Sociality in Web Forum

Bing Wu
School of Economics and Management, Tongji Univeristy, Shanghai, China

Keywords: Sociality, Web Forum, Exponential Random Graph, Sentiment.

Abstract: We use exponential random graph to examine the generative processes that give rise to widespread sociality patterns in web forums. We apply the methods to Yahoo finance Wal-Mart message board from 1999 to 2008 to investigate authors' propensities to establish relationship increase by activity. Research results shows that although having the lowest percentage in Web forums, medium activity authors are more social than high activity authors, which shows the consistent pattern with few core members contributing the majority of content. Considering sentiment, objective authors have the highest sociality, followed by negative subjective authors, which is proportional to the constitution of author sentiment group. Similar situation happened in author sociality by class, authors' tendency to establish relationship is quite different. We conclude with a discussion of how exponential random graph may contribute to our understanding of social interaction structure and the processes that create it.

1 INTRODUCTION

A long line of research has shown the importance of social behavior in web forum, and many of the same themes continue to be explored today. Despite the proliferation of web forums, not much is currently known regarding sociality patterns. Tools are needed to provide a means of identifying and assessing the characteristics of social interaction patterns, and how social interaction patterns further our understanding of the network-scale sociality in web forums. Currently Exponential Random Graph Models (ERGM) provides a better way for construction of a comprehensive and dynamic model for the following reasons. Accordingly our research goals are to develop a general framework for understanding sociality in web forum based on ERGM and then to provide a statistically principled approach to the systematic exploration of several features simultaneously. And potential applications of this study can be useful for managing a community and its members as well (Campbell and Cecez-Kecmanovic, 2011).

2 LITERATURE REVIEW

To form a comprehensive understanding of our study problem, a literature review was conducted in two aspects, including social interaction network in web forum and corresponding modeling methods.

2.1 Social Network in Web Forum

A web forum details a sequence of discussions through threads of messages. Therefore, web forums create implicit ties that connect senders and receivers in what is often called a "reply network" or "reply graph." The reply networks can be analyzed to identify important relationships (Lee and Lee, 2006). Consequently the social relationships in web forums can be described as social networks, which consist of finite sets of authors and the relationships between them.

As key indicators can be used to understand the status of social network in web forums, we examine and classify previous papers in terms of attributes in web forums, including: author activity, message sentiment and topics

To establish of any relationships between measures of web forum discussions and author's behavior. The author activity analysis is based on Activity Theory, which is the idea that human activity is a dialectic relationship between subject (person) and object (purpose) (Levy, 2008). In Activity theory, the constituents of activity are not fixed, but can dynamically change with conditions (Kaptelinin and Nardi, 1997). The unit of analysis in

Wu B..
Sociality in Web Forum.
DOI: 10.5220/0004425000540057
In *Proceedings of the 15th International Conference on Enterprise Information Systems* (ICEIS-2013), pages 54-57
ISBN: 978-989-8565-61-7

Activity Theory is an activity directed at an object which motivates activity.

Sentiment analysis (Das and Chen, 2007); (Antweiler and Frank, 2004) web forum discussions is considered to deal with computational evaluation of expressions of opinion, sentiment, and subjectivity in text (Pang and Lee, 2008). Broadly speaking, there are two types of sentiment analysis approaches. Machine learning-based approaches require a pre-coded training dataset which consists of texts and labeled sentiments. Obtaining the training data and training is a time intensive task (Aue and Gamon 2005). In contrast, lexicon-based approaches are faster provided an appropriate dictionary is available. One of the advantages is that they can take negation (e.g. not) and intensification (very) into account. Some well-known lexicons include Senti Word Net (SWN) (Esuli and Sebastiani 2006) and the Harvard-IV-4 dictionary (Tetlock, 2007; 2008).

2.2 Corresponding Modelling Methods

Empirical studies validate a model with real data. The model is parameterized based on real data, and the interaction process depicted with estimated parameters is compared to the real interaction process. In empirical research of web forums, findings show that there is a consistent pattern of participation with a few core members contributing the majority of content, many peripheral members contributing infrequently, and a large number of lurkers (Nonnecke and Preece, 2000) who benefit by overhearing the conversations of others (Hansen, 2009). The nature the conversation depends largely on the type of web forum. (Hansen et al., 2010).

