Fourth International Conference on Semantics, Knowledge and Grid

Community Discovery Based on Social Actors' Interests and Social Relationships

Fei Yan [1], Jing Jiang [2], Yang Lu [3], Qingjun Luo [4], Ming Zhang [5]

School of Electronics Engineering and Computer Science, Peking University, P.R.China

[1]fei.yan929@gmail.com
[2]jiangjing@net.pku.edu.cn
[3]yanglu@pku.edu.cn
[4]ydnaol.luo@gmail.com
[5]mzhang@net.pku.edu.cn

***Abstract*— The increasing availability of social network data has motivated more computational research in social network analysis. Recently, discovering community from social networks came into the big picture of social network study. In this paper, we propose a novel approach for community discovery based on the contents of social actors' personal interests and their social relationships. Our dataset is a populated ontology created by integrating the publicly available FOAF (Friend-of-a-Friend) documents. We firstly discover the communities based on the social actors' personal interests. Then we use the social relationship to extend the communities for further discovery. Finally we conduct extensive experimental study to demonstrate that our approach is effective for community discovery.**

I. Introduction

Nowadays, there are two cultures affecting the future evolution of the Web: the Semantic Web and Web 2.0. On one hand the Semantic Web was formally presented in [1] and became the research hotspot immediately. Many efforts have been made by the researchers to boost the Web shifting to the Semantic Web form: RDF [2], RDFS [3], OWL [4] and SPARQL [5] have become the standards of the world. Academic research contributed methodologies for ontology engineering [6][7], evolution [8], debugging [9], modularisation [10] and other related areas [11][12][13][14]. On the other hand the Web 2.0 technologies, as outlined in [15] and exemplified by websites such as Wikipedia [16], Youtube [17], flickr [18] and Facebook [19], focus on the easier distributed collaboration of the Web. This version of the Web promotes users to the center status through the operation process. In the Web 2.0 world, users are not only the consumers, but also the producers, so they have a substitutional name called contributors. At the same time, Web 2.0 technologies allow contributors to collaborate and share information easily through one of the Web 2.0 characteristic features called Community. Both the Semantic Web and Web 2.0 are regarded as the rudiments of the next generation Web.

Social network analysis, which is one of the most important Web 2.0 application, derived from the sociology. With the increasing availability of social network data, social network analysis has received more and more attention from the Web community. Analysis of social networks can be applied to many domains, such as viral marketing [20] and the evaluation of importance of social actors [21].

An important application of social networks is the community discovering: how social actors gather into groups such that they are intra-group close and inter-group loose [22]. Community is primarily proposed to provide a sharing platform for people with common interests. Community activities change the social network and affect the social behavior greatly. There are two types of community discovering methods. One is based on the common interests, and this kind of methods address problems related to the similarity measurement of social actors' interests. The other one is based on the social relationships between social actors. For example, Tyler et al. [23] and Culotta et al. [24] extract community structures from email corpora, and the social network is constructed to measure the intensity of contacts between email users.

However, both content based and social relationship based methods just considered one aspect of the character of community. In fact, contents and social relationships are both of high importance to community in interest sharing and community relationship constructing. We propose a novel method which combines contents and social relationships to discover the communities based on an open and actual dataset collecting from the Web, i.e. the aggregation of FOAF [25] documents.

In order to analyze social network, a representative and available dataset is necessary. FOAF, i.e. Friend of a Friend, is a machine-readable ontology which describes people, their activities, their relationships and objects. The aggregation of such FOAF documents by means of the "knows" relationship forms a social network. FOAF documents, which are representative of Semantic Web data, are created independently by users. They consist of links among real-world persons, and they are publicly available. Therefore, the aggregation of FOAF documents is the ideal dataset for our social network analysis.

Our contributions are summarized as follows:

- We introduce the Semantic Web technologies to process Web 2.0 applications - community discovery, and combine two cultures in a smooth way.
- We propose a novel three-step approach to discover community, which fully considers the content network

978-0-7695-3401-5/08 $25.00 © 2008 IEEE
DOI 10.1109/SKG.2008.38

and social network.

- We conduct extensive experiments to demonstrate that our method is effective.

The rest of this paper is organized as follows. Section II introduces the motivation and background of our work. Section III presents our main work of dataset preparation. Community discovery and extension will be explained in Section IV, followed by the performance evaluation in Section V. Finally, Section VI concludes the paper.

II. Motivation and Background

This paper aims to discover communities by combining content network and social network, which is an open problem in the social network research. We introduce the Semantic Web technology to deal with the problem. Our work is to characterize the common engineering and research challenges of building practical Semantic Web applications rather than contribute to the theoretical aspects of Semantic Web or social network. Furthermore, our work provides a clue to deal with the open problem of combining the content network and social network, and boosts Semantic Web technologies deploying in current Web. To achieve this goal, we should follow the process similar to the engineering process of developing Semantic Web application. The three-step process are summarized as follows and illustrated in Figure 1.

