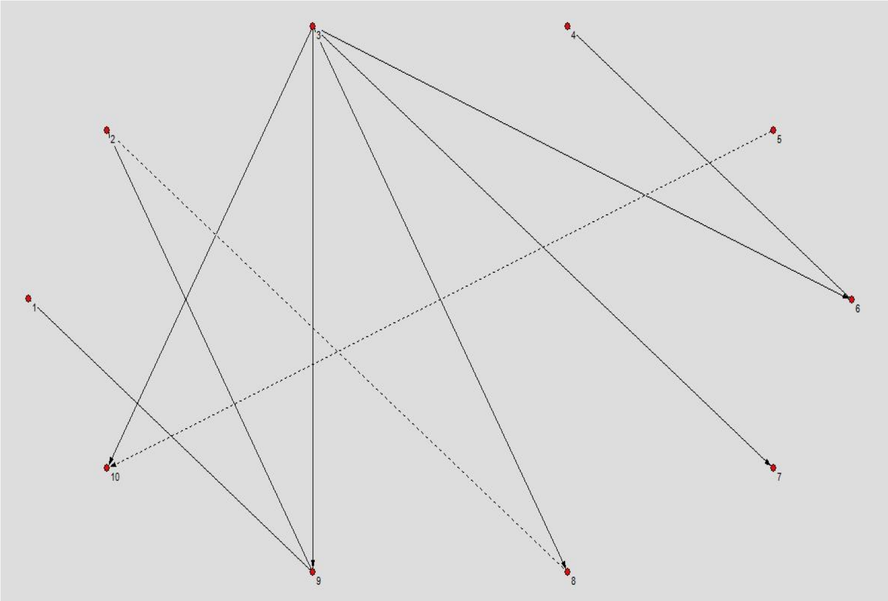
to allow wet lab work for in/validation for disease networks.

Data

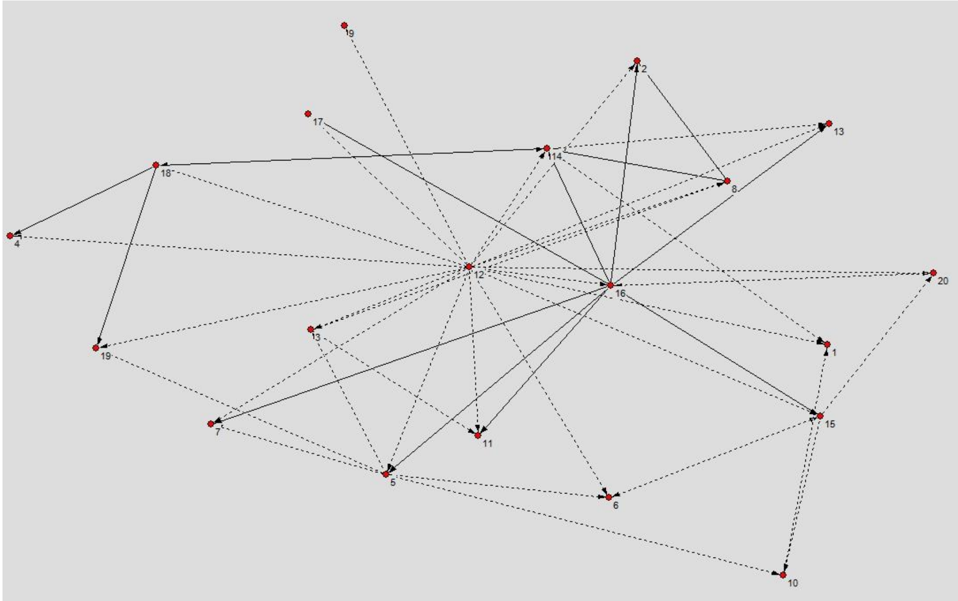
- ➢ SysGenSIM software for synthetic data
- ➢ Scale free networks of some average node degree
- ➢ Networks of size 10, 20, 50 and 100
- ➢ Both 500 and 1000 timesteps for ever network size

$$\frac{\mathrm{d}\,G_g}{\mathrm{d}\,t} = Z_g^c \cdot V_g \cdot \theta_g^{syn} \cdot \prod_k \left(1 + A_{k,g} \frac{G_k^{h_{k,g}}}{G_k^{h_{k,g}} + (K_{k,g}/Z_k^t)^{h_{k,g}}}\right) - \lambda_g \cdot \theta_g^{deg} \cdot G_g$$

