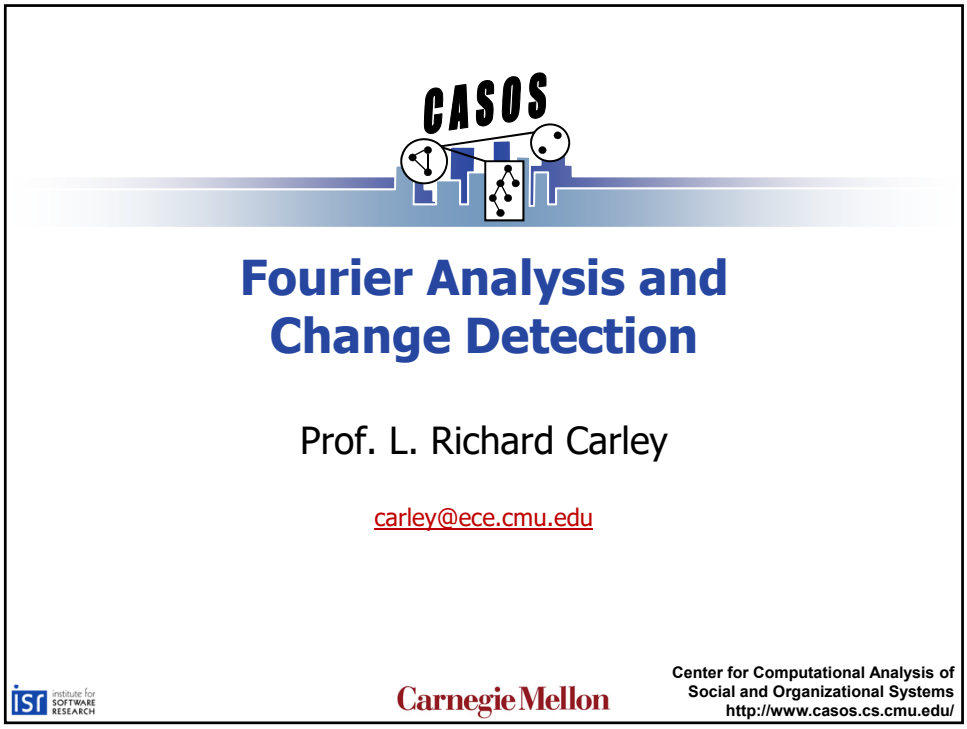
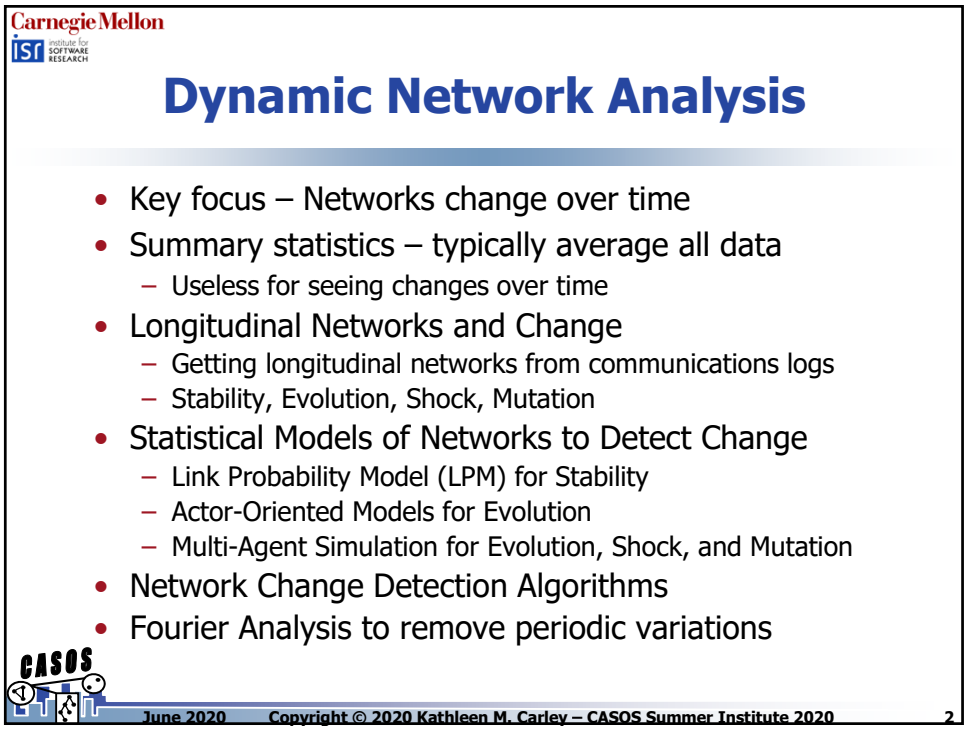

CASOS
Fourier Analysis and
Change Detection
Prof. L. Richard Carley
carley@ece.cmu.edu
institute for SOFTWARE RESEARCH
Carnegie Mellon
Center for Computational Analysis of
Social and Organizational Systems
http://www.casos.cs.cmu.edu/


Carnegie Mellon
institute for SOFTWARE RESEARCH
Dynamic Network Analysis
• Key focus – Networks change over time
• Summary statistics – typically average all data
– Useless for seeing changes over time
• Longitudinal Networks and Change
– Getting longitudinal networks from communications logs
– Stability, Evolution, Shock, Mutation
• Statistical Models of Networks to Detect Change
– Link Probability Model (LPM) for Stability
– Actor-Oriented Models for Evolution
– Multi-Agent Simulation for Evolution, Shock, and Mutation
• Network Change Detection Algorithms
• Fourier Analysis to remove periodic variations
CASOS
June 2020
Copyright © 2020 Kathleen M. Carley – CASOS Summer Institute 2020
2

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

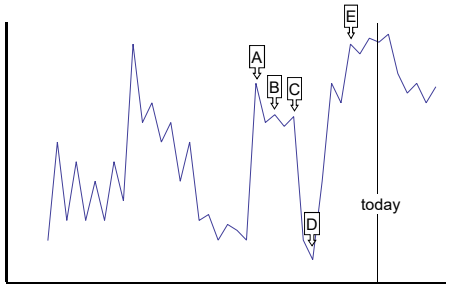
Types of Changes in Network Data

- Stability: Relationships remain statistically the same over time
  - If you are a signal processing person, the Network is "Ergodic"
- Evolution: Interaction among agents cause the relationships to change over time.
  - All link weights / costs are evolving over time during obervations
- Shock: Change is exogenous to the social group.
  - E.g., like an earthquake hits Southern California
- Mutation: A shock stimulates evolutionary behavior.
  - E.g., after earthquake, people form many new links trying to survive

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Changes in Network Data

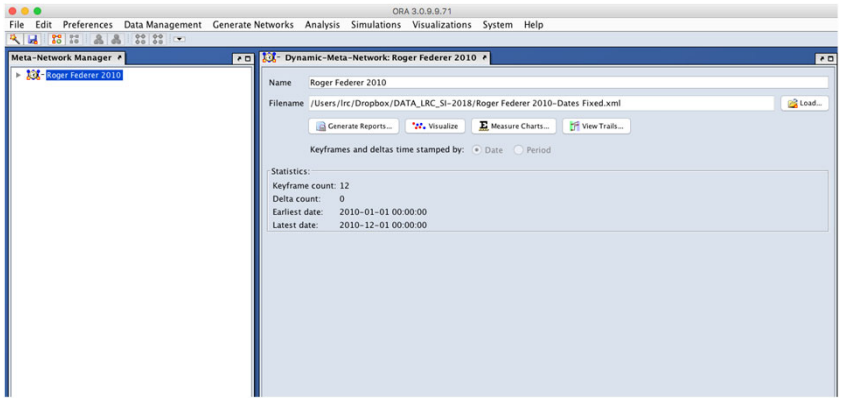
- Various measures of a network are taken for a window at each time point.
- Change detection: quickly determine *that* a change occurs.
- Change point identification: *when* did the change occur.



