Soc. Netw. Anal. Min.
DOI 10.1007/s13278-012-0054-z

ORIGINAL ARTICLE

Temporal dynamics of communities in social bookmarking systems

Said Kashoob · James Caverlee

Received: 1 July 2011 / Revised: 5 November 2011 / Accepted: 9 February 2012

Abstract Unprecedented growth in social bookmarking systems is making accessible the perspectives of millions of users on online content. This makes possible the ability to detect temporal group formation and their transient interests in online social systems. Here, we introduce a community evolution framework for studying and analyzing social bookmarking communities over time. We apply this framework to a large set of social bookmarking data, over 13 million unique postings, collected over a period of 15 weeks. We inspect the temporal dimension of social bookmarking and explore the dynamics of community formation, evolution, and dissolution. We show how our approach captures evolution, dynamics, and relationships among the discovered communities, which has important implications for designing future bookmarking systems, and anticipating user's future information needs.

Keywords Social bookmarking · Temporal · Model · Community

1 Introduction and background

Social bookmarking systems are prime examples of the proliferating and increasingly popular web-based social systems, in which user activity is recorded and traceable on a massive scale. These systems provide new opportunities to explore user perspectives and interests, and to examine the relationship between social interactions and traditional web content. Social annotations (or tags) are simple keywords or phrases that can be attached to an object as informal user-specific metadata. For example, on the Delicious social bookmarking system, a user could tag the web resource http://www.espn.com with tags like "sports", "my-favorites", and "scores". While seemingly a simple mechanism without a clear incentive structure for inducing users to bookmark web resources in the first place, much less share bookmarks with strangers, the Delicious bookmarking site alone has grown to over 5 million users who have bookmarked over 180 million unique URLs. In addition to Delicious, similar bookmarking sites have sprung up, including Flickr with more than 5 million images, CiteuLike with around 5 million scholarly articles, and StumbleUpon with over 9 million users. In isolation, a user's annotations can help organize a single user's bookmarks. But as these tags are shared and since many users independently assign tags to the same resource, there is a great opportunity to investigate the presence of latent structures, hidden communities, and the potential impact of these communities on information sharing and knowledge discovery.

Our hypothesis is that an underlying social collective intelligence is embedded in the uncoordinated actions of users on social bookmarking services, and that this social collective intelligence can be leveraged for enhanced web-based information discovery and knowledge sharing. Concretely, we posit the existence of underlying *implicit communities* in these social bookmarking systems that drive the social bookmarking process and can provide a foundation for community-based organization of web resources.

S. Kashoob (✉)
Department of Engineering, Salalah College of Technology,
P.O. Box 608, Salalah PC 211, Oman
e-mail: skashoob@gmail.com

J. Caverlee
Department of Computer Science and Engineering,
Texas A&M University, College Station, TX 77843, USA
e-mail: caverlee@cse.tamu.edu

Published online: 01 March 2012

The notion of community is fundamental to the social web—be it friendships on Facebook, groups of similarly-interested users who comment on YouTube videos, collections of Wikipedia contributors who specialize in certain topics, and so on. Social bookmarking systems aggregate what would appear to be the independent and uncoordinated tagging actions of a large and heterogeneous tagger population, meaning that it is not obvious that communities of users exist or are even detectable. In contrast to explicitly declared group memberships in social systems (which are often stale and fail to reflect the vibrant activity of the system), these implicit communities are necessarily *hidden* from us, but could provide a window into the real-time and dynamic self-organization of these systems.

Toward uncovering and leveraging community on the social web, this work addresses the following research questions:

- Can we capture community structure in temporal social bookmarking systems' data? And how can we model and effectively extract implicit communities from the large, heterogenous, and uncoordinated actions of millions of users?
- Can we observe and capture community evolution in social bookmarking data? Do different communities evolve differently? Are there any underlying features that differentiate these differently evolving communities?
- When communities evolve, how does that reflect in their constituent elements, i.e., users and tags? Are there dynamic communities and other static (stable) ones? Can we draw any conclusions about their underlying features?
- Can we measure cross-community relationships? What are examples of related communities and why are they related?

To address these research questions, this paper makes two unique contributions:

- The first contribution of this paper is a community evolution framework for studying and analyzing social bookmarking communities over time. This models posits that the observed tagging information in a social bookmarking system is the product of an underlying community structure, in which users belong to implicit groups of interest.
- Second, we explore the temporal dimension of social bookmarking and explore the dynamics of community formation, evolution, and dissolution. We show how this approach captures evolution, dynamics, and relationships among the discovered communities, which has important implications for designing future bookmarking systems, anticipating user's future information needs, and so on.

By uncovering these communities and analyzing them over time, we can enable new avenues of transformative research, including the study of how social knowledge networks are self-organized, a deeper understanding and appreciation of the factors impacting collective intelligence, and the creation of new information access algorithms for leveraging these communities.

Unlike top-down hierarchical information architectures that are often brittle and quickly outdated, this social web promises a flexible bottom-up (emergent) approach to organizing and managing information centered around people and their social connections to other people and information resources. This people-centric approach to information management can lead to large-scale user-driven growth in the size and content in the system, bottom-up discovery of citizen-experts with specialized knowledge, serendipitous discovery of new resources beyond the scope and intent of the original system designers, and so on. Indeed, this promise is attracting significant strategic investment and support by public health agencies, emergency responders, federal, state and local governments, major companies, and universities, among many others, and has already encouraged new advances in web-based social information sharing (Twitter 2010), online commerce (Litner and Grechenig 2008), governance (Willard 2009), citizen journalism (Digital journal 2010), and education (Bercovitz et al. 2009).

In the following section we survey related work on social bookmarking systems, on web information organization, on topic modeling, and on community discovery, before turning our sights towards modeling and analyzing the temporal dynamics of communities in social bookmarking systems.

2 Related work

The investigation of social bookmarking and its role in modern computing and information systems has been the topic of many research works over the past years. Previous work has addressed various aspects of social bookmarking systems, see (Macgregor 2006) for a discussion of the prospects, limitations and value of social bookmarking data for information and knowledge organization. In the following, we review research efforts that have dealt with exploiting tagged resources to create semantics and ontologies, model the dynamics of bookmarking systems, and enhance search and retrieval.

Structured data is created by professional curators that have formal education and training. Some examples of structured data systems include the Dewey Decimal System, and the Library of Congress Classification System. These systems consist of taxonomies, ontologies and controlled vocabularies that permit high-quality cataloging, categorization and classification of information and resources. However, they are considered to be costly, static, and unscalable.

On the other hand, social bookmarking systems have a very low barrier to entry, and minimal expertise and education requirements as can be seen in Delicious, and Flickr among others. These systems employ free-style tagging with no vocabulary restrictions, no coordination among taggers, and no experts. These systems are inexpensive, dynamic, and scalable.

Now with the emergence of social bookmarking systems, some research works have looked into the effectiveness of social bookmarking systems in producing useful metadata (Marlow et al. 2006), semantics (Markines et al. 2009, Wu et al. 2006), and their usefulness in web classification (Noll 2008) versus expert classification. Methods for augmenting structured data with free-style user contributed data (Mika 2005) aim to combine the advantages of both worlds (Christiaens 2006) and allow for the creation of emergent knowledge, "knowledge not contained in any one source" (Gruber 2008). However, social bookmarking systems introduce serious issues such as vocabulary growth and reuse (Farooq et al. 2007), quality selection (Sen et al. 2009), spam (Liu et al. 2009; Markines et al. 2009; Neubauer et al. 2009) and relevance to content and query (Carman et al. 2009).

In one of the earliest studies of social bookmarking, Golder and Huberman (Golder 2005) found a number of clear structural patterns in Delicious, including the stabilization of tags over time, even in the presence of large and heterogeneous user communities. This stabilization (which might be counterintuitive, especially in contrast to the tightly controlled metadata produced by domain experts) suggests a shared knowledge in tagging communities. These results are echoed by Halpin et al. (2007), who found a power-law distribution for Delicious tags applied to web pages—meaning that in the aggregate, distinct users independently described a page using a common tagging vocabulary. Similar results can be found elsewhere, including Cattuto et al. (2006, 2007), Li et al. (2008), and Veres (2006).

The past few years have seen an increased interest in modeling social annotations. Several works that adapt topic-modeling-based approaches for modeling social annotations include mapping tags, users, and content to a single underlying conceptual space (Wu et al. 2006), mapping combined content and tags to an underlying topic space (Zhou et al. 2008), mapping content, tags and additional link information to multiple underlying topic spaces (Ramage et al. 2009). Additionally, in Plangrasopchok (2007) and (2008), the authors assume hidden structures of interests and topics that generate tags for resources. They then are able to discover related resources based on their relevance (distributions) to interests and topics. A topic in our context is the unifying or prominent theme conveyed by a collection of terms. Specifically, our modeling approach results in grouping of terms. The identification of the theme or topic for a group of terms is done by inspection of the top 20 terms in the group. Works such as Yin et al. (2011), Zhang et al. (2009) have also studied the temporal aspect of social bookmarking systems. These results motivate our interest in uncovering hidden communities that could help us understand social bookmarking systems better.

Tagging's most basic function is to organize resources as a step towards improved browsing and search (Begelman et al. 2006; Li et al. 2007). Once tagging activities are shared they result in an impressive source of knowledge that can be used in numerous ways. For example, it can be used to complement link-based search methods (Kolay 2009; Yanbe et al. 2007; Heymann et al. 2008), to measure resource popularity (Bao et al. 2007), to build language models for retrieval (Zhou et al. 2008), and to detect trends (Wetzker et al. 2008). Social bookmarking data has also shown potential for improved personalization (Carman et al. 2008; Guan et al. 2009; Bateman et al. 2009; Xu et al. 2008; Hamouda 2011), query expansion (Wang 2008), and recommender systems (Vig et al. 2009; Song et al. 2008; Esslimani et al. 2011).

3 Modeling community

Toward modeling and analyzing the temporal dynamics of communities in social bookmarking systems, we propose in this section a community-based tagging model. The term community as is commonly defined brings to mind unifying notions of similarity, coordination, and purpose. Identifying coherent communities in social bookmarking systems is challenging, however, for a number of reasons:

- *User Heterogeneity* In contrast to "controlled vocabularies" applied by domain experts to organize web resources (e.g., like the Open Directory Project or Yahoo's web directory), social bookmarking systems rely on a heterogenous bookmarking population that may apply tags that vary greatly in purpose and quality.
- *Lack of Coordination* Coupled with the user heterogeneity is the lack of coordination in social bookmarking systems. Since bookmarks are typically made in isolation and without explicit coordination with other users (except perhaps implicitly, by viewing the prior tags applied by users on a resource), it is not obvious if the aggregate bookmarks applied by millions of different users should provide any overarching "meaning" to web resources.