Regression analysis evaluates models by testing their feasibility and equilibrium status based on mathematical variables and models. Results from analyses demonstrate that the social context, including pre-existing social networks, groups, and intergroup boundaries, significantly constrained the flow of information interaction pattern across intercultural CMC (computer mediated communication, CMC) groups. And in addition the influence of the social context on CMC collaboration could be moderated by other contingent factors such as national culture and individuals' expectancies of Internet use (Hichang Cho, Jae-Shin Lee, 2008)

As social network analysis provides powerful ways to summarize networks and identify key people or other objects that occupy strategic locations and positions within the matrix of links. The threaded conversation structure in web forums leads itself well to social network analysis. The basic properties of social networks include the size, density, centrality, degree, reach ability (Hanneman, 2001), connectivity (Stocker et al., 2001) and multiplicity (Emirbayer and Goodwin, 1994).

Pioneering contributions to model social interaction using ERGM have been made in the study of sociological implications of friendship network. ERGM was used to model friendship formation as a selection process constrained by individual's sociality (propensity to make friends), selective mixing in dyads (friendships within race, grade, or sex categories are differentially likely relative to cross-category friendships), and closure in triads (a friends' friends are more like to become friends), given local population composition), so that socio demographic structure and the processes that creates it can be understood (Goodreau, 2007); (Goodreau, 2009). Furthermore in order to acquire sociological implications for single-gender and cross-gender influences on teenagers' behavior, ERGM framework including social network techniques were used to examine gender clustering in a complete network of teenagers and their friends (Kirke, 2009).

3 RESEARCH DESIGN

The Yahoo! Finance Forum is chosen as the test bed due to the large amount of message postings in this platform. Wal-Mart was selected due to its prominence in the market, societal presence, and active collection of stakeholder groups. As social interaction pattern in web forum should be considered over the long-term, the time span for analysis covers from January 1999 to June 2008.

Each author's sentiment can be obtained by the average sentiment of all messages posted or replied to by whom. In this study, sentiment of each author will be normalized as three sentiment groups, negative subjective author (ps1), objective author (ps2) and positive subjective author (ps3).

The author group depends on the result of key-phrase extraction for all of his/her all messages posted or replied to by whom. Thereby the author class is employees (pt1), investors (pt2) and customers (pt3) in Yahoo finance forum.

In this study we normalize activity as three activity groups delegating low activity (pa1), medium activity (pa2) and high activity (pa3).

4 RESEARCH RESULTS

Since the dataset spans 10 years, from Jan 1999 to June 2008, we will use a graphical display, i.e., box plots. Box plots enclose the interquartile range of the data in a box displaying the median, giving an idea about the center, variability, and degree of asymmetry of a sample. Interquartile range is between the 75th percentile (upper quartile) and the 25th percentile (lower quartile). Box plots are used in this study to present the variation for each measures group with a given degree across the 10-years span, the vertical solid line in each horizontal box denotes the median value of each given group with lower quartile and upper quartile on both ends. Figure 1 charts the sociality estimates from the ERGM. As the vertical solid line in each horizontal box is on the right side of the 0 axis, sociality estimates for each category are positive.

In the lower part of Figure 1, based on low activity authors (pa1), the positive median values of medium activity authors (pa2) and high activity authors (pa3) mean authors' propensity to establish relationship increases by activity. This is consistent with the pattern of a few core members contributing the majority of content, similar to research results of Nonnecke and Preece's (2000).

The positive median values of negative subjective authors (ps1) and objective authors (ps2) denote authors' preference to post messages increases based on objectivity (Figure 1). Thus, objective authors (ps2) have the highest sociality, followed by negative subjective authors, which is proportional to the constitution of the author sentiment group.

The upper part of Figure 1 shows that authors' tendency to post messages is the strongest among investor authors (pt2), then employee authors (pt1), based on the positive median values. This tendency is consistent with the proportions of the author group as well.

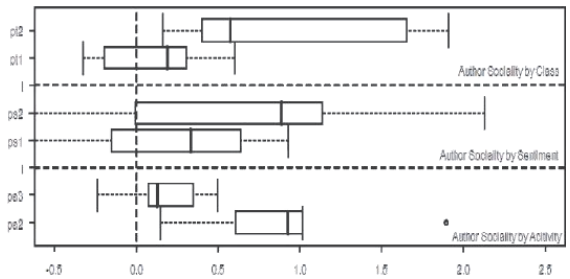

Figure 1: Sociality estimates plotted across 10 years.

