NATIONAL UNIVERSITY OF SINGAPORE School of Computing PH.D DEFENCE - PUBLIC SEMINAR

Title:	Influence analysis for online social networks
Speaker:	Mr Xu Enliang
Date/Time:	16 October 2014, Thursday, 03:30 PM to 05:00 PM
Venue:	Executive Classroom, COM2-04-02
Supervisor :	Dr Wynne Hsu, Professor, School of Computing

Abstract:

The prevalence of online social media such as Facebook, Twitter, LinkedIn and YouTube has attracted considerable research in social influence analysis with applications in viral marketing, online advertising, recommender systems, information diffusion, and experts finding. Social influence occurs when one?s emotions, opinions, or behaviors are affected by others. Most of the works on social influence analysis have largely been focused on validating the existence of influence, studying the maximization of influence spread in the whole network, inferring the ?hidden? network from a list of observations, modeling direct influence in homogeneous networks, mining topic-level influence on heterogeneous networks, and conformity influence.

In this thesis, we perform influence analysis for online social networks by addressing three important issues in the discovery of influential nodes and influence relationships, which have been given little attention by existing works: influential path, topic-level influence and consistent influencer. We outline our approaches as follows.

First, we focus on influential path discovery. We show that influential paths can capture the dynamics of information diffusion better compared to influential edges. We propose a generative influence propagation model based on the Independent Cascade Model and Linear Threshold Model, which mathematically models the spread of certain information through a network. We formalize the top-k maximal influential path inference problem and develop an efficient algorithm, called TIP, to infer the top-k maximal influential paths. TIP makes use of the properties of top-k maximal influential paths to dynamically increase the support and prune the projected databases. As databases evolve over time, we also develop an incremental mining algorithm, named IncTIP, to maintain the set of top-k maximal influential paths efficiently. We evaluate the proposed algorithms on two real world datasets (MemeTracker and Twitter). The experimental results show that our algorithms are more scalable and more efficient than the base line algorithms. In addition, influential paths can improve the precision of predicting which node will be influenced next.

Next, we investigate topic-level influence. We show that in many applications the underlying networks are not explicitly modeled, and temporal factor plays an important role in determining social influence, which is ignored by existing works. We take into account the temporal factor in social influence to infer the influential strength between users at topic-level. Our approach does not require the underlying network structure to be known. We propose a guided hierarchical LDA approach to automatically identify topics without using any structural information. We then construct the topic-level social influence network incorporating the temporal factor to infer the influential strength among the users for each topic. Experimental results on two real world datasets (Twitter and MemeTracker) demonstrate the effectiveness of our methods. Further, we show that the proposed topic-level influence network can improve the precision of user behavior prediction and is useful for influence maximization.

Finally, we propose to identify k-consistent influencers. We show that finding influential users at single time point only cannot capture whether the users are consistently influential over a period of time. We devise an efficient algorithm that utilizes a grid index to scan the users in the 2D personal-preference consistency space, thereby obtaining the rank of these users at a given time point. Then we design the TCI algorithm to identify the k-consistent influencers for a given time interval. We conduct extensive experiments on three real world datasets (Citation, Flixster and Twitter) to evaluate the proposed methods. The experimental results demonstrate the effectiveness and efficiency of our methods. We show that the proposed k-consistent influencers is useful for identifying information sources and finding experts.