NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

PH.D DEFENCE - PUBLIC SEMINAR

Title: **User Profiling and Privacy Preserving from Multiple Social Networks**

Speaker: Ms Song Xuemeng

Date/Time: 21 September 2016, Wednesday, 11:00 AM to 12:30 PM

Venue: Executive Classroom, COM2-04-02

Supervisor : Dr Chua Tat Seng, Professor, School of Computing

Abstract:

User profiling, which aims to infer users' unobservable information based on observable information such as individual's behaviour or utterances, is the basis for many applications, such as personalized recommendation and expert finding. Traditional user profiling conducted with traditional medium, such as document records, is often hindered by the limited data sources. In recent years, the proliferation of social media has opened new opportunities for user profiling. Moreover, as different social networks provide different services, an increasing number of people are involved in multiple social networks, in which different aspects of users can be revealed by different social networks. Therefore, to comprehensively learn users' profiles, it is time to shift from a single social network to multiple social networks. Therefore, this thesis aims to investigate user profiling across multiple social networks. In particular, it covers studies in general scenarios of user profiling, in which a single task and multiple tasks are involved, respectively. Meanwhile, as user profiling would potentially put users at high privacy risks, this thesis also proposes a framework for privacy preserving.

In general, multi-social network learning involves two main steps: 1) social account mapping, and 2) multi-source learning. The first step aims to identify the same users across different social networks, while the second step targets at effectively aggregating multiple sources. This thesis will not address the social account mapping problem, and concentrate instead on the second step.

This thesis first proposes a novel scheme for multi-source mono-task learning to infer users' attributes, such as volunteerism tendency, which involves a single task. In particular, this proposed scheme is able to tackle the missing data problem, which is due to the fact that users may not be active enough in certain social networks. In addition, this scheme is capable of modeling both the source confidence and source consistency simultaneously. This thesis then proposes a multi-source multi-task learning scheme to infer users attributes, such as interest, where multiple related tasks can be involved. The proposed scheme jointly

regularizes two important aspects: source consistency and task relatedness. Finally, this thesis also develops a framework for privacy preserving to reduce users' privacy risks on social media. In particular, it proposes a taxonomy to comprehensively characterize users' personal aspects. With the guidance of such a taxonomy, we correspondingly propose a multi-task learning scheme to identify the potential privacy leakage.

Extensive experiments have been conducted on the real-world datasets. The experimental results enable us to draw the following key findings. First, utilizing multiple social networks does improve the performance of user profiling problems. Second, it is important to take source consistency and source confidence into consideration when dealing with multiple social networks. Third, in the context of user profiling with multiple tasks, taking task relatedness into account is plausible. Fourth, LIWC and Sentence2Vector features are the most discriminate features regarding privacy leakage detection. Last, the privacy leakage via UGCs holds certain temporal patterns and distinct behaviour patterns.