

SNAKDD 2008 - Social Network Mining and Analysis Post-Workshop Report

Haizheng Zhang and
Marc Smith
Microsoft
One Microsoft Way
Redmond, WA 98052
hazhan@microsoft.com

C. Lee Giles and John
Yen and Henry Foley
College of Information Science
and Technology
Pennsylvania State University
University Park, PA 16802
jyen,giles@ist.psu.edu

ABSTRACT

In this report, we summarize the contents and outcomes of the recent SNAKDD 2008 workshop on Social Network Mining and Analysis that was held in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), August 24-27, 2008, in Las Vegas, Nevada. The SNAKDD 2008 is the second in a successful series of workshops on social network mining and analysis. It solicited papers on broad overview subject "Social Network Mining and Analysis". The workshop drew a large number of participants from academia and industry. In this report we reflect on the results of the workshop, as captured in the regular paper session as well as the poster session.

Keywords

Social network analysis, community discovery, web mining, evolution

1. INTRODUCTION

The second SNAKDD '2008 is the second in a successful series of workshops on social network mining and analysis. It aims to bring together practitioners and researchers with a specific focus on the emerging trends and industry needs associated with a variety of social networking systems. The first SNAKDD workshop was co-held with WebKDD 2007 at SIGKDD 2007 conference in San Jose, California.

2. THEME

The second SNAKDD workshop focuses on the emerging trends and industry needs associated with a variety of social networking systems, encompassing lessons learned over the past years and new challenges for the years to come. The workshop received submissions including experimental and theoretical work on social network mining and analysis.

3. SUBMISSIONS

There were 32 papers submitted to SNAKDD 2008. Each of the papers was strictly reviewed by at least three reviewers from the SNAKDD 2008 program committee. Due to the quality and diversity of the submissions, we accepted 11 papers as regular papers and 8 as poster papers. The workshop papers are divided into four categories based on their topics: (1) Social network models, (2) Community discovery and extraction, (3) Social dynamics and evolution, and (4) Social network applications.

4. WORKSHOP

The workshop attracted interests from a large number of conference participants. Among them included participants from academia and government agencies, Web related industry, as well as practitioners from other companies.

4.1 Session 1. Social Network Models

The paper, "Combining Transitive Trust and Negative Opinions for better Reputation Management in Social Networks" [5], authored by Debora Donato, Stefano Leonardi, and Mario Panicia, focuses on reputation management model and investigates the role of users' negative opinions in decentralized systems. In this paper the authors propose a new approach to combine negative and positive opinions expressed by peers to reach a global consensus on trust and distrust values for each peer of the network. The experiments show that these strategies outperform previous strategies in terms of number of successful transactions against a large set of sophisticated threat attacks posed by coalitions of malicious peers.

The paper, "Pseudolikelihood EM for Within-Network Relational Learning", authored by Rongjing Xiang and Jennifer Neville [17], compares three in-sampling learning approaches [17], including disjoint learning with independent inference, disjoint learning with collective inference, and joint learning with collective inference, by developing a pseudolikelihood-EM method that facilitates learning collective classification models with both unlabeled and labeled data in the same network. The authors discuss the experiment results on the three approaches.

The paper, "Nonparametric Relational Learning for Social Network Analysis" [18], by Zhao Xu, Volker Tresp, Shipeng Yu, and Kai Yu, discusses how the infinite hidden relational

model (IHRM) model can be used to model and analyze social networks, where each edge is associated with a random variable and the probabilistic dependencies between these variables are specified by the model based on the relational structure. The hidden variables are able to transport information such that non-local probabilistic dependencies can be obtained. The IHRM provides effective relationship prediction and cluster analysis for social networks. The experiments demonstrate that this model can provide good prediction accuracy and capture the inherent relations among social actors.

4.2 Session 2: Community discovery and extraction

The paper, “Adequacy of Data for Characterizing Caller Behavior” [15], by Santi Phithakkitnukoon and Ram Dantu, investigates phone users’ behaviors and constructs a probability models based on callers call arrival, inter-arrival, and talk time from the call logs in order to predict or estimate the future observation. The authors focus on the question that how much historical data is adequate for prediction purpose.

The paper, “Social Topic Models for Community Extraction” [14], by Nishith Pathak, Colin DeLong, Arindam Banerjee, and Kendrick Erickson, presents a Bayesian generative model for community extraction which incorporates both the link and content information in the social network. The model assumes that actors in a community communicate on topics of mutual interest, and the topics of communication determine the communities. The model is evaluated on the Enron email corpus and the results demonstrate that the model is able to extract well-connected and topically meaningful communities.

