Abstract - Social bookmarking systems are a promising tool for classifying web resources. To be able to use such systems intelligently, it is important to understand how they work. This study takes a look at one such system to examine tagging and bookmarking patterns. The existing literature tends to view these patterns somewhat simplistically. We specifically examine three questions: (1) at what rate do bookmarks accumulate? (2) To what extent do early tags influence later tagging behavior? (3) How do the top 10 tags evolve? Given the influence of the literature on the direction of future research on social bookmarking systems, a more careful analysis may be of use.

Keywords: Tagging patterns, Social Bookmarking Systems, Tagging behavior.

1 Introduction

With the development of the social web, many interesting tools that focus on the social interaction and collaboration between users have appeared including Facebook, Delicious, Flicker, CiteULike, Twitter etc. These social systems provide new forms of interaction between users and resources. Social bookmarking systems (SBS) such as Delicious, CiteULike, Dogear, etc. are sometimes defined as “Public link management systems”[1]. They were first created to provide users with bookmark portability. They have shown the ability to attract a significant number of users who have collected a large set of resources. They have also attracted the research community because of their ability to generate a significant amount of content in a short period of time. Unlike library classification systems, social bookmarking systems provide tools to classify documents in an unrestrict way where “classification” arises from uncoordinated agreement among users on terms that describe documents. This classification has been called ‘Folksonomy’ by Thomas Vander Wal [2]. To optimally use the data generated by these systems it is important to understand the nature of the systems. This research takes a closer look at the nature of such systems by utilizing Delicious, which is one of the early social bookmarking systems with a significant number of public users.

2 Previous Work

Social Bookmarking Systems (SBSs) contain a large number of bookmarks that associate users and resources. Tagging provides a mechanism to both organize resources and facilitate browsing and discovery of new resources. Tags are also being used to supplement indexing and ranking as described in [3-6]. Some research focuses on understanding how folksonomies develop. Other researchers study tagging behavior: how users tag resources, what motivate users toward those systems, what makes some resources more popular than others, how the vocabulary evolves in SBSs, etc.

An early study done by Golder and Huberman [7] analyzed a small set of resources (212 resources) and found many regularities in the system. They found that tagging patterns tend to stabilize after about 100 bookmarks. Furthermore, they found that early tags tend to be more popular. They observed that early tags are more general and that they provide a sense of agreement among users which also implies that later tags patterns where mainly influenced by earlier ones. They relate the stabilization of this pattern to two factors: imitation by other users and shared knowledge. They argue that since Delicious provides users with a list of tag suggestions; initial tags influence later taggers[7].

A similar study done by Wetzker et al. [8] analyzed a data set of 142,341,551 bookmarks. In their work; they looked at delicious system activity as a whole. They found that

• Users activity follows a power law distribution
• Burst in the popularity of resources are caused by their appearance on delicious main page or another popular site or it could be the similar interests of users on the network.
• “Delicious community pays attention to new resource for a very short period of time. As a result URLs receive most of their posts very quickly and disappear shortly afterwards”.

Halpin, Robu and Shepherd [9] looked at the dynamics of SBSs and the possibility of defining collaborative systems as complex systems. They found that tagging frequency follows a power law distribution. Marlow et al. [10] conducted a study that classified systems. They argued that user behavior is influenced by many factors such as: tagging rights, tagging support, aggregation, type of object being tagged, source of material, etc. They also discuss two motivations behind using SBSs: organizational and social.

Farooq et al. [11] looked at set of data from CiteULike and Delicious; in their work, they identified six tags metrics:

1. Tag growth: looking at how the tag vocabulary evolves over time
2. Tag reuse: measuring how many times tags in the system have been reapplied
3. Tag non-obviousness: comparing the tag to its resource to see if the tag helps describe the resource.
4. Tag discrimination: looking at how the tag discriminates a resource
5. Tag frequency: characterizes tags based on their usage frequencies
6. Tag patterns: how users tag resources over time

Some of these studies were done during the early stages of Delicious’ history when the system did not have much data[7]. Kipp and Campbell [12] focused on frequency and co-word analysis to investigate whether collaborative tagging could support library cataloging and indexing. Farooq et al. [11] focused their analysis on academic paper social bookmarking which might be applicable to other SBS such as Delicious. Although they have mentioned many useful metrics which we are looking at, they did not examine the patterns that can arise from those metrics. Golder and Huberman [7] looked at different tagging patterns and usage in SBS. They have also analyzed user’s behavior but their analysis was limited to two small sets (212 of popular resources with their bookmarks, 229 users with their bookmarks) and their analysis was also done when Delicious was still in its infancy. The study by Wetzker et al. [8] looked at Delicious users and resources as a whole.

