Ranking Bookmarks and Bistros: Intelligent Community and Folksonomy Development

Benjamin Szekely, Elias Torres
Harvard University
{bszekely,torres}@fas.harvard.edu
May, 2005

Abstract

Online communities are an important forum for people to share information on the internet. Historically, online communities have been centered around free-text sharing and simple user rating systems. With the shadow of the Semantic Web looming behind many new innovations on the internet today, online communities are beginning to adopt semantic techniques such as tagging and folksonomies as a vehicle for the sharing and classification of information. Gourmetvillage.org is a prototype online community that uses tagging to provide an innovative restaurant rating mechanism. Users may provide free-text reviews and tags for any aspect of a restaurant. **UserRank** is an algorithm based on Google's [**?**] **PageRank** that provides a ranking of users based on whose taggings are most often followed. **TagRank** provides a ranking of tags based on the ranking of users. **UserRank** may in fact be used to rank any entity in a community based on user association. del.icio.us contains a significant dataset of users and tagged entities with which to test these algorithms. **UserRank** and **TagRank** yield different rankings than their count-based counterparts. When drilling down into the results, the top user according to **UserRank** only made a few taggings, yet hundreds of users in the system agreed with his taggings, showing that these algorithms capture community consensus fairly well.

1 Introduction

Online communities are rapidly adopting folksonomies and tagging to classify and and share information. In this work, we introduce algorithms that operated on tagged bodies of knowledge with the goals of finding significant information, folksonomy convergence, and automated expert discovery. The first section provdies a detailed introduction to folksonomies and tagging. The next section introduces a prototype online community that relies heavily on tagging as a motivating example. In the third section, we summarize our design goals based on the prototype community. Next, we describe our algorithms in detail. Finally, we evaluate our algorithms using an existing real-world dataset and draw conclusions.

2 Folksonomy and Tagging

It would be futile to attempt to create a new class of online communities without first looking at the latest social computing trends. Currently, folksonomies and tagging are the most recent, successful innovations in this space and so have become the focus of our research and development efforts. We first introduce these concepts before showing their use in intelligent community development.

2.1 Folksonomy

"Folksonomy is a neologism for a practice of collaborative categorization using freely chosen keywords" [1]. The etymology of the word is "a taxonomy created by the people." Folksonomies are gaining importance in online communities because their results are the product of the combined efforts of the actual consumers of the information. This product more accurately reflects the community consensus than previous professionally organized classification efforts. This community-based set of terms can be considered extremely valuable for communities that wish to capture and share knowledge. For example, CitySearch's online restaurant guide provides a single category for all Indian cooking. In a metropolitan area with a large number of Indian restaurants, it would beneficial to differentiate between general North Indian cuisine and the more specific Punjab style. The core belief behind the notion of folksonomy is that only the people as a community can arrive at such a complete and detailed classification system

With minimal effort, folkosonomies enable online communities to classify large, bodies of information and resources for future search, retrieval, archiving and sharing. Folksonomies empower the community members to define the vocabulary that identifies them in the broader Internet. Next, we will present the definition of tags and their perceived benefits to online communities.

2.2 Tagging

What is tagging? We present two complimentary definitions. At the most basic level, tagging is *the mindless process of ascribing keywords to objects for the purpose of categorization*. Holistically, however, tagging is a *potentially powerful method of harnessing the collective classification efforts of large knowledge-centric social networks.* In other words, tagging is simple, yet powerful method for realizing a folksonomy. What's a tag? A tag is any keyword, usually belonging to a folksonomy, that may be ascribed to an entity or concept. For example, one might tag a vacation photo with 'beach', 'sunset', and 'cameraphone'. Tags have already proven useful in folksonomies on the web such as Flickr and del.icio.us. Tags are beginning to show great potential in many areas of social computing, creating new benefits in three important categories: networked individualism, emergent intelligence and social discovery [2].

Before discussing these esoteric categories, we provide a few examples of tag- based online communities: Flickr, del.icio.us and Technorati. Flickr allows users to upload, organize and share photos. Flickr uses tags for many purposes. Users may tag a photo to make it easier to find later or use tags so others can find their photos. Flickr uses tags for personal, popular or standardized grouping of photos. For a personal example, the 'mynecktie' tag groups the author's pictures of all of the neckties he wears. Popular and group tags are more obvious, such as geographical tags or 'whatsinyourbag' tag that depicts the kind of technical gadgets people carry in their bags. Standardized tags are those whose meaning is well understood by many such as the 'me' tag used to group pictures of the folks who post to Flickr. Del.icio.us is very similar to Flickr though it uses tagging for bookmark management. Lastly, Technorati is a special search engine that utilizes tagged content from Flickr, del.icio.us and others to categorize its contents and provide visitors with more organized content. Clearly, tags provide an important catalyst for social computing. We now describe the three benefit categories of tagging [2].

Networked individualism: Tagging is a key supporter of networked individualism. In an age where individuals are spending most of their time online, we need ways to provide incentives for participation such as attention, feedback, implicit reputation and group forming. Individuals must have the ability to belong to multiple groups and communities. In short, tagging provides users an easy, yet powerful method to express themselves within a community.

