

Tutorial on Over-Time Networks

Dr. L. Richard Carley
Netanomics
rick.carley@netanomics.com



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

1

1

Future Help / Questions

ORA Google Group

<https://groups.google.com/forum/#!forum/ORA-google-group>

This group is to support the sharing of information about network science, and the network science tools ORA (PRO or LITE), NetMapper, SM-deidentifier, AutoMap and Construct. All members should feel free to post and answer questions.

You can request membership in this Google Group by sending e-mail to:

ORA-google-group@googlegroups.com

Over 700 Members who can and do help answer questions from new members



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

2

2



Outline of Tutorial

- Network Representation for Dynamic Networks
 - Using ORA as a example vehicle
- Over-Time Network Stability vs. Change
 - How networks change over time
- Measures on Network Structures that vary over time
 - How to analyze Networks that change over time to get measures
 - Change Detection on Network measures
- Tracking Changes in Geo-Temporal Network Structures
 - How to analyze Networks that change over time and space



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

3

3

Start with an Example

- This dataset is a mapping of the 1981 movie “Raiders of the Lost Ark”
- It is a dynamic meta-network containing 27 different time periods
- Each time period contains the same 19 locations and 9 characters
- Each differs from another in two ways
 - The links present/not present
 - The timestamp

The screenshot shows the Meta-Network Manager interface. The left pane displays a tree view of the network structure, listing 27 time periods (keyframes) such as 'Raiders, part 01-Peru', 'Raiders, part 02-new_england', and 'Raiders, part 20-caro-omars_square'. The right pane shows the 'Dynamic-Meta-Network Rai' configuration, including the name 'Raiders of the Lost Ark', the filename 'E:\Summer Institute 2015\De', and a 'Generate Reports...' button. Below the configuration, statistics are shown: 'Keyframe count: 27' and 'Delta count: 0'. A small movie poster for 'Indiana Jones and the Raiders of the Lost Ark' is visible in the bottom right corner of the screenshot.



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

4

4



Raiders of the Lost Ark (1981)

- Movie
 - Set in 1936.
 - Archeologist Indiana Jones is hired by the US government to find the Ark of the Covenant before the Nazis.
 - The Nazis have hired Rene Belloq to do their bidding.
 - Indy partners with Marion Ravenwood.
 - Indy is 2nd most popular film hero according to American Film Institute.
- Data
 - Relational data based on draft script of Raiders
 - Relations coded among characters, resources, knowledge, locations per scene.
 - Incomplete w/minimal task representation



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

5

5

Click <+> on tree widget to show Keyframes in dynamic meta-network

Each keyframe represents interactions that occurred in a scene

83 scenes are represented

Keyframe dates range from 5/1/1936 – 6/11/1936

NOTE – you can create Keyframes in a Meta-Network by right clicking in the white space in this panel while the desired Meta-Network is selected.

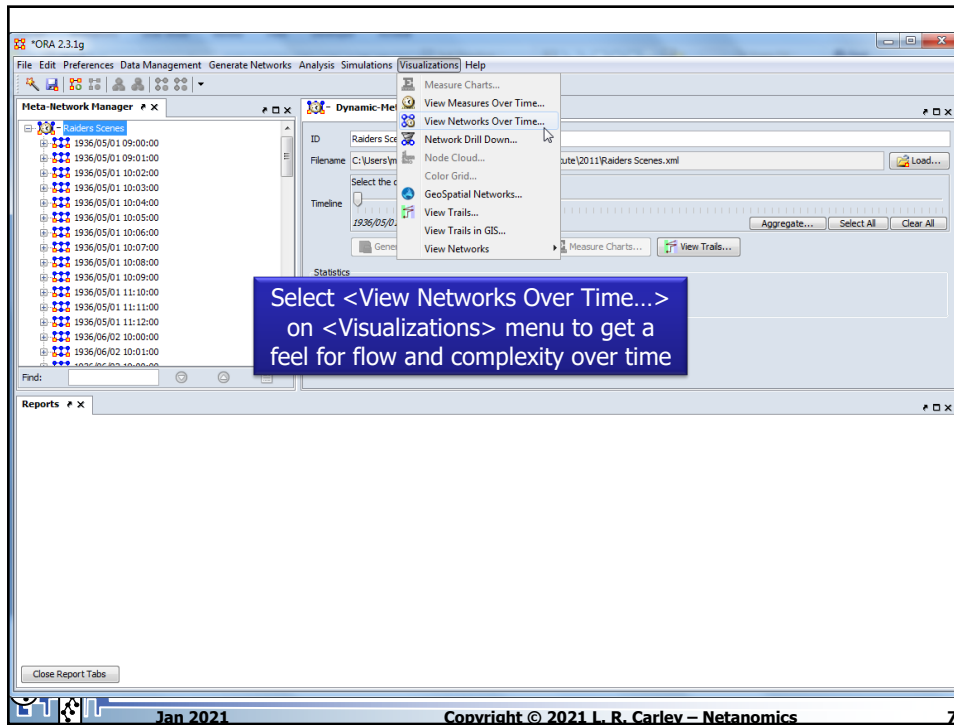
Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

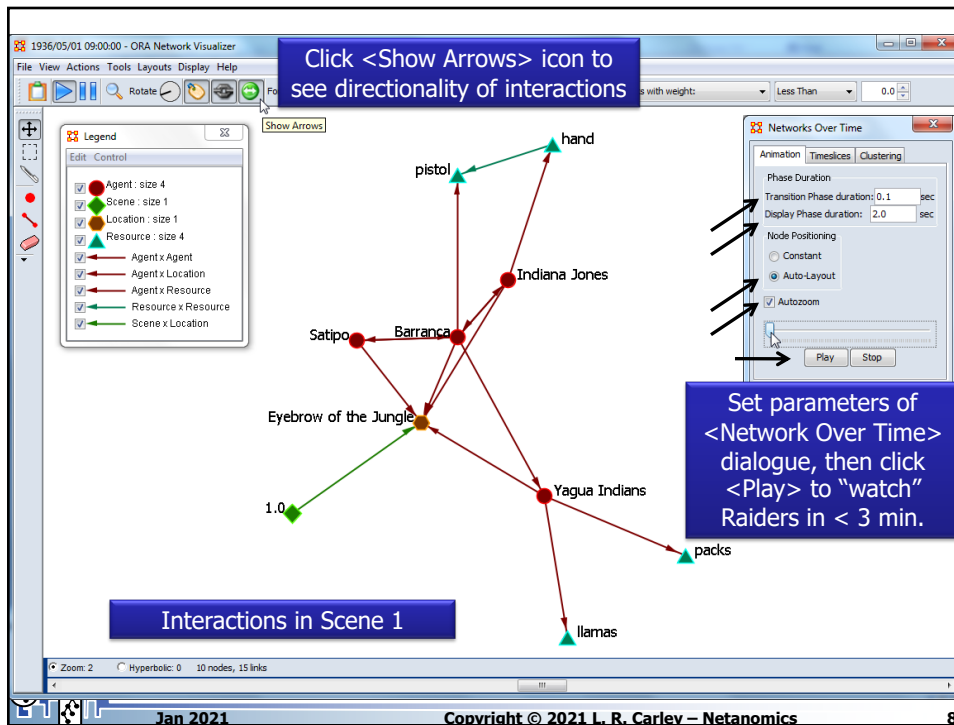
6

6





7



8



What Have We Learned?

- Indy is an important character, given a variety of relevant measures
 - Indy ranked in top 3 in 94% of measures calculated
 - Marion Ravenwood, Sallah, & Rene Belloq are also important (i.e., top-ranked in a high percentage of measures)
 - German Agents, while identified as important, is an entity that represents various extras who wore Nazi uniforms in bit parts
- Knowing of the Well of Souls & the Ark of the Covenant is important
- The Ark of the Covenant is the most important resource in the movie
- The Raven Saloon & Tanis Ruins are important locations



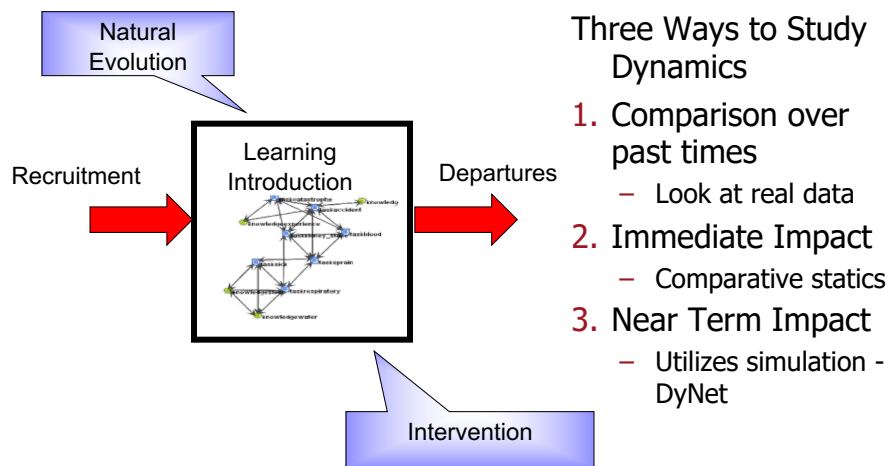
Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

9

9

Changes in Networks over Time



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

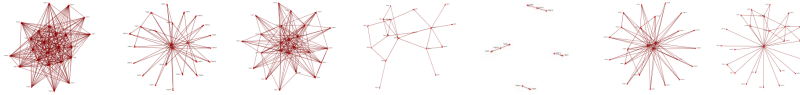
10

10



Longitudinal (Over Time) Networks

- Consider watching communications on a network, such as email. Mark a link between agents if communicated.



