

Research Statement

LIM Ee-Peng
School of Information Systems, Singapore Management University
Tel: (65) 6828-0781; Email: eplim@smu.edu.sg
Updated on 22 (Day) 1 (Month) 2012 (Year)

Background

- Provides a context of your research interests. Why does it matter?

My ongoing research is motivated by the needs to uncover knowledge about user behaviors in the data traces of user-user and user-item interactions left at social media sites and in company databases. Such kind of knowledge is of utmost importance to the understanding of user preferences, their mutual influence and to the design of new “socially friendly” products and services.

With the establishment of Living Analytics Research Centre (LARC), my research scope has further expanded to include more industry relevant problems and co-experiments. I have already started several projects in collaboration with the LARC’s industry partners, including Buzzcity, Brandtology and JamiQ. In these projects, I collaborate with several SMU/CMU faculty members to address the research problems together.

In the first year of LARC, I have also spent significant time helping the centre to establish the procedures and guidelines for coordinating research projects, resource allocation, and CMU exchange. In addition to ensuring the new LARC’s physical lab is conducive for research, I am also overseeing all the research activities happening in the centre. As the LARC director, I keep a close link with the industry and government bodies, to organize two steering committee meetings each year, and to report the centre’s research progress to MDA IDMPO on a regular basis.

In 2011, I also successfully organized the *Third International Conference on Social Informatics (SocInfo2011)* in SMU. SIS is the main organizer of the conference and several colleagues from SMU, NUS and NTU have also committed their time and efforts to organize the event. We attracted more than 200 participants from both computing and social science disciplines to join the conference. The conference program featured highly renowned social computing researchers as keynote speakers. It also included the usual research paper presentations, demos and posters. SocInfo2011 received generous support from AFOSR/AOARD, MDA, Lee Foundation, Poland Japan Institute of Information Technology (PJIT) and Singapore Internet Research Centre (SiRC).

In the past one year, my main focus is get the LARC research projects to start off well. I am the principal investigator of two steering committee approved projects, namely the (a) actionable pattern mining for dynamic social networks, and (b) Twitalytics: Interactive Social Analytics for Twitter Data. Both projects have started off quite well with the respective industry partners. Outside LARC, I participated in a new Defence Innovation Research Programme (DIRP) project proposal (entitled "Intelligent Network Search Using Interactive Topic Discovery") as a collaborator and the proposal has been awarded funding of by DSTA in November 2011.

Research Areas

- If you have multiple research areas, describe each area under a separate section.
- For each research area, please describe
 - o what you have been doing recently and currently. What are the research problems and the proposed solutions? How does your research contributed to your field.
 - o In what direction do you plan to go. What are the research problems you plan to attack? Why are these important or interesting problems to your field, or area of practice, or to society at large?

Social Network Mining. In social network mining, our objective is to discover user and object characteristics, and interesting link patterns from multidimensional and dynamic networks.

Users of online social networks not only create and maintain friendships, they also utilize the social networking sites to gain access to a wide range of products, advertisements and services. This creates a trend of information access which sees the users increasingly rely not only on their own preferences, but also on friends'. We call this new way of information access *social information processing*. In this research, we investigate the effects of social correlation on users' adoption of items. Given a user-user social graph and an item-user adoption graph, we seek to answer the following questions: 1) whether the items adopted by a user correlate to items adopted by her friends, and 2) how to incorporate social correlation in order to improve prediction of unobserved item adoptions. We propose the Social Correlation model based on Latent Dirichlet Allocation (LDA) that decomposes the adoption graph into a set of latent factors reflecting user preferences, and a social correlation matrix reflecting the degree of correlation from one user to another. This matrix is learned (rather than pre-assigned), has probabilistic interpretation, and preserves the underlying social network structure. We further devise a Hybrid model that combines a user's own latent factors with her friends' for adoption prediction. Experiments on Epinions and LiveJournal data sets show that our proposed models outperform the approach based on latent factors only (LDA).

Twitter is another online social network with large number of users following others so as to receive messages from their followees. Twitter users can use mobile phones or desktop machines to generate short messages to be shared with others in realtime. Twitter has seen heavy usage in many recent international events including Japan earthquake, Iran election, etc. In such events, many tweets may become viral for different reasons. In this paper, we study the virality of socio-political tweet content in the Singapore's 2011 general election (GE2011). We collected tweet data generated by

about 20K Singapore users from 1 April 2011 till 12 May 2011, and the follow relationships among them. We introduce several quantitative indices for measuring the virality of tweets that are retweeted. Using these indices, we identify the most viral messages in GE2011 as well as the users behind them. Some of the indices have been further extended to allow to determine user characteristics.

Trust Mining. The huge presence of online communities on the Web and high volume of online transactions have created fresh possibilities for businesses to expand their customer base, to reach out to niche customer markets and for individuals to find better products or services. Businesses will have to acquire knowledge about trust relationships among users and to develop new applications that can determine the needs of their existing and new customers so as to provide more personalized search and recommendation services. This research project aims to develop novel models for predicting trust among online users so as to assemble a more complete Web of trust for realizing better search and recommendation.