Emergent intelligence: Tagging is also a mechanism by which intelligence can emerge as a social byproduct of community participation. Wikis are already a perfect example of this social phenomenon. We would like to build on theories found in *The Wisdom of Crowds* by James Surowieck, which holds that group decisions are better than those of individual experts. We are determined to show that invaluable information surface in a community that individual restaurant reviewers could never gather in a single visit to an establishment as reviewers are unlikely to try every dish in a restaurant discover dish variations or overcome their own biases. How can a reviewer like Zagat validate the rumor that a restaurant tends to use lower quality ingredients across several of their dishes? These and other questions are the kinds of issues we would like our application to address.

Social discovery: Lastly, tagging provides a unique tool for forming personal connections and groups. Implicitly, users who consisently use the same set of tags have formed a group and can seek each other out using simple search techniques. In addition tagging is an exellent method for discovering higher- level user preference. For example, a restaurant community could easily find members with very similar tastes.

2.3 Spagging: the problems of tagging

Any system on the Internet stands to be abused for profit or personal satisfaction. Like anything user input system, tagging can be automated to cause havoc. Spam-tagging, or *spagging* has several negative effects such as false classification, dilution of meaning, and overabundance of advertising. Online communities are also susceptible to flooding with offensive content. These issues are important to consider when designing tag-based online communities.

3 gourmetvillage.org

In this section, we introduce gourmetvillage.org, a prototype online community that makes use of folksonomies and tagging to evaluate restaurants. Presently, the most popular restaurant guides such as Zagat are based on one-time surveys and do not allow granular user feedback about arbitary aspects of a restaurant or a particular dish. Other online forums such as epinions.com allow freeform user comments but only a single evaluation score, failing to accurately capture the quality of restaurant.

The first-order object in our community is the restaurant. However, each restaurant may contain many *facets* such as location, decor and service, as well as any number of dishes, which themselves may have facets. Users may enter freeform, detailed comments on as many facets of a restaurant or dish as desired, creating additional facets as necessary. In addition to textual comments, users may apply specific tags to facets. Tags may indicate quality, characterization, categorization, or any property. Again, users may choose from an existing set of tags or add new tags. Our belief is that our user community will settle on a core set of facets and tags that will result in a richer restaurant evaluation system than would any set of predefined feedback forms designed by comittee. Our community will, in a sense, be defined by the set of facets and tags that users have decided are necessary to evaluate restaurants.

Like most web applications, the desktop web browser will be the primary access point to gourmetvillage.org as seen in Figure 1. The most basic functions will include:

1. Browse restaurants by facet, i.e. location, service, food quality, specific dish quality, etc...
2. Search for a restaurant using a complex query
3. Add freeform comments to a facet of a restaurant or dish
4. Tag a facet of a restaurant or dish
5. View top restaurants, dishes or users

In Figure 1, the user is povided type-ahead assistance while tagging. This listing is provided by the **TagRank** algorithm. This ranking will help the user choose tags relevant to the current facet. In this way, the community will converge on a folksonomy of tags. We will discuss these ideas in detail in a later section.

3.1 A Simple Scenario

H.T. Kung, a Harvard Computer Science Professor, needs to bring a handful of prospective PhD students to dinner. H.T. enters gourmetvillage.org and does a facetted search of **location:harvard** and **cuisine-type:indian**. Among others, Bombay Club in Harvard Square shows up in the search results. H.T. navigates to the page for Bombay Club and finds that many other users have tagged facets of Bombay Club with favorable tags such as **food:wonderful**. In particular, H.T.'s favorite dish has been tagged with positive tags and an encouraging freeform description. H.T. decides to take the students to Bombay club.

3.2 onthetown.gourmetvillage.org: entering the village with a cell phone

Entering gourmetvillage.org via a cell phone (onthetown.gourmetvillage.org) is certainly a compelling use case. Satisfied gourmets might want to capture their thoughts on the ambiance of a restaurant while still basking in the dim light surrounded by the smell of savory food. On the other hand, disgusted customers might want to enter their opinions with the bad taste still in their mouths. Entering opinions in onthetown.gourmetvillage.org can also be a good ice-breaker on a tough date or a way to pass the time waiting for a movie to start.

In particular, tagging is a perfect paradigm for the mobile phone because input can be achieved by simple scroll and select from a list provided by **TagRank** More patient or tech-savvy users might be willing to enter freeform text from a Blackberry, while others may wait until they have access to a PC to add more details.

Let's continue with H.T's story. After selecting a restaurant on gourmetvillage.org, H.T. takes the prospective students to Bombay Club restaurant. He finds that the decor is sophisticated. However, H.T. decides to be adventurous and orders Lamb Rogan Josh that turns out to be terrible instead of ordering his favorite dish. Eager to show the new students how cool he is, H.T. pulls out his Sprint Vision enabled Sony Treo and enters onthetown.gourmetvillage.org. He navigates to Indian restaurants in Boston and quickly finