Evolution of Social Networks Based on Tagging Practices

Hak-Lae Kim, John G. Breslin, *Member, IEEE*,
Han-Chieh Chao, *Senior Member, IEEE*, and Lei Shu, *Member, IEEE*

Abstract—Websites that provide content creation and sharing features have become quite popular recently. These sites allow users to categorize and browse content using "tags" or free-text keyword topics. Since users contribute and tag social media content across a variety of social web platforms, creating new knowledge from distributed tag data has become a matter of performing various tasks, including publishing, aggregating, integrating, and republishing tag data. In this paper, we introduce an object-centered social network based on tagging practices across different sources, and then we show how this network can be built and emerged over time.

Index Terms—Social tagging, folksonomy, social network, online community

1 Introduction

Web 2.0 sites that allow the creation of content by users or communities have become popular in recent years. The content of these sites can be virtually anything: Blog entries, message board posts, videos, audio, images, wiki pages, user profiles, bookmarks, events, and so on. Each content item on a Web 2.0 site can be a source of creating links not only between people, but also between people and content, or among content items. From this perspective, this content can enhance social relationships among participants who have an underlying shared interest within a community [1]. According to [2], object-centered sociality refers to individuals grouped together around a shared object that mediates the ties between them. Each content item can often be viewed as a social object that indicates why people affiliate with others or participate in communities [3]. Therefore, we can assume that each content item on a Web 2.0 site can be a source of social connectivity, catalyzing social networking in virtual spaces [4], [5].

However, creating new knowledge or new social relationships from various social objects remains a big challenge. Since a content item is a heterogeneous resource, with many different associated type(s), and will be defined by different sets of metadata on distributed sites, a consistent way of collecting shared interests from independent and heterogeneous sites is required [6]. Thus, we need to determine which objects can be used for interlinking among different social content across heterogeneous sites [1]. A key feature of user-contributed content in Web 2.0 sites is that the content item may be tagged, and can be shared with and commented upon by others. From this perspective, tags can be seen as objects for sharing, exchanging, and integrating a user's interests through tags attached to social objects on various Web 2.0 sites. Although a few words alone cannot identify user interests, a culture of mass participation leads to social interaction among users, and influences the use of terms in a community [7], [8]. Tagging itself is a simple process, tapping into an existing cognitive process without adding much cognitive cost [9], [10]. Tagging does not aim to create a strict classification of objects, but rather allows a user to categorize an object according to their own interests with their own keywords [9], [11]. In this paper, we aim to explore a hidden structure of tagging practices and to build implicit social network that consists of people and their using patterns. For this purpose, we use social network analysis that reveals the "structure of social relationships" in a group or a community between people [12]. For example, social networking sites such as Facebook, MySpace, or LinkedIn have strongly impacted to people, who want to connect with others or to share interests on the web. Social content sharing sites such as a microblog to Twitter, a photo uploaded to Flickr, are currently one of the fastest growing sources for creating social connections on the web. Note that because these content are fundamentally leveraging interactions by sharing or collaboration between users and sites or services, overall activities on these sites can be a path for creating new sociality.

The main contributions of this paper are summarized as follows:

1. In Section 4, we show that strong power law tailed distributions exist tag usage in the given data set,

- H.-L. Kim is with the Software R&D Center, Samsung Electronics Co., LTD., 416, Maetan-3dong, Yeongtong-gu, Suwon-city, Gyeonggi-do 443-742, Korea. E-mail: haklae.kim@gmail.com
- J.G. Breslin is with the College of Engineering and Informatics, and Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland. E-mail: john.breslin@nuigalway.ie.
- H.-C. Chao is with the Institute of Computer Science and Information Engineering and Department of Electronic Engineering, National Ilan University, No. 1, Sec. 1, Shen-Lung Road, I-Lan, Taiwan 260, and the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan. E-mail: hcc@niu.edu.tw.
- L. Shu is with the Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan. E-mail: lei.shu@ist.osaka-u.ac.jp.

Manuscript received 22 May 2010; revised 18 Jan. 2011; accepted 13 Sept. 2011; published online 6 Feb. 2012.
For information on obtaining reprints of this article, please send e-mail to: tsc@computer.org, and reference IEEECS Log Number TSC-2010-05-0079.
Digital Object Identifier no. 10.1109/TSC.2011.54.

and characterize several tags have high centrality measures in the network.

2. In Section 5, we find dynamics of network structures according to both tagging usage and user activities in PlanetRDF. From 2006 to early 2008, tagging activities in the community has been become remarkably increased, and a variety of tags has been used. On the other hand, different users in each period have highest betweenness measures. In this sense, the users in the community (although relatively small size) have slightly differentiated interests, and this makes several clusters from tagging activities.
3. In Section 6, we apply a principal component layout to the data set, which is subdivided by three months, and then evaluate how the clusters consist of users have been changed over time. We show that each cluster consists of different users based on their tagging practices. The resulting analysis reveals that tagging activities have strong efforts on changing relationships between users.

The paper is organized as follows: After a brief review of related work (Section 2), the data set and analysis methods are introduced in Section 3, and the results and findings from data analysis are exposed in Section 4. Section 5 shows how the network structure has changed over time.

2 Related Works

One of the first quantitative studies on folksonomies is presented by Golder and Huberman [13], who discuss the structure and the kinds of tags, such as what the resource is about, who owns the resource, or characteristics of the resource, and so on. They also discuss the distinction between collaborative tagging and taxonomies, whereas [11] conduct a preliminary analysis of using a set of tags between users and propose a model that tags are a kind of an object for connecting users and resources. Similarly, [14] explore social links on folksonomies that users with similar tags are more likely to be friends, and [4] investigate user-induced links between resources in collaborative tagging systems. These papers may consider that tags are methods for describing users' interests. In particular, [15] points out that folksonomies reflect the population's conceptual model and as side benefits enable both "possibilities of aggregation and analysis" and the "creation of communities."

Although many researchers investigate social relationships by analyzing folksonomies or tags [4], [7], [16], these relationships have a distinction with conventional social network [17], [18]. The idea of social networks has been studied in the areas of social science for decades. The focus lies on analyses associated with the ties between people in families, organizations, and even towns or countries. However, with the emergence of the social web, social networks on online communities have created new connection types based on user's activities or interests. The way of constructing social networks between people is classified according to the medium of connection: Ego centered and object centered. An ego-centered network is about individuals associating with each other by means of their personal network in which they have a specific relationship. A connection can be built by the profiles of people on the web. Examples of this type of network include Facebook, LinkedIn, MySpace, or Friendster. On the other hand, an object-centered network refers to individuals' grouped together around a shared object that mediates the ties between them [19]. People have been interested in who has similar patterns in terms of using tags, videos, or bookmarks on social websites. In this sense, these objects can be considered as the key that makes and encourages social ties. Examples of object-centered sociality include Delicious (bookmarks), Flickr (photos), YouTube (videos), SlideShare (presentations), or Last.fm (music).

Engeström [20] argued that social network theory fails to recognize such real-world dynamics because most notions of social networks are restricted to people. He stressed that social networks on the web should consist of people who are connected by a shared object on behalf of making up of only people. Recently, several works proposed different approaches to construct social networks from folksonomies. Mika [21] uses a graph-based approach to build a network, analyzing users-tags association graphs, and Cattuto et al. [22] introduce a network of tag co-occurrence using classical network measures such as characteristic path length and clustering efficient. Both efforts show statistical properties of folksonomy-based networks collected from Delicious or Flickr. However, the dynamics between users in a community according to tag usage are not investigated in detail. For example, users assign a variety tags based on their interests, and tags can be differentiated over time. In this sense, social networks from folksonomies should be investigated over time. Those works tend to show a power law and a long tail effect from folksonomies, which is a result of tagging activities by many people. These results can be useful to investigate popular items or trends in large groups of people using a few popular tags. Our approach is similar in spirit to several works on social networks on folksonomies. However, unlike those works, which take a broad view of folksonomies, we will focus on investigating tag usage patterns in a small community and then on revealing a change of social structure according to them over time.

3 Experimental Data Sets

3.1 Data Collection

For this study, we collected RSS feed data from Planet RDF, which is an aggregation of blogs belonging to Semantic Web enthusiasts and hackers and is updated hourly. The selection of blogs was performed by one of the individuals—Dave Beckett, and most of them are very famous and/or active in the Semantic Web area. The 58 members included blogs of individual users as well as organizations. The following shows some users we will mention in this paper:

Alexandre Passant, Andrew Matthews, Benjamin Nowack, Bob DuCharme Clark and Parsia, Dan Brickley, Dan Connolly, Danny Ayers, Danny Weitzner, Dave Beckett, Dean Allemang, Edd Dumbill, Elias Torres, Frederick Giasson, Gregory Williams, Henry Story, Ian Davis, Ivan Herman, Jeen Broekstra, John Breslin, John Goodwin,

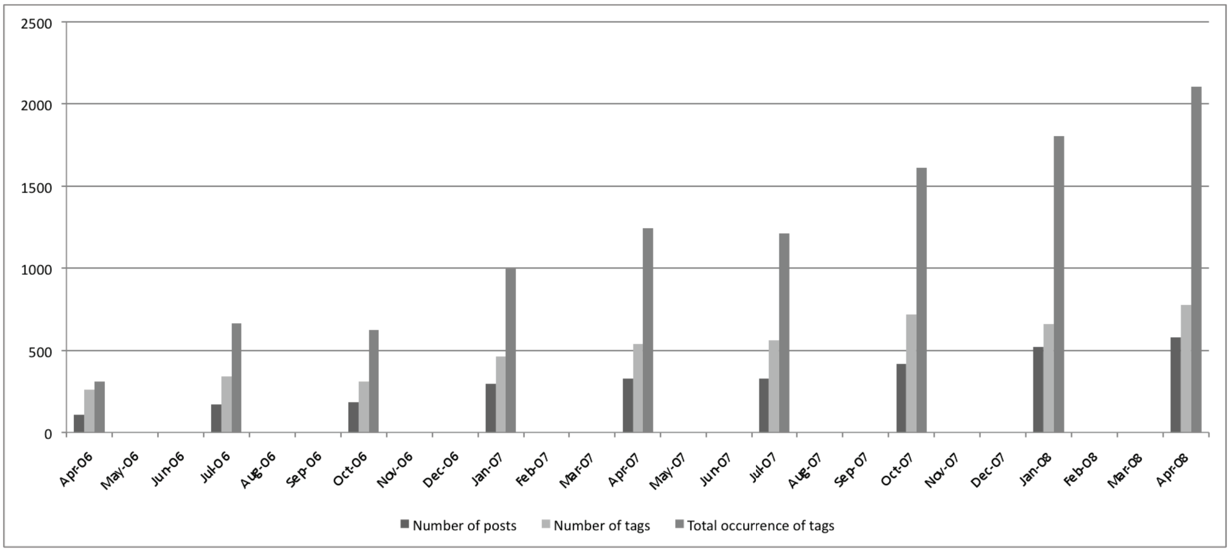

Fig. 1. Descriptive statistics of data over time. The total number of members in the given data set has not been changed, while tagging activities have been increased. In particular, after 2007, members have used tags on their resources.

Kingsley Idehen, Lee Feigenbaum, Leigh Dodds, Libby Miller, Morten Frederiksen, Norm Walsh, Ora Lassila, Orri Erling, Phil Dawes, Richard Cyganiak, Tim Berners-Lee, Uldis Bojars.

The data set makes up of a folksonomy with $|U| = 58$ users, $|T| = 2,750$ tags, and $|R| = 4,386$ resources, related by in total $|Y| = 18,317$ tag assignments, covering January 2004 to the beginning of April 2008. As we described, this paper aims to investigate dynamic change of network structures on tagging practices. In this sense, the data set from the beginning of 2006 to 2008 seems about right because there was less tagging activity in this community in 2006, but tagging activity has increased since the middle of 2007.

Fig. 1 illustrates the number of tags, posts, and members for each of the given periods (by three monthly intervals). For example, F_{0604} contains all tag assignments performed from February 2006 to the end of April 2006, together with the involved users, tags, and resources. The posts vary from the lowest level of 261 in April 2006 to the highest one of 719 in October 2007. We did not consider the data before January 2006 because of lack of data. The number of users observed in the given periods range from 28 in April 2006 to 43 in April 2008. The tags assigned by the users are diverse and are not limited only to Semantic Web-related topics. On average, each user published 12.6 posts with 1.7 associated tags every three months. The quarterly average number of tags assigned by a user ranges from the lowest of 3.9 in 2006 to the highest of 12.7 in January 2008. The number of members differed according to their level of activity, and some members who were included at the beginning of the period were omitted by the end, whereas others had joined in the meantime. As a consequence, some individuals are not present in all periods. Although most of the published blog feeds were associated with tags, few of them did not have any.

3.2 Analysis Methods

For analyzing social networks from tagging practices, two mode networks, which aim to obtain ties between two sets of entities, can be applied. That is, tag-based network refers to a network about subsets of actors who participate in the same social tagging activities (i.e., they have a set of same tags). Formally, two mode networks have a structure $A = (M, N, R, w)$, where M and N are disjoint sets of nodes, R is the set of arcs from an initial node in the set M to a terminal node in the set N, and $w : R \rightarrow IR$ is a weight. The set R can be viewed also as a relation $R \subseteq M \times N$. Thus, the network model for tagging practices can be defined as a set of users who tag (e.g., bloggers) $N = \{n_1, n_2, \ldots, n_i\}$ as the first of the two modes, and $M = \{m_1, m_2, \ldots, m_i\}$ as the second mode. For analyzing their dynamics over time, we transformed the user-tag two-mode network data into one-mode networks $A_{user} = (M_1, \varepsilon_1, \omega_i)$ and/or $A_{tag} = (N_2, \varepsilon_2, \omega_2)$. This approach simplifies the analysis of the characteristics of users and tags, respectively, as the analysis of two-mode networks is usually complex. In particular, this improves the readability for a large data set and offers an intuitive interpretation for our data.

There were a number of possible approaches that we could adopt to analyze this network. In our analysis, we focus on individual properties of a network, in particular, *centrality measures*. Centrality is a structural attribute of nodes in a network, and there are two major centrality measures [23], *degree* and *betweenness* centrality:

1. *Node degree centrality*. Actors at the core have more connections than actors at the periphery of a network. These actors occupy a structural position that exchanges larger volumes of information. Thus, actors who have high-degree centrality are in a position of power within a network. According to Wasserman and Faust [24], the degree centrality of an actor v is given by $C_{D(v)} = d(v)/n - 1$, where $d(v)$ is the degree of the actor v and *n-1* is the number of the remaining nodes in the network. Thus, $C_{D(v)}$ is the normalized index of the degree of the actor.
2. *Node betweenness centrality*. Centrality is a structural attribute of nodes in a network, and there are two

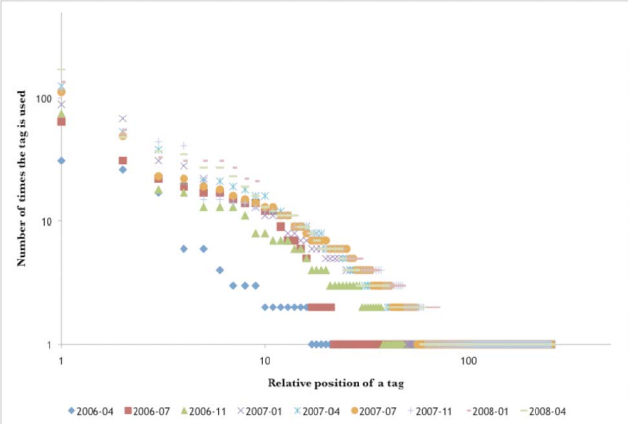

Fig. 2. Tag frequencies over time. The horizontal scale gives the logarithm of the relative position (where the most used tag is in position 1, the second most used tag is in position 2, etc.), while the vertical scale gives the logarithm of the frequency of use.

major centrality measures [23]. Betweenness refers to the ability to bridge intermediary actors or events. A high betweenness index means an actor or an event occupies a broker role within the network. An actor with a high index could control the flow of information or exchange of resources. A betweenness centrality for actor v is the sum of the probabilities for all pairs of actors j and k that actor v is involved in the pair's geodesic(s), which refers to the shortest route between two points [12] [24]. $C_B(v) = \Sigma\sigma_{jk}(v) / \sigma_{jk} \ (j \neq v \neq k \in V \text{ and } j \neq k)$, where σ_{jk} is the number of geodesic paths from j to k, and $\sigma_{jk}(v)$ is the number of shortest geodesic paths from j to k that are formal through actor v.

4 Data Analysis

As expected, the tagging behavior in the given data set displays a power law tailed distribution. Fig. 2 shows power law distributions in the use of tags generated from our data set. The points are nearly in a straight line in the log-log scale, indicating that the distribution follows the power law. The distribution shows that, in most cases, users only use each tag a few times, and it has been observed by others that very popular tags are the product of thousands of users agreeing to use the same tag to label the same concept [25]. When a particular tag becomes very popular, it is used even more (e.g., "Web 2.0" instead of "Web 20"). The strong correlation between the number of users using a particular tag and its popularity suggests that there is a strong sense of agreement on common tags among users. Not surprisingly, the most popular tag (e.g., "Semantic Web") was used more than 900 times by 53 different users altogether in the given data. However, different tags have been used, meaning that all the members may have different interests of topics. More details are presented in Section 5.

We also analyze who the most active taggers are in this community. Fig. 3 shows the distribution of users' tagging activities in a plotted log-log scale. Compared to Fig. 2, this figure does not show a power law distribution that appears in each position over time. One possible reason is that many

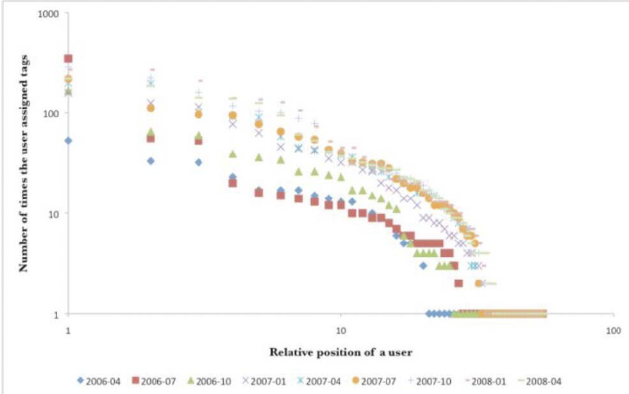

Fig. 3. Frequency of tag assignment per user. In 2006, only a few members used a number of tags on their resources, whereas more users assigned a set of tags in 2007 and 2008.

users within the community published their posts with a set of tags regularly. It can be interpreted that tagging is a common activity in this community. In fact, most users used tags when describing their posts in given periods (except the beginning of 2006), and there are some very active members. For example, Idehen, Breslin, and Powers have considerably high assignment scores $|Y|_{\text{Idehen}} = 2{,}178$, $|Y|_{\text{Breslin}} = 2{,}068$, and $|Y|_{\text{Powers}} = 1{,}340$, respectively, whereas the mean of $|Y|$ for the duration of the study is 333.02.