- *Sparsity* Elements making social bookmarking systems vary wildly in their behavior, and contribution to the systems. Therefore their importance and impact also vary. For example, a small percentage of the tag vocabulary is extremely popular while the majority of tags experience minimal exposure. This presents a challenge as well as an opportunity.
- *Granularity* Due to the intense bookmarking activity a decision must be made to determine on which time scale to observe and draw conclusions about the resulting bookmarking behavior. Viewing the data in hourly, daily, or weekly intervals is expected to reveal different aspects and criteria of the system and may also lead to alternate conclusions.

With these challenges in mind, we next present a community-based tagging model designed to discover community structures that underlie social bookmarking systems.

3.1 The segmented community-based tagging model

Our approach to handle social bookmarking data is based on a text-based topic modeling approach, where we consider the document unit to be the collection of all tags and users applied to a particular resource (Kashoop et al. 2009). We call this collection of tags and users applied to a resource its *social tagging document*.

Definition: Social tagging document For a resourceo $r \in \mathcal{U}$, we refer to the collection of tags assigned to the resource as the resource's social tagging document S, where S is modeled by the set of users and the tags they assigned to the resource: $S = \{\langle \text{user}_j, \text{tag}_j \rangle\}$.

In this approach, we posit the existence of L communities that are implicit in the universe of discourse $\mathcal{U}$, where each community is composed of users and tags that are representative of the community's perspective. Since community membership is not fixed, we model membership as a probability distribution, where each user and tag has some probability of belonging to any one community.

Definition: Social tagging community A social tagging community c is composed of (1) a probability distribution over users in U such that $\sum_{u \in U} p(u|c) = 1$, where $p(u|c)$ indicates membership strength for each user u in community c; and (2) a probability distribution over tags in the vocabulary T such that $\sum_{t \in T} p(t|c) = 1$, where $p(t|c)$ indicates membership strength for each tag t in community c.

To capture the temporal nature of social bookmarking services we view the basic unit to be the collection of tags and users applied to a particular resource but limited to those occurring in a specific time period (e.g., by hour, day, week, etc.). Now all tagging activity that occurs in a given time interval is assumed to be drawn from global latent structures of communities of users and their associated tag vocabulary. In addition, these global structures can now change from one time interval to the next. We call this approach, the Segmented Community-based Tagging (SCTAG) Model.

Our intuition is that reasonably long time intervals (for example a week) will contain a mixture of tagged resources that can potentially reveal the current global interests as well as a classification of the different taggers and the tags they use. Additionally, observing the system over consecutive time intervals will reveal the evolution of interests, user groups, as well as tag groups.

The SCTAG model partitions the annotations applied to a single resource into K segments based on the time during which the annotation was applied. For example, Fig. 1 shows a sample resource (the CNN page, http://www.cnn.com) being split into June and July 2010 segments. This processing step is performed for all resources. It results in collections of social tagging documents ordered by time. Each collection is then fed to the SCTAG model, shown in Fig. 2, along with the latent structures learned from the preceding time segment. This way, we allow the latent structures to evolve and also capture the changes from one time segment to the next.

The generative process for the SCTAG model in Fig. 2 works as follows:

– for time segment $k = 2, \ldots, K$

1. for each community $c = 1, \ldots, L$
   - Select U dimensional $\tau_c^k \sim$ Dirichlet $(f(\tau_c^{k-1}, \alpha))$
   - select V dimensional $\phi_c^k \sim$ Dirichlet $(f(\phi_c^{k-1}, \gamma))$
2. for each object $\mathbf{S}_i, i = 1, \ldots, D$
   - Select L dimensional $\theta_i^k \sim$ Dirichlet $(f(\theta_i^{k-1}, \beta))$
   - For each position $\mathbf{S}_{i,j}, j = 1, \ldots, N_i$
     - Select a community $\mathbf{c}_{i,j} \sim$ multinomial (θ_i^k)
     - Select a user $\mathbf{S}_{i,j}^u \sim$ multinomial $(\tau_{\mathbf{c}_{i,j}}^k)$
     - Select a tag $\mathbf{S}_{i,j}^t \sim$ multinomial $(\phi_{\mathbf{c}_{i,j}}^k)$

The first time segment, $k = 1$, has no prior latent structure. Each consecutive time segment augments the prior structure with current observation points allowing for evolutionary behaviors to be observed. Next we use Gibbs sampling to estimate the latent structures of the SCTAG generative process.

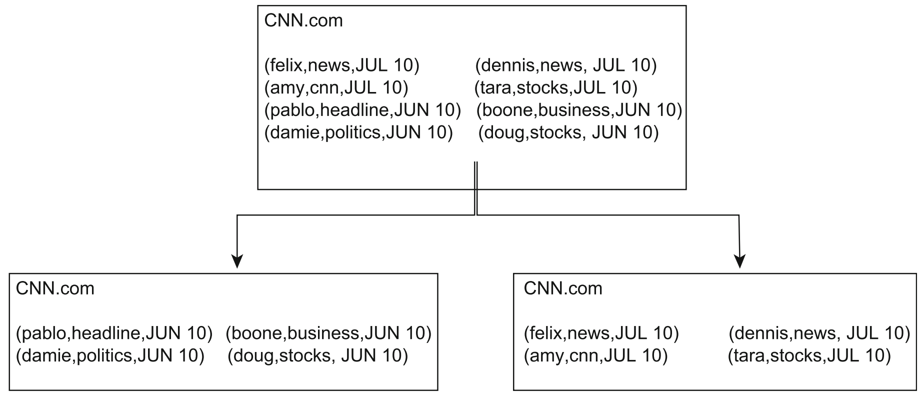


Fig. 1 Splitting annotations to time segments (month)

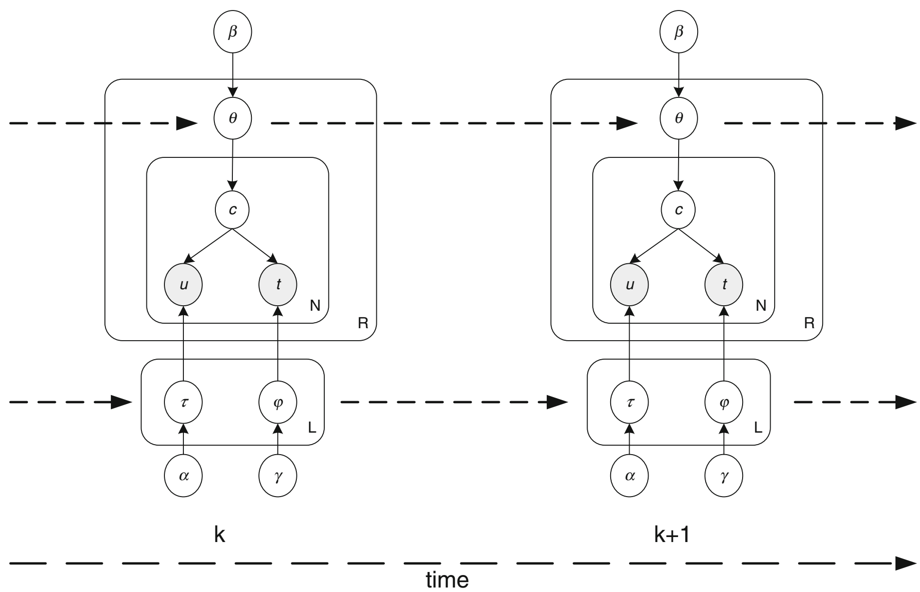


Fig. 2 Segmented community-based tagging model (SCTAG)

3.2 Parameter estimation with Gibbs sampling

The generative process shown above describes how temporal social tagging documents are created. Our goal here is to take collections of social tagging documents that we assume are the product of such a generative process and recover the underlying hidden structures of communities, their users and their tags. More specifically, we learn model parameters τ, θ, and ϕ (the distributions over communities, users, and tags, respectively) for each time segment.

Let $\mathbf{S}$ and $\mathbf{c}$ be vectors of length $\sum_i^R N_i$ representing $\langle user, tag \rangle$ pair, and community assignments, respectively, for the collection in a single time segment, k. Also let u and t be user and tag variables. Suppose the latent structures learned in the previous time segment $k-1$ is known, $\mathcal{M}_{k-1}$. Following the approach used in Heinric (2005), we derive the following Gibbs sampler's update equation for assigning communities to user, tag pairs in time segment k:

$$p(\mathbf{c}_i = l | \mathbf{c}_{\neg i}, S^t, S^u, \mathcal{M}_{k-1}) \propto \frac{n^u_{l,\neg i} + \tilde{n}^u_l + \alpha_u}{\sum_{u=1}^{U} n^u_{l,\neg i} + \tilde{n}^u_l + \alpha_u} \times \frac{n^t_{l,\neg i} + \tilde{n}^t_l + \gamma_t}{\sum_{t=1}^{V} n^t_{l,\neg i} + \tilde{n}^t_l + \gamma_t} \times \frac{n^l_{S,\neg i} + \beta_l}{\left(\sum_{l=1}^{L} n^l_S + \beta_l\right) - 1} \quad (1)$$

where $n^{(\cdot)}_{(\cdot),\neg i}$ is a count excluding the current position assignments of $\mathbf{c}_i$ (e.g., $n^t_{l,\neg i}$ is the count of tag t generated by the lth community excluding the current position). $\tilde{n}^{(\cdot)}_{(\cdot),\neg i}$ are prior counts from the preceding time segment. The prior counts are set to zero for the first time segment as well as for new users and tags that appear for the first time in later segments.

3.3 Segmented community-based tagging framework

To summarize, the segmented community-based tagging model is the key component of the overall framework for uncovering communities and in studying their dynamics. The overall community-based tagging framework is illustrated in Fig. 3 and includes the collection of raw social tagging data, the pre-processing and partitioning of the raw data into appropriate segments, the application of the SCTAG model for discovering communities, and then the subsequent analyses (e.g., community dynamics) and applications that can build on these discovered communities. In the following section, we explore several scenarios in which the framework can reveal community evolution and community dynamics.

4 Experiments

Now that we have introduced our model, we present how it can be applied to real data collected from the Delicious social bookmarking system. First, we introduce and describe the dataset. Second, we demonstrate how our model is able discover communities, capture evolution, and observe dynamics in this dataset.

