Leveraging the Potential of Social Hub Services

Qingyuan Gong, Xiaolong Yu, Chao Xu, Zhichun Guo, Yang Chen, Xin Wang

School of Computer Science, Fudan University, China

{chenyang,xinw}@fudan.edu.cn

ABSTRACT

Given the diverse focuses of emerging online social networks (OSNs), it is quite common that a user has accounts on multiple OSN sites. Social hub services, a.k.a., social directory services, help users manage their accounts on different sites. In this work, we conduct a data-driven study by crawling more than one million profiles from about.me, a representative social hub service. Our study aims to gain insights by referring to the crawled data. According to our analysis, we can see a user has a high probability to provide consistent information across different OSN sites. In addition, we demonstrate the usefulness of cross-OSN information aggregation.

CCS CONCEPTS

• **Human-centered computing** → **Social network analysis**;

KEYWORDS

Online social networks, social hub services, measurement

1 MOTIVATION AND BACKGROUND

Nowadays online social networks are developing rapidly around the world. In addition to supporting social interactions among users, each OSN site has its own focus. For example, Twitter acts as a new medium for information sharing, LinkedIn plays an important role in job search, and Foursquare supports location-centric activities. Because of such diversity in services provided by OSNs, it is very common that a user has accounts on multiple OSN sites [3].

Social hub services have been introduced to help a user manage her accounts on different OSNs. In this paper, we focus on about.me [7], a representative social hub service, allowing a user to add her OSN accounts to her profile page. About.me has attracted millions of users around the world. This service provides us a unique and informative source to aggregate each user's different online identities together. In this work, we conduct a data-driven study by crawling the profile pages of more than one million about.me users. We aim to answer two key questions. On one hand, we study the

user coverage of social hub services. By investigating the composition of about.me users, we demonstrate who are more interested in using such services. On the other hand, we show the usefulness of the social hub services in understanding the user behavior. With the help of the about.me data, we are able to obtain a user's public information on multiple OSN sites. We evaluate the consistency of a user's demographic information on different sites. Furthermore, we study the cross-OSN information aggregation. Our findings can help construct more comprehensive social footprints of users, which can be further applied in different practical applications, including urban computing [6], wellness profiling [4] and malicious account detection [5].

2 DATA COLLECTION AND ANALYSIS

To obtain a massive set of about.me profiles, we use a list of 1.11 million about.me user IDs in [1]. Based on this list, we implement a Python-based crawler to fetch the profile pages of all these users. We did the crawling during March 2017, using 5 virtual instances on the Vultr cloud. For each profile page, we extract the demographic information including the name, interest tags, occupation, and location. In particular, we get the list of a user's linked social accounts, for example, her accounts on Facebook, Twitter, LinkedIn, and etc.

Based on the crawled profiles, we study the composition of about.me users from different aspects. First, we look at the interest tags of each user. Note that a user is allowed to add multiple interest tags. The percentages of the top 10 interest tags are shown in Fig. 1(a). In particular, "Technology", "Photograph" and "Music" are the most popular interest tags. In addition, we examine the occupations of the users. We list the the percentages of the top 10 occupations in Fig. 1(b). The top three occupations are "Consultant", "Writer" and "Software Engineer". Last but not least, we explore the corresponding OSN sites of linked social accounts. We list the top 10 linked OSN sites of about.me users in Fig. 1(c). The top three sites are Twitter, Facebook, and LinkedIn.

About.me provides us an opportunity to access each user's linked social accounts on multiple OSN sites. Therefore, we are able to further crawl a user's social accounts on different sites to get a comprehensive view of this user. We focus on two aspects, i.e., *cross-OSN information consistency* and *cross-OSN information aggregation.* For consistency, we aim to know whether a user prefers to disclose consistent information on different OSN sites or not. For aggregation, we are interested in examining the usefulness of putting a user's information on different OSN sites together. We get a subset of 30,000 randomly selected about.me accounts and further crawl their public profiles on Facebook, Foursquare, Google+ and Blogger.

To evaluate the cross-OSN information consistency, we use the gender and location fields, which are two common information fields of the profiles on all these four OSN sites. For any two OSN sites, we define the "consistency" metric as the fraction of users

(a) Top 10 Ranks of Interests

(b) Top 10 Ranks of Occupations
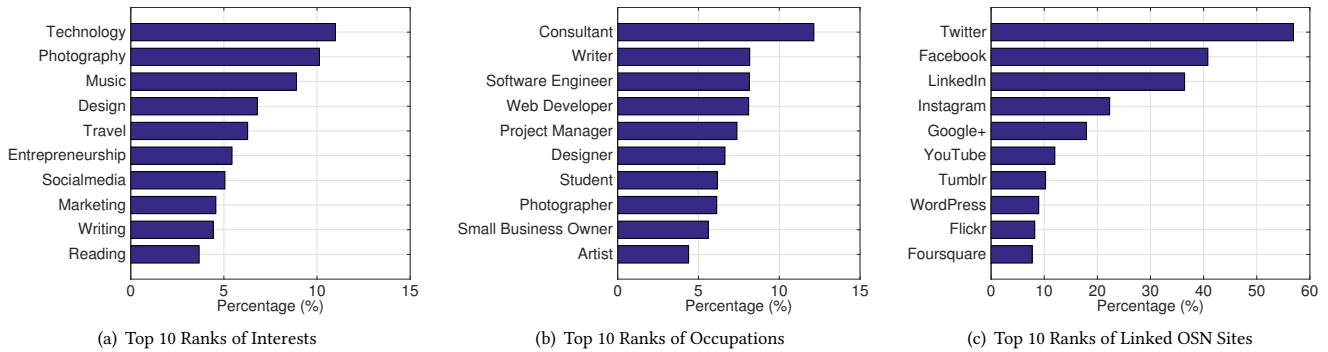
(c) Top 10 Ranks of Linked OSN Sites

Figure 1: The distribution of popular interests, occupations, and social links on about.me

Table 1: Consistency of each OSN pair (Gender, Location)

Websites	Facebook	Foursquare	Google+	Blogger
Facebook	\	(0.99, 0.92)	(0.99, 0.94)	(0.98, 0.89)
Foursquare	(0.99, 0.92)	\	(0.99, 0.93)	(0.99, 0.91)
Google+	(0.99, 0.94)	(0.99, 0.93)	\	(0.98, 0.89)
Blogger	(0.98, 0.89)	(0.99, 0.91)	(0.98,0.89)	\

that set the same values for a certain information field. In particular, for the location field, we evaluate the consistency at the country level. For each pair of the four selected OSN sites, we use a two-tuple to represent the consistency values in terms of gender and country. The results are shown in Table 1. It can be observed that each pair of OSN sites shows a high consistency concerning the gender and location fields. In short, we can see that users tend to provide consistent information across different OSN sites.

Regarding cross-OSN information aggregation, we can see each OSN site has its own set of available information fields. Some of them are mandatory and some are optional. Intuitively, putting the information fields of different OSN sites together would provide a more detailed view of a user's online footprints [2]. Note that on different OSN sites, the same information field might have different names, such as "work" on Facebook and "occupation" on Blogger. We manually examine all information fields of the four selected OSN sites and merge these fields. Finally, we get 11 different information fields across all four sites. In Fig. 2, we see the complementary cumulative distribution function (CCDF) of the number of activated information fields of each of the four OSN sites. In addition, an aggregated set of a user's enabled information fields on different OSN sites are shown as "All". We can see that through the cross-OSN information aggregation by referring to about.me profiles, the number of activated information fields of each user has been increased largely. In other words, putting a user's disclosed information on different OSN sites together, we can know more about this user. This demonstrates the great potential of the social hub services.

3 DISCUSSION AND FUTURE WORK

In this work, we conduct a measurement-based study of social directory services, using about.me as an example. We gain insights from
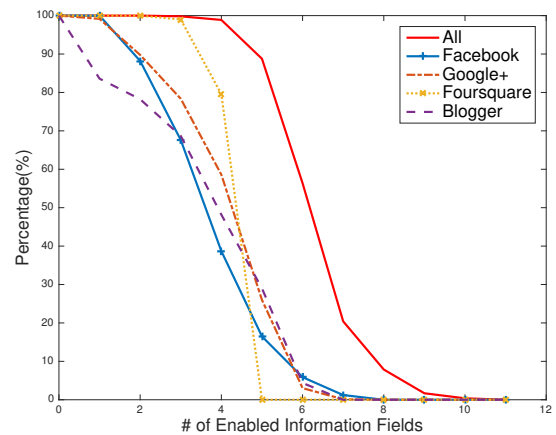

Figure 2: Cross-OSN information aggregation

showing the composition of users on about.me. In addition, among the collected about.me profiles, we refer to each about.me user's linked accounts on different OSN sites. We study both cross-OSN consistency and cross-OSN aggregation problems. Our findings reveal the great potential of the linkage of a user's accounts on multiple OSN sites.

For future work, we aim to make good use of the linkage of user accounts on different OSNs. First, by aggregating a user's information across multiple OSNs, we will be able to achieve a more informative view by getting a better "social footprint" of a user. We can leverage the behavioral data from multiple sites to study the user behavior deeply, instead of relying on a single site. In addition, we plan to formalize and study the cross-OSN prediction problem. We will explore the potential of using a user's activities on site A to predict her behavior on site B. Such cross-OSN prediction will be useful in different kinds of user classification problems, such as malicious account detection. Last but not least, since social hub services provide a more comprehensive view of each user, it will naturally bring some challenges to user privacy. We will investigate the privacy preserving challenges introduced by the linkage between accounts on different OSN sites.

REFERENCES

[1] X. Cao and Y. Yu. BASS: A Bootstrapping Approach for Aligning Heterogenous Social Networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2016.

[2] T. Chen, M. A. Kaafar, and et al. Is More Always Merrier? A Deep Dive Into Online Social Footprints. In *Proc. of ACM WOSN*, 2012.

[3] Y. Chen, C. Zhuang, Q. Cao, and P. Hui. Understanding cross-site linking in online social networks. In *Proc. of the 8th ACM Workshop on Social Network Mining and Analysis (SNAKDD)*, 2014.

[4] A. Farseev and T. Chua. TweetFit: Fusing Multiple Social Media and Sensor Data for Wellness Profile Learning. In *Proc. of AAAI*, 2017.

[5] G. Venkatadri, O. Goga, C. Zhong, B. Viswanath, K. P. Gummadi, and N. Sastry. Strengthening Weak Identities Through Inter-Domain Trust Transfer. In *Proc. of WWW*, 2016.

[6] N. J. Yuan, F. Zhang, D. Lian, K. Zheng, S. Yu, and X. Xie. We Know How You Live: Exploring the Spectrum of Urban Lifestyles. In *Proc. of ACM COSN*, 2013.

[7] C. Zhong, H.-w. Chang, D. Karamshuk, D. Lee, and N. Sastry. Wearing Many (Social) Hats: How Different are Your Different Social Network Personae? In *Proc. of AAAI ICWSM*, 2017.