5 Network Analysis

In this section, we discuss the results of our social network analysis with respect to tagging practices observed through Planet RDF. Note again that a tag-based network refers to the social structure that people are connected when they use the same set of tags. We will first have a look at the changes in community structure at a given time. This reveals how tagging practices influence the community structure. We then describe the dynamics of tag usage and its popularity over time, followed by observations regarding the changes of individual users' tagging activities. We also present the correlations between users and tags via an observation of the dynamics of these indirect interest-based relationships over time.

5.1 Community Structure

Table 1 shows three measures for describing network structures of this community. The clustering coefficient is a measure of degree to which nodes in a graph tend to cluster together, whereas geodesic distance refers to the distance between two vertices in a graph is the number of edges in a shortest path connecting them. Density is a measure of the general level of links among the vertices in a graph. This network has the so-called *small-world property* [26], which means that the average separation between the nodes is rather small: A short path along the links between most pairs of nodes [27]. The scores for the average clustering coefficient $C_G = 56.72$ and density $\Delta_N = 18.04$ in 2008 are higher than the ones in 2006 and 2007, while the indices for geodesic distance $d(i,j)$ have decreased. The models from October 2007 to April 2008 demonstrate a high degree of clustering, but they also have their shortest

TABLE 1
Clustering Coefficient, Geodesic Distances, and Density over Time

	Apr-06	Jul-06	Oct-06	Jan-07	Apr-07	Jul-07	Oct-07	Jan-08	Apr-08
C_G	11.51	20.8	37.98	30.01	42.00	31.25	69.81	55.12	58.31
$d(i,j)$	1.64	1.65	1.52	1.50	1.33	1.52	1.46	1.42	1.39
Δ_N	2.15	3.71	6.58	7.70	11.74	9.12	15.14	16.07	20.01

distances between random nodes. One can assume that tag-based networks change considerably over time. This means that the users in the community can be tied among different users by using their tags, although the users may have social network relationships within the community. In addition, collective tagging activities, which consist of the relationships of users, tags, and items may reveal how social structure based on tags change in the given period. To address this assumption, we need to consider the individual user's activities to observe the network evolution. We now perform a detailed analysis to learn about the evolution of the community in regards to the tagging practices taking place within.

5.2 Tag Centrality

The popularity of 2,750 tags used by the community in the given period varied, although some trends were clearly visible. Not surprisingly, "*Semantic Web*," "*rdf*," "*Web 2.0*," "*sparql*," and "*web*" are the core tags in this set of tags over time, and technology-oriented tags such as "*python*," "*xml*," *and* "*internet*" also demonstrated a consistent high ranking throughout. Note that the tag "general" has high scores in the analyzed period. We assume that most members use WordPress, and this tag is the default for categorization in this system. Table 2 shows the 10 tags, which have higher scores in degree centralities. Due to formatting limitations, some tags are shortened: sw, sm, sn, dp, and ws refer to Semantic Web, social media, social network, data portability, and web services, respectively.

In the early periods, the tags that have the highest degree centralities (e.g., work, general, bicycle, Ireland), are not strongly associated with Semantic Web and its technologies. In this perspective, these tags do not contribute much to the picture of social structure or characteristics of the community-based tagging practices. After 2006, the centralities of these tags faded. However, the domain of the tags becomes broader in 2007. More specifically, there is a surge in tags related to social technologies, such as "*social software*," "*social media*," "*social network*," and "*twitter*." This reflects the increase in popularity of social software and technologies at the time, and it is also reflected by an increase in posts related to the "social web" domain, which became popular in 2007. Similarly, 2008 sees the introduction of "*data portability*," whereas the centralities of "*rdf* ($\mathbf{C}_D = 3.3$)" and "foaf ($\mathbf{C}_D = 2.5$)" in F_{0801}, and "foaf ($\mathbf{C}_D = 3.3$)," and "sioc ($\mathbf{C}_D = 2.2$)" in F_{0804} occur significantly, converging toward more mainstream topics. Note that Semantically Interlinked Online Community (SIOC) and Friend of a Friend (FOAF) are one of the most popular RDF vocabularies in the Semantic Web area. These

TABLE 2
Top 10 Most Central Tags over Time

Ranks	Apr-2006	Jul-2006	Oct-2006	Jan-2007	Apr-2007	Jul-2007	Oct-2007	Jan-2008	Apr-2008
1	general (5.6)	sw (10.5)	sw (4.5)	sw (2.9)	sw (6.1)	sw (5.3)	web (2.5)	sw (5.4)	sw (6.5)
2	sw (3.5)	web2.0 (8.1)	ws (2.4)	readify (1.7)	rdf (2.7)	sm (3.7)	sw (2.4)	tech (4.7)	foaf (3.3)
3	python (3.2)	dataspace (6.9)	web2.0 (2.2)	general (1.7)	internet (2.1)	web (3.1)	sm (2.3)	rdf (3.3)	web2.0 (2.9)
4	work (2.9)	rdf (6.6)	rdf (2.1)	web2.0 (1.7)	web2.0 (1.9)	rdf (2.7)	web2.0 (1.5)	foaf (2.5)	sm (2.7)
5	bicycle (2.1)	sparql (6.1)	dataspace (1.7)	rdf (1.6)	blog (1.7)	web3.0 (1.6)	general (1.3)	sm (2.1)	sioc (2.2)
6	rss (1.4)	howto (5.3)	xml (1.6)	.net (1.4)	ss (1.6)	web2.0 (1.5)	rdf (1.0)	web2.0 (1.7)	dp (1.9)
7	planetrdf (1.0)	javascript (5.3)	general (1.5)	internet (1.1)	ireland (1.6)	general (1.5)	tech (0.8)	sparql (1.6)	ss (1.8)
8	dig (1.0)	ws (5.3)	sparql (1.5)	blog (1.1)	.net (1.4)	ai (1.4)	blog (0.8)	general (1.4)	rdf (1.4)
9	privacy (1.0)	os (5.2)	.net (1.4)	ireland (1.0)	sn (1.4)	blog (1.4)	galway (0.6)	image (1.4)	general (1.4)
10	xml (1.0)	screencast (5.2)	sioc (1.1)	ws (1.0)	deri (1.3)	owl (1.1)	web3.0 (0.6)	ggg (1.2)	sn (1.3)

The numbers in the table show degree centrality scores.

vocabularies are positioned to make all social web data interoperable at a semantic level. To sum up, many tags were used to describe higher levels of abstraction of the resources in 2006, whereas members began using similar tags in their posts beginning in 2007, which this is a possible reason for the change in community structure.

