Case Western Reserve University
Great Lakes Energy Institute

An Information Theoretic Framework and Self-organizing Agent-based Sensor Network Architecture for Power Plant Condition Monitoring

Kenneth A. Loparo, Ph.D.
Richard M. Kolacinski, Ph.D.

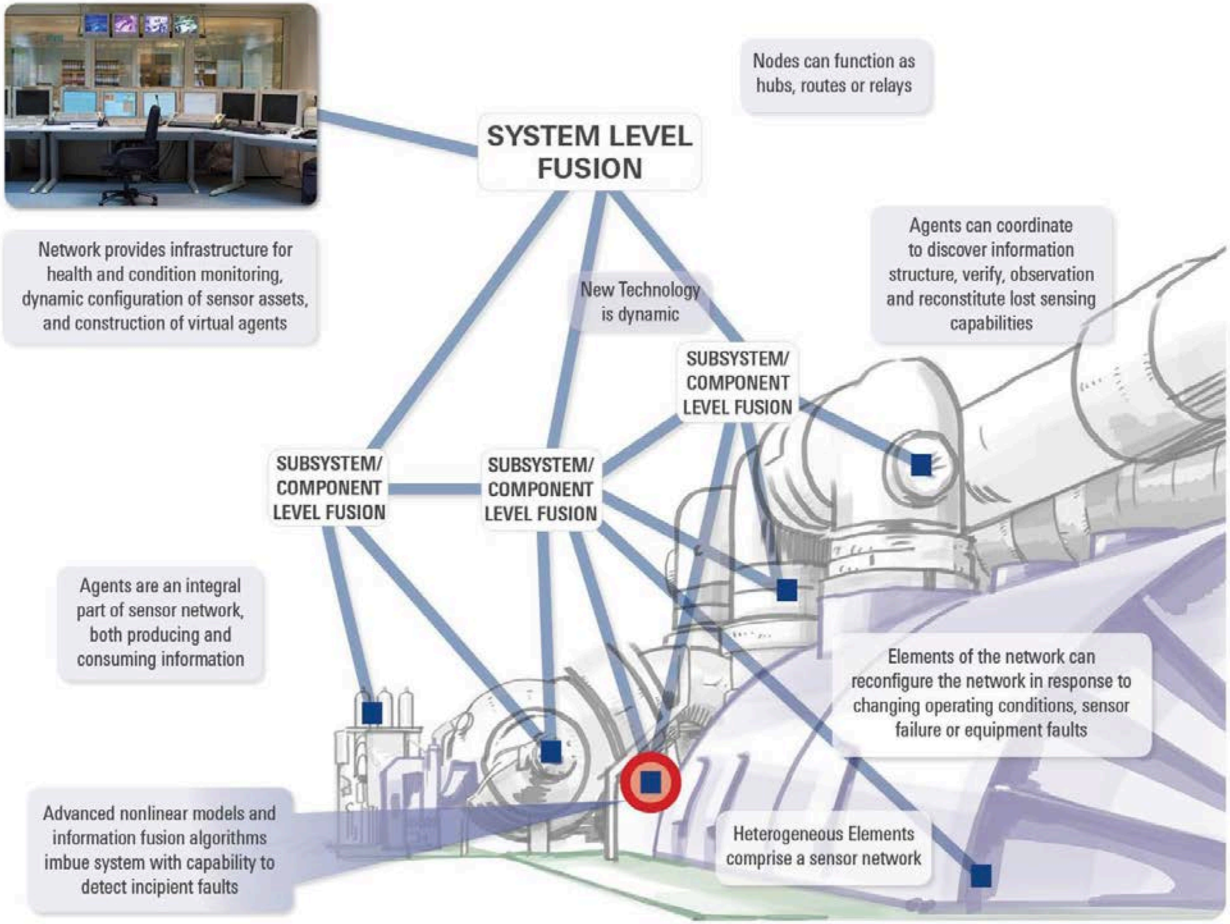
2016 Crosscutting Research Review
April 20, 2016

Agenda

- Introduction
  - Distributed Health and Condition Monitoring
- Information Measures
  - Entropy/Information
  - General information measures
- Information Structure of Systems
  - Properties of information
  - System decomposition
  - Computation of information measures
  - Detecting changes in system structure

Introduction

Production Systems



4

Information Measures

Information Theory

- **Information is the amount of surprise contained in the data;**
  - Data that tells you what you already know is not informative,
  - Not all data is created equal.
- **The fundamental measure of information is *Shannon entropy* is**

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_d p(x),$$

where $X \in \mathcal{X}$ is a discrete R.V., $\mathcal{X}$ is a finite set known as the alphabet, and $p(x) = \Pr\{X = x\}$.

CASE WESTERN RESERVE UNIVERSITY | GREAT LAKES ENERGY INSTITUTE

The Multivariate Case

- **For a pair of discrete R.V.'s (X,Y) with joint and conditional distributions $p(x,y)$ and $p(x|y)$, the joint and conditional entropies are, respectively:**

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2 p(x,y)$$

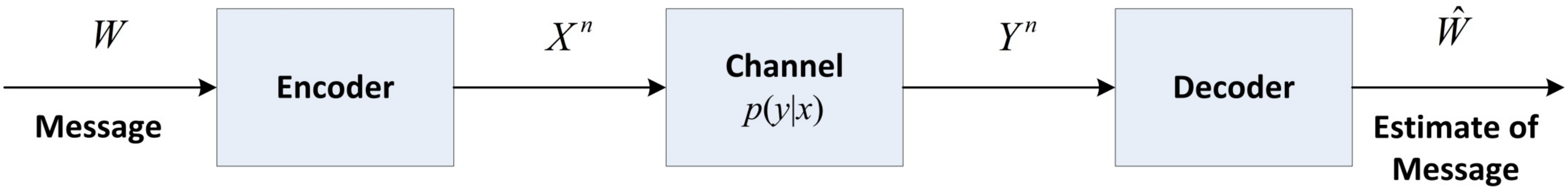
$$H(X|Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2 p(x|y)$$

- **The relationship between these R.V.'s is captured by *Mutual Information*:**

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)}$$

- ***Mutual Information* and *Shannon Entropy* are related by:** $I(X;Y) = H(X) - H(X|Y)$

Information Channels



- **Let $\mathcal{X}$ and $\mathcal{Y}$ be the *input alphabet* and *output alphabet*, respectively, and let S be the set of channel states. An information channel is a system of probability functions:**

$$p_n(\beta_1, \ldots, \beta_n | \alpha_1, \ldots, \alpha_n : s)$$

where $\alpha_1, \ldots, \alpha_n \in \mathcal{X}$, $\beta_1, \ldots, \beta_n \in \mathcal{Y}$, and $s \in S$ for $n = 1, 2, \ldots$.

