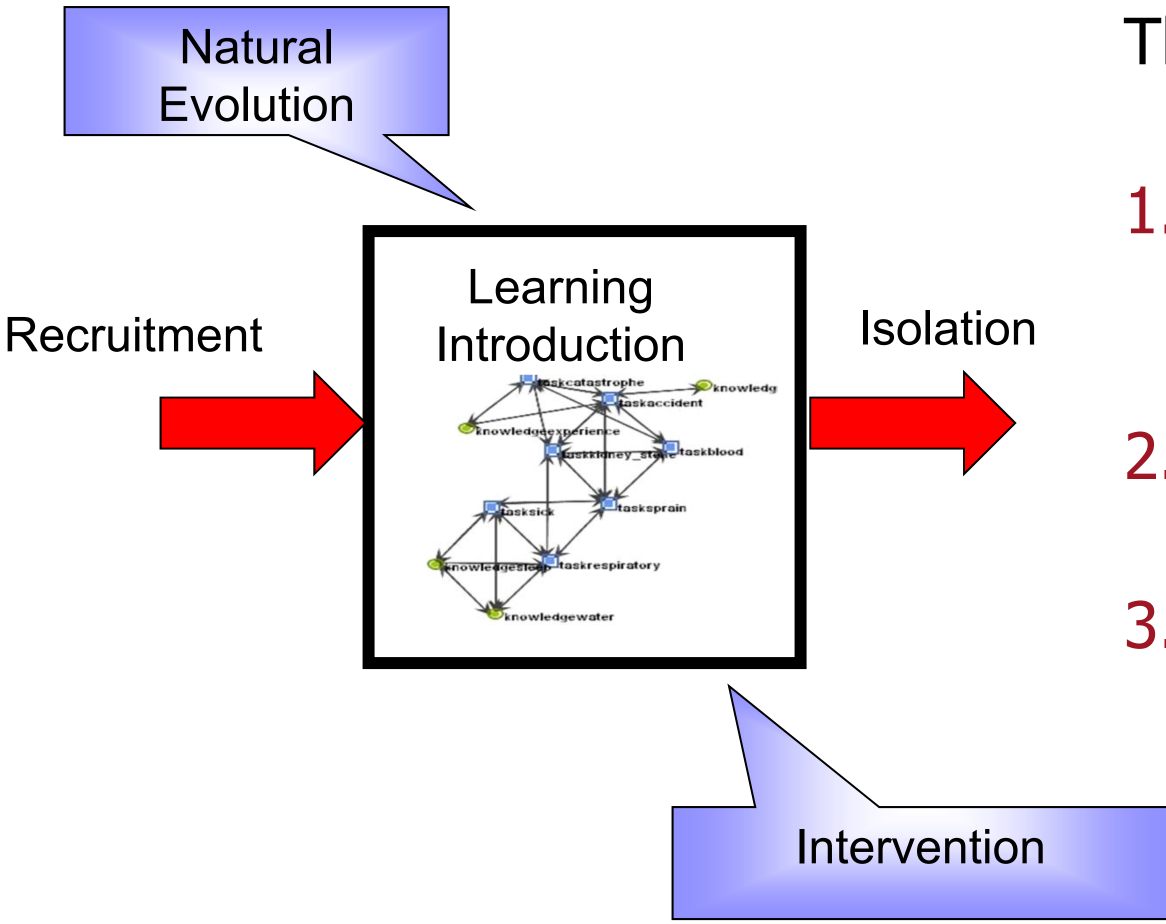
Dynamics: Reasoning About Networks Over Time

Prof. L. Richard Carley

carley@ece.cmu.edu

Change in Networks

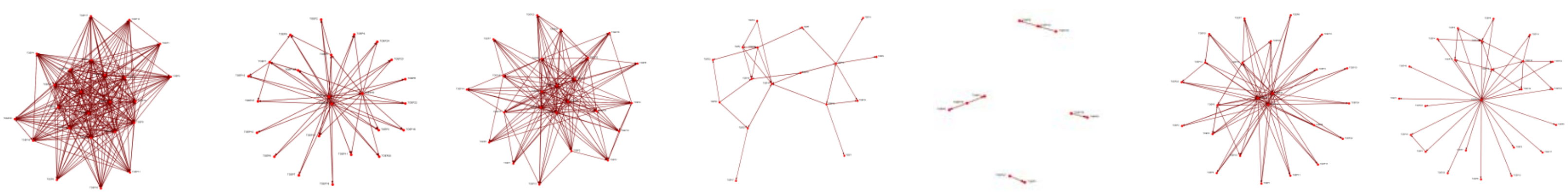


Three Ways to Study Dynamics

1. Comparison over time
   - Look at real data
2. Immediate Impact
   - Comparative statics
3. Near Term Impact
   - Utilizes simulation - DyNet

Longitudinal Networks

- Consider a communication network, such as email.



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

Basic Issue

- 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
- Not a right answer
  - It depends on what you want to know
  - Often try two different approaches and see how much they differ

Types of Change in Network Data

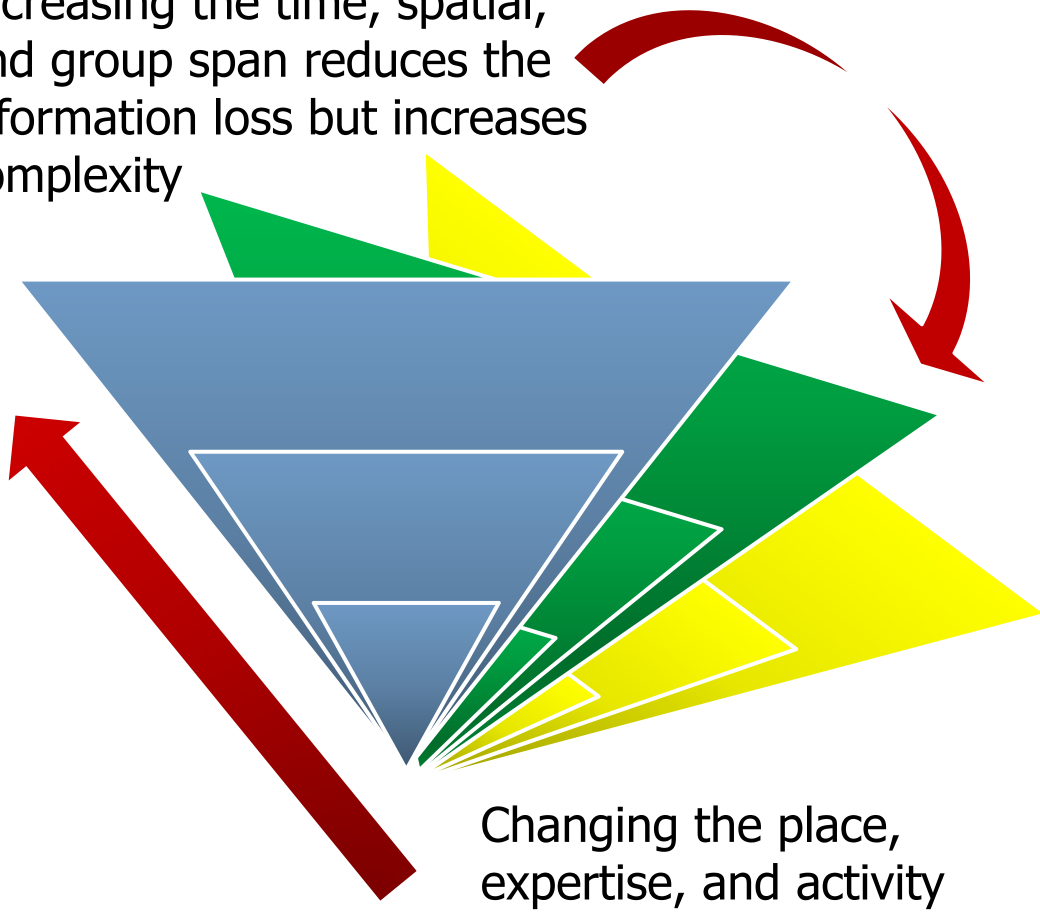
- Stability: Relationships remain the same over time.
- Evolution: Interaction among agents cause the relationships to change over time.
- Shock: Change is exogenous to the social group.
- Mutation: A shock stimulates evolutionary behavior.

Models Used to Study Change

- Stability: LPM , ERGM, repeated measures
  - LPM is Link Probability Model
  - ERGM is Exponential Random Graph Model
- 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

Dimensions of Relevance

Increasing the time, spatial, and group span reduces the information loss but increases complexity



Changing the place, expertise, and activity alters the structure of the network

First: Determine what occurs

- Place
- Expertise
- Activity

Impacts analysis

- Measurement effects
- Interaction
- Adaptation

Social Networks are Continuously Emerging Structures

- Networks emerge from intersecting trails
  - Constrained and enabled
- Networks reinforce some trails
  - Secondary emphasis to some constraints
- Slices across trails is the “measured” or “observed” social network
- The level of aggregation determines the “width” of the slice
  - The greater the width – the higher the density

Aspects of Trails of Interest

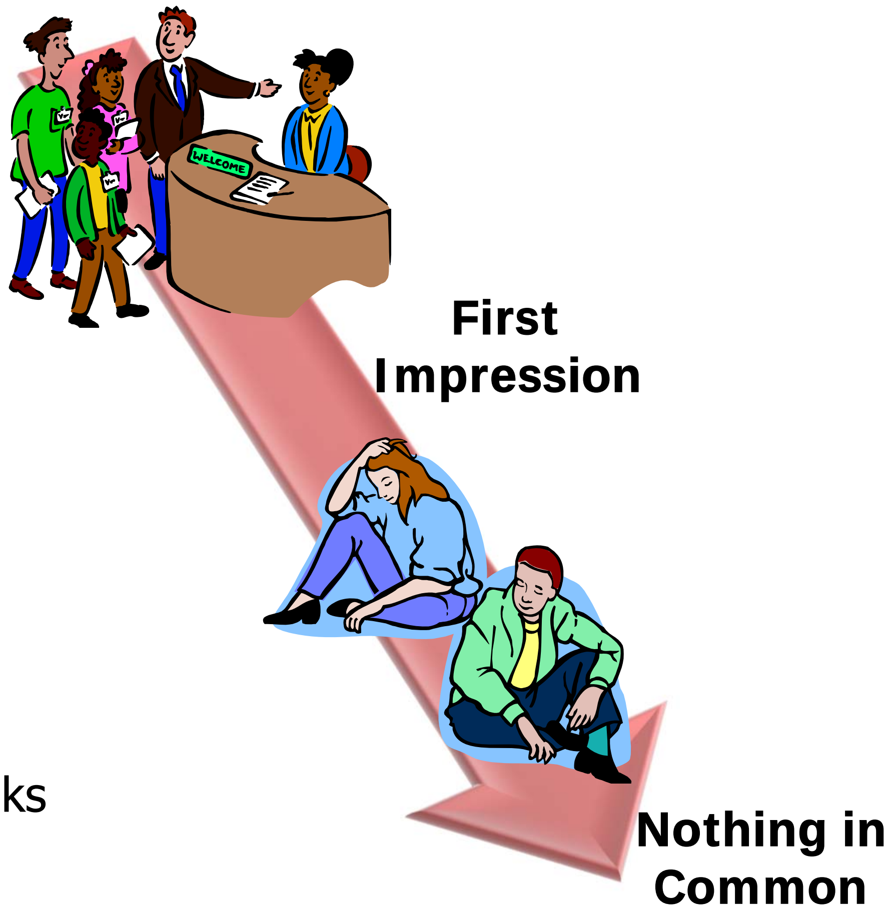
- PLACE – Physical
  - Who was where when
  - doing what (how (to whom (why)))
- EXPERTISE – Cognitive
  - Who was providing what information when
  - how (to whom (from where (why)))
- ACTIVITY – Occupation
  - Who was doing what when
  - how (with whom (where (why)))

Trails Provide Meta-Network Information

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Social Dynamics due to Learning

- Implicit link
  - seen together
  - common sources
  - seniority
- Explicit link
  - information exchange
  - learned from each other
  - mentoring



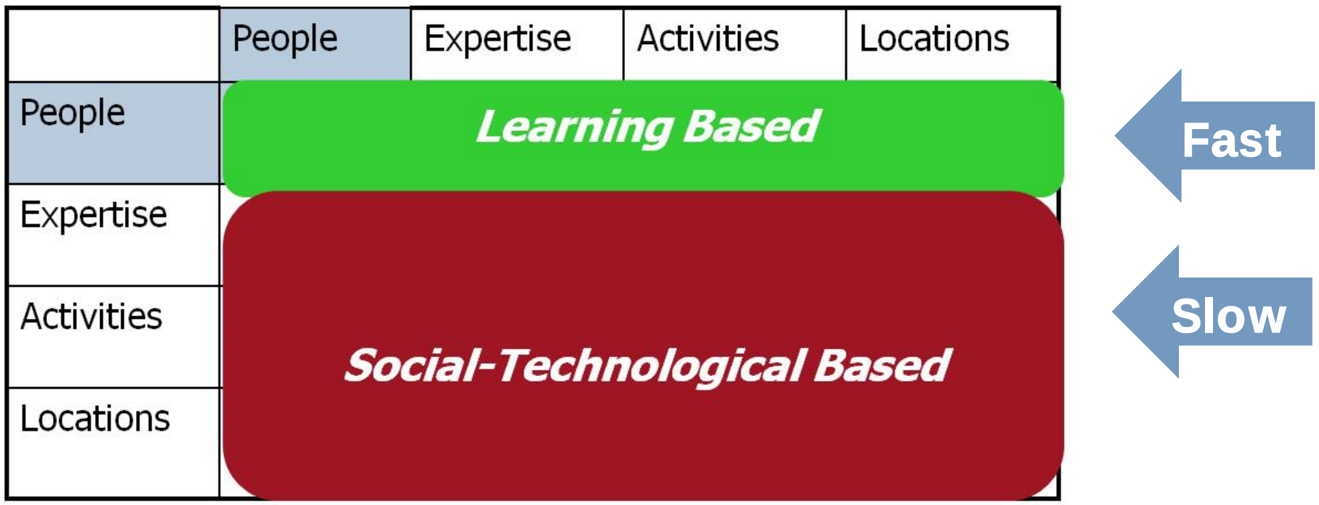
- When meeting a new person
  - Infer expertise based on implicit links
  - Baseline for trust
  - Social shakeout occurs as you move from implicit to explicit links

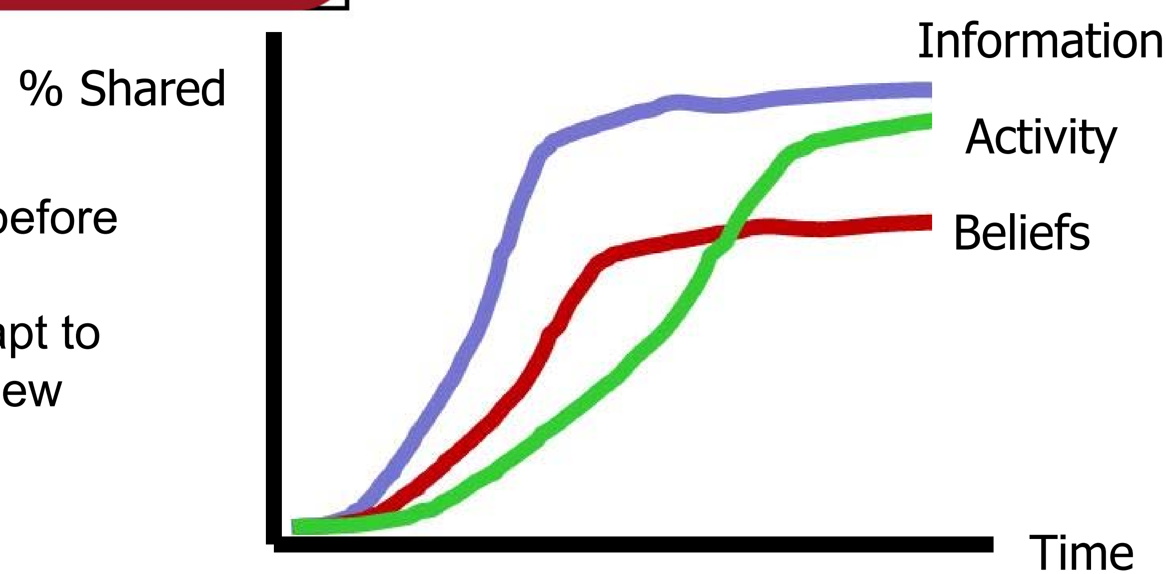
CASOS

Dynamics

- Networks emerge, adapt, etc.
- Two sources of dynamics
  - Adaptation - Actual
  - Activation - Perceived
- Adaptation
  - Networks change as people and their environments change
  - Multiple logics
- Activation
  - Networks appear different because trails are constrained
  - Multiple constraints

Rates of Adaptation





You can change behavior before changing beliefs
At group level, people adapt to technology faster than new technology emerges

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)

Topic 1: What if Key Actors are Removed?