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Change Detection Hands-On

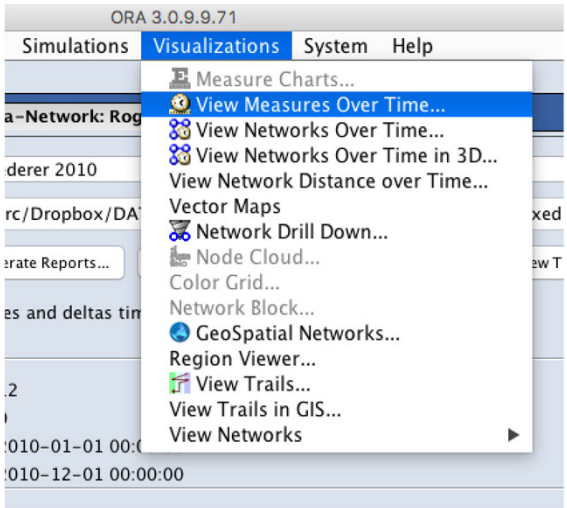
- Based on Roger Federer 2010 data



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

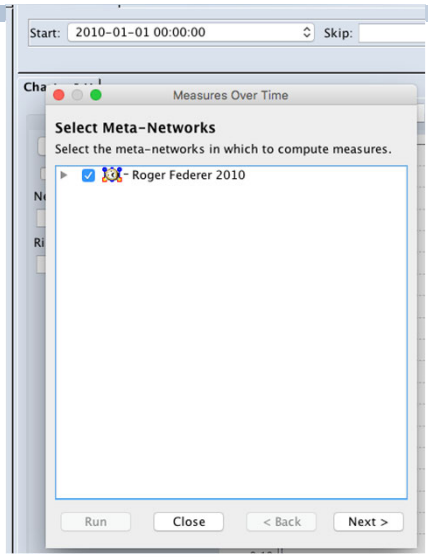
Change Detection Hands-On

- Analysis uses over-time changes in "measures" based on the network data



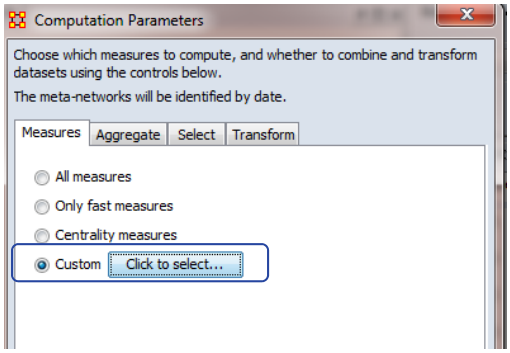
Carnegie Mellon
institute for SOFTWARE RESEARCH

Select The Metanetwork



Carnegie Mellon
institute for SOFTWARE RESEARCH

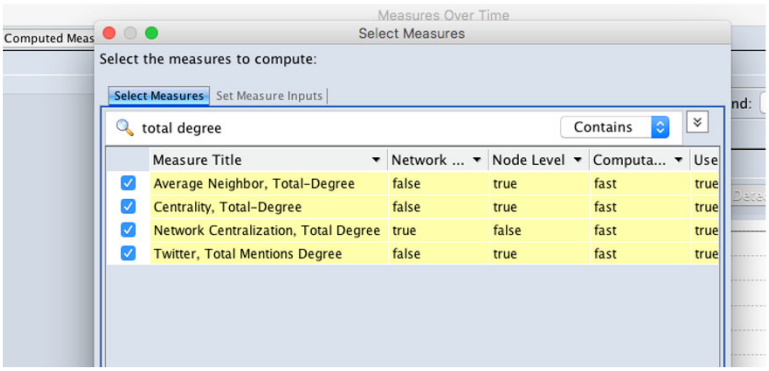
Custom Measure Selection



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

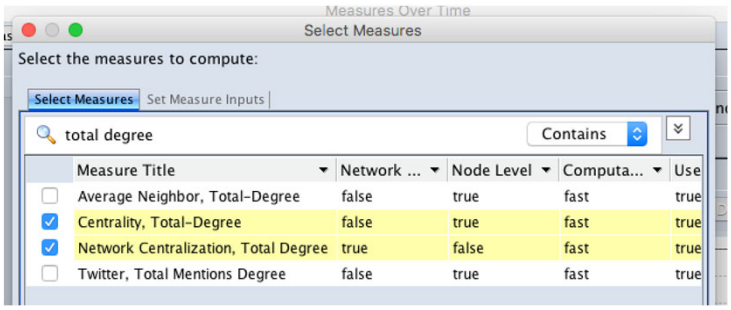
Use Search to Find Measure

Hint: Click Select Box at bottom to deselect all measure,
Then use search to find the ones you want



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Two measures selected to run

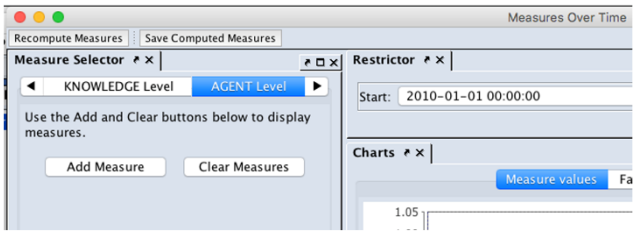


Add Measure – Agent Based Measure – select “Centrality, Total Degree”

