**Guest editorial: Special Issue on Social Computing and its Applications**

**Guandong Xu[[1]](#footnote-1), Aoying Zhou[[2]](#footnote-2), Nitin Agarwal[[3]](#footnote-3)**

Nowadays the emergence of web-based communities and hosted services such as social networking sites – Facebook, LinkedIn, wikis – Wikipedia, microblogging - Twitter and folksonomies – Delicious, Flickr and so on, brings in tremendous freedom of Web autonomy and facilitate collaboration and sharing between users. And along with the interactions between users and computers, social Web is rapidly becoming an important part of our digital experience, ranging from digital textual information to rich multimedia formats. Social networks have played an important role in different domains for about one decade, particularly involved in a broad range of social activities like user interaction, establishing friendship relationships, sharing and recommending resources, suggesting friends, creating groups and communities, commenting friend activities and opinions and so on. Recent years, has witnessed the rapid progress in the study of social networks for diverse applications, such as user profiling in Facebook and group recommendation via Flickr.

These aspects and characteristics form the most active and challenging parts of Web 2.0. A large amount of challenges and opportunities have arisen with the propagation and popularity of new applications and technologies. A prominent challenge lies in modeling and mining this vast volume of data to extract, represent and exploit meaningful knowledge, and to leverage structures and dynamics of emerging social networks residing in the social Web, especially social media. Social networks and social Web mining combines data mining with social computing as a promising direction and offers unique opportunities for developing novel algorithms and tools ranging from text and content mining to link mining and community detection and so on.

This special issue has gained overwhelming attention and received 52 submissions from researchers and practitioners working on social network analysis and social media mining. After initial examining of all submissions, 42 papers are selected into the regular rigorous review process and each submission has been reviewed by at least three reviewers. After 2-3 round reviews, eventually 10 quality papers are recommended to be included into this special issue, which are summarized as below.

The paper, titled “Mesoscopic Analysis of Networks with Genetic Algorithms”, presents a genetic based approach to discover communities in networks is proposed. The algorithm optimizes a simple but efficacious fitness function to identify densely connected groups of nodes with sparse connections between groups, thus sensibly reducing the search space of possible solutions. Experiments on synthetic and real life networks show the ability of the method to successfully detect the network structure.

Complex networks have received increasing attention by the scientific community, in line with the increasing availability of real-world network data. Apart from the network analysis that has focused on the characterization and measurement of local and global properties of graphs, such as diameter, degree distribution, centrality, and so on, the multidimensional nature of real world networks has discovered, i.e. many networks containing multiple connections between any pair of nodes have been analyzed. The paper “Multidimensional Networks: Foundations of Structural Analysis” discusses the basis for multidimensional network analysis by presenting a solid repertoire of basic concepts and analytical measures, which take into account the general structure of multidimensional networks. The framework has been tested on different real world multidimensional networks, showing the validity and the meaningfulness of the measures introduced, that are able to extract important and non-random information about complex phenomena in such networks.

In “A Time Decoupling Approach for Studying Forum Dynamics”, authors propose an approach that decouples temporal information about users into sequences of user events and inter-event times. Online forums are rich sources of information about user communication activity over time. Finding temporal patterns in online forum communication threads can advance our understanding of the dynamics of conversations. The main challenge of temporal analysis in this context is the complexity of forum data. There can be thousands of interacting users, who can be numerically described in many different ways. Moreover, user characteristics can evolve over time. Authors aim to develop a new feature space to represent the event sequences as paths, and model the distribution of the inter-event times. They study over 30, 000 users across four Internet forums, and discover novel patterns in user communication.

The paper, titled “Who Blogs What: Understanding the Publishing Behavior of Bloggers”, investigates the bloggers’ publishing style and impact by grouping bloggers based on the analysis of topical coverage and comparing their publishing behaviors. From a blog website with more than 370,000 posts, first two types of bloggers are identified: specialists and generalists. Then they study and compare the respective publishing behaviors in the blogosphere, finding that bloggers with different topical coverage do behave in different ways. Specialists generally make more contributions than generalists and tend to publish more on weekdays, during business hours, and on a more regular basis. Moreover, they also observe that specialists also have different publishing behaviors, with only a small fraction creating a large “buzz” or producing a voluminous output.

Online discussion threads are conversational cascades in the form of posted messages that can be generally found in social systems that comprise many-to-many interaction such as blogs, news aggregators or bulletin board systems. The paper, “A likelihood-based framework for the analysis of discussion threads”, proposes a framework based on generative models of growing trees to analyze the structure and evolution of discussion threads. The authors consider the growth of a discussion to be determined by an interplay between popularity, novelty and a trend (or bias) to reply to the thread originator. The relevance of these features is estimated using a full likelihood approach and allows to characterize the habits and communication patterns of a given platform and/or community. They apply the proposed framework on four popular websites: Slashdot, Barrapunto (a Spanish version of Slashdot), Meneame (a Spanish Digg-clone) and the article discussion pages of the English Wikipedia, to evaluate their model.

