Carnegie Mellon

Ph.D. Program in Computation, Organizations & Society

CASOS

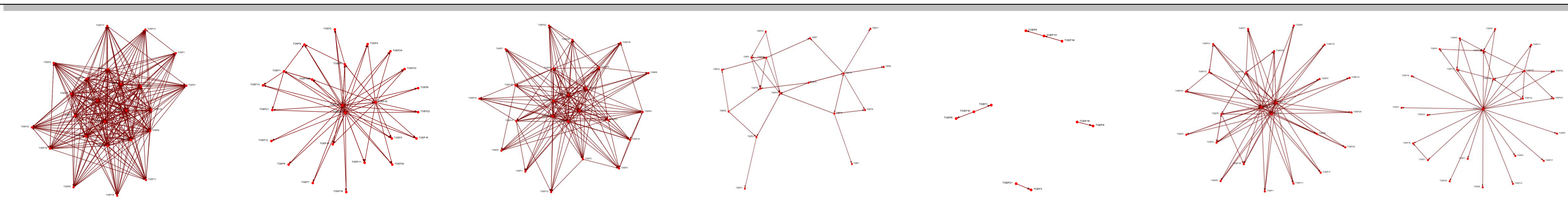
Social Network Change Detection

Ian McCulloh
imccullo@cs.cmu.edu

Prof. Kathleen M. Carley
kathleen.carley@cs.cmu.edu

Abstract

Changes in observed social networks may signal an underlying change within an organization, and may even predict significant events or behaviors. The breakdown of a team's effectiveness, the emergence of informal leaders, or the preparation of an attack by a clandestine network may all be associated with changes in the patterns of interactions between group members. The ability to systematically, statistically, effectively and efficiently detect these changes has the potential to enable the anticipation of change, provide early warning of change, and enable faster response to change. By applying statistical process control techniques to social networks we can detect changes in these networks. Herein we describe this methodology and then illustrate it using three data sets, of which the first is data simulated using multi-agent simulation, the second dealing with the email communications among graduate students and the third the perceived connections among members of al Qaeda based on open source data. The results indicate that this approach is able to detect change even with the high levels of uncertainty inherent in these data.



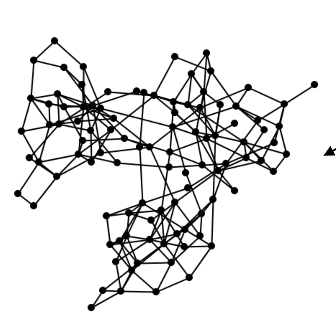
These networks represent e-mail communication over a period of 14 weeks. What is an average network? When did a statistically significant change occur?

Method: Social Network Measures

Network measures can be calculated from the entire graph or for each individual node. Centrality network measures such as betweenness and closeness are widely used for their easily applied practical applications in determining how information spreads through a social network. For illustration this paper will use one graph level measure, density (Coleman and Moré, 1983); and two individual node measures averaged over the graph, closeness (Freeman, 1979) and betweenness (Freeman, 1977). These are chosen because they are commonly used in the literature and represent a range of the types of measures available for change detection.

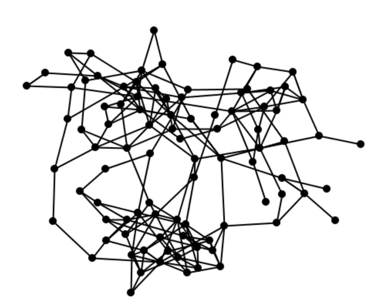
Performance on Simulated Data

Simulated data is used in order to inject an organizational change at a defined point in time. With simulated data, SNCD can be explored in a more controlled series of virtual experiments. For this initial investigation, we use a multi-agent simulation of a 100 node network, using the Construct simulation model.

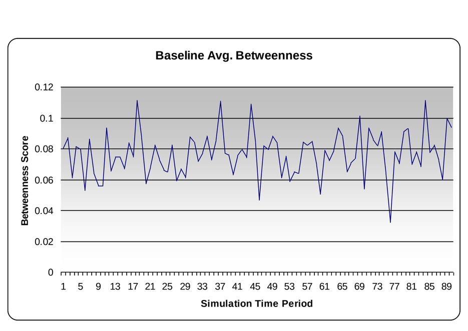


Before Change

After Change

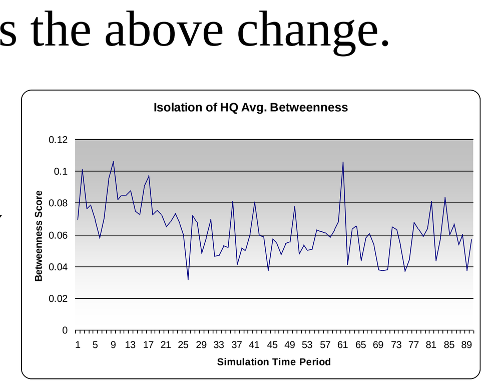


For illustration, the Average Betweenness measure over time is shown for both a network that does not change as well as the above change.