4.1 Temporal social bookmarking data and features

Our dataset was collected from Delicious' recent feed over a period of 15 weeks (November 11th, 2009 to March 1st, 2010). It consists of 13,405,322 unique postings over 3,778,338 unique URLs, performed by 641,021 users using 1,504,147 unique tags.

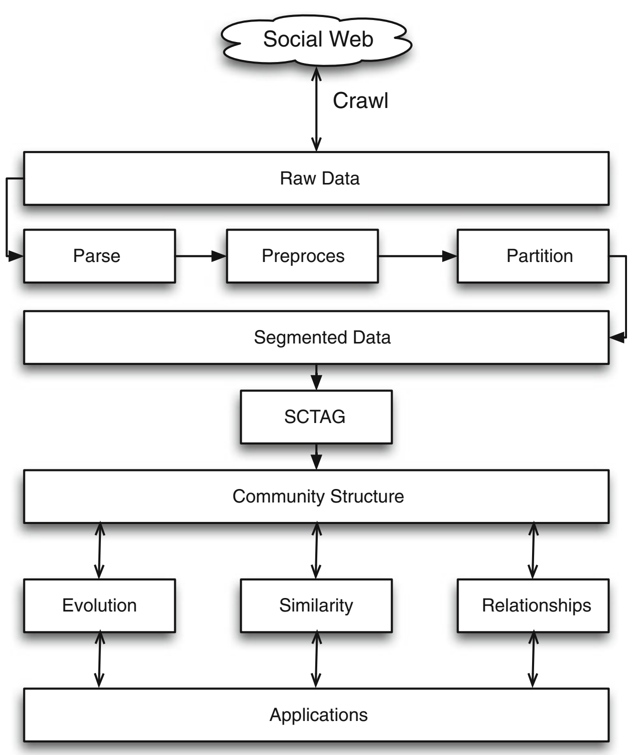


Fig. 3 Segmented community-based tagging framework

Some interesting observations about the dataset are the number of tagged resources over time, the number of taggers over time, and the number of tags used. In Fig. 4a we show resource activity over the observed period. The *x*-axes show the time interval in hours and the *y*-axes show the number of resources. On average, there are 4,726 unique resources tagged every hour. Figure 4b shows the number of unique active taggers per hour. On average, there are 3,353 unique taggers every hour. Figure 4c shows that on average, taggers use 5,644 unique tags every hour.

Another interesting observation concerns the new taggers, tags, and resources seen for the first time. In Fig. 5, we present the number of each of these three types which have not been seen before. Notice that on average there are 1,000 resources which have not been observed in any previous time interval. And there on average 100 taggers and 500 tags per hour which have not been observed in any previous time interval.

Generally, we observe that the overall Delicious social bookmarking community gets around 110K resources, 80K taggers, and 130K tags on a daily basis. We also can conclude that this community maintains growth in the number of resources, users and tags; with resources having the highest growth rate followed by tags then by taggers. The drop in activity around the 1,000th hour is due to Christmas time and the New Year. Notice that a drop in the number of new taggers arriving into the system occurs around the 700th hour. We take this as an indicator that most taggers are active at least once a month. A similar drop but less pronounced occurs with new tags around the same time. A much slighter drop is seen in the case of resources. A one month duration is sufficient to capture the majority of taggers and tags but not of resources.

Next, we turn to examine the re-occurrence of resources, taggers and tags over time during the observed period. For each resource, tagger, and tag, we want to observe the corresponding number of hours in which they were active at least once.

In Fig. 6a, we plot the count of resources and the corresponding time intervals in which they appeared at least once. It shows us that there are more than 1M resources that were observed in only one time interval each (1 h), while there are 10 resources that were observed in at least 300 time intervals each. These features can be of interest for identifying popular resources or spam.

Similarly, for taggers we plot in Fig. 6b the number of taggers and the corresponding number of hours in which they were active. We see that there are more than 100K taggers that were active only in one time interval. On the other hand, 10 taggers were seen in between 200 to 300 time intervals each. These features can be of interest for

Fig. 4 Tagging activity over a 15-week period (Nov. 11, 2009 to Mar. 1, 2010). **a** Unique Resources tagged per hour. **b** Unique active Taggers per hour. **c** Unique applied Tags per hour

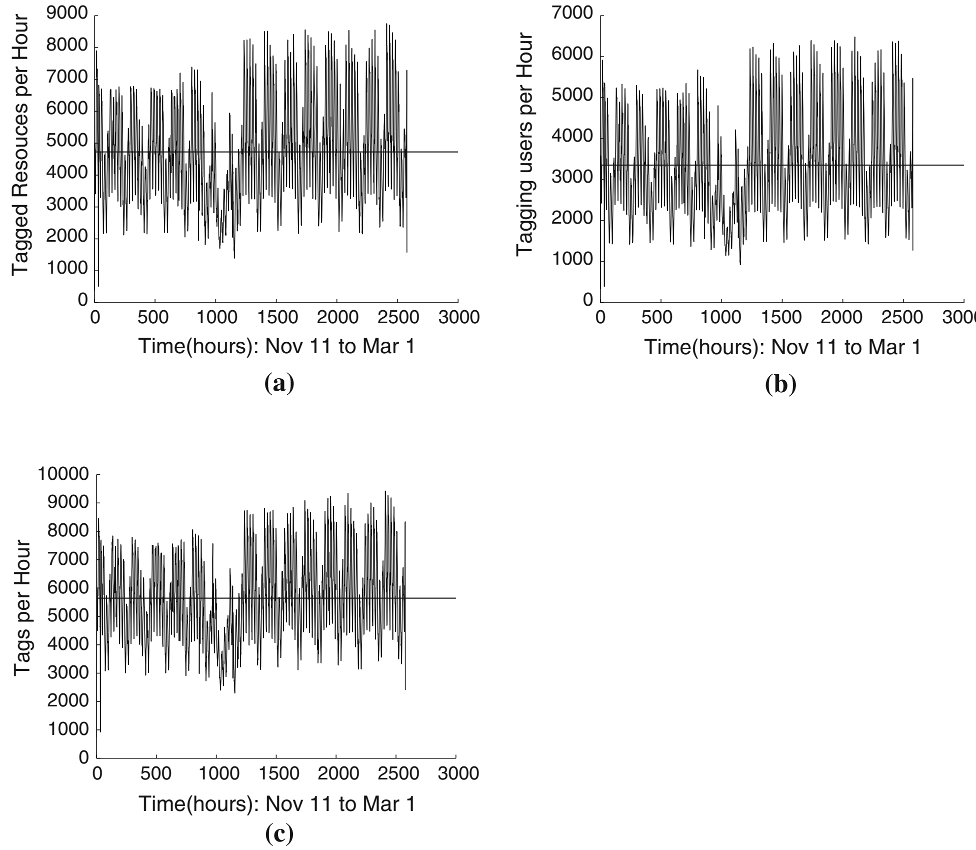
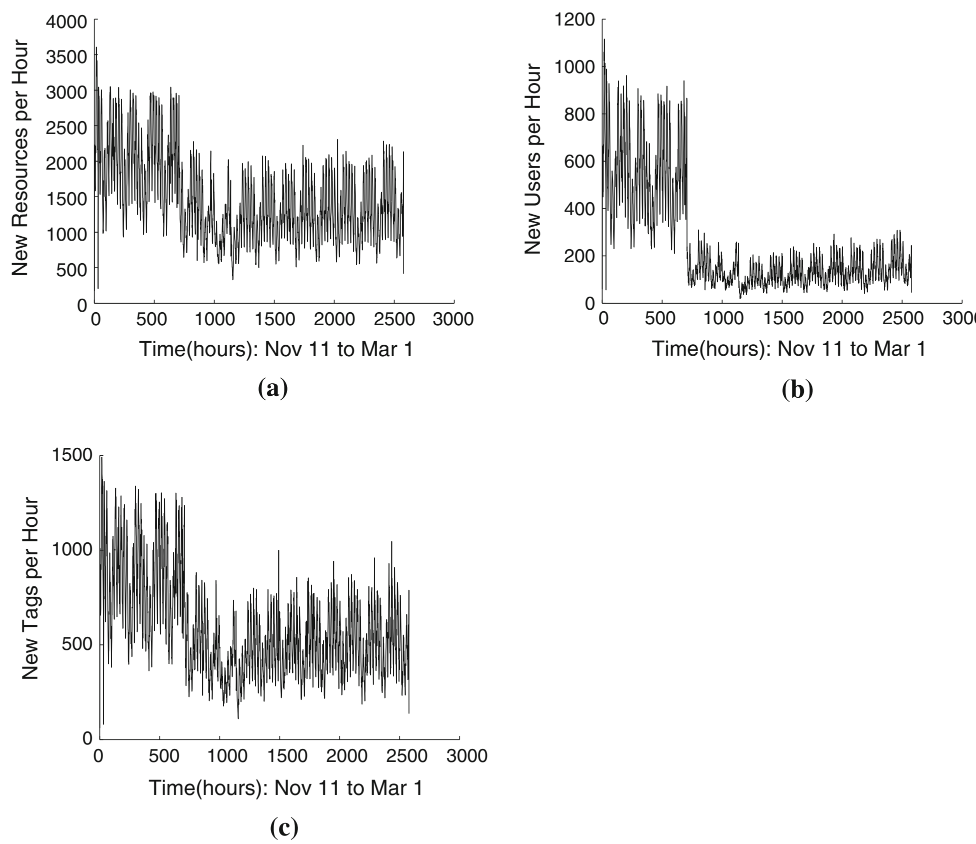


Fig. 5 Tagging activity over time: newcomers observed for the first time. **a** New unique Resources tagged per hour. **b** New unique active Taggers per hour. **c** New unique applied Tags per hour



identifying prolific taggers, spammers, trend makers, as well as bots.

Finally, Fig. 6c shows the number of tags and the count of hours in which they were used at least once. The figure indicates that there are about 1M tags that were seen in only one time interval each, while about 10 tags appeared in 1,000 time intervals each. These features can be useful to determine trends, classification, and spam terms.

We can see that resources, taggers, and tags re-occurrence follow the common power-law distribution where a few elements are very active and the majority have very low re-occurrence. Notice that the resources and the tags re-occurrence form a straight line on the log–log scale while the taggers re-occurrence forms a curved line (slower decline in the number of taggers for increasing time interval counts) which indicates that a large number of taggers are active unlike the case for active tags and active resources.