- **Mutual information between the input and output provides a measure of *channel transmittance*:**

$$T(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y})$$

- **The maximum over all distributions is known as the *channel capacity*.**

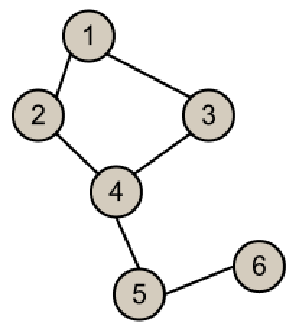
Information Structure

Information Structure

- **The communications topology determined by available observation processes;**
  - fusing information from multiple sensors,
  - Reconstituting lost or degraded sensing,
  - Detect system changes reflected in changing communication topology.
- **Identify "correlative" structure of sensor data;**
  - Provides means of identifying relevant (possibly abstract) subsystems,
  - Basis for mesoscopic models and "summary" variables.

System Structure

Undirected Graph & Adjacency Matrix



Undirected Graph

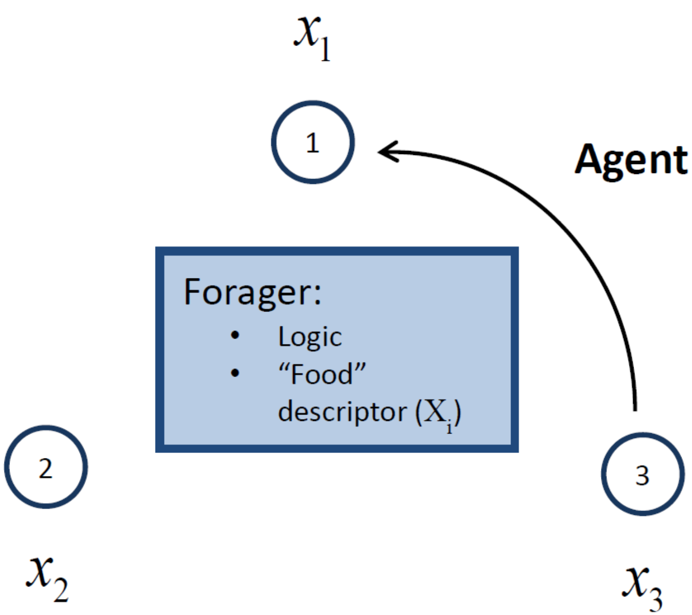
	1	2	3	4	5	6
1	0	1	1	0	0	0
2	1	0	0	1	0	0
3	1	0	0	1	0	0
4	0	1	1	0	1	0
5	0	0	0	1	0	1
6	0	0	0	0	1	0

Adjacency Matrix

$$M_I = \begin{bmatrix} \hat{I}(Z_1;Z_1) & \hat{I}(Z_1;Z_1) & \dots & \hat{I}(Z_1;Z_n) \\ \hat{I}(Z_2;Z_1) & \hat{I}(Z_1;Z_1) & & \hat{I}(Z_1;Z_1) \\ \vdots & & \ddots & \vdots \\ \hat{I}(Z_n;Z_1) & \hat{I}(Z_1;Z_1) & \dots & \hat{I}(Z_1;Z_n) \end{bmatrix}$$

Where Z_i represents i^{th} sensor in power plant

Agents and Observations



$\boldsymbol{x_i}$: a time series which is the observations at node ***i***

$\boldsymbol{w_i}$: a time series which is the partial observations of $\boldsymbol{x_i}$ at node ***i***

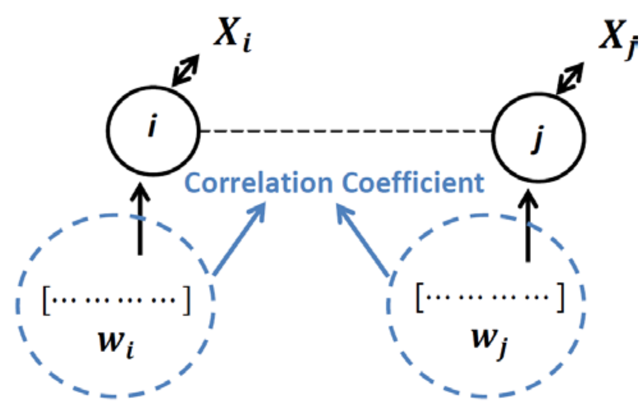
<u>Agent going from home node 3 to node 1:</u>

Behavior:

The agent carries the data, w_3 ,from its home node to the next node.

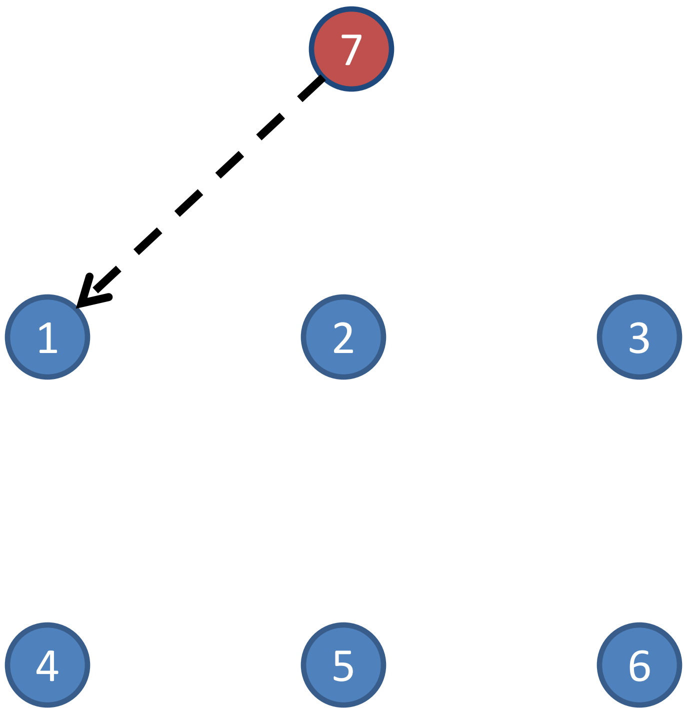
Food Definition:

The similarity, Correlation Coefficient, between the time series w_3 and w_1.