5.3 User Centrality

Idehen, Brickley, and Breslin are the highest indices of degree centrality in the network, and their role changed constantly. For example, Idehen has the most central position because he is the most active ($\mathbf{C}_D = 11.37$ in July 2006) in early periods, whereas Breslin and Brickly take top spots after July 2007. Dawes and Barstow ($\mathbf{C}_D = 12.88$ and 6.34 in April 2006, respectively) were active in the first quarter of 2006, but their degree centrality plummeted after the beginning of 2007. In contrast, new individuals gained higher centrality (e.g., Ebiquity, Powers, Passant, and Breslin).

A high betweenness index for individuals suggests that they serve as an intermediary between others. While only seven users in F_{0604} have valid indices, other periods have many more. The rank order has changed significantly over time. For example, Dawes and Idehen are the highest index in F_{0604}, as shown in Fig. 4, but they are not ranked in F_{0704} and F_{0804}. On the other hand, Breslin and Dumbill have the highest betweenness centralities in the network in F_{0704}, and Brickley and Ebiquity in F_{0804} are ranked first, respectively. This suggests that they have a significantly high number of connections to other users in the network. It is also interesting to note how many users have 0 betweenness indices in the set of 58 users in given periods. More members in (B) and (C) appear to be relatively more powerful than few members in (A) April 2006 by this measure. A low betweenness index indicates that actors have no power over the flow of communication in a network. As tagging activities have become popular, members have much betweenness power in the network and they are connected among them by using a set of same tags. In this sense, tagging practices by members may play a critical role for the flow of communication within the network.

6 Tag-Based Networks over Time

We contrast the graphs for these tag-based networks during the given periods to observe whether specific groups of individuals tend to share a number of tags and whether they "stick together" in a "common-interest" cluster over extended periods of time. We apply a principal components layout to construct a cluster based on tagging practices. Instead of describing the network along a single dimension (e.g., betweenness centrality), the principal components layout decomposes the relationship matrix, placing nodes near each other to the extent that they are structurally equivalent. Using a principal components layout, the networks are clearly divided into several areas according to tagging practices (users and their tag assignments) over time. As shown in Fig. 5, clusters in the network varied in given time, and changes occurred between April 2006 and April 2007. Thus, we focus on three six-month periods: F_{0704}, F_{0710}, and F_{0804}. One can see that there are three clusters in F_{0704} and F_{0710}, respectively, while F_{0804} has four

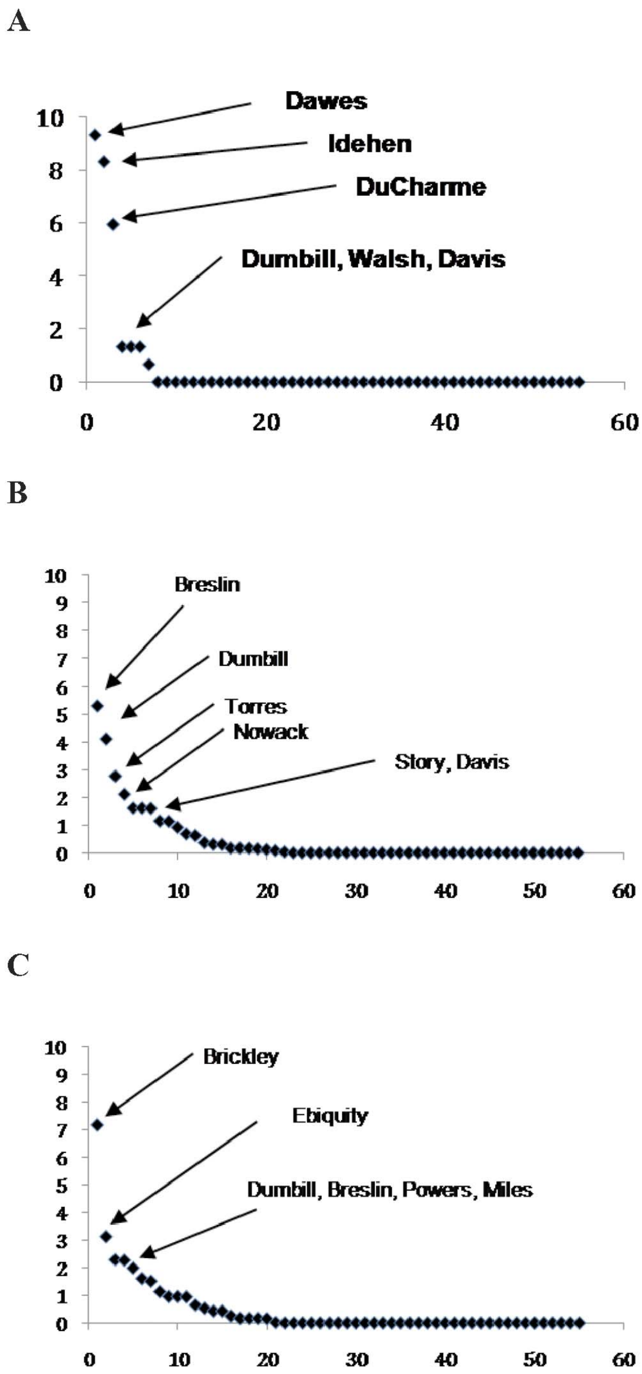

Fig. 4. The highest betweenness centrality of users for the periods (A) April 2006, (B) April 2007, and (C) April 2008.

clusters. Table 3 shows the details regarding users and tags for each cluster in Fig. 5. AT first glance, it can be noticed that, in general, the number of tags in each cluster during these periods went through a considerable change, compared to F_{0604}.