4.3 Session 3. Social dynamics and evolution

The paper, “Communication Dynamics of Blog Networks” [9], by Mark Goldberg, Malik Magdon-Ismael, Stephen Kelley, Konstantin Mertsalov, and William Wallace, studies the communication dynamics of Blog networks by exploring the Russian section of LiveJournal. The two fundamental questions that this paper is concerned with include (1) what models adequately describe such dynamic communication behavior; and (2) how does one detect changes in the nature of the communication dynamics. The paper leverages “stable statistics” in their research in disclosing the dynamics of the networks and characterizing the locality properties.

The paper, “Ontology Engineering and Feature Construction for Predicting Friendship Links in the Live Journal Social Network” [2], by Vikas Bahirwani, Doina Caragea, Waleed Aljandal, and William Hsu, addresses the problem of building an interest ontology and predicting potential friendship relations between users in Live Journal, using features constructed based on the interest ontology. The goal of this paper is to organize users’ interests in an ontology and to use the semantics captured by this ontology to improve the performance of learning algorithms at predicting if two users can be friends. We have designed and implemented a “hybrid clustering” algorithm, which combines hierarchical agglomerative and divisive clustering paradigms, and automatically builds the interest ontology. Furthermore, we have

explored the use of this ontology to construct interest-based features and shown that the resulting features improve the performance of various classifiers for predicting friendship links.

The paper, “Wikis as Social Networks: Evolution and Dynamics” by Ralf Klamma and Christian Haasler [11], analyzes wikis as social networks by applying dynamic network analysis techniques. In particular, the authors handle the complex data management problems arising when dealing with different wiki engines and different sizes of wiki dumps. The analysis and visualization of evolving wiki networks allow wiki stakeholder to research the social dynamics of their wikis.

The paper, “Overcoming Resolution Limits in MDL Community Detection” [4] by Karl Branting, studies the resolution limit problem with two compression-based algorithms that were designed to overcome such limit. The author identifies the aspect of each approach that is responsible for the resolution limit and proposes a variant, SGE that addresses this limitation. The paper demonstrates on three artificial data sets that (1) SGE does not exhibit a resolution limit on graphs in which other approaches do, and that (2) modularity and the compression-based algorithms, including SGE, behave similarly on graphs not subject to the resolution limit.

4.4 Session 4. Social network applications

The paper, “Group Proximity Measure for Recommending Groups in Online Social Networks” [16], by Barna Saha and Lise Getoor. investigates the problem of defining proximity measure between groups/communities of OSN in order to reveal new insights into the structure of the network. The authors use this proximity measure in a novel way with other structural features of online groups to predict the evolving state of the network and propose a new framework for recommending groups to users of OSNs. The authors evaluate the effectiveness of our methods in three real OSNs: Flickr, Live Journal and You Tube.

The paper, “A Social Query Model for Decentralized Search” [3], by Arindam Banerjee and Sugato Basu, introduces a novel Social Query Model (SQM) for decentralized search, which factors in realistic elements such as expertise levels and response rates of node. In the context of the model, the authors design a query routing policy that is simultaneously optimal for all nodes, in that no subset of nodes will jointly have any incentive to use a different local routing policy. For computing the optimal policy, the paper presents an efficient distributed approximation algorithm that is almost linear in the number of edges in the network. Extensive experiments on both simulated random graphs and real small-world networks demonstrate the potential of our model and the effectiveness of the proposed routing algorithm.

4.5 Poster Session

Due to the number of high quality submissions that we have received and limited workshop time, we create a poster session in addition to the regular paper session. There were 8 papers that were presented poster papers. The topics of these papers include: graph mining, such as “A Resampling Technique for Relational Data Graphs” [6] (by Hoda El-