- Has this organization changed significantly?
- Has it evolved?
- Have people changed their position in the network?



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

11

11

One Issue: the node set

- Over time the set of nodes change
- What should you do?
 - Compare just nodes present in all time periods
 - For core group – how has it changed
 - Create a master network of all nodes
 - How has the flux altered the groups
 - Use whatever nodes are available
 - What are the natural dynamics
 - Note – choice changes many measures that are scaled by size
- No single right answer
 - Right answer depends on what you want to know
 - Often try two different approaches and see how much they differ



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

12

12



Taxonomy of Change in Network Data

- Stability: Relationships remain the same over time.
 - But will still have significant “random” variations with time
- Evolution: Interaction among agents cause the relationships to change over time.
 - Normal state of affairs with humans beings as agents
 - Still has “random” variations as well
- Shock: Change is exogenous to the social group.
 - This is crucial for many real world applications
- Mutation: A shock stimulates evolutionary behavior.
 - This is longer term response of organization to changing environment



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

13

13

Models Used to Classify “Change”

- Stability: LPM , ERGM, repeated measures
 - LPM is Link Probability Model
 - ERGMs are Exponential Random Graph Models
- Evolution: SIENA, multi-agent simulation (CONSTRUCT), or both
- Shock: Change detection in real-world applications
Multi-agent simulation for experimentation
- Mutation: Change detection coupled with SIENA for real world applications
Multi-agent simulation for experimentation



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

14

14



Dynamic Analysis Techniques

- Visualization
- Comparative Statics – Immediate Impact
- Longitudinal Networks and Change
 - Stability, Evolution, Shock, Mutation
- QAP (Quadratic Assignment Procedure) and MRQAP (Multiple Regression QAP), Longitudinal QAP
- Statistical Models of Networks
 - Link Probability Model (LPM) for Stability
 - Actor-Oriented Models for Evolution
 - Multi-Agent Simulation for Evolution, Shock, and Mutation
- Exponential Random Graph Models
- SIENA
- Statistical Process Control
- Network Change Detection
- Fourier Analysis
- Simulation (Agent-Based Dynamic Network)



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

15

15

Critical Issue: Slicing and Dicing

- Approach 1: Cumulative network
 - Each time period is all prior links plus new
 - Good for data where links don't go away – e.g., citation networks
- Approach 2: Divide based on external shock
 - Number of time windows depends on external events e.g., before and after a referendum
 - Good for data where there is a major known change
- Approach 3: Divide into uniform periods
 - Number of time windows depends on collection and time slice
 - Good for large data and for doing periodicity studies
- Approach 4: Streaming
 - Only show most recent data using some moving average
 - Good when data too large to be stored – least developed



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

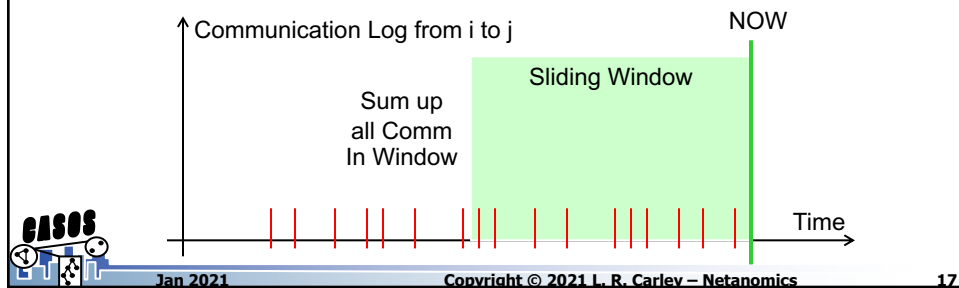
16

16



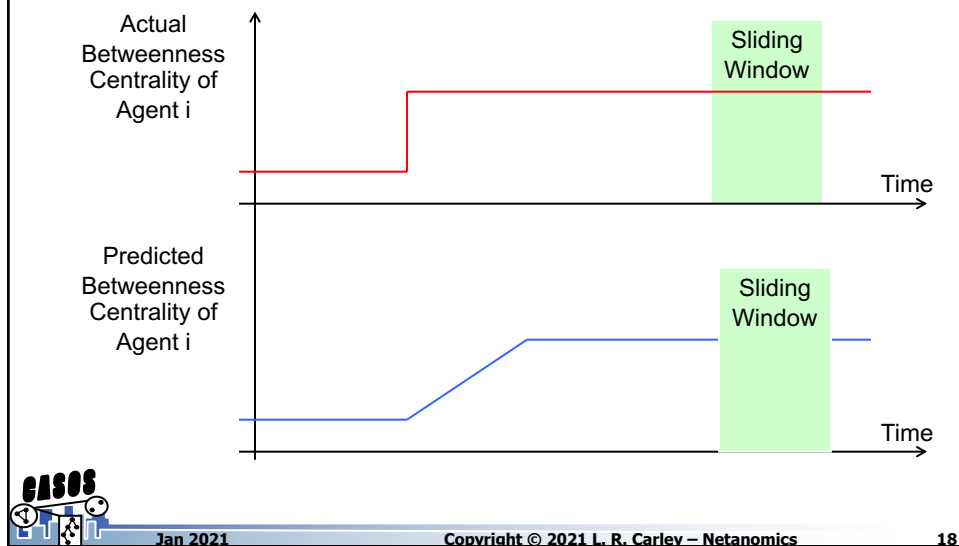
Sliding Window for Over-Time Links

- Estimator for Link Weight (a.k.a. Link Cost)
 - Add up # of Communication Events between x & y in window
 - Take reciprocal. If # is 0, there is no Link between that pair
 - Then move window forward by a time step and repeat
 - Alternatives possible:
 - Incorporate duration of communication
 - Weight different communications channels differently



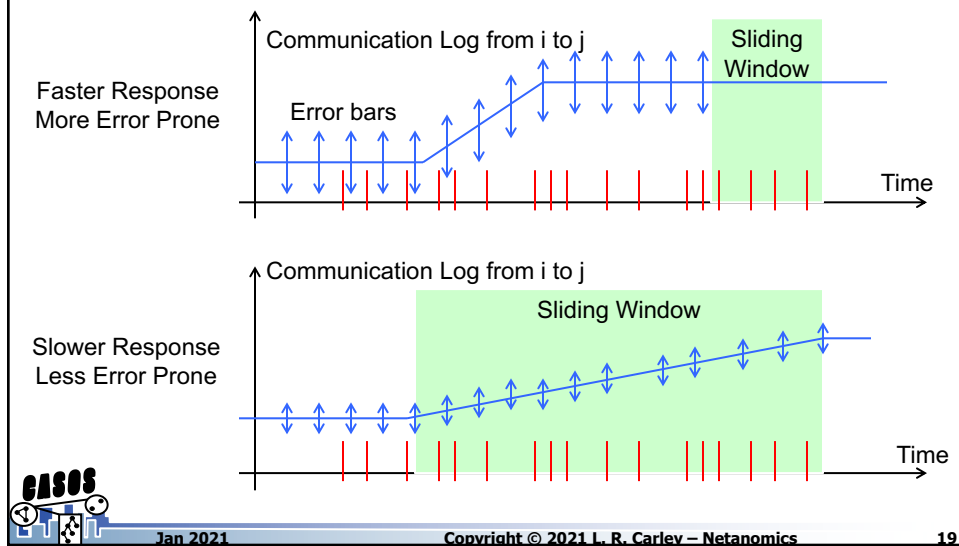
17

Smoothing Effect of Sliding Window



18

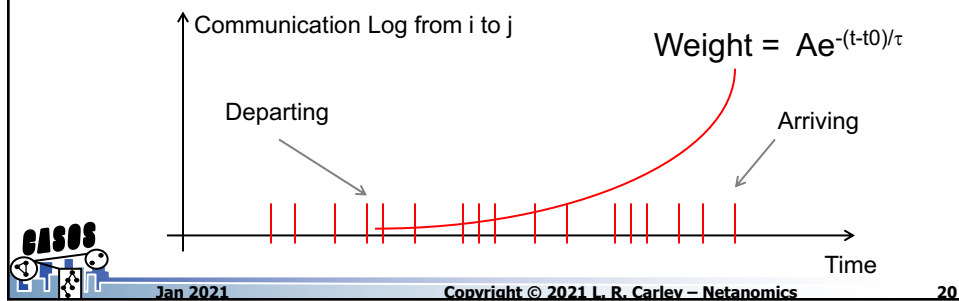
Adjusting Window Size



19

Mathematically Better Window

- Improved tradeoff between smoothing and averaging
 - Mathematically, Exponentially Weighted Moving Average (EWMA)
 - Considers all past known events in estimating current network
 - Old events receive smaller and smaller weighting
 - New events receive highest weighting
 - Exponential time constant – τ – sets how quickly past attenuates vs. how much averaging reduces variance of network