Carnegie Mellon
institute for SOFTWARE RESEARCH

Now Select Display

First step is to select type of variables to display
– AGENT Level in this case

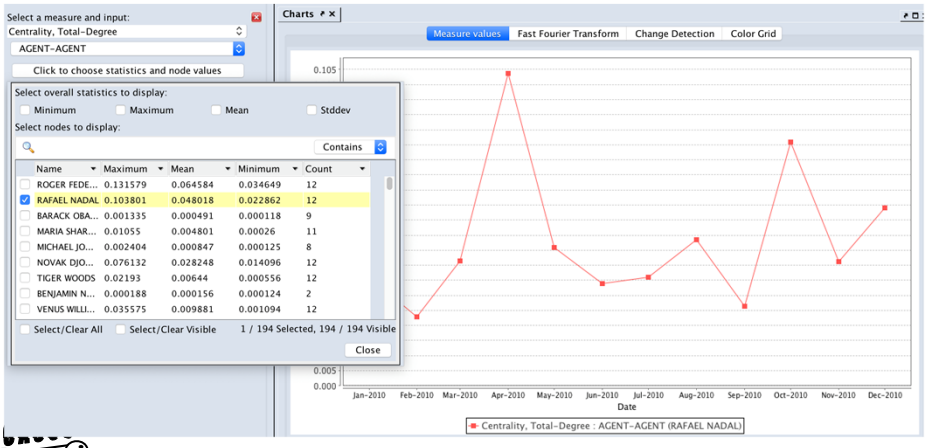


Then click on “Add Measure” to add a new plot line

Carnegie Mellon
institute for SOFTWARE RESEARCH

Select Agent for Measure

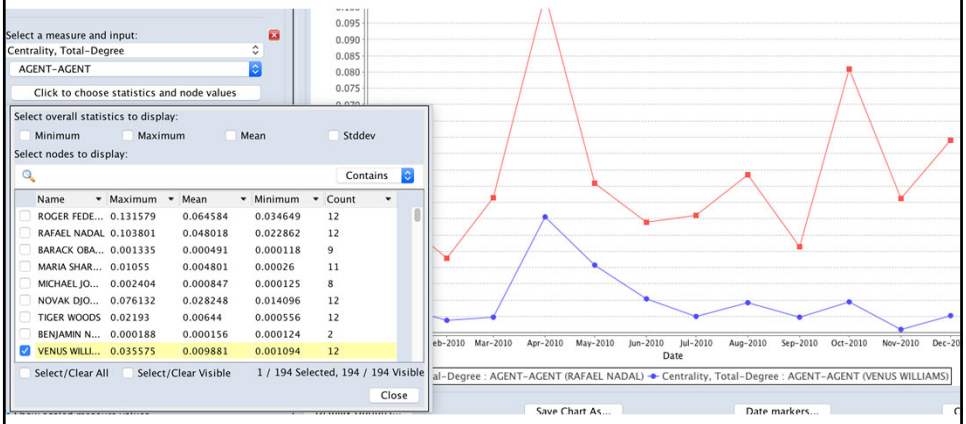
Click on the “Click to choose ….” button
and select second agent for analysis (Federer is always primary)



Carnegie Mellon
institute for SOFTWARE RESEARCH

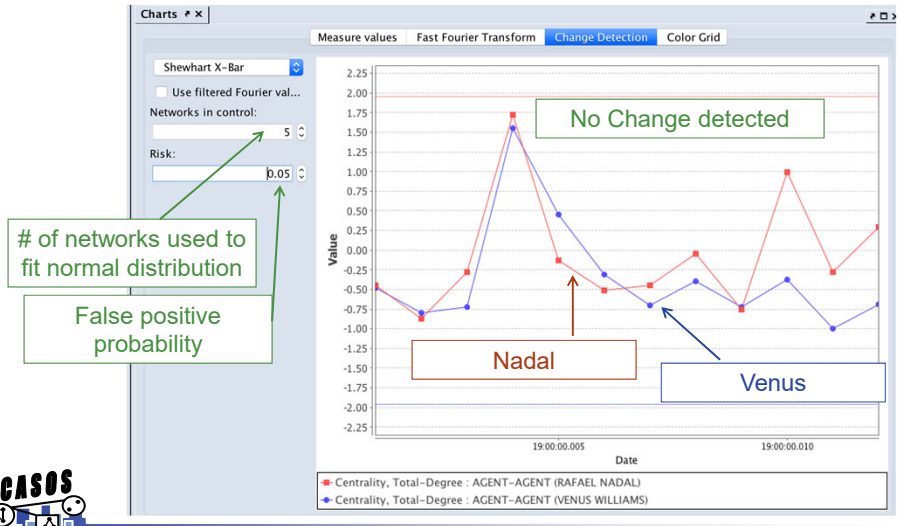
Add Another Agent Measure Plot

Click on Add a Measure button again to add another line



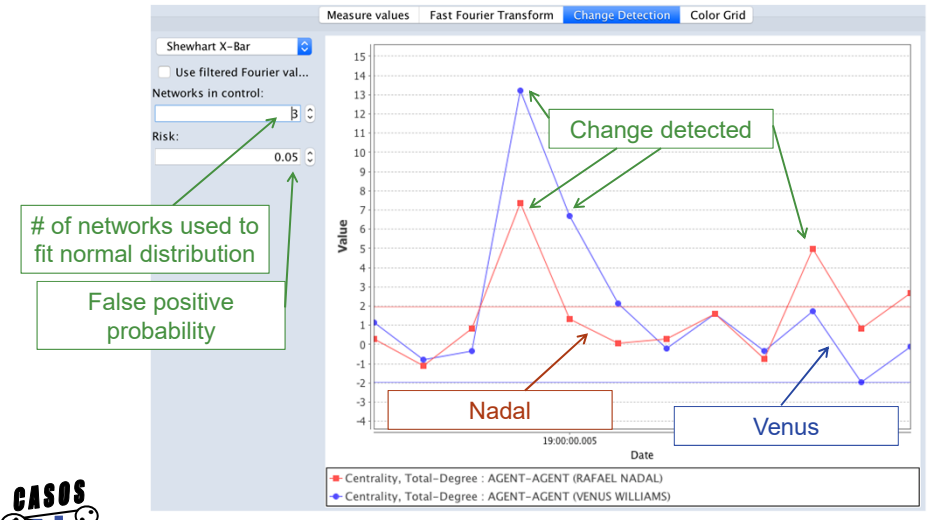
Carnegie Mellon
institute for SOFTWARE RESEARCH

Change Detection Hands-On
The Shewhart X-Bar Chart



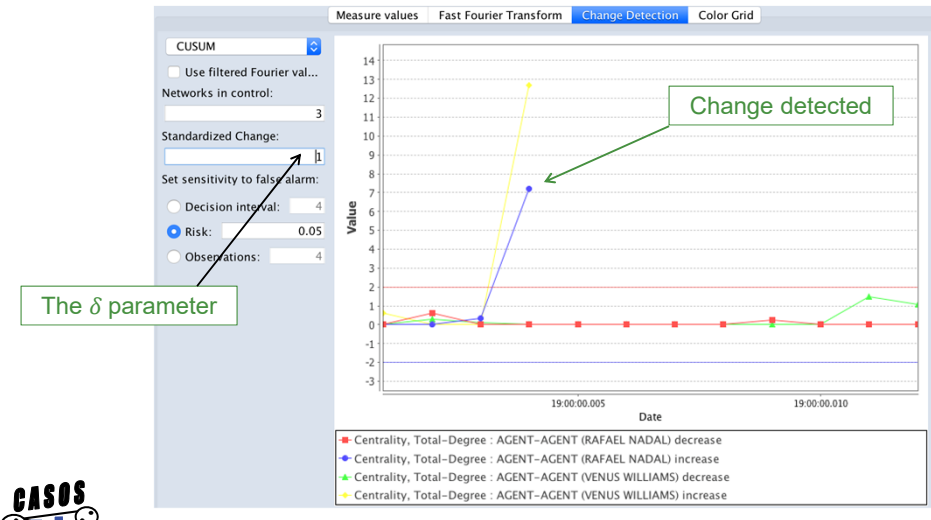
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Change Detection Hands-On
The Shewhart X-Bar Chart