- Preparing high quality FOAF dataset: such data is often not available, and hence we should collect data from the internet, clean it and set up the dataset in the form of populated ontology.
- Content based community discovery: in this step, the information of people's interest provided by the dataset is used to discover the communities.
- Community extension based on social relationships: due to the lack of adequate interest information, we use the social relationships extracted from the dataset to extend the communities.

In the remaining section, we introduce the background of our work. Section II-A is about the Semantic Web application used by describing person and the social relationship between people - FOAF. We collect the FOAF documents from the Web to set up our social network dataset. In Section II-B, we will discuss the methods used in community discovering, and introduce two kinds of community models: CM1 and CM2.

A. FOAF

FOAF [26] is a machine-readable ontology describing people, their activities, their relationship and objects. FOAF allows groups of people to describe social networks without a centralized database. FOAF is an extension to Resource Description Framework (RDF) and is defined using OWL Web Ontology Language. Each profile has a unique identifier, (e.g. the person's e-mail addresses, a URI of the homepage or weblog of the person), which is used to define these relationships.

Through FOAF documents, information of document individuals can be obtained, especially the people who have direct relationships between them. If these people also publish FOAF documents, a social network based on friend relationship can be created by using the FOAF documents.

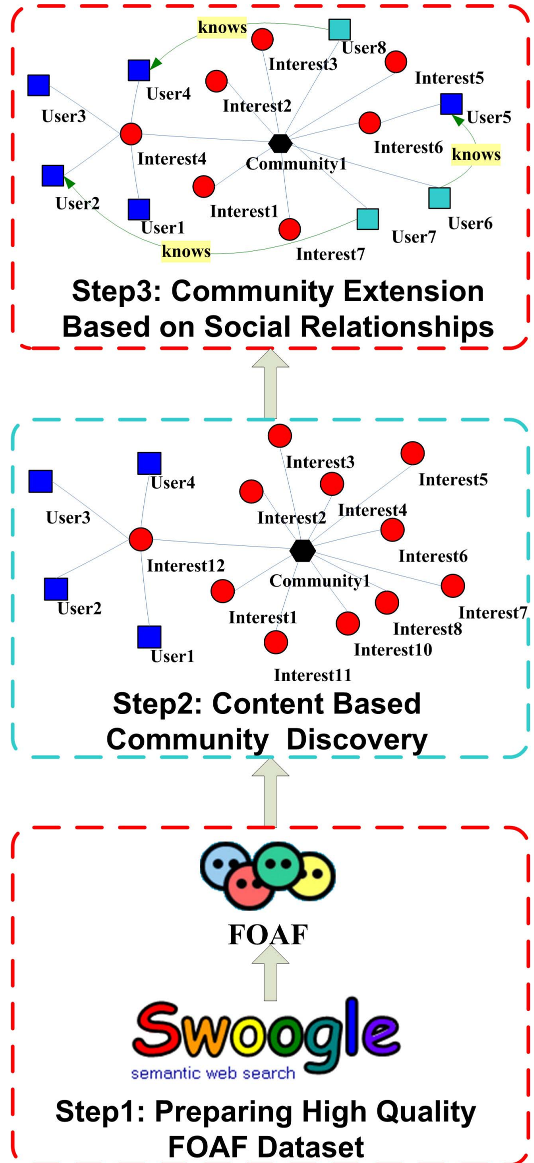

Fig. 1. Three-step process

B. Community discovery

Discovering community from general networks is of obvious interest. Early methods include graph partitioning [27] and hierarchical clustering [28]. Recent algorithms [29] and [30] addressed several problems related to prior knowledge of community size, the precise definition of inter-vertices similarity measure and improved computational efficiency [22]. Zhou et al. [31] proposed a method using probabilistic models for discovering communities. In general, semantic similarity in social networks is the meaning or reason behind the network connections.

According to [31], there are two kinds of community models

(CM). The first category CM1 is modeling community with users. In this model, the social relationships between people are mainly considered to discover communities. Community is a collection of people with a set of interests. Figure 2 shows the structure of CM1. The other one is modeling community with interest. In contrast to CM1, CM2 community consists of a set of interests, which are relative with respective user groups. In this paper, we will address the community discovery based on CM2, whose structure is illustrated in Figure 3.