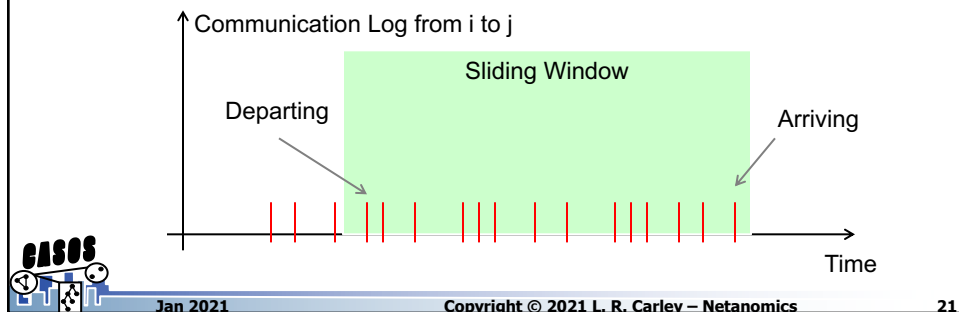


20



Incremental Sliding Window

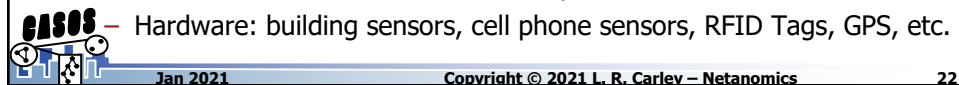
- Sliding Window is Synergistic with Incremental Analysis
 - As window moves forward in time
 - New events “arrive” and must be processed
 - Old events “fall out” of trailing edge of window and must be processed
 - BUT – all of the data in middle of window remains unchanged
 - Incremental algorithms fast because only small part of data changes



21

Communications as a Proxy

- “Ideal approach” – directly sample network each time period
 - E.g., have every member of society fill out survey every time period
 - Limited to very small societies and really motivated subjects
- Or, tracking changes over time using communications data
 - Communication is “proxy” for a network tie
 - Tracking large amounts of communication data gives approximate picture of the underlying social network structure
 - Can use it to find Key Entities and other Network measures
- Communication log data available from many sources
 - Cell Phone Service Providers – call logs, txt msg logs
 - E-mail Data logs – available within organizations
 - Software: Twitter, Facebook, FourSquare, etc.
 - Hardware: building sensors, cell phone sensors, RFID Tags, GPS, etc.



22



Communications Log Data

- Data on who you talk to over monitored means, but NOT what you say (decreased privacy concerns relative to full text monitoring)
- Researchers often only have access to logs from 1 or 2 communications channels – not all possible channels
 - Missing data is substantial
- Communication event is taken as a proxy for a link
 - But this may not always be the case; e.g., calling a wrong #



Jan 2021

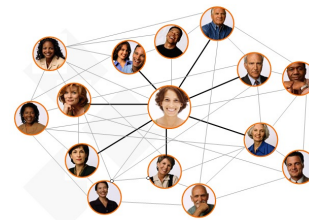
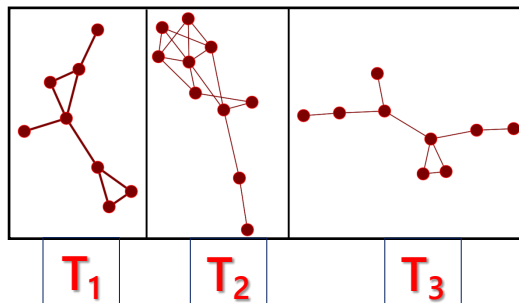
Copyright © 2021 L. R. Carley – Netanomics

23

23

Dynamic Metrics on Over-Time Data

- *Identifying central nodes in a network*



Dynamically Changing Network Structure!!!

Focus on calculating measures for each network
Then evaluate how measures change over time



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

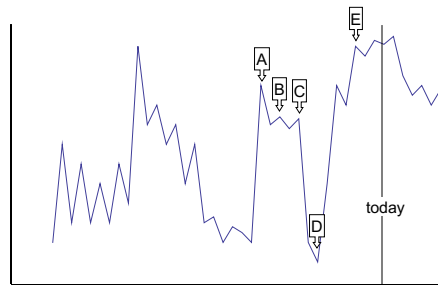
24

24



Changes in Network Data Measures

- Various measures of a network are calculated for a window of network data at a multiple points in time
- Change detection: quickly determine *that* a change occurs.
- Change point identification: *when* did the change occur.



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

25

25

Change Detection

- Goal: Rapidly detect that a change has occurred
- Detect *shocks*, not evolutionary changes
 - Evolutionary change: change due to interaction among actors in a network
 - Example: change of interaction patterns over time among new students as they get to know each other
 - Shock: change reason is exogenous to the network
 - Example: change of interaction patterns among students after they graduate
 - Another way to say it: detect “fast” change not “slow” change
- Another goal is to identify *change point*
 - Likely time when change occurred
 - Limits the scope of explanation for network change



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

26

26



Statistical Process Control (SPC)

- Change detection can be based on SPC
- What is Statistical Process Control?
 - Used in manufacturing to maintain quality control
 - Monitors a process to detect potential changes
 - Calculates a statistic from observed measurements of a process and compares it to a decision interval
 - If the statistic exceeds the decision interval, it is said to “signal”, that a potential change may have occurred
 - A quality engineer will then begin to search for the specific cause of change



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

27

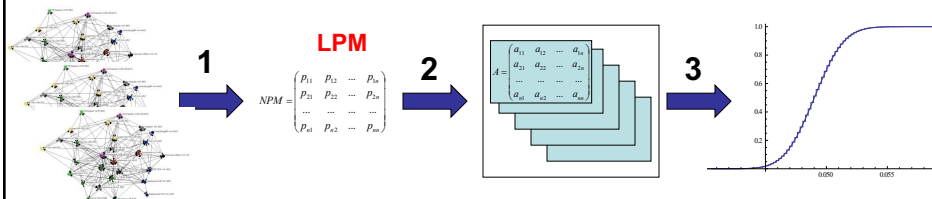
27

Statistical Models of Networks

Link Probability Model (LPM) for Stability

- LPM is a model for a network in *Stability*
- The probability that an email is sent from i to j within some period of time t is:

$$p = \int_0^t f_{ij}(x | \theta_{ij}) dx$$
 - (p , as a function of t , is a CDF: f is the PDF that best fits cell ij in an NPM)
- LPM can be used to simulate stable longitudinal networks



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

28

28



Statistical Models of Networks

Link Probability Model (LPM) for Stability

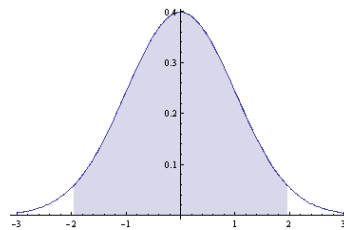
LPM simulated networks are compared to empirical networks and are shown to represent the network well.

| M | δ | N | 60000 | | | |
|----------|----------|----------|----------|--|----------|------|
| e_mean | e_stdev | s_mean | s_stdev | | t-val | p |
| 409.2857 | 38.5604 | 358.0939 | 12.77466 | | 3.754923 | 0.00 |
| 365.8571 | 18.2978 | 320.0974 | 12.7394 | | 7.073195 | 0.00 |
| 365.8571 | 29.04266 | 320.1638 | 12.79331 | | 4.449958 | 0.00 |
| 377.8571 | 38.24669 | 330.6744 | 12.77289 | | 3.489244 | 0.00 |
| 375.2857 | 36.10039 | 328.3765 | 12.79551 | | 3.675254 | 0.00 |
| 349.8571 | 38.15944 | 306.0783 | 12.7845 | | 3.244918 | 0.00 |
| 373.8571 | 48.45076 | 327.0728 | 12.82622 | | 2.731135 | 0.01 |
| 362.4286 | 55.63529 | 317.1509 | 12.77754 | | 2.301849 | 0.02 |



Probability Background

- Consider a normal distribution with $\mu=0$ and $\sigma=1$.
- 95% of the time, observations are between ± 1.9597
- When an observation occurs in the tail, we don't believe it and think that something unusual might be going on.



Statistical Process Control

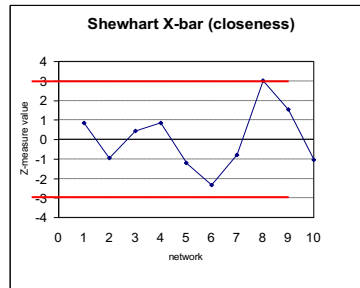
- Manufacturing processes are: stochastic, dependent, non-ergotic, complex, and involve human interaction.
- Shewhart (1927) X-bar Control Chart proposed to monitor change of any process
- Calculate Z_t transform value for each time-period, t .

$$Z_t = (x_t - \mu_0) / \sigma$$

- Calculate a control limit, L , based on risk for false alarm.

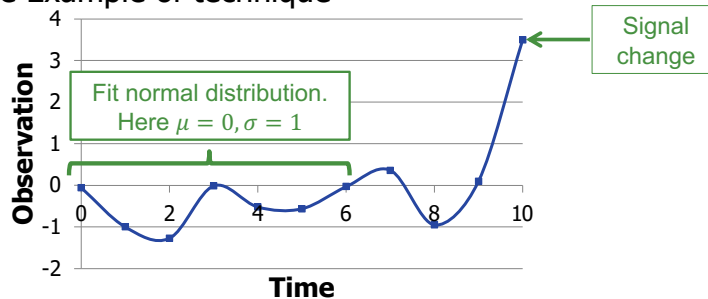
$$\int_L^{\infty} f(x) dx = \alpha$$

- Chart Signals when Z exceeds control limit, L .



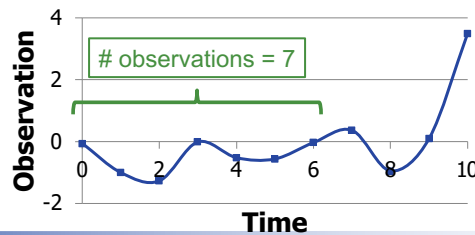
The Shewhart X-Bar Chart

- Overview
 - Fit normal distribution on "control period" (early observations) > assumed to represent the "normal state"
 - Signal change if a subsequent observation is outside confidence interval
- Simple Example of technique



The Shewhart X-Bar Chart

- Parameters
 - # observations used to fit distribution (the "normal" period)
 - False positive risk or decision interval
 - Trade-off between False positive risk & detection speed
- Assumption
 - Observations are normally distributed as independent random vars
 - Shewhart X-Bar chart used even when assumption is violated. However, false positive risk probability may be inaccurate



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

33

33

Statistical Process Control (cont.)

