Accepted by IEEE Communications Magazine

Cross-Community Sensing and Mining (CSM)

Bin Guo[1], Zhiwen Yu[1], Daqing Zhang[1,2], Xingshe Zhou[1]
[1] School of Computer Science, Northwestern Polytechnical University, P. R. China
[2] Institut TELECOM SudParis, France
guob@nwpu.edu.cn

Abstract

With the developments in ICT techniques, people are involving in and connecting via various forms of communities in the cyber-physical space, such as online communities, opportunistic (offline) social networks, and location-based social networks. Different communities have distinct features and strengths. With humans playing the bridge role, these communities are implicitly interlinked. In contrast to the existing studies that mostly consider a single community, this paper addresses the interaction among distinct communities. In particular, we present an emerging research area – cross-community sensing and mining (CSM), which aims to connect heterogeneous, cross-space communities by revealing the complex linkage and interplay among their properties and identifying human behavior patterns by analyzing the data sensed/collected from multi-community environments. The paper describes and discusses the research background, characters, general framework, research challenges, as well as our practice of CSM.

Keywords: cross-community mining, opportunistic communities, mobile social networks, hybrid social networking, cross-space

1. Introduction

People contact and interact with each other via social networks (or communities). With the development in ICT technologies, people are connecting via various forms of communities in the cyber-physical space. At least two forms of social communities have gained popularity in the past decade: *online communities* in the cyber space, where people are connected by sharing content, opinions, and experiences in online social networks (OSNs, e.g., Facebook, Flickr), and *opportunistic* (or *offline*) *communities* in the physical space [1], which exploit opportunistic contacting and ad hoc connection between pairs of devices (e.g., mobile phones, vehicles) to share each other’s content (e.g., local traffic information) and resources. It mimics the way in which people seek information via social networking through direct, face-to-face contacts.

Different communities have distinct features and strengths. First, *these communities have different technical features that lead to distinct kinds of interactions*, such as patterns of comments/likes in online communities and co-location/movement in offline ones. Second, *their infrastructure support is different*. Online communities rely on services in the network infrastructure, while opportunistic communities are developed based on mobile ad hoc networks. Though possessing many differences, there are also intricate similarities between the two forms of communities. For instance, they all facilitate information sharing and dissemination among peers. Thanks to the bridge role that humans play, the properties of different communities are implicitly interlinked. For example, it has been reported that human movement (a spatial property) in the physical world is highly associated with their online connections [2].

The online community or the offline community is not new and has been studied extensively in the past few years (refer to Section 2). However, they follow separate research lines, and quite few of them address the interaction and interplay of different forms of communities. Since people are involved in multiple, heterogeneous communities and often traverse them in daily lives, we cannot give a comprehensive understanding and prediction of human social behaviors by using the information from a single community. For example, at one moment, Bob is staying at a place with Internet connection and he can communicate with his online friends (in the *online community*), while at another moment, he may travel by train with merely ad hoc connection with nearby passengers (forming an *offline community*). During this process, the information gained through interaction can be floated from online to offline communities. It is thus not difficult to understand that a number of issues need to be investigated when addressing the aggregated effects of heterogeneous communities. For some examples, *are there any relations between offline human mobility and online social interactions; will the aggregation of data from different communities bring new opportunities for human-centric services; can the knowledge learned from one community be transferred to another; can we leverage the interaction of online/offline community channels to enhance information dissemination?*

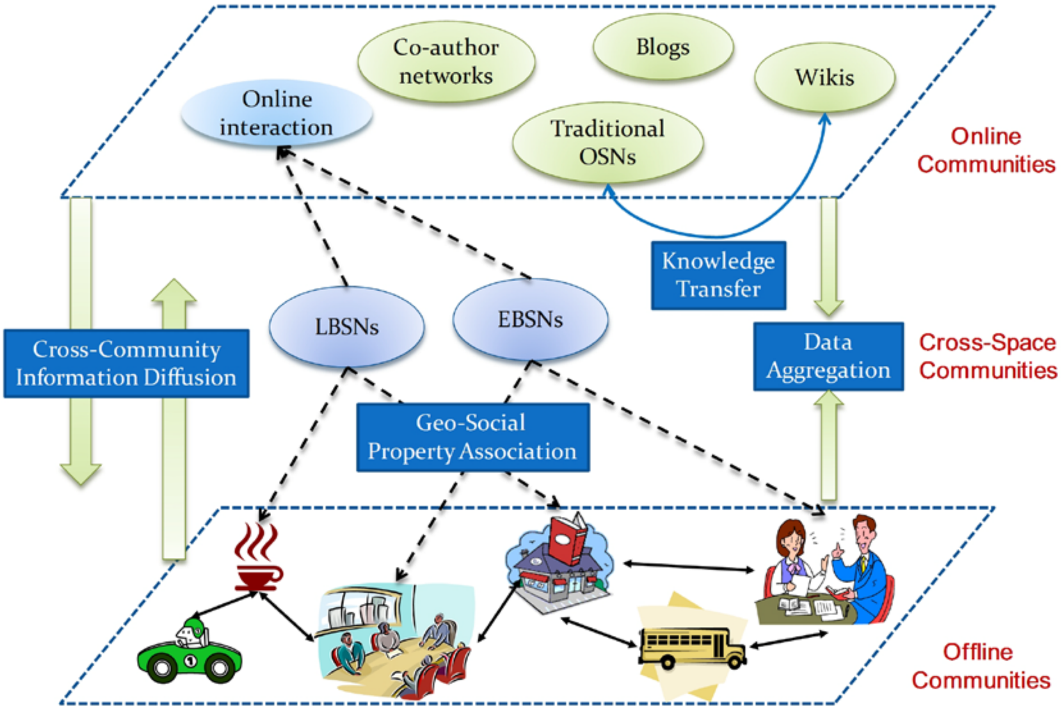

Figure 1. Cross-Community Sensing and Mining.

Rather than viewing online and offline communities as independent, we see them as complementary (due to their distinct features) and correlated, and thus in this paper we suggest cross-community sensing and mining (CSM). We are inspired by the multi-community involvement and cross-community traversing nature of modern people. CSM aims to connect heterogeneous communities by revealing the complex linkage and interplay among their properties (e.g., spatial or social) and identifying human behavior patterns by analyzing the data sensed/collected from multi-community environments. Furthermore, we envision the rapid development of *cross-space communities* in recent years, which try to bridge the gap between human interactions in the physical and virtual worlds. Significant examples are location-based social networks (LBSNs, e.g., FourSquare[1]) and event-based social networks (EBSNs, e.g., MeetUp[2]), which interlink online human interaction with offline human behaviors (check-ins, activities). CSM lays special emphasis on emerging cross-space communities, and explores the interaction of online/offline behaviors over them. We make an illustration of the CSM concept and its relationship with online/offline communities in Fig. 1, but more details will be depicted

[1] https://foursquare.com/
[2] http://www.meetup.com

in Section 3. Contrary to other closely related research areas, the unique characteristics of this new research area can be embodied in the following aspects:

- *Data*. The data sources are multimodal and heterogeneous, collected from either physical or virtual spaces. Different properties of people and communities can be extracted from the raw data collected from online/offline communities. The properties from different communities are correlated and can be integrated.
- *Technology*. The core technologies for CSM include mobile phone sensing, community analysis, data mining, and so on. The objective of sensing and data processing goes from capturing the multimodal data of user interactions and extracting physical/social features to cross-space feature fusion and association analysis.
- *Applications*. It aims to enable innovative services at society level like social recommendation, target advertising, location-dependent information diffusion, social event detection, and so on.

2. Research Background

Research on CSM can benefit from the ongoing and past research outcomes in *online community analysis*, *opportunistic social networking*, and *mobile social networking*.

Table 1 A taxonomy of CSM related research areas

Name	***Description***	***Representative Examples***
Traditional online social networks (OSN)	Social interaction and information sharing services on the online world	Wikis, Blogs, Skype, LinkedIn, Flickr
Opportunistic social networks (OPSN)	Facilitating the interaction among co-located people in the physical world	Find & Connect [5]
Location-based social networks (LBSN)	Exploring the link between "online" social interaction and "offline" check-ins	FourSquare, Facebook Places, Twitter
Event-based social networks (EBSN)	Exploring the link between physical events and online interactions	Meetup
Object-based social networks (OBSN)	Connecting people through their shared experience with physical objects	MemPhone [7]
CSM	Studying the interaction and association between different forms of communities	Hybrid SN [1] Social Contact Manager [14]

2.1. Online Community Analysis

During the last decade, we have observed an explosive growth of online social services such as e-mail, instant messaging, etc., which have changed the way in which people share/get information and communicate with each other. Leveraging on those services, a large body of work on human interaction analysis springs up. More recently, as the Internet steps into the era of Web 2.0, researchers turn their attention to the online social communities, such as online social networking (OSN) sites, wikis, and blogs. A lot of work has been done on online community and social behavior analysis. For example, Sheth's research group terms Web 2.0 service users as "citizen sensors" and has done much work on social event detection from user-contributed contents [3].

2.2. Opportunistic Social Networking

Social interactions in the physical world have always been important in sociology. During the past decade, the widespread use of sensor-equipped mobile phones has offered unprecedented opportunities to sense and gain insights on community behaviors in the physical world [4]. One of the significant research areas for offline community study is opportunistic social networking (OPSN) [1].

People present in the same place often miss opportunities to leverage social affinities for instant interaction owing to lack of awareness. OPSN, however, enables friend recommendation among opportunistically encountered people using ad hoc networking techniques (e.g., Bluetooth, Wi-Fi). Find & Connect [5] is one such system that supports ephemeral user interaction based on user profiles and physical proximity. Besides friends making, OPSN also facilitates local information sharing. For instance, people often want to be aware of nearby events (e.g., local traffic information) or need to distribute location-dependent information to others in the proximity (e.g., selling an unused ticket near the train station). Such information are better to be disseminated within the local-area community, without leveraging the global network/Internet.