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Change Detection Hands-On
CUMSUM Method



Carnegie Mellon
institute for SOFTWARE RESEARCH

Fast Fourier Transform (FFT)

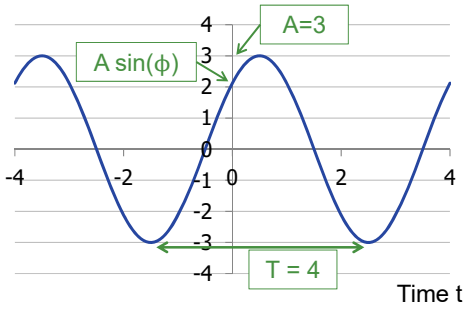
- Goal: detect periodicity in over-time data
- Examples
  - Weekly periodicity in email data
  - Time of the day effects
- Fourier's theorem
  - Any time signal can be represented by a sum of sinusoidal functions with different frequencies, amplitudes and phase shifts
- Fourier transform finds sinusoids that decompose a signal
  - Analogy: given a dish, find the ingredients
  - Sinusoids have the advantage that they are orthogonal

Carnegie Mellon
institute for SOFTWARE RESEARCH

Sinusoidal Function

- A sinusoidal function $y = A\sin(2\pi ft + \phi)$ has
  - A amplitude
  - f frequency ($T = \frac{1}{f}$ is the period)
  - ϕ phase

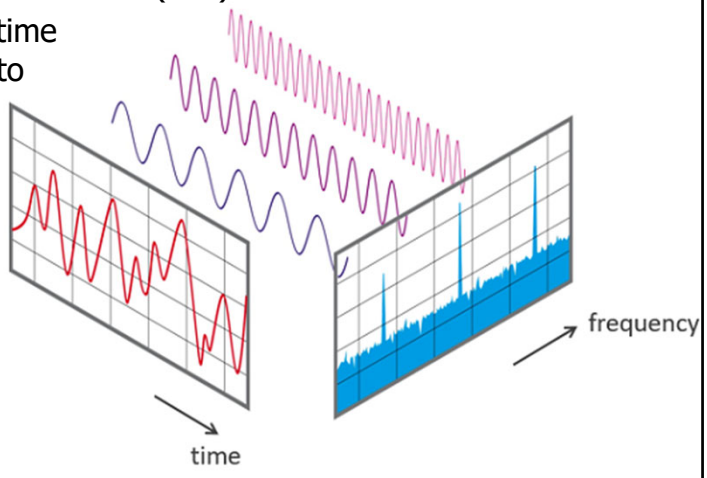
y = 3 sin(2π(0.25)(t + 0.5))



Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Frequency Domain: The FFT

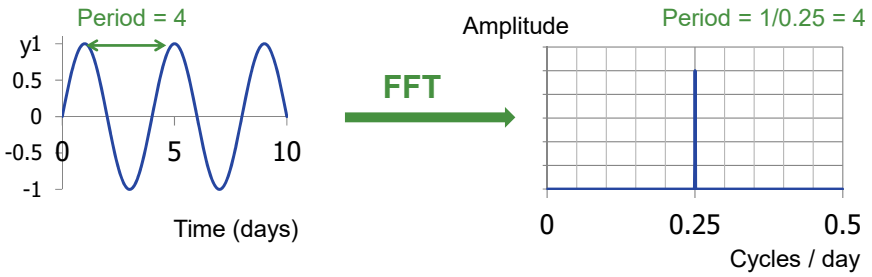
- Fast Fourier Transform (FFT)
- Decompose time waveform into sum of sinewaves



Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

FFT Example: Sum of Sinusoids

- Fast Fourier transform of sinusoidal function is a spike at the sinusoidal frequency
- Example $y = sin(2\,\pi\,0.25\,t)$



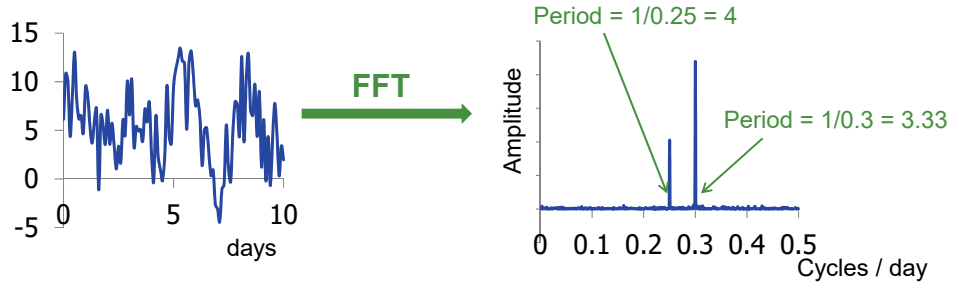
Over-time signal

FFT

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

FFT Example 2

- FFT finds periodicities that may be unclear in over-time signal



Over-time signal

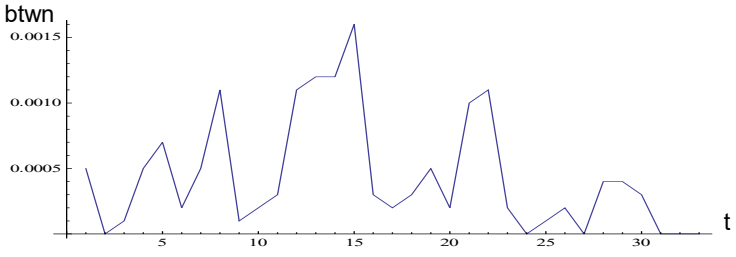
FFT

Hidden "recipe": over – time signal computed as
y(t) = 2 sin(2 pi 0.25 t) + 3 sin(2 pi 0.3 t + 0.2) + noise

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

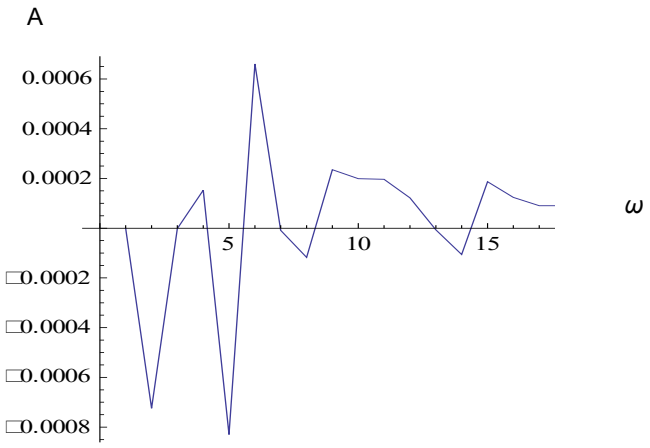
Fourier Analysis Example 3

- 24 cadets in a regimental chain of command agreed to have their email monitored to form a social network data set known as IkeNet3.
- The betweenness was calculated based on the e-mail communications observations over the first month in their duty positions.