Networks

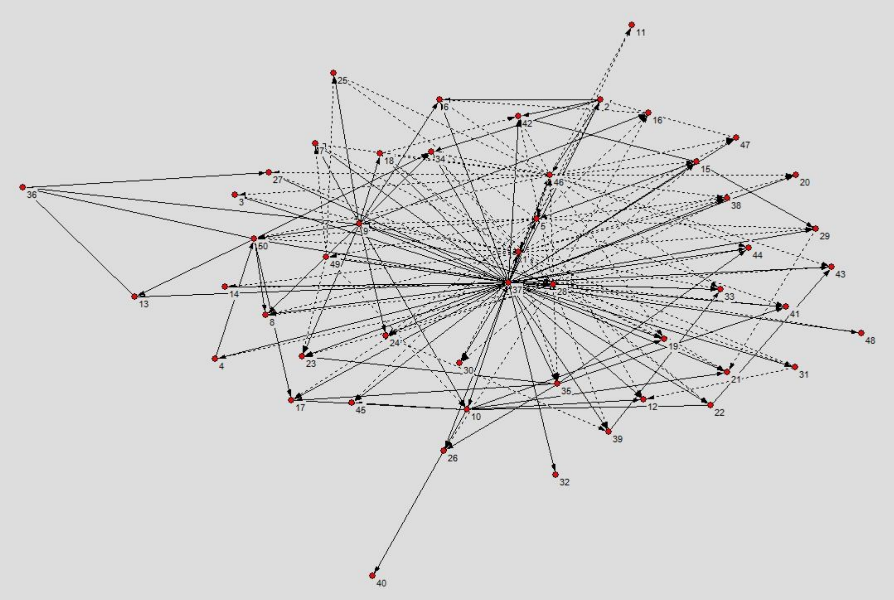


Network Size 10

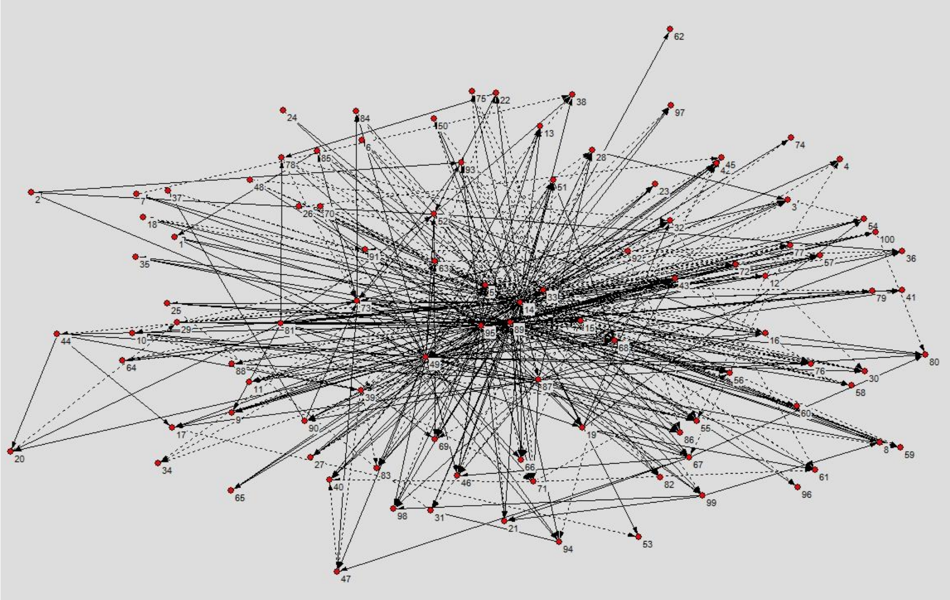


Network Size 20

Networks

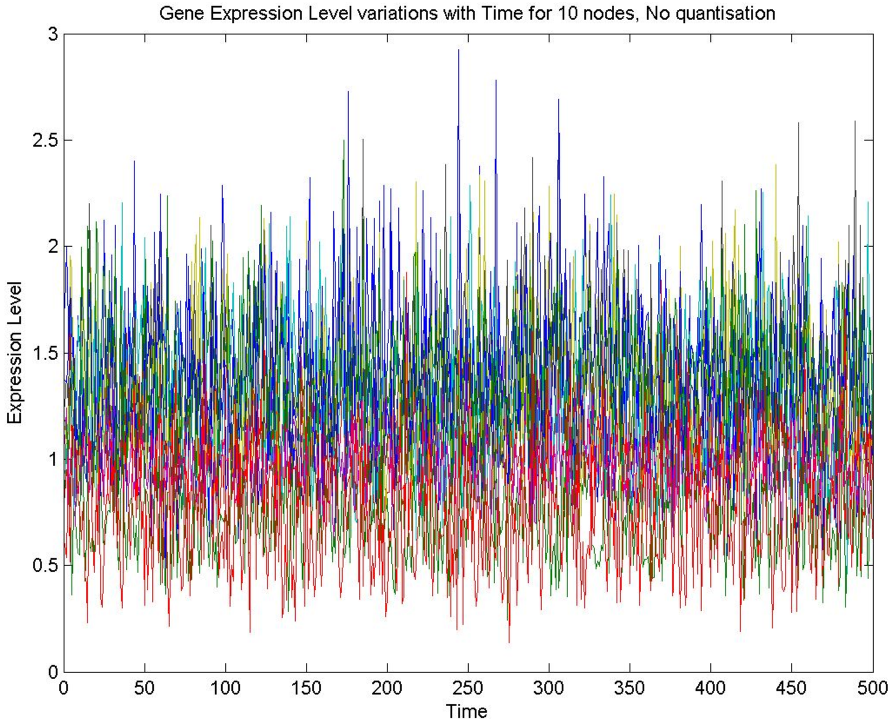


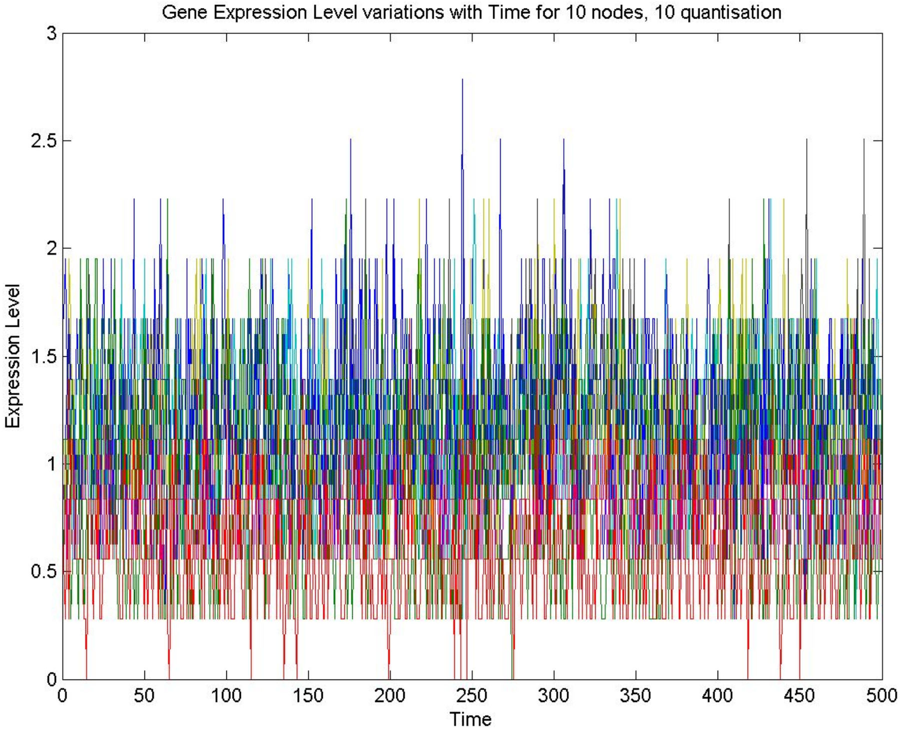
Network Size 50



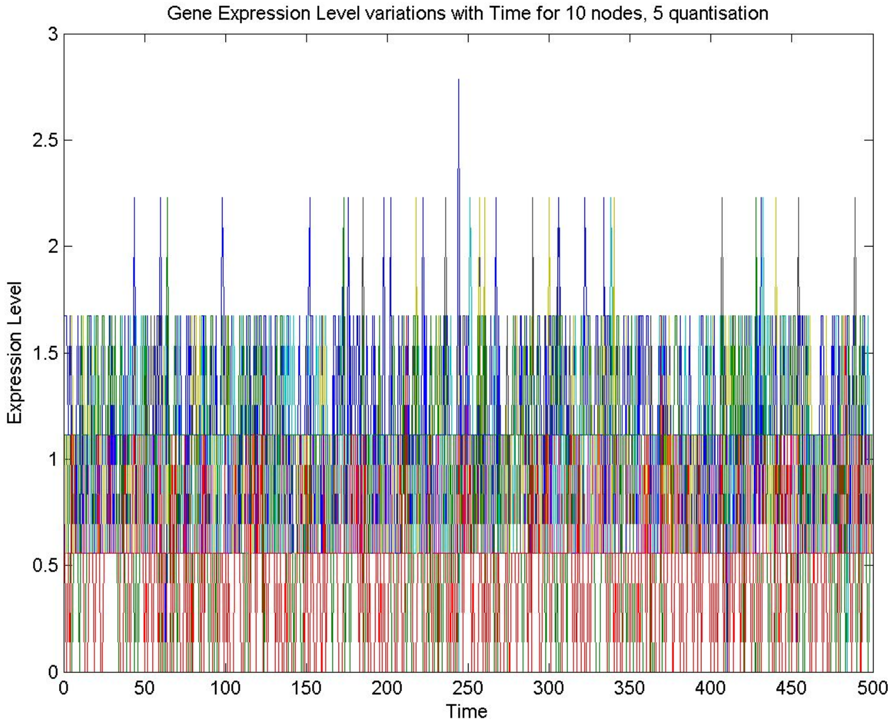
Network Size 100

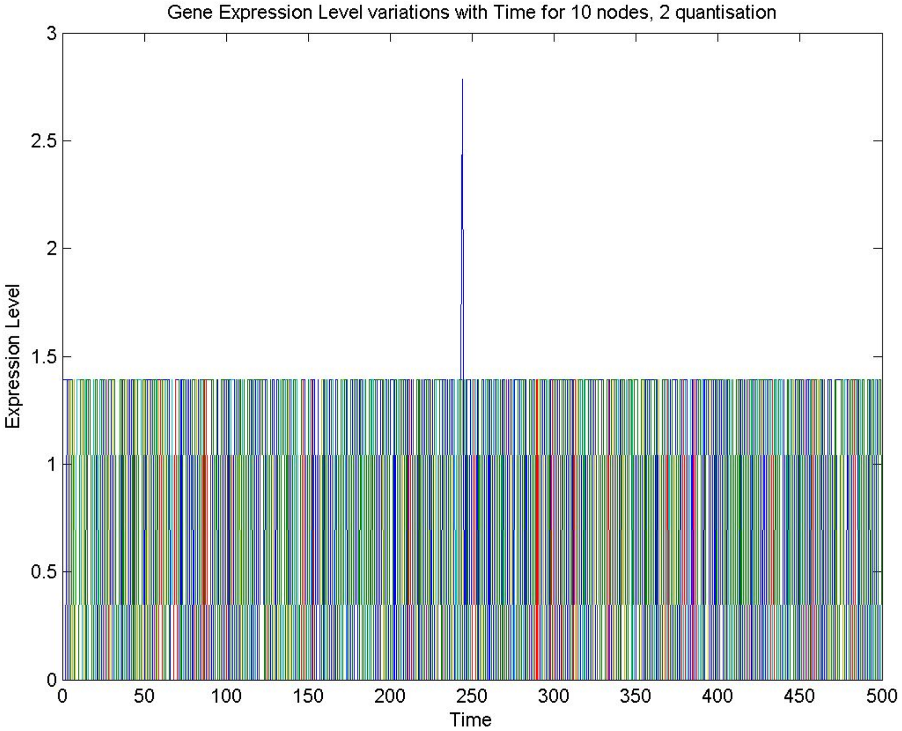
Data Quantisation





Data Quantisation





Techniques

- ➢ **Correlation** - lagged correlation coefficient for time series data, along with lag value to infer directionality
- ➢ **Granger Causality** - X causes Y iff prediction of Y is significantly better, given past X and Y, as compared to with Y alone
    - Regression residue analysis
    - F-statistic
- ➢ **Mutual Information** - measure of mutual dependence between X and Y
- ➢ **Transfer Entropy** - X causes Y iff prediction of Y is significantly better, given past X and Y, as compared to with Y alone
    - Probabilistic prediction analysis

Techniques In a Nutshell

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

Parameters

- Size
- Quantisation Levels
- Time Series Length
  - Time Lag
- Techniques
  - Smoothing

Parameters

- Size [10, 20, 50, 100]
- Quantisation Levels [2, 5, 10, 20]
- Time Series Length [500, 1000]
  - Time Lag
- Techniques [ra, co, gc, mi, te]
  - Smoothing
- > 150 experiment runs

Parameters - Time Lag

- Model Selection using Bayesian Information Criterion, for Granger Causality

$$BIC = -2 \log(L) + k \log(n)$$

- Maximum possible information transfer, for others (max_lag = 5)