Social recommender systems largely rely on user-contributed data to infer users’ preference, which might introduce unreliability to recommenders as users are allowed to insert data freely. Although detecting malicious attacks from social spammers has been studied for years, detecting Noisy but Non-Malicious Users (NNMUs), which refers to those genuine users who may provide some untruthful data due to their imperfect behaviors, remains an open research question. The paper “Noisy but Non-Malicious User Detection in Social Recommender Systems”, studies how to detect NNMUs in social recommender systems. Based on the assumption that the ratings provided by a same user on closely correlated items should have similar scores, the authors propose an effective method for NNMU detection by capturing and accumulating user’s “self-contradictions”, i.e., the cases that a user provides very different rating scores on closely correlated items. They show that self-contradiction capturing can be formulated as a constrained quadratic optimization problem w.r.t. a set of slack variables, which can be further used to quantify the underlying noise in each test user.

The paper, titled “SocialSearch+: Enriching Social Network with Web Evidences”, addresses the problem of searching for social network accounts, *e.g.*, Twitter accounts, with the rich information available on the Web, *e.g.*, people names, attributes, and relationships to other people. Existing solutions building upon naive textual matching inevitably suffer low precision due to false positives (*e.g.*, fake impersonator accounts) and false negatives (*e.g.*, accounts using nicknames). To overcome these limitations, the authors leverage “relational” evidences extracted from the Web corpus, namely web-scale entity relationship graphs that extracted from name co-occurrences of Web and web-scale relational repositories, such as Freebase with complementary strength. Using both textual and relational features obtained from these resources, a ranking function is learned to aggregate these features for the accurate ordering of candidate matches. Another key contribution of this paper is to formulate confidence scoring as a separate problem from relevance ranking. The proposed system is evaluated by using real-life internet-scale entity-relationship and social network graphs.

The recommender systems utilizing Collaborative Filtering (CF) as the key algorithm are vulnerable to shilling attacks which insert malicious user profiles into the systems to push or nuke the reputations of targeted items. There are only a small number of labeled users in most the practical recommender systems, while a large number of users are unlabeled because it is expensive to obtain their identities. In “Shilling Attack Detection Utilizing Semi-supervised Learning Method for Collaborative Recommender System”, a new semi-supervised learning based shilling attack detection algorithm, namely Semi-SAD, is proposed to take advantage of both types of data. It first trains a naive Bayes classifier on a small set of labeled users, and then incorporates unlabeled users with EM to improve the initial naive Bayes classifier. Experiments on MovieLens datasets indicate that Semi-SAD can better detect various kinds of shilling attacks than others, especially against obfuscated and hybrid shilling attacks.

Mobile and pervasive computing technologies enable us to obtain real-world sensing data for sociological studies, such as exploring human behaviors and relationships. In “Understanding Social Relationship Evolution by Using Real-World Sensing Data”, the authors present a study of understanding social relationship evolution by using real-life anonymized mobile phone data. Through the study the authors show that social relationships (not only reciprocal friends and non-friends, but non-reciprocal friends) can be likely predicted by using real-world sensing data. In terms of the friendship evolution, they verify that the principles of reciprocality and transitivity play an important role in social relation evolution.

The paper titled “Can Predicate-Argument Structures be used for Contextual Opinion Retrieval from Blogs?” presents the use of predicate-argument structures for contextual opinion retrieval. Different from the keyword-based opinion retrieval approaches, which use the frequency of certain keywords, their solution is based on frequency of contextually relevant and subjective sentences. They use a linear relevance model that leverages semantic similarities among predicate argument structures of sentences. The model features with a linear combination of a popular relevance model, the proposed transformed terms similarity model, and the absolute value of a sentence subjectivity scoring scheme. The predicate-argument structures are then derived from the grammatical derivations of natural language query topics and the well-formed sentences from blog documents. Evaluation and experimental results demonstrate the feasibility of using predicate-argument structures for contextual opinion retrieval.

Finally, we would like to appreciate all authors who submitted manuscripts for consideration, and over 120 anonymous dedicated reviewers for their criticism and time to help us making final decisions. Without their valuable and strong supports, we cannot make this special issue successful. Our sincere gratitude will also go to the WWWJ EiC, Prof. Yanchun Zhang, Ms. Jennylyn Roseiento, Mr. Hector Nazario from the Springer Journal Editorial Office for helping us to presenting this special issue to readers.

1. University of Technology Sydney, Australia, email: Guandong.xu@uts.edu.au [↑](#footnote-ref-1)
2. Chinese University of Hong Kong, China, email: yu@se.cuhk.edu.hk [↑](#footnote-ref-2)
3. Inha University, Korea, email: trinity@inha.ac.kr [↑](#footnote-ref-3)