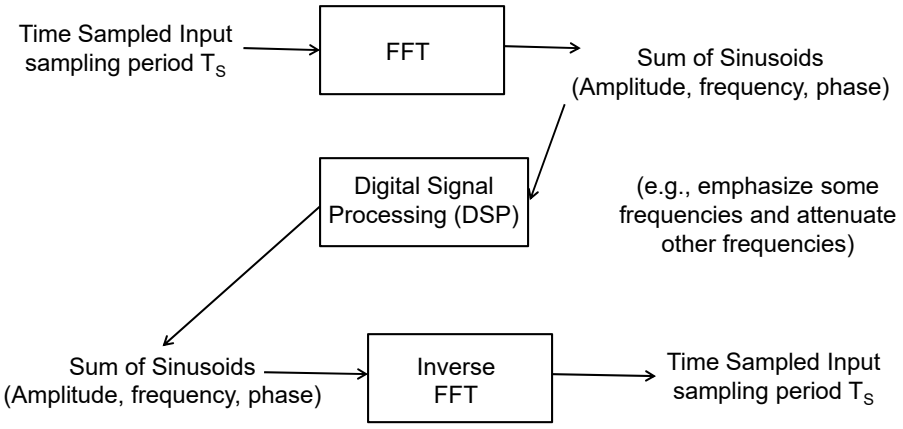
Carnegie Mellon
SEI institute for SOFTWARE RESEARCH

Fourier Analysis – Example 3



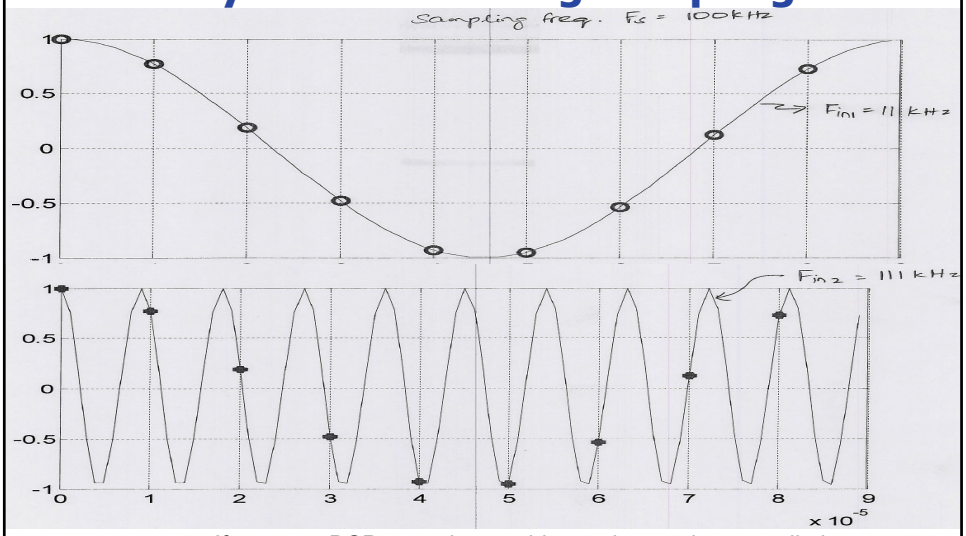
Carnegie Mellon
SEI institute for SOFTWARE RESEARCH

Digital Signal Processing



Carnegie Mellon
institute for SOFTWARE RESEARCH

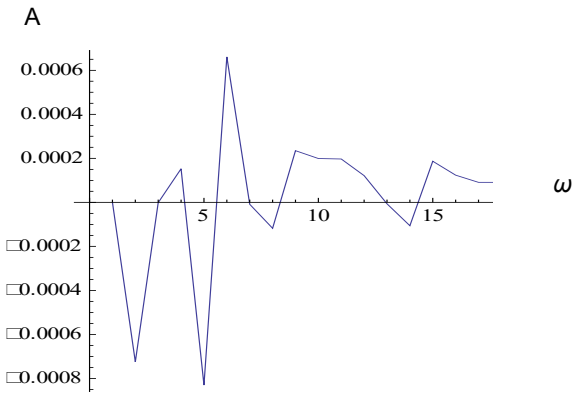
One Key Issue: Aliasing Sampling Limits



If we want DSP to work unambiguously, we choose to limit maximum input frequency to ≤ 1 / (2 T_S) = F_S / 2 = "Nyquist Frequency"

Carnegie Mellon
institute for SOFTWARE RESEARCH

Digital Signal Processing

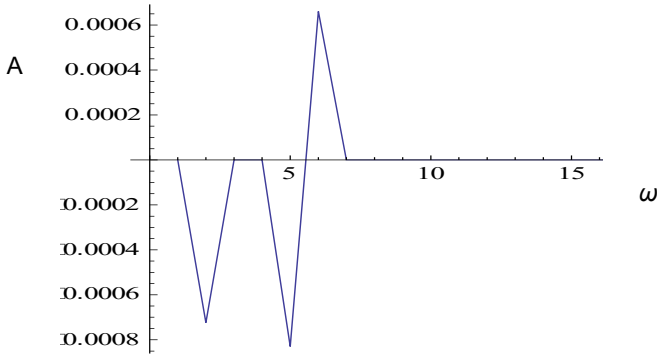


How do we use this?

One possible approach – big peaks are periodic "background"

Carnegie Mellon
institute for SOFTWARE RESEARCH

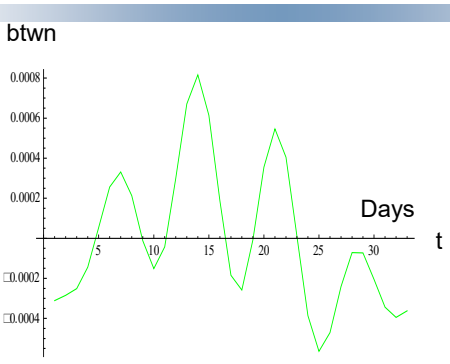
Nonlinear Filtering



e.g., identify 3 main (high magnitude) components
keep them and remove FFT components
at all other frequencies