2.3. Location-based Social Networking

With the development of mobile phone sensing techniques, mobile social networking (MSN), which showcases the power of merging social networking with various sensed physical elements, has rapidly grown up as a new kind of social interaction service. Contrary to OPSN, which is based on infrastructure-free networks, MSN services are usually based on backend servers and Internet connections. Further, by introducing physical contexts into traditional OSNs, MSN services are enabled to blend offline behaviors with online interactions. To date, the most prevalent and successful MSN service is LBSN.

LBSNs contain both “online” social interactions and “offline” check-in (location, can be obtained from on-board GPS sensors) information. Since location plays an essential role in our everyday lives, the addition of location dimension in traditional OSNs helps in bridging our online and offline worlds. Moreover, as location is one of the most important components of user context, extensive knowledge about an individual’s interests, behaviors, and relationships with others can be learned from his/her location. Significant examples of LBSNs include FourSquare, Facebook Places, and so on. Twitter also allows users to add location on their tweets. The development of LBSNs enables many novel applications that change the way in which we live, such as friend suggestion and location recommendation. For instance, adjacent location check-ins have been explored to cue potential social connections [2].

2.4. Other MSN Services

In addition to LBSNs, there are several other forms of MSN services that try to link the physical and virtual worlds. The EBSN is one of them, which tries to build the link between physical events (or offline social interactions) and online interactions [6]. Over EBSN services, on the one hand, people may propose social events (e.g., dinning out, workshops) and share them over online facilities, which can promote face-to-face social interactions. On the other hand, users’ participation in the same event in offline environments can also be captured. Object-based social network (OBSN) is another form of MSN services, which brings a new opportunity to integrate physical contexts with SNs. As reported in [7], OBSNs try to connect people and strengthen their relationship through their shared experience with physical objects, under the support of Internet of Things techniques (e.g., mobile tagging with RFID).

We have made a summary of the above-related research areas of CSM in Table 1. Similar to these related research areas, CSM takes human factors and social interaction analysis as key dimensions. However, it goes beyond all these areas in terms of its focus and research challenges. Contrary to the areas that focus on a single community, CSM particularly studies the interaction and association between different forms of communities, trying to bridge the gap between online behaviors and offline social interactions.

3. Characterization of CSM

Having presented the research background of CSM, this section presents its terminology and definition, and characterizes the features of this new research direction.

3.1. Terminology and Definition

We describe three concepts that are crucial for this work: *community*, *property*, and *cross-community mining*.

Community. Individuals are tightly connected via various social and physical processes, thus forming different forms of communities. This paper focuses on the communities enhanced by ICT technologies, which at least include the following types (as illustrated in Fig. 1):

- *Offline community*. Communities of co-located users that opportunistically form during everyday activities (e.g., during a meeting).
- *Online community*. It represents virtual communities who interact and share information through online social media, including social communication services such as Flickr and blogs, crowdsourcing communities such as Wiki, and collaboration networks such as DBLP (a co-author network).
- *Cross-space community*. This refers to the communities that integrate both online and physical elements. Various mobile social networks (e.g., LBSNs, EBSNs) discussed in Section 2 are typical examples of cross-space communities.

Property. Various types of properties can be extracted from tech-enhanced human communities, the significant ones being temporal, geographical, social, and thematic properties. Geo-features represent the ones that can be learned from human check-ins and GPS trajectories, such as points of interest (POIs) and mobility patterns. Social-features highlight interpersonal interaction patterns, including social topology, social popularity, and so on. Thematic properties focus on the topic and event contexts.

CSM. CSM emphasizes on the interaction among different forms of communities, addressing the association and fusion of the multimodal data extracted from distinct communities. The power of CSM can be explored from different aspects, such as *property interplay and association*, *data aggregation*, *cross-community information diffusion*, and *knowledge transfer* (as illustrated in Fig. 1). We characterize each of them in the following subsections.

3.2. Property Interplay and Association

Our society is founded on the interplay of human relationships and interactions. Since people exist and traverse among different communities, the properties (social, geographical, thematic) of distinct social networks are thus interweaved and highly associated. For instance, since every person is tightly embedded in our social structure, more and more evidence shows that when we want to model the behavior of a person, the best predictor is often not based on the person himself/herself, but rather on his/her social links [2]. The correlation between human social ties and geographic coincidences has also been investigated in [4]. Flap [8] studied the bidirectional relationship between social ties and user check-in places on Twitter. CSM emphasizes the correlation and interplay between different properties extracted from communities. Several references on property correlation are summarized in Table 2.

Table 2 References on CSM property correlation

Project or Work	*Community*	*Property Correlation*
Cho et al. [2]	LBSN (Brightkite, Gowalla)	Mobility and friendship
Reality Mining[4]	OPSN	Co-location and friendship
Flap [8]	LBSN (Twitter)	Social ties and check-in places

3.3. Cross-Community Data Aggregation

Data from different communities often present different attributes and strengths; moreover, they are often complementary. CSM explores the integration of data from different communities to demonstrate their aggregated power for various purposes. Here, we describe three distinct examples to showcase the effects of data integration from distinct communities, yet there are many more that can be explored.

(1) Sensor-based activity recognition enhanced by Web-mined knowledge. Knowledge obtained from the Web can be used to assist activity recognition in the physical world. For instance, Philipose et al. extracted the activity-relevant objects from the Web (Wikis, HowtoDos), which is then used in RFID-based human activity recognition [9].

(2) Merging the data from heterogeneous communities to develop new social apps. Data from different spaces often characterize one facet of a situation; thus, the fusion of data sources often draws a better picture of the situation. For example, by integrating the mined theme from user posts and the revealed location information from GPS-equipped

mobile phones, Twitter has been exploited to support near real-time report of earthquakes in Japan [10].

(3) Composition of social networks to optimize the prediction model. Prediction of human behaviors is important for the social software. However, an effective model often cannot be trained using data from a single community. Pan et al. [11] proposed a novel model that can infer an optimal composite network with the power from several candidate SNs, which is used to predict the mobile application installation behaviors of users.

3.4. Information Diffusion across Communities

Understanding the dynamics of data diffusion is critical for community-related studies. Through the combined effects of various SNs, information is diffused across local areas and cities. Previous works on information diffusion are single-community-oriented. For instance, for online data dissemination, researchers find that the nodes having high-degree centrality have a high impact. For offline communities, social features such as human popularity and community structure can significantly affect the diffusion process [1, 5].

In modern life, people are simultaneously involved in multiple online and offline communities. Contrary to the existing studies, CSM lays particular emphasis on the issues raised when data is disseminated over heterogeneous communities, especially under the complex interactions among online/offline community properties. Many fundamental problems need to be studied more comprehensively. For example, *to what extent does the geographic distribution of friendships in LBSNs affect where the content will be potentially propagated; are we able to determine which users in a community are structurally central in delivering information to a specific spatial region; how will information be disseminated over the interplay of online and offline human behaviors?* To answer these questions, CSM needs to study new data diffusion models and metrics that can capture and quantify the correlation in geo-social networks.

4. The Generic Framework

To facilitate the development of CSM applications, a generic system framework is essential. It should provide a set of mechanisms for multi-community data sensing and property extraction, property association, cross-space property fusion, and application development. We have proposed a conceptual framework for CSM systems, as shown in Fig. 2. It can be a starting point to build CSM applications with framework support.

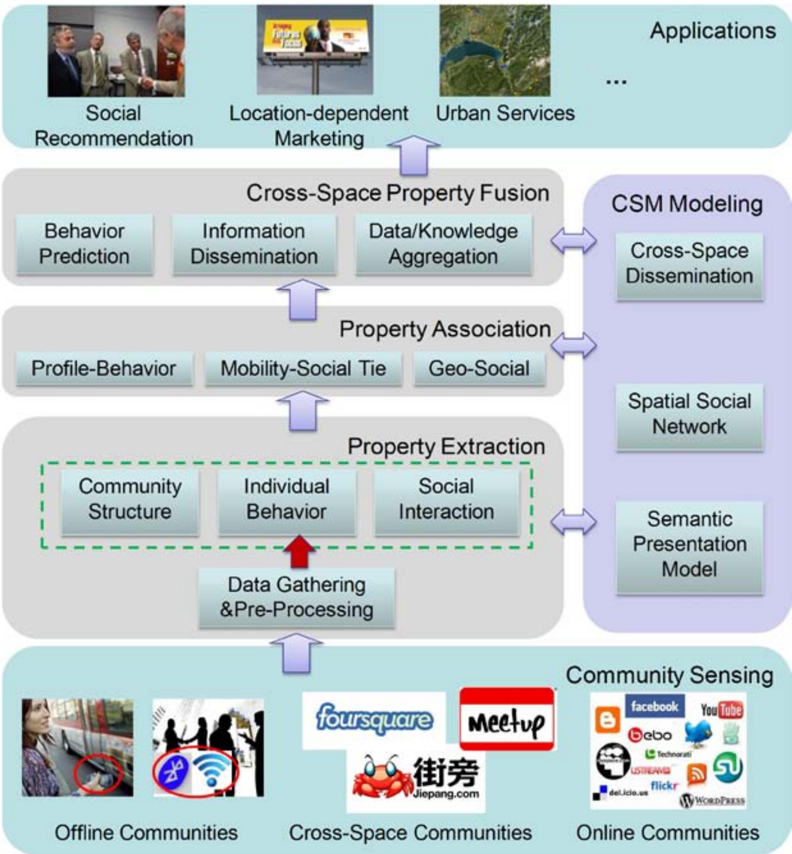

Figure 2. A Generic Framework for CSM.