Online social networks are dynamic in nature. While links between users are seemingly formed and removed randomly, there exists some interested link formation behaviors demonstrated by users performing link creation and removal activities. Uncovering these behaviors not only allows us to gain deep insights of the users, but also pave the way to decipher how social links are formed. In this paper, we propose a general framework to define user link formation behaviors using well studied local link structures (i.e., triads and dyads) in a dynamic social network where links are formed at different timestamps. Depending on the role a user plays in a link structure, we derive different types of link formation behaviors. We develop models for these behaviors and measure them for a set of users in an Epinions dataset.

There has been a recent increase of interest in analyzing trust and friendship networks to gain insights about relationship dynamics among users. Many sites such as Epinions, Facebook, and other social networking sites allow users to declare trusts or friendships between different members of the community. In this work, we are interested in extracting direct antagonistic communities (DACs) within a rich trust network involving trusts and distrusts. Each DAC is formed by two sub-communities with trust relationships among members of each sub-community but distrust relationships across the sub-communities. We develop an efficient algorithm that could analyze large trust networks leveraging the unique property of direct antagonistic community. Case studies on myGamma and Epinions are used to demonstrate that our approach can discover interesting DACs which provide useful insights into how the different subgroups exist within the Web 2.0 communities.

Future research

In what direction do you plan to go. What are the research problems you plan to attack? Why are these important or interesting problems to your field, or area of practice, or to society at large.

Looking ahead, I envisage that the future of computer science research will be very much dependent on the ability of computer scientists to develop new models, algorithms, and systems to solve problems relevant to emerging killer applications. Such application relevant problems will likely involve domain knowledge and require a combination of solution techniques from different research fields. My future research plan therefore is to use my past research experience in data integration and data/text mining, as well as my collaboration with domain experts in education and homeland security to identify these

applications and their related research problems. I can then: (a) propose solutions to address these problems; (b) construct large datasets from the Web and Web 2.0 data so as to establish a test-bed for evaluating different solution techniques; and (c) build tools and systems that demonstrate the viability of our proposed solutions in real life applications. In other words, I would like to focus on both the “science” and “engineering” dimensions of computer science research in my work. In addition, I would like to expand my collaboration with other SMU faculty members so as to develop a strong group in social network analysis and mining research.

Selected Publications and Outputs

1. Hady Wirawan Lauw, Ee-Peng Lim, Ke Wang: Quality and Leniency in Online Collaborative Rating Systems, ACM Transactions on Web (TWEB), to appear.
2. Hanbo Dai, Feida Zhu, Ee-Peng Lim, and Hwee Hwa Pang. Detecting Extreme Rank Anomalous Collections. SIAM International Conference on Data Mining (SDM2012), Anaheim, California, April 2012.
3. Bing Tian Dai, . Freddy Chong Tat Chua, . Ee-Peng Lim. Structural Analysis in Multi-Relational Social Networks. SIAM International Conference on Data Mining (SDM2012), Anaheim, California, April 2012.
4. Freddy Chong Tat Chua, Hady W. Lauw, Ee-Peng Lim. Mining Social Dependencies in Dynamic Interaction Networks. SIAM International Conference on Data Mining (SDM2012), Anaheim, California, April 2012.
5. Didi Surian, Nian Liu, David Lo, Hanghang Tong, Ee-Peng Lim and Christos Faloutsos, Recommending People in Developers' Collaboration Network, Working Conference on Reverse Engineering (WCRE2011), Lero, Limerick, Ireland, October 2011.
6. Tuan-Anh Hoang, Ee-Peng Lim, Palakorn Achananuparp, Jing Jiang, Feida Zhu: On Modeling Virality of Twitter Content. International Conference on Asian Digital Libraries, October 2011.
7. David Lo, Didi Surian, Kuan Zhang, Ee-Peng Lim, Mining Direct Antagonistic Communities in Explicit Trust Networks, 20th ACM International Conference on Information and Knowledge Management (CIKM2011), Glasgow, October 2011.
8. Freddy Chong Tat Chua and Ee-Peng Lim, Modeling Bipartite Graphs Using Hierarchical Structures, International Conference on Advances in Social Network Analysis and Mining (ASONAM2011), Kaohsiung, Taiwan, July 2011.
9. Freddy Chong Tat Chua, Hady Wirawan Lauw, Ee-Peng Lim. Predicting Item Adoption Using Social Correlation. 2011 SIAM International Conference on Data Mining (SDM2011), Phoenix/Mesa Arizona, April 2011.
10. Viet-An Nguyen, Cane Wing-Ki Leung, and Ee-Peng Lim. Modeling Link Formation Behaviors in Dynamic Social Networks. 2011 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP2011). College Park, Maryland, March 2011.
11. Wayne Xin Zhao, Jing Jiang, Jing He, Yang Song, Palakorn Achananuparp, Ee-Peng Lim, Xiaoming Li: Topical Keyphrase Extraction from Twitter. The 49th Annual

Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL 2011), Portland, Oregon, June 2011.

12. Tuan-Anh Hoang, Ee-Peng Lim, Palakorn Achananuparp, Jing Jiang, Loo-Nin Teow. Modeling Socialness in Dynamic Social Networks, International Conference on Advances in Social Network Analysis and Mining (ASONAM2011), Kaohsiung, Taiwan, July 2011.
13. Aixin Sun, Ying Liu, Ee-Peng Lim: Web classification of conceptual entities using co-training. *Expert Syst. Appl.* 38(12): 14367-14375, 2011.