

# An Introduction to the Special Issue on Cross-Community Mining

Bin Guo<sup>1</sup>, Zhiwen Yu<sup>1</sup>, Francisco Camara Pereira<sup>2</sup>

<sup>1</sup> School of Computer Science, Northwestern Polytechnical University, P. R. China

<sup>2</sup> Singapore-MIT Alliance for Research and Technology, SMART/MIT

E-mail: guob@nwpu.edu.cn, zhiwenyu@nwpu.edu.cn, camara@smart.mit.edu

## 1. Introduction

People now live in heterogeneous social communities within cyber-physical spaces - both *online communities* (e.g., Flickr, Google+, LinkedIn) and social networks where digital content is exchanged, and *opportunistic* or *offline communities* that exploit opportunistic relationships between pairs of networked devices to exchange content (built on mobile ad hoc networking techniques) [1]. These communities have different technical features which lead to distinct kinds of interaction - such as patterns of comments and likes in online communities and co-location in offline communities, or issues of friendship, trust and influence in online communities and social popularity and movement patterns in offline communities. We further envision the rapid development of *cross-space communities* in recent years, which try to bridge the gap between human interactions in the physical world and virtual world (by merging social elements in online social networks with physical contexts in offline communities). Significant examples include: location-based social networks (LBSNs, e.g., FourSquare, Jiepan) [2], which interlink online human interaction with offline check-ins; event-based social networks (EBSNs, e.g., MeetUp), which try to build the link between physical and online events [3].

Rather than viewing online communities and offline communities as competing, we see them as complementary and suggest cross-community mining (CCM), which aims to connect different forms of communities and study the interaction among them. In general, CCM emphasizes the interaction and interplay among different forms of communities, addressing the aggregation, association, and fusion of the multi-modal data extracted from cross-space, heterogeneous community environments.

## 2. The Coverage of CCM

The power of CCM is demonstrated from different aspects, including *property association*, *data aggregation*, *knowledge transfer*, *cross-community information dissemination and forwarding*, and so on, as summarized in Table 1.

**Table 1** The Coverage of CCM

<i>Research topics</i>	<i>Description</i>	<i>Example</i>
<i>Property association</i>	It refers to the correlation and interplay between different properties (temporal, geo, social,) extracted from communities.	The interplay between mobility and friendship [2].
<i>Data aggregation</i>	It explores the integration of data from different communities to demonstrate their aggregated power in various purposes.	Composition of social networks to optimize prediction model [4]
<i>Knowledge transfer</i>	We can borrow and transfer the available knowledge from the source community to enhance the performance of the target one.	Using supervised information from the e-mail network to help infer social ties in a mobile phone network.
<i>Information dissemination &amp; forwarding</i>	It investigates how data is disseminated and forwarded over heterogeneous communities, especially under the complex interactions among online/offline community properties.	Leveraging the merits of online and opportunistic communities to augment information diffusion [5].

## 3. In This Issue

The articles in this issue highlight relevant current technologies and approaches pertaining to cross-community mining, including theoretical studies, practical issues, emerging technologies and innovative applications. Submissions to this special issue come from invited papers and open call for papers. We received a total of 12 submissions, including three invited papers and nine regular submissions. A large number of reviewers assisted us in the review process. In order to ensure high reviewing standards, three to four reviewers evaluated each paper. Six papers (three are regular submissions) were selected after the two-round review process. The six selected papers fall into four main research topics of CCM: three are about cross-space data aggregation, one about property association, one about information forwarding, and another one about cross-community context management.

In the opening paper of this special issue, “Enabling Large-scale Human Activity Inference on Smartphones using Community Similarity Networks (CSN)”, Nic Lane et al. propose a community-oriented approach to enhance human activity recognition using mobile phone data. Conventional approaches to training classifiers struggle to cope with

the diverse user populations routinely found in large-scale popular mobile applications. Differences between users (e.g., age, sex, behavioral patterns, lifestyle) confuse classifiers, which assume everyone is the same. The Community Similarity Networks (CSN) is proposed to address this issue, which incorporates inter-personal similarity measurements into the classifier training process. In other words, CSN exploits crowd-sourced sensor-data to personalize classifiers with data contributed from other similar users. Experiments show CSN outperforms existing approaches to classifier training under the presence of population diversity.

The second paper, “Mining City Characteristics based on Community Detection and Profiling in Location-Based Social Networks”, authored by Zhu Wang et al. addresses the community detection problem in cross-space communities. Different from traditional online social networks (e.g., Flickr, Facebook) which have explicit groups for users to join, LBSNs usually have no explicit community structure. Meanwhile, unlike homogeneous social networks which only contain a single type of social interaction, the co-existence of both online and offline social interactions in LBSNs makes the community detection problem much more challenging. Based on the heterogeneous social networks of LBSNs users, this paper presents a novel edge-centric co-clustering framework to discover overlapping communities in LBSNs, leveraging the aggregated power of online and offline network features. By analyzing the community structures of three different cities (using data from FourSquare), the authors further show that the proposed community detection framework can be used to investigate city characteristics.