Critical Personnel

- Individual whose absence will dramatically decrease performance of organization
  - Only person who can do a task
  - Only person with certain organizationally critical knowledge
  - Person who keeps others in line, supported, feeling good about the organization
  - Person who is the only access point to certain organizationally critical knowledge
  - Only person who knows key people
  - Person who knows almost everything
- Examples
  - Lead scientist
  - Life time administrative assistant
  - One person lab/technician
  - Lone visionary

Critical Personnel

- Key players, network elite
- Those with power
- Those who, were they to leave, would reduce the organizations performance, adaptability, competence …
- Direct identifiers
  - The centralities: e.g., degree
  - The exclusivities: e.g., task
  - The integrators: e.g., simmelian ties
  - The loads: e.g., workload and cognitive demand
- Indirect
  - Those who have access to, can influence, those who are critical

Immediate Impact - Prediction

- *What if ? Remove top 5 emergent leaders*
- Change in performance – many possible measures
  - e.g., change in information diffusion rate
    - Anticipated increase – 67% percentage difference
  - New emergent leaders
    1. 0.0174 said_mortazavi
    2. 0.0137 kamal_kharazi
    3. 0.0127 reza_asefi
    4. 0.0120 morteza_sarmadi
    5. 0.0100 hashemi_shahroudi
  - Value of "lowest" old emergent leader was .0246
- But this is only a very simple view of dynamic networks
  - Only "time" effect is removal of one or more agents or links

Networks Heal Themselves

- BUT, Organizations adapt to the loss of agents
  - e.g., Manager losing admin will "borrow" support from other admins in org.
- Networks, particularly cellular networks, can withstand high levels of turnover
- Agents the are in structurally "equivalent positions" are replaceable by others that are "equivalent"
  - Connected to same others
- Agents in specialized positions, e.g., those with high cognitive load, are harder to replace and will cause longer lasting performance drop
- Newcomers typically enter as neither structurally equivalent with a key actor nor high in cognitive load
  - Low transactive memory
  - Few pre-existing ties
  - "start off on simple tasks"

What Happens When Critical Personnel are Removed

- Decrease redundancy
- Decrease or remove intellectual property
- Alter performance
- Alter adaptability
- Alter information diffusion

How to be less vulnerable and more adaptive

- Increase redundancy
  - Decrease number of tasks – outsource
  - Decrease number of skills/resources/knowledge – do simpler tasks, employ skill reduction technologies
  - Increase number of personnel
  - More highly trained personnel (each knows more)
  - Increase workload
  - Redistribute workload – retask individuals
  - Redistribute resources – retrain individuals
  - Redistribute employees – reassignment

Topic 2: Over Time Networks - Setup

Basic Issue

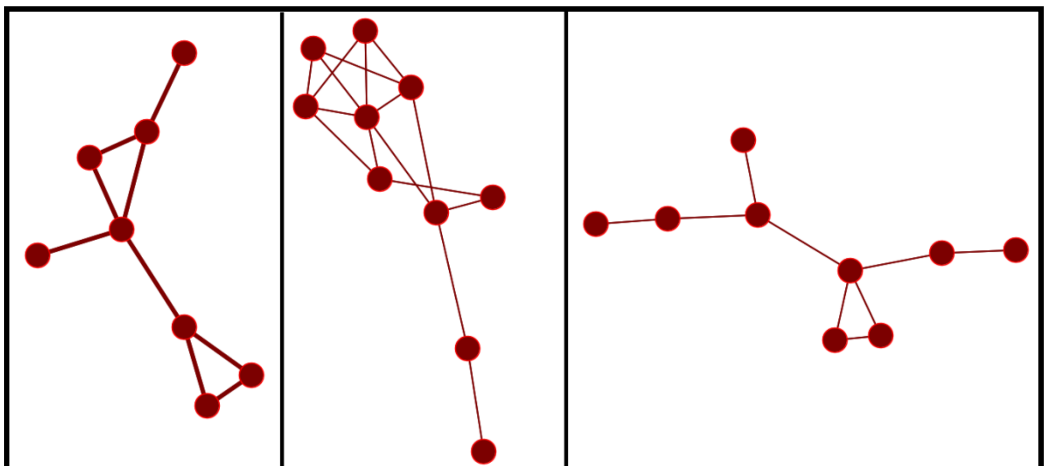
- Real Social Networks are not time independent
  - They are NOT "stationary" processes in the statistical sense
- Over time the set of nodes change
  - Agents die, agents are born
  - If data set has limited geographic focus,
    - Agents can enter region under study
    - Agents can leave region under study
- Network connections between agents can change
  - A network link between two agents can disappear
    - Two family members have a fight and refuse to talk to each other
  - A new network link can be created
    - People meet new people and form new relationships
    - Advertising campaigns can convince people to follow companies

Types of Changes in Network Data

- Stability: Relationships remain statistically the same over time
  - If you are a signal processing person, the process is "stationary" and the Network is "Ergodic"
- Evolution: Interaction among agents cause the relationships to change gradually over time
  - All link weights / costs are evolving over time during observations
- 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

Dynamic Metrics on Over-Time Data

- *Identifying central nodes in a network*



T_1 T_2 T_3

Dynamically Changing Network Structure!!!

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

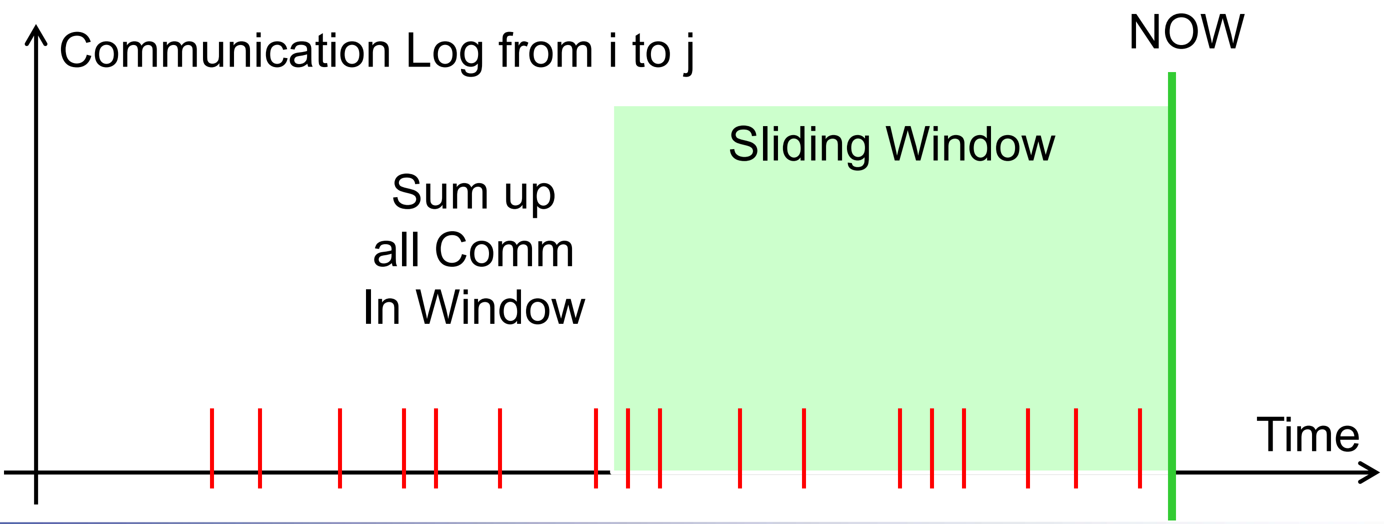
Slicing and Dicing

- 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

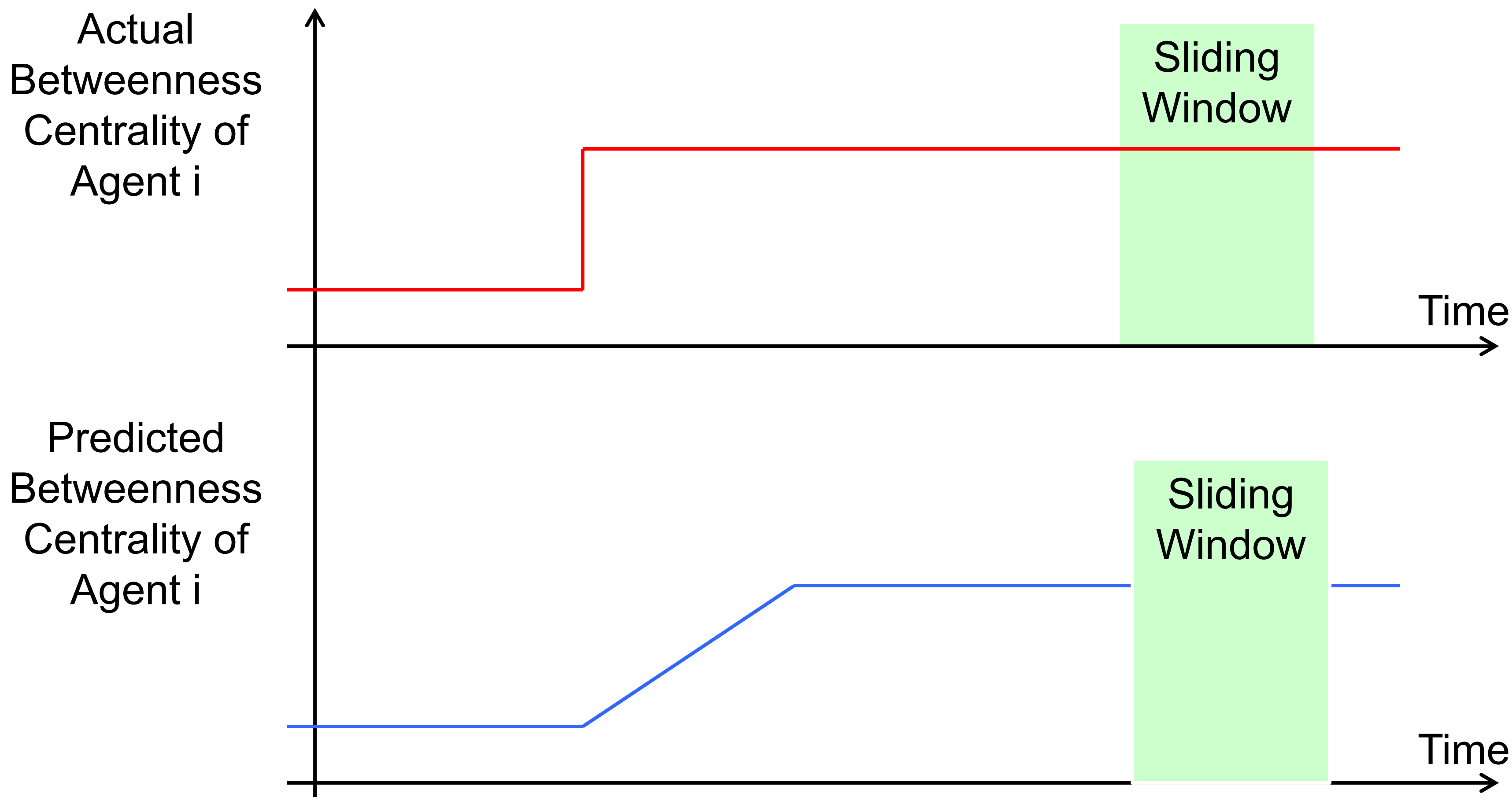
CASOS

Sliding Window for Over-Time Links

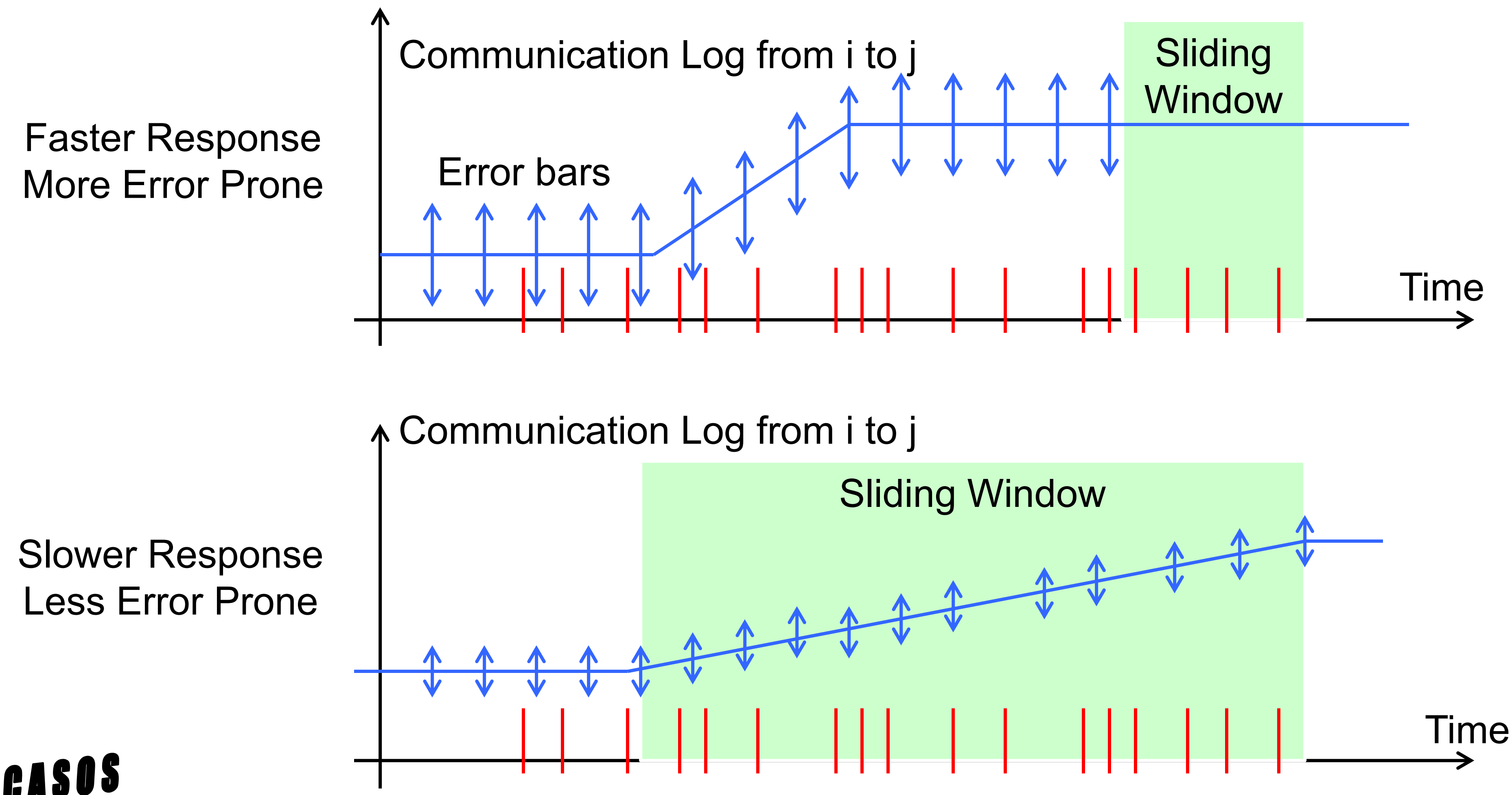
- Estimator for Link Weight (a.k.a. Link Cost)
  - Add up # of Communication Events between x & y in window
  - Take reciprocal. If # is 0, there is no Link between that pair
  - Then move window forward by a time step and repeat
  - Alternatives possible:
    - Incorporate duration of communication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



Smoothing Effect of Sliding Window

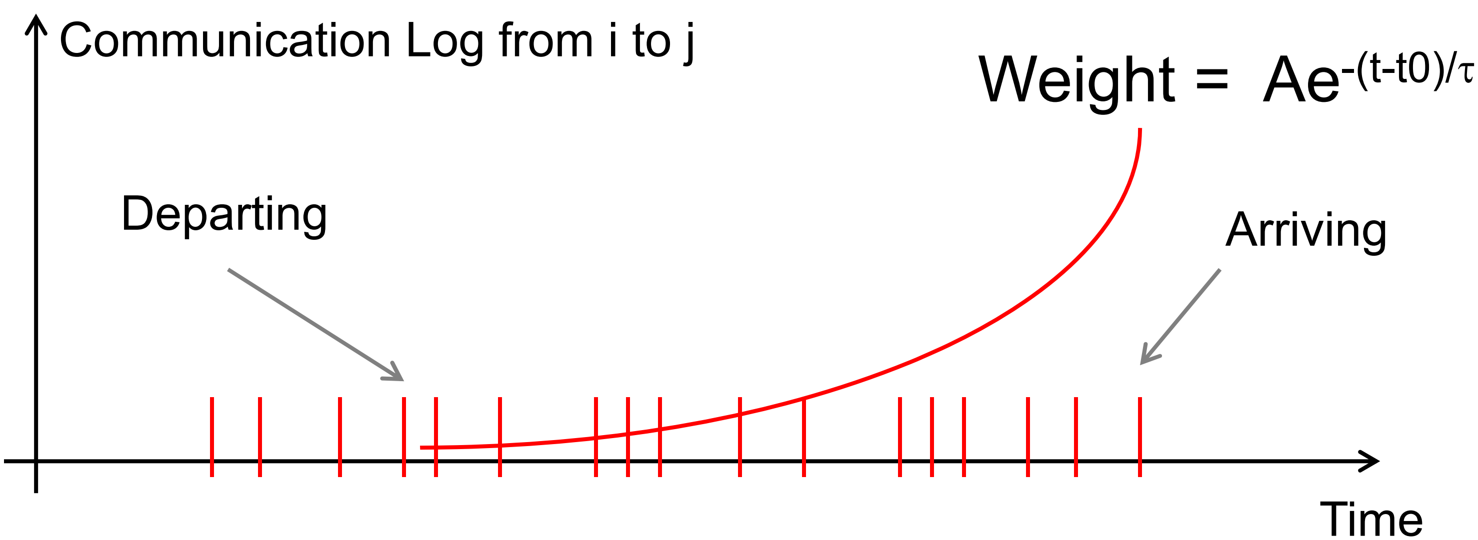


Adjusting Window Size



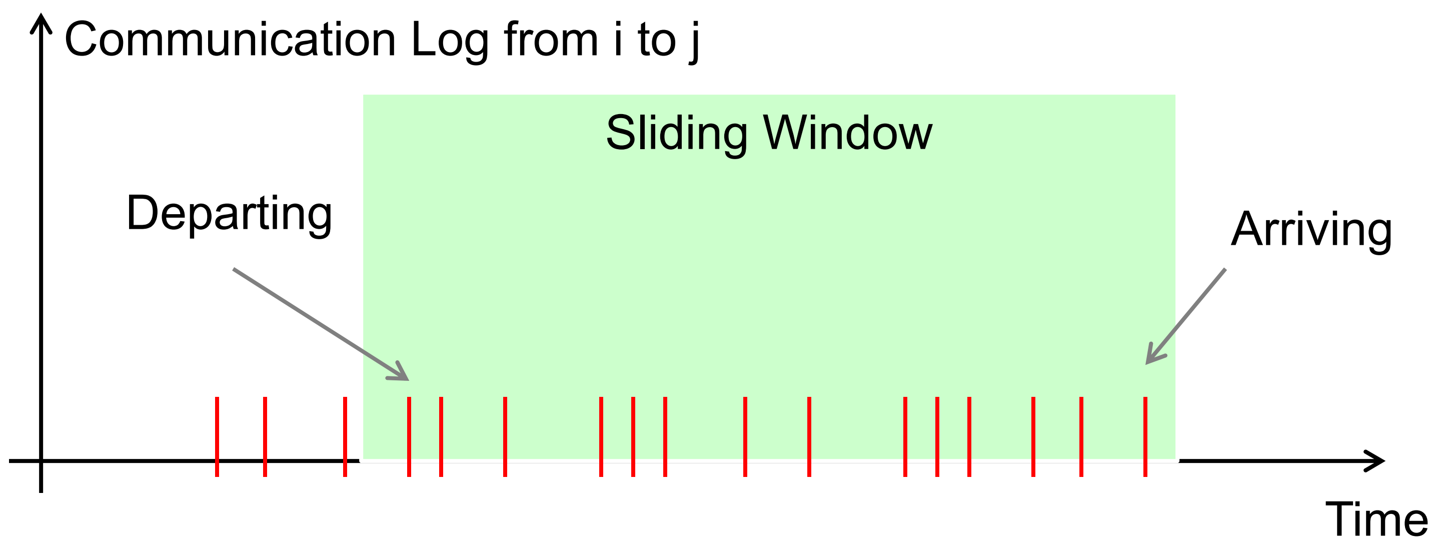
Mathematically Better Window

- 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



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



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.

Communications Log Data

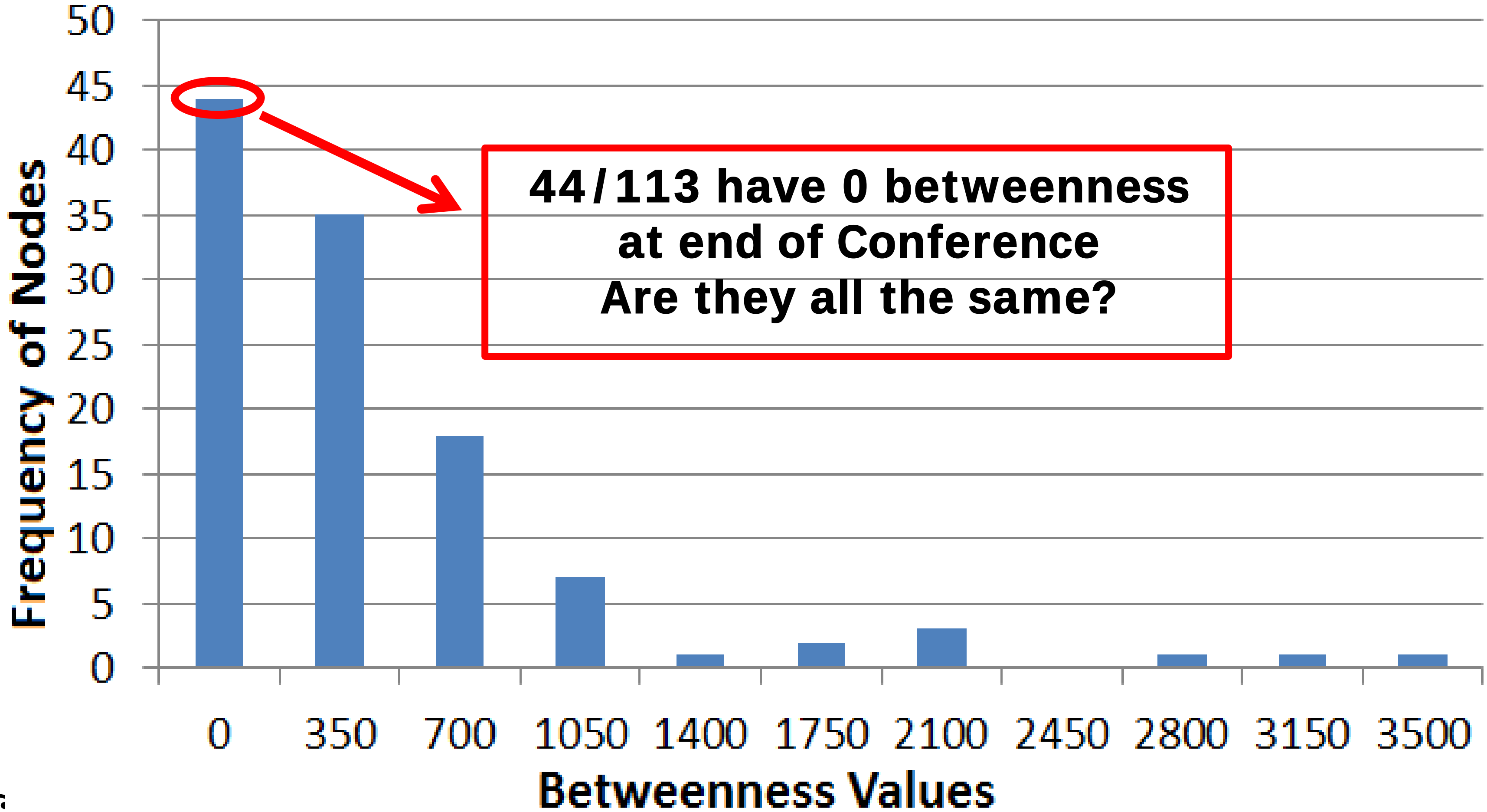
- Data on who you talk to over monitored means, but NOT what you say (decreased privacy concerns relative to full text monitoring)
- Researchers often only have access to logs from 1 or 2 communications channels – not all possible channels
  - Missing data is substantial
- Communication event is taken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

Example: Temporal Social Network

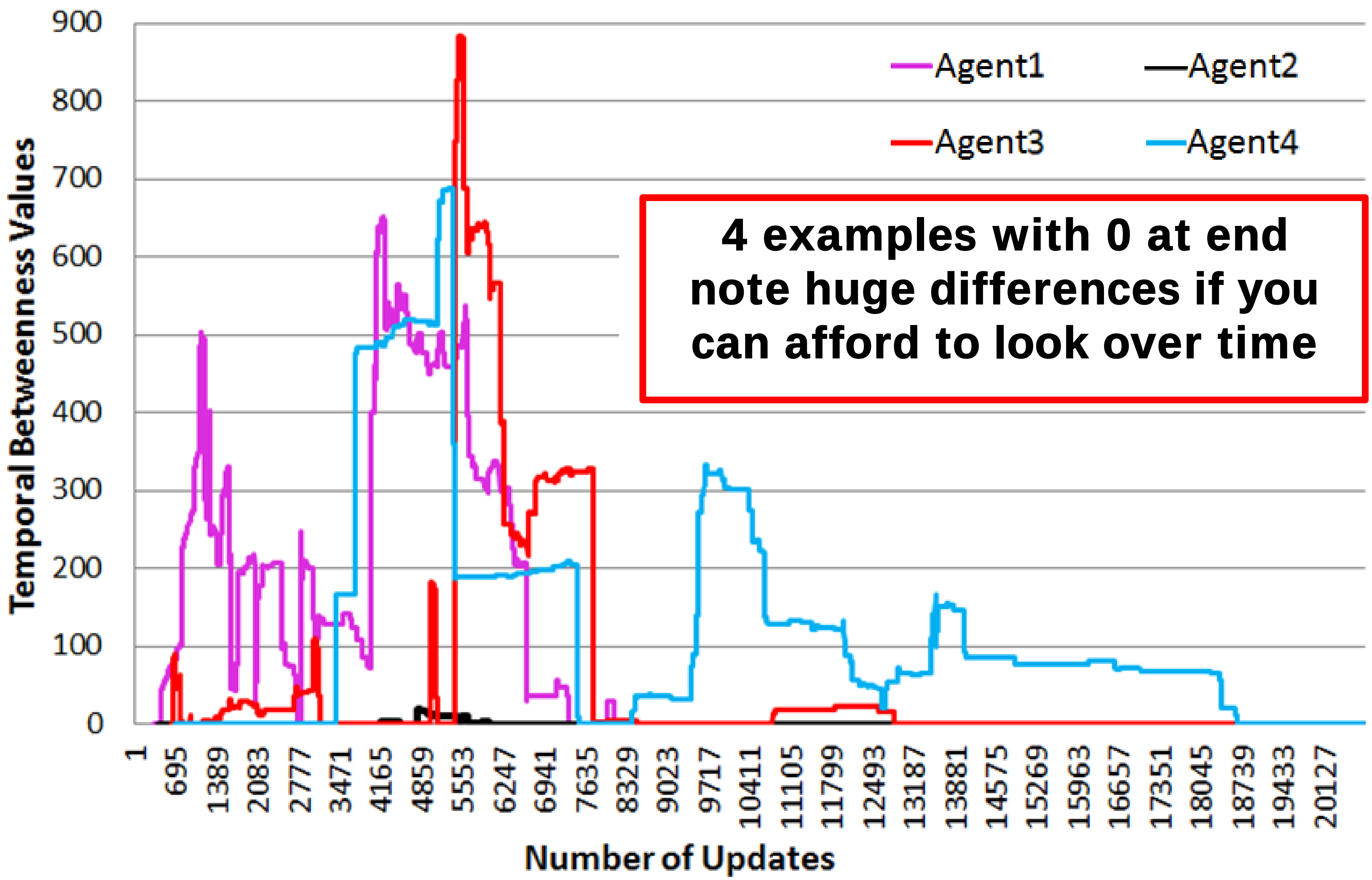
- ACM Hypertext 2009 Conference
  - Badges with active RFID Tags that sense each other
- Tag-to-Tag Detection Only at Close Range
  i.e., detecting Face-to-Face Contact, proxy for interacting
  - Range only 1 - 1.5 meters of one another
  - Human body acts as an RF shield so no front-to-back detection
- Collect sensor data every 20 seconds for 2.5 days
  - 20,818 real time data updates
  - 113 participants, 2196 undirected, weighted links

New Interactions → **Stronger Relations** → **Different Interactions**

Socio-Patterns: Betweenness Centrality Distribution



Socio-Patterns: Betweenness Over-Time Trends



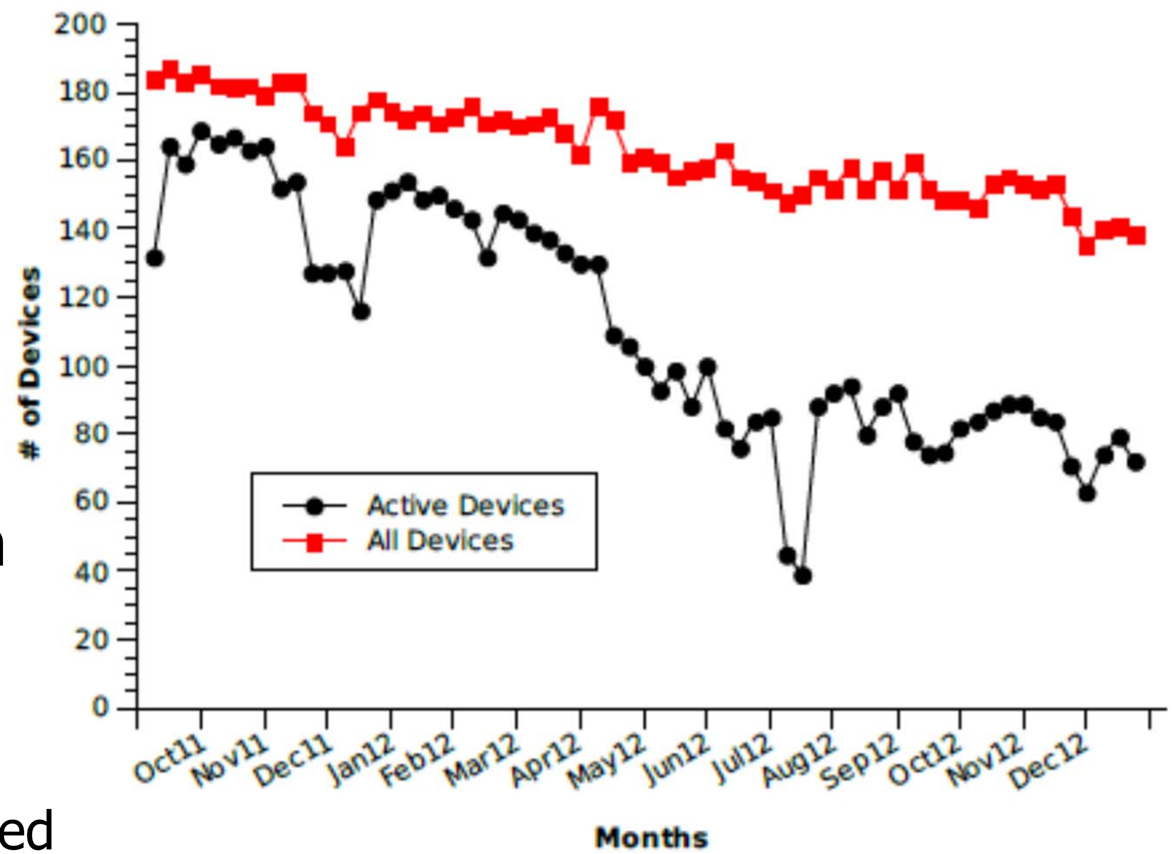
4 examples with 0 at end note huge differences if you can afford to look over time

Is Com Log Data a good Proxy ?

- Example: 2011-2013 NetSense Data Set from Notre Dame
  - Aaron Striegel, Shu Liu, Lei Meng, Christian Poellabauer, David Hachen, Omar Lizardo, "Lessons Learned from the NetSense Smartphone Study," Proceedings of HotPlanet'13, August 16, 2013, Hong Kong, China.
- They recruited 180+ incoming freshmen/freshwomen in 4 dorms to join study
  - Students received free cell phone (including phone plan)
  - Students had to agree to use provided Android cell phone as their primary cell phone
  - Students agreed to having calls and txt msgs logged
  - Students agreed to filling out monthly surveys
- Data collected from study for 4 academic semesters
  - Data from Summer survey too unreliable to use because many students were away from campus for summer

NetSense Details

- NetSense study surveyed all participants monthly + an extra long survey at end of each semester
  - Survey return rate nearly 100%
  - This work focused on the long survey at end each sem.
  - Long Survey Question asked top 10 people you interact with
- NetSense population changes over time
  - Students either quit or violate terms of study and are removed

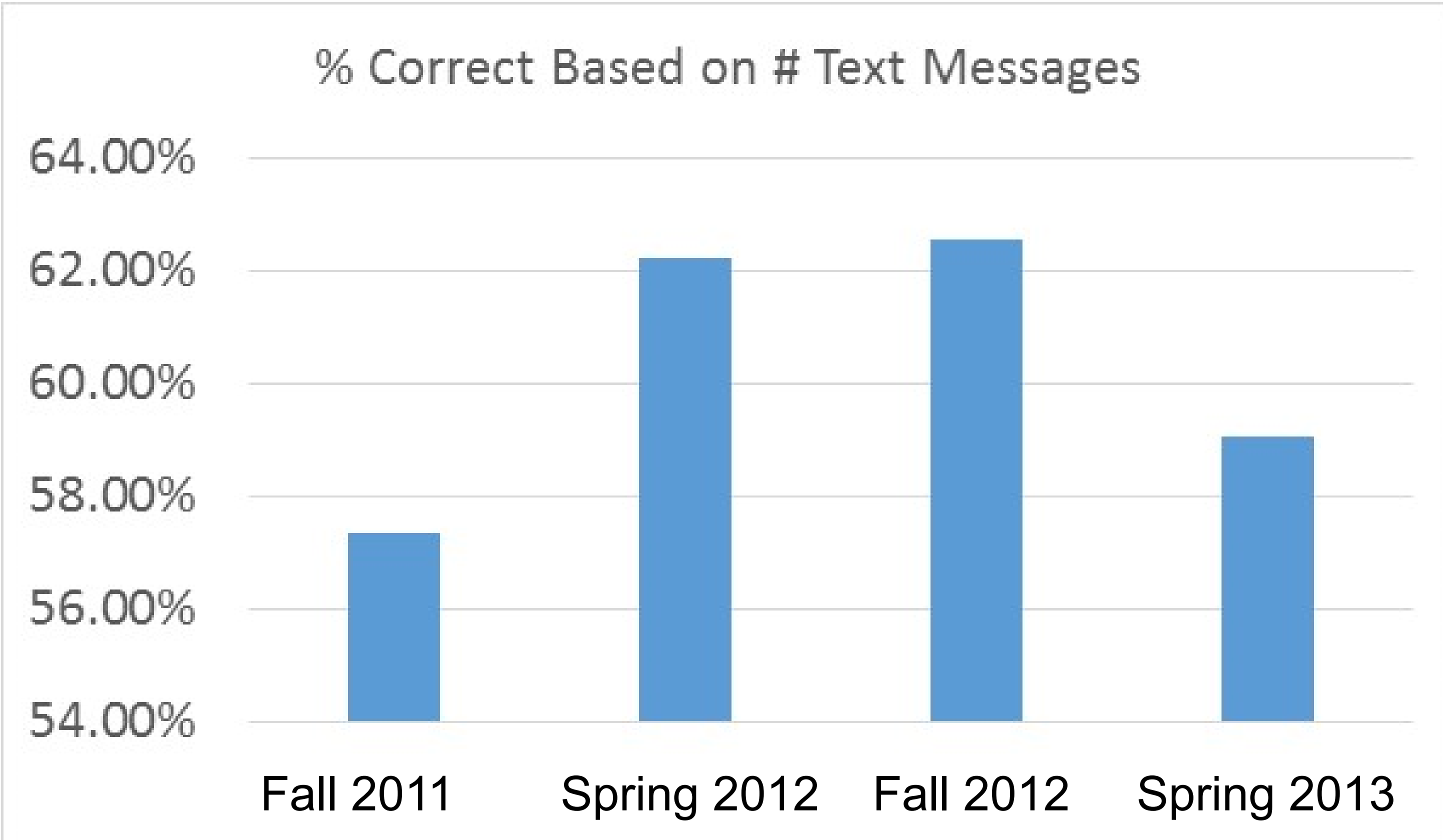


Methodology

- Question to be studied:
  - Accuracy of phone logs relative to survey for predicting network
- Survey
  - Asked students to list top 10 people they interact with regularly
  - Students didn't have to fill in all 10 slots
  - May of those listed were people outside of study (e.g., parents)
  - Keeping only those in study gave list of 0-10 others in the study that the surveyed individual considered strong interaction targets
- Cell Phone Data
  - Looked at # txt msgs, # txt chars, # phone calls, # secs on calls
  - Ranked in-study interactors based on these metrics
- Predictor Quality
  - Probability individual listed as one of N in-study individuals in survey is in the top N based on cell phone data

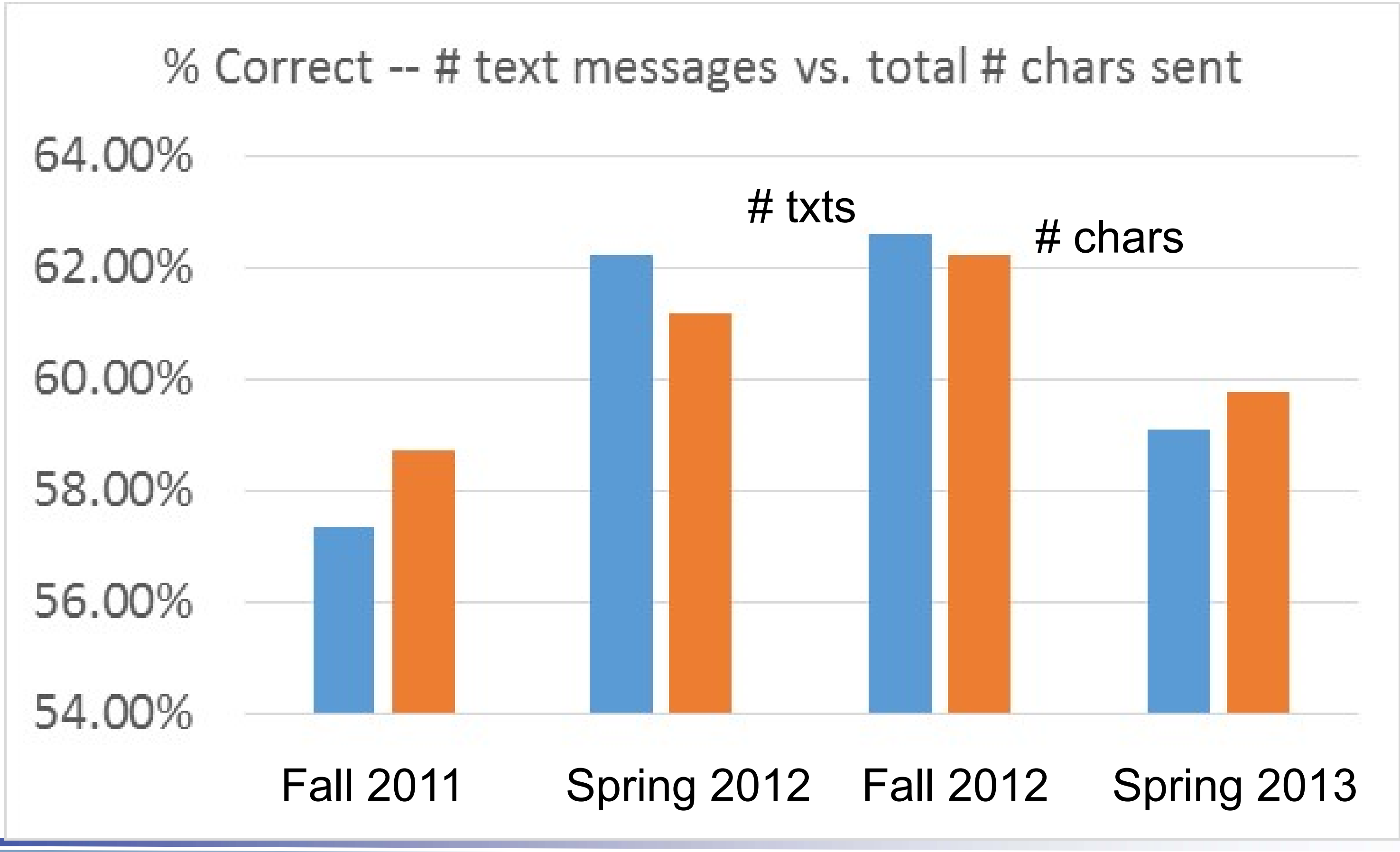
Text Messages

- This generation lives on text messages
  - Overall, # txt msgs accurage about 60% of the time



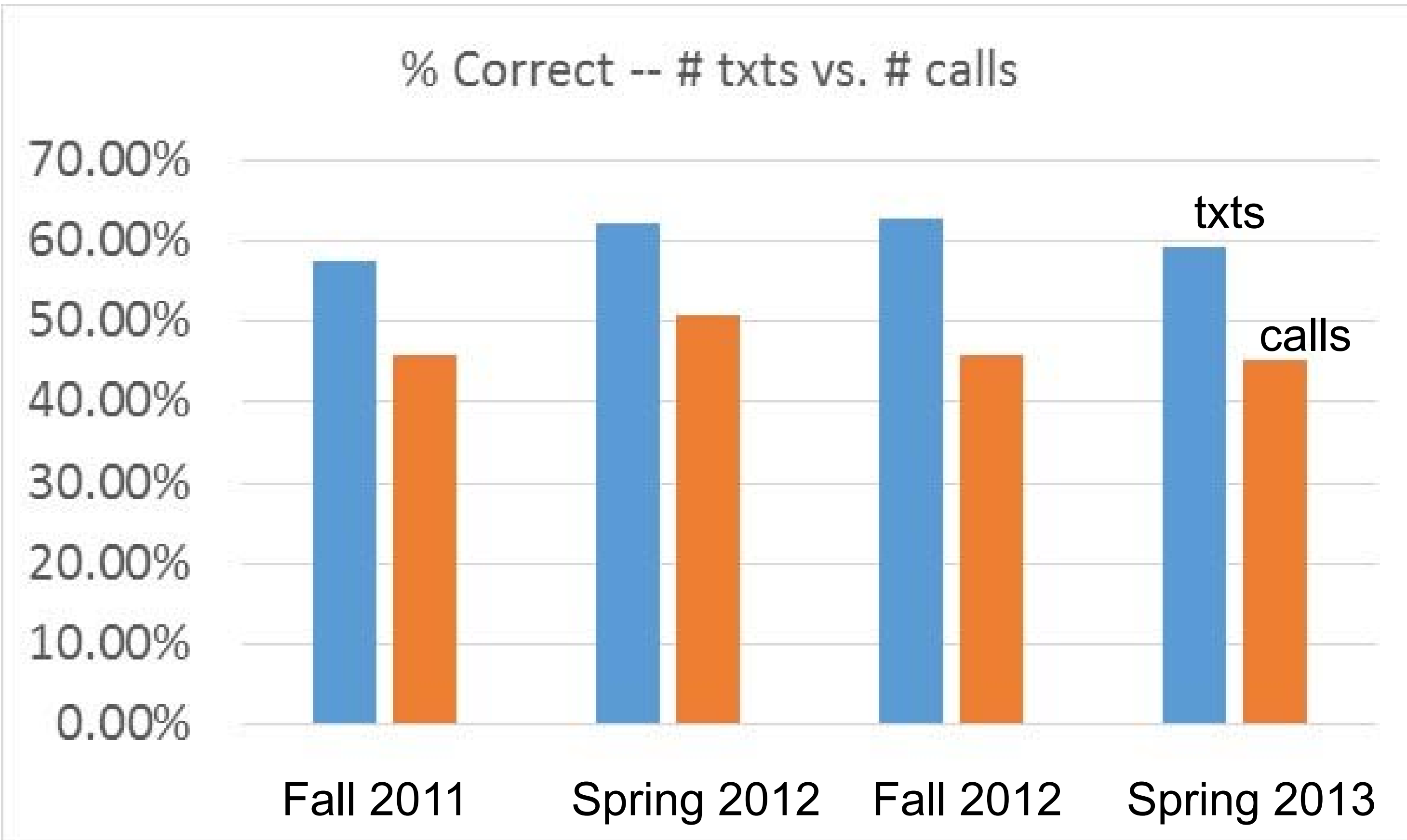
Chars in Text Messages

- Length of text message does not seem significant
  - Using # chars instead of # txt msgs is actually slightly worse



Phone Calls

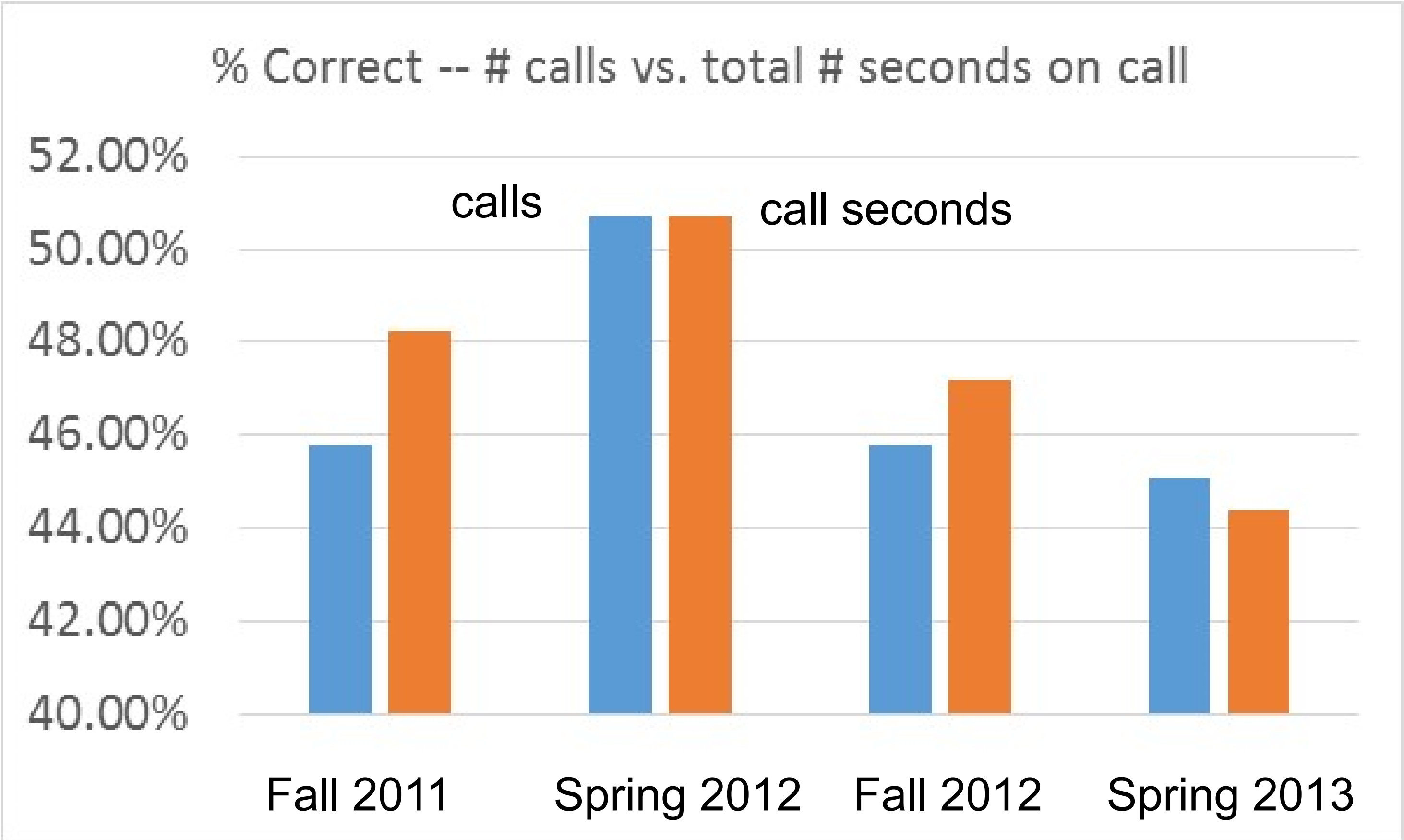
- How about using # phone calls instead of # txt msgs
  - Not good at all – only about 45% on average for phone calls



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

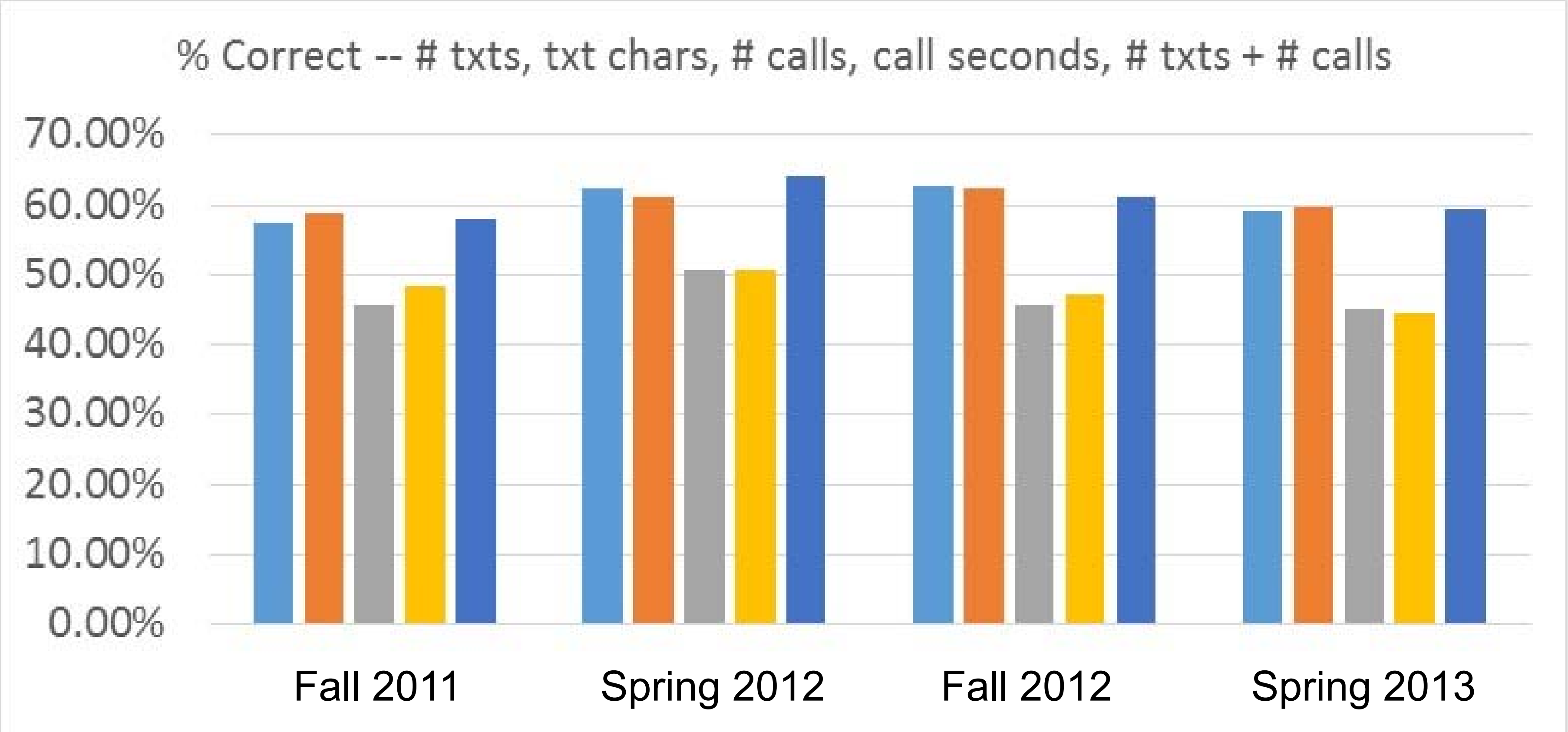
Call Seconds vs. # Calls

- How about using # call seconds instead of # txt msgs ?
  - A little better – about 48% on average for calls seconds



Composite Measures

- Comparing all to (#txts + #calls) as measure ?
  - This combination achieves highest score, above 60% accurate



Conclusions: Com Logs can be OK Proxy for Network Ties

- # txt msg is good proxy for interaction propensity for this cohort
- Combinations of comm data metrics can slightly increase accuracy, but only a little
- Accuracy level of about 60% indicates that many interactions are mediated by other communications channels (e.g., face-to-face).
- Results of this analysis may vary widely for different communities – 2011 freshmen/freshwomen are highly attached to txt msgs for communication
- Note, self-reporting errors may influence these results – e.g., participants took final survey less seriously

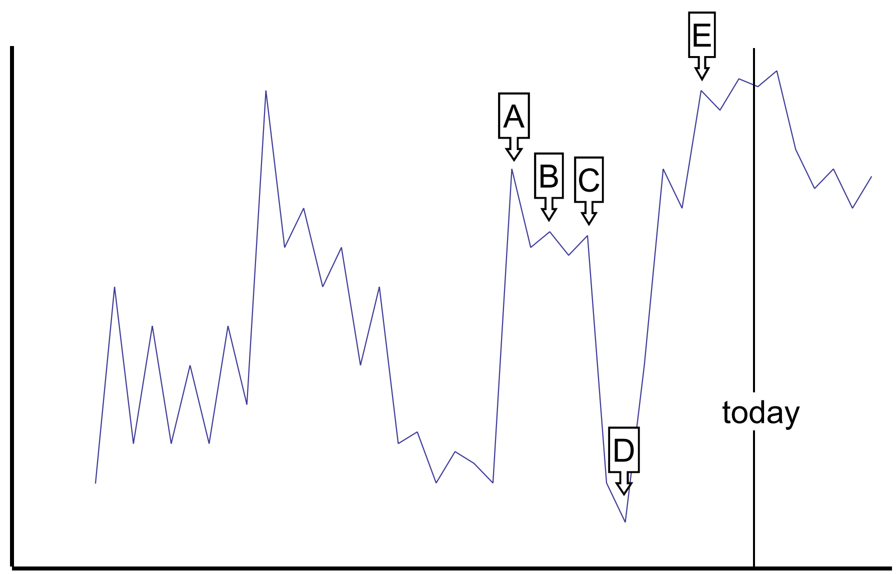
Carnegie Mellon

isr institute for SOFTWARE RESEARCH

Topic 3: Detecting Change and when it occurs

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.