Carnegie Mellon
institute for SOFTWARE RESEARCH

Inverse Fast Fourier Transform

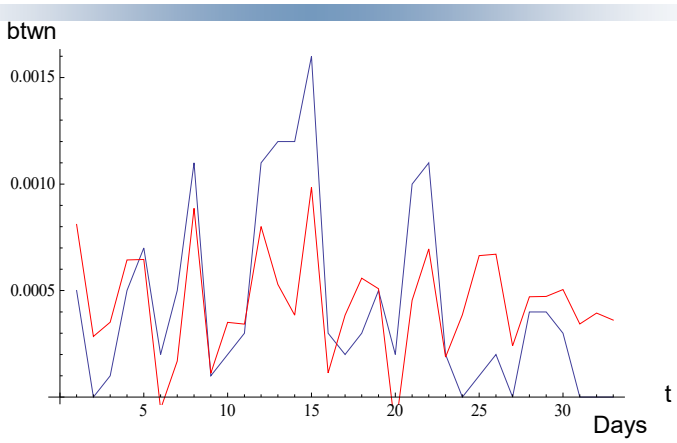


This is the inverse Fourier transform of just the 3
selected components, which are then reconverted to time waveform

There is a weekly, two week and three week cycle

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Anomaly Detection

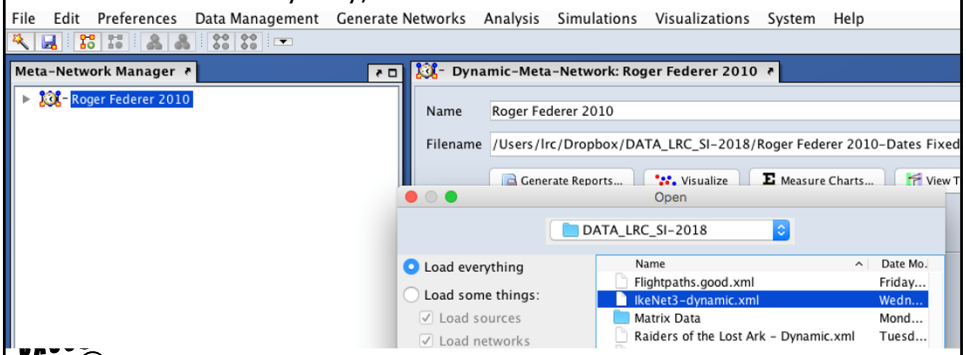


The filtered pattern has been subtracted from the original
The red is what is left – the anomalies

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

FFT Example Hands-On

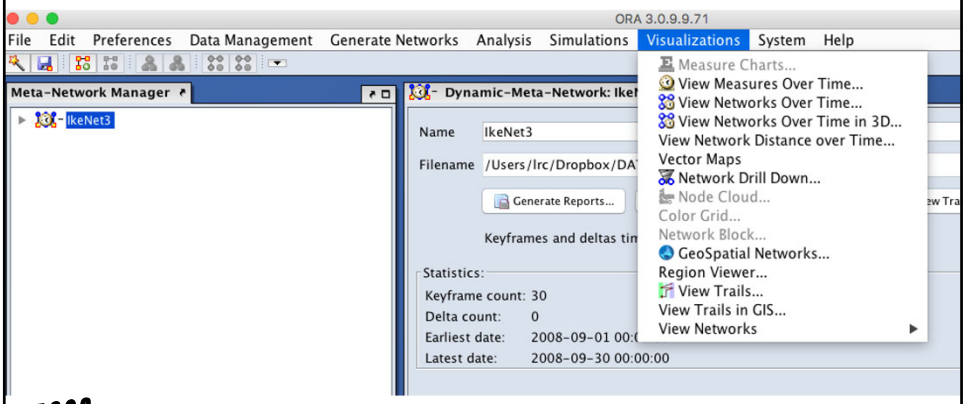
- IkeNet data (IkeNet3-dynamic.xml)
  - Email exchange data among mid-career officers in a one-year graduate program at Columbia University
  - Granularity: day; Duration: month



Carnegie Mellon
institute for SOFTWARE RESEARCH

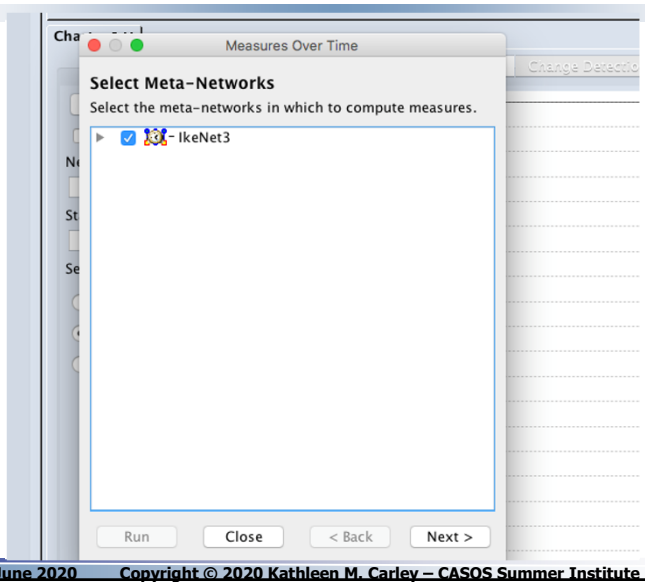
Over Time Measures

Again View Measures Over Time



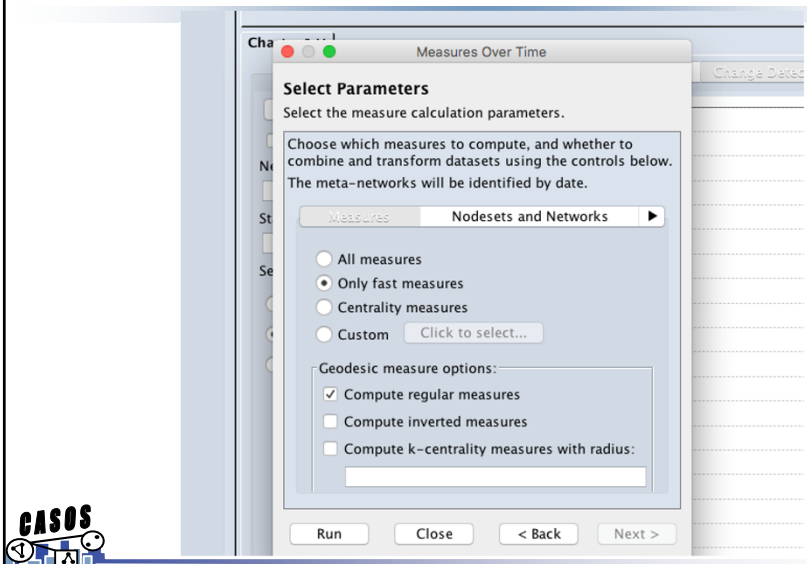
Carnegie Mellon
institute for SOFTWARE RESEARCH