- Newer approaches detect change in fewer observations subject to the same rate of false positives.
 - Scan Statistic (Fisher, 1934)
 - Exponentially Weighted Moving Average (EWMA) (Roberts, 1959)
 - Good at detecting small changes in mean over time
 - Performs well on time series with closely spaced data samples
- $$w_t = \lambda \bar{x}_t + (1 - \lambda)w_{t-1} \quad \mu_0 \pm L\sigma_{\bar{x}} \left(\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2T}] \right)^{1/2}$$
- Cumulative-Sum (CUSUM) Control Chart (Page, 1961)
 - Good at detecting small changes in mean over time
 - Built-in change point detection
 - Two Charts (To Detect Increase and Decrease)



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

34

34



Cumulative Sum (CUSUM)

- Cumulative-Sum Control Chart
 - Good at detecting small changes in mean over time
 - Built-in change point detection
- Calculate Z_t transform for each time-period, t


$$Z_t = (x_t - \mu_0) / \sigma$$

- Two Charts (To Detect Increase and Decrease)

$$C_t^+ = \max \left\{ 0, Z_t - \frac{\delta}{2} + C_{t-1}^+ \right\}$$

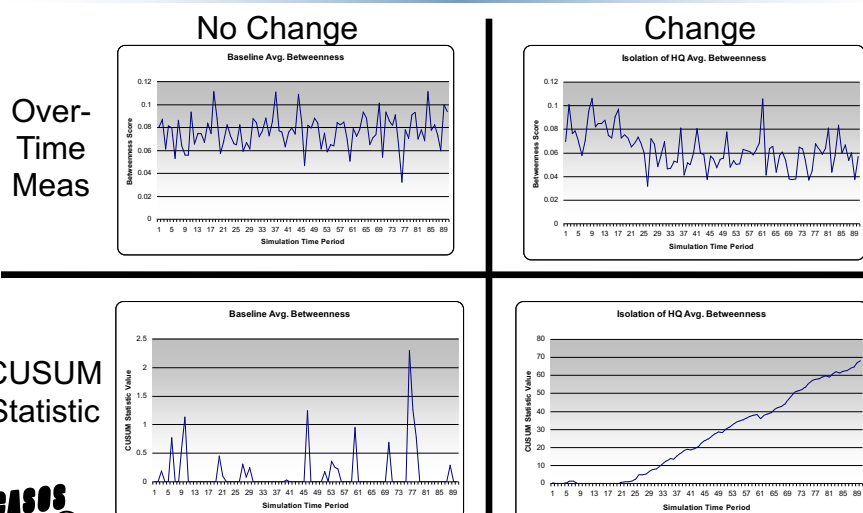
- Chart Signals when C^+ or C^- statistic exceeds decision interval

$$C_t^- = \max \left\{ 0, -Z_t - \frac{\delta}{2} + C_{t-1}^- \right\}$$

CASOS Sensitivity in CUSUM due to discrete integration of error
 Jan 2021 Copyright © 2021 L. R. Carley – Netanomics 35

35

Comparison of Change Detection Approaches



CASOS Jan 2021 Copyright © 2021 L. R. Carley – Netanomics 36

36



Comparison of Change Detection Approaches

| | CUSUM $k = 0.5$ | EWMA $r = 0.1$ | EWMA $r = 0.2$ | EWMA $r = 0.3$ | Scan Statistic |
|-------------------------|--------------------|-------------------|-------------------|-------------------|-------------------|
| Average Betweenness | 9.32 | 8.24 | 10.16 | 11.52 | 6.76 |
| Maximum Betweenness | 14.36 | 14.72 | 15.72 | 17.08 | 13.24 |
| Std Dev. Betweenness | 16.44 | 16.24 | 16.92 | 18.52 | 15.24 |
| Average Closeness | 10.68 | 9.08 | 13.60 | 17.52 | 10.48 |
| Maximum Closeness | 8.76 | 6.00 | 10.60 | 37.96 | 8.64 |
| Std Deviation Closeness | 34.48 | 34.72 | 34.52 | 35.68 | 27.08 |
| Average Eigenvector | 31.28 | 31.28 | 31.28 | 31.28 | 24.00 |
| Minimum Eigenvector | 14.36 | 14.36 | 14.28 | 15.56 | 14.88 |
| Maximum Eigenvector | 5.24 | 5.40 | 5.80 | 7.52 | 4.00 |
| Std. Dev Eigenvector | 5.92 | 4.88 | 6.40 | 6.96 | 3.64 |



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

37

37

Network Change Detection: Analysis of Real World Data

| | # Nodes | Time Periods | Method of Collection | Type of Relation | Design | Known Change |
|------------|---------|-----------------|-------------------------|---------------------|--------|-----------------|
| Fraternity | 17 | 15 | Survey | Ranking | Fixed | Yes |
| Leav 07 | 68 | 8 | Survey | Rating | Free | Yes |
| Leav 05 | 158 | 9 | Survey | Rating | Free | None |
| Al-Qaeda | 62-260 | 17 | Text | Rating | Free | Yes |
| Winter C | 22 | 9 | Observation & Survey | Rating | Fixed | Yes |
| Winter A | 28 | 9 | Observation & Survey | Rating | Fixed | Yes |
| IkeNet 2 | 22 | 46 | Email | Count Msg | Free | Yes |
| IkeNet 3 | 68 | 121 | Email | Count Msg | Free | Yes |



Jan 2021

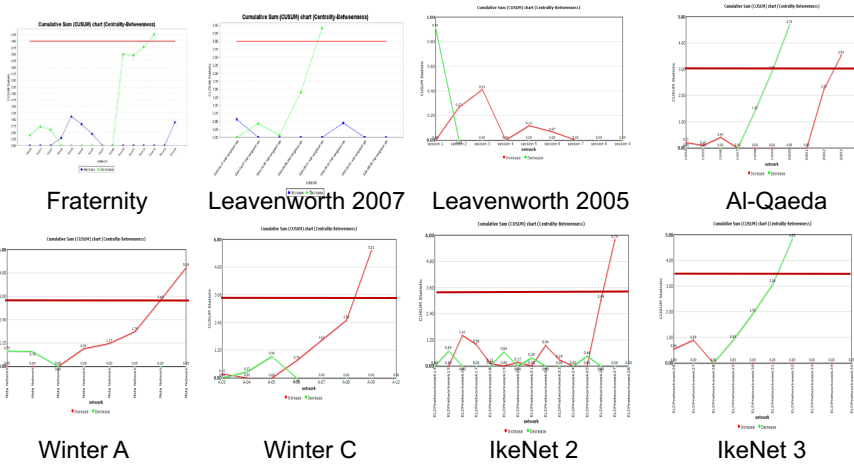
Copyright © 2021 L. R. Carlev – Netanomics

38

38



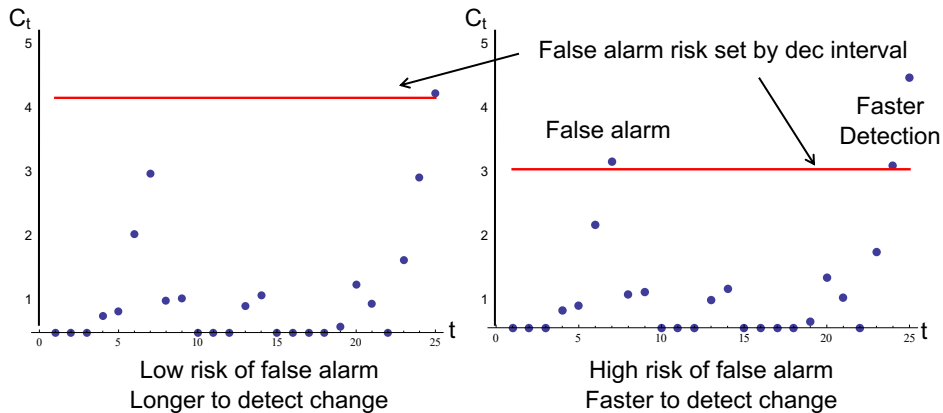
Network Change Detection: Analysis of Real World Data



39

Summary of Change Detection Across Data Sets

There is a trade-off between false positive and rapid detection



40



Summary of Change Detection Across Data Sets

Too little risk may prevent change detection all together

| Data | Change | $\alpha = 0.05$ | $\alpha = 0.02$ | $\alpha = 0.01$ | $\alpha = 0.005$ | $\alpha = 0.001$ |
|------------|--------|-----------------|-----------------|-----------------|------------------|------------------|
| Fraternity | 8 | 10 | 10 | 10 | 13 | Never |
| Leav 07 | 3 | 5 | 5 | 5 | Never | Never |
| Leav 05 | None | No F.A. | No F.A. | No F.A. | No F.A. | No F.A. |
| Al-Qaeda | 1997 | 1999 | 1999 | 2000 | 2000 | Never |
| Winter C | May | Sept | Sept | Oct | Oct | Never |
| Winter A | May | Aug | Sept | Sept | Sept | Oct |
| IkeNet 2 | 25 | 26 | 26 | 27 | 27 | 27 |
| IkeNet 3 | 14 | 15 | 18 | 19 | 19 | 20 |



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

41

41

Change Detection Hands-On

- Based on Roger Federer 2010 data

ORA 3.0.9.9.71

File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Meta-Network Manager

Dynamic-Meta-Network: Roger Federer 2010

Name: Roger Federer 2010

Filename: /Users/lrc/Dropbox/DATA_LRC_SI-2018/Roger Federer 2010-Dates Fixed.xml

Generate Reports... Visualize Measure Charts... View Trails...