To further illustrate the characteristics of social bookmarking systems, we examine the co-occurrences among resources, taggers, and tags. We start by looking at how resources interact with taggers and tags. Figure 7a plots the number of resources and the corresponding number of taggers and tags that they co-occurred with. Notice in the top subfigure that there are more than 1M resources that are tagged by just one tagger each, while there are around 100 resources that are tagged by about 200 taggers each. In the bottom subfigure we see that there are about 1M resource with only one tag each and about about 100 resources with about 150 tags each. Once again we observe the phenomenon that a few resources are popular, attracting many users and tags, while the majority get minimal exposure to users and very few tags.

Now we look at how taggers interact with resources and tags. In Fig. 7b we show similar observation for taggers. In the top subfigure, we observe that there are about 1M taggers that tag just one resource each, while there are 10 taggers that tag more than 1,000 resources each. In the bottom subfigure, we see that there are about 100K taggers that use just one tag each, and there are 10 taggers that use about 1,000 tags each.

Finally, we look at how tags interact with taggers and resources. Figure 7c presents similar counts for tags. In the top subfigure we observe that there are more than 100K tags that are used by just on tagger each, while there are 10 tags that are used by about 1,000 taggers each. The bottom subfigure shows that there are about 1M tags that appear on only one resource each, while there are 100 tags that appear on about 1,000 resources each.

Although the Delicious feed could be throttled, filtered or delayed internally the figures above suggest a minimum activity rate for all three types (resource, user, tag) as well a minimum growth estimate.

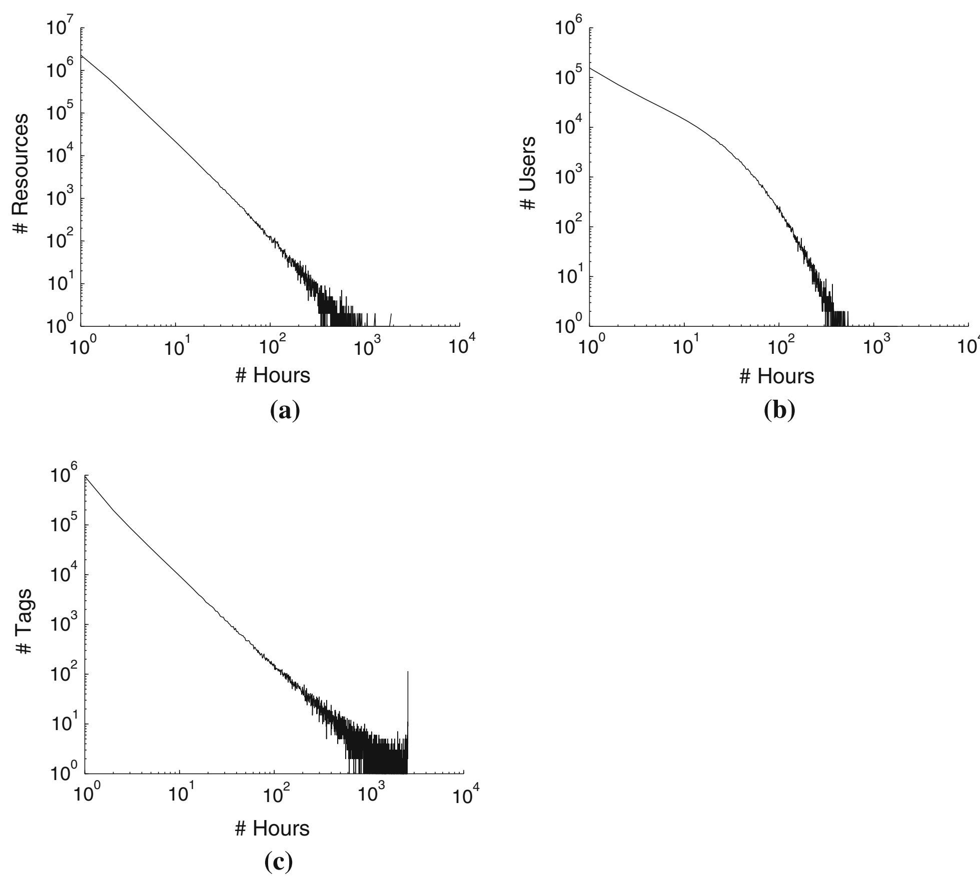
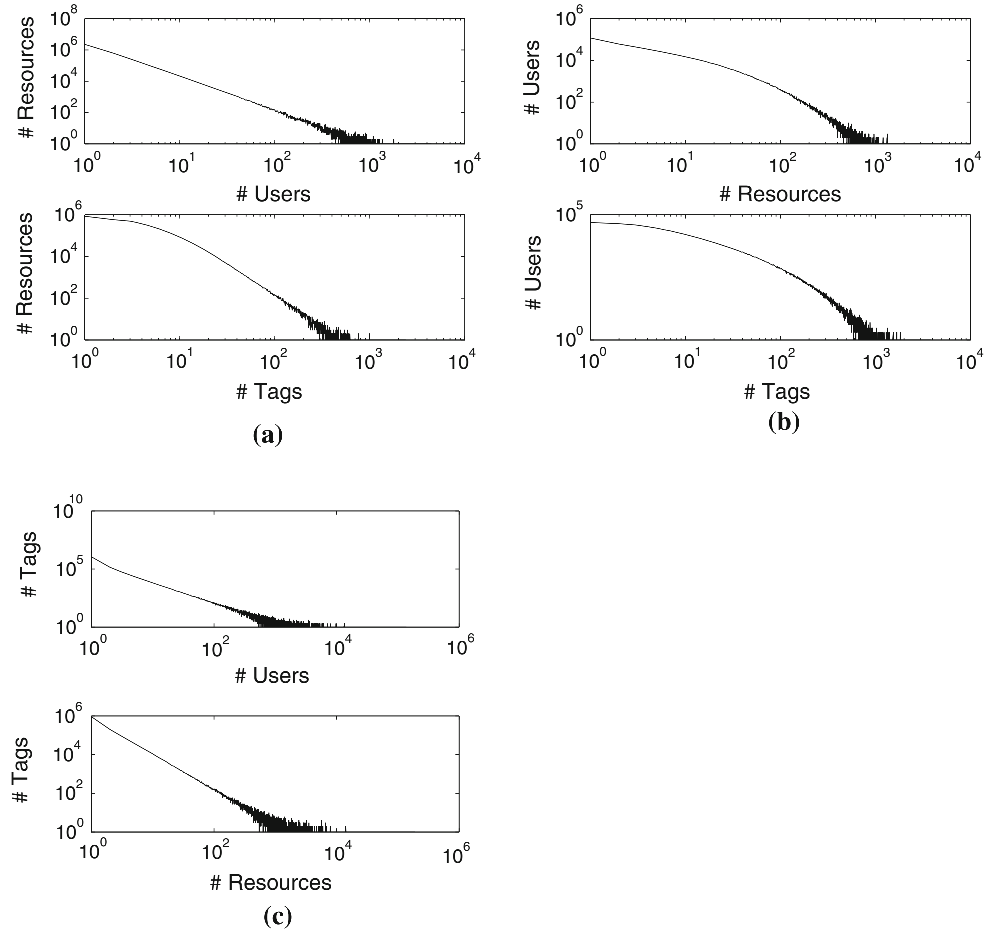


Fig. 6 Tagging activity (re-occurrence): the corresponding number of hours in which they were observed at least once. **a** Resources, **b** taggers, **c** tags

Fig. 7 Interactions in Delicious social bookmarking systems. **a** Resources versus (users, tags). **b** Users versus (resources, tags). **c** Tags versus (users, resources)



4.2 Community discovery

In our experiments, we segment the dataset into weekly time segments as was discussed in previous sections. This results in 15 sub-collections each with 90,000 social tagging documents on average. We run the SCTAG model on the first sub-collection to determine the number of potential communities. We set the model hyperparameters to ($\alpha = 0.9$, $\beta = 0.1$, $\gamma = 0.01$) and vary the number of communities from 20 to 160.

Our goal here is to first confirm that the model works and to also determine an appropriate number of communities. Since the focus is not on optimizing the number of communities discovered, we elect to fix the community parameter to 100 since it resulted in the most cohesive top-20 tags with the least overlap across discovered communities. A sample of the discovered communities top tags are shown in Table 1.

To illustrate the benefit of using this model, we compare these discovered communities to the top frequency tags observed during 1-h intervals in our dataset. Table 2 shows a sample of most frequently used tags per hour. Notice the overwhelming presence of "web design" and "programming". Contrary to the impression one gets of a lack of community structure based on the most frequent tags, a topic modeling-based approach reveals some interesting communities.

After the initial step of determining the number of communities, the SCTAG model is run on the remaining time segments where a preceding time segment result serves as a prior for the following time segment.

4.3 Community evolution

Previously we have used user, tag, and resource interactions to capture communities of users and their tag vocabulary. Now, we observe these communities over time and try to capture how they change. We define change based on two aspects of the community: (1) the users forming the community and (2) the tags representing the community perspective.

Let us now inspect how community interests change over time. An example of community interest evolution over time segments is shown in Fig. 8. We present two sample communities about "Google tools" and "Health" along with their top-10 tags over the 15-week period.

Table 1 A sample top tags in communities

	Top tags
Comm 0	Video movi film stream youtub media entertain cinema televis onlin
	Free documentari subtitl towatch anim multimedia clip download review
	Watch live show list filmmak recommend vimeo seri flv…
Comm 1	Learn educ elearn web2 train teach research technolog onlin
	Moodl resourc knowledg open pedagogi eportfolio opensourc virtual
	Blog commun ict instruct secondlif lm assess virtualworld dog theori…
Comm 2	Map api googl geographi googlemap gp gi mashup geo locat
	Data local geoloc visual googleearth geocod earth cartographi
	World develop refer foursquar geologi geotag tool mapa webservic…
Comm 3	Socialmedia facebook market social media socialnetwork
	Trend brand web2 busi strategi advertis roi mashabl casestudi 2010
	Research measur socialmedia twitter digit internet polici…
Comm 4	Fashion shop blog cloth magazin design style inspir shirt
	Vintag cultur beauti trend tshirt shoe moda men accessori store
	Retro art bag hipster lifestyl jewelri cool…

Table 2 Top frequency tags per hour

Hour								
1	2	…	1001	1002	…	2001	2002	…
tools	design	…	design	design	…	tools	design	…
design	tools	…	tools	blog	…	webdesign	webdesign	…
google	programming	…	blog	webdesign	…	tools	tools	…
software	webdesign	…	webdesign	inspiration	…	blog	inspiration	…
programming	inspiration	…	programming	software	…	video	blog	…
reference	blog	…	software	video	…	inspiration	web	…
inspiration	software	…	video	programming	…	web	video	…
blog	google	…	inspiration	art	…	programming	resources	…
webdesign	web	…	free	tutorial	…	free	programming	…
web2.0	reference	…	tutorial	reference	…	css	reference	…
science	tutorial	…	reference	web	…	software	software	…
tutorial	free	…	art	development	…	reference	free	…
resources	development	…	development	photography	…	tutorial	css	…
web	art	…	music	music	…	jquery	development	…
research	video	…	howto	howto	…	web2.0	art	…
opensource	education	…	web2.0	web2.0	…	resources	javascript	…

Notice how the "Google tools community" is initially concerned with "Google wave" applications and "collaboration" in weeks 1–9 (November–December 2009, during which Google Wave was released) then switched to "Google china", "Google buzz", and "privacy" in weeks 11–15 (January–February 2010, during which the Google China hack scandal and Google Buzz privacy issues occurred). On the other hand, the "health community" can be seen as more stable with interests continually represented by tags such as "food", "exercise", "health care", and "medicine".