We can characterize types of tags by their role in the network. Clusters are formed via a number of *defining tags*, e.g., the defining tags for cluster $\mathbf{C}_{10}$ are *Semantic Web* ($R = 39.2\%$), *social media* (13.2 percent), *Web 2.0* (8.6 percent), *web* (8.5 percent), *database* (8.6 percent), *rdf* (4.0 percent), *owl* (3.5 percent), *security* (2.5 percent), *mobile computing* (2.5 percent), *cs* (2.0 percent), *foaf* (1.5 percent), *ai* (1.5 percent), and *openid* (1.0 percent). Clusters are bridged via *bridging*

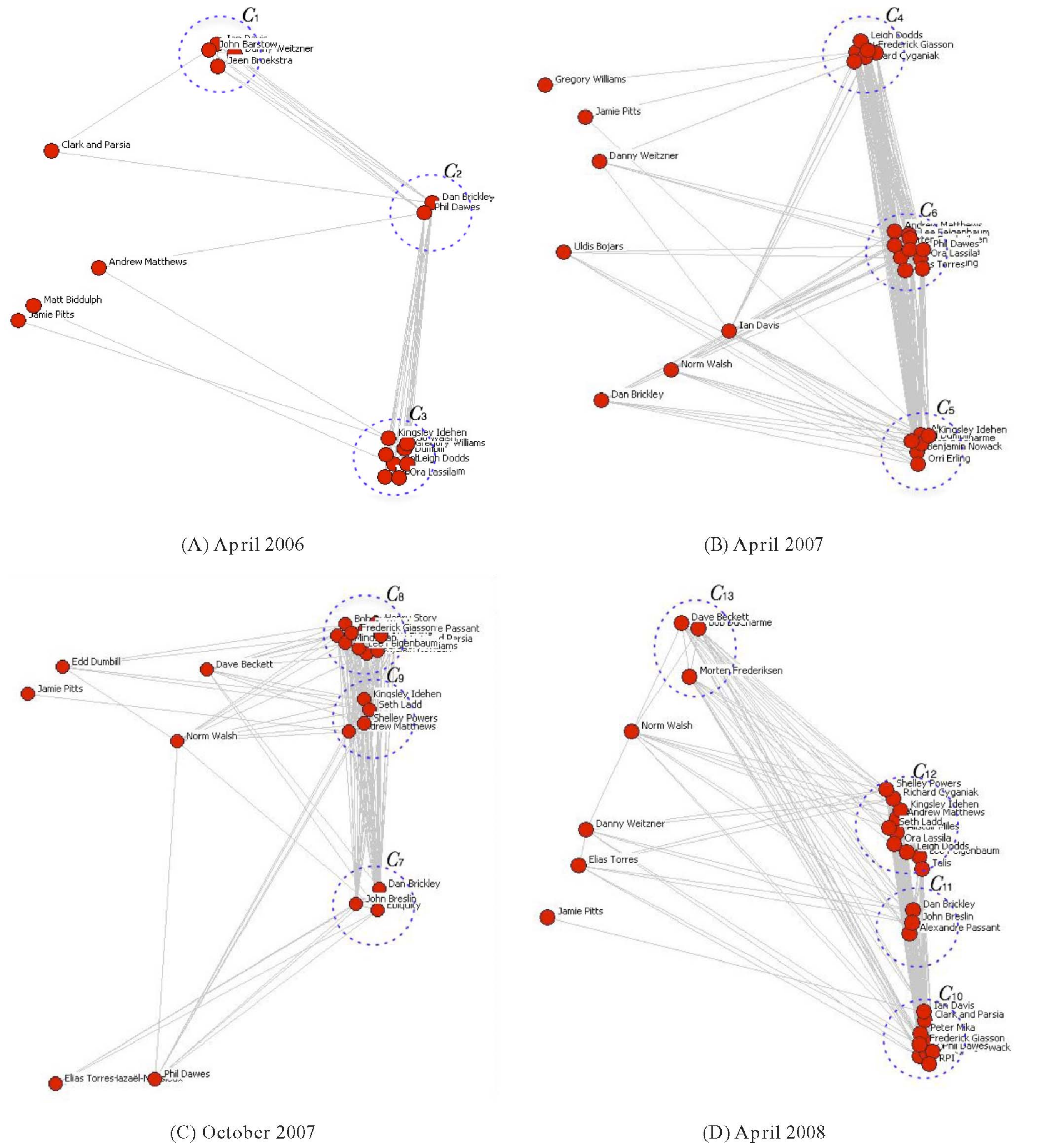

Fig. 5. Clusters in April 2008 using the principle components layout.

tags—tags that are popular with individual users within multiple clusters. A set of tags play a role in defining a cluster, and one tag can be used in different clusters. Although clusters are identified by the overall tagging activities of users, a set of tags can be linked across clusters. For example, the tags *Semantic Web*, *foaf* and *rdf*, weave clusters together (e.g., $\mathbf{C}_{11}$ and $\mathbf{C}_{12}$). However, a bridging object does not necessarily need to be popular with the majority of cluster users, e.g., *general* is popular only with the minority of users in clusters $\mathbf{C}_1$ and $\mathbf{C}_2$, yet it brings them together. Intercluster tags tend to represent very general concepts, e.g., *programming*, *web*, *Web 2.0*, *general*, *twitter*, *internet*, and *rss*.

We see that there was a higher use of tags in 2008, and more of these tags had a bridging role between clusters. Considering individual clusters, we see that there are some similarities in between. In particular, Passant and Breslin are included within both cluster $\mathbf{C}_7$ (October 2007) and cluster $\mathbf{C}_{11}$ (April 2008). Cluster $\mathbf{C}_7$ was defined by the following tags: *Semantic Web, Web 2.0, web, rss, twitter*, and

TABLE 3
Subgroups Based on the Same Tags in the Given Period

Period	Cluster	Users	Tags
F_{0604}	C_1	Barstow, Broekstra, Davis, Weitzner	general
	C_2	Brickley, Dawes	foaf, general, rdf, rss, semantic web
	C_3	Dodds, DuCharme, Dumbill, Feigenbaum, Idehen, Lassila, Williams, Walsh	semantic web, xml
F_{0704}	C_4	Breslin, Cyganak, Dodds, Giasson, Passant, Story, Davis	flickr, foaf, general, rss, twitter, semantic web
	C_5	Dumbill, DuCharme, Erling, Idehen, Miles, Nowack	foaf, javascript, metadata, microformat, microsoft, openid, programming, rdf, semantic web, sparql, virtuoso, xml
	C_6	Clark and Parsia, Dawes, Ebiquity, Feigenbaum, Fredenksen, Lassila, Matthews, Torres	database, general, openid, programming, python, rdf, semantic web, software, sparql
F_{0710}	C_7	Breslin, Brickley, Passant	foaf, general, internet, linked data, rdf, semantic web, sioc, social network, sparql, web2.0
	C_8	Miles, DuCharme, Clark and Parsia, Ebiquity, Story, Erling	database, dublincore, java, metadata, ontology, owl, rdf, retrieval, search, semantic web, skos, web2.0, xml
	C_9	Matthews, Nowack, Idehen, Dawes, Ladd, Powers	ajax, foaf, php, politics, programming, rdf, ruby, semantic web, social network, sparql, web2.0
F_{0804}	C_{10}	Nowack, Clark and Parsia, Ebiquity, Story, Davis, Erling, Mika, Talis, RPI	ai, cs, database, foaf, mobile computing, openid, owl, rdf, security, semantic web, social media, web, web2.0
	C_{11}	Passant, Brickley, Breslin	data portability, foaf, general, google, linkeddata, music, openid, rdf, semantic web, sindice, sioc, social network, sparql
	C_{12}	Miles, Matthews, Idehen, Dawes, Cyganiak, Ladd	atom, foaf, general, identity, openlink, python, rdf, semantic web, social networking, web2.0, xpath
	C_{13}	DuCharme, Beckett, Frederiksen	ebooks, metadata, music, php, publishing, rdf, sparql, xml

