Human Interactions on Online Social Media
Collecting and Analyzing Social Interaction Networks

Fredrik Erlandsson

Akademisk avhandling
som för avläggande av teknologie doktorsexamen vid Blekinge Tekniska Högskola kommer att offentligt försvaras i sal J1650, Campus Gräsvik, den 15 januari, 13.00.

Handledare:  
Henric Johnson, Professor, BTH  
Martin Boldt, Assistant professor, BTH  
Piotr Bródka, Assistant professor, Wroclaw University of Science and Technology, Poland  
Felix S. Wu, Professor, University of California Davis, USA

Opponent:  
Sune Lehmann Jørgensen, Associate professor, Technical University of Denmark

Betygsnämnd:  
Lisa Kaati, Swedish Defence Research Agency (FOI)  
Matteo Magnani, Uppsala University  
Przemyslaw Kazienko, Wroclaw University of Science and Technology, Poland

Blekinge Tekniska Högskola  
Institutionen för datalogi och datorsystemteknik, DIDD
Abstract

Online social media, such as Facebook, Twitter, and LinkedIn, provides users with services that enable them to interact both globally and instantly. The nature of social media interactions follows a constantly growing pattern that requires selection mechanisms to find and analyze interesting data. These interactions on social media can then be modeled into interaction networks, which enable network-based and graph-based methods to model and understand users' behaviors on social media. These methods could also benefit the field of complex networks in terms of finding initial seeds in the information cascade model. This thesis aims to investigate how to efficiently collect user-generated content and interactions from online social media sites. A novel method for data collection that is using an exploratory research, which includes prototyping, is presented, as part of the research results in this thesis.

Analysis of social data requires data that covers all the interactions in a given domain, which has shown to be difficult to handle in previous work. An additional contribution from the research conducted is that a novel method of crawling that extracts all social interactions from Facebook is presented. Over the period of the last few years, we have collected 280 million posts from public pages on Facebook using this crawling method. The collected posts include 35 billion likes and 5 billion comments from 700 million users. The data collection is the largest research dataset of social interactions on Facebook, enabling further and more accurate research in the area of social network analysis.

With the extracted data, it is possible to illustrate interactions between different users that do not necessarily have to be connected. Methods using the same data to identify and cluster different opinions in online communities have also been developed and evaluated. Furthermore, a proposed method is used and validated for finding appropriate seeds for information cascade analyses, and identification of influential users. Based upon the conducted research, it appears that the data mining approach, association rule learning, can be used successfully in identifying influential users with high accuracy. In addition, the same method can also be used for identifying seeds in an information cascade setting, with no significant difference than other network-based methods. Finally, privacy-related consequences of posting online is an important area for users to consider. Therefore, mitigating privacy risks contributes to a secure environment and methods to protect user privacy are presented.