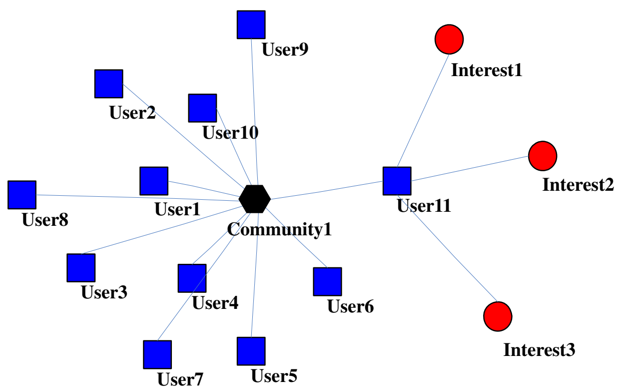

Fig. 2. Structure of CM1

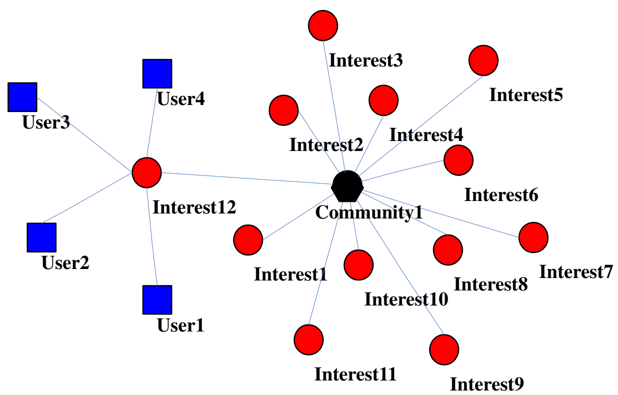

Fig. 3. Structure of CM2

III. FOAF Dataset Preparation

The Friend of a Friend (FOAF) data source, which is representative of Semantic Web data, is created independently by many individuals because anyone can use the FOAF vocabulary to publish information about themselves and their social relationships. For example, a person entity may include identity-properties such as email and homepage, additional personal-properties such as name and personal photo using foaf:name and foaf:depiction respectively, and friendship-properties marked by foaf:knows. All the information can be encoded using an RDF/XML syntax thus making the corresponding social network information "machine processable". Many people maintain this type of social networks information in the FOAF world. For this reason, we can expect that people will use various sets of properties and that the values of such properties will be written using different conventions. These FOAF documents which we used as our dataset were discovered by Swoogle Semantic Web search engine [32].

A. FOAF dataset cleaning

After we collect the FOAF documents from the Web, we should analyze the documents and wipe the noise data. Ding et al. in [33] analyzed the characteristic patterns which were implied by the ontological semantics and the empirical usage of FOAF vocabulary. They gave us a formal definition of a strict FOAF document D with the following four characteristic patterns:

- D is a valid RDF document. This can be validated by a RDF parser.
- D uses the FOAF namespace.
- D contains an RDF graph pattern as shown in Figure 4. X and Z are two different instances of rdfs:Resource and Y is an instance of rdf:Property using FOAF namespace.
- D defines only one instance of foaf:Person without referencing it as an object in any triples within D. D may additionally have some other instances of foaf:Person; however, each of them must be referenced as an object in at least one triple in D.

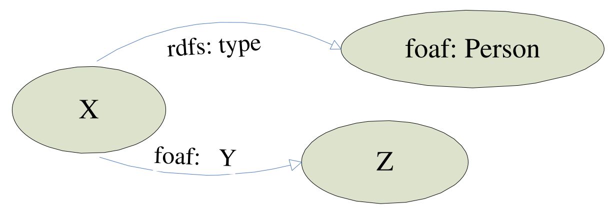

Fig. 4. FOAF document pattern

The above patterns, especially the fourth pattern, are quite strict and exclude many documents not dedicating to a person. Therefore, by removing the fourth pattern, Ding et al. defined a general FOAF document as long as it contains instances of foaf:Person.

We use the definition of general FOAF document as the criterion to refine the FOAF dataset. The cleaning process is not the keystone of this paper, so interested readers can find the method in [33].

B. FOAF dataset analysis

After the data cleaning, we have set up a populated ontology as our dataset. There are in total 201,612 RDF triples in the dataset, which contains 36,724 instances of foaf:person and 33,266 relations of foaf:knows. The foaf:knows vocabulary represents the social relationship of a person. It is also the most important property of the FOAF project. The foaf:knows relationships will be used to extend the communities discovered by contents discussed in next section.

Another important property related to our propose is the foaf:interest property. It represents the person's interest in the real world. We can use the contents of foaf:interest to discover communities which consist of people with similar interests.

According to the statistic work mentioned by [33], foaf:interest is the 7^{th} most frequently used vocabulary. The wide usage of the foaf:interest property guarantees the validity of our work. The dataset contains 4,215 triples of foaf:interest instances, most of which are URLs of some websites. However, there are incomprehensible characters existing in the object of the foaf:interest triple. This phenomenon may be caused by the FOAF documents owners' mistake when creating them, or something wrong happened during the FOAF documents transportation or download process. The quantities of the error foaf:interest instances are very small, and hence it will lead little impact on our research, so we simply delete them from the dataset. After wiping of the error triples and the redundancy, we get 1,101 instances of foaf:interest. The first step of community discovery will be based on this foaf:interest dataset.