The SCTAG model gives us a per community distribution over users as well as a distribution over tags. Focusing on just one community over a sequence of time segments, we can measure changes on both distributions over users and over tags using a measure like the Jensen–Shannon (JS) distance. The JS distance compares two probability distributions p and q over an event space X:

$$\begin{aligned} JS(p,q) &= 0.5[KL(p,m) + KL(q,m)] \\ KL(p,q) &= \sum_{x \in X} p(x) \cdot \log(p(x)/q(x)) \end{aligned} \tag{2}$$

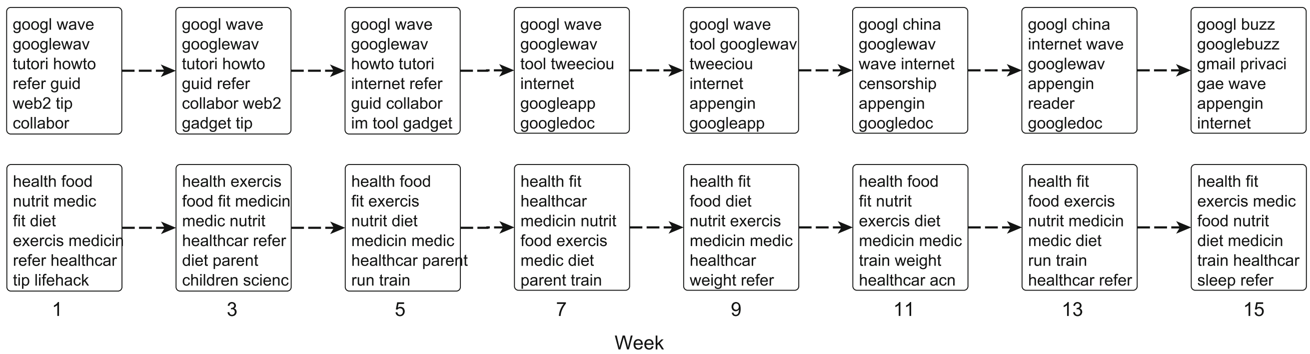


Fig. 8 A community's topic evolution over time

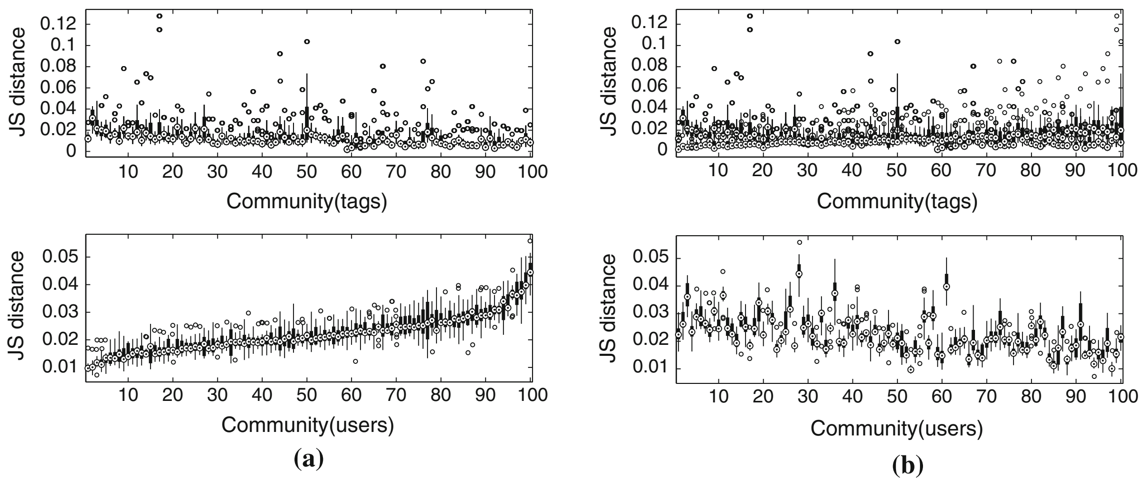


Fig. 9 JS distance per community over time. **a** Sorted by users, **b** sorted by tags

where $m = 0.5\ (p + q)$ and $KL(p, q)$ is the Kullback–leibler divergence. The event space X for our purposes can be either the tag vocabulary, the user vocabulary or a combination of both.

We present the results of the JS distance per community over time in both user and tag spaces in Fig. 9a. The results are sorted by the distance over the user space. Notice that communities vary in their distance over the user space as shown in the bottom subfigure. This indicates evolutionary dynamics of user membership in the communities. This however cannot be said about the tag space as is shown in the top subfigure. Notice that all communities have relatively similar distances in the tag space despite their distances in the user space, meaning that communities with high user churn and those with low user churn all have relatively stable interests.

The same trend can be observed when the results are sorted by the distance in the tag space, as is shown in Fig. 9b. Based on this we can conclude that communities tend to evolve more on their user space than on their tag space. That is, users tend to change their community membership over time more often than do tags. This is an expected result as user membership in communities represents the user transient interest while tag membership in communities represents a thematic classification of the tags.

4.4 Community dynamics

In the above sections we have shown the ability of our model to detect communities and to observe community evolution over time. In the following, we address these remaining research questions:

- When communities evolve, how does that reflect in their constituent elements, i.e., users and tags? Are there dynamic communities and other static (stable) ones? Can we draw any conclusions about their underlying features?
- Can we measure cross-community relationships? What are examples of related communities and why are they related?

To determine inter-community relationships we can apply the same Jensen–Shannon distance to measure the overlap over users and tags between two communities. This overlap can also be measured as a Jaccard or cosine distance. However, these methods are too expensive as they require pairwise comparisons over the community space. Alternatively, we develop a simpler method for comparing communities focusing on the user and the tag spaces and their community assignment. This method is similar in spirit to the bit matrix approach taken in Asur et al. (2007).

Our method takes the results of the SCTAG model and tracks each user and tag, and their top community assignment in each time segment. From this step, we can determine the paths that users or tags take over time in relation to communities. For example, in Fig. 10, we have three communities spread over three time segments. An example path is ($x0$, $x1$, $y2$), representing users or tags assigned to community x at times 0 and 1; and to community y at time 2. By simply counting the number of users or tags transitions between communities over time segments, we can capture community evolution, similar communities, communities that evolve together and others that deviate over time.

Formally, let there be a graph $G = (V, E)$ where the set of nodes V represents the community space spanning all time segments, and the set of edges E represents users' community assignment transition from one time segment to the next. An edge $e = (x_0, y_1)$ indicates that a user with community assignment x at time 0 got assigned to community y at time 1. The weight of edge e, $w(e)$, represents the number of users or proportion of users that had made the same community assignment transition. For example, in Fig. 10, 13 users assigned to community z at time 0 were assigned to community z at time 1, while 3 of these users got assigned to community y at time 2.

Using this graph setup we ask questions of community relationships and stability. Are there stable communities? Do stable communities have core members? And what is the size of this stable core? We consider stability to be influenced by the following factors: the community core, the community parters, and the community joiners.

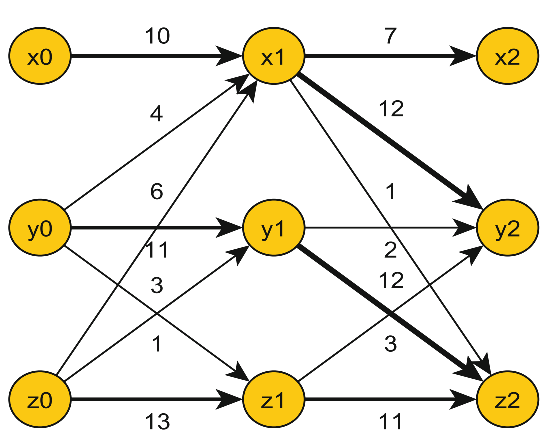


Fig. 10 Example transitions across communities over time

Definition: Community core A social tagging community c has a core consisting of users or tags that are successively assigned to community c in two consecutive time segments, $CC(c) = \{u \in \cap_k^{k+1} U_k \text{ s.t. } c = arg\ max_{c=1,\dots,L}\ \tau_{c,u}^k\}$. For tags, $CC(c) = \{t \in \cap_k^{k+1} T_k \text{ s.t. } c = arg\ max_{c=1,\dots,L}\ \phi_{c,t}^k\}$.

Definition: Community Parters A social tagging community c parters are a set of users or tags that are assigned to community c at time segment $k - 1$ but not at time segment k:

$PA(c) = \{u \in \cap_k^{k+1} U_k \text{ s.t. } c = arg\ max_{c=1,\dots,L}\ \tau_{c,u}^{k-1}$ and $c \neq arg\ max_{c=1,\dots,L}\ \tau_{c,u}^k\}$.

And for tags:

$PA(c) = \{t \in \cap_k^{k+1} T_k \text{ s.t. } c = arg\ max_{c=1,\dots,L}\ \phi_{c,t}^{k-1}$ and $c \neq arg\ max_{c=1,\dots,L}\ \phi_{c,t}^k\}$.

Definition: Community joiners A social tagging community c joiners are a set of users or tags that are not assigned to community c at time segment $k - 1$ but are in time segment k:

$JO(c) = \{u \in \cap_k^{k+1} U_k \text{ s.t. } c \neq arg\ max_{c=1,\dots,L} \tau_{c,u}^{k-1}$ and $c = arg\ max_{c=1,\dots,L}\ \tau_{c,u}^k\}$.