In this study we extend their work. We examine some of the findings of previous work by analyzing a set of 41,469,488 resources. With this data set we examine three questions: (1) at what rate do bookmarks accumulate? (2) To what extent do early tags influence later tagging behavior? (3)How do the top 10 tags evolve?

3 The Data Set

Delicious was created by Joshua Schachter in 2003 and acquired by Yahoo in 2005. In 2008, Delicious announced that they had reached 5.3 million users and 180 million unique bookmarked resources. The exact number of users when Delicious was re-sold to AVOS on 2011 has not been found in the literature.

Our data set was crawled from November, 2009 to January, 2010, then again from May, 2010 to August, 2010. The data set is described in Table 1

| Total resources | 41,469,488 |
| Total tags      | 214,420,801 |
| Total bookmarks | 73,216,330 |
| Total users     | 723,342 |
| Min bookmarks   | 1 |
| Max bookmarks   | 9099 |
| Avg bookmarks per resource | 2 |
| Avg tag per bookmark | 3 |

4 Data Analysis

Using the dataset described above, various issues addressed or discussed in the literature, either explicitly or implicitly were examined. The goal was to verify what had been reported in the literature or qualify those findings. The issues explored included the influence of early tagging behavior, the propagation of bookmarks on resources, and the adoption of dominant tags. Each of these issues is addressed below.

4.1 At what rate do bookmarks accumulate?

Golder and Huberman [7] noted that among their 212 most popular resources they found that most resources receive their bookmarks in the first 10 days while some others stay in the system for at least six month before they experience a burst in popularity. They observed that all resources experience “peak in popularity”. We investigated those pattern and found that both patterns exists on our data set “Early and Late adoption” but we have also found one more “Gradual pattern”.

4.1.1 Gradual Pattern

In this pattern, the relationship between the number of bookmarks associated with a resource and time is generally linear. The resource accumulates bookmarks gradually with time. This indicates ongoing attention from users. Figure1 shows this linear relationship. Other than the users’ interest, this also can be an indication of the type of this resource. In Figure 1 the resource is the search engine 'Bing', which is a resource that might be used by users on a regular basis and whose popularity is increased through recommendations of users

![Figure 1: Gradual Bookmarks accumulation](URL 5620779)

4.1.2 Early adoption

Under this pattern a resource accumulates bookmarks very quickly, in some cases a thousand bookmarks in one day, and then a very slow growth rate ensues (see figure 2). These kinds of resources attract attention instantly. We suspect that external events may be more important here. For example this can be a news article or an important event. This pattern was confirmed by [7, 8]
4.1.3 Late adoption

Under this pattern a resource is introduced and not bookmarked frequently. Suddenly, attention is attracted to it (Figure 3). This pattern can be an indication of many things: a sudden increase in the importance of that resource or external events that might lead to the popularity of this resource.

Note that Figure 2 and Figure 3 shows a set of events related to Delicious as a system. There is some indication that some resources grow in popularity related to these system events.

4.2 To what extent do early tags influence later tagging behavior?

Golder and Huberman [7] found that users tend to use more general tags first and those general tags are the most frequently used meaning that the tagging behavior of later users is heavily influenced by early taggers. They also pointed out that after 100 bookmarks the pattern of tagging stabilizes due to imitation and shared knowledge[7]. While this current study is generally supportive of the Golder and Huberman findings, the actual structure is somewhat more complex. To examine how tagging evolved over time, we looked at the resource at different points in its lifetime (a) one week (b) one month (c) three months (d) six months (e) one year and (f) the most recent date for which we have information (as little as one year and as much as six years). We also considered the tag set at this point as a mature set. The accumulation of tags over time was visualized as shown in Figure 4 where the resource is represented by a blue circle in the middle and each tag is placed on a spiral ordered by creation date moving clockwise. The size of the triangle increases as the frequency of use increases. Tags that have been used more than once are labeled. Running this visualization on a sample, we found a number of different patterns illustrated in the next section.

4.2.1 Early Tag Dominance Pattern:

In this pattern, as the bookmarks accumulate for that resource, early tags are used most frequently. If we consider the popularity of a tag by its usage, earlier tags are more popular in this pattern. This pattern was first described in [7] and it is illustrated in (Figure 5a-5f). The selected resource accumulates bookmarks slowly and the increase in the tags usage is gradual so that we can see at the end that early tags are the most popular. The analysis shows this pattern to be far from the dominant pattern. Over a sample of 53 resources, this pattern was observed about 49% of the time.