business; whereas $\mathbf{C}_{11}$ was defined by *Semantic Web* ($R = 22\%$), *foaf* (21 percent), *sioc* (10 percent), *data portability* (10 percent), *social network* (6.8 percent), *sparql* (5.1 percent), *rdf* (3.4 percent), and *openid* (2.8 percent), among others. This suggests that while the focus of cluster $\mathbf{C}_7$ changed, it still maintained its members as seen in $\mathbf{C}_{11}$ and also included a new user (i.e., Brickley). Another interesting similarity can be observed between clusters $\mathbf{C}_8$ and $\mathbf{C}_9$ (2007) and clusters $\mathbf{C}_{10}$ and $\mathbf{C}_{12}$ (2008). In 2007, Miles and DuCharme are in cluster $\mathbf{C}_8$, whereas Nowack and Idehen are both in cluster $\mathbf{C}_9$. Then, in 2008, Nowack and DuCharme are placed in $\mathbf{C}_{10}$ and $\mathbf{C}_{13}$, respectively, while Miles is located in $\mathbf{C}_{12}$ with Idehen and Dawes. By analyzing the defining tags of these clusters, we can deduce that users are located in different clusters according to their tag usage. We found that some of the tags from $\mathbf{C}_4$ and $\mathbf{C}_8$ are indeed bridging tags for $\mathbf{C}_{10}$ and $\mathbf{C}_{11}$ (e.g., *foaf*, *rdf*, *web2.0*). Thus, it is possible that the increased participation in community discussions led to clusters from 2007 being reorganized into multiple interest clusters in 2008.

As noted, although F_{0604} has some clusters (i.e., $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_3$), it seems difficult to suggest that tags can identify clusters due to the lower usage of tags. On the other hand, other clusters can be considered meaningful cohesive subgroups, reflecting their interests based on tagging activities. This seems to indicate that social relationships cannot be driven out by only a few tags; however, by looking at tagging practices, an aggregated set of tags and their popularity can reflect users' interests. In addition, via the analysis of these patterns over time, one can see how a user's interests can change, even though the community is relatively small and topic centered.

7 Conclusion

Social networks help us work together over common activities or interests. Currently, social websites offer common functionalities to construct relationships based on various objects: networks of not only online profiles consisting of friends or colleagues but also online activities such as blogging, comments, or tagging, and so on. User-generated content is an important medium not only for supporting social interaction and communication but also for incorporating networking among individuals participating in information activities for organizing and sharing a variety of resources.

We focused on analyzing the hidden structure of tagging practices for community members in a small and well-organized community—PlanetRDF. As stated in the beginning of this paper, tagging practices in online communities are quite dynamic, even if the community is relatively small. Via the longitudinal analysis, we found out that the interests of the community have evolved and this is influenced by the relationships among users. Using social network analysis, we demonstrated the popular tags and users in the given data. Degree and betweenness centrality measures are used. In particular, the betweenness measure showed which tags or people played a role in connecting other tags or users. We also investigated clusters based on the tagging practices of all users. This analysis showed the relationships between users and their tagging practices. We identified several clusters and characterized their features. Our results showed that the top popular tags connected most of the users and resources, which motivates us to utilize tags for sharing social interests among users in the community.

However, the result of analyzing tagging practices does not make it easier to express this evolution of relationships in an explicit manner. Currently, most tagging systems do not provide a standardized format to share, exchange, and reuse tag data among users or communities. The limitation in terms of representation can be corrected via Semantic Web technologies by providing more specific ontological terms to represent people, relationships, and their behaviors. Once all this information is exposed to machines, intelligent technology can be employed to constantly elicit new knowledge by observing online social practices and behaviors. Then, our next goal is to draw similarities between the implicit social structure of the community given their tagging practices (i.e., via tag-based network) and the explicit social structure as Semantic Web descriptions of a person and their social network. Furthermore, our next challenge will be to investigate the dynamic social structure of tagging practices from microblogs such as Twitter. Many users generate a large amount of content that covers a variety of topics with limited characters as well as interaction with others. Some efforts already revealed the social influences by analyzing the follower and following of Twitter. However, as we demonstrated, a tag-based network can be more dynamic feature.

Acknowledgments

This work was funded by the Science Foundation Ireland under grant SFI/08/CE/I1380 (Lion 2). Lei Shu's research work was supported by Grant-in-Aid for Scientific Research (S)(21220002) of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] H.L. Kim, "Leveraging a Semantic Framework for Augmenting Social Tagging Practices in Heterogeneous Content Sharing Services," PhD thesis, Dept. of Eng., Nat'l Univ. of Ireland, Galway, 2010.

[2] K.K. Cetina, "Sociality with Objects: Social Relations in Post-Social Knowledge Societies," *Theory, Culture, and Soc.*, vol. 14, no. 4, pp. 1-30, 1997.

[3] J. Breslin and S. Decker, "The Future of Social Networks on the Internet: The Need for Semantics," *IEEE Internet Computing*, vol. 11, no. 6, pp. 86-90, Nov.-Dec. 2007.

[4] C. Au Yeung, N. Gibbins, and N. Shadbolt, "User-Induced Links in Collaborative Tagging Systems," *Proc. 18th ACM Conf. Information and Knowledge Management (CIKM '09)*, pp. 787-796, 2009.

[5] Y.L. Jeng, T.C. Huang, and Y.M. Huang, "Web 2.0 Based Learning Management System for Supporting SCROM & QTI," *J. Internet Technology*, vol. 9, no. 5, pp. 307-312, 2008.

[6] T. Gruber, "Collective knowledge systems: Where the Social Web-Meets the Semantic Web," *J. Web Semantics*, vol. 6, no. 1, pp. 4-13, 2008.

[7] S. Sen, J. Vig, and J. Riedl, "Tagommenders: Connecting Users to Items through Tags," *Proc. 18th Int'l Conf. World Wide Web (WWW '09)*, pp. 971-980, 2009.