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

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

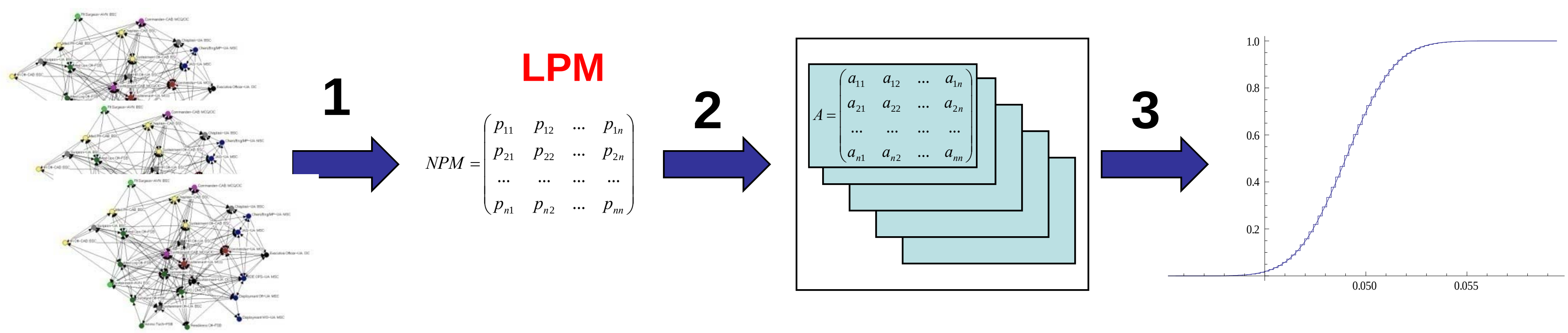
Statistical Models of Networks

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:

$$p = \int_0^t f_{ij}(x \mid \theta_{ij})\,dx$$

  - (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



Statistical Models of Networks

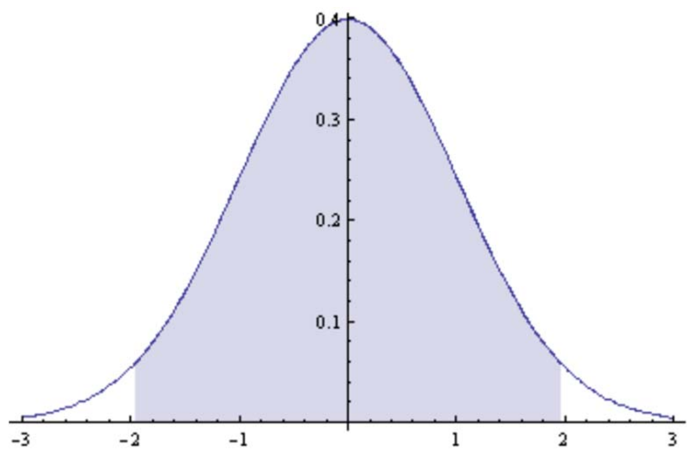
Link Probability Model (LPM) for Stability

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

M	*8*		*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 μ=0 and σ=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

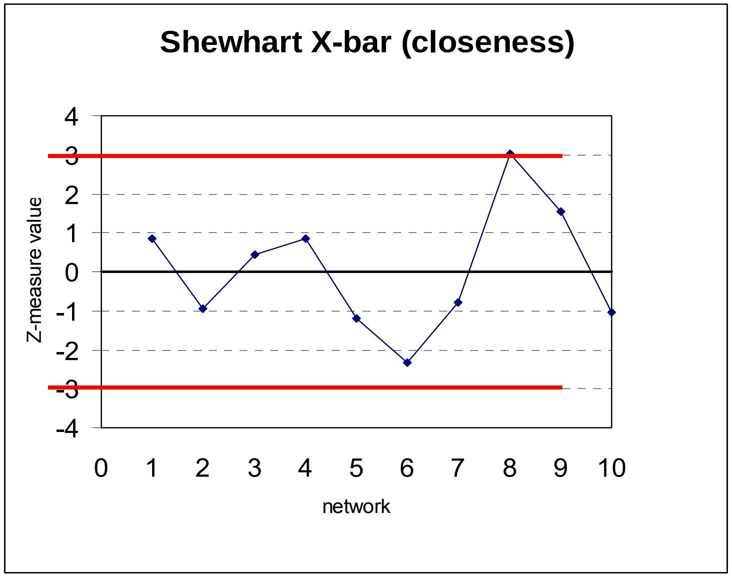
- Manufacturing processes are: stochastic, dependent, non-ergotic, complex, and involve human interaction.
- Shewhart (1927) X-bar Control Chart proposed to monitor change of any process
- Calculate Z_t transform value for each time-period, *t*.

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

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

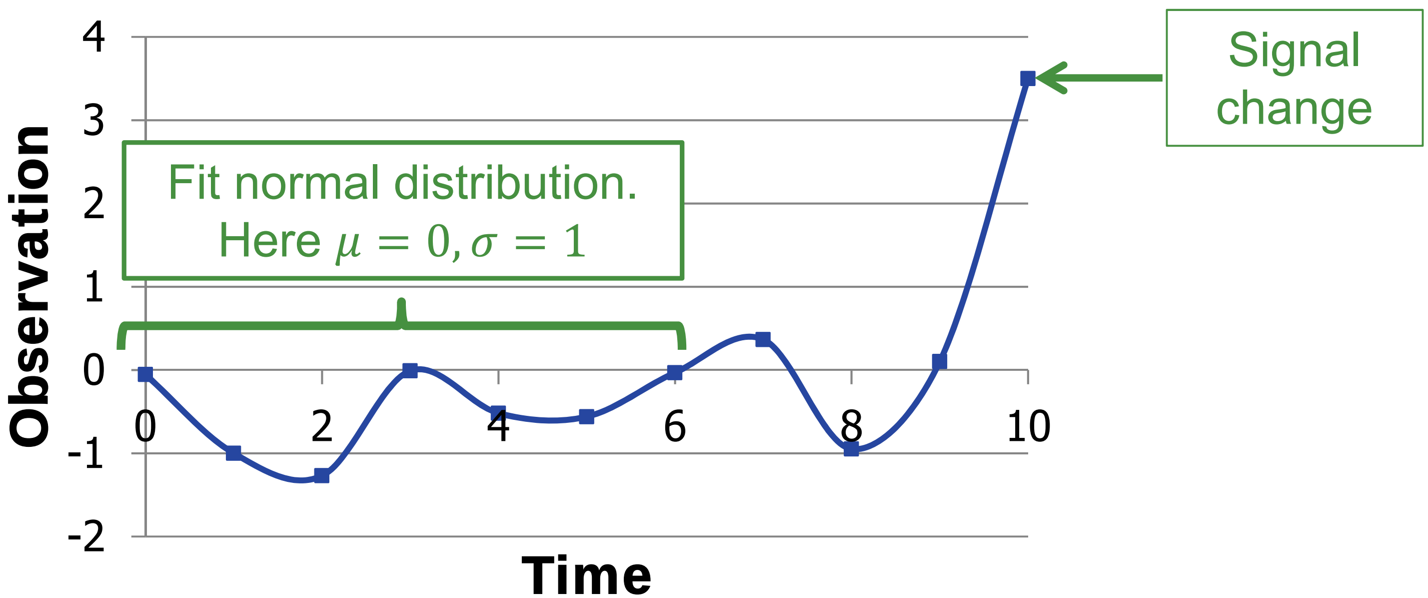
$$\int_L^\infty f(x)dx = \alpha$$

- Chart Signals when *Z* exceeds control limit, *L*.



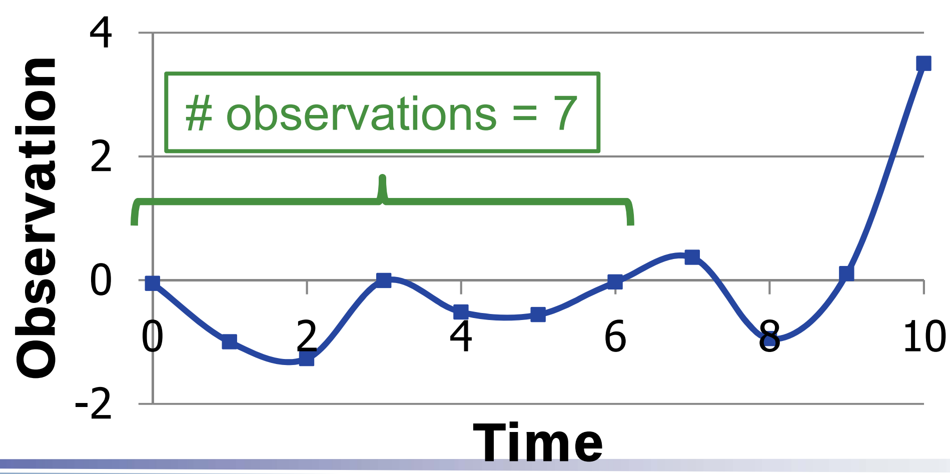
The Shewhart X-Bar Chart

- Overview
  - Fit normal distribution on “control period” (early observations)
  - 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 independent random vars
    - Shewhart X-Bar chart used even when assumption is violated. However, false positive risk proability may be inaccurate



Carnegie Mellon

isr institute for SOFTWARE RESEARCH

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

$$w_t = \lambda \bar{x}_t + (1-\lambda) w_{t-1} \qquad \mu_0 \pm L\sigma_{\bar{x}} \left(\frac{\lambda}{2-\lambda} \left[1-(1-\lambda)^{2T} \right] \right)^{1/2},$$

- 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)

$$C_t^+ = \max\{0,\ Z_t - k + C_{t-1}^+\} \qquad C_t^- = \max\{0,\ -Z_t - k + C_{t-1}^-\}$$

CASOS

Cumulative Sum (CUMSUM)

- 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

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

- Two Charts (To Detect Increase and Decrease)

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

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

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

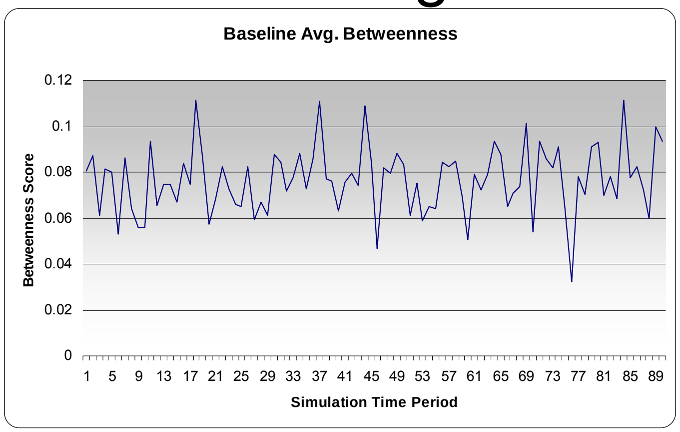
- Sensitivity in CUSUM due to discrete integration of error

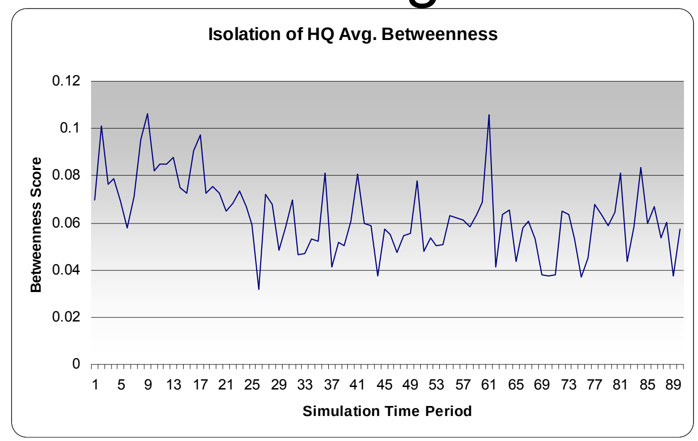
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Comparison of Change Detection Approaches

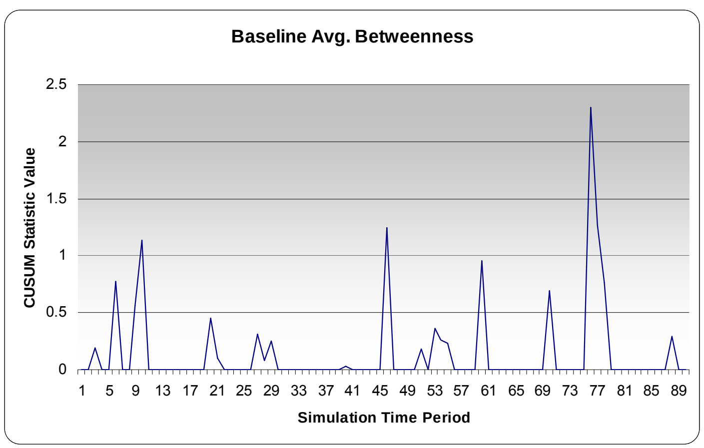
No Change | Change

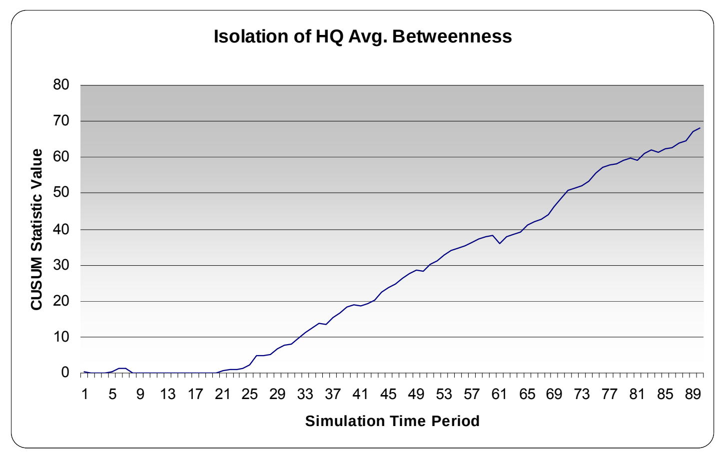
Over-Time Meas





CUSUM Statistic





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

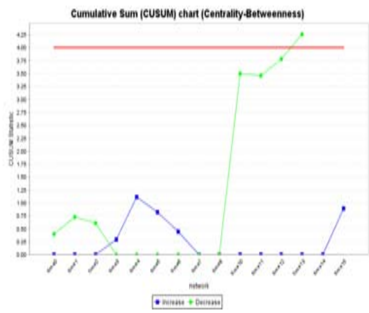
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Network Change Detection: Analysis of Real World Data

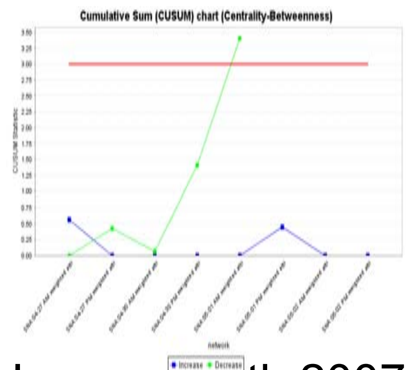
	No Nodes	Time Periods	Method of Collection	Type of Relation	Design	Known Change
Fraternity	17	15	Survey	Ranking	Fixed	Yes
Leav 07	68	8	Survey	Rating	Free	Yes
Leav 05	158	9	Survey	Rating	Free	None
Al-Qaeda	62-260	17	Text	Rating	Free	Yes
Winter C	22	9	Observation & Survey	Rating	Fixed	Yes
Winter A	28	9	Observation & Survey	Rating	Fixed	Yes
IkeNet 2	22	46	Email	Count Msg	Free	Yes
IkeNet 3	68	121	Email	Count Msg	Free	Yes

CASOS

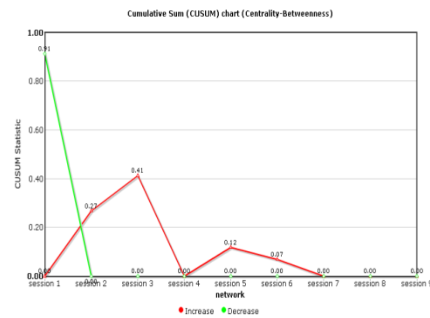
Network Change Detection: Analysis of Real World Data