The framework consists of the following components: The *community sensing* layer is responsible for collecting data from different forms of communities (using available Web APIs, crawling, or pervasive sensing techniques, as depicted in Section 5.1). The *CSM modeling* layer is the fundamental layer that handles cross-space social network analysis, multi-community data dissemination modeling, and knowledge/semantic sharing. The *property extraction* layer applies diverse data mining techniques to convert the low-level, single-modality sensing data into high-level properties, such as user behaviors/preferences and mobility patterns. The *property association* layer studies the linkage and interplay among various properties, such as the correlation between human mobility and social ties. The *cross-space property fusion* layer leverages the aggregated power of hybrid, cross-space features to predict human behaviors and improve the performance of data dissemination. Finally, the *application layer* includes a variety of potential services that can be enabled by the availability of CSM.

5. Key Research Challenges

Different challenges are faced by each layer of the CSM framework, in this section we discuss the key challenges of CSM. They are presented in line with the functional layers of the framework.

5.1. Community Sensing and Data Collection

Sensing and data collection are the basic steps for CSM research. There are two general ways for CSM data preparation: (1) We can collect data from emerging cross-space communities, which not only contain online social interactions as in other conventional OSNs but also include valuable offline human behaviors. For some cross-space communities (e.g., FourSquare), open APIs are available. However, there are often query rate limits when using open APIs, which may result in missing some user-contributed data. Several approaches have been proposed to address this issue, such as Web crawling techniques or the multi-IP collaboration method used in [8]. (2) Aggregating the data collected from online communities (e.g., OSNs) and offline communities (using mobile phone sensing). The major problem is *identity resolution*, to connect user online (e.g., Twitter ID) and offline identities (e.g., mobile phone number), and merging the data of the same entity from different data sources.

5.2. Modeling and Measurement Metrics

We turn our attention to cross-space community modeling and measurement metrics, which is important for studying the complex interaction between social and spatial dimensions. Though there is still no general approach for this, there have been several models/metrics proposed for dealing with new forms of social communities. In LBSNs, the "spatial social network," formed by a combination of social and spatial elements, is leveraged in [12]. A set of new geo-social metrics has also been proposed to spatial social networks, such as *spatial degree centrality* and *spatial closeness centrality*. The two metrics are an extension of the concepts of degree centrality and closeness centrality in traditional social networks, and both of them can be indicators of how the influenced audience of a user is geographically close to a given spatial neighborhood. Different from LBSNs, EBSNs contain not only location but time and people involved. Liu et al. have proposed a new modeling method for EBSNs, where two forms of social interactions (online and offline) are incorporated [6].

5.3. Geo-Social Property Association

Analysis of cross-space community data indicates community properties from different dimensions: temporal, geographical, and social. Under the bridge role of humans (a mixture of online and offline activities), these properties are highly associated. Recently, several works have explored the correlation between geo-social properties of OSNs, focusing for example on the correlation between geography and social topology [4, 8], friendship and mobility [2], and so on. Because of the complex nature of cross-space communities, more investigations should be made to explore the interplay between online interactions and real-world phenomena/events [5], such as interaction patterns and emergency events, and social interaction and economic development.

5.4. Inter-Community Knowledge Transfer

The knowledge we obtain from different communities often differs. For instance, we can learn trust relationships in reviewer networks, identify friendships from the communication network, and study user preferences from their check-ins in LBSNs. One of the promising questions is whether we can borrow and transfer the available knowledge from source communities to enhance the performance of target ones. Quite recently, several initial studies have been conducted to explore knowledge transfer among social networks. For instance, [13] have studied the knowledge transfer for inferring social ties in different social networks (e-mail, co-author, mobile). A transfer-based factor graph model, incorporating social theories (social balance, structural hole, social status, etc.) into a semi-supervised learning framework, is used to transfer supervised information from the source network to help infer social ties in the target network. Knowledge transfer has also been widely used in other data-processing-related domains, the theatrical foundation of which originates from the “transfer learning” concept in the machine-learning community.

5.5. Cross-Community Information Dissemination

In the past years, significant research efforts have been made on facilitating information diffusion in both online and offline communities. There have been very few works that explore information diffusion in cross-space, multi-community environments. Lima and Musolesi [12] presented a set of metrics that quantitatively capture the effects of online social links on the spreading of information in a given area. Liu et al. [6] investigated the dissemination models in EBSN services. The dissemination process over multiple communities can be very complex, and the interplay between geo-social features additionally affects the data diffusion process. The existing works have explored simple models that leverage the interaction among different communities, yet there is lack of a generic model that can address the aggregated effect of distinct factors. For instance, data dissemination in real settings often demonstrates the mixture of cascade and parallel process over online/offline channels [6], which still remains unexplored.

5.6. Privacy Concerns

Privacy preservation has been an important research field since the development of data-sharing techniques. The interplay between different communities in CSM raises new, yet more implicit privacy issues, with information (e.g., location, POIs) being sensitive and vulnerable to privacy attacks. For instance, Sadilek et al. investigated the correlation between social ties and location, and show that even if people keep their data private, their location can be inferred from the location of their friends [8]. The conclusion is that individuals who choose to reveal small amount of public information may be strong signals that leak the information they want to keep private. We should address new privacy issues in parallel with the development of CSM systems.

6. Our Practice of CSM

The study of CSM brings new potentials in many application areas. We make a summary of our ongoing works in the following and present our insights on how to address the challenges faced by CSM. Their development follows the framework proposed in Section 4.

6.1. Social Contact Manager

Social contact manager (SCM) demonstrates how to *leverage the aggregated power of heterogeneous data sources from online and offline communities*. As the number of contacts increases, people often find it difficult to maintain their contact network using human memory alone. We are frequently beset with questions like "*Who is that person? I think I met him in Tokyo last year*." The existing contact tools make up for the unreliability of human memory by storing contact information in digital format; however, manual input of contact data can burden the users. To address this issue, we develop SCM, an intelligent social contact management system [14]. It supports the auto-collection of rich contact data (e.g., profile, face-to-face meeting contexts) from both online and opportunistic communities, leveraging the aggregated power of *Web intelligence* and *pervasive sensing* techniques (as shown in Fig. 3).

We employ a mobile card-scanner to extract basic information from the collected business cards (forming an opportunistic community). The scanned basic information is then used to extract other contact information from the Web (i.e., the online community) using a hybrid of heuristic rules and the Conditional Random Field (CRF) [14]. In addition to the contact information from online community, we also extract various contextual information regarding the meeting event through mobile phone sensing. For example, we characterize the meeting-location context from two aspects. One is depicted at the city level, which is obtained from the embedded GPS in mobile phones. The other is at the semantic level, which characterizes the categories of meeting places (e.g., workplaces, leisure places). It is recognized by processing the audio signals captured by the microphones in mobile phones. All collected data will be put in the right form and aggregated (the *data aggregation* module in the CSM framework) in a *contact data repository* (a *semantic presentation* model in the framework). These information can be leveraged to manage their contacts better, especially for efficient contact retrieval in name-slipping situations.

To validate the performance of SCM, we recruited 12 subjects from our university and divided them into two subgroups (6 in each). To make a comparative study, one subgroup was asked to use the SCM, and the other one to use a traditional contact tool — the Gmail contact (GC). We introduced 6 contacts to the subjects at the beginning and, after a certain period (e.g., 1 or 2 weeks), we asked them to use the given tool to locate the correct contacts (by showing their photos). Experiment results show that SCM save up to 50% time on contact search than when using GC. The performance gap grows with the increase in elapsed time since the introduction of the contact. The result indicates that when compared with using merely profile information in traditional contact tools, the usage of aggregated information (profile, contextual cues, etc.) from heterogeneous data sources can improve the performance on contact recall.

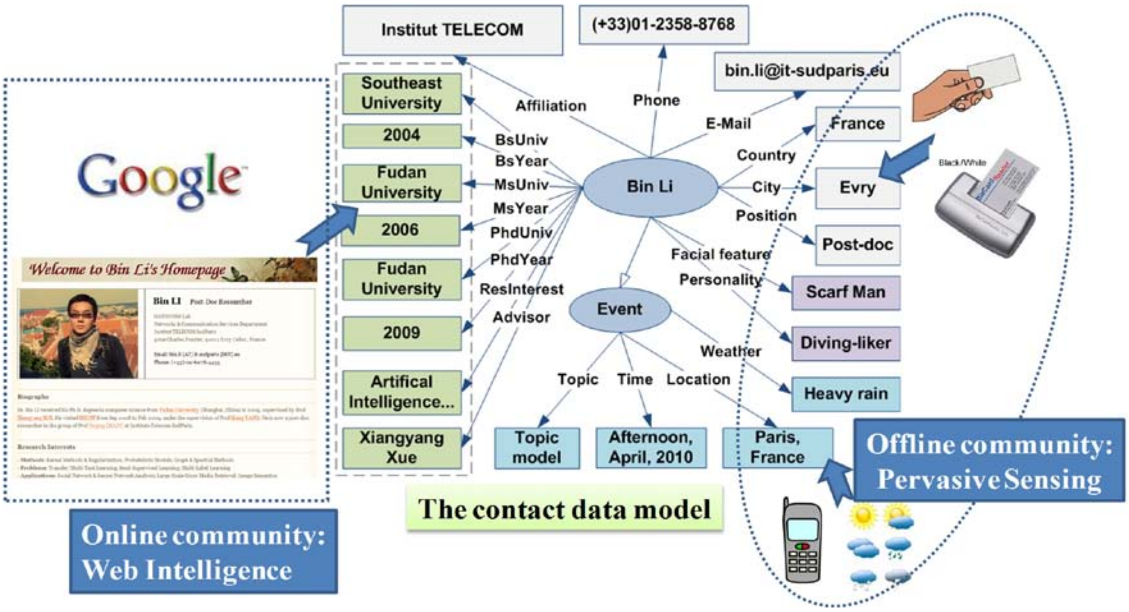

Figure 3. The Social Contact Manager.

6.2. Heterogeneous Social Networking

Heterogeneous Social Networking (HSN) [1] leverages *the interaction between online and opportunistic communities to augment information dissemination.* Information sharing and dissemination have been addressed in both online and opportunistic communities. People connected via online communities can share data. However, since the linkage relations in online communities change slowly, the “audience” covered is often limited. Sharing and dissemination of information in opportunistic communities can opportunistically cover more people with or without pre-ties, but usually comes at long dissemination delays. HSN is designed in a way that links online and opportunistic SNs to enhance them both (i.e., a *cross-space dissemination* model in the CSM framework).

In opportunistic networking studies, to facilitate information dissemination, brokers (usually popular nodes encountered in the physical world) are often used to carry and forward information. However, brokers may not be willing to contribute their resources (e.g., storage, battery) and thus the data can be dropped. One of the key features enabled by HSN is the *popularity-based online broker selection protocol*. It is based on the social willingness phenomena, which indicates that people are willing to help their friends. HSN allows users to advertise their predicted popularity in the online community (online friends are willing to act as brokers), and a publisher can choose the ones with highest-popularity among them. We compared the performance of HSN with single-community-dependent methods (e.g., the pure opportunistic method). Experiment results show that great performance improvement is obtained when using HSN: the success ratio to cover interested nodes increases by an average of 25% and the match latency decreases by around 60%. It is because the integration of an online community shortens the broker selection process, and increases the opportunity to select brokers with high popularity. Detailed experiment settings and discussions can be found in [1].

6.3. Cross-Space Community Analysis

This research demonstrates *how to extract the community structure leveraging the fusion of heterogeneous, online/offline features from LBSNs*. Community detection can facilitate various applications, such as target marketing and social recommendation. Owing to the multi-role play and intercommunity traversing nature (e.g., families, football teams, research groups, etc.), it is more reasonable to cluster users into overlapping communities rather than disjoint ones. Most of the existing community-detection approaches are based on network structural features (e.g., social links). However, structural information in OSNs is often sparse and weak; it is thus difficult to detect interpretable overlapping communities by considering only online network structural information. Contrary to the traditional OSNs, the heterogeneous interactions in LBSNs provide rich information about users, venues, and the connections between them (i.e., check-ins), making it possible for group users according to different clustering metrics.

Figure 4 gives a *spatial social network* model (referring to the CSM framework) to characterize the interactions in LBSNs, where we have four users and four venues represented as two types of nodes, and each check-in is represented as an edge between a user and a venue. For this spatial-element-enhanced social network, if we cluster users based solely on the *check-in network structure*, we can get two overlapping communities: *Group 1* (Lee, Jerry) and *Group 2* (Jerry, Lin, Lucy). According to the semantic of associated venues, we can roughly categorize *Group 1* as a friend community and *Group 2* as a colleague community. However, for an LBSN, there is not only the check-in network, but additional properties that can be learned from check-in venue history, such as the radius of gyration (*r* for short) of a user. As we can read from

Fig. 4, because Jerry and Lin travel frequently, the *r* of them is 500 km and 600 km, respectively. In contrast, Lucy often stays locally and the *r* of her is only 60 km. By using a combination of check-in network and human mobility (e.g., the radius of gyration) features, we can get three overlapping communities: *Group 1* (Lee, Jerry), *Group 2* (Jerry, Lin), and *Group 3* (Lucy). In this case, even though Jerry, Lin, and Lucy have similar check-in patterns, they are further grouped into two separate communities. Here, we can probably label *Group 1* as a friend community, *Group 2* as a researcher community, and *Group 3* as a staff community.

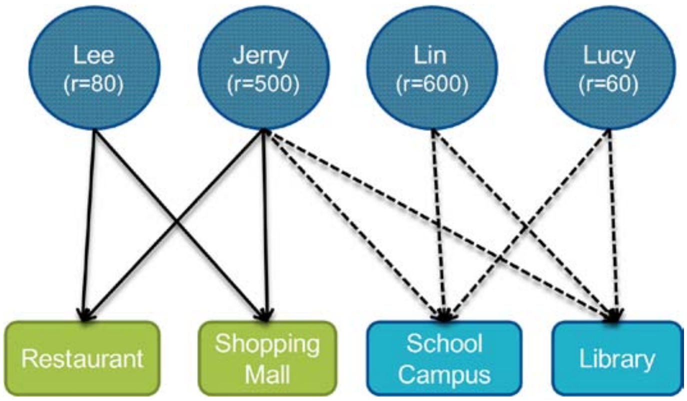

Figure 4. An example of user check-in network.

Based on the heterogeneous social networks of LBSNs, we come out with a novel edge-centric, co-clustering approach to discover overlapping communities in LBSNs [15]. By employing a combination of cross-space features (e.g., online check-in network structure, mobility patterns in the physical world), the proposed approach is able to group like-minded users from different social perspectives and granularity. The efficacy of our approach is validated by intensive empirical evaluations based on the collected FourSquare dataset of 720,000 users with 3 million check-ins [15].

7. Conclusion and Future Vision

This paper has presented CSM, a new research area that emphasizes on the interaction of heterogeneous communities, addressing the aggregation and association of the multimodal data extracted from cross-space, heterogeneous community environments. As an emerging area, the prevalence and development of CSM still face numerous challenges, such as community data collection and integration, cross-space community modeling and measurement, approaches for data association, inter-community knowledge transfer, and information dissemination in multi-community environments. All these challenges present substantial research opportunities for academic researchers, industrial technologists, and business strategists. We further present three of our ongoing projects/applications on CSM, including social contact management, opportunistic social networking, and cross-space community analysis, and demonstrate our experience to address the challenges.

With the rapid development of various ICT-enhanced community services and the increasing availability of data dimensions (enabled by wireless sensing techniques) that link physical and virtual worlds, we believe that the research scope of CSM will expand and its applications will multiply in the next decade. First, the future CSM will unify the virtual world and the physical world by linking online social communities and offline spontaneously formed communities. Second, with the increase in the large-scale data collected from heterogeneous communities, advanced techniques on complex network modeling, data mining and knowledge transferring, data association and aggregation, and semantic fusion techniques will become more and more important. Third, the complementary nature of heterogeneous communities will bring new opportunities to develop new human-centric services. Finally, we should investigate how heterogeneous communities evolve as a result of people's behaviors conducted in different communities.

Acknowledgements

This work was partially supported by the National Basic Research Program of China 973 (No. 2012CB316400), and the National Natural Science Foundation of China (No. 61332005, 61373119, 61222209).

References

[1] B. Guo *et al.*, "Hybrid SN: Interlinking Opportunistic and Online Communities to Augment Information Dissemination," *Proc. of UIC'12 Conf.*, Fukuoka, Japan, 2012, pp. 188-195.

[2] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and Mobility: user movement in location-based social networks," *Proc. of KDD'11 Conf.*, 2011, pp. 1082-1090.

[3] A. Sheth *et al.*, "Citizen Sensing, Social Signals, and Enriching Human Experience," *IEEE Internet Computing*, vol. 13, no. 4, 2009, pp. 87-92.

[4] N. Eagle *et al.*, "Inferring Social Network Structure using Mobile Phone Data," *PNAS*, vol. 106, no. 36, 2007, pp. 15274-15278.

[5] X. Zuo,et al., "Connecting People at a Conference: A Study of Influence Between Offline and Online Using a Mobile Social Application," *Proc. IEEE CPSCom'12 Conf.*, 2012, pp. 277-284.

[6] X. Liu *et al.*, “Event-based social networks: linking the online and offline social worlds,” *Proc. of KDD’12 Conf.*, Beijing, China, 2012, pp. 1032-1040.
[7] B. Guo *et al.*, “MemPhone: From Personal Memory Aid to Community Memory Sharing using Mobile Tagging,” *Proc. of PerCom Workshops*, San Diego, USA, 2013.
[8] A. Sadilek, H. Kautz, and J.P. Bigham, “Finding your friends and following them to where you are,” *Proc. of WSDM’12 Conf.*, 2012, pp. 723-732.
[9] M. Philipose *et al.*, “Inferring Activities from Interactions with Objects,” *IEEE Pervasive Computing*, vol. 3, no. 4, 2004, pp. 50-57.
[10] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors,” *Proc. WWW’10 Conf.*, 2010, pp. 851-860.
[11] W. Pan, N. Aharony, and A. Pentland, “Composite social network for predicting mobile apps installation,” *Proc. of AAAI’11 Conf.*, 2011.
[12] A. Lima, and M. Musolesi, “Spatial dissemination metrics for location-based social networks,” *Proc. of UbiComp’12 Conf.*, 2012, pp. 972-979.
[13] J. Tang *et al.*, “Inferring Social Ties across Heterogeneous Networks,” *Proc. of WSDM’12 Conf.*, 2012, pp. 743-752.
[14] B. Guo *et al.*, “Enhancing Memory Recall via an Intelligent Social Contact Management System,” *IEEE Transactions on Human-Machine Systems*, 2013 (to appear).
[15] Z. Wang *et al.*, “Cross-Domain Community Detection in Heterogeneous Social Networks,” *Personal and Ubiquitous Computing*, 2013 (to appear).