As a large scale of FOAF users are Semantic Web researchers, the co-author relation is an important factor to indicate the social relationship tightness between two researchers. Our original intention is to introduce the co-author relationship, which is determined by the use of foaf:publication vocabulary, combined with foaf:knows relationship to measure the tightness between two individuals. Unfortunately, our dataset only contains 101 foaf:publication triples. It is hard to evaluate whether our method is valid by such a small dataset. However, if we introduce the DBLP bibliography [34], which provides collaboration network data by virtue of the explicit co-author relationships among authors, to our work, it will make sense. This will be further discussed in Section VI.

IV. Community discovering and extension

This paper intends to combine content and social network to discover communities based on the FOAF dataset. In Section IV-A, we explain the process of discovering communities by clustering interests. In Section IV-B, we extend the discovered communities by person-to-person and person-to-community social relationship calculations.

A. Communities discovery based on interests

Community is primarily proposed to provide a sharing platform for people with common interests. Community members, whether friends or not, are encouraged to contribute and share more contents. Through this process, the chances of members to become friends will be promoted. Community activities change the social network and affect the social behavior tremendously. In our dataset, the foaf:interest vocabulary represents the FOAF document owner's personal interest. Due to the large usage of the vocabulary, we can discover communities through the content of foaf:interest. However, almost all the foar:interest contents in our dataset are URLs. It is hard to extract the representative information of the interest from the URLs directly. Thus, in the first step, we should analyze the websites denoted by the URLs to extract the representative information, i.e. some feature words, of the interest. The next step is to cluster the dispersed interests into groups, which are the same concepts as communities, based on the similarity among the interests. Then the corresponding people are divided into communities. After that, the communities are formed. The last step is to give the communities their own labels. We label the communities with the most frequently occurring feature words, and find the most famous person (MFP) in each community. From the view of degrees, MFP is the person with the highest indegree calculated by foaf:knows vocabulary. Thus MFPs are the advertisements of the communities to attract more people to join.

1) Tagging the contents of interests: How to get the representative information from the website? The traditional approach uses tf*idf technique to extract the feature of a document. There are many open source softwares that help us extract the feature of a website based on the traditional approach. However, most foaf:interest websites are not the article-kind websites whose main contents are articles. They can be hub-liked pages with many out links to other websites. It is hard to analyze the contents of hub-liked websites by machine. Accompanying with the rapid development of Web 2.0, more and more users take part in the construction of the Web. They upload video, share pictures, tag the website and so on. The users' social behaviors of tagging the websites provide a new way to find the content features of the website. Tags generated by users represent their apprehension of the content of the websites. They are more expressive and accurate than the features extracted by machine. Can we use user-generated tags instead of the features extracted by machine automatically to represent the contents of a websites? Yahoo researchers [35] testified the validity of our hypothesis by sufficient experiments. Based on the hypothesis, we boost our work to find user-generated tags to represent the contents of the foaf:interest websites.

We tried several ways to obtain the tags of the foaf:interest website generated by users. At first we tended to find the tags from the famous Web 2.0 website-Del.icio.us [36]. Del.icio.us provides social bookmarking Web service for storing, sharing, and discovering Web bookmarks. It owns more than three million users and 100 million bookmarked URLs. Unfortunately only a few foaf:interest websites were marked in Del.icio.us, the rest of the websites should be tagged manually. Therefore, we developed a tagging tool, and invited three Peking University students to help us tag the foaf:interest websites. Ultimately, we obtained 1,101 high quality tag files. Each tag file contained three to eight representative words corresponding to a foaf:interest website. The fileset will be used to discover the communities.

2) Discovering the community: Based on the foaf:interest feature fileset, we used the k-means clustering algorithm to discover the communities. The k-means algorithm is an algorithm to cluster n objects based on attributes into k partitions, $k < n$. The algorithm attempts to find the centers of natural clusters in the data. It assumes that the object attributes form a vector space. The objective is to minimize total intra-cluster variance or the squared error function [37](illustrated by formula 1)

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (1)$$

where there are k clusters S_i $(i = 1, 2, \cdots, k)$, and μ_i is the centroid or mean point of all the points $x_j \in S_i$.

In this paper, the clustering objects are the foaf:interest files. And the object attributes are the tags generated by people. Yale[38] is a famous and free Text-mining software. After downloading the software and the clustering plugin, it is easy to get the clustering result by inputting the foaf:interest fileset and the parameter of the number of cluster k. The clustering result is a file which contains the clusters with the their member object-ids. By mapping the object-id to foaf:interest file, and then mapping the foaf:interest file to foaf:person, the communities are discovered.

The type of community model discovered by us is CM2 shown in Figure 3.

3) Labeling the communities: We use k-means clustering algorithm and successfully discover the communities based on the FOAF users' interests. However, the discovered communities are just sets of users, lacking descriptive information such as community name, famous members and so on. In this step we label the communities and find the most famous person of high social affection as the advertisements.

To name the communities, we use the word frequency calculation method. The communities clustered by the contents of FOAF users' personal interests are the CM2 pattern discussed in Section II. Each interest file contains some feature words. For each community, we collect all the feature words from the interest files and count the words occurrence frequency. Then we choose top n most frequently occurrence words as the name of the community.

Name can be used to identify a community, and the most famous person (MFP) with high social affection can be the advertisement of a community. We judge whether a member is famous by counting the indegree of the foaf:knows relation. If none of members has indegree, the person with highest outdegree will be promoted to be the MFP. For each community, we will choose only one member as the MFP. The MFPs with high social affection are used easily to attract other social actors to join their communities.

B. Community extension based on social network analysis

One problem we must address in this paper is that some FOAF documents do not use foaf:interest vocabulary. For those FOAF users, we can't divide them into any communities discovered based on foaf:interest. A community in a small size would be meaningless according to its initial purpose, i.e. sharing contents and making friends. How to improve it? From the view of content, the resources are limited, and we have already made full use of the contents of foaf:interest. We can hardly improve the performance from this aspect. However, when it comes to friends making, we can make some efforts to invite more people to join the communities. FOAF means Friend of a Friend, and the friendship is represented mainly by the foaf:knows relationship. The foaf:knows relationships between people form the social network. In this section, we attempt to enrich the communities discovered based on contents clustering by analyzing the social network of the FOAF dataset. We first calculate the person-to-person social relationship of the dataset. Then we explore person-to-community social relationship. At last, we distribute the individual to join the communities with high scores of person-to-community social relationship. After the community extension, the structure of community model is changed, which combines CM1 and CM2, and adds the foaf:knows relationship between users. We name the new model CM3 and its structure is shown by Figure 5.

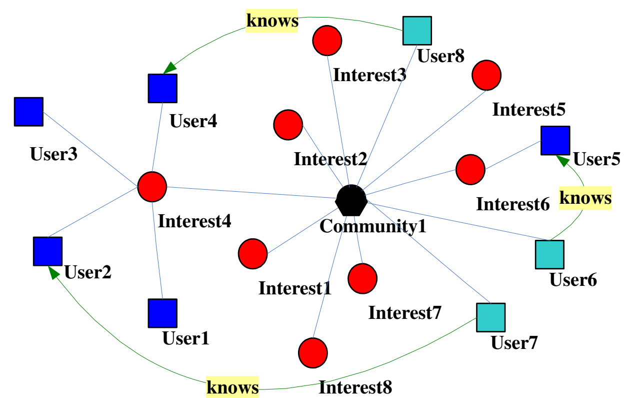

Fig. 5. Structure of CM3

1) Strength of person-to-person social relationship calculation: We need to quantify the social relationship in order to calculate the tightness of relationship between people. The foaf:knows relationship is the most important relationship to represent the social network. It is used to explicitly list the people known by someone. We get the foaf:knows relationship by querying the dataset ontology using SPARQL. We assign the strength of foaf:knows relation from person A to person B to be 1, if A assert foaf:knows to B; otherwise we assign 0. The example SPARQL query for the names of subjects and objects with foaf:knows relationship is presented in Figure 6.

```
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>
PREFIX rdf:   <http://www.w3.org/1999/02/
22-rdf-syntax-ns#>

SELECT ?subjectName, ?objectName
WHERE
 ( ?person1 foaf:name ? subjectName)
 ( ?person1 rdf:type foaf:Person )
 ( ?person2 foaf:name ? objectName)
 ( ?person2 rdf:type foaf:Person )
 ( ?person1 foaf:knows ? person2)
```

Fig. 6. An example of SPARQL query

2) Strength of person-to-community social relationship calculation: Our main idea to enrich the discovered communities is based on the common sense that now that friends are people sharing common interests, interests between friends have relatively higher similarity. For example, Jack is a good friend of Peter, and Jack likes soccer very much. We can deduce that Peter has greater chance to take part in some soccer related activities with Jack. Based on the common sense, we can use the friendship to extend the communities discovered based on contents of interests. We calculate the social relationship from a person to a community, and determine whether to invite the person to join the community by the computing result. The person-to-community social relationship strength is calculated through the person-to-person (community member) social relationship. The formula 2 illustrates the calculation process:

$$R_{ij} = \sum_{v_k \in C_j} r_{ik} \tag{2}$$

In formula 2, R_{ij} represents the strength of social relationship between person i and community C_j, and r_{ik} represents the strength of social relationship from person i to member k of C_j. After obtaining the strength of social relationship from the person to all the communities, we can select the high score communities to invite the person to join. Figure 7 shows a segment of this calculation process.

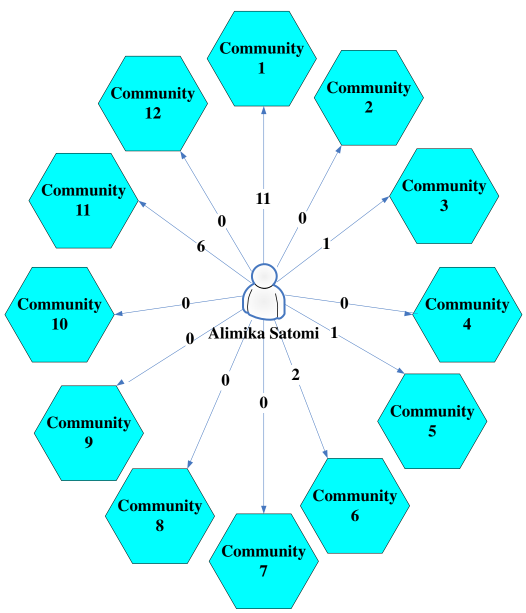

Fig. 7. An example of person-to-community social relationship computing

V. Experiments

In this section, we evaluate the performance of our method by extensive experimental study on the FOAF ontology dataset. The method is implemented in the Java language and runs on a 2.0GHz Intel CPU with 1GB memory running Windows XP.

We use k-means clustering algorithm to discover the CM2 type of communities. Through extensive experiments, we choose 12 as the clustering coefficient k which indicates the number of communities the algorithm outputs. Because the size of each community is relatively more balanced than other assignment of the coefficient k. We choose top 3 frequent occurrence feature words that occur at least 10 times as the community label. The communities discovered based on contents of personal interests are listed as Table I shows.

As to community extension, we invite a person to communities if the strength of the person-to-community social relationship is greater than 1. To evaluate the effect of community extension, we design three experiments. The first one is to measure the number of each community members before and after the community extension. The second one is to measure distribution of people join specific number of communities before and after community extension. The last one is to measure the force of MFP of each community.

A. Comparison of number of members of each community

Measuring the number of members of each community is the intuitional approach to evaluate the effect of community extension. We compare the community size before and after extension in Figure 8. Community 1 benefit most from the extension, where 159 newcomers joined in. This is caused by the greatest base of the community 1. It is easier to extend the friendship to the outside due to a large number of community members. Therefore, community 1 got the greatest increase of the community size. However, community 8 and community 10 have no increase. This may be caused by the relatively small number of community members before extension. To our delight, community 5 gets one increase based on community size of two, at the rate of 50% which is the highest increase rate of the total extension.

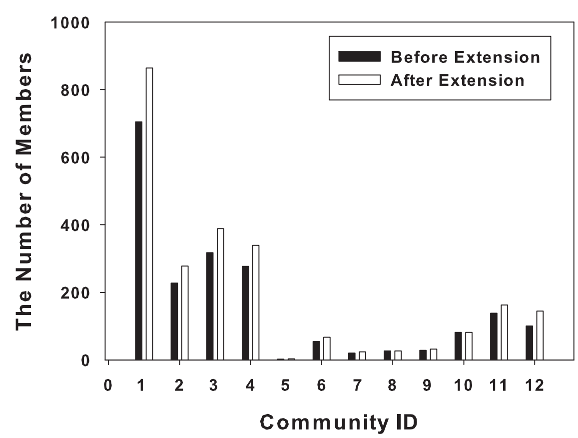

Fig. 8. Community size comparison

TABLE I

COMMUNITIES DISCOVERED BASED ON INTERESTS

Community ID	Community Label	MFP	Number of Members
1	movie-w3c-project	Tim Nowaczyk	705
2	web-semantic	Aza	227
3	technology-tv-conference	Oskar Welzl	318
4	developer-conference	Martin Pittenauer	277
5	profile	Dac	2
6	music-band-rock	Oskar Welzl	55
7	DevDaysBloggers-Locations	Martin Pittenauer	20
8	japan	jackola	27
9	network-friend	Christoph Goern	28
10	game	jackola	82
11	japanese-software-language	Josef Petr	138
12	computer-science	Ian Davis	100

B. Comparison of distribution of people join specific number of communities

How does the extension affect the individuals in the dataset? Figure 9 illustrates the effect. The horizontal axis represents the number of communities individual joined. The vertical axis represents the number of people who joined a specific number of communities. We can see from the curve that the community extension indeed affect the number of community individual joined. For instance, about 828 people join one community before extension while 1139 people join one community after, with an increase of 311.