Keyframes and deltas time stamped by: Date Period

Statistics:

Keyframe count: 12

Delta count: 0

Earliest date: 2010-01-01 00:00:00

Latest date: 2010-12-01 00:00:00



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

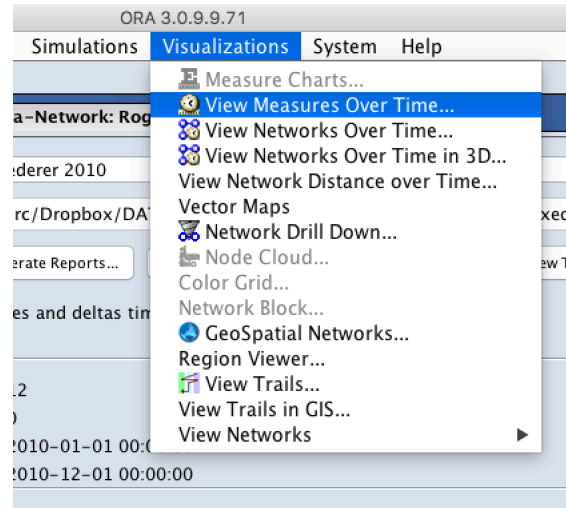
42

42



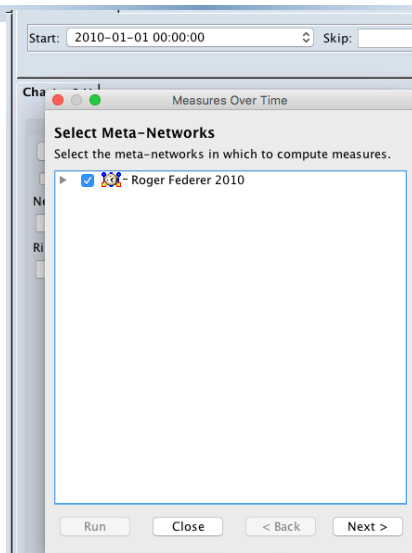
Change Detection Hands-On

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



43

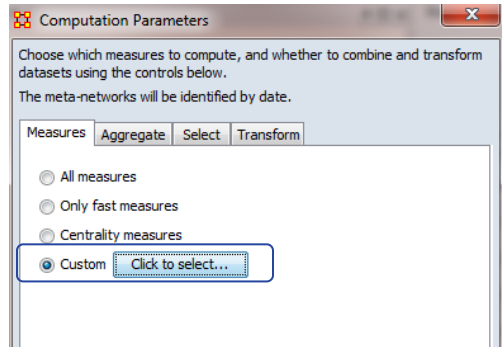
Select The Metanetwork



44



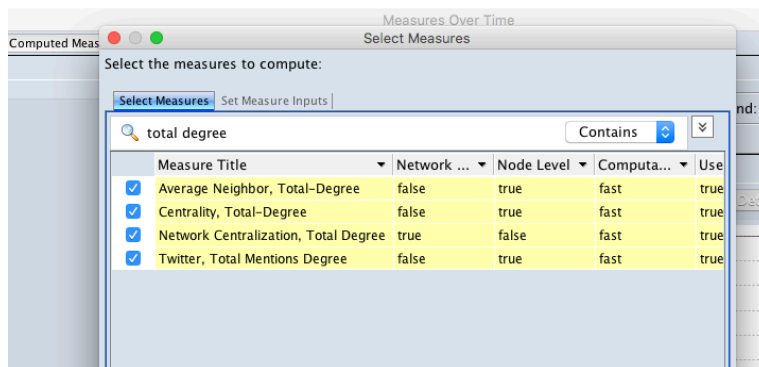
Custom Measure Selection



45

Use Search to Find Measure

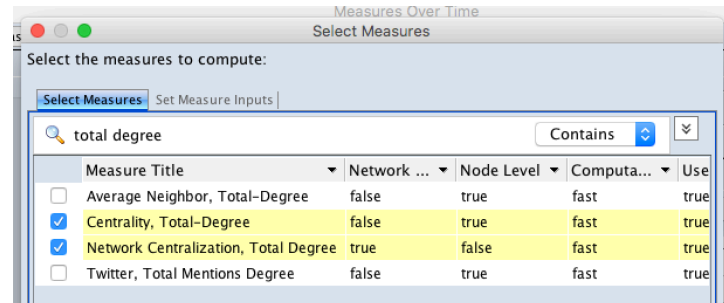
Hint: Click Select Box at bottom to deselect all measure,
Then use search to find the ones you want



46



Two measures selected to run



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



Jan 2021

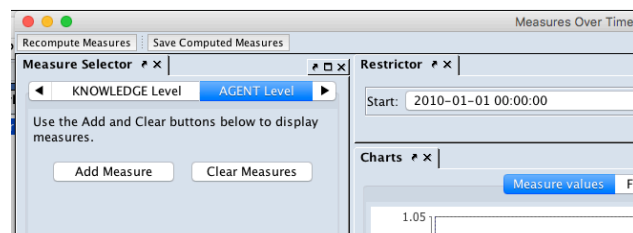
Copyright © 2021 L. R. Carlev – Netanomics