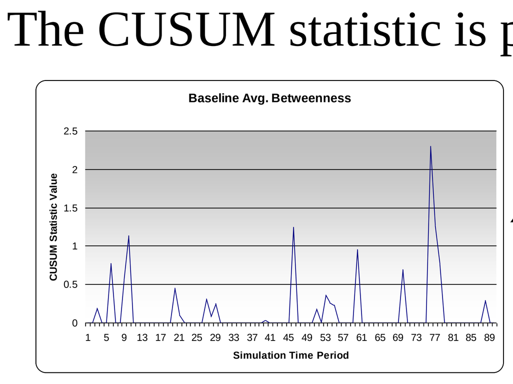


No Change

Change at 20

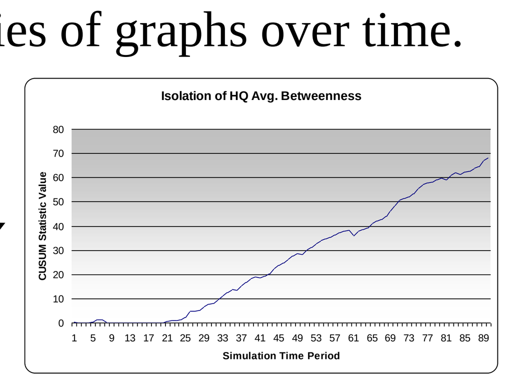


The CUSUM statistic is plotted for the same series of graphs over time.



No Change

Change at 20



Social Network Change Detection provides a much more dramatic indication of the change at time period 20.

Method: Statistical Process Control

Statistical Process Control (SPC) is a technique used by quality engineers to monitor industrial processes for change. Once a change has been detected, the engineers determine the most likely time the change occurred to re-examine and reset the process to avoid financial loss for the company. The CUSUM control chart is a widely used SPC method derived from the sequential probability ratio test (SPRT) (Page, 1961). The CUSUM control chart sequentially compares the statistic *C* against a control limit *h* until $C > h$. Since we are not interested in concluding that the network is unchanged, the cumulative statistic is given by,

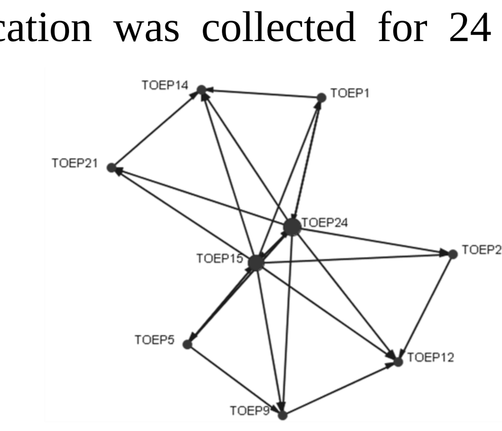
$$C_t^+ = \max\{\, 0, Z_t - k + C_{t-1}^+ \}$$

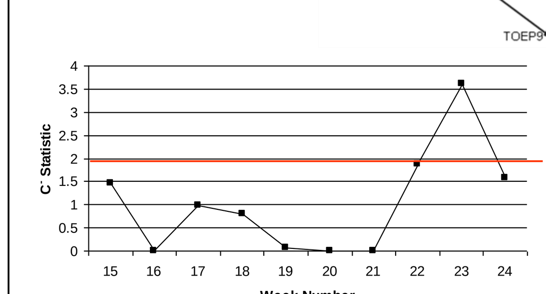
The statistic is compared to the constant control limit, *h*. If $C^+ > h$, then the control chart signals that an increase in a network measure has occurred. Since this rule only detects increases in the mean, a second cumulative statistic rule must be used to detect decreases in the mean, which signals a decrease in a network measure's mean when,

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

Performance on Real-World Data

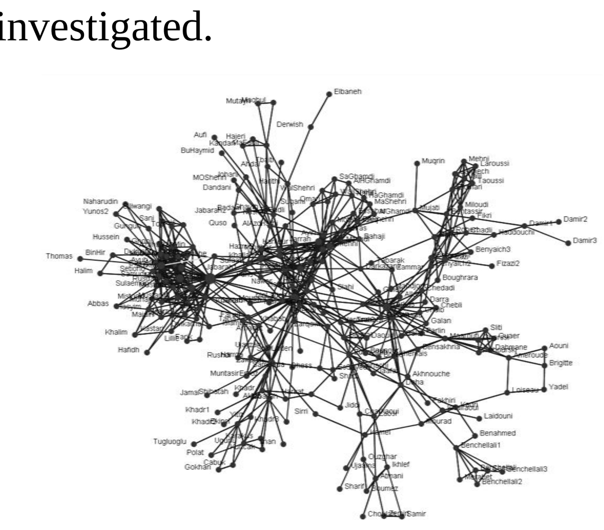
The Tactical Officer Education Program (TOEP) is a one-year graduate program run as a joint effort by the United States Military Academy (USMA) and Columbia University. Social network data on email communication was collected for 24 weeks.

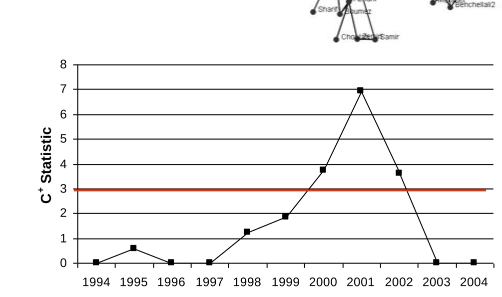




The CUSUM detects the week of the comprehensive exam. The change point was the week that study questions were sent to the students.

Open source snapshots of the annual communication between members of the al Qaeda organization from its founding in 1988 until 2004 are investigated.





The CUSUM signals a change in 2001. The change point, 1997, was the year that Al-Qaeda organized the Islamic Front and planned offensive operations.

This work is part of the Dynamics Networks project at the center for Computational Analysis of Social and Organizational Systems (CASOS) of the School of Computer Science (SCS) at Carnegie Mellon University (CMU). Support was provided, in part, by National Science Foundation (NSF) Integrative Graduate Education and Research Traineeship (IGERT) program, The Army Research Institute for the Behavioral and Social Sciences, the Army Research Lab, and the U.S. Military Academy. The views and proposal contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the, the National Science Foundation, the Army Research Institute, the Army Research Lab or the U.S. government.

institute for SOFTWARE RESEARCH