Results - Correlation

Size : 10

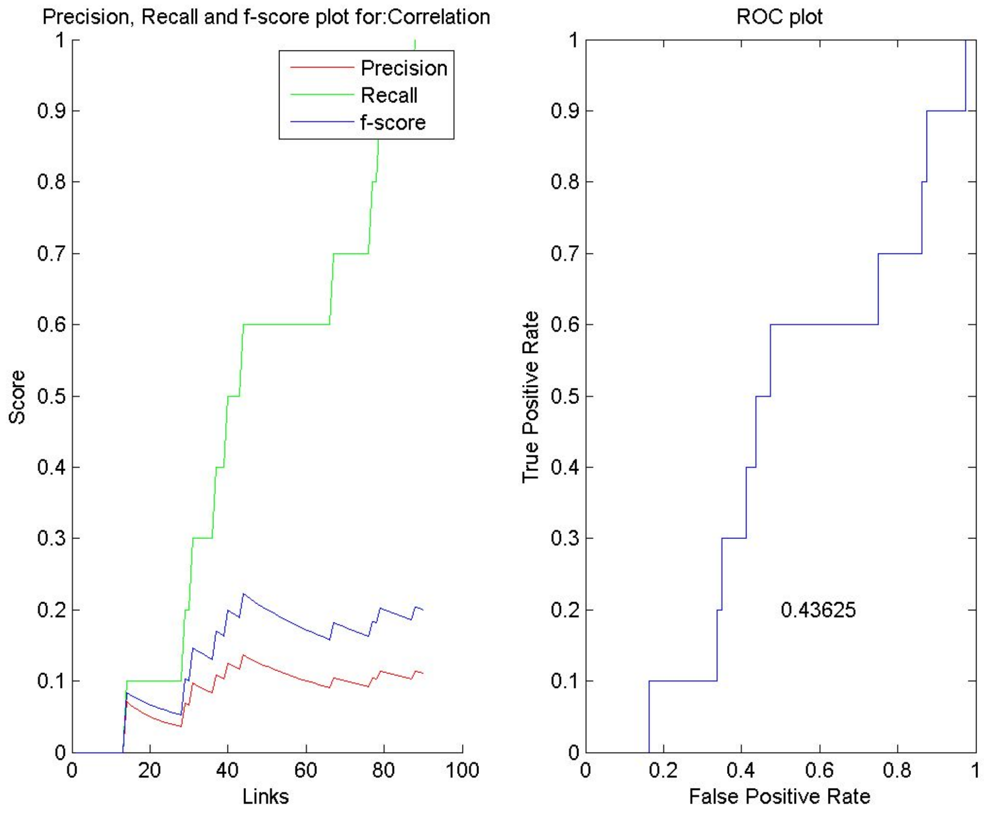
Type : Scale free

Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 2



Results - Correlation

Size : 10

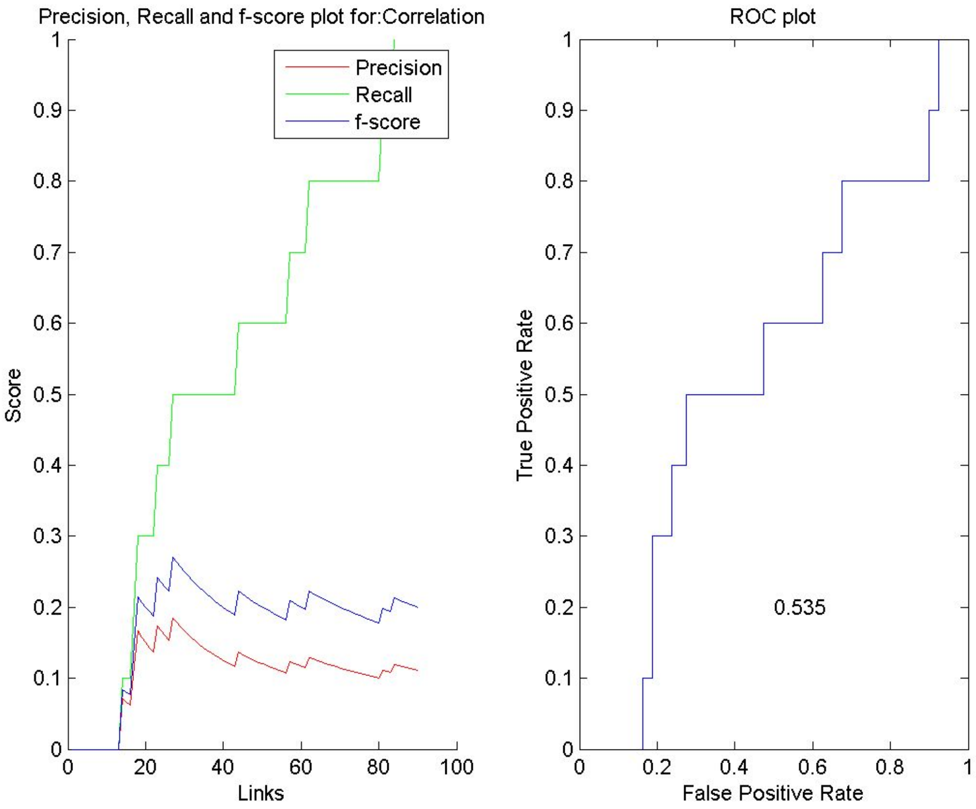
Type : Scale free

Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 5



Results - Correlation

Size : 20

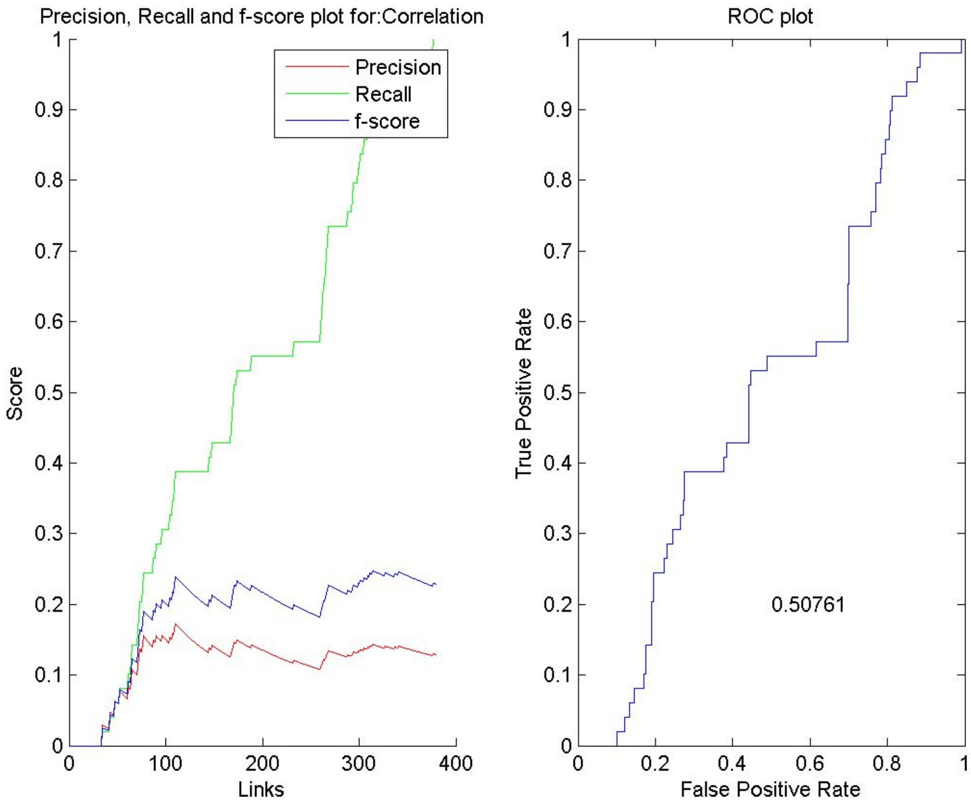
Type : Scale free

Average Degree : 4

Linearity : Non linear

Data points : 500

Bins : 2



Results - Correlation

Size : 20

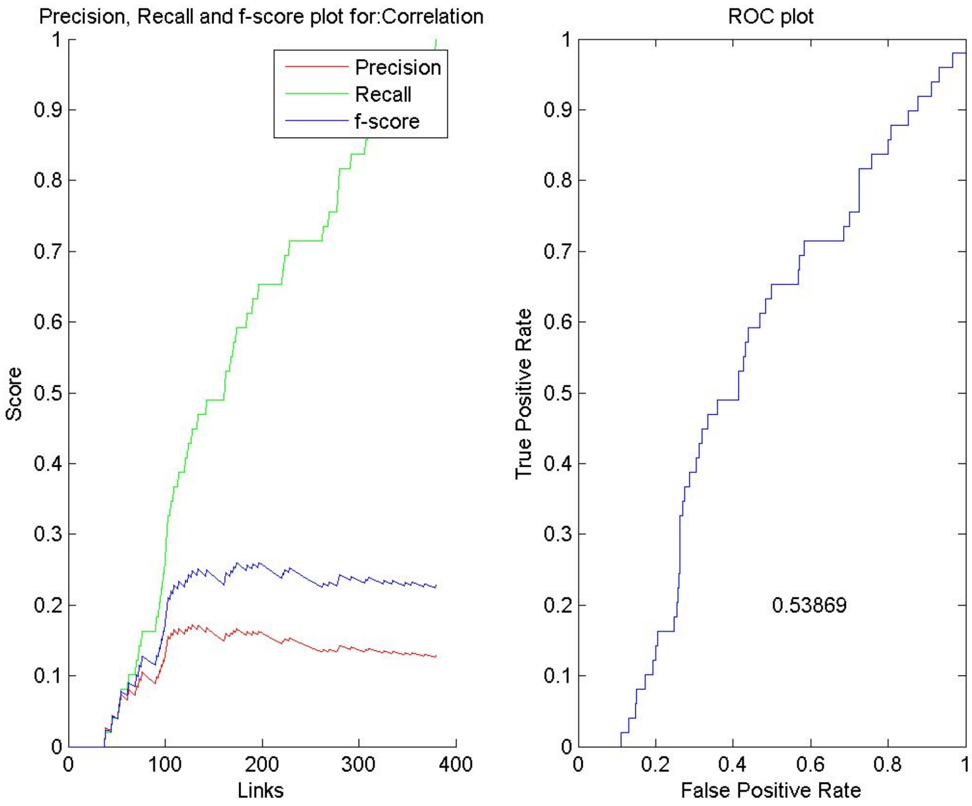
Type : Scale free

Average Degree : 4

Linearity : Non linear

Data points : 500

Bins : 5



Explanation - Correlation

Correlation --/--> Causality

- ➢ Strong correlation between independent genes with common ancestors, especially if they are siblings
- ➢ Scaling of correlation to [0, 1] as correlation coefficient reduces difference

Results - Pairwise Granger Causality

Size : 10

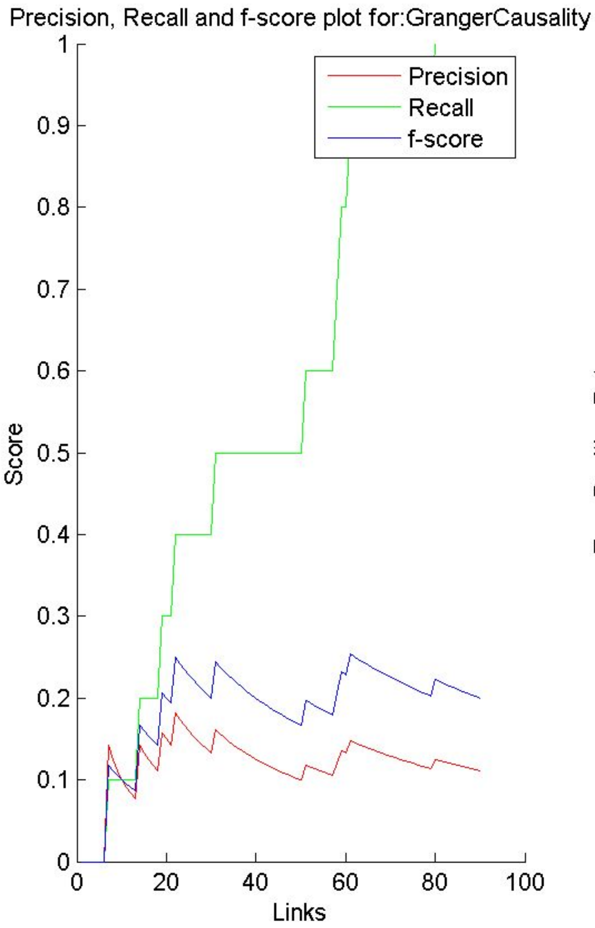
Type : Scale free

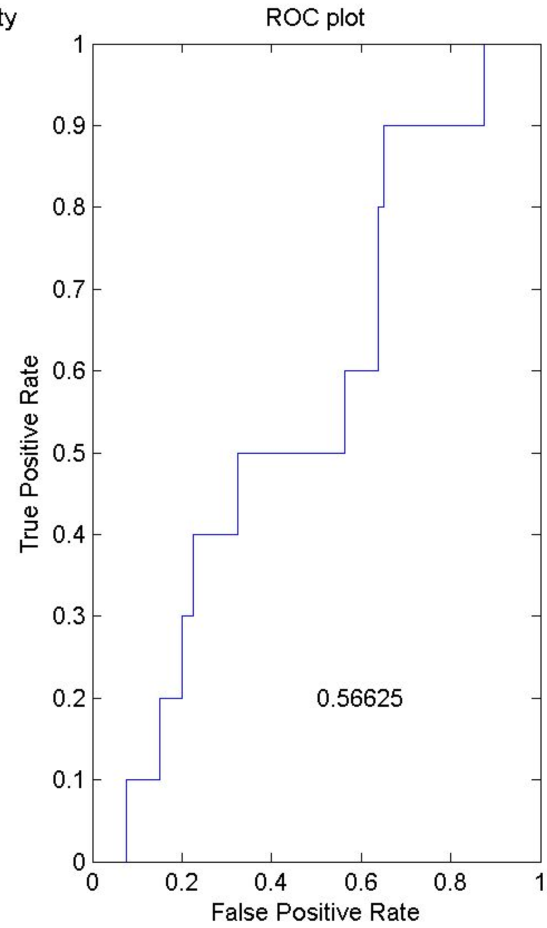
Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 2





