Visualizing the Evolution of Social Networks

Weizhong Zhu, Chaomei Chen, and Robert B. Allen

College of Information Science and Technology, Drexel University

wz32@drexel.edu, cc345@drexel.edu, rba@drexel.edu

Introduction

We are particularly interested in how a social network evolves over time. We construct diagrams of social networks as follows. Vertices in a network represent individuals. Relationships between the individuals in these networks cover a wide variety of types, for instance, exchanging emails and appearing in the same news. The strengths of such relationships are represented along edges. The evolution of a social network over a period of time, ranging from days in networks derived from news to years in networks associated with a working group, is studied by studying network snapshots taken from a series of consecutive time intervals within the entire period of time. Participating actors in the network are clustered and weighted based on attributes such as email send-reply pairs, interaction thread, concurrence frequency of name entitles in news and time-based centrality measures [1]. We demonstrate two examples of how emergent linkage patterns can be identified so as to improve our understanding of the social dynamics of the underlying group. One example is based on an email archive of a working group; the other is based on a set of synthesized news data. For example, one may want to find out the most active group member through email conversations during a given period. We will demonstrate how our approach can help not only to understand the roles of actors in terms of their influence and contributions in social networks, but also to generate hypotheses and evidence for visual analytics.

Time Series Analysis

Time series analysis divides a long period into a number of consecutive time frames. Betweenness is a way to measure the distance from a group to a node outside the group [2]. It reflects the influence of each actor in the social network. In this study, we expect to find emergent communication patterns that can lead to insights into the social dynamics of the underlying social network. The communication pattern is defined and computed as following:

Step1: Each person is presented as a vector of interaction frequencies in a time series. A linkage is defined as another person with who the person communicates in a certain time window.

Step2: Cosine similarities between these vectors are calculated.

Step3: Based on these similarity scores and certain threshold, a similarity network could be generated and displayed with small-world network model. Betweenness score are obtained through Betweenness analysis in Pajek [3], which is an implementation of Brandes' algorithm [4].

Evolution of the W3C URI Working Group

The World-Wide Web Consortium (W3C) working group email archive contains emails exchanged between group members between 1994 and 2004. Our focus is on the URI working group because the history of this working group is closely related to the growth of W3C itself. 4460 email transactions are included in this dataset. The results of time series analysis on this data collection are showed below. We can see how the social network changes in reaction to important events.

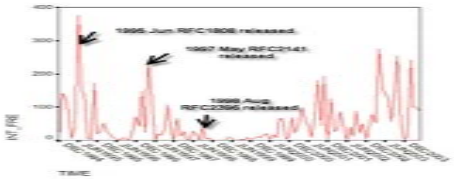

Fig.1 1994 Dec-2004 May Email Send-Reply Interaction Flows of W3C URI Working Group. It Turns Out these Peaks in the Diagram Matches with the Historical Events of URI Working Group. For Instance, the Highest Peak Matches the Date of URI RFC1808 Release.

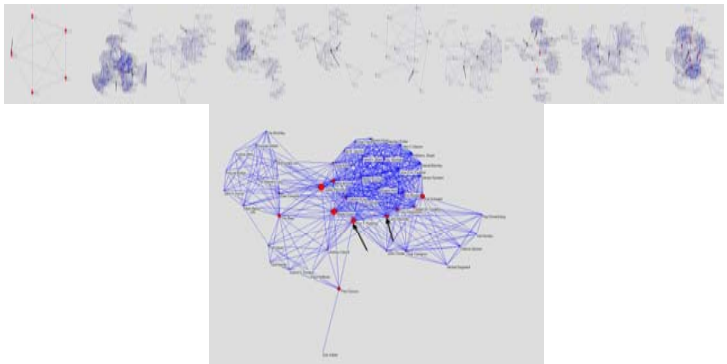

Fig.2 1994 Dec -2004 May Social Movement of W3C URI Working Group based on Yearly Time Series Analysis. The Sizes of Nodes Represent the Score of Time-based Betweenness Measure.

Rank	BW	T_BW
1	Larry Masinter	Larry Masinter
2	Roy T. Fielding	Roy T. Fielding
3	Michael Mealling	Dan Connolly
4	Martin Duerst	Michael Mealling
5	Dan Connolly	Martin Duerst
6	Paul Hoffman	Al Gilman
7	Al Gilman	Harald Tveit Alvestrand
8	Patrick Stickler	Paul Hoffman
9	Daniel LaLiberte	Daniel LaLiberte
10	Aaron Swartz	Leslie Daigle

Table1. A Comparison between the Top 10 Ranking Lists of Betweenness (BW) and Time-based Betweenness (T_BW) across the Ten Years. According to the Evaluation of Expert Search Task in TREC Enterprise 2005 Competition that Provides the List of Most Influential Members in the URI Working Group , Time-based Betweenness Performs Better.

Extracting Social Networks from Daily News

This task attempted to extract name entities (i.e., person name, location) from daily news with information extraction tools (e.g., LingPipe [5]) and measure the relationships between people based on concurrence analysis of these name entities. The testing set of the IEEE VAST contest is used to evaluate our visual dynamic analytic techniques. The dataset consists 1082 news stories from Alderwood's city newspaper from 2002 to 2004, along with some multimedia materials and background information. This is an on-going project and some preliminary results are showed below:

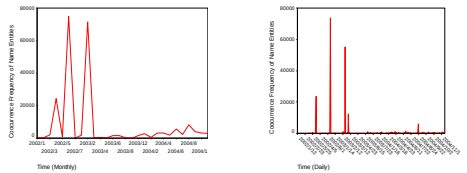

Fig.3 2002-2004 Monthly and Daily Person Name Entities Concurrence Flows of VAST Testing Set. It Turns Out that These Highest Peaks in the Diagram Reflect the Similar Events.

Fig.4 2002 May Social Movement of the Social Network Extracted from Daily News based on Monthly Time Series Analysis. The Sizes of Nodes Represent the Score of Betweenness Measure.

Conclusion

1. We developed a new approach to extend Betweenness centrality. With the additional predictive factor of time, Betweenness estimates the influence of actors in Social Networks based on the interaction flows and gives the better results.
2. Visual time series analysis depicts the dynamics of the communication patterns in email conversations or daily news and reflects the dynamic change of people's roles and social structures. It makes the results more understandable.
3. Trend analysis in this study shows that information interaction flows map the historical events or similar events and could be a predictor for events in the future.

Reference

1. Zhu, W.Z., Allen, R.B., and Chen, C.. Discovering Influential Members in Email Conversations Using Static and Dynamic Analysis with PageRank and Betweenness Centrality. Submitted for publication.
2. Freeman, L. C.. A set of measures of centrality based on betweenness. *Sociometry*, 40:35-41,1977.
3. Batagelj, V., and Mrvar, A.. Pajek - Analysis and Visualization of Large Networks. In Jünger, M., Mutzel, P., (Eds.) *Graph Drawing Software*. Springer, Berlin, 2003.
4. Brandes, U., A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology* 25(2):163-177, 2001.
5. http://www.alias-i.com/lingpipe/