And for tags:

$JO(c) = \{t \in \cap_k^{k+1} T_k \text{ s.t. } c \neq arg\ max_{c=1,\dots,L} \phi_{c,t}^{k-1}$ and $c = arg\ max_{c=1,\dots,L}\ \phi_{c,t}^k\}$.

For example, consider the communities x, y, z shown in Fig. 10 with 3 time segments. The core for community x is at most 7, its parters are 13 and its joiners are 10. Respectively for communities y and z, their cores are at most 2 and 11, their parters are 19 and 6, and their joiners are 6 and 14.

4.4.1 Users and tags in communities

To inspect how community evolution reflects in its constituent users and tags, we start by showing how the user space is assigned to the discovered communities. Figure 11 shows the average number of users assigned to each community over the observed time period. The average number of users assigned to a community fall in the range [12, 300].

Transforming the SCTAG discovered communities into the community dynamics graph introduced earlier allows us to observe the community core, community parters, and community joiners. We present the results for users (core, parters, joiners) per community in Fig. 12. The figure shows the mean proportion of users and its variance for all

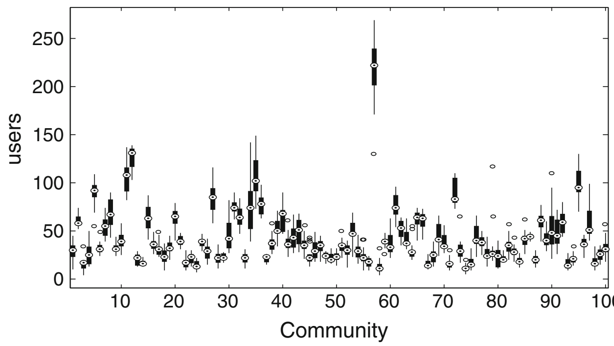


Fig. 11 Average number of users assigned to community

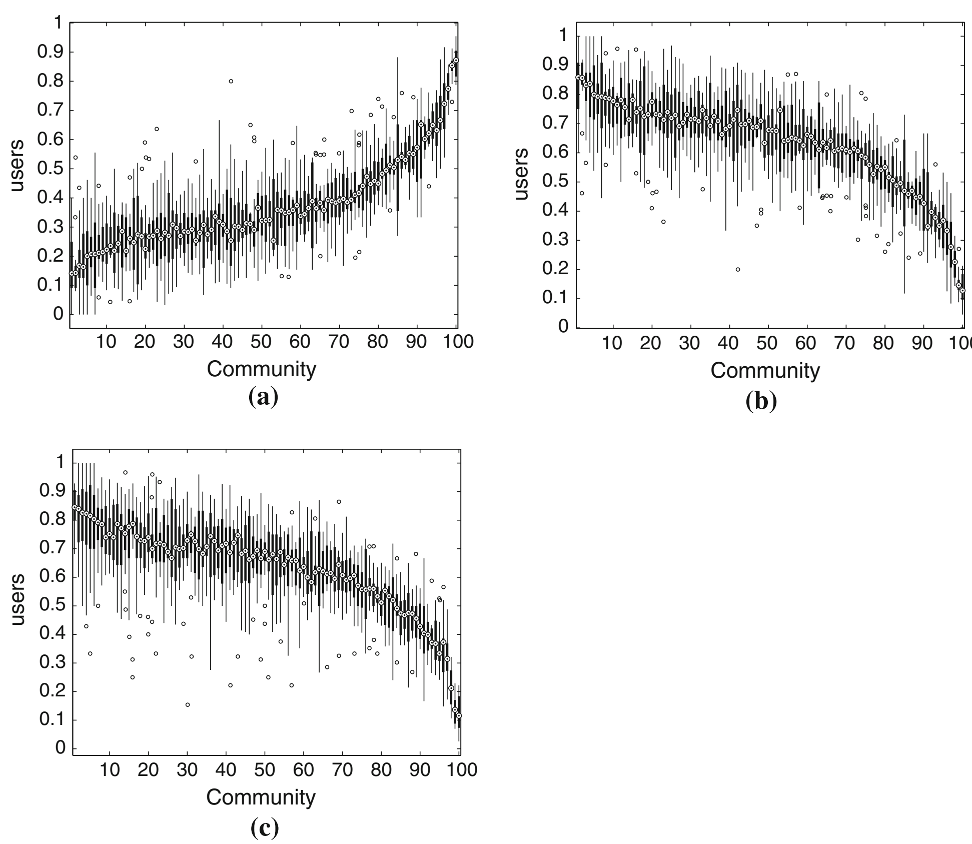
three types. Notice the communities with small core user proportion less than 0.25 are the ones that have high mean proportion of joiners and parters, as is expected. We call these communities *high user churn communities*. These communities form about one-fourth of the total number of communities. Similarly, communities with high core user proportion greater than 0.6 are the ones that have low mean proportion of joiners and parters. We call these *low user churn communities*. These communities form a smaller fraction of the discovered communities, about one-tenth. The majority of communities have a mean core user proportion in the range [0.25, 0.60], meaning the majority of communities maintain between one-quarter to one-half of their user memberships over time. This indicates a community stability over time which a number of applications (e.g., prediction) can exploit. Similar conclusions about tags in communities can be drawn from Fig. 13. Next we take a closer look at low and high user churn communities.

4.4.2 Low and high user churn communities

In this experiment we focus on community behavior (static or low churns vs. dynamic or high churns).

In Table 3, we present five low-churn communities and five high-churn communities. The first column shows the mean core user proportion for the community over the observed time period, the second column shows the standard deviation, and the third shows the average core user size. The fourth column, shows the top tags representing the community interest. By our definition, the low-churn communities have high mean core user proportion and high-churn communities have low mean core user proportion. Notice how the low-churn communities have lower variance around the mean compared to high-churn communities. Also low-churn communities cores are much larger than high-churn communities. Now by inspecting the top tags for the low-churn communities we notice that they have specialized, and narrow interests. For example, the top one is about “Fan Fiction Stories”, the following one is

Fig. 12 Proportion of users per community. **a** Core, **b** parters, **c** joiners



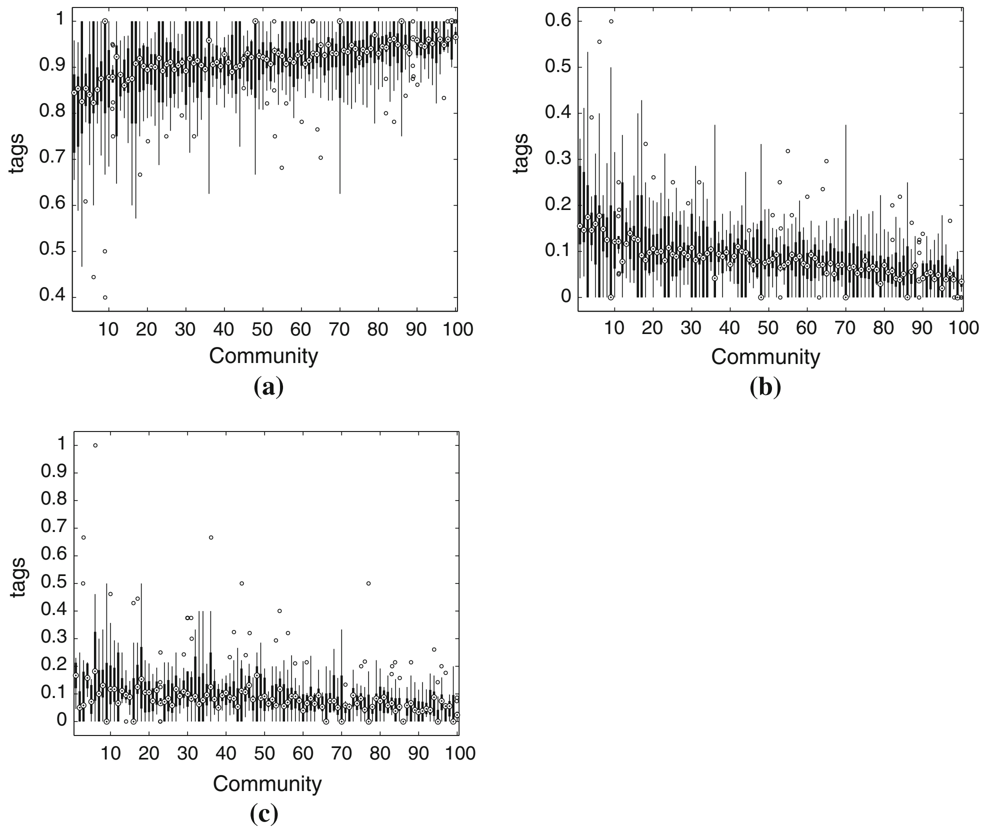


Fig. 13 Proportion of tags per community. **a** Core, **b** parters, **c** joiners

about "Cooking and Recipes", and the third is about "Web Programming". On the other hand, high-churn communities have more generic interests like "Shopping", "Search tools", and "Videos". Also in Table 3, we highlight core tags (in italics) among the top tags for each community. In general, low-churn communities have higher counts of core tags compared to high-churn communities.

Next we look at how low and high-churn communities top tags evolve over time. We present in Table 4 a sample low-churn community interested in "Politics" and another high-churn community interested in "Audio and Sound". For each community we list the top tags in time segments {1, 5, 9, 13}. We also highlight the core tags in italics. Generally, we see no dramatic differences between how top tags of low and high-churn communities evolve over time. But we can see that over time more core tags are represented in the communities top tags for both low and high-churn communities. The observed distinguishing factors

Table 3 Community user churn

Mean	SD	Average size	Tags
Low churn			
0.868	0.051	44.13	*bandom brendon spencer* fic *ryan patd* author slash *gerard frank* pair fandom *torchwood jack pete*
0.643	0.087	59.46	*recip food cook dessert bake vegetarian chocol* blog *cake chicken pasta bread cooki breakfast*
0.642	0.059	60.8	*net drupal asp* develop program tutori microsoft *modul mvc* cm howto facebook *silverlight* pattern
0.623	0.074	100.2	*rubi rail rubyonrail* program develop tutori test *gem* plugin web xmpp deploy howto github api
0.607	0.046	125.8	*socialmedia* market *facebook social media* twitter socialnetwork web2 advertis blog busi brand trend
High churn			
0.168	0.1212	12.6	document write refer program tutori howto wiki tip latex vim django develop python manual editor
0.173	0.1229	16.8	*shop ecommerc commerc import* busi *url* onlin websit web *paypal* magento *auction* internet *info*
0.193	0.1178	16.2	*search* googl *present searchengin* tool *powerpoint* web2 engin internet web visual *research* bing
0.201	0.1330	14.3	*confer event forum* commun collabor video calendar access present web2 tool web *webconferenc*
0.218	0.0856	28.8	*video movi stream film* onlin free *youtub televis* media entertain download search *cinema*

Core tags are shown in italics

Table 4 Community evolution

Week	Type
Low churn	
1	*polit govern* histori new econom law *usa* cultur blog refer research *war* activ *obama* media statist militari *crime women polici*
5	*polit govern* econom *usa* histori china new economi cultur *obama* activ *war polici women* law militari intern world *gender femin*
9	*polit govern* econom *usa* histori activ *terror* cultur *war* new law *women* economi *obama* societi *femin polici gender islam crime*
13	*polit govern usa* econom activ cultur *obama* new histori law *women femin war polici gender* militari *race* india economi corrupt
High churn	
1	*audio podcast sound* music book free read video mp3 literatur resourc librari blog *audiobook* web2 tool educ *record speech*
5	*audio podcast sound* free video speech mp3 music *multimedia audiobook record* bbc radio resourc text award *listen* teen
9	*audio podcast* video music *sound* free *multimedia* bbc mp3 *record audiobook speech* radio *sampl* award resourc media text
13	*audio podcast sound* music video text mp3 *audiobook speech multimedia record* free award radio voic ipod media *sampl* nois

Core tags are shown in italics

between low and high-churn communities are important in applications such as targeted advertising, and recommender systems.