47

47

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



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

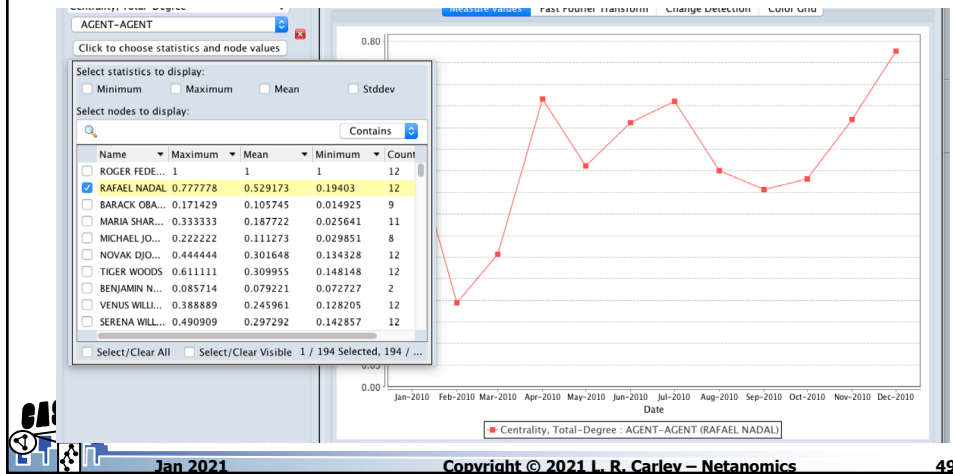
48

48



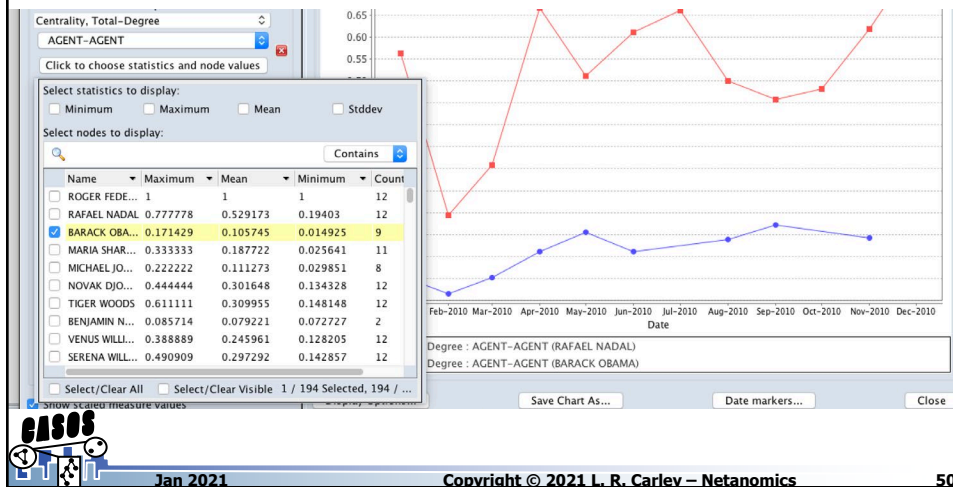
Select Agent for Measure

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

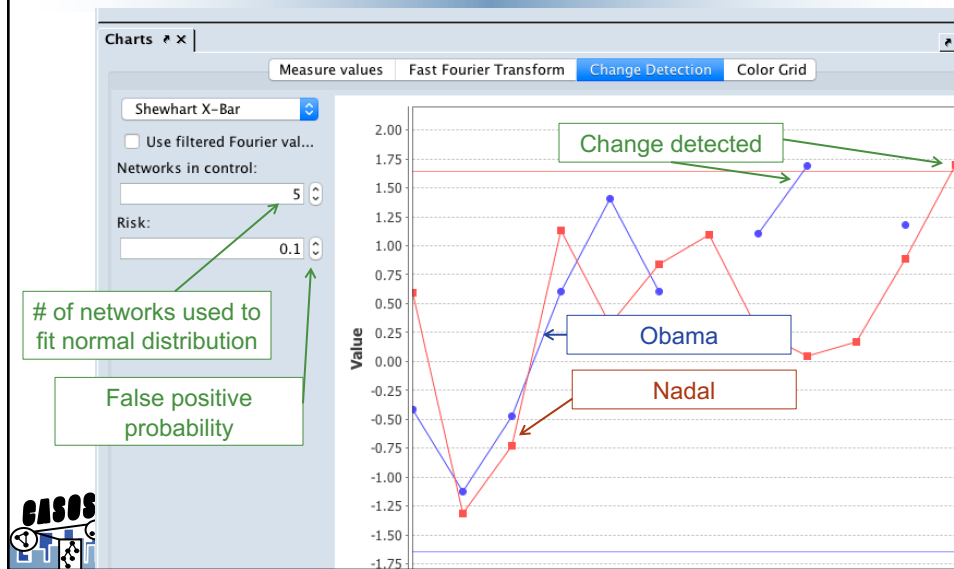


Add Another Agent Measure Plot

Click on Add a Measure button again to add another line

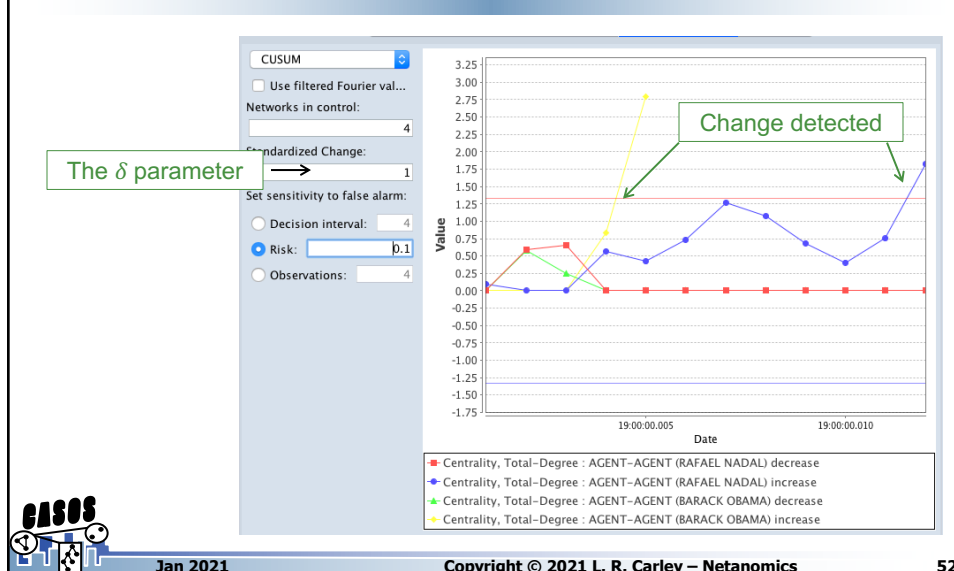


Change Detection Hands-On The Shewhart X-Bar Chart



51

Change Detection Hands-On CUSUM Method



52



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



Jan 2021

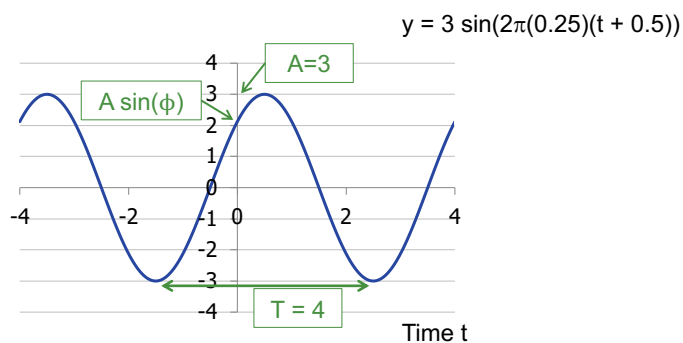
Copyright © 2021 L. R. Carlev – Netanomics

53

53

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



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

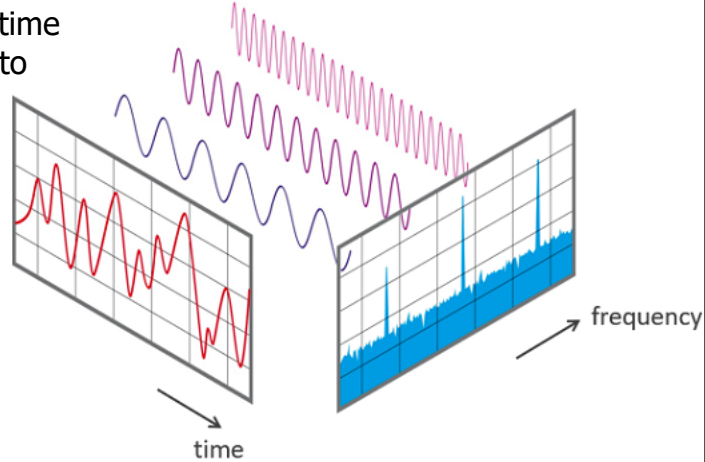
54

54



Frequency Domain: The FFT

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



Jan 2021

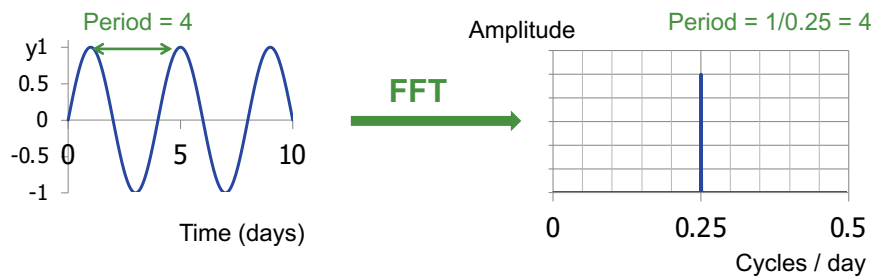
Copyright © 2021 L. R. Carlev – Netanomics

55

55

FFT Example: Sum of Sinusoids

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



Jan 2021

Copyright © 2021 L. R. Carlev – Netanomics

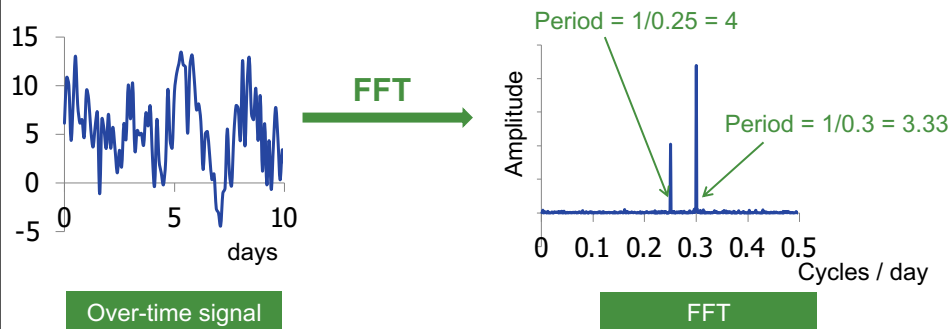
56

56



FFT Example 2

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



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



Jan 2021

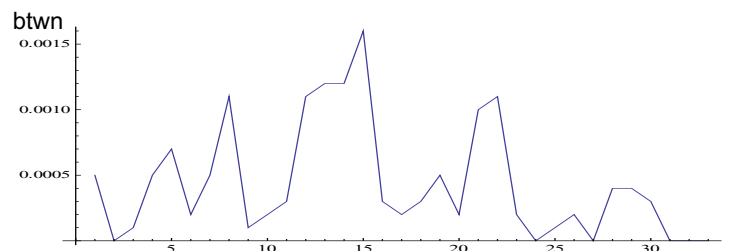
Copyright © 2021 L. R. Carley – Netanomics

57

57

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.



Jan 2021

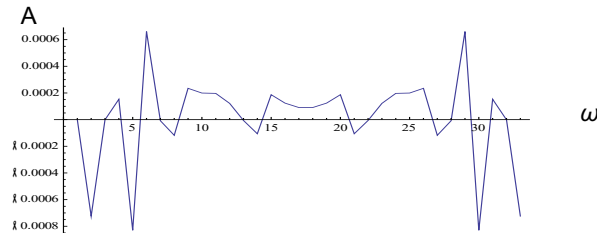
Copyright © 2021 L. R. Carley – Netanomics