Results - Pairwise Granger Causality

Size : 10

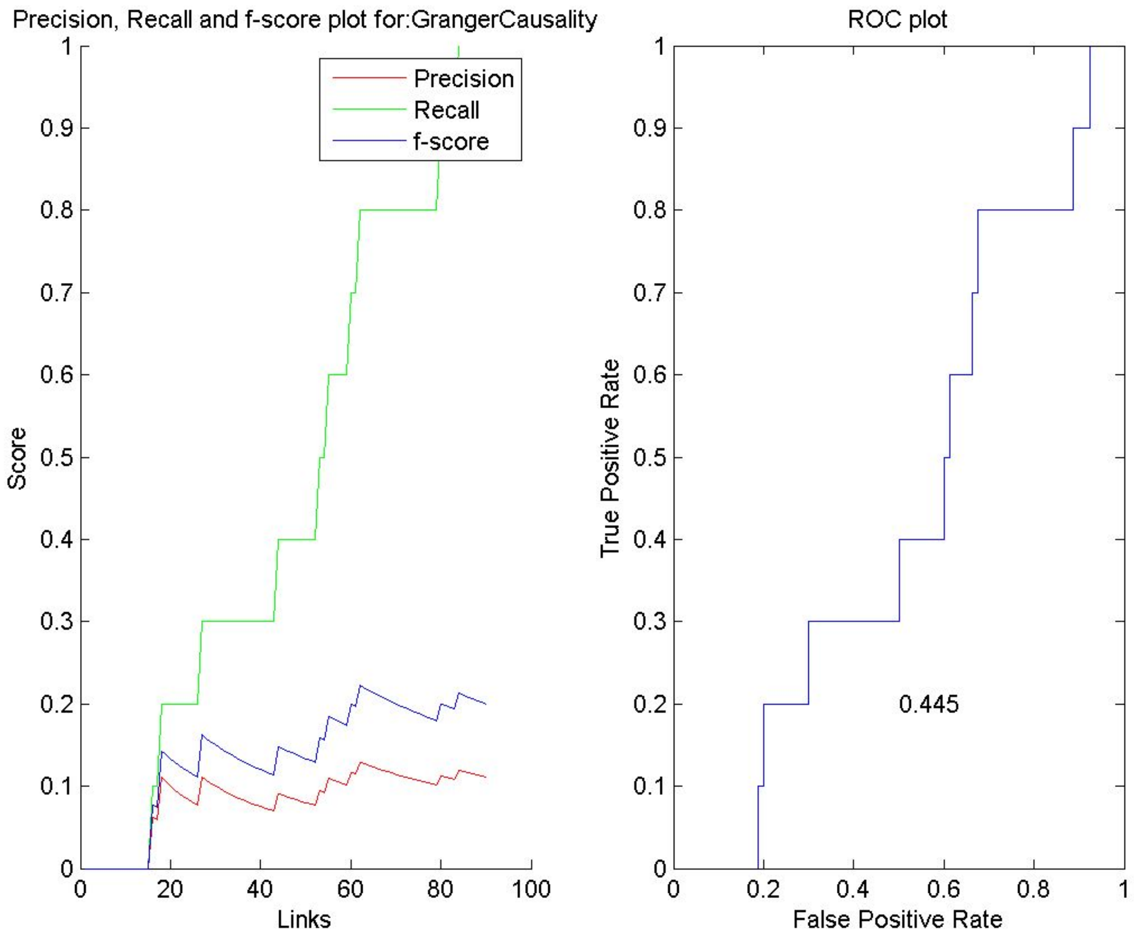
Type : Scale free

Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 5



Results - Pairwise Granger Causality

Size : 20

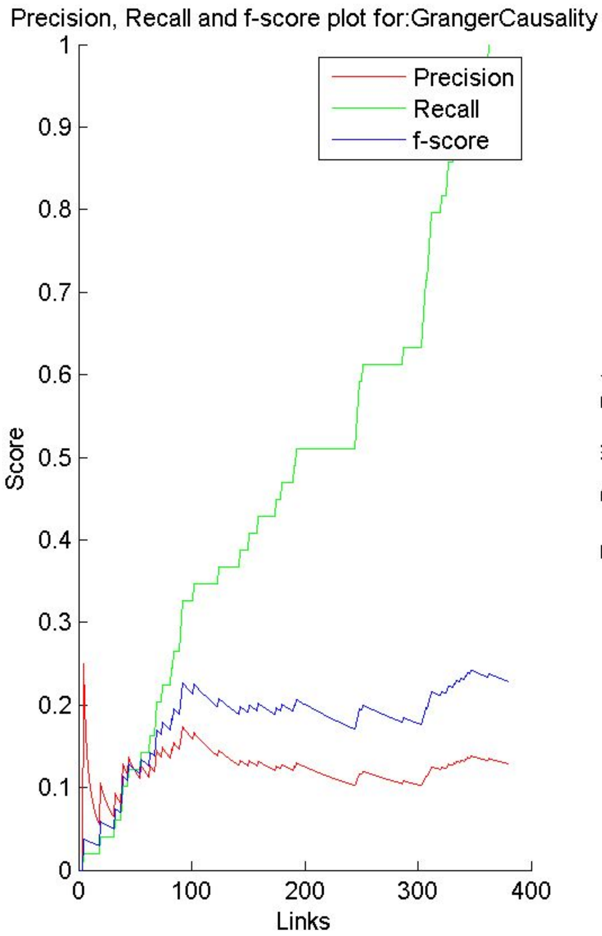
Type : Scale free

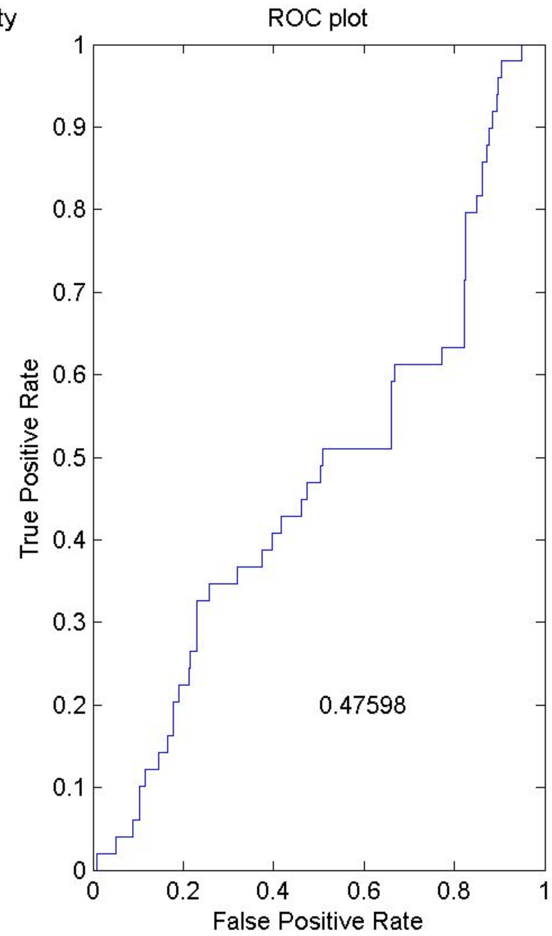
Average Degree : 4

Linearity : Non linear

Data points : 500

Bins : 2





Results - Pairwise Granger Causality

Size : 20

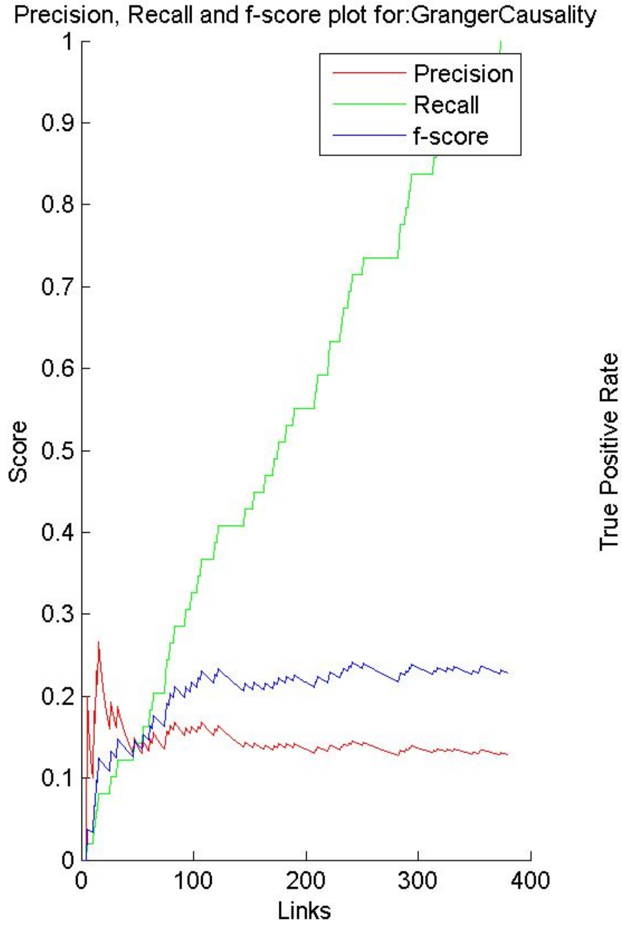
Type : Scale free

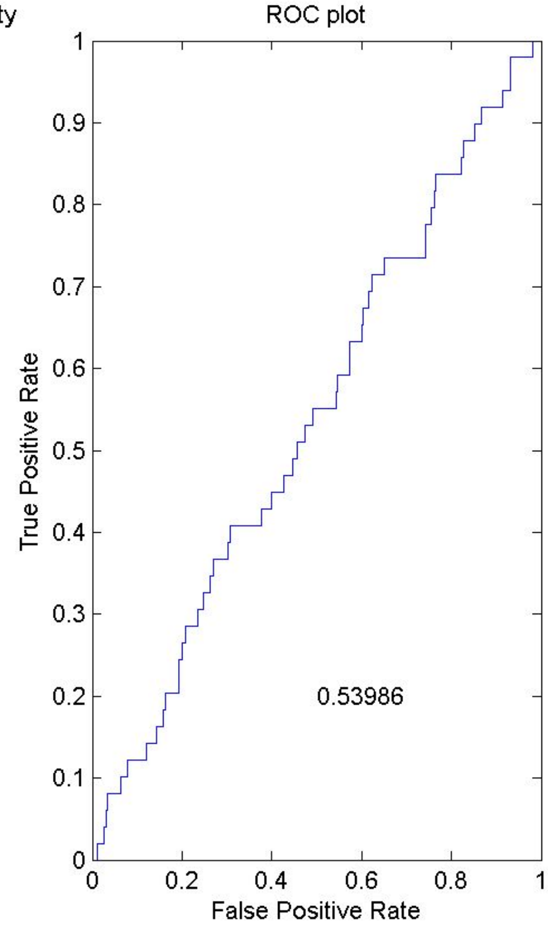
Average Degree : 4

Linearity : Non linear

Data points : 500

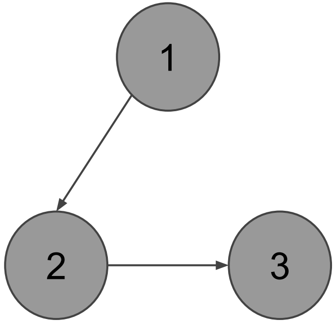
Bins : 5

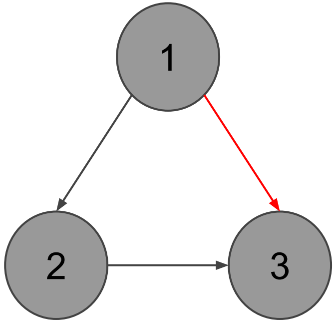


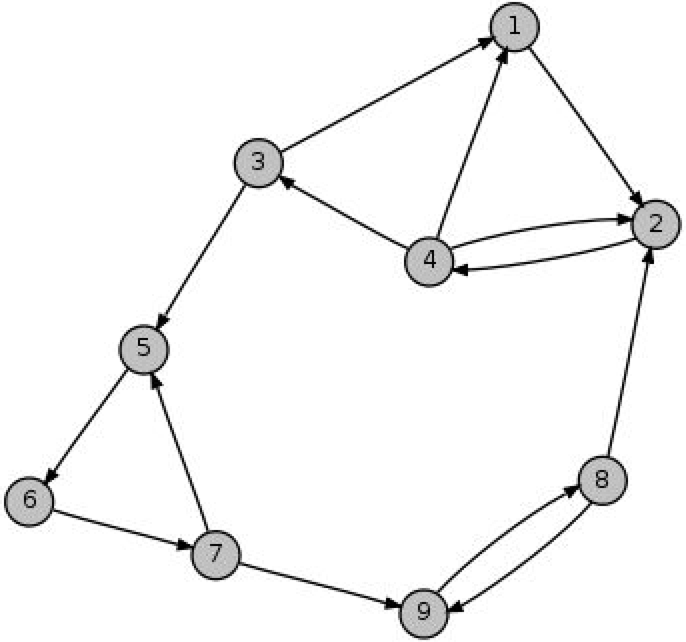


Explanation - Granger Causality (Pairwise)

- ➢ PGC cannot differentiate between direct and indirect causalities and thus, though recall is high, precision is low
- ➢ Non linear data







Results - Mutual Information

Size : 10

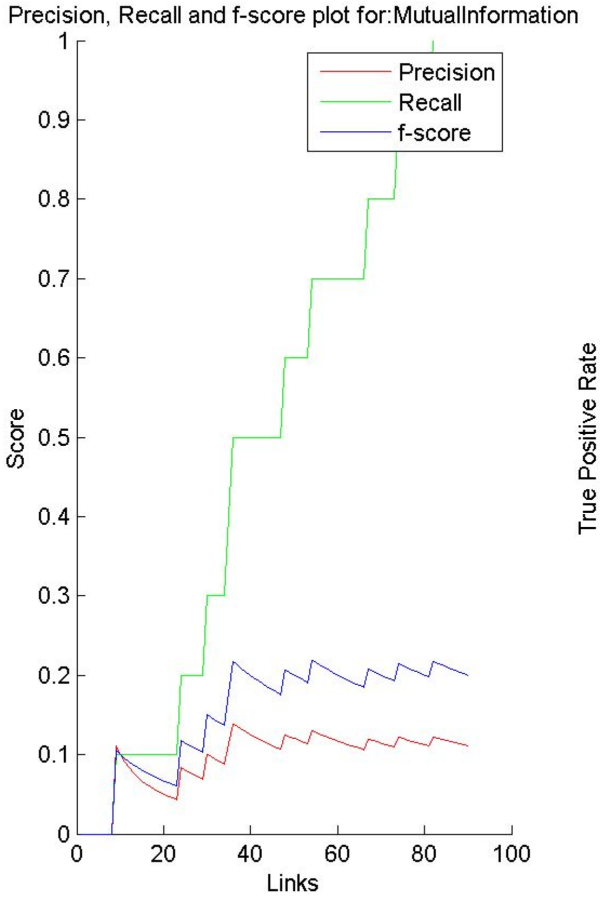
Type : Scale free

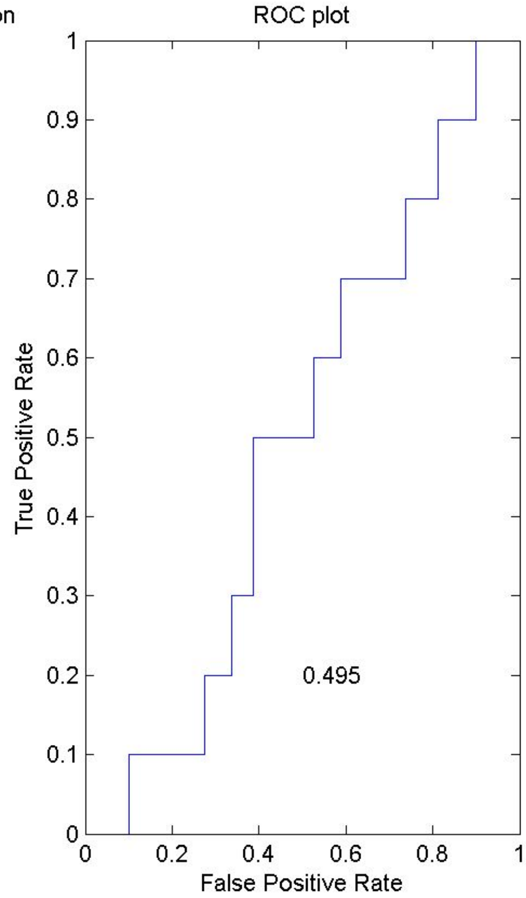
Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 2