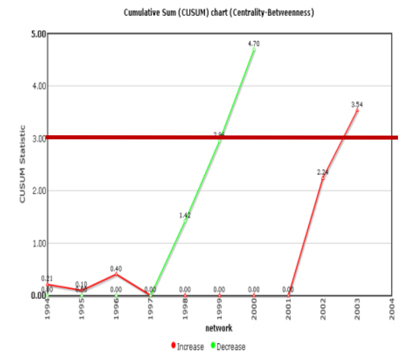
Fraternity



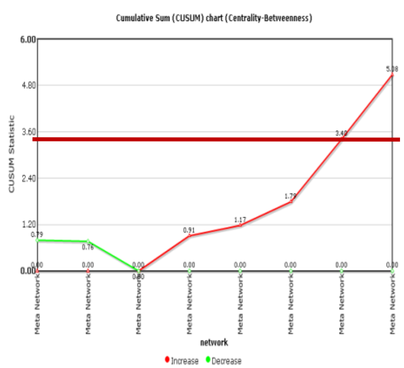
Leavenworth 2007



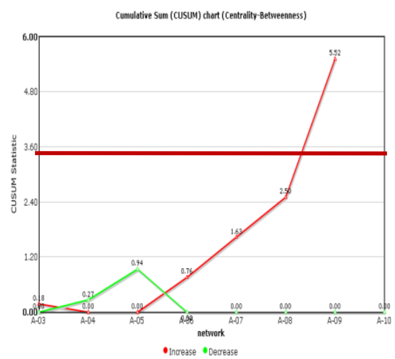
Leavenworth 2005



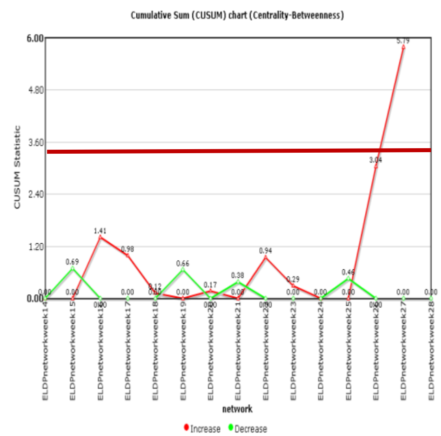
Al-Qaeda



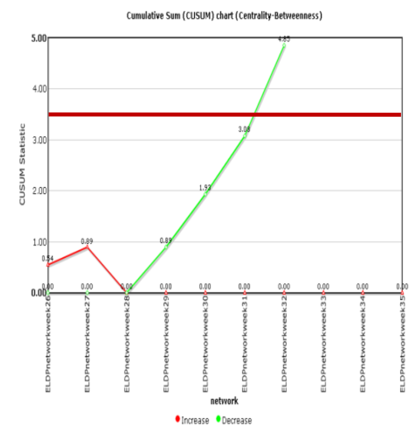
Winter A



Winter C



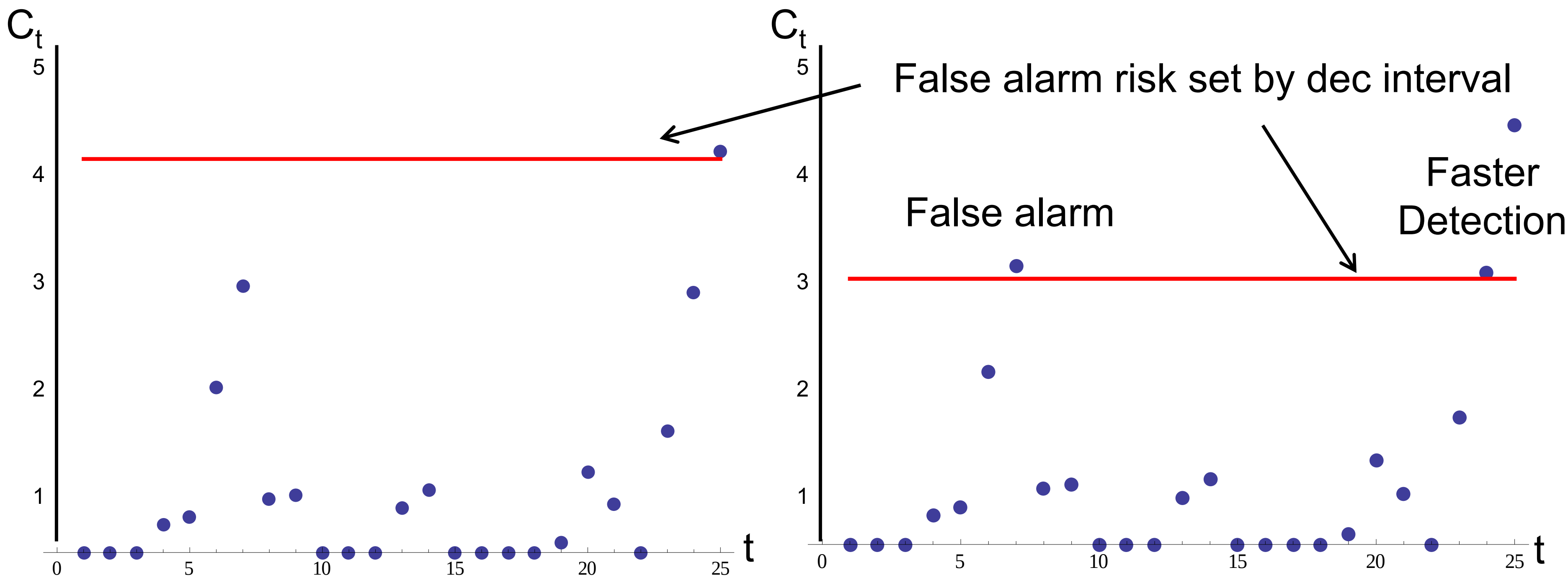
IkeNet 2



IkeNet 3

Summary of Change Detection Across Data Sets

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



Low risk of false alarm
Longer to detect change

High risk of false alarm
Faster to detect change

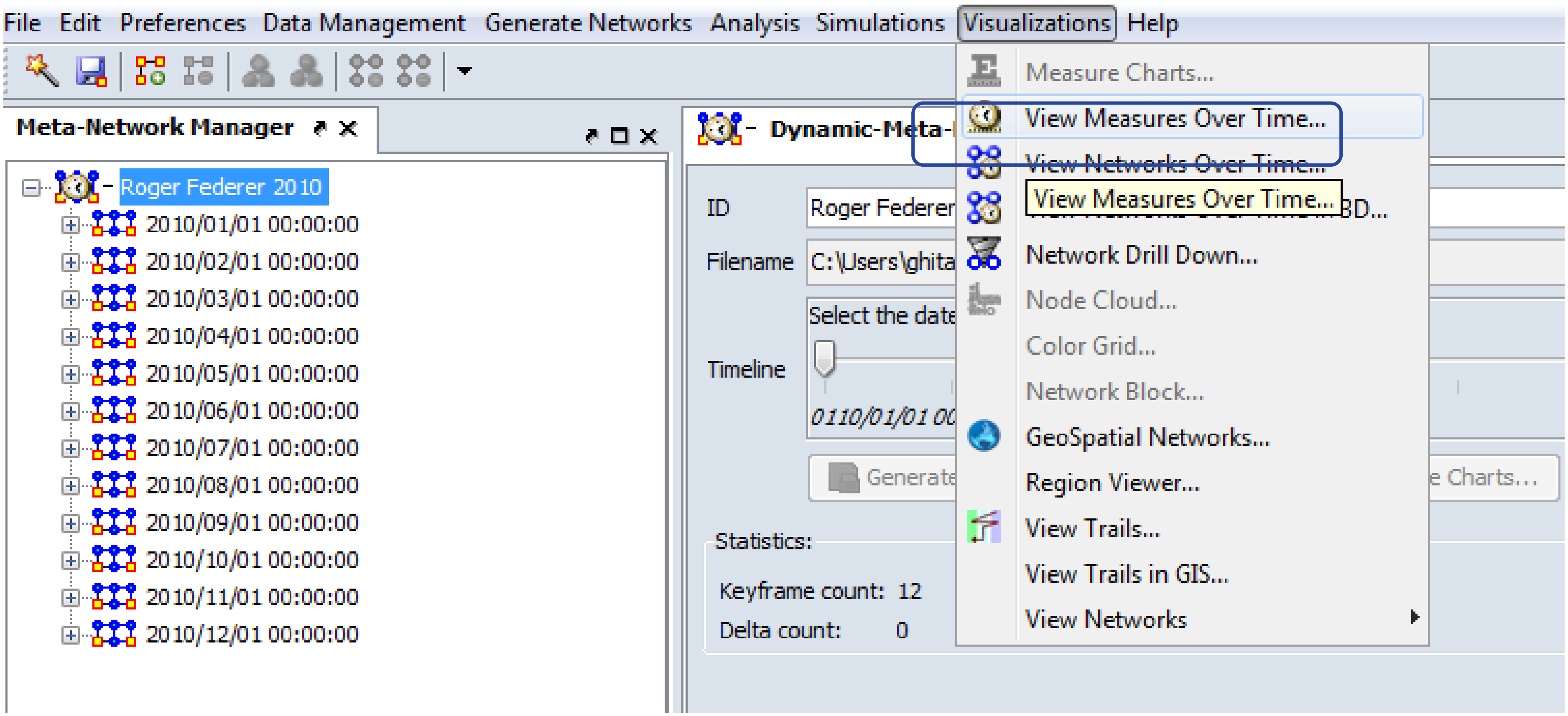
Summary of Change Detection Across Data Sets

Too little risk may prevent change detection

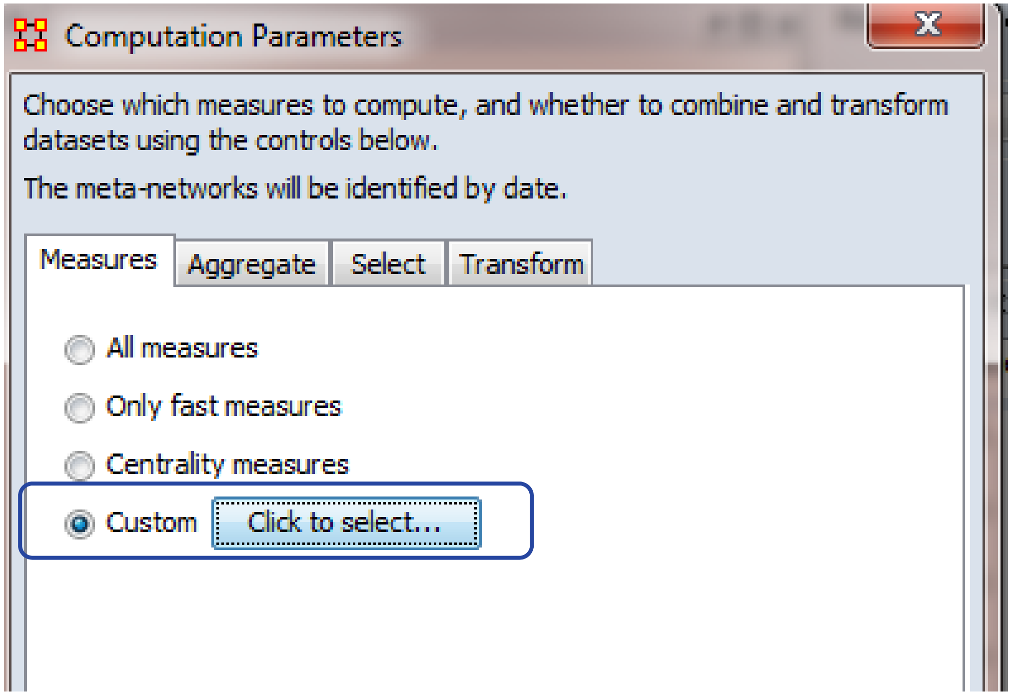
Data	Change	α = 0.05	α = 0.02	α = 0.01	α = 0.005	α = 0.001
Fraternity	8	10	10	10	13	Never
Leav 07	3	5	5	5	Never	Never
Leav 05	None	No F.A.	No F.A.	No F.A.	No F.A.	No F.A.
Al-Qaeda	1997	1999	1999	2000	2000	Never
Winter C	May	Sept	Sept	Oct	Oct	Never
Winter A	May	Aug	Sept	Sept	Sept	Oct
IkeNet 2	25	26	26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

Change Detection Hands-On

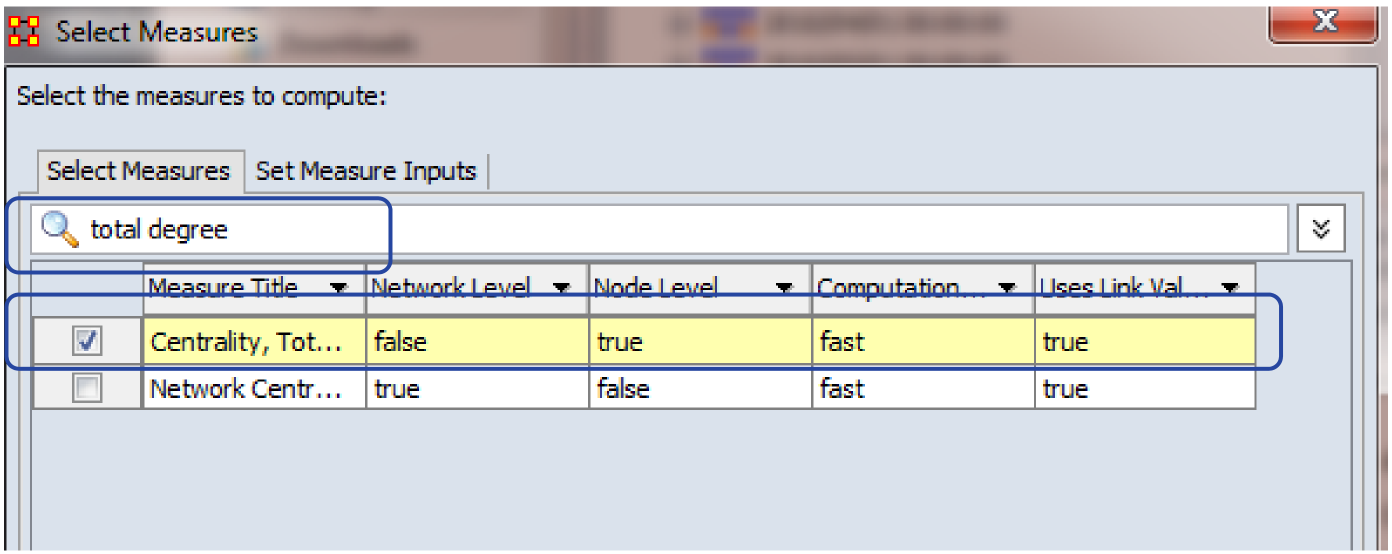
- Based on Roger Federer 2010 data



Change Detection Hands-On

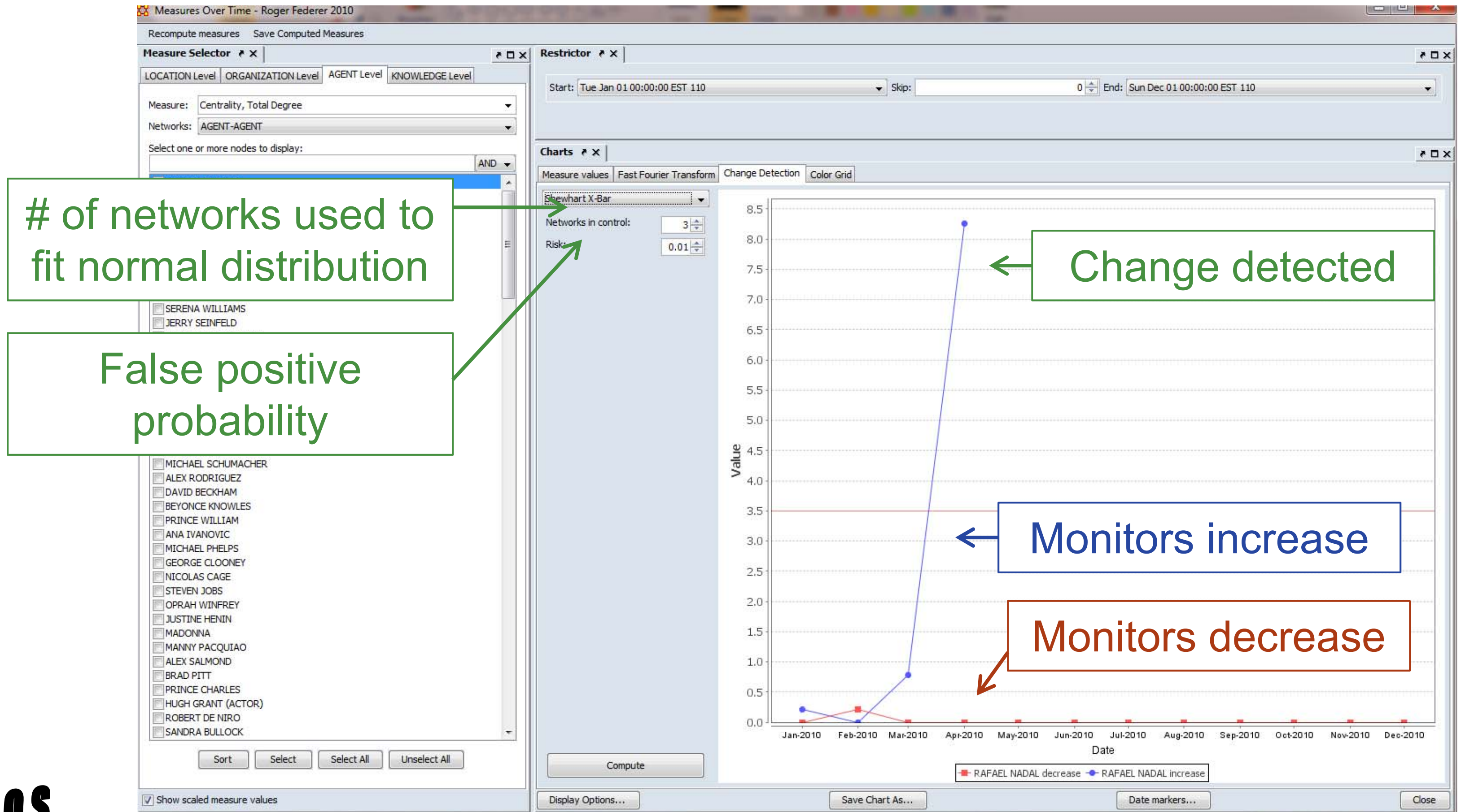


Change Detection Hands-On

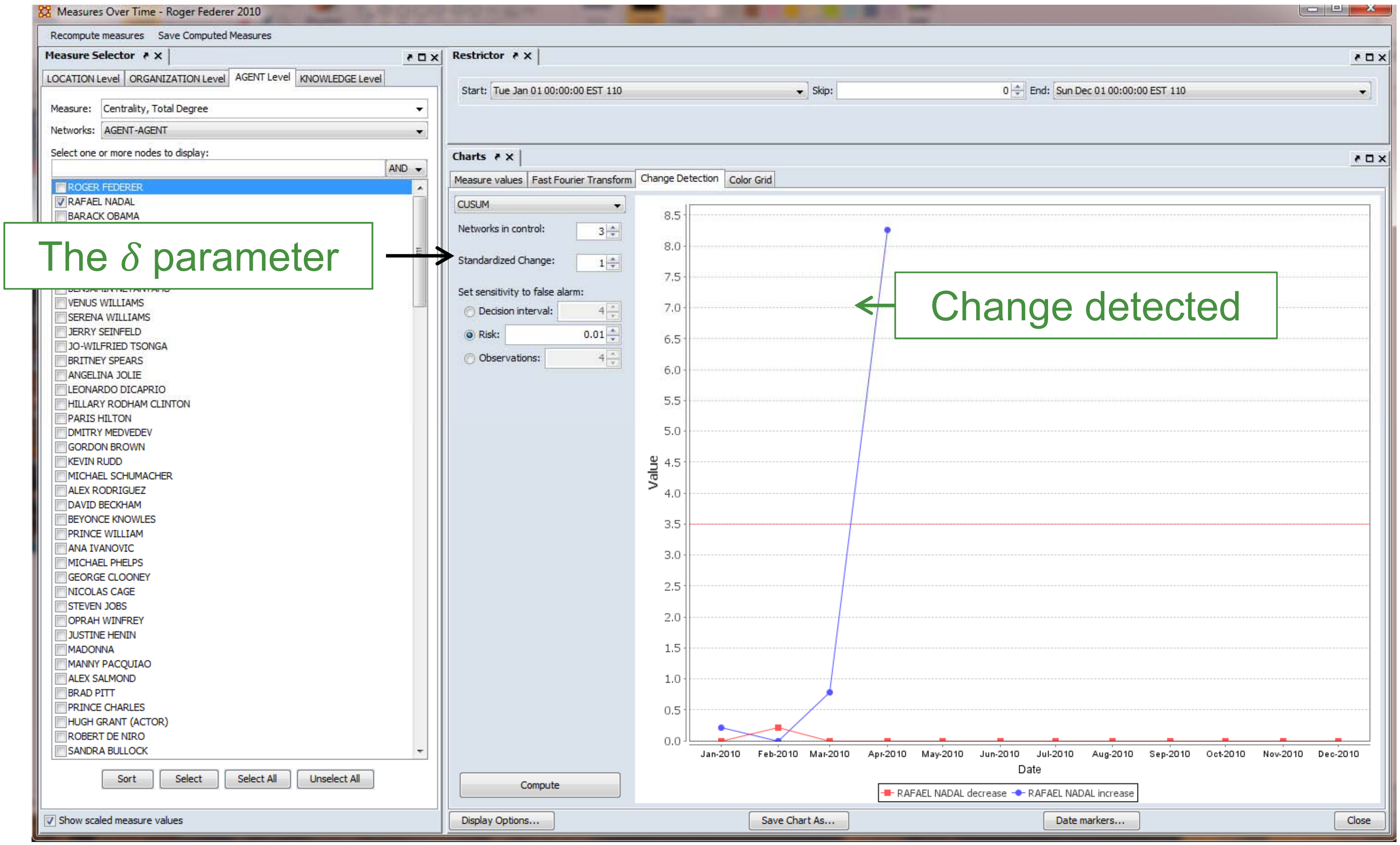


Add Measure – Agent Based Measure – select "Centrality, Total Degree"