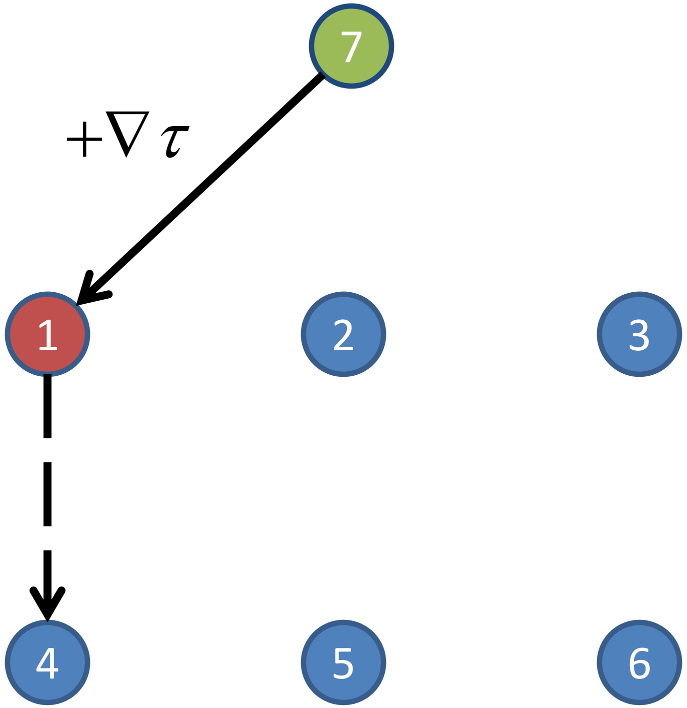
Initialization

Time Duration at this Iteration = [0:T]
Home Node = 7
Current Node = 7
Carrying Data = X_7 ([0:T])
Forward Flag = 1
Selected Next Node = 1



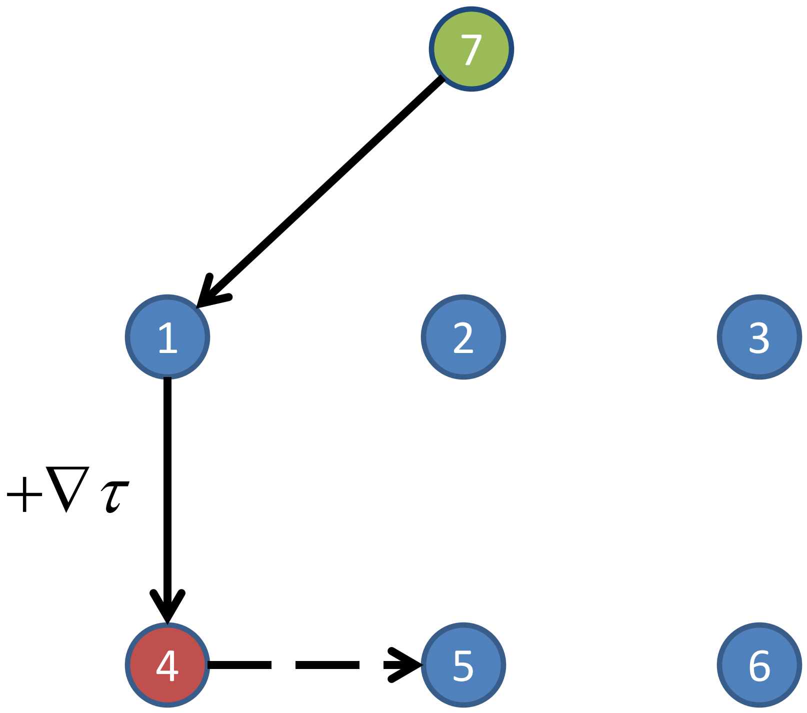
Iteration 1

Time Duration at this Iteration = [0:T]
Home Node = 7
Current Node = 1
Carrying Data = X_7 ([0:T])
Correlation Coefficient to Calculate is between: { X_7 ([0:T]) , X_1 ([0:T]) }
Correlation = Not High
Forward Flag = 1
Selected Next Node = 4



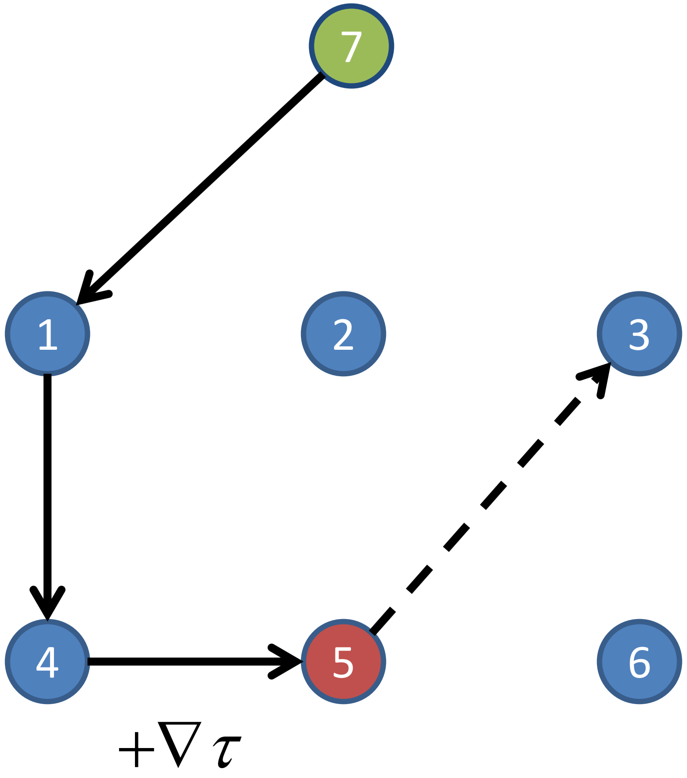
Iteration 3

Time Duration at this Iteration = [T:2T]
Home Node = 7
Current Node = 4
Carrying Data = X_7 ([0:T])
Correlation Coefficient to Calculate is between: { X_7 ([0:T]) , X_4 ([T:2T]) }
Correlation = Not High
Forward Flag = 1
Selected Next Node = 5



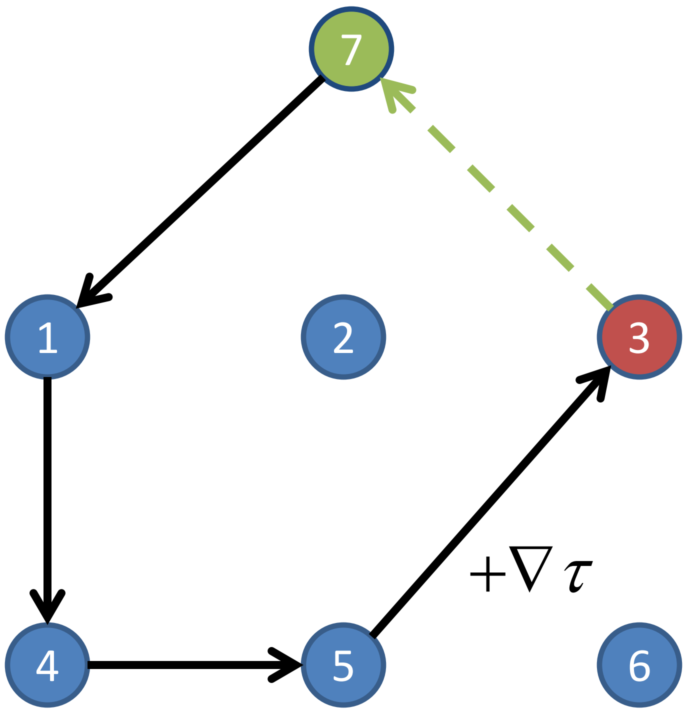
Iteration 4

Time Duration at this Iteration = [2T:3T]
Home Node = 7
Current Node = 5
Carrying Data = X_7 ([0:T])
Correlation Coefficient to Calculate is between: { X_7 ([0:T]) , X_5 ([2T:3T]) }
Correlation = Not High
Forward Flag = 1
Selected Next Node = 3



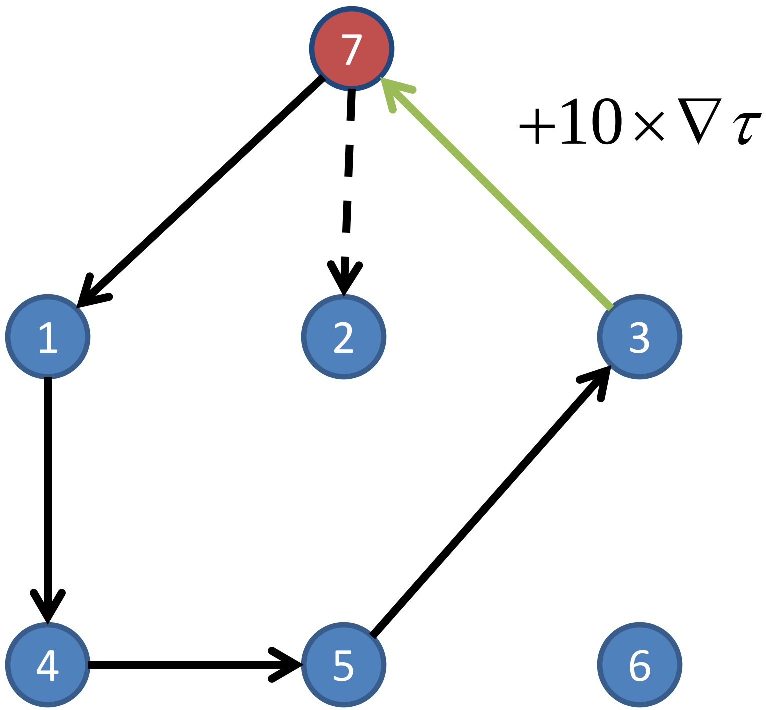
Iteration 5

Time Duration at this Iteration = [3T:4T]
Home Node = 7
Current Node = 3
Carrying Data = X_7 ([0:T])
Correlation Coefficient to Calculate is between: { X_7 ([0:T]) , X_3 ([3T:4T]) }
Correlation = High
Forward Flag = 0
Selected Next Node = 7 (Home Node)



Iteration 6

Time Duration at this Iteration = [4T:5T]
Home Node = 7
Current Node = 7
Carrying Data = X_7 ([0:T])
Forward Flag = 1
Selected Next Node = 2



Graph Similarity

- **Our graphs are weighted bidirectional graphs where $w_{ij} \neq w_{ji}$.**
- **In this case the Laplacian matrix is not symmetric and therefore its eigenvalues are not necessarily real positive numbers. This makes some problems in calculating the spectral distance with complex numbers.**
- **Use symmetrized Laplacian**

$$L(G) = D(G) - \left(A(G) + A(G)^T \right)$$

Spectral Distance Formula

$$d_k(G,H)=\begin{cases}\sqrt{\dfrac{\sum_{i=N-k}^{N-1}(\lambda_i-\mu_i)^2}{\sum_{i=N-k}^{N-1}\lambda_i^{\,2}}} & \text{if } \sum_{i=N-k}^{N-1}\lambda_i^{\,2}\le\sum_{i=N-k}^{N-1}\mu_i^{\,2}\\[2ex]\sqrt{\dfrac{\sum_{i=N-k}^{N-1}(\lambda_i-\mu_i)^2}{\sum_{i=N-k}^{N-1}\mu_i^{\,2}}} & \text{if } \sum_{i=N-k}^{N-1}\lambda_i^{\,2}>\sum_{i=N-k}^{N-1}\mu_i^{\,2}\end{cases}$$

– λ_i represents the eigenvalues of the Laplac matrix for graph G

– μ_i represents the eigenvalues of the Laplac matrix for graph H

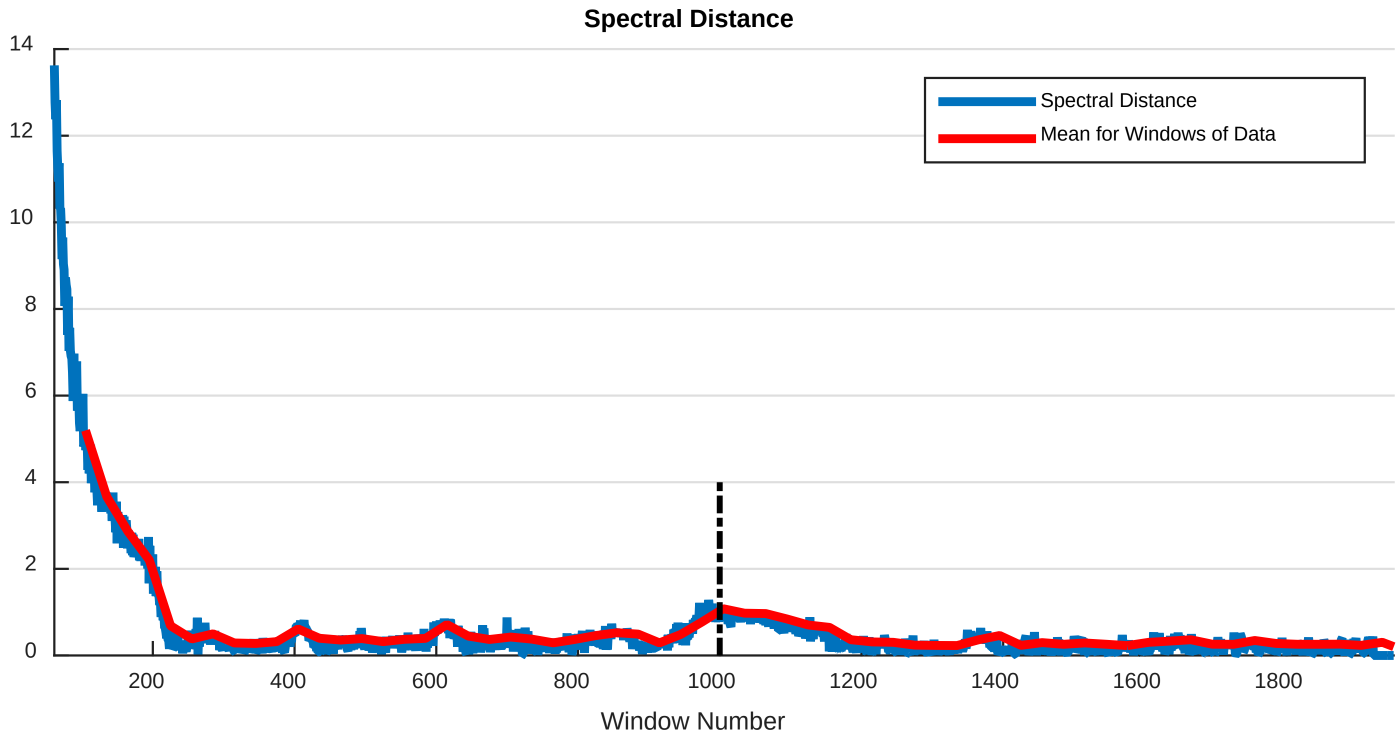
– use $\left\lfloor \frac{N}{2} \right\rfloor$ largest eigenvalues

Synthetic Test Data

Gaussian Distribution 1 with 1000 Iterations	Gaussian Distribution 5 with 1000 Iterations

2000 Iterations

Results



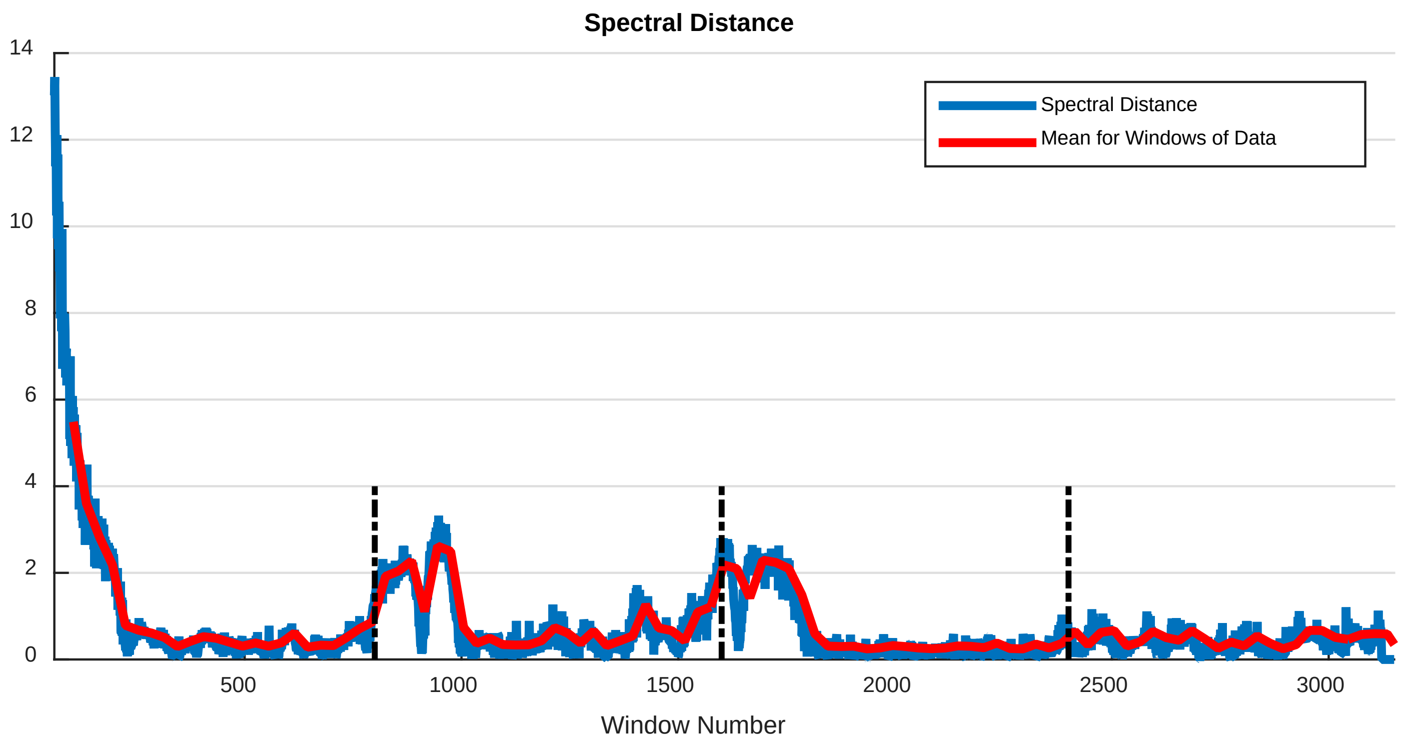
Spectral Distance
14
12
10
8
6
4
2
0
200
400
600
800
1000
1200
1400
1600
1800
Window Number
Spectral Distance
Mean for Windows of Data

Test Data

Gaussian Distribution 3 with 800 Iterations	Gaussian Distribution 4 with 800 Iterations	Gaussian Distribution 5 with 800 Iterations	Gaussian Distribution 3 with 800 Iterations

3200 Iterations

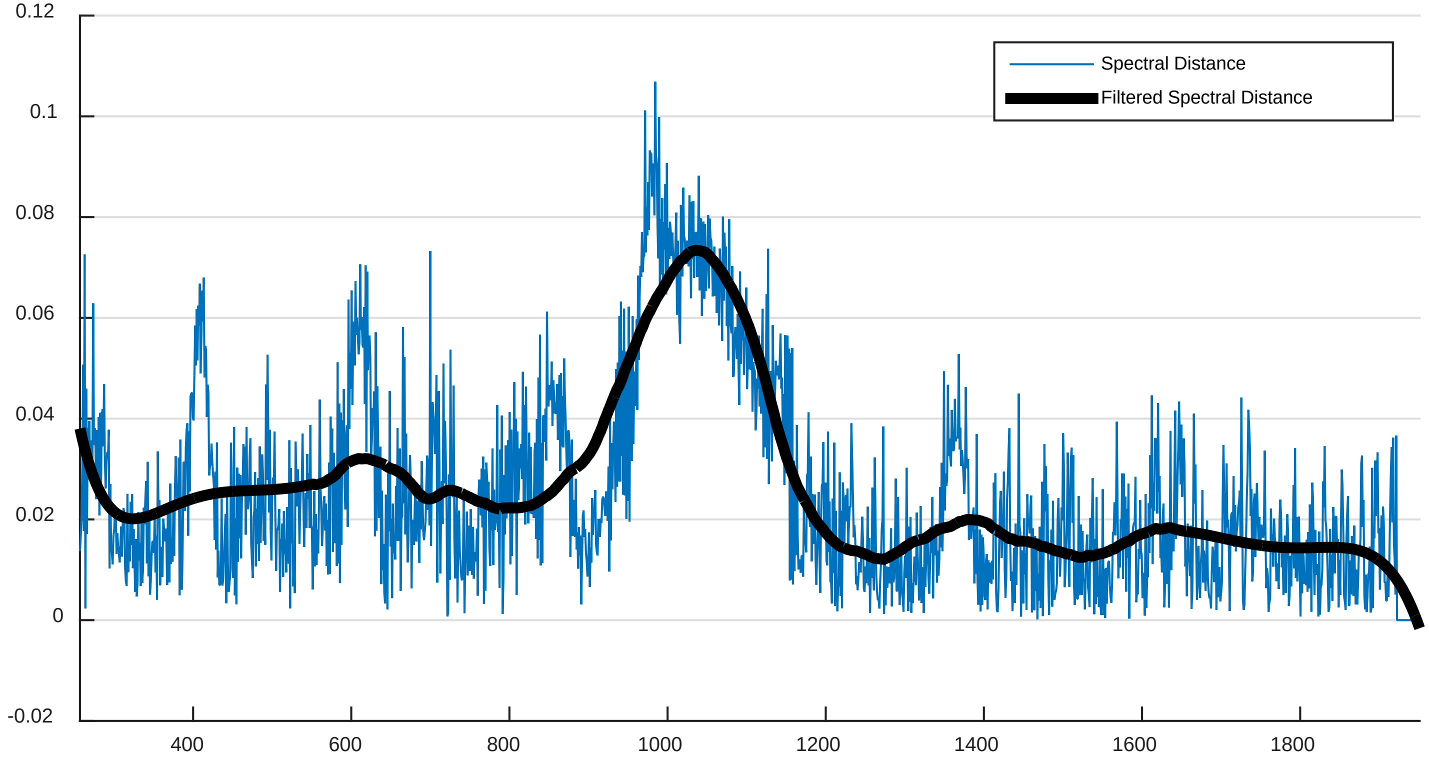
Results


Spectral Distance
14
12
10
8
6
4
2
0
500
1000
1500
2000
2500
3000
Window Number
Spectral Distance
Mean for Windows of Data