4.2.2 Mixed Tag Dominance Patterns

In this pattern, as the resource accumulates more bookmarks, we see that both early and later tags are dominant. This pattern is illustrated in Figure 6. In other cases it is also possible that the later tags are the most dominant.

This pattern is illustrated in Figure 6. The first bookmark that belongs to this resource was added to delicious on 02-11-2004. After one week this resource accumulated 8 tags with usage frequency of 1 for each tag. A month later, there were no changes in this pattern. Three months later (Figure 6b), two new tags were added. After six months (Figure 6c), the tag usage started to increase. Then, at one year, more tags accumulated and the tag usage has increased. At the last known bookmark for this resource on our data set (10-8-2010), the most popular tag is 'programming' which was
added to the system at about three months as were other popular tags (Figure 6e). Variations of this pattern were observed for 38% of the 53 resources that were examined.

Figure 5: Early tag dominance pattern

4.2.3 Rapid versus slow growth patterns

As illustrated in Figure 7, some resources grow rapidly accumulating a large set of tags from their first day in the system while other start with few bookmarks. Figure 7 shows two snapshots of two different resources after one week from the day they were first added to the system. The resource on the left has about 100 bookmarks at the end of the first week while the resource on the right has two bookmarks with 8 tags. The visualization on their tag set at the latest date shows that both resources develop a similar pattern even though one grew slowly and the other rapidly.

Figure 7: Fast vs slow growth

Golder and Huberman [7] hypothesized that the order of tag growth is influenced by imitation and shared knowledge. Other studies followed that tried to build a model to simulate the behavior of those systems focusing on imitation as the only factor that affected the behavior [9, 13]. Another study [14] built a model in which they simulated the effect of imitation and shared knowledge, their model performed better than the previous models. We show here that the
underlying structure may have more factors that influence the behavior. Figure 8 shows that the growth of the top 10 tags were influenced by imitation. The first four tags that were added to the system were ranked by their order (1st, 2nd, 3rd and 4th) and all the tags that were added to the system on the first day were the most popular tags. Note that ‘webdesign’, ‘icon’ and ‘design’ were all added to the system on the first day and they were fourth on the sequence and their crosses on the chart are overlapping. Figure 9 shows the timing order for top 10 tags for the resource www.freelance.com. We can see that the most frequently used tag was added to the system on day 133 and it was the second on the list of tags provided by the user ‘finrod’. We can also see that ‘outsourcing’ was added to the system on day 266 and it was ranked 3rd. One would expect that the user would provide more descriptive tags first but surprisingly, ‘outsourcing’ was the seventh on the list of tags provided by the user and it was ranked 3rd. Furthermore, Delicious provides a list of seven recommended tags which are the most popular for that resource and the seven most frequently used tags for that user. If users just imitate each other, we would expect the tagging behavior to be somewhat similar to Figure 8 but Figure 9 shows a totally different behavior.

5) Conclusions about Tagging Patterns: Resources start with a set of tags. The most popular tags may be from the initial offering or they may appear later. When a resource attracts a lot of early attention there is even less chance that dominant tags will be the first used. Clearly, there are many factors influencing the tagging pattern other than the influence of early tags. Does it mean that each group of users perceives the resource differently or is it a change in the users’ interests? To further answer this question, we are planning an examination of differences between the semantics of tags that become dominant and those that do not.

### Table 1: Top 10 Tags Ordered by Time

<table>
<thead>
<tr>
<th>Rank</th>
<th>Tag</th>
<th>Day Order</th>
<th>List Order</th>
<th>Frequency</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>favicon</td>
<td>1</td>
<td>1</td>
<td>152</td>
<td>barganburglar</td>
</tr>
<tr>
<td>2</td>
<td>generator</td>
<td>1</td>
<td>2</td>
<td>107</td>
<td>barganburglar</td>
</tr>
<tr>
<td>3</td>
<td>tools</td>
<td>1</td>
<td>3</td>
<td>82</td>
<td>barganburglar</td>
</tr>
<tr>
<td>4</td>
<td>webdesign</td>
<td>1</td>
<td>4</td>
<td>82</td>
<td>barganburglar</td>
</tr>
<tr>
<td>5</td>
<td>icon</td>
<td>1</td>
<td>4</td>
<td>58</td>
<td>biscotheque</td>
</tr>
<tr>
<td>6</td>
<td>design</td>
<td>1</td>
<td>4</td>
<td>53</td>
<td>conque</td>
</tr>
<tr>
<td>7</td>
<td>icons</td>
<td>1</td>
<td>2</td>
<td>47</td>
<td>conque</td>
</tr>
<tr>
<td>8</td>
<td>image</td>
<td>1</td>
<td>7</td>
<td>34</td>
<td>biscotheque</td>
</tr>
<tr>
<td>9</td>
<td>web</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>conque</td>
</tr>
<tr>
<td>10</td>
<td>webservice</td>
<td>1</td>
<td>6</td>
<td>23</td>
<td>biscotheque</td>
</tr>
</tbody>
</table>