Change Detection Hands-On
The Shewhart X-Bar Chart



Change Detection Hands-On CUMSUM Method



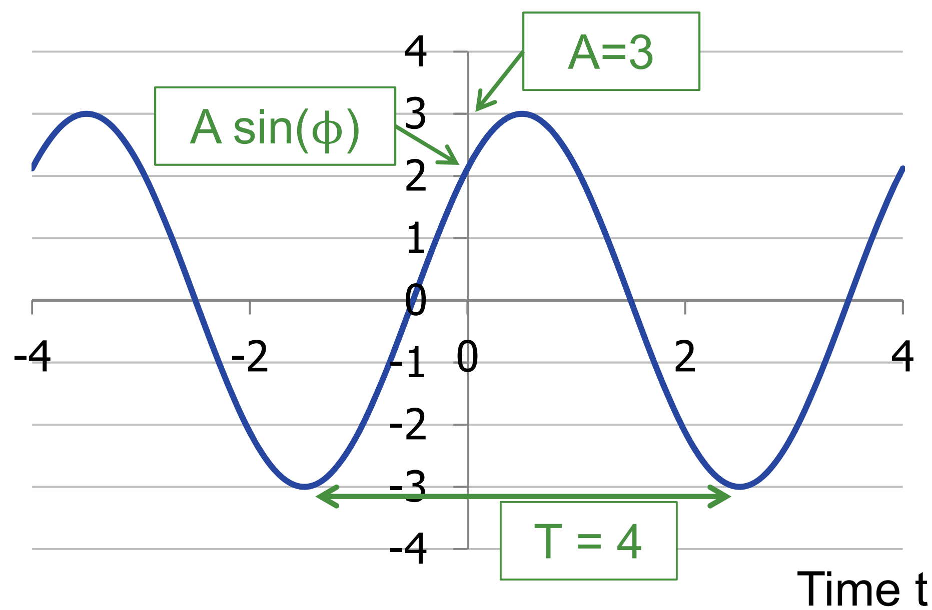
Fast Fourier Transform (FFT)

- 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

Sinusoidal Function

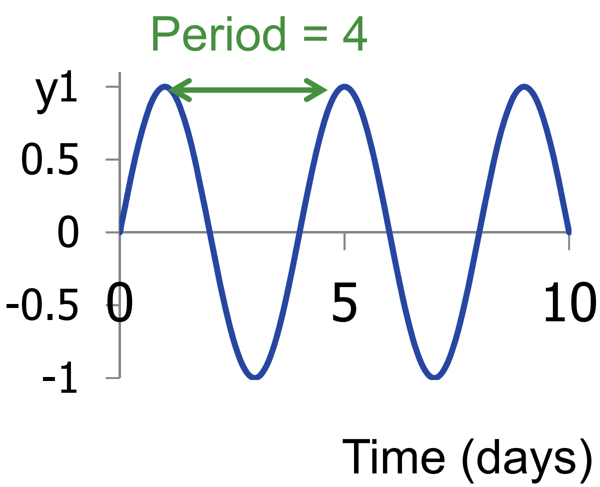
- A sinusoidal function $y = A\sin(2\pi f t + \phi)$ has
  - A amplitude
  - f frequency $\left(T = \dfrac{1}{f}\right.$ is the period)
  - ϕ phase

y = 3 sin(2π(0.25)(t + 0.5))

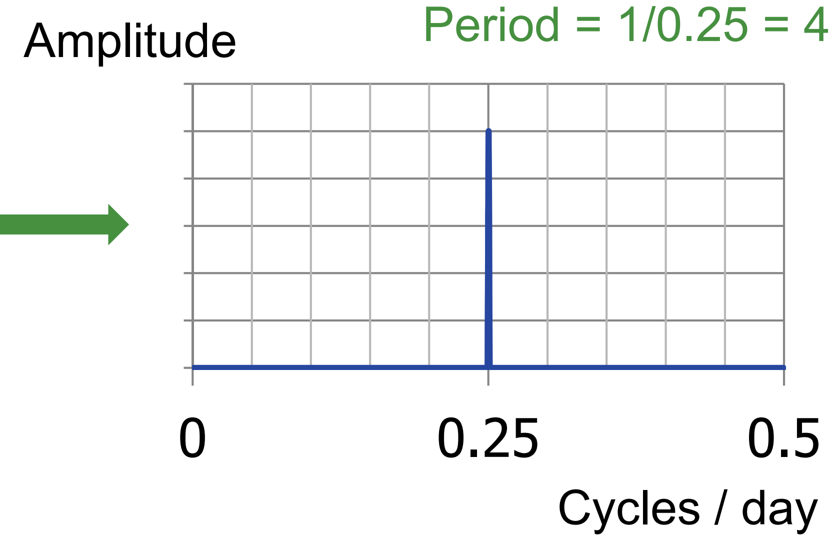


FFT Example: Sinusoidal Function

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



FFT

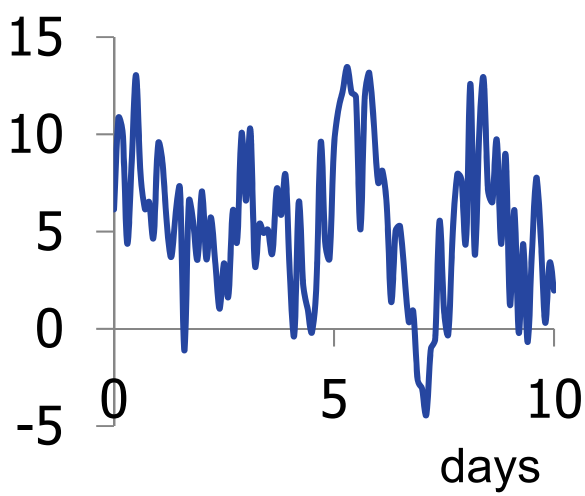


Over-time signal

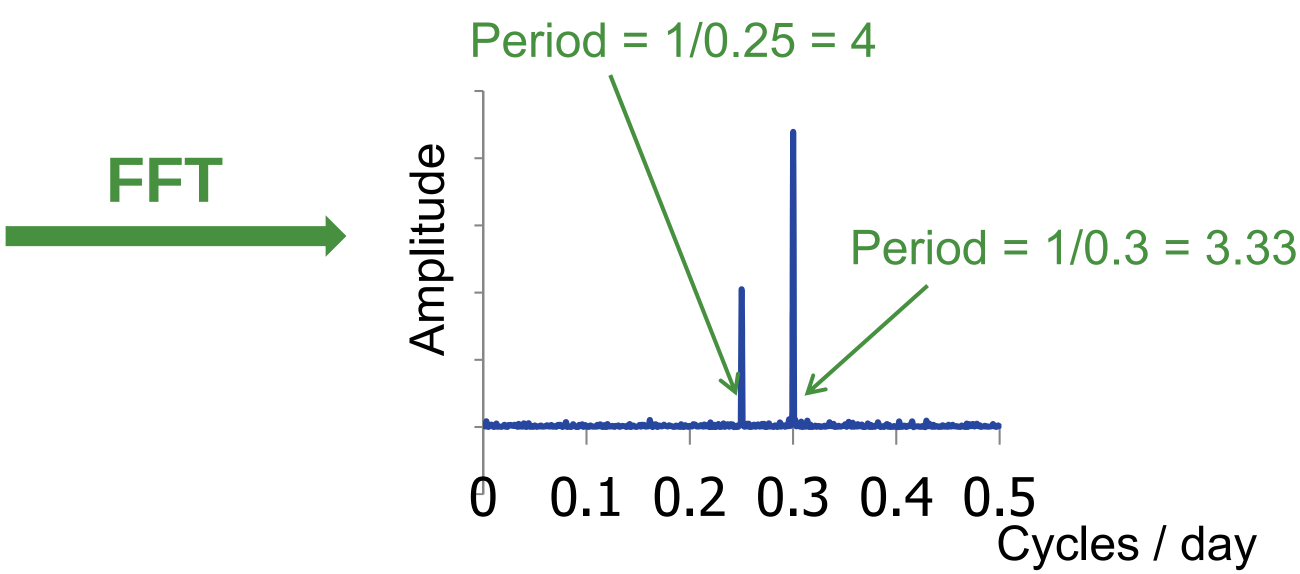
FFT

FFT Example 2

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



Over-time signal

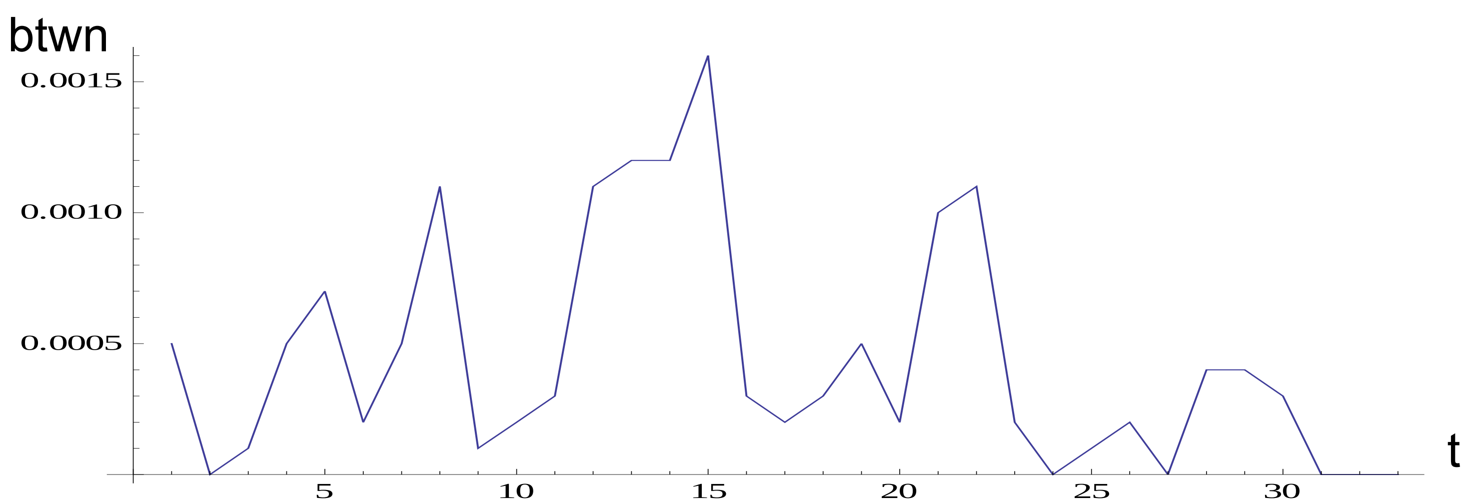


FFT

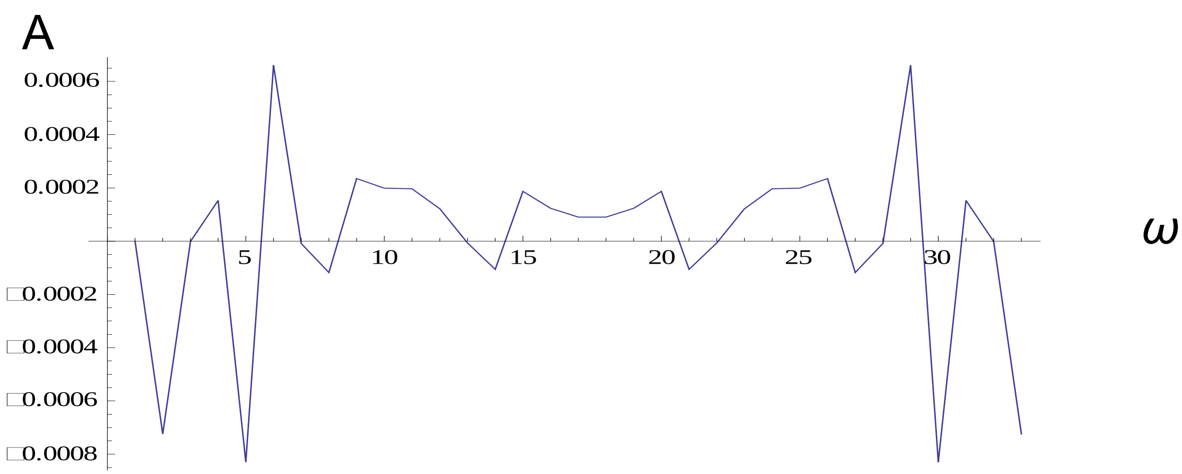
Hidden "recipe": over – time signal computed as
y(t) = 2 sin(2 pi 0.25 t) + 3 sin(2 pi 0.3 t + 0.2) + noise

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.



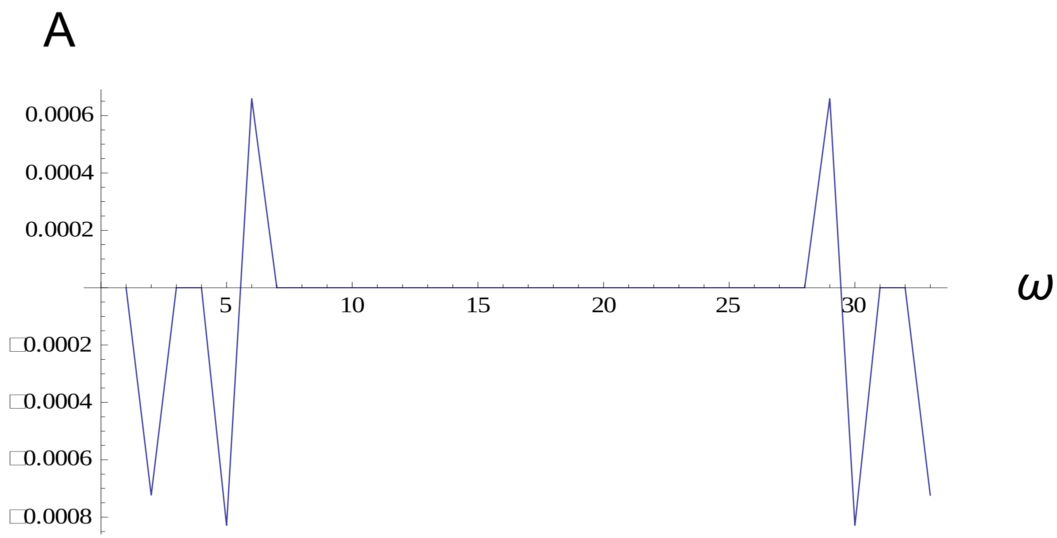
Fourier Analysis – Example 3



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

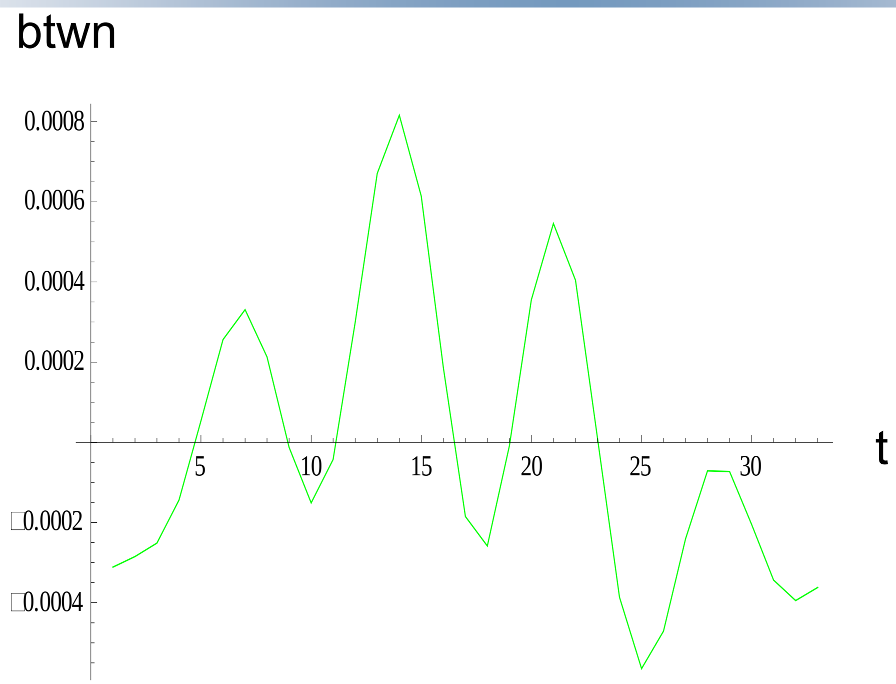
Filtering



3 main (high magnitude) components picked out
The others have been clipped out

ω Frequency in radians per second

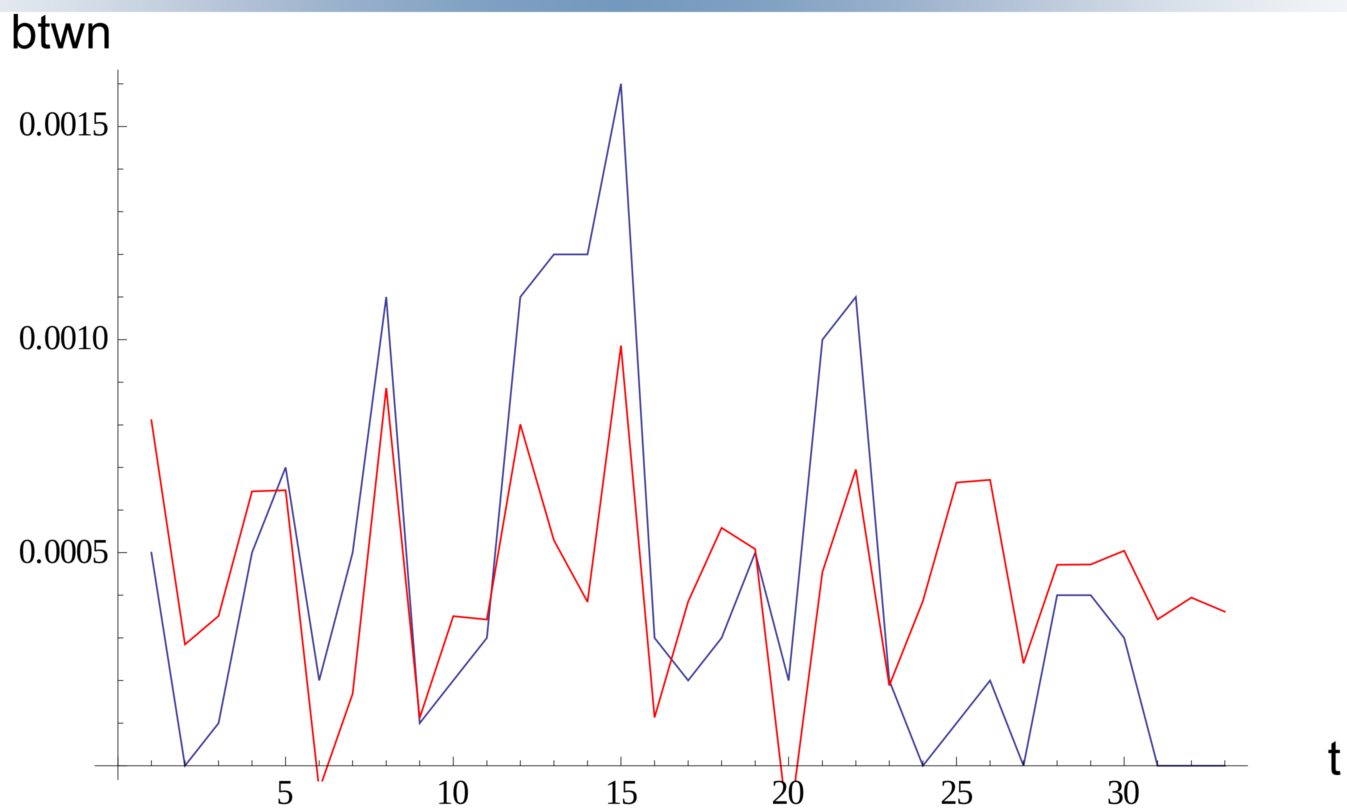
Inverse Fast Fourier Transform



This is the inverse Fourier transform
The filtered 3 components have been reconverted to time
There is a weekly, two week and three week cycle