58

58



Fourier Analysis – Example 3

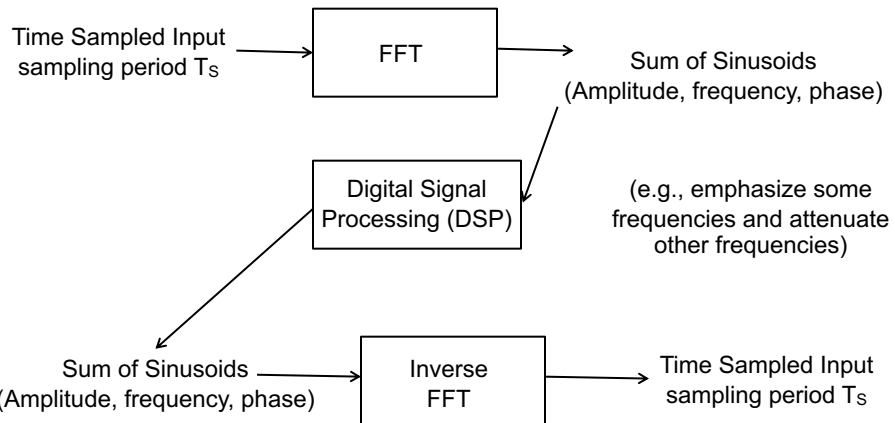


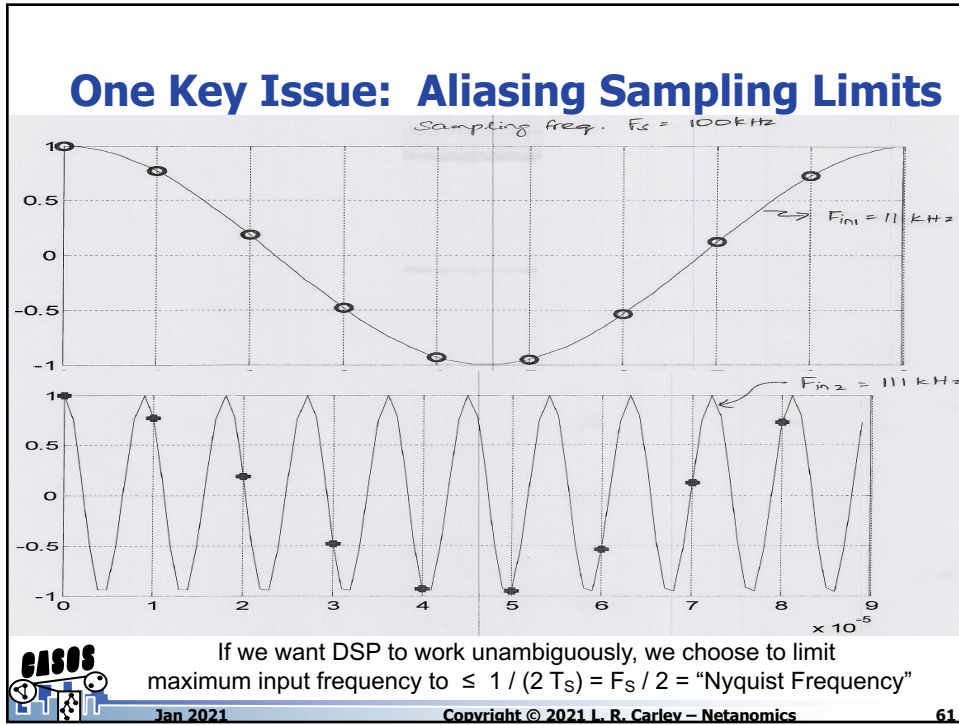
Fourier transform
Symmetric around the midpoint
3 main components (in terms of magnitude)

That is why we typically only display from origin up to midpoint

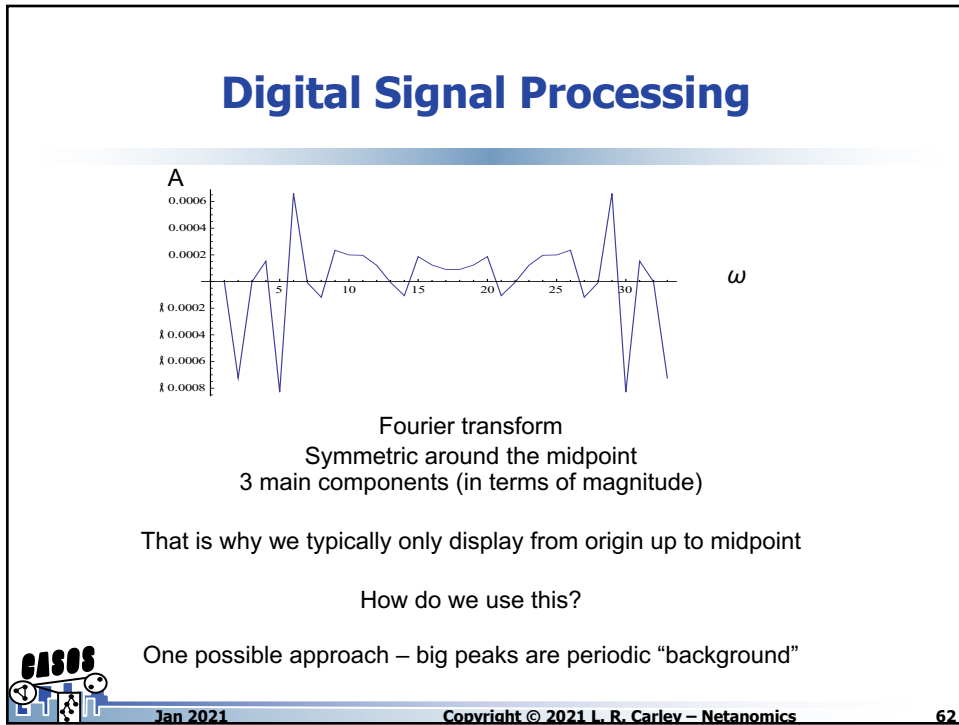


Digital Signal Processing





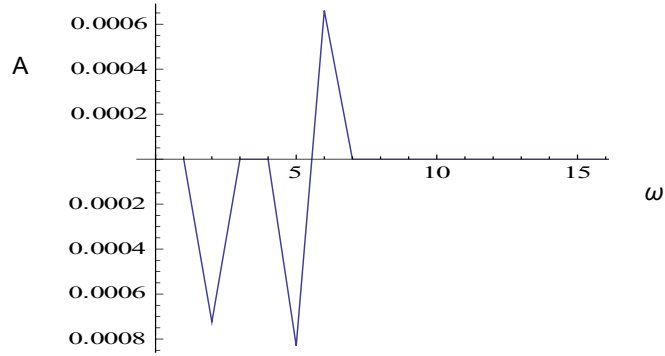
61



62



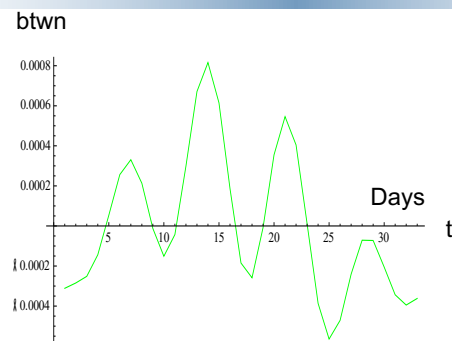
Filtering



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



Inverse Fast Fourier Transform

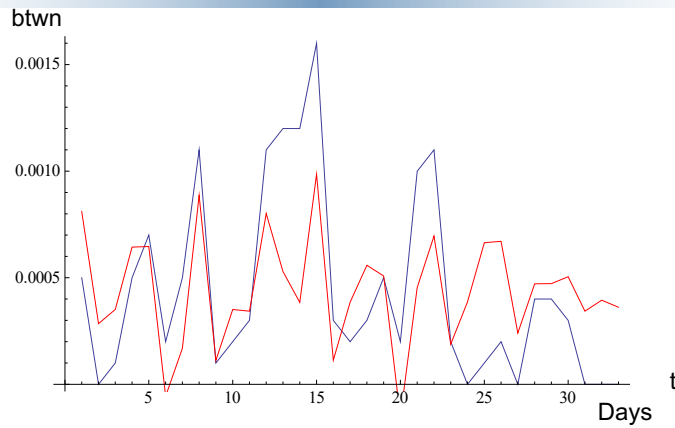


This is the inverse Fourier transform of just the 3
selected components, which are then reconverted to time waveform

There is a weekly, two week and three week cycle



Anomaly Detection



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



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

65

65

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



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

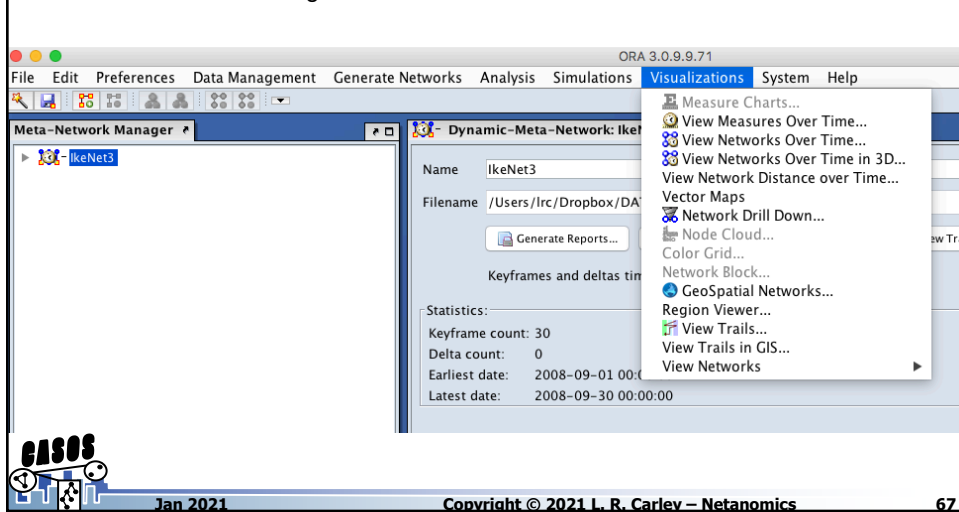
66

66



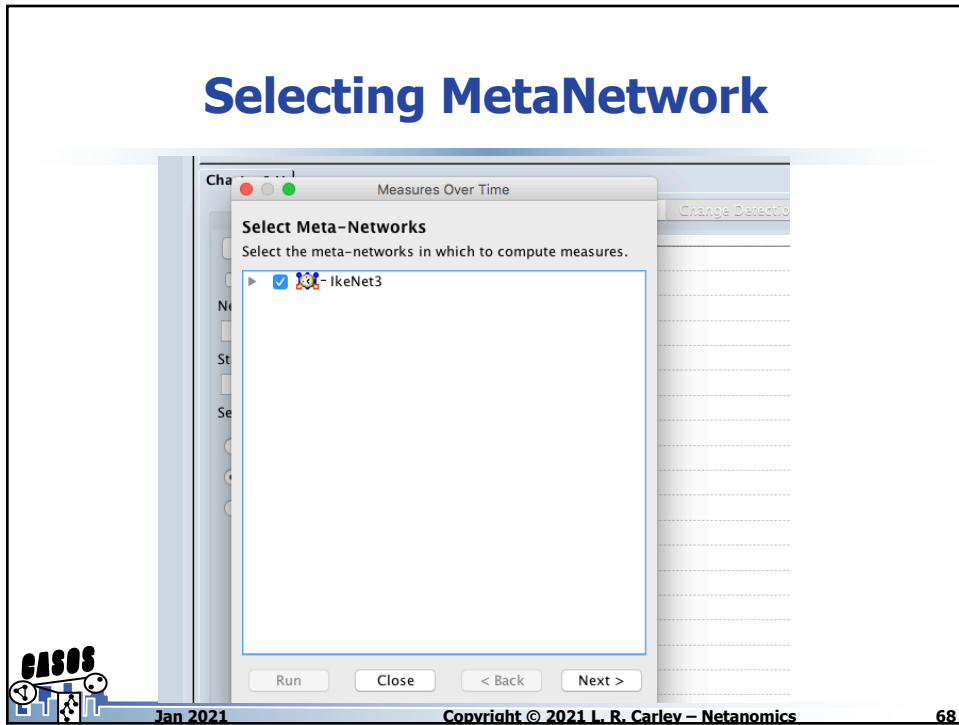
Over Time Measures

Again View Measures Over Time



67

Selecting MetaNetwork



68



Selecting Measures

Select Parameters
Select the measure calculation parameters.