dardiry and Jennifer Neville); and “Finding Spread Blockers in Dynamic Networks” [10] (by Habiba, Yintao Yu, Tanya Berger-Wolf, and Jared Saia). , community discovery and detection, such as the paper, “Community Detection using a Measure of Global Influence” [8] (by Rumi Ghosh and Kristina Lerman), and the paper, “Large-Scale Principal Component Analysis on LiveJournal Friends Network” [12] (by Miklos Kurucz, Andras A. Benczur, and Attila Pereszlenyi); as well as prediction and classification, including the paper, “Using Friendship Ties and Family Circles for Link Prediction” [19] (by Elena Zheleva, Lise Getoor, Jennifer Golbeck, and Ugur Kuter), “Comparative analysis of tag suggestion algorithms” [13] (by Steffen Oldenburg, Martin Garbe, Clemens Cap and Lukas Zielinski), “A Shrinkage Approach for Modeling Non-Stationary Relational Autocorrelation” [1], (by Pelin Angin and Jennifer Neville); and “Leveraging Label-Independent Features for Classification in Sparsely Labeled Networks: An Empirical Study” [7] (by Brian Gallagher and Tina Eliassi-Rad).

5. CONCLUSIONS

The SNAKDD 2008 workshop continued a tradition of exposing strong and ever more innovative research in critical areas of social network mining and analysis, including social network modeling, growth and dynamics, applications, and community discovery. The workshop also rendered discussions that helped raise several questions and bring to light several interesting future directions and challenges in the area of social network mining and analysis.

Acknowledgment

We would like to thank the authors of all submitted papers. Their creative efforts have supported the competitive technical program of SNAKDD 2008. We would also like to express our gratitude to the members of the Program Committee for their vigilant and timely reviews, namely: Lada Adamic, Aris Anagnostopoulos, Arindam Banerjee, Tanya Berger-Wolf, Yun Chi, Aaron Clauset, Isaac Council, Tina Eliassi-Rad, Lise Getoor, Mark K. Goldberg, Larry Holder, Andreas Hotho, Kristina Lerman, Wei Li, Yi Liu, Gueorgi Kossinets, Ramesh Nallapati, Jennifer Neville, Dou Shen, Bingjun Sun, Jie Tang, Andrea Tapia, Alessandro Vespigani, Xuerui Wang, Stefan Wrobel, Michael Wurst, Xiaowei Xu. We extend our thanks and encouragements to the entire Social network analysis community, including all those who attended the workshop and contributed in the interesting discussions.

6. REFERENCES

[1] P. Angin and J. Neville. A shrinkage approach for modeling non-stationary relational autocorrelation. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[2] V. Bahirwani, D. Caragea, W. Aljandal, and W. Hsu. Ontology engineering and feature construction for predicting friendship links in the live journal social network. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining (KDD 2008), August 2008.

[3] A. Banerjee and S. Basu. A social query model for decentralized search. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[4] K. Branting. Overcoming resolution limits in mdl community detection. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[5] D. Donato, S. Leonardi, and M. Paniccia. Combining transitive trust and negative opinions for better reputation management in social networks. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[6] H. Eldardiry and J. Neville. A resampling technique for relational data graphs. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[7] B. Gallagher and T. Eliassi-Rad. Leveraging label-independent features for classification in sparsely labeled networks: An empirical study. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[8] R. Ghosh and K. Lerman. Community detection using a measure of global influence. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[9] M. Goldberg, M. Magdon-Ismael, S. Kelley, K. Mertsalov, and W. Wallace. Communication dynamics of blog networks. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

[10] Habiba, Y. Yu, T. Berger-Wolf, and J. Saia. Finding spread blockers in dynamic networks. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.

- [11] R. Klamma and C. Haasler. Wikis as social networks: Evolution and dynamics. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [12] M. Kurucz, A. A. Benczur, and A. Pereszlenyi. Large-scale principal component analysis on livejournal friends network. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [13] S. Oldenburg, M. Garbe, C. Cap, and L. Zielinski. Comparative analysis of tag suggestion algorithms. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [14] N. Pathak, C. DeLong, A. Banerjee, and K. Erickson. Social topic models for community extraction. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [15] S. Phithakkitnukoon and R. Dantu. Adequacy of data for characterizing caller behavior. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [16] B. Saha and L. Getoor. Group proximity measure for recommending groups in online social networks. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [17] R. Xiang and J. Neville. Pseudolikelihood em for within-network relational learning. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [18] Z. Xu, V. Tresp, S. Yu, and K. Yu. Nonparametric relational learning for social network analysis. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
- [19] E. Zheleva, L. Getoor, J. Golbeck, and U. Kuter. Using friendship ties and family circles for link prediction. In *Proceedings of SNAKDD 2008: KDD Workshop on Social Network Mining and Analysis, in conjunction with the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008)*, August 2008.
-