ONLINE SOCIAL NETWORKS: EVOLUTION AND SEARCH

PhD thesis submitted by Saptarshi Ghosh
(Roll number: 09CS9501)
Department of CSE, IIT Kharagpur

Abstract

Online Social Networks (OSNs) are among the most popular sites on the Web today. They are presently used by millions of users, not only to communicate with friends and others sharing common interests, but also to exchange and search for real-time information and news. In this scenario, it is important to understand the ongoing dynamics in OSNs, such as how users create social links, how they form social groups based on common interests, how they interact with online resources or contents, and so on. Further, given the enormous amount of information being shared in OSNs, search and recommendation services need to be developed over the OSN platforms to enable users to locate content that is relevant to their interests. In this thesis, we study some of the features of OSNs related to the above issues; specifically, we propose analytical models describing the evolution dynamics of OSNs, and methodologies to improve search on OSNs. As a specific example of an OSN over which the various issues are studied, this thesis focuses on Twitter which is one of the most popular OSNs at present.

For studying the evolution dynamics of OSNs, we consider the evolution of (i) the social network, where nodes are users and edges are the social links among the users, and (ii) social systems such as the group memberships of users or users' interactions with online resources, which are modeled as bipartite user-group or user-resource networks. Though several evolution models have already been proposed for such networks, none of these consider some of the practical issues found in real OSNs, such as large heterogeneity in user behavior, or the presence of restrictions on the number of social links of individual users. We develop analytical frameworks, using tools from complex network theory, to model the evolution of the networks. We demonstrate that the developed models can accurately explain the degree distributions of several real OSNs, such as the degree distributions of users in the Twitter social network, and the degree distributions of bipartite social systems such as the memberships of popular users in Twitter Lists, memberships of users in popular social groups, and annotations of popular resources by users in social tagging sites.

In order to improve search on the Twitter OSN, we study the linking strategies employed by spammers to farm links (followers) in the Twitter social network, and propose methodologies to prevent spammers from acquiring high influence ranks according to graph-based influence metrics such as PageRank (which are used by search engines). Moreover, we propose a novel methodology to discover topical attributes of users in Twitter, with the objective of improving topical search (i.e., search for information on specific topics) on Twitter. The proposed methodology, which utilizes social annotations, can accurately infer topical attributes for millions of popular Twitter users, and can also be used to identify topical experts on specific topics.

We believe that the analytical frameworks developed in this thesis not only have significant contributions towards modeling the evolution dynamics of OSNs, but also they advance the state-of-the-art in modeling of network evolution in general. In addition, the studies on link farming and identification of topical attributes of users provide several insights that can be used to build better search services on OSNs.