Choose which measures to compute, and whether to combine and transform datasets using the controls below. The meta-networks will be identified by date.

Measures Nodesets and Networks ▶

All measures
 Only fast measures
 Centrality measures
 Custom Click to select...

Geodesic measure options:

Compute regular measures
 Compute inverted measures
 Compute k-centrality measures with radius:

Run Close < Back Next >

Jan 2021 Copyright © 2021 L. R. Carley – Netanomics 69

69

Selecting Measures to Plot

Density, Weighted & eagent2eagent

Use the Add and Clear buttons below to display measures.

Select a measure and input:
Density, Weighted
eAgent2eAgent

Add Measure Clear Measures

Charts * x | Measure values Fast Fourier Transform Change Detection Color Grid

Value

Date

Density, Weighted : eAgent2eAgent

Jan 2021 Copyright © 2021 L. R. Carley – Netanomics 70

70

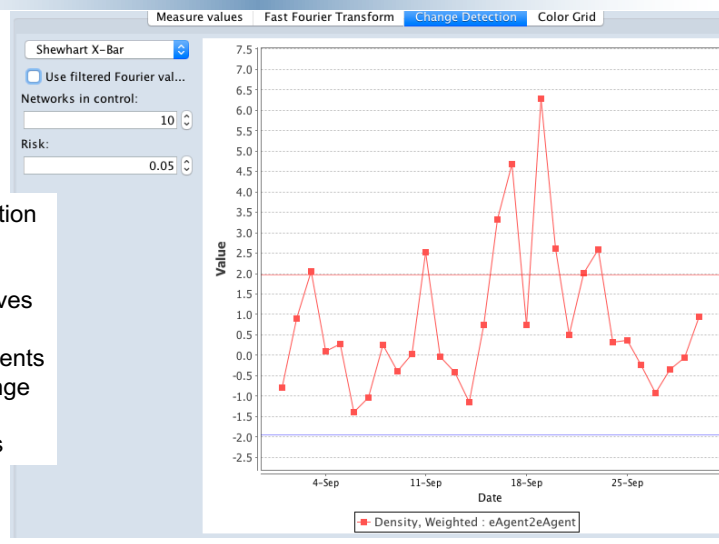


Run Shewhart to Detect Change

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

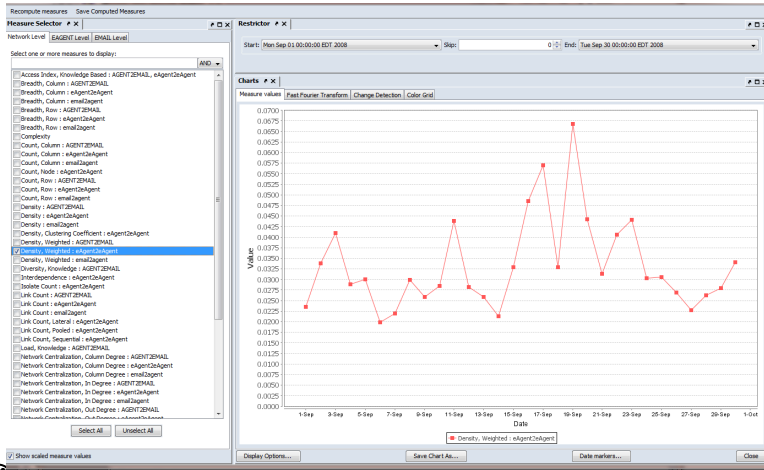
Lots of apparent events
 That could be Change

All upward changes



FFT Can Help with Periodic Patterns

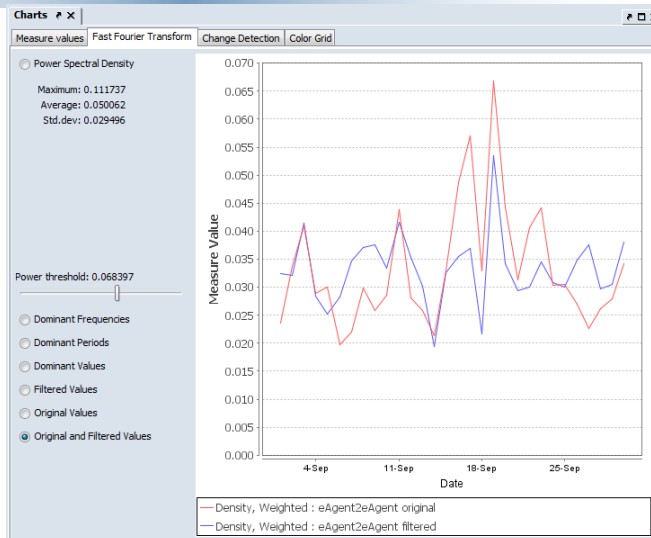
Network Level / Density, Weighted / eAgent2eAgent



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



75

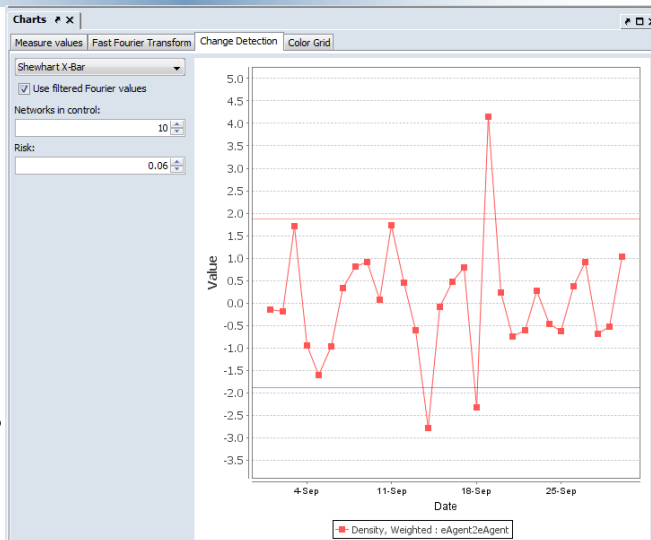
Shewhart on FFT subtracted data

Now use the Filtered Data

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

Only 1 + event and
2 - events

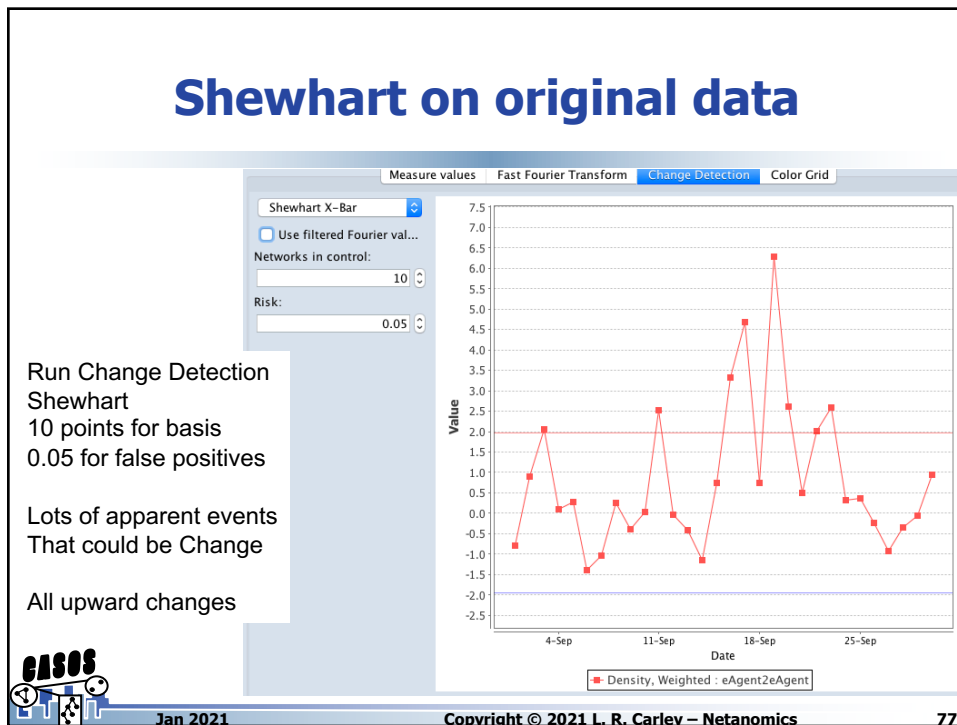
Open Question – is this better than original or not?



76



Shewhart on original data



77

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.



78



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 $\#T(\log(\#T))$



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

79

79

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.



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

80

80



Summary

- 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.



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

81

81

Questions



Jan 2021

Copyright © 2021 L. R. Carley – Netanomics

82

82