Another interest phenomena can be found in Figure 9. Before the number of communities individual joined reaches 5, the number of people joined specific number of communities is growing after community extension, but the growing trend is decreasing. After the 5, two curves are almost identical. In both curves, only very few people joined more than 5 communities. This phenomena is easy to explain, since most individuals in the dataset has lower degree, hence most newcomers tend to be invited to join one or two communities. Only a few famous individuals with rich social relationships may be invited to more communities.

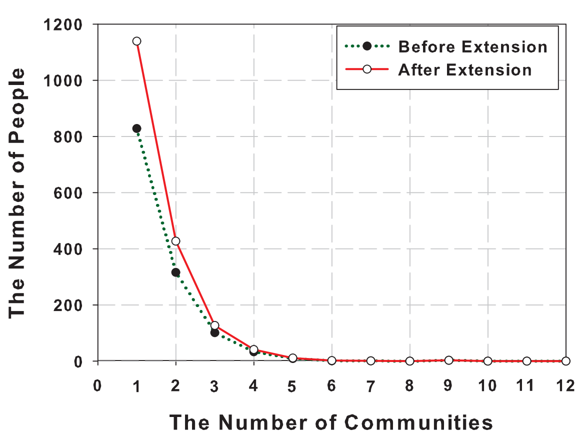

Fig. 9. Measurement of number of people joined specific number of communities before and after community extension

C. The force of MFP

This experiment aims to measure the power of MFPs of the communities. As Section IV-A.3 analyzes, the MFP is calculated by indegree of the community members in the global dataset scope. The result would not be changed after community extension, because our dataset is close. However, in the local scope within the community, the indegree of members can be changed by newcomers that take part in. Measuring indegree of MFP in local scope before and after community extension can discover the social power of the MFP.

Figure 10 shows the local scope indegree of each community's MFP. 0 local indegree of the MFP means the MFP may be known by the people outside rather than the community members. This is caused by the communities that were discovered based on the interests, without considering the social relationship. The community extension, in another aspect, supplies the social relationship of the communities. We can see that Tim Nowaczyk, who is the MFP of community 1, brought 22 newcomers to the community, whose local indegree increased from 7 to 29. He has the strongest social power among all the MFPs. Compared with Figure 8 in Section V-A, the size of community 8 and community 10 have no increase. So the indegree of their MFPs remain unchanged. The same thing happened on the MFP of community 7 due to low increase of community members, just 4.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel method of discovering communities using FOAF ontology dataset collected from the Web. The Semantic Web technologies are applied to collect the data, clean the data, and set up the dataset ontology. Then we combine contents of personal interests and social relationships to discover and extend the communities. The experimental results show that the community extension is very effective.

Much work still remains. We have focused exclusively on the foaf:knows relationship; many other factors should be considered to form the social network, such as the working place, the organization, and the co-author relationship of papers and so on. All above factors could enhance the strength

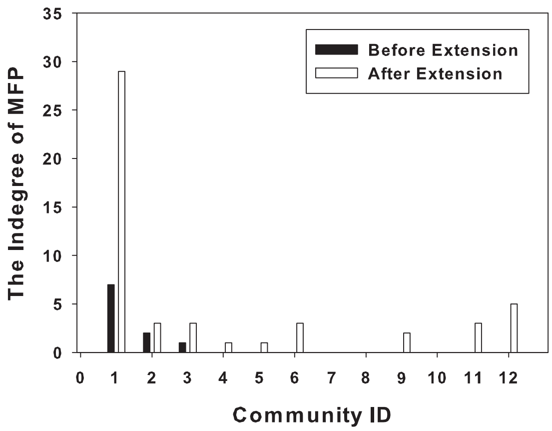

Fig. 10. Measurement of community indegree of MFPs before and after community extension

of the social relationship between social actors. More factors considered, the real world social relationship will be presented more precise by the model. As a large scale of FOAF users are Semantic Web researchers, the co-author relationship is also an important social relationship between them. DBLP bibliography (dblp.unitrier.de/) provides collaboration network data by virtue of the explicit co-author relationships among authors. In the future work, we will combine the FOAF dataset and DBLP dataset to set up a unified ontology dataset to analyze social relationships.

ACKNOWLEDGMENT

This work is supported by the Specialized Research Fund for the Doctoral Program of Higher Education ("SRFD") under Grant No.20070001073; the National Natural Science Foundation of China under Grant Nos. 90412010 and 60773162.

REFERENCES

[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," in *Scientific American*, 2001.