Results - Mutual Information

Size : 10

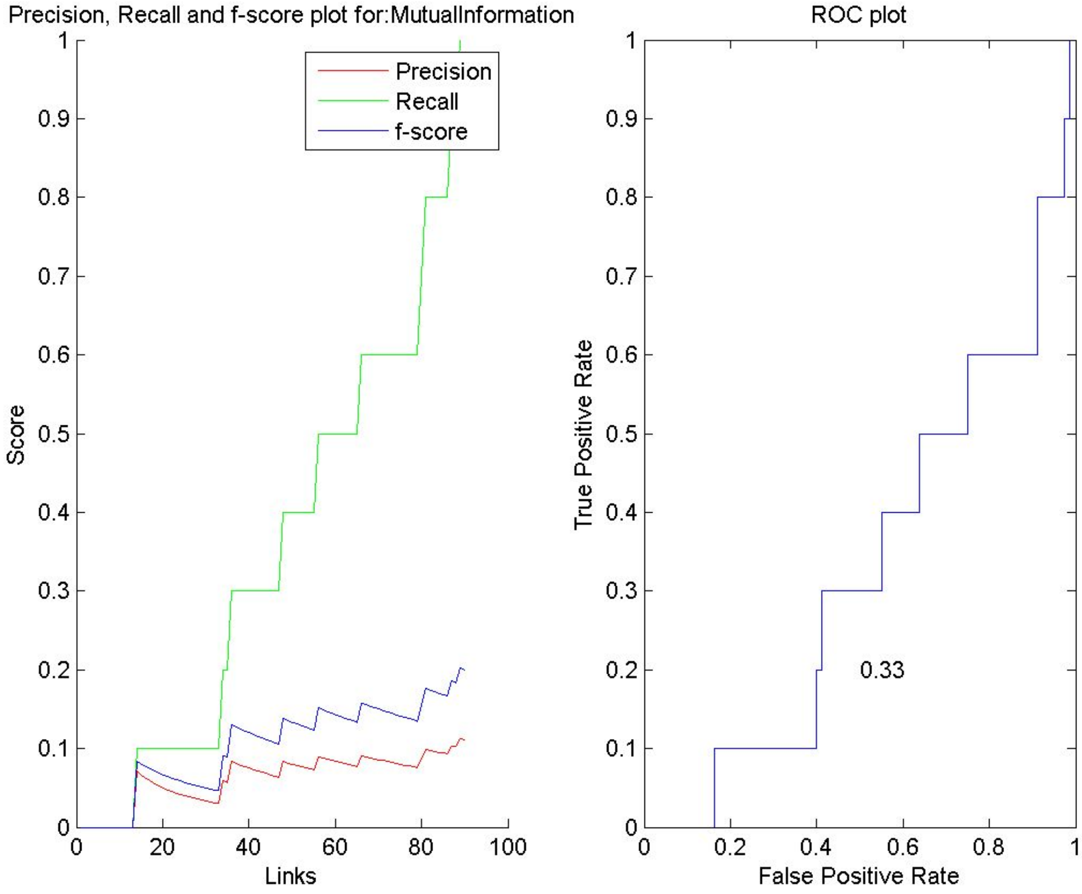
Type : Scale free

Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 5



Results - Mutual Information

Size : 20

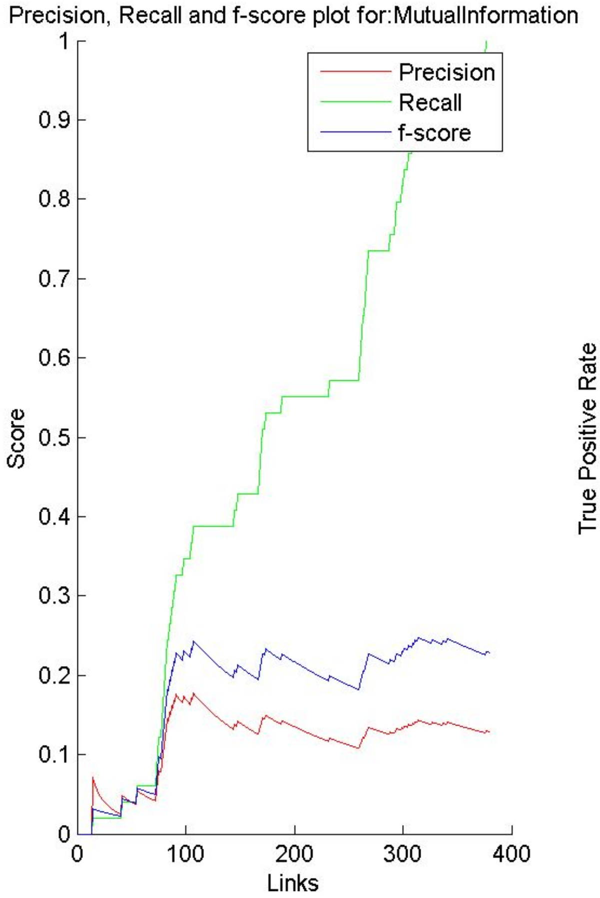
Type : Scale free

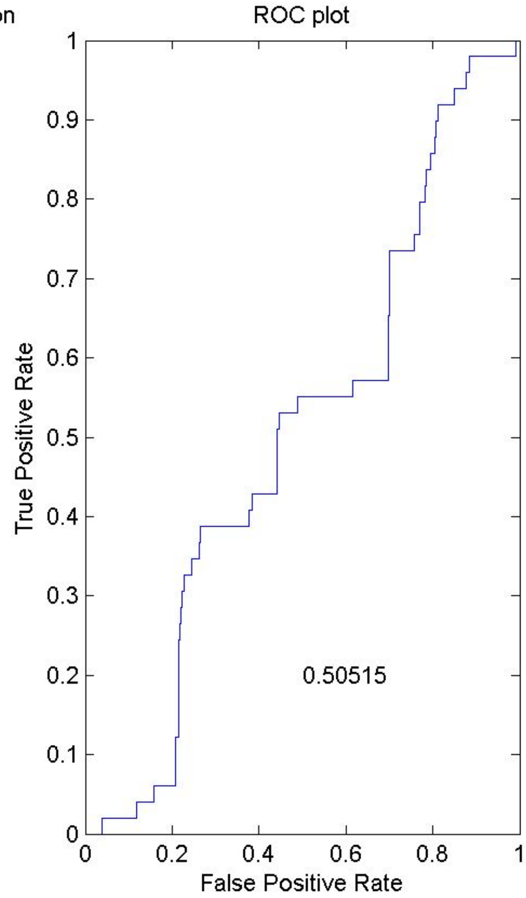
Average Degree : 4

Linearity : Non linear

Data points : 500

Bins : 2





Results - Mutual Information

Size : 20

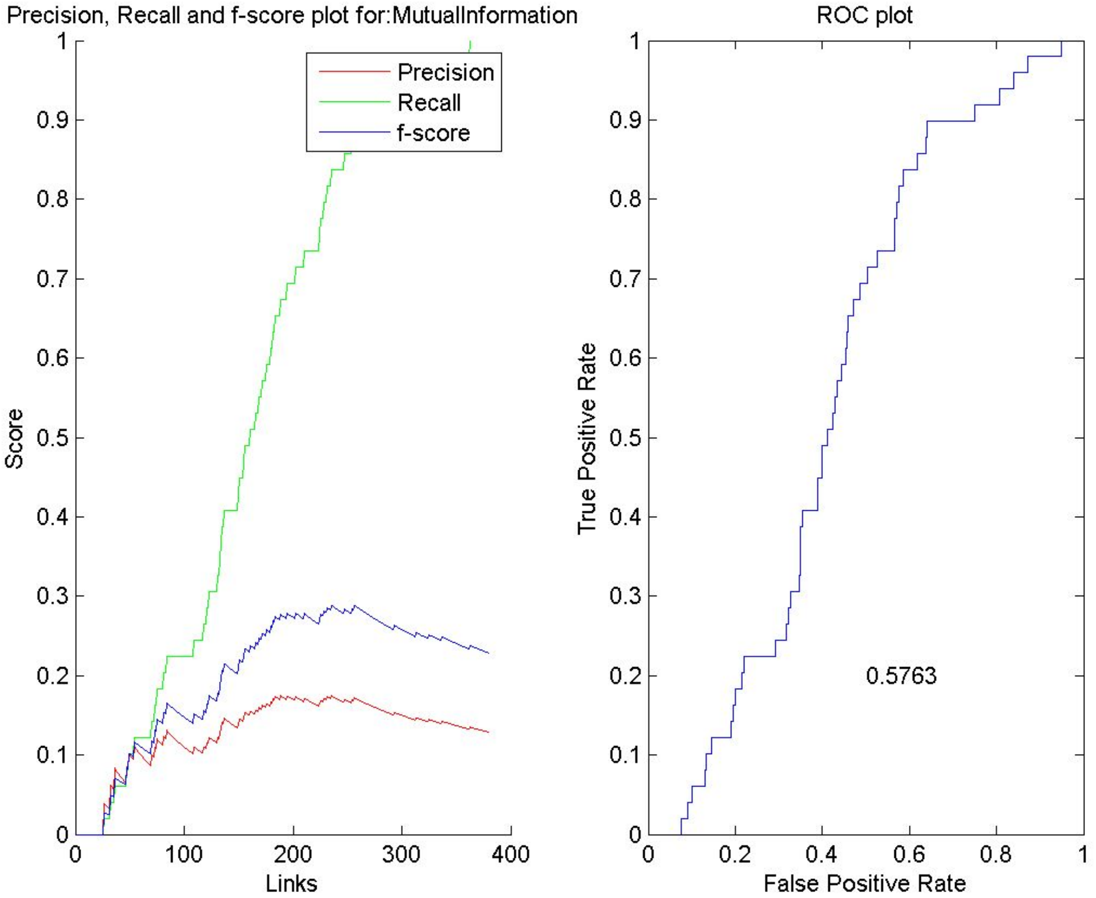
Type : Scale free

Average Degree : 4

Linearity : Non linear

Data points : 500

Bins : 5



Explanation - Mutual Information

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

- MI is not effective at predicting future events from current data. It is symmetric.
- It does not indicate the direction of the flow of information (unless we use the lag direction).

Results - Transfer Entropy

Size : 10

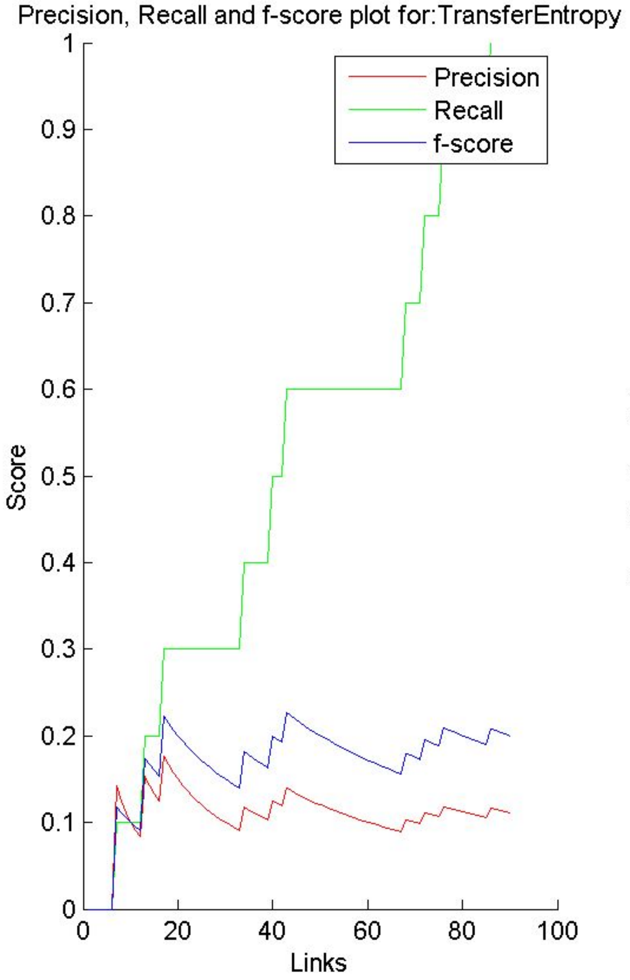
Type : Scale free

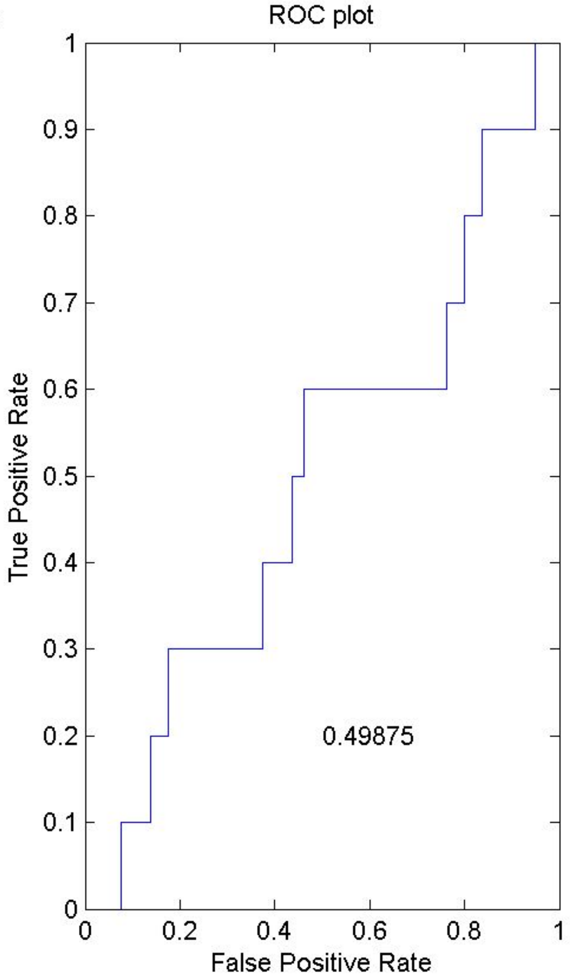
Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 2