[8] J. Trant, "Studying Social Tagging and Folksonomy: A Review and Framework," *J. Digital Information*, vol. 10, no. 1, pp. 1-42, 2009.

[9] A. Mathes, "Folksonomies—Cooperative Classification and Communication through Shared Metadata," http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html, 2004.

[10] H. Halpin, V. Robu, and H. Shepherd, "The Complex Dynamics of Collaborative Tagging," *Proc. 16th Int'l Conf. World Wide Web (WWW '07)*, pp. 211-220, 2007.

[11] C. Marlow, M. Naaman, D. Boyd, and M. Davis, "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read," *Proc. ACM 17th Conf. Hypertext and Hypermedia (HYPERTEXT '06)*, pp. 31-40, 2006.

[12] S.P. Borgatti and M.G. Everett, "Network Analysis of 2-Mode Data," *Social Networks*, vol. 19, no. 3, pp. 243-269, 1997.

[13] S. Golder and B.A. Huberman, "The Structure of Collaborative Tagging Systems," *J. Information Sciences*, vol. 32, no. 2, pp. 198-208, 2006.

[14] R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer, "Folks in Folksonomies: Social Link Prediction from Shared Metadata," *Proc. Third ACM Int'l Conf. Web Search and Data Mining (WSDM '10)*, pp. 271-280, 2010.

[15] E. Quintarelli, "Folksonomies: Power to the People," *Proc. Int'l Soc. Knowledge Organization. Meeting*, 2005.

[16] C. Cattuto, D. Benz, A. Hotho, and G. Stumme, "Semantic Grounding of Tag Relatedness in Social Bookmarking Systems," *Proc. Seventh Int'l Conf. Semantic Web*, pp. 615-631, 2008.

[17] U. Farooq, T.G. Kannampallil, Y. Song, C.H. Ganoe, J.M. Carroll, and L. Giles, "Evaluating Tagging Behavior in Social Bookmarking Systems: Metrics and Design Heuristics," *Proc. Int'l ACM Conf. Supporting Group Work (GROUP '07)*, pp. 351-360, 2007.

[18] A. Mislove, H.S. Koppula, K.P. Gummadi, P. Druschel, and B. Bhattacharjee, "Growth of the Flickr Social Network," *Proc. First Workshop Online Social Network (WOSP '08)*, pp. 25-30, 2008.

[19] K.K. Cetina and U. Bruegger, "The Market as an Object of Attachment: Exploring Postsocial Relations in Financial Markets," *Canadian J. Sociology*, vol. 25, no. 2, pp. 141-168, 2000.

[20] J. Engestrm, "Why Some Social Network Services Work and Others Don't—Or: The Case for Object-Centered Sociality," http://www.zengestrom.com/blog/2005/04/why-some-social.html, 2005.

[21] P. Mika, "Ontologies Are Us: A Unified Model of Social Networks and Semantics," *J. Web Semantics*, vol. 5, no. 1, pp. 5-15, 2007.

[22] C. Cattuto, C. Schmitz, A. Baldassarri, V.D.P. Servedio, V. Loreto, A. Hotho, M. Grahl, and G. Stumme, "Network Properties of Folksonomies," *AI Comm.*, vol. 20, no. 4, 245-262, 2007.

[23] L.C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40, no. 1, pp. 35-41, 1977.

[24] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications.* Cambridge Univ. Press, 1994.

[25] P. Heyman and H. Garcia-Molina, "Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems," technical report, Stanford Univ., 2006.

[26] D.J. Watts and S.H. Strogatz, "Collective Dynamics of Small-World Networks," *Nature*, vol. 393, no. 6684, pp. 409-410, 1998.

[27] A.L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the Social Network of Scientific Collaborations," *Physica A*, vol. 311, nos. 3-4, pp. 590-614, 2002.

Hak-Lae Kim received the BS degree in business administration, the MS and PhD degrees in management information systems from Dankook University in 2004 and 2008, and the PhD degree in computer science from the National University of Ireland (NUI), Galway, in 2010. He is currently a senior engineer at the Media Solution Center at Samsung Electronics Co., LTD. He is also a director of the Web Science Workgroup in Korea. He was a researcher at the Digital Enterprise Research Institute at NUI, where his research efforts focused on creating social spaces based on tagging activities. He is the founder of the Social Semantic Cloud of Tags Project and ontology. In addition, he aims to provide semantic relations for tagging data and has implemented the int.ere.st website for sharing tagging data. He has served as an editor of the *Journal of Web Science*.

John G. Breslin is currently a lecturer in electronic engineering at the School of Engineering and Informatics, National University of Ireland (NUI), Galway. He is also an associate researcher and leader of the Social Software Unit at the Digital Enterprise Research Institute at NUI Galway, researching semantically enabled social networks and community portals. He is the founder of the Semantically-Interlinked Online Communities Initiative, which aims to semantically interlink online communities. He has received a number of awards for website design, including two Net Visionary Awards from the Irish Internet Association for the Irish community website www.boards.ie, which he cofounded in 2000. He is a member of the IEI, the IET, and the IEEE.

Han-Chieh Chao received the MS and PhD degrees in electrical engineering from Purdue University in 1989 and 1993, respectively. He is a jointly appointed professor in the Department of Electronic Engineering and the Institute of Computer Science and Information Engineering, National Ilan University, I-Lan, Taiwan. He also holds a joint professorship in the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan. He is serving as an IPv6 Steering Committee Member and the deputy director of the R&D division of the National Information and Communications Initiative Committee, Taiwan, and cochair of the Technical Area for IPv6 Forum, Taiwan. His research interests include high-speed networks, wireless networks, and IPv6-based networks and applications. He is a senior member of the IEEE and a fellow of both the IET and the BCS.

Lei Shu received the BSc degree in computer science from the South Central University for Nationalities, China, in 2002, the MSc degree in computer engineering from Kyung Hee University, Korea, in 2005, and the PhD degree from the Digital Enterprise Research Institute, National University of Ireland, Galway, in 2010. He is currently a specially assigned research fellow in the Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Japan. He has published more than 90 papers in related conferences, journals, and books. He received the MASS 2009 IEEE TCs Travel Grant, the Outstanding Leadership Award of the EUC 2009 as Publicity Chair, and the GlobeCom 2010 Best Paper Award. He has served as editor and guest editor of more than 20 international journals. He has served as cochair for more than 30 various international conferences and workshops, TPC member of more than 100 conferences, and as a reviewer of more than 50 journals. His research interests include semantic sensor networks, wireless sensor network, context aware, and sensor network middleware. He is a member of the IEEE.