Selecting MetaNetwork



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

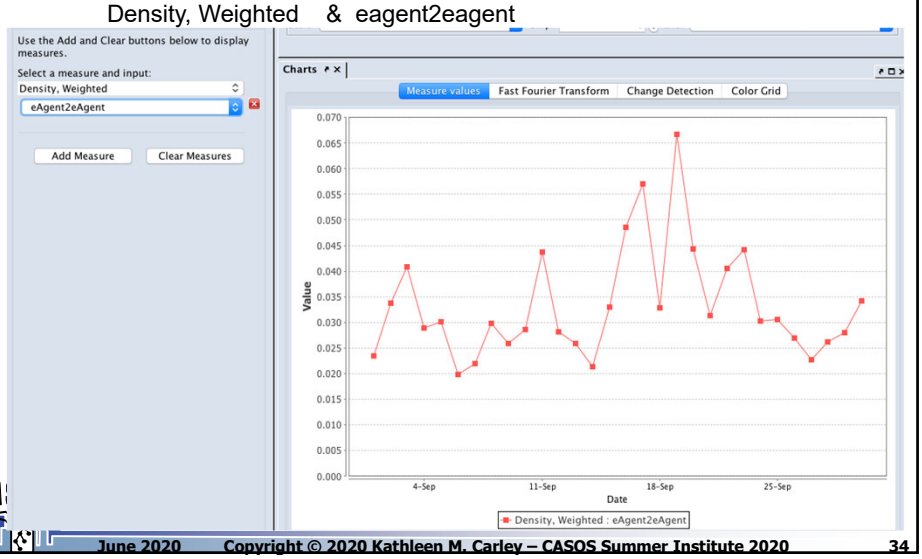
Selecting Measures



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

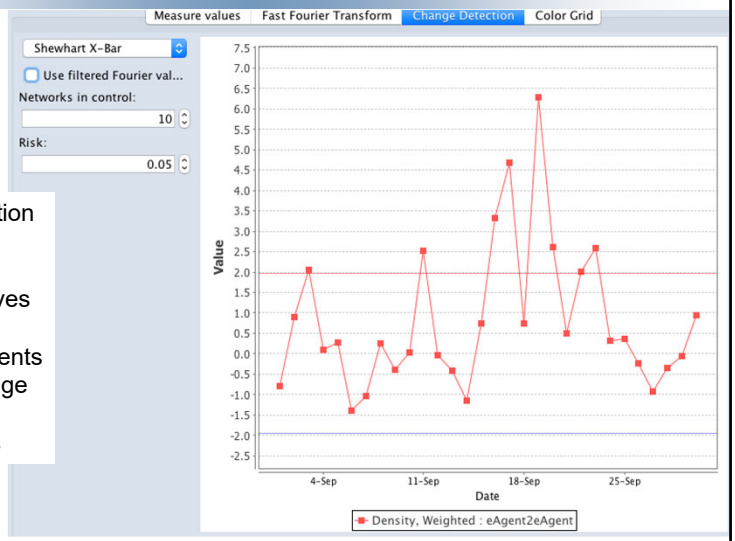
Selecting Measures to Plot

Density, Weighted & eagent2eagent



Carnegie Mellon
institute for SOFTWARE RESEARCH

Run Shewhart to Detect Change



Run Change Detection
Shewhart
10 points for basis
0.05 for false positives

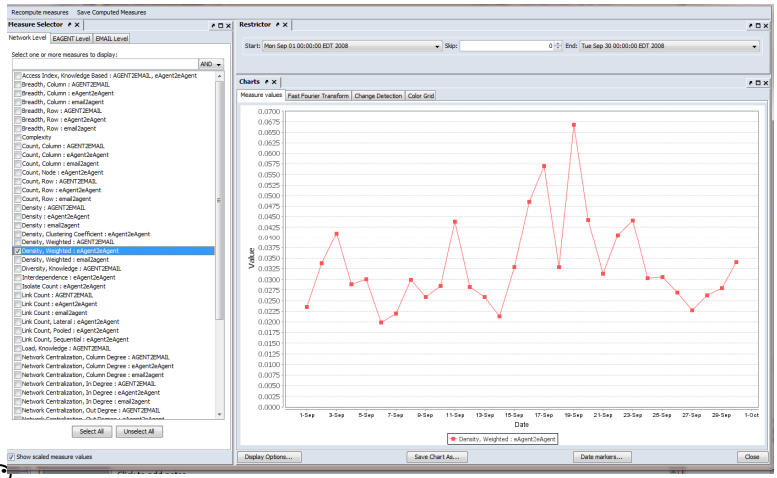
Lots of apparent events
That could be Change

All upward changes

Carnegie Mellon
institute for SOFTWARE RESEARCH

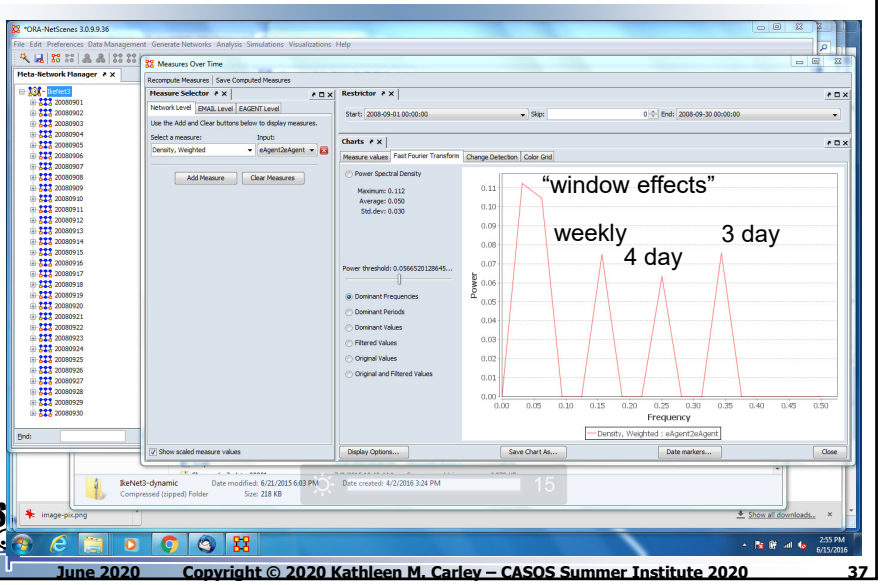
FFT Can Help with Periodic Patterns

Network Level / Density, Weighted / eAgent2eAgent



Carnegie Mellon
institute for SOFTWARE RESEARCH

FFT Power Spectral Density



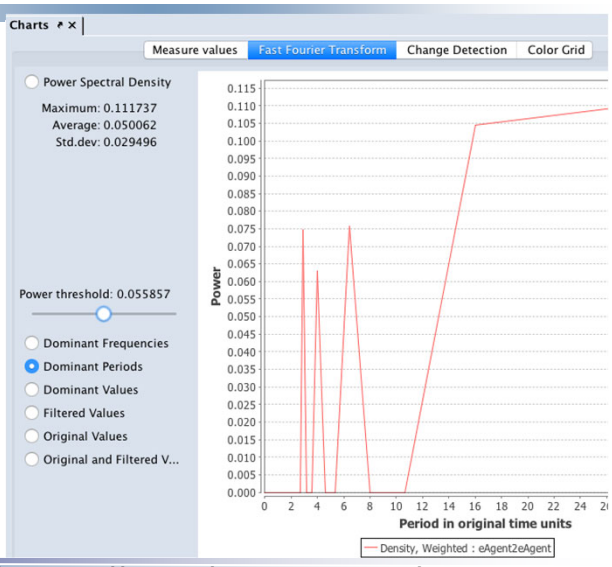
Carnegie Mellon
institute for SOFTWARE RESEARCH