### Figure 8: Top 10 tags ordered by time

### Figure 9: Top 10 tags ordered by time

4.3 How do the top 10 tags evolve?

We define the top 10 tags as the 10 most frequently used tags regardless of when they were introduced in the system. We inspected the frequency of each of those tags at various points in time – every 30 days. In general we observed two
patterns: exponential popularity growth and linear popularity growth. The exponential pattern where a tag stays unpopular for a while and then its popularity explodes (Figure 10). On the other hand, the linear pattern experiences a sudden growth at the beginning and then a stable growth (Figure 11). The exponential pattern corresponds to the early adoption pattern found in the bookmark accumulation while the linear pattern corresponds to the late adoption. We tried to relate those tag growth patterns to the observed bookmarking patterns (Section IV-A). We found that gradual and late adoption patterns tend to follow the exponential growth pattern while the early adoption pattern tends to follow the linear growth pattern. This is not to generalize the case because it is not true all the time but it is to raise the point that social bookmarking system behavior needs further investigation. Sometimes the top 10 tags will start to grow exponentially and then stabilize to a linear pattern of growth. We are clearly in need for a further understanding of these different cases so we can make more coherent conclusions on those tags.

![Figure 10: Exponential Popularity](URL145150)

![Figure 11: Linear Popularity](URL124576)

5 Discussion:

This paper reexamined some of the assumptions about how users interact with resources in social bookmarking systems. This analysis examined how resources accumulate tags. We found that some resources accumulate tags rapidly while others accumulate them slowly. Which tags become the most popular was also examined. In folksonomies, tags are measures of agreement among users. It is not necessarily the case that early tags are the most dominant. If the early tags are the most dominant, it may simply be a matter of users imitating early tags or it may be the case that there is agreement among users. When tags that appeared later become the most dominant it may indicate a disagreement among users, or it may be the case that users are more discriminating about which are the correct tags. This implies that the tagging practice is not as influenced by early taggers or Delicious’ suggested tags as some have suggested [7, 8].

Analyzing how bookmarks accumulate overtime, we found that different behaviors may attribute different meanings. A resource that attracts attention gradually might not be as useful as a resource which attracted attention early on. For example: in searching for extreme events such as natural disasters we might need to consider a resource that follows the early adoption pattern. On the other hand, searching for a popular restaurant or popular shopping sites we might want to consider those resources that grow gradually because those sites popularity might grow slowly through recommendations by users.

Both bookmark and tag evolution patterns need to be considered when using SBSs. For example, if we are using tags for ranking, we should consider how tags evolve and how the resource accumulates bookmarks. Furthermore, when ranking documents, if we consider only early popular tags, we might be losing an important description that can be obtained from the later popular tags and vice versa. Similarly, tag accumulation rate should also be considered. A resource that accumulates tags early should not be treated the same as a resource which accumulated tags slowly.

We were able to confirm some of the findings in [7] and [8] in what we call an early adoption pattern. Wetzker et al [8] states that “Delicious community pays attention to new resources only for a very short period of time. As a result, these resources receive most of their posts very quickly and disappear shortly afterwards”. We confirmed the existence of this pattern as well as other patterns such as the late adoption patterns described in [7] in which a resource experiences a sharp growth after a period of suppression. We have also found that other resources experience gradual growth over time which might be due to continued user interest.

We found that there are different possible scenarios of how those tags grow over time. There are two clear patterns: exponential and linear but we can’t generalize the finding without further investigation of those patterns and their meanings. A tag that grows linearly might mean a rapid agreement between users while a tag that grows exponentially means a sudden change in this agreement. The growth of late tags might as well be a combined effect of imitation and shared knowledge in which a user might add all the recommended tags first and then add their tags which can attract more users later.
6 Conclusions

This paper examined how tags accumulate over time and how those tags grow in popularity. We have confirmed the existence of the patterns described in early research [7-9, 12] but we have also observed other patterns that were not discussed in the literature which might be important. It would be useful to examine the semantics of dominant tags, potentially through a time analysis of when specific semantics become dominant. Also, more careful analysis of the cognitive aspects of user behavior might be useful in understanding the semantics of the list of tags provided by the user to describe a specific resource.

Research reported in this publication was supported by Saudi Arabian Cultural Mission to the U.S.

7 References