

Chapter 15

Issues and Methods for Access, Storage, and Analysis of Data From Online Social Communities

Christopher John Quinn
Purdue University, USA

Matthew James Quinn
Williams College, USA

Alan Olinsky
Bryant University, USA

John Thomas Quinn
Bryant University, USA

ABSTRACT

This chapter provides an overview for a number of important issues related to studying user interactions in an online social network. The approach of social network analysis is detailed along with important basic concepts for network models. The different ways of indicating influence within a network are provided by describing various measures such as degree centrality, betweenness centrality and closeness centrality. Network structure as represented by cliques and components with measures of connectedness defined by clustering and reciprocity are also included. With the large volume of data associated with social networks, the significance of data storage and sampling are discussed. Since verbal communication is significant within networks, textual analysis is reviewed with respect to classification techniques such as sentiment analysis and with respect to topic modeling specifically latent semantic analysis, probabilistic latent semantic analysis, latent Dirichlet allocation and alternatives. Another important area that is provided in detail is information diffusion.

DOI: 10.4018/978-1-5225-3142-5.ch015

INTRODUCTION

Online social networks have become increasingly popular as a medium for information exchange and dialog for diverse communities. Twitter, a microblogging platform, has seen rapid growth and currently processes hundreds of millions of new messages each day with as many active users. Given the volume of activity and availability of data, there is a large amount of computational social science research investigating social networks and interactions in Twitter.

Some questions aim to characterize the structural properties of the social graph or “who follows whom.” The follower model employed by Twitter facilitates celebrity-fan communities. For instance, as of June 2014, USA President Barack Obama and musician Lady Gaga both have over forty million followers yet follow less than one million. Professional basketball player LeBron James has over twelve million followers and follows about 250 accounts. However, the number of followers is only one statistic of the network. There is interest in exploring how the connections are structured. For instance, do the users who follow LeBron James follow each other or at least follow some other player in common? Could they be considered a community or are they too separated? Are the followers of Barack Obama or Lady Gaga better interconnected? This is important not only for understanding the common interests of users but also for how easily information might be shared between them.

A significant question is “who is influential” in the social graph. With online social networks, influence is most easily measured by tangible phenomena such as reposting content deemed of interest (such as links to a video or article) or adopting a common hashtag. While the number of followers is an important statistic, research has shown that alone does not entail influence. The network topology needs to be taken into account as well as the followers themselves – some users are more likely to spread content than others.

There has also been significant research in exploring the communication of users. Microblogging platforms such as Twitter in particular promote novel linguistic features due to brevity constraints (each message must be less than 140 characters). Abbreviations, acronyms, hashtags (such as “#fail”) and excessive punctuations are all common. This presents challenges for automatic processing of the text for supervised classification tasks such as identifying the gender, age or political orientation of a user as well as unsupervised learning of the topics in a set of messages. Since the purpose of Twitter is to enable users to communicate, it is important for researchers to develop methods to be able to process and analyze that communication.

This chapter begins with a review of important techniques in social graph analysis, followed by a discussion of data access and sampling, textual analysis and finally, information diffusion. Software programs to assist processing social network data are referenced in an appendix.

Social Graph Analysis

Boyd and Ellison (2007) detail the history of social networks sites. They relate that SixDegrees.com could be considered one of the first of the online social communities dating back to 1997. Since then there has been a large growth in these communities. There are now well over 100 social networking sites with some of them having millions of registered users. In order to investigate these sites, network analysis is often utilized. The analysis of online social communities has become a major area of research. For instance, one work examined the information spread among 253 million subjects in the online social community Facebook (Bakshy, Rosenn, Marlow & Adamic, 2012). They found, as expected, that those

29 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/issues-and-methods-for-access-storage-and-analysis-of-data-from-online-social-communities/198772

Related Content

A Meta-Analysis Comparing Relational and Semantic Models

Keng Siau, Fiona F.H. Nahand Qing Cao (2013). *Innovations in Database Design, Web Applications, and Information Systems Management* (pp. 394-409).

www.irma-international.org/chapter/meta-analysis-comparing-relational-semantic/74401

CMU-WEB: A Conceptual Model for Designing Usable Web Applications

Akhilesh Bajajand Ramayya Krishnan (1999). *Journal of Database Management* (pp. 33-43).

www.irma-international.org/article/cmu-web-conceptual-model-designing/51225

Inherent Fusion: Towards Scalable Multi-Modal Similarity Search

Petra Budikova, Michal Batko, David Novakand Pavel Zezula (2016). *Journal of Database Management* (pp. 1-23).

www.irma-international.org/article/inherent-fusion/178633

Active Federated Database Systems

Genoveva Vargas-Solar (2005). *Encyclopedia of Database Technologies and Applications* (pp. 5-10).

www.irma-international.org/chapter/active-federated-database-systems/11114

Versioning Approach for Database Evolution

Hassina Bounif (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 91-101).

www.irma-international.org/chapter/versioning-approach-database-evolution/20692