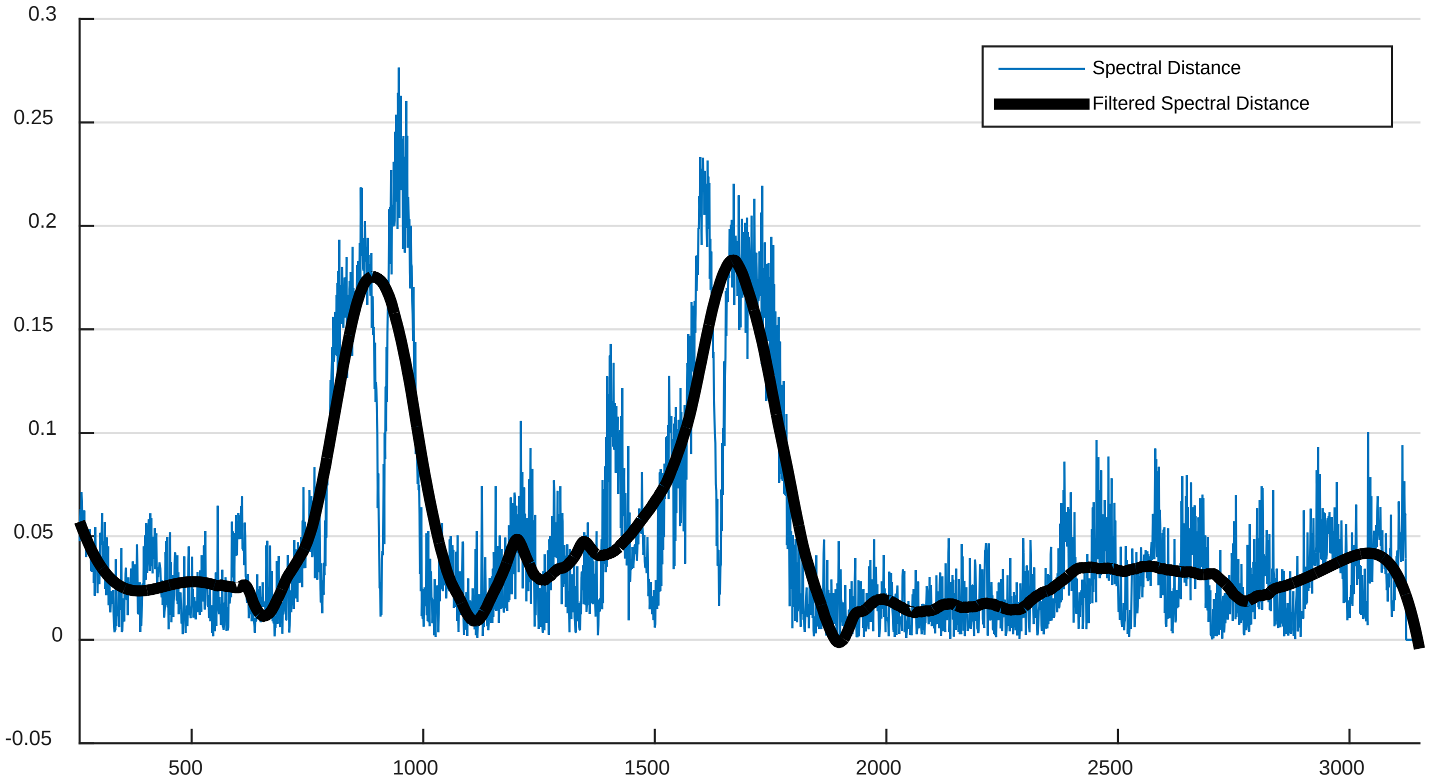
Change Point Detection

- **Before calculating the change points, smoothing the distance vector eliminates small fluctuations.**
- **Filtered instead of Averaging is recommended.**
- **We suggest using "Savitzky-Golay FIR Smoothing Filter".**

Results: First Data Set



Results-Second Data Set



Change Point Detection Methods

- **Minimum Mean Square Error (MMSE)**
- **Cumulative Summation (CUSUM)**

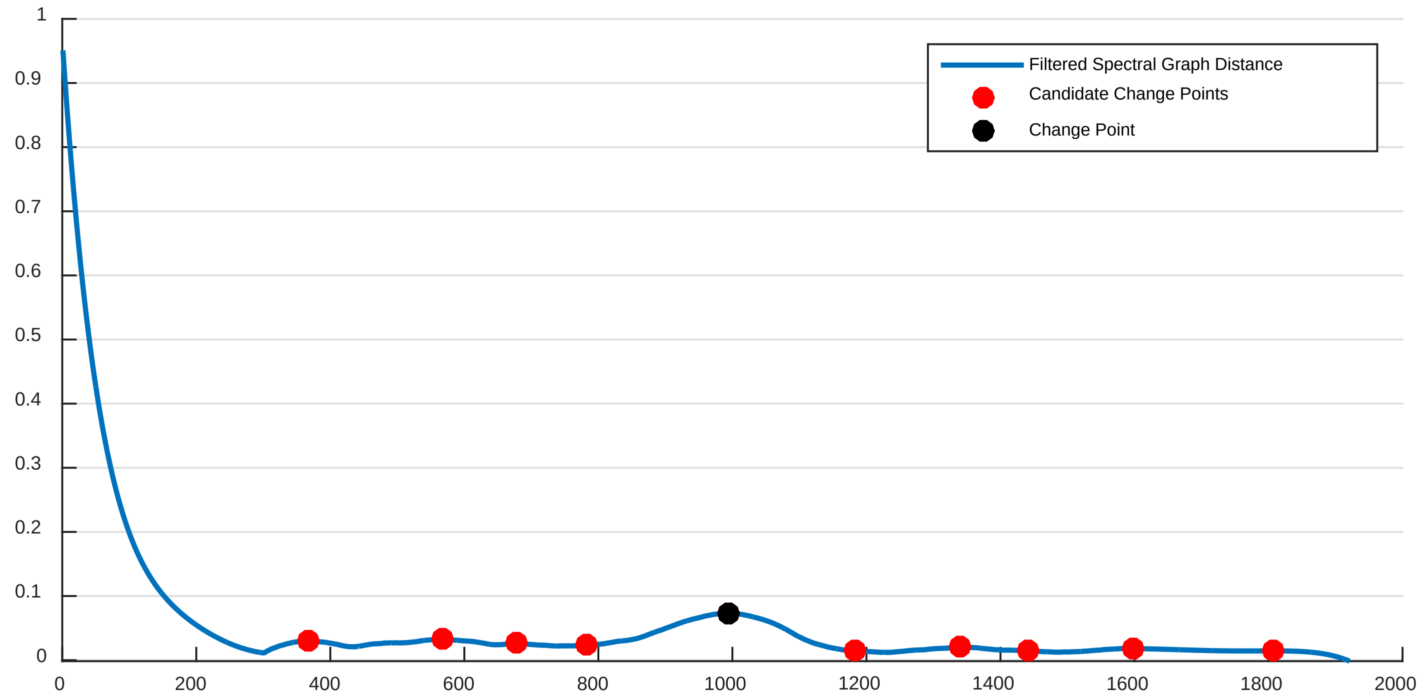
- **Combine above methods using bootstrapping and with confidence level calculations to eliminate false change points.**