FFT Power vs. Period

Adjust Power Threshold to just select "background" periodic variations to be removed

NOTE – viewing the Dominant Periods is often more intuitive than PSD

The long period values are artifact of finite length of data

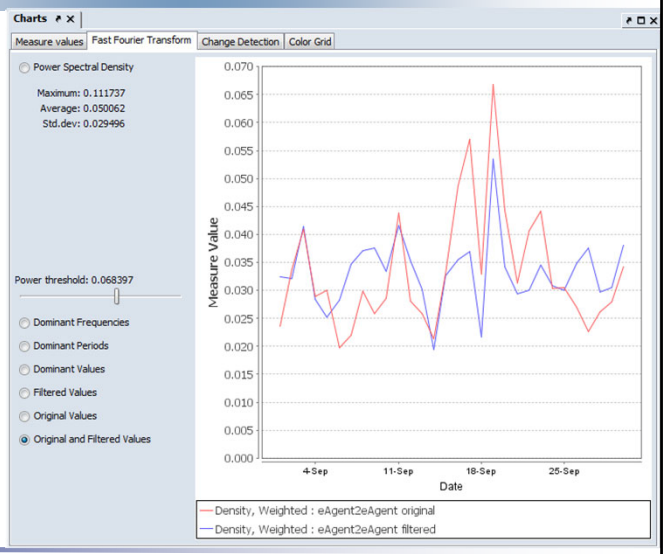


Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Original vs. FFT subtracted data

Remaining FFT components are converted back to time values (inverse FFT) and subtracted

This removes some of the periodic patterns covering up the fundamental changes to be detected



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

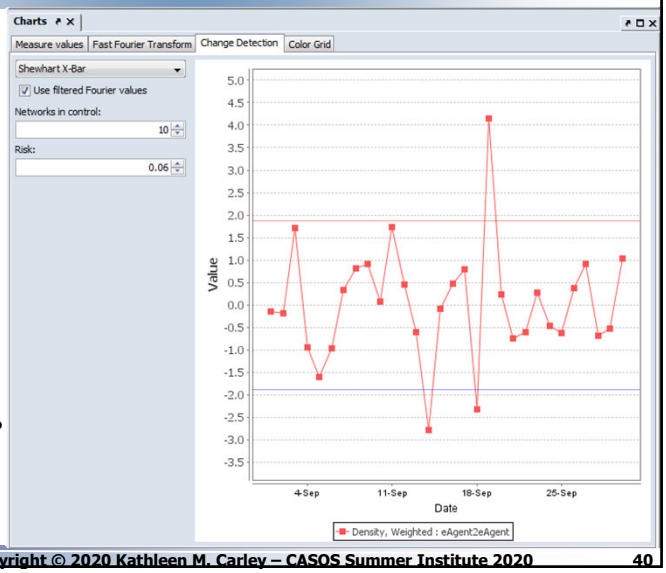
Shewhart on FFT subtracted data

Now use the Filtered Data

Run Change Detection Shewhart
10 points for basis
0.05 for false positives

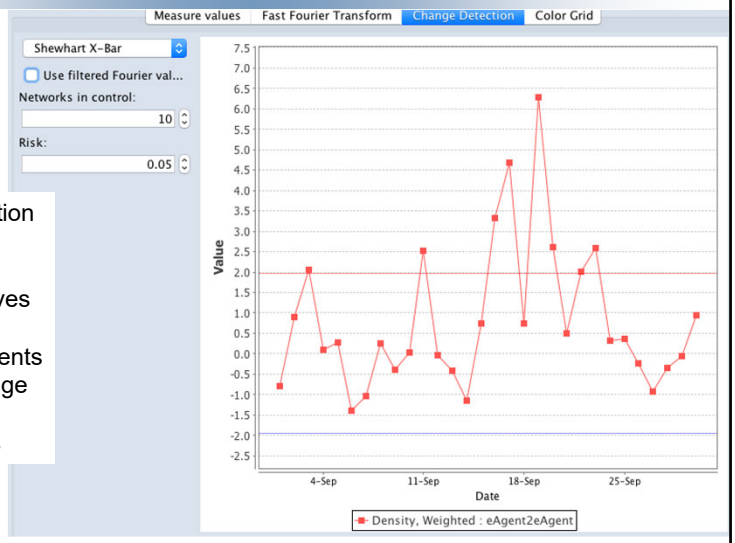
Only 1 + event and
2 – events

Open Question – is this better than original or not?



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Shewhart on original data



Run Change Detection
Shewhart
10 points for basis
0.05 for false positives

Lots of apparent events
That could be Change

All upward changes

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Fourier Analysis to Handle Periodicity

- Fourier analysis can effectively identify periodic trends in longitudinal network data.
- Identification of periodic trends can allow the analyst to aggregate relational data over the period to remove over-time dependence.
- The inverse Fourier transform of the significant period can be used to subtract off periodicity from longitudinal network data measures over time.
- Further exploration of wavelets may produce greater insights in to network dynamics.

Carnegie Mellon
institute for SOFTWARE RESEARCH

Scalability

- The change detection algorithm is linear, thus the time consuming part is calculating network measures.
- Networks with less than 20 nodes tend to have a higher variance in over time measures. When a link is added or removed, it affects (n-1)(n-2) triads.
- Requires at least 3 time periods: >2 to determine typical behavior and 1 to compare at each time point. In practice, 10+ network time points are preferred.
- No difference in number of required networks for each technique: CUSUM, EWMA, Scan Statistic, x-bar, etc.
- Wavelet/Fourier based approach needs many more time periods and complexity grows roughly as #T(log(#T))

Carnegie Mellon
institute for SOFTWARE RESEARCH

Limitations

- View findings on data with caution
- Slicing and dicing can distort conclusions
- Examine errors associated with technique through extensive simulations.
- Investigate more real world data sets.
- Investigate the degree to which network measures are correlated to understand the effects of compounding error.
- Investigate multi-dimensional network properties such as the cosine similarity between the triad census at different time periods.

Carnegie Mellon
institute for SOFTWARE RESEARCH

Summary of Change Detection

- Rapid change detection may allow an analyst to get inside a decision cycle and shape network evolution.
- Simulation is important for modeling longitudinal network behavior.
- Isolating when networks change enables more focused study on the causes of evolution, shock, and mutation, which may lead to future predictive analysis.
- Statistical process control is a useful tool for understanding social behavior.

Carnegie Mellon
institute for SOFTWARE RESEARCH

Conclusions

- Change detection
  - Detect occurrence of shocks i.e. change due to reasons exogenous to the network
- Fourier analysis
  - Detect periodicity in over-time data