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

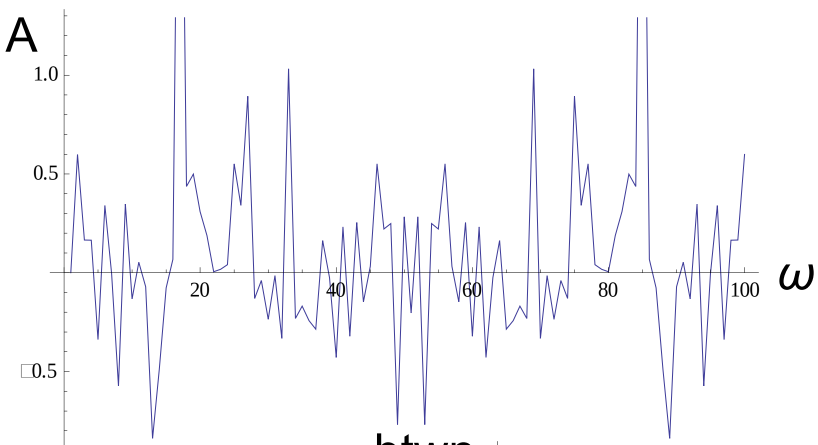
Anomaly Detection

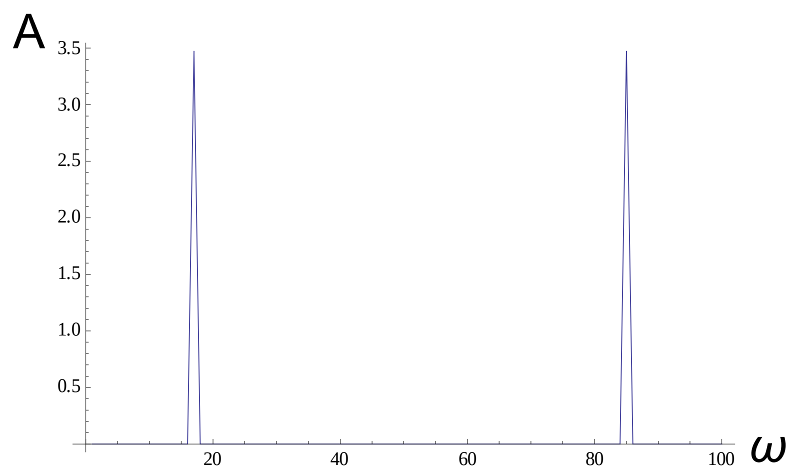


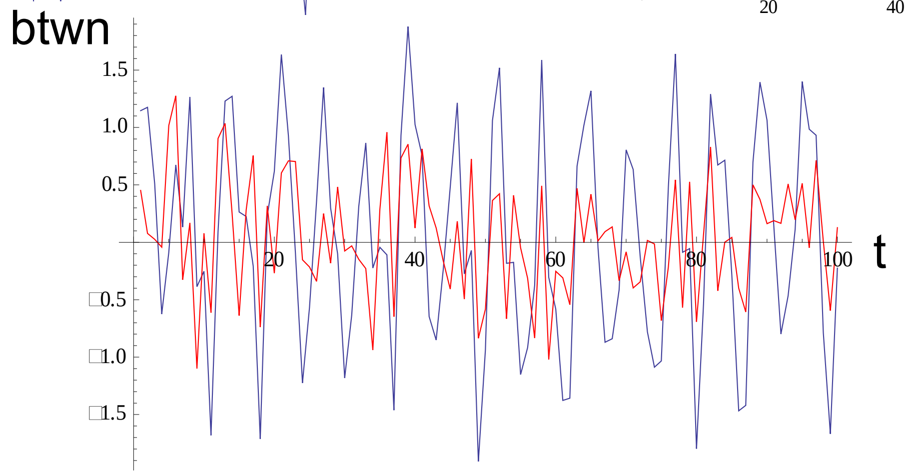
The filtered pattern has been subtracted from the original
The red is what is left – the anomalies

Fourier Analysis to Handle Periodicity

Fourier Analysis on IkeNet data from AY2007-2008

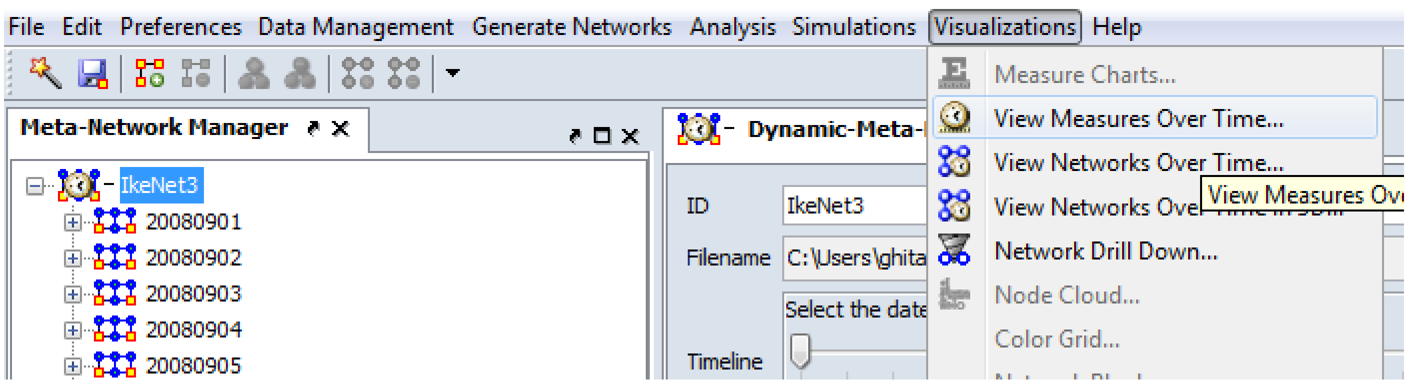




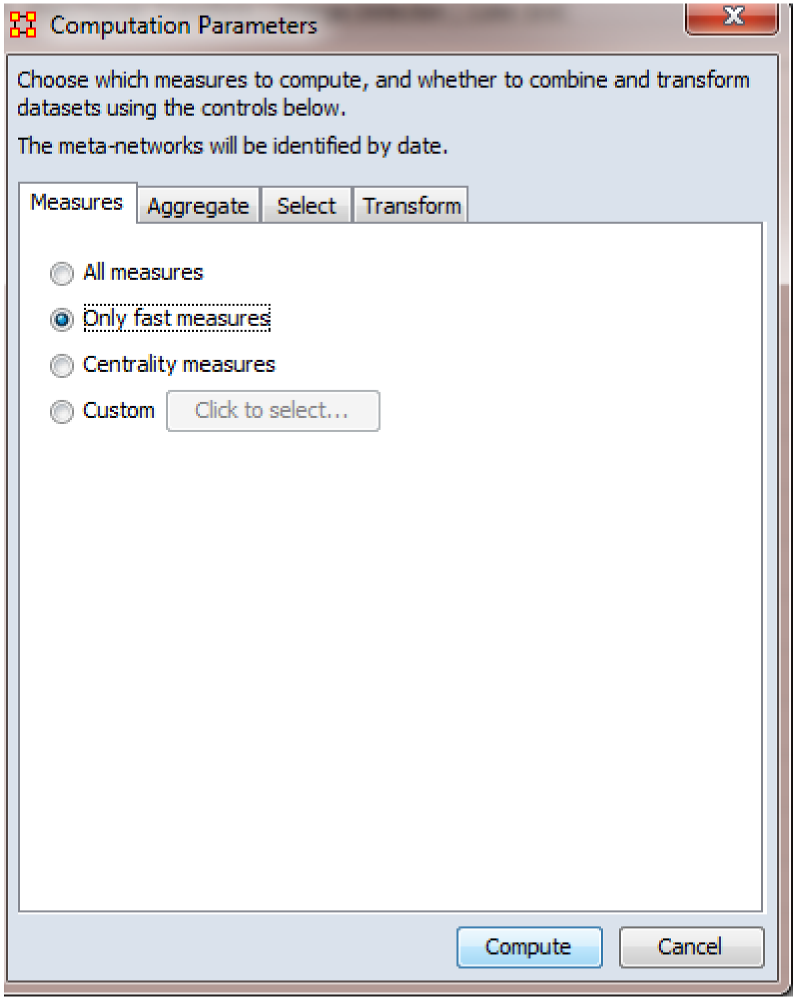


FFT Example Hands-On (1/4)

- 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

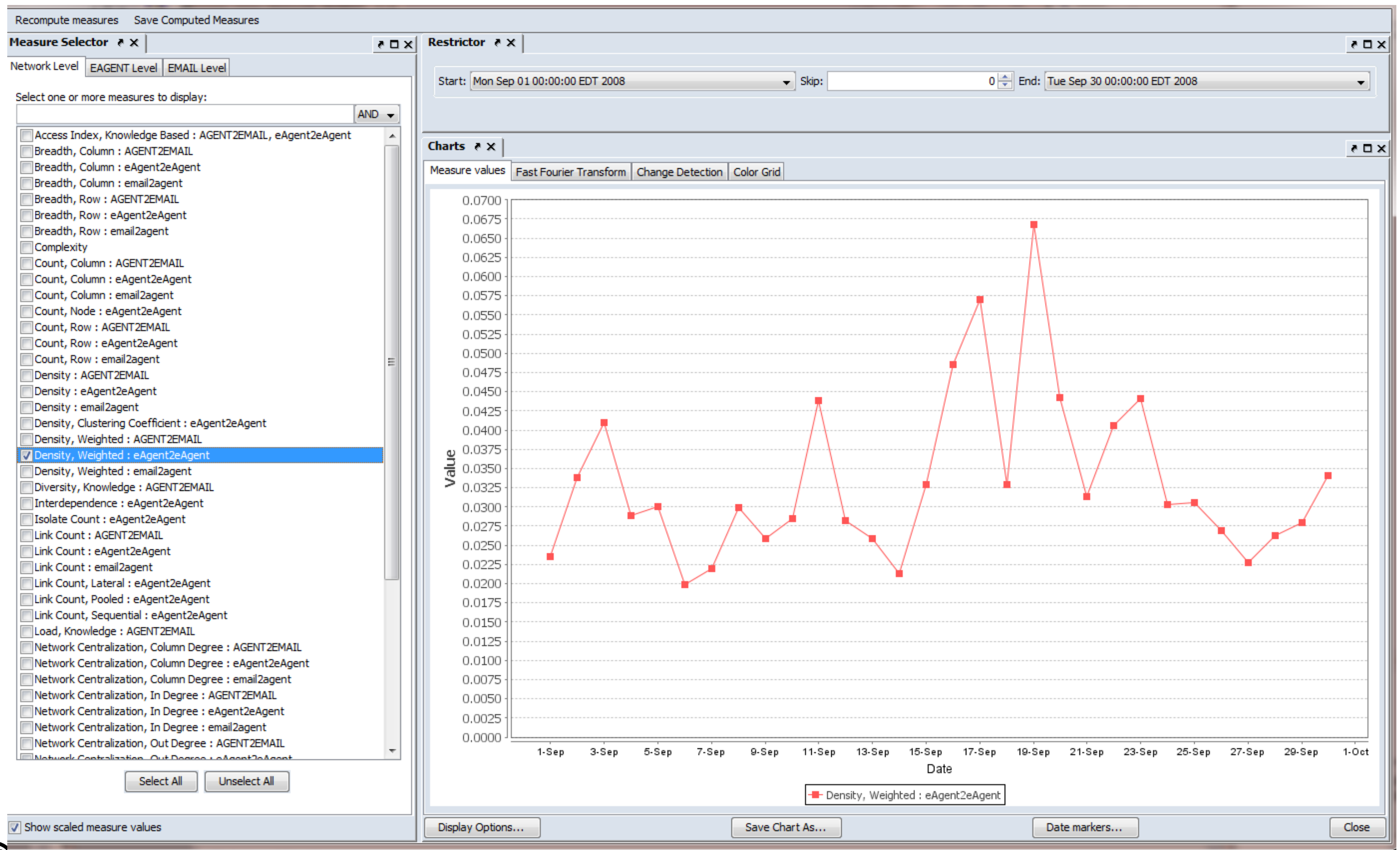


FFT Example Hands On (2 / 4)

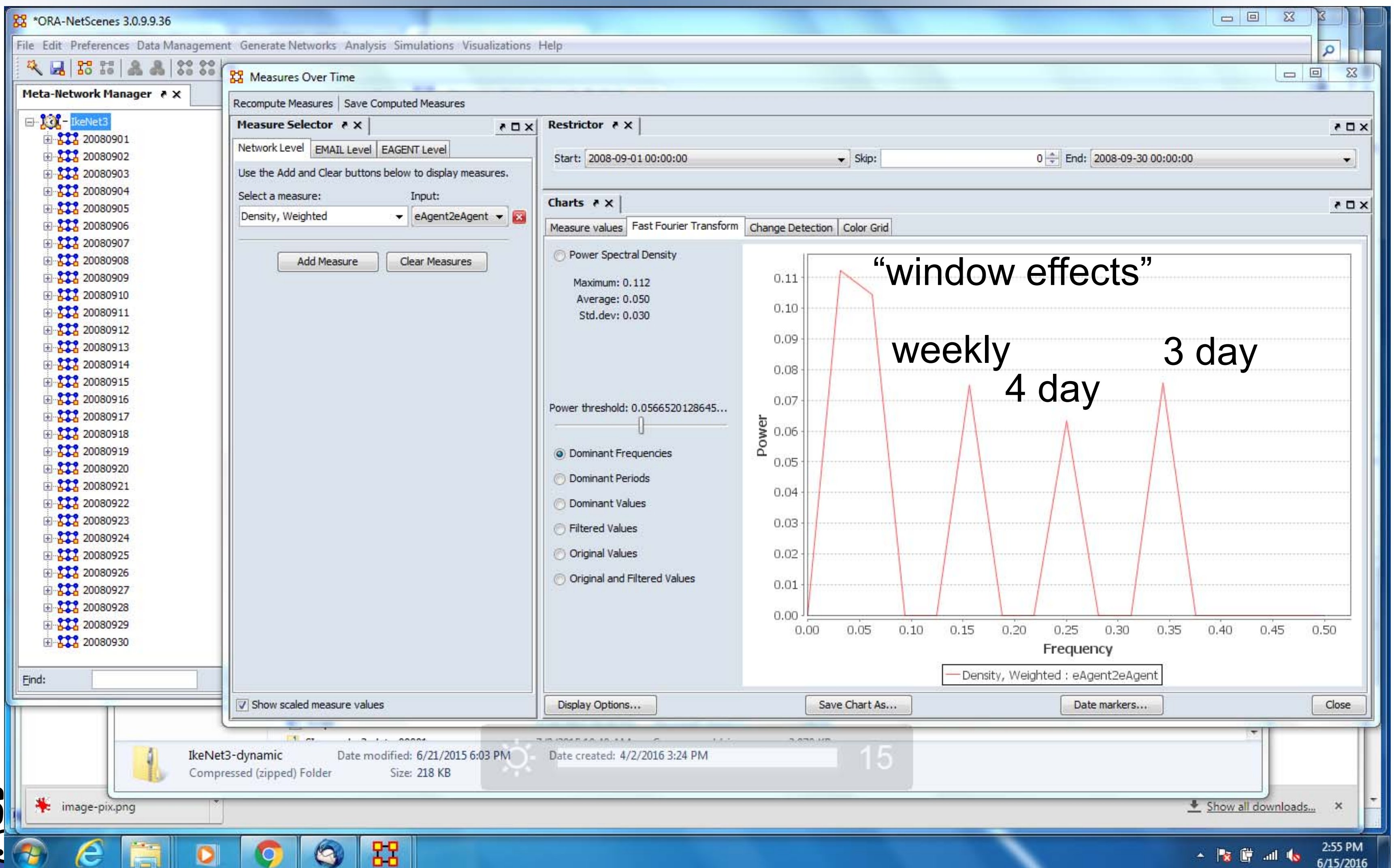


FFT Example Hands On (3 / 4)

Network Level / Density, Weighted / eAgent2eAgent



FFT Example Hands On (4/4)



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 filter out periodicity from longitudinal network data.
- Further exploration of wavelets may produce greater insights in to network dynamics.

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. In practice, 10+ network time points are preferred.
- No difference in number of required networks for each technique: CUSUM, EWMA, Scan Statistic, x-bar, eyeball
- Wavelet/Fourier based approach needs many more time periods

Limitations

- View findings on data with caution.
- 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.

Change Detection 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.

Conclusions

- Networks change over time
  - Approximate underlying network from available data
- Change detection
  - Detect occurrence of shocks i.e. change due to reasons exogenous to the network
- Fourier analysis can be used as a type of filter
  - Detect periodicity in over-time data