4.4.3 Core users and tags in communities

This leads us to the question of how core tags and core users behave in low and high-churn communities. To observe this behavior we compute the number of core tags and core users and the corresponding number of time segments in which they occur for both low and high-churn communities.

In Fig. 14 we show the results for core tags. Notice that in low-churn communities about 20 core tags occur in all 15 time segments (the entire observed period). On the other hand, for high-churn communities, no single core tag occurred in all 15 time segments. Generally, we observe, in low-churn communities, that more core tags occur over many time segments. In high-churn communities we see that core tags occur over fewer time segments.

We present the results for core users in Fig. 15. For low-churn communities, more than 20 core users occur in 15 time segments while no single core user occurred in more than 3 time segments in high-churn communities. Again, we observe that core users occur over many time segments in low-churn communities while majority of core users occur in only one single time segment for the case of high-churn communities. Notice that the behavior of core users in high versus low-churn communities is more stark than that of core tags.

To illustrate how a community's core tags and core users correlate, we view the communities on the *xy*-coordinates, with the *x*-axis representing a community's core user proportion and the *y*-axis representing its core tag proportion. The results are shown in Fig. 16. Notice that all

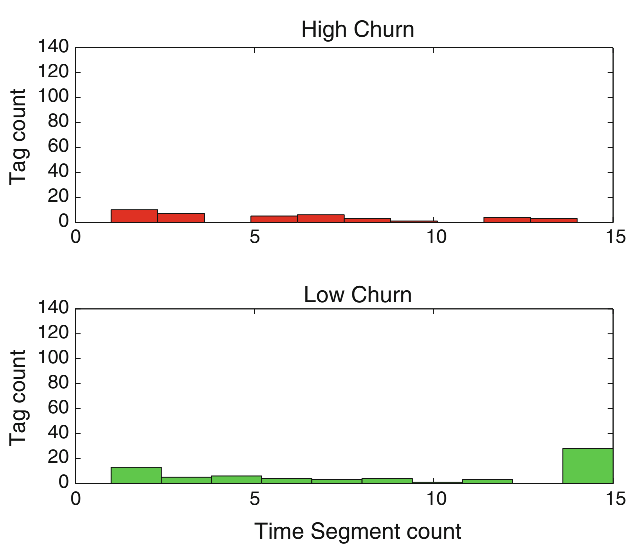


Fig. 14 Core tags behavior in low and high-churn communities

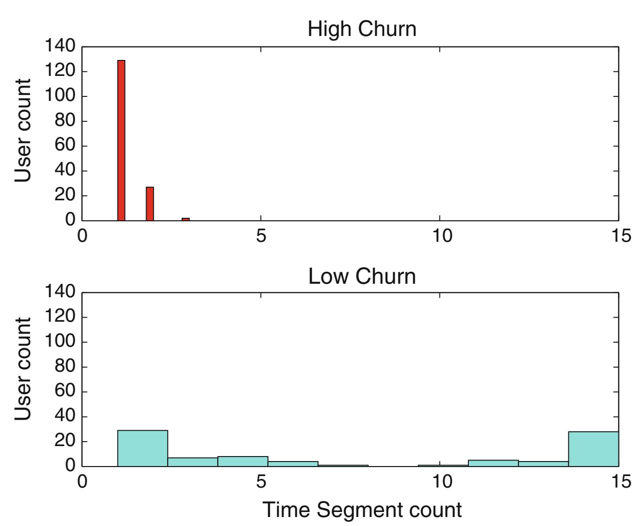


Fig. 15 Core users behavior in low and high-churn communities

communities have core tag proportions greater that 0.8 while the core user proportion vary widely.

The behavior of core tags and core users in low and high-churn communities calls for further investigation. Profiling of node behavior and characteristics to discern the role of individual nodes (users, tags) in forming communities and the possible social factors involved is an interesting future work.

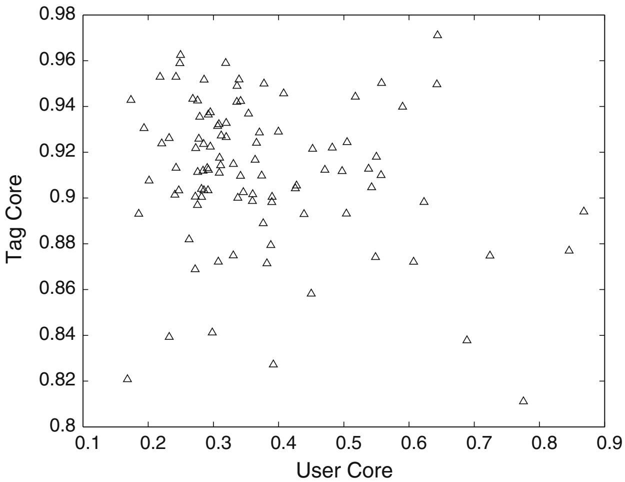


Fig. 16 Core users and core tags' relation to community evolution

4.5 Community relationships

In the above sections, we considered individual communities and how they change over time based on their users and tags. Now we look at how communities relate and interact with each other. For example, suppose we are given a community that is interested in the topic "Health". Here we ask, how does this community relate to other communities, for example "Cooking", or "Politics"? What are the closest communities to it, and how does their relationship behave over time?

Our approach focuses on the Joiner and the Parters per community. In Fig. 17, we show the "Politics" community and how it relates to other communities over time. For this purpose we consider the users that at some time segment have been members of the "Politics" community. For these users we look at what transitions they make over time, which communities do they transition to when leaving the "Politics" community and which communities do they come from when joining it? The figure shows us other communities that users with interests in "Politics" have shown interest in. Some examples of these communities are "Jobs", "Climate", "Social Media", "Culture", "Finance" and "Health".

We show similar results for the "Health" community in Fig. 18. The figure shows examples of communities that users who were members of the "Health" community were

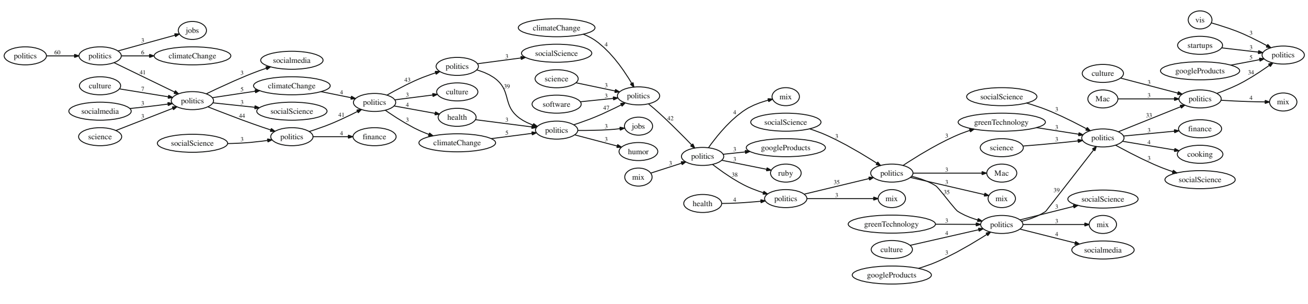


Fig. 17 "Politics" community users transition to/from other communities over time

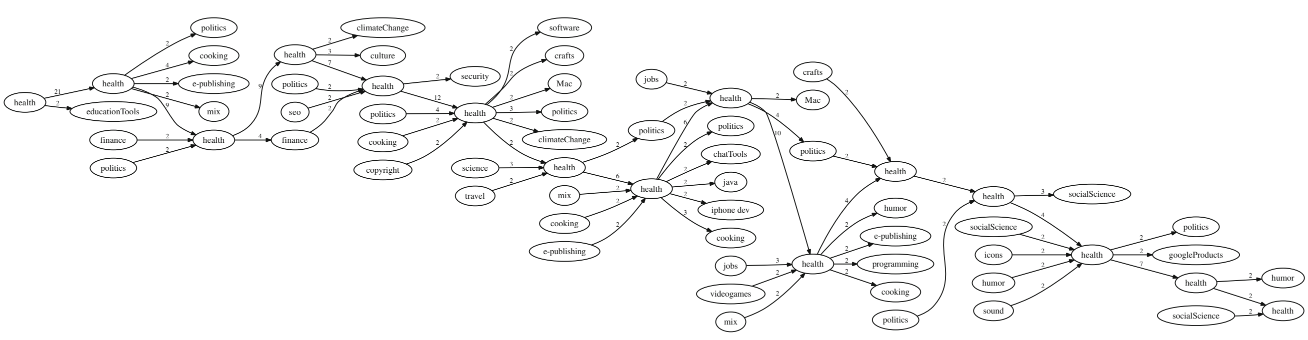


Fig. 18 "Health" community users transitions to/from other communities over time

also interested in, such as “Politics”, “Finance”, “Cooking”, “Science”, “Travel”, and “Iphone-dev”. In both examples, we see that user transitions across communities is an indicator of some implicit bond between the communities. Identifying communities and their relationships to other communities is important in prediction, focused or specialized search, and recommenders applications.