Results - Transfer Entropy

Size : 10

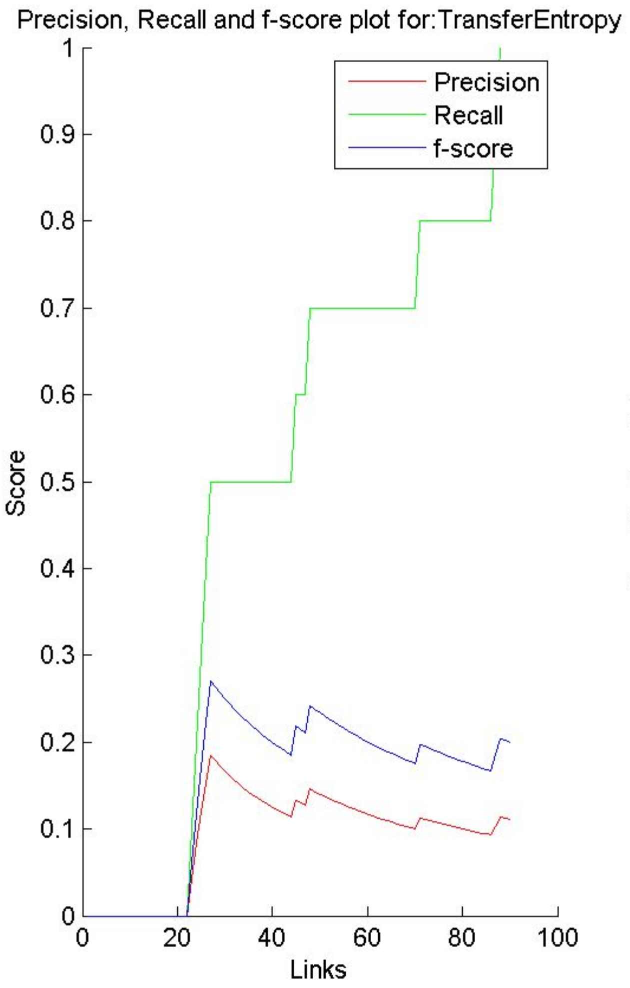
Type : Scale free

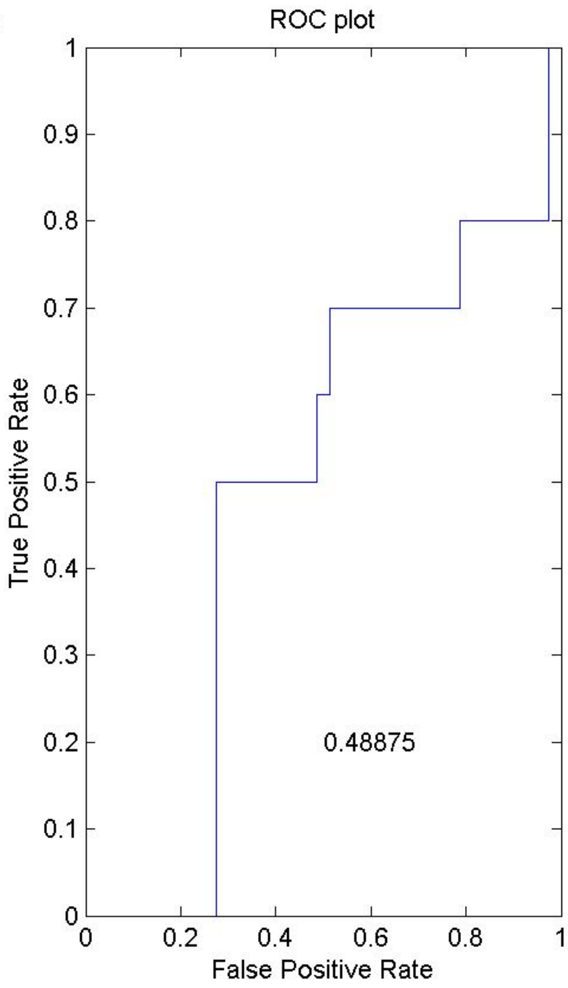
Average Degree : 3

Linearity : Non linear

Data points : 500

Bins : 5





Results - Transfer Entropy

Size : 20

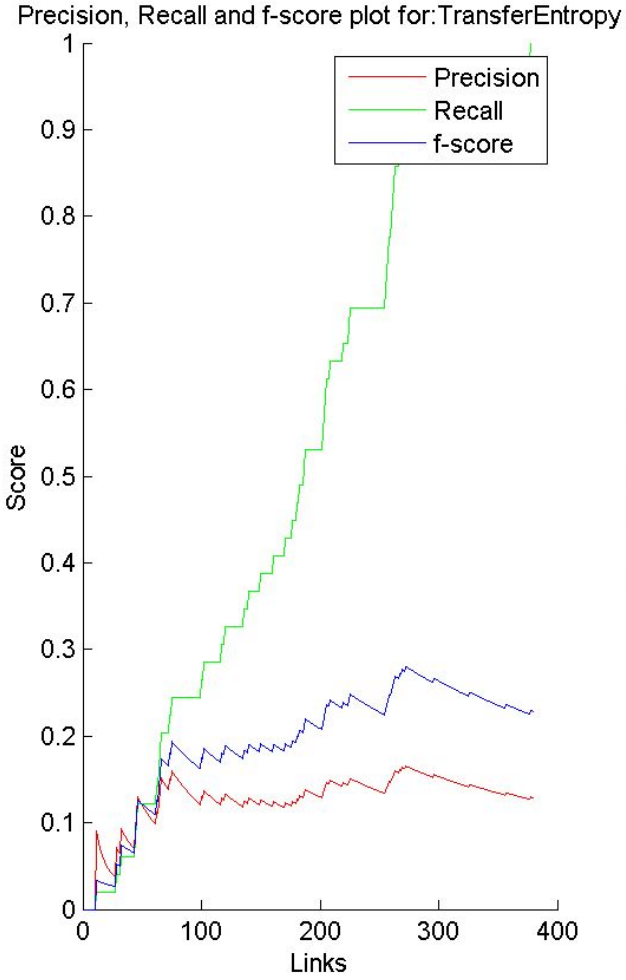
Type : Scale free

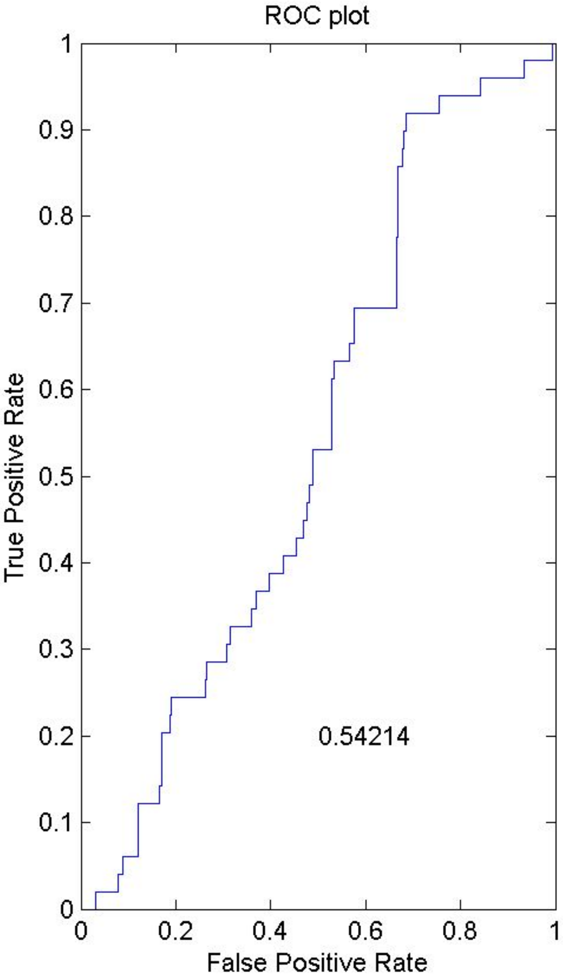
Average Degree : 4

Linearity : Non linear

Data points : 500

Bins : 2





Results - Transfer Entropy

Size : 20

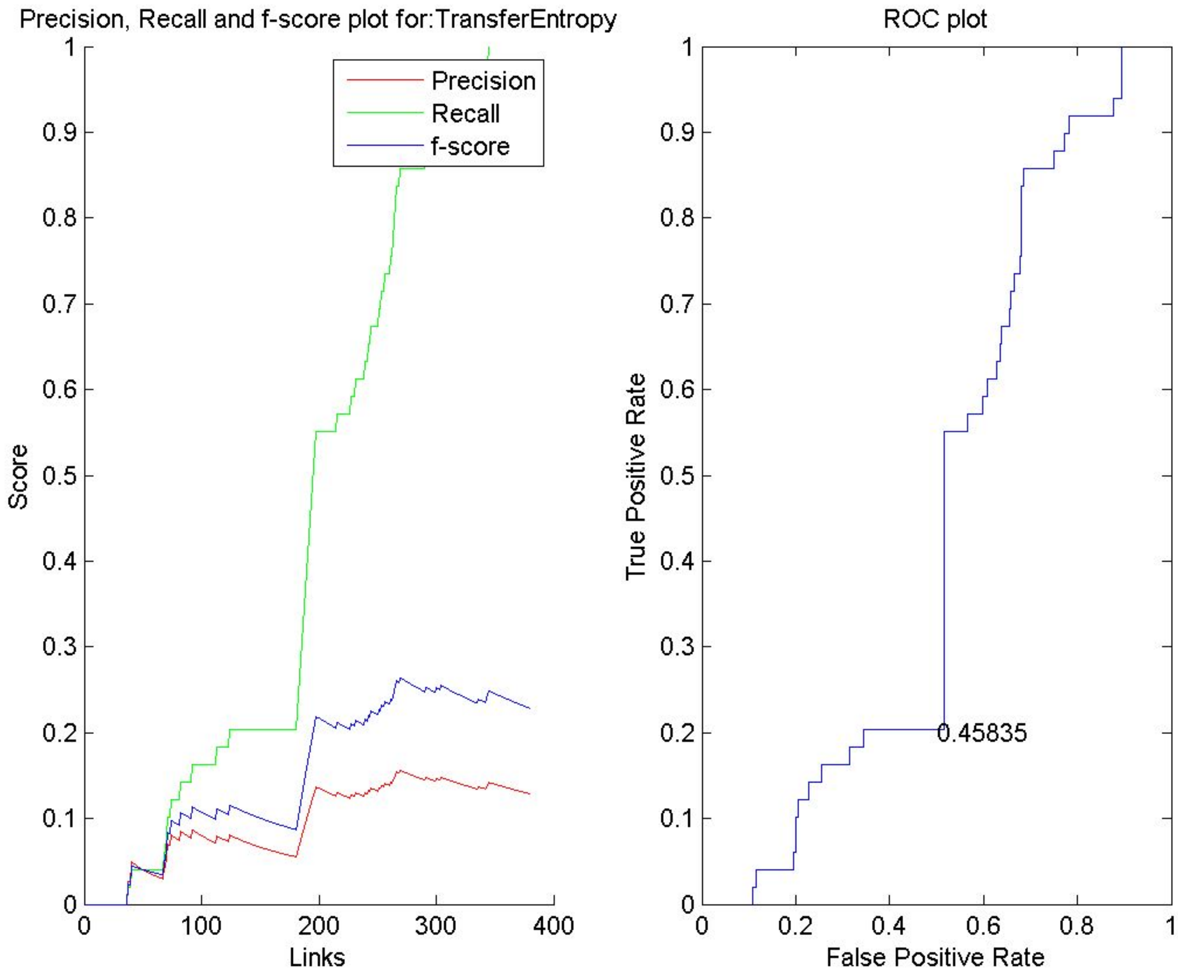
Type : Scale free

Average Degree : 4

Linearity : Non linear

Data points : 500

Bins : 5



Parameters - Additive Smoothing

$$T_{J \to I} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) \log \left(\frac{p(x_{n+1}, x_n, y_n) \cdot p(x_n)}{p(x_n, y_n) \cdot p(x_{n+1}, x_n)} \right)$$

- Problem: signal is too short!

P(X=x) = (1+favourable(x))/(size(X)+total)