5 CONCLUSIONS

Based upon these findings, we state the conclusions and implications of this study below. In this case study, other attributes in addition to author activity, author sentiment and author class, such as author's residence time, reply interval, etc., could be added, which will be useful for further exploration of dynamic social interaction pattern in future. In addition, selective modeling by vertex attributes, holistic feature modeling and goodness of fit will be in future ERG modeling, so that complex social interaction networks in Web forum can be explored by most appropriate ERG model. Meanwhile an iterative exploratory technique of progressively increasing the model complexity should be used in ERG modeling process. Moreover other experiments for different test beds are on the agenda of future direction.

ACKNOWLEDGEMENTS

This work was supported in part by NSFC under Grant Nos. 71071117.

REFERENCES

Antweiler, W. and Frank, M.: Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, vol. 59, no.3, pp. 1259-295, 2004.

A. Aue and M. Gamon: Customizing sentiment classifiers to new domains: A case study. *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, 2005.

B. Nonnecke and J. Preece: Lurker Demographics: Counting the Silent.*CHI Conference*, pp. 73-80, 2000.

Das, S. and Chen, M. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, vol.53, no.9, pp. 1375-1388, 2007.

Deirdre M. Kirke. Gender clustering in friendship network: some sociological implications. *Methodological Innovations online* (4), pp. 23-26, 2009.

Derek Hansen et al.: Analyzing Social Media Networks with NodeXL. *Morgan Kaufmann*, 2010.

D. L. Hansen. Overhearing the crowd: an empirical examination of conversation reuse in a technical support community. *Proceedings of the Fourth international Conference on Communities and Technologies*. ACM, New York, NY, pp.155-164, 2009.

Emirbayer, M., & Goodwin, J. Network analysis, culture, and the problem of agency. *American Journal of*

Sociology, vol.99, no.6, pp.1411-1454, 1994.

Esuli, A. and Sebastiani, F. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of Conference on Language Resources and Evaluation*, 2006.

Hannerman, R. Introduction to Social Network Methods, retrieved from http://www.faculty.ucr.edu/~hanneman/, 2001.

Hichang Cho, Jae-Shin Lee. Collaborative information seeking in intercultural computer-mediated communication groups: testing the influence of social context using social network analysis. *Communication Research*, vol.35, no.4, pp. 548-573, 2008.

John Campbell, Dubravka Cecez-Kecmanovic. Communicative practices in an online financial forum during abnormal stock market behavior. *Information & Management, vol.48*, no.1, pp.37-52, 2011.

Kaptelinin, V., Nardi, B.A. Activity theory: basic concept and application, *CHI*, 1997.

Pang, B. and Lee, L. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, vol.2, no.1/2, pp. 1-135, 2008.

Sang Jun Lee, Zoonky Lee. An Experimental Study of Online Complaint Management in the Online Feedback Forum. *Journal of Organizational Computing and Electronic Commerce*, vol.13, no.1, pp. 65-85, 2006.

Steven M Goodreau. Advances in exponential random graph (p *) models applied to a large social network. *Social networks*, vol. 29, no.2, pp. 231-248, 2007

Steven M. Goodreau, James A. Kitts, Martina Morris. Birds of a Feather, Or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography*, vol.46, no.1, pp.103-125, 2009.

STOCKER et al. Consensus and cohesion in simulated social networks. *Journal of Artificial Societies and Social Simulations*, http://jasss.soc.surrey.ac.uk/4/4/5.html, vol. 4, no.4, 2001.

Tetlock P. Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance*, vol.62, no.3, pp. 1139-1168, 2007.

Tetlock P., Teschansky M. and Macskassy S. More than words: Quantifying language to measure firm's fundamentals. Journal of Finance, vol.63, no.3, pp. 1437-1467, 2008.

Xiaojun, W. and Jianguo, X. Exploiting neighborhood knowledge for single document summarization and keyphrase extraction. *ACM Transactions on Information Systems*, vol.28, no.2, pp.1-34, 2010.

Yair Levy. An empirical development of critical value factors (CVF) of online learning activities: An application of activity theory and cognitive value theory. *Computers & Education*, vol.51, no.4, pp.1664–1675, 2008.

Zachary M. Saul_and Vladimir Filkov. Exploring Biological Network Structure Using Exponential Random Graph Models. *Bioinformatics Advance* Access published July 20, pp. 1-7, 2007.