Two Simple Methods

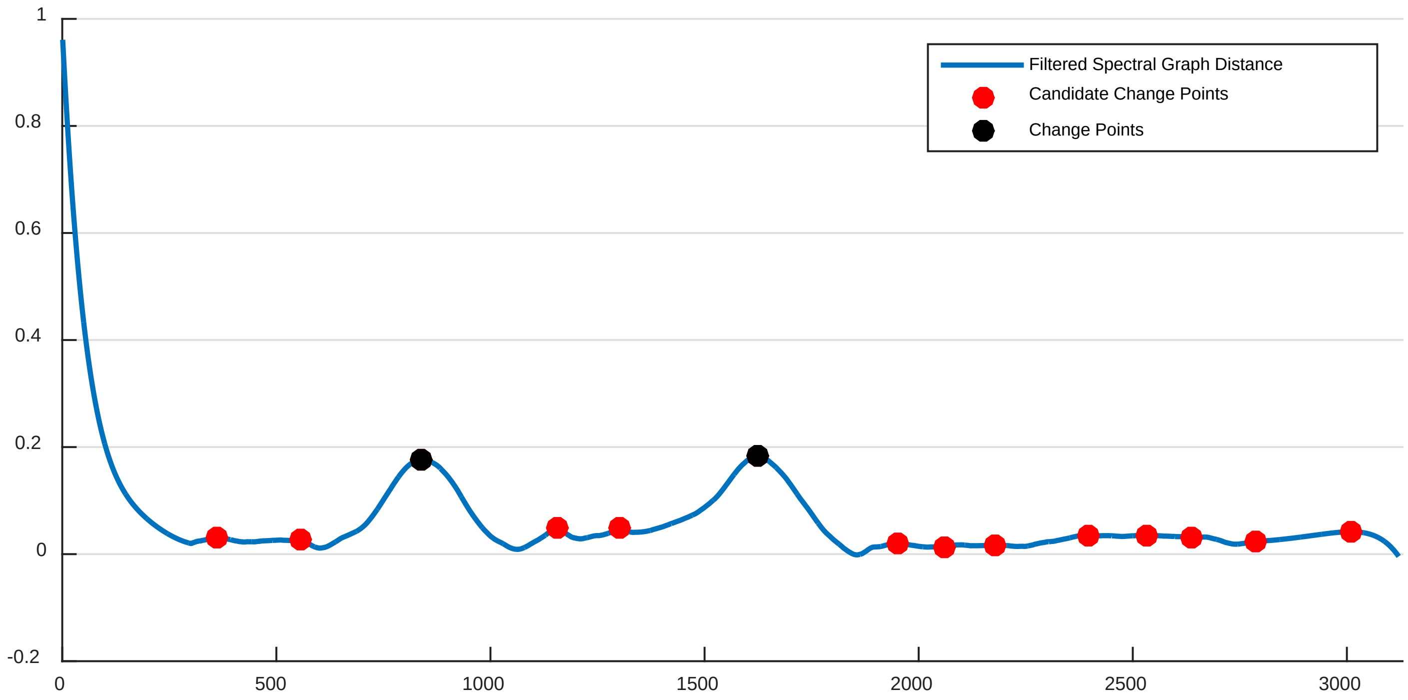
1. **Check the average of candidate change points**
   - Change point is the point with a value 30% higher than the average.
2. **Calculate the angle between the line connecting two consequent candidate change points**
   - Change point is the point with angle above 70 degrees.

- **Both methods give similar results**

Results: First Dataset



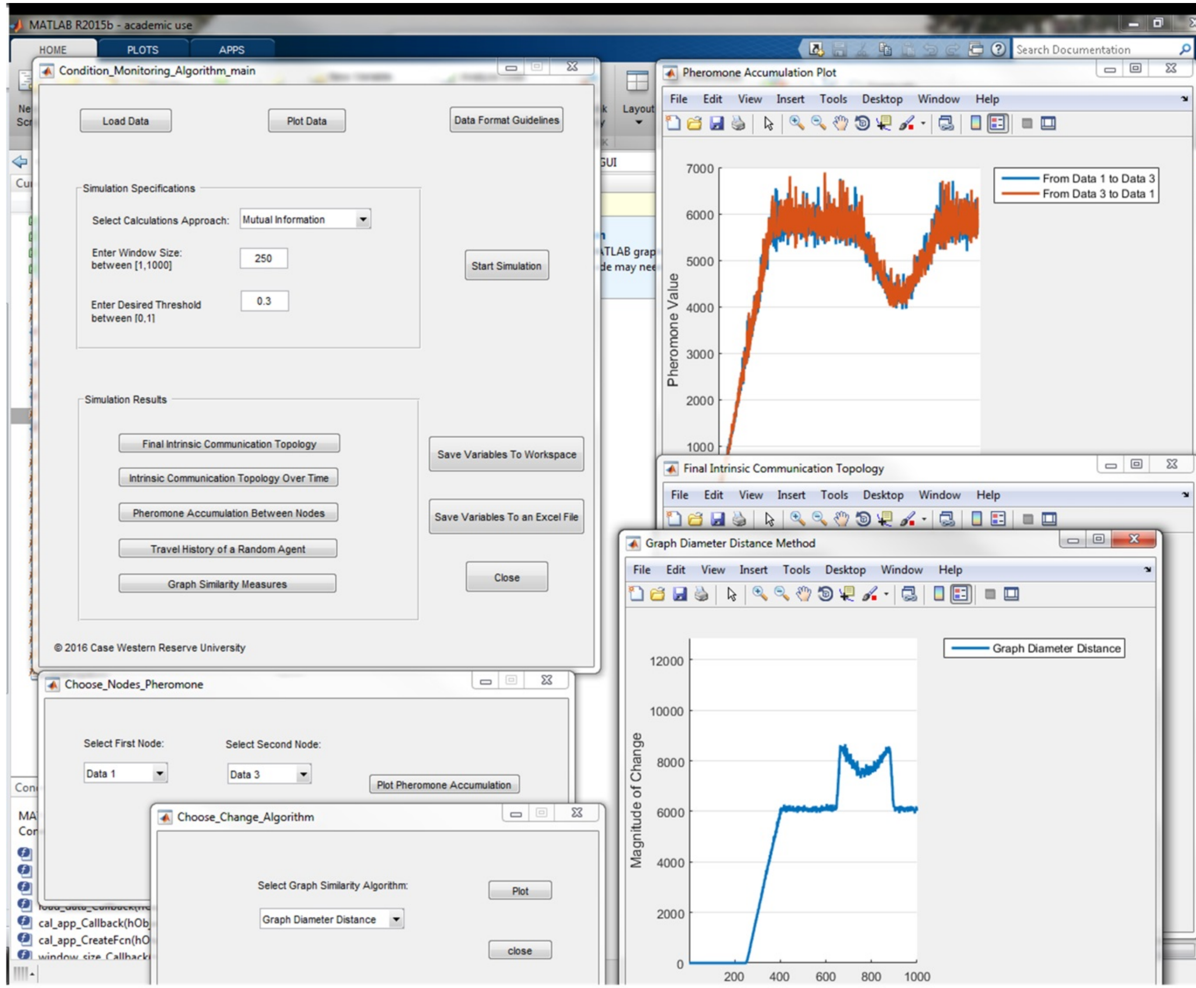
Results: Second Dataset



Interactive Network Detection Tool

Case Western Reserve University | Great Lakes Energy Institute

- Data analytics GUI
- Takes file inputs
- User driven analysis
  - *Context driven options*
  - *Context menus for simulation and visualization*
- Integrates all tools
  - *Mutual information*
  - *Self-organizing network discovery*
  - *Change detection*
- Basis for future demonstration efforts



CASE WESTERN RESERVE UNIVERSITY | GREAT LAKES ENERGY INSTITUTE

Future Application: Ultra-Supercritical Steam Plant

- Simulation of a 1000 Mwe Steam Power Plant
  - Main steam flow: 600˚C at 58 bar g
  - Net heat rate: 9,045 kJ/kWh

Questions?

"Say . . . what's a mountain goat doing way up here in a cloud bank?"