Results - Mutual Information

Size : 10

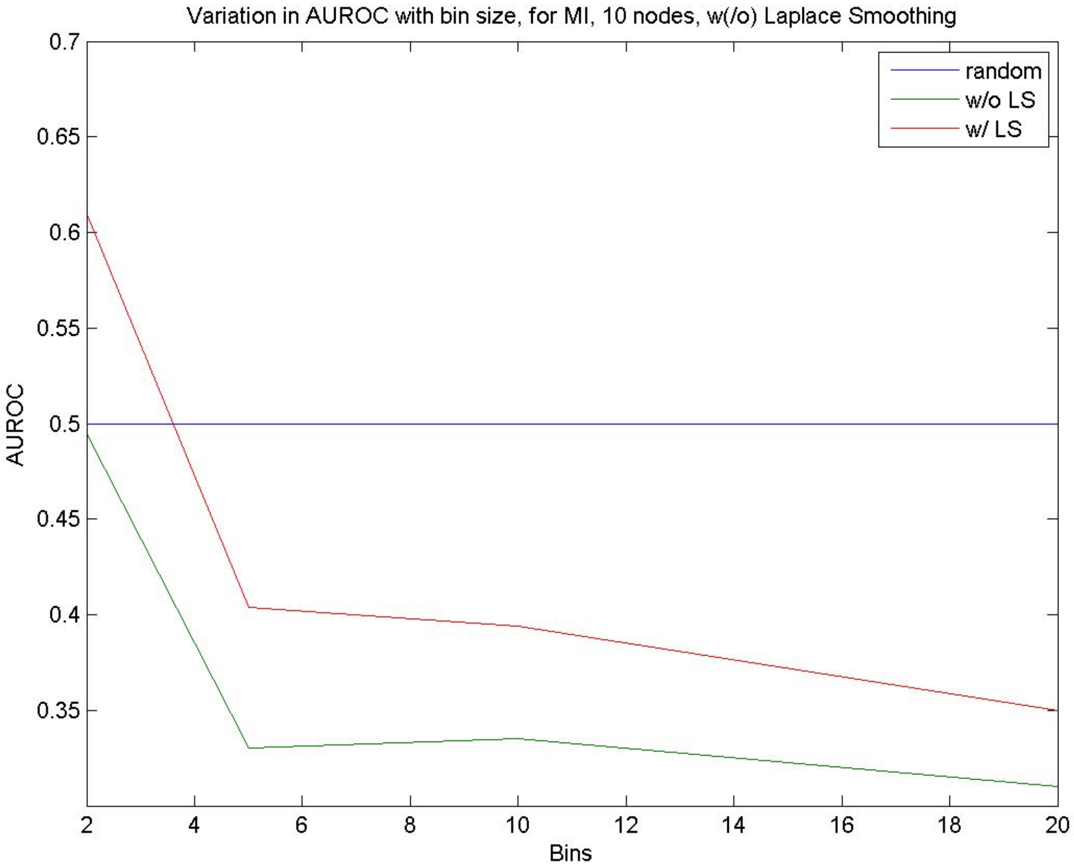
Type : Scale free

Average Degree : 3

Linearity : Non linear

Data points : 500

Smoothing : Additive



Results - Mutual Information

Size : 20

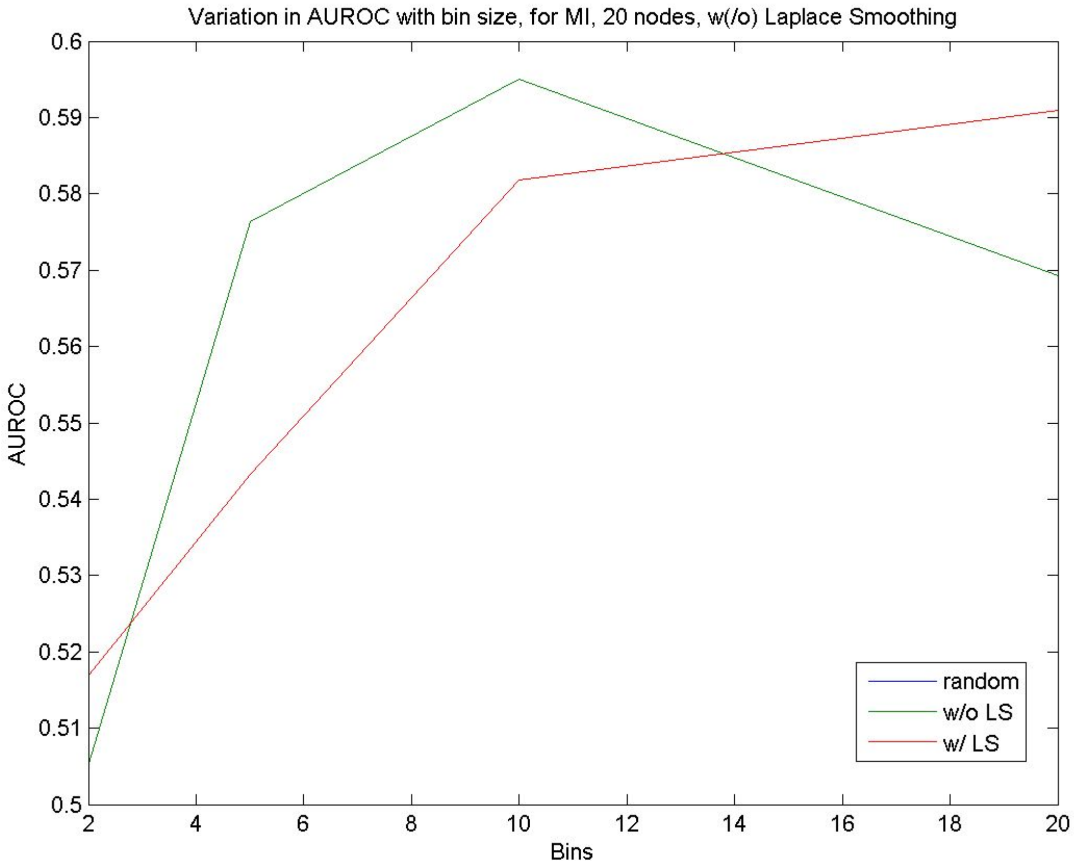
Type : Scale free

Average Degree : 4

Linearity : Non linear

Data points : 500

Smoothing : Additive



Results - Transfer Entropy

Size : 10

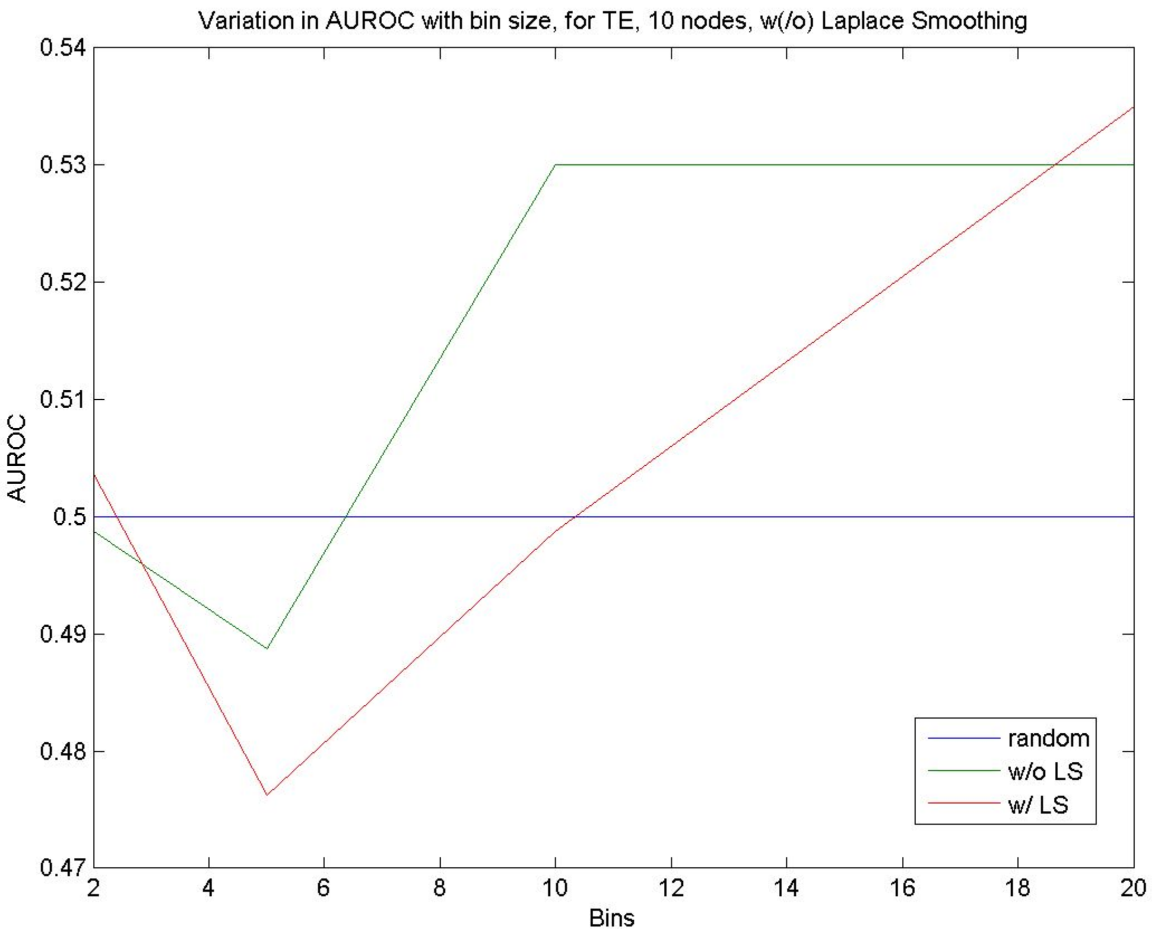
Type : Scale free

Average Degree : 3

Linearity : Non linear

Data points : 500

Smoothing : Additive



Results - Transfer Entropy

Size : 20

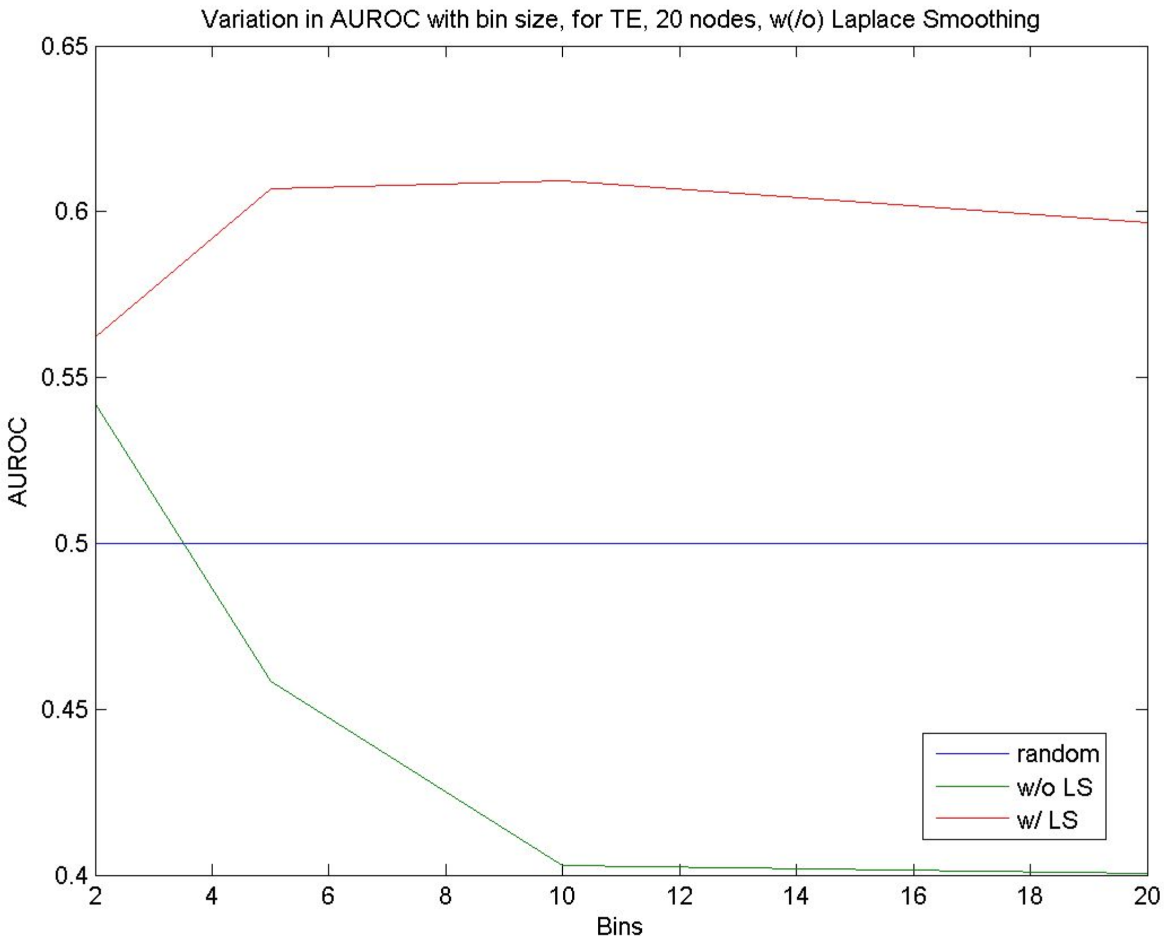
Type : Scale free

Average Degree : 4

Linearity : Non linear

Data points : 500

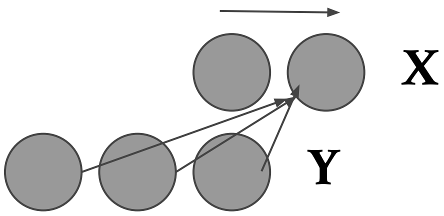
Smoothing : Additive



Explanations and Questions - Transfer Entropy

$$T_{X \rightarrow Y} = H(Y_t | Y_{t-1:t-L}) - H(Y_t | Y_{t-1:t-L}, X_{t-1:t-L})$$

- TE is the more generalised case for Granger Causality
- Does not assume linearity of the system being studied
- Additive Smoothing is a must, because of the large domain of the probability distributions involved
- Model Order can be increased?
- Optimal level of quantisation?



Summary of our Experience

- Poverty of data → Smoothing
- Abundance of parameters → Grid Search
- No strict trends in any direction, but largely:

$$\text{TE} \sim \text{MI} > \text{GC} > \text{Correlation}$$

- These methods, standalone, are not a good measure for the discovery of GRNs with high confidence
- Combination of multiple methods (akin to FP Correction, Structure Learning) can enhance the performance

Future Work

- Correction for False Positives: interplay of GC, MI, TE
  - Feature Weighting
- Treat signals being causally related if they belong to same dynamical system

As Random Variable

Correlation	Mutual Information
Granger Causality	Transfer Entropy

As Dynamical System

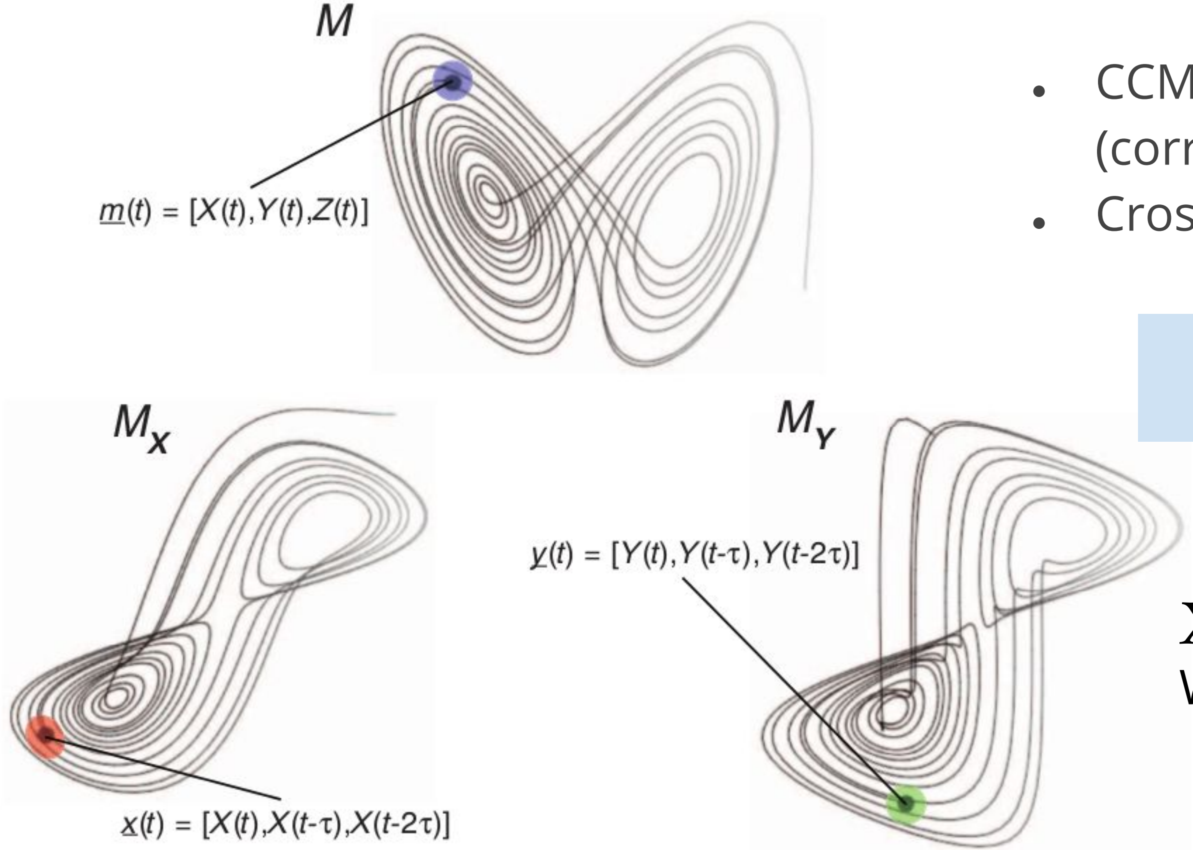
Convergent Cross Map

- Incorporate Network Substructures
- Use real Biological Datasets to validate so-formed technique

Thank You

Questions?