5 Summary

Observing the social bookmarking process over time offers many interesting insights. A 1-month duration can capture the majority of taggers and tags but not resources. Resources, taggers, and tag re-occurrences follow the common power-law distribution where a few elements are very active and the majority have very low re-occurrences. Co-occurrences among the three types show that a few resources, users and tags are popular while the the majority have minimal exposure. The plots, however, suggest a sustained daily tagging activity indicating growth in tags, users, and resources.

Our modeling approach for social bookmarking services over time involves segmenting the social tagging document for each resource. The model is then used to uncover structures in each time segment and uses a preceding time segment structure’s as a prior to determining the structures of the following time segment. This modeling process allows for the detection of communities and their evolution. A further inspection shows that communities are more dynamic along their user distributions than along their tag distributions.

We have also introduced a community dynamics representation of communities and their users and tags transitions that allows us to further inspect the behavior of users and tags in relation to communities. Based on this representation we observe low and high user churn communities. We see that low user churn communities have specialized and narrow interests versus the more generic interests of high user churn communities. We also observe that over time core tags gain more prominence in the community’s top tags.

Our examination of how core users and tags behave over time in both low and high user churn communities reveals that: (1) core tags and core users in low user churn communities are present in many time segments; (2) core tags and core users in high-churn communities are present in only a few time segments. We also examine the correlation between core user proportions and core tag proportions over communities and conclude that no correlation is present. Our illustration of how different communities are related based on user interest transitions over time using the community dynamics graphs shows that we can identify related communities that are meaningful.

We envision a number of applications that can benefit from our approach. For example, the developed model could be used to track users across communities over time, allowing for the ability to predict the probability of the user’s future community assignment. Such a predictive community assessment could support the prediction of future user interests, narrow down the scope of search (to reflect the user’s “current’ community), provide recommendations, improve the design of social bookmarking systems, and inform targeted community-oriented marketing and advertising campaigns. In our continuing work, we are considering these scenarios as well as considering: (1) an investigation of the different factors causing community formation and the role of individual nodes in forming communities—what are the first causes of community formation, and can these factors be discerned? and (2) A study of community evolution variation and the characteristics of community stability—why do some communities change greatly, while others are much more static?

References

Asur S, Parthasarathy S, Ucar D (2007) An event-based framework for characterizing the evolutionary behavior of interaction graphs. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining, KDD ’07, ACM, New York, pp 913–921

Bao S, Xue G, Wu X, Yu Y, Fei B, Su Z (2007) Optimizing web search using social annotations. In: Proceedings of the international conference on World Wide Web, pp 501–510

Bateman S, Muller MJ, Freyne J (2009) Personalized retrieval in social bookmarking. In: Proceedings of the ACM 2009 international conference on supporting group work, GROUP ’09, pp 91–94

Begelman G, Keller P, Smadja F (2006) Automated tag clustering: improving search and exploration in the tag space. In: Proceedings of the international conference on World Wide Web

Bercovitz B, Kaliszan F, Koutrika G, Liou H, Mohammadi Zadeh Z, Garcia-Molina H (2009) Courserank: a social system for course planning. In: Proceedings of the ACM SIGMOD international conference on management of data, pp 1107–1110

Carman MJ, Baillie M, Crestani F (2008) Tag data and personalized information retrieval. In: Proceedings of the 2008 ACM workshop on search in social media, SSM ’08, pp 27–34

Carman MJ, Baillie M, Gwadera R, Crestani F (2009) A statistical comparison of tag and query logs. In: Proceedings of the international ACM SIGIR conference on research and development in information retrieval, pp 123–130

Cattuto C, Baldassarri A, Servedio VDP, Loreto V (2007) Vocabulary growth in collaborative tagging systems. http://arxiv.org/abs/0704.3316

Cattuto C, Loreto V, Pietronero L (2006) Collaborative tagging and semiotic dynamics. http://arxiv.org/abs/cs.CY/0605015

Christiaens S (2006) Metadata mechanisms: from ontology to folksonomy … and back. In: OTM workshops, Lecture Notes in Computer Science, pp 199–207

Digital journal (2010) http://www.digitaljournal.com/

Esslimani I, Brun A, Boyer A (2011) Densifying a behavioral recommender system by social networks link prediction methods. Social Netw Anal Mining 1(3):159–172

Farooq U, Song Y, Carroll JM, Giles CL (2007) Social bookmarking for scholarly digital libraries. IEEE Internet Comput 11(6):29–35
Golder S, Huberman BA (2005) The structure of collaborative tagging systems. http://arxiv.org/abs/cs/0508082
Gruber T (2008) Collective knowledge systems: where the social web meets the semantic web. In: Web semantics: science, services and agents on the World Wide Web 6, pp 4–13
Guan Z, Bu J, Mei Q, Chen C, Wang C (2009) Personalized tag recommendation using graph-based ranking on multi-type inter-related objects. In: Proceedings of the international ACM SIGIR conference on research and development in information retrieval, pp 540–547
Halpin H, Robu V, Shepherd H (2007) The complex dynamics of collaborative tagging. In: Proceedings of the international conference on World Wide Web, pp 211–220
Hamouda S, Wanas N (2011) Put-tag: personalized user-centric tag recommendation for social bookmarking systems. Soc Netw Anal Min 1:377–385
Heinrich G (2005) Parameter estimation for text analysis. http://www.arbylon.net/publications/text-est.pdf
Heymann P, Koutrika G, Garcia-Molina H (2008) Can social bookmarking improve web search? In: Proceedings of the ACM international conference on Web search and data mining, pp 195–206
Kashoob S, Caverlee J, Khabiri E (2009) Probabilistic generative models of the social annotation process. In: Proceedings of the IEEE international conference on computational science and engineering, pp 42–49
Kolay S, Dasdan A (2009) The value of socially tagged urls for a search engine. In: Proceedings of the international conference on World Wide Web, pp 1203–1204
Leitner P, Grechenig T (2008) Collaborative shopping networks: Sharing the wisdom of crowds in e-commerce environments. In: 21st Bled eConference, pp 321–335
Li R, Bao S, Yu Y, Fei B, Su Z (2007) Towards effective browsing of large scale social annotations. In: Proceedings of the international conference on World Wide Web, pp 943–952
Li X, Guo L, Zhao YE (2008) Tag-based social interest discovery. In: Proceedings of the international conference on World Wide Web, pp 675–684
Liu B, Zhai E, Sun H, Chen Y, Chen Z (2009) Filtering spam in social tagging system with dynamic behavior analysis. In: Proceedings of the 2009 international conference on advances in social network analysis and mining, pp 95–100
Macgregor G, McCulloch E (2006) Collaborative tagging as a knowledge organisation and resource discovery tool. Libr Rev 55(5):291–300
Markines B, Cattuto C, Menczer F (2009) Social spam detection. In: Proceedings of the 5th international workshop on adversarial information retrieval on the Web, AIRWeb '09, pp 41–48
Markines B, Cattuto C, Menczer F, Benz D, Hotho A, Gerd S (2009) Evaluating similarity measures for emergent semantics of social tagging. In: Proceedings of the international conference on World Wide Web, pp 641–650
Marlow C, Naaman M, Boyd D, Davis M (2006) Ht06, tagging paper, taxonomy. Flickr, academic article, to read. In: Proceedings of the ACM conference on hypertext and hypermedia, pp 31–40
Mika P (2005) Ontologies are us: a unified model of social networks and semantics. In: International semantic Web conference, Lecture Notes in Computer Science, vol 3729, pp 522–536. International semantic web conference 2005
Neubauer N, Wetzker R, Obermayer K (2009) Tag spam creates large non-giant connected components. In: Proceedings of the 5th international workshop on adversarial information retrieval on the Web, AIRWeb '09, pp 49–52
Noll MG, Meinel C (2008) Exploring social annotations for web document classification. In: Proceedings of the ACM symposium on applied computing, pp 2315–2320
Plangprasopchok A, Lerman K (2008) Modeling social annotation: a bayesian approach. http://arxiv.org/abs/0811.1319
Plangrasopchok A, Lerman K (2007) Exploiting social annotation for automatic resource discovery. http://arxiv.org/abs/0704.1675
Ramage D, Heymann P, Manning CD, Molina HG (2009) Clustering the tagged web. In: Proceedings of the ACM international conference on Web search and data mining, pp 54–63
Sen S, Vig J, Riedl J (2009) Learning to recognize valuable tags. In: Proceedings of the 13th international conference on intelligent user interfaces, IUI '09, pp 87–96
Song Y, Zhuang Z, Li H, Zhao Q, Li J, Lee WC, Giles CL (2008) Real-time automatic tag recommendation. In: Proceedings of the international ACM SIGIR conference on research and development in information retrieval
Twitter (2010) http://twitter.com/
Veres C (2006) The language of folksonomies: what tags reveal about user classification. Nat Lang Process Inf Syst 3999/2006:58–69
Vig J, Sen S, Riedl J (2009) Tagsplanations: explaining recommendations using tags. In: Proceedings of the 13th international conference on intelligent user interfaces, IUI '09, pp 47–56
Wang J, Davison BD (2008) Explorations in tag suggestion and query expansion. In: Proceedings of the 2008 ACM workshop on search in social media, SSM '08, pp 43–50
Wetzker R, Plumbaum T, Korth A, Bauckhage C, Alpcan T, Metze F (2008) Detecting trends in social bookmarking systems using a probabilistic generative model and smoothing. In: Proceedings of the international conference on pattern recognition (ICPR), pp 1–4
Willard T (2009) Social networking and governance for sustainable development. http://www.iisd.org/pdf/2009/social_net_gov.pdf
Wu X, Zhang L, Yu Y (2006) Exploring social annotations for the semantic web. In: Proceedings of the international conference on World Wide Web, pp 417–426
Xu S, Bao S, Fei B, Su Z, Yu Y (2008) Exploring folksonomy for personalized search. In: Proceedings of the international ACM SIGIR conference on research and development in information retrieval, pp 155–162
Yanbe Y, Jatowt A, Nakamura S, Tanaka K (2007) Can social bookmarking enhance search in the web? In: Proceedings of the joint conference on digital libraries, pp 107–116
Yin D, Hong L, Xue Z, Davison BD (2011) Temporal dynamics of user interests in tagging systems. In: Proceedings of the twenty-fifth AAAI conference on artificial intelligence (to appear)
Zhang D, Mao R, Li W (2009) The recurrence dynamics of social tagging. In: Proceedings of the 18th international conference on World Wide Web, pp 1205–1206
Zhou D, Bian J, Zheng S, Zha H, Giles CL (2008) Exploring social annotations for information retrieval. In: Proceedings of the international conference on World Wide Web, pp 715–724