In “Routing with Multi-Level Cross-Community Social Groups in Mobile Opportunistic Networks”, Fan Li et al. investigate the routing problem in opportunistic communities. Routing in opportunistic networks is challenging since it has to handle tough issues such as network partitioning, long delays, etc. Most existing routing methods for opportunistic networks avoid the costly flooding by selecting “good” relays from all encounters. To achieve efficient delivery of messages at low costs, this paper proposes a novel group-based routing protocol in which the relay node is selected based on multi-level, cross-community social group information. A group formation method that rely on a combination of online (user profiles) and offline information (historical

encounters in physical world) is introduced. The effectiveness of the protocol is validated via simulation experiments with several real-life data sets.

In the fourth paper, “Understanding the Locality Effect in Twitter: Measurement and Analysis”, Roberto Gonzalez et al. conduct a comprehensive analysis to understand the geographical characteristics of Twitter using the CCM techniques. Specifically, they study the locality effect of Twitter through its three main elements: *users*, *relationships* and *information flow*. Some interesting findings are obtained. For instance, they find that Twitter users present a high locality profile since they typically perform their activity from few cities separated at most few hundred kilometers within the same country. The interplay between physical world elements (e.g., language or culture) and online behaviors (locality of social ties) of Twitter users has also been investigated.

The fifth paper, “SpinRadar: A Spontaneous Service Provision Middleware for Place-aware Social Interactions”, authored by Byoungoh Kim et al. presents a middleware that supports spontaneous interaction of co-located users. The development of mobile social applications to enable opportunistic social interactions has become popular in recent years. One of the challenging problems in such interactions is to discover interaction opportunities with nearby users. Existing works focus on properties related to mobile users (e.g., user preferences/profile) in order to find similar users in the surroundings. However, the social implications of a place where the interaction is taking place is not considered to cue opportunistic connections. This paper presents SpinRadar, which empowers the recommendation of spontaneous interactions among co-located users by taking into account the semantics of a place. The evaluation results show that the proposed scheme meets the users’ satisfaction much higher than existing schemes.

The last paper, “Cross community context management in Cooperating Smart Spaces”, authored by Nikos Kalatzis et al., addresses the context management problem in multiple community environments. Aiming to bridge this gap and couple the advantages of pervasive computing and social networking, the Cooperating Smart Space concept has been introduced that addresses the needs of individual users, as well as of heterogeneous communities of users. The proposed framework contains innovative mechanisms for cross-community context management, such as the estimation of

community and cross-community context, the support for context inheritance across the members of hierarchical communities, and the exploitation of user context similarity for the discovery of new context-based communities. The evaluation results indicate that this framework is quite efficient in correctly detecting new communities and provides necessary mechanisms to identify appropriate members via a cross-community context similarity evaluation.

In concluding this overview, we would like to address our special thanks to Dr. Peter Thomas, the Editor-in-Chief, for his great support and effort throughout the whole publication process of this special issue. We are also grateful to all the authors for submitting their papers and the reviewers for their professional and timely work in making it possible to publish this special issue.

## References

- [1] Guo B, Zhang D, Wang Z, Yu Z, Zhou X (2013) Opportunistic IoT: Exploring the Harmonious Interaction between Human and the Internet of Things. *Journal of Network and Computer Applications (JNCA)*, Elsevier.
- [2] Cho E, Myers SA, Leskovec J (2011) Friendship and Mobility: user movement in location-based social networks. In: *Proceedings of the 17th ACM international conference on Knowledge discovery and data mining (KDD-11)*, pp. 1082-1090.
- [3] Liu X, He Q, Tian Y, Lee W, McPherson J, Han J (2012) Event-based social networks: linking the online and offline social worlds. In: *Proceedings of the 18th ACM international conference on Knowledge discovery and data mining (KDD-12)*, pp. 1032-1040.
- [4] Pan W, Aharony N, Pentland A (2011) Composite social network for predicting mobile apps installation. In: *Proceedings of the 25th Conference on Artificial Intelligence (AAAI-11)*, pp. 821-827.
- [5] Guo B, Zhang D, Yu Z, Zhou X (2012) Hybrid SN: Interlinking Opportunistic and Online Communities to Augment Information Dissemination. In: *Proceedings of the 9th IEEE International Conference on Ubiquitous Intelligence and Computing (UIC'12)*, pp. 188-195.