The Way Ahead - Convergent Cross Mapping



- CCM involves convergence (correlation =/= causation)
- Cross-map in:

One Direction

Both Directions

$X \rightarrow Y$

We can predict X from Y

The Hybrid Approach

Information Theory

- ✗ For stochastic, non-linear systems
- ✓ Can differentiate, through thresholding

Theory of Manifolds

- ✓ For possibly synergistic, deterministic systems
- ✗ No difference between first order and transitive causality

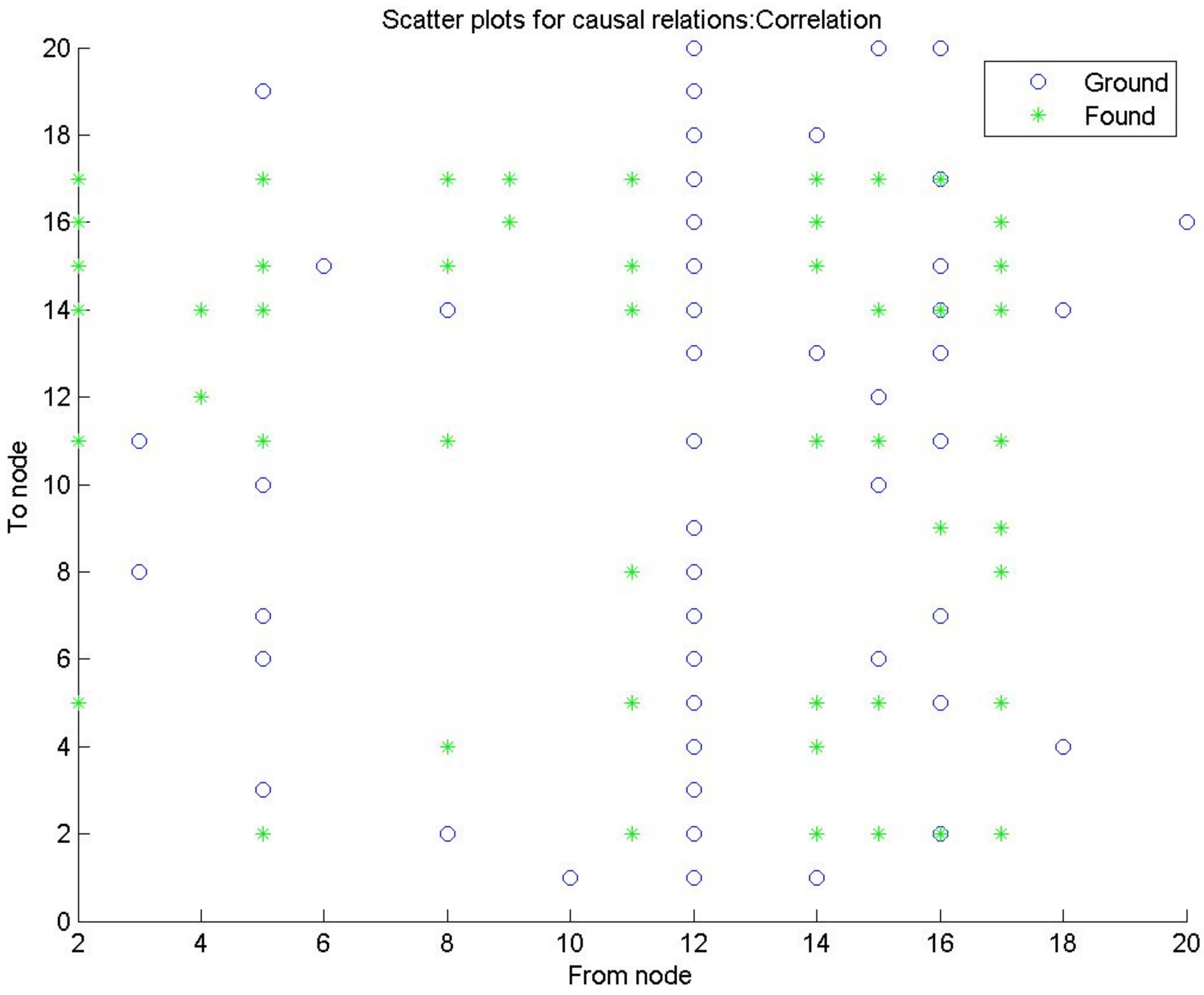
Scatter

Size : 20

Data Points : 500

Bins : 2

Method : Correlation



Scatter

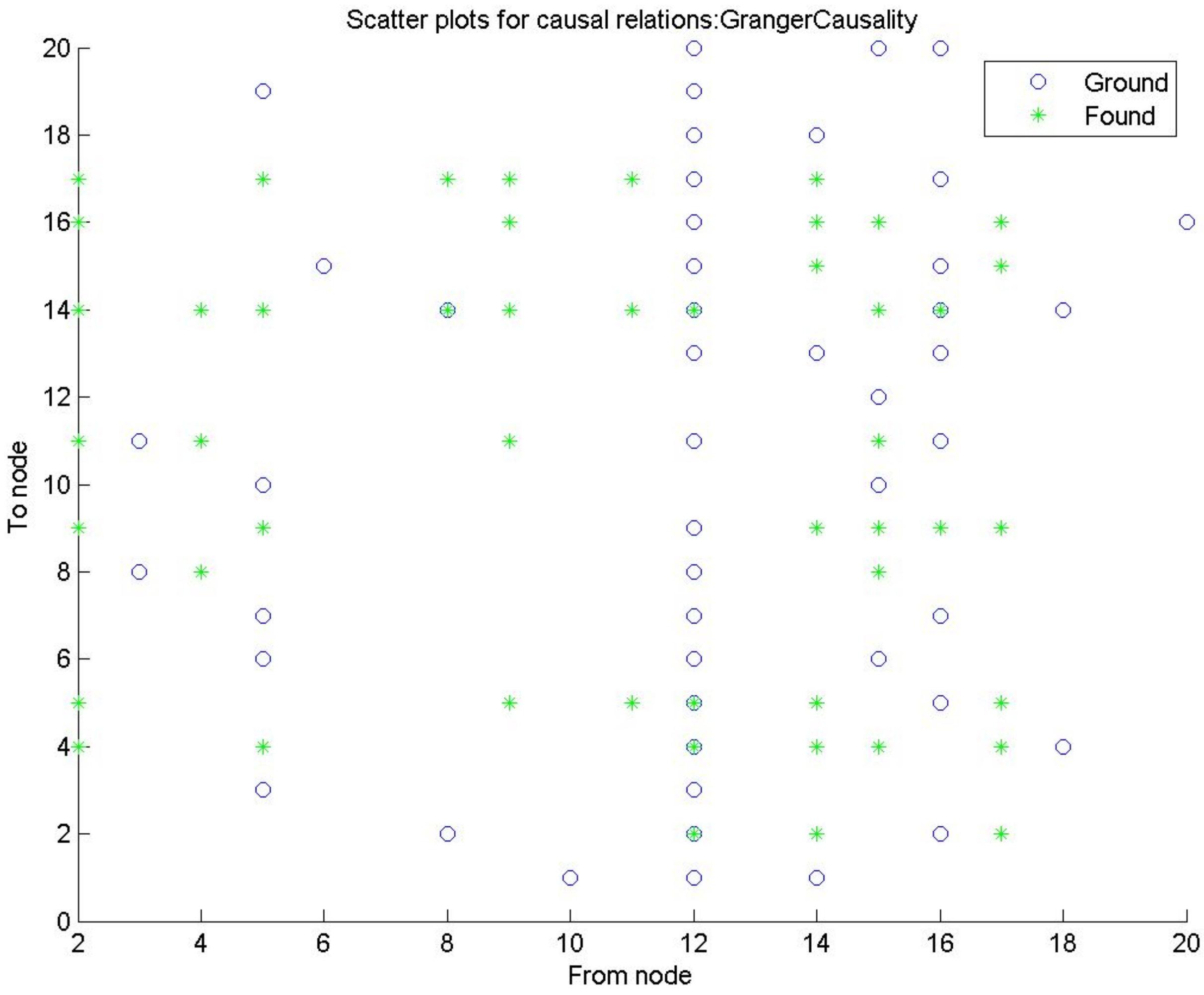
Size : 20

Data Points : 500

Bins : 2

Method :

Granger Causality



Scatter

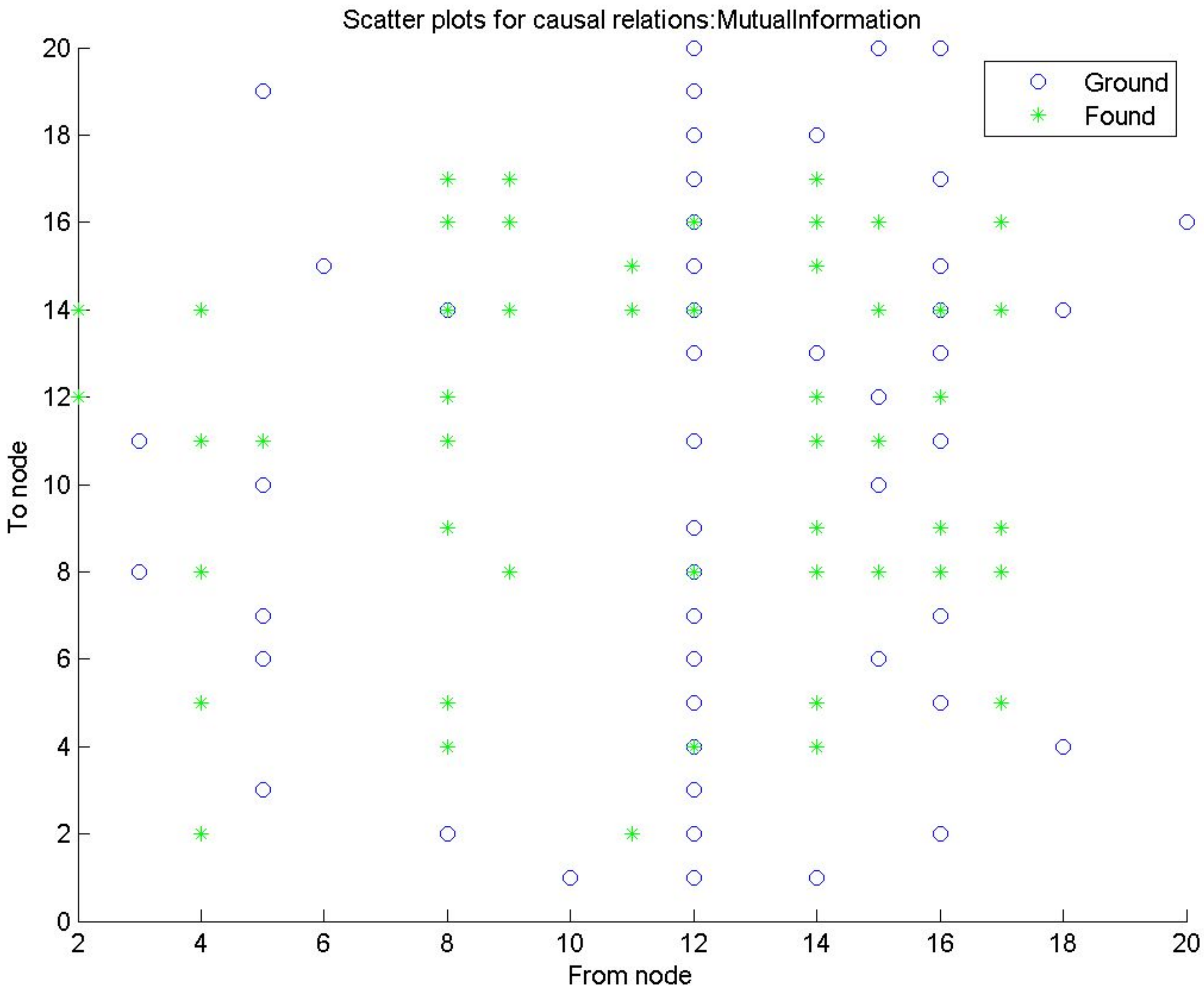
Size : 20

Data Points : 500

Bins : 2

Method :

Mutual Information



Scatter

Size : 20

Data Points : 500

Bins : 2

Method :

Transfer Entropy