[2] F. Manola and E. Miller. Resource description framework (rdf) primer. w3c recommendation 10 february 2004. [Online]. Available: http://www.w3.org/TR/rdf-primer/

[3] D. Brickley and R. V. Guha. Rdf vocabulary description language 1.0: Rdf schema. w3c recommendation 10 february 2004. [Online]. Available: http://www.w3.org/TR/rdf-schema/

[4] M. K. Smith, C. Welty, and D. McGuinness. Owl web ontology language guide, 2004. w3c recommendation 10 february 2004. [Online]. Available: http://www.w3.org/TR/owl-guide

[5] A. Seaborne and E. Prud'hommeaux. Sparql query language for rdf. w3c recommendation 15 january 2008. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/

[6] Y. Sure, S. Staab, and R. Studer, *On-to-knowledge methodology*. Springer, 2003.

[7] D. Vrandecic, H. Pinto, Y. Sure, and C. Tempich, "The diligent knowledge processes," in *Journal of Knowledge Management*, 2005, pp. 85–96.

[8] A. Maedche, B. Motik, and L. Stojanovic, "Managing multiple and distributed ontologies in the semantic web," in *VLDB journal*, 2003, pp. 286–302.

[9] A. Kalyanpur, B. Parsia, E. Sirin, and J. Hendler, "Debugging unsatisfiable classes in owl ontologies," in *Journal of Web Semantics*, 2005.

[10] J. Seidenberg and A. Rector, "Web ontology segmentation: Analysis, classification and use." in *Proc. 15th International Conference on World Wide Web (WWW 2006), Edinburgh, Scotland*, 2006.

[11] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[12] H. Zhuge and X. Li, "Peer-to-peer in metric space and semantic space," in *IEEE Transactions on Knowledge and Data Engineering*, 2007.

[13] H. Zhuge, "Communities and emerging semantics in semantic link network: Discovery and learning," in *IEEE Transactions on Knowledge and Data Engineering*, 2008.

[14] H. Knublauch, R. W. Fergerson, N. F. Noy, and M. A. Musen, "The protege owl plugin: an open development environment for semantic web applications," in *Proc. 3rd International Semantic Web Conference(ISWC04)*, 2004.

[15] T. O'Reilly. (2005) What is web 2.0 - design patterns and business models for the next generation of software. [Online]. Available: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

[16] Wikipedia. [Online]. Available: http://wikipedia.org

[17] Youtube. [Online]. Available: http://www.youtube.com

[18] flickr. [Online]. Available: http://www.flickr.com

[19] Facebook. [Online]. Available: http://www.Facebook.com

[20] P. Domingos and M. Richardson, "Mining the network value of customers," in *7th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, 2001, pp. 57–66.

[21] S. White and P. Smyth, "Algorithms for estimating relative importance in networks," in *9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 266–275.

[22] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, 2004.

[23] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman, "Email as spectroscopy: Automated discovery of community structure within organizations," in *Communities and Technologies*, 2003, pp. 81–96.

[24] A. Culotta, R. Bekkerman, and A. McCallum, "Extracting social networks and contact information from email and the web," in *First Conference on Email and Anti-Spam*, 2005.

[25] D. Brickley and L. Miller. FOAF vocabulary specification, revision, 24 May 2007. [Online]. Available: http://xmlns.com/foaf/0.1/

[26] FOAF-Software. [Online]. Available: http://en.wikipedia.org/wiki/FOAF(software)

[27] B. W. Kernighan, "An efficient heuristic procedure for partitioning graphs," *Bell System Technical Journal*, vol. 49, pp. 291–307, 1970.

[28] J. Scott, *Social Network Analysis: A Handbook, 2nd edition*. Sage, 2000.

[29] M. Girvan and M. Newman, "Community structure in social and biological networks," in *Proceedings of National Academic Science*, 2002, pp. 7821–7826.

[30] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, 2004.

[31] D. Zhou, E. Manavoglu, J. Li, C. L. Giles, and H. Zha, "Probabilistic models for discovering e-communities," in *Proceedings of ACM WWW*, 2006, pp. 173–182.

[32] L. Ding, T. Finin, A. Joshi, R. Pan, R. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs, "Swoogle: A search and metadata engine for the semantic web," in *International Conference on Information and Knowledge Management*, 2004.

[33] L. Ding, L. Zhou, T. W. Finin, and A. Joshi, "How the semantic web is being used: An analysis of foaf documents," in *HICSS*, 2005.

[34] DBLP. [Online]. Available: http://dblp.uni-trier.de/db/index.html

[35] X. Li, L. Guo, and Y. Zhao, "Tag-based social interest discovery," in *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 675–684.

[36] del.icio.us. [Online]. Available: http://del.icio.us/

[37] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *Applied Statistics*, vol. 28, no. 1, pp. 100–108, 1979.

[38] Yale. [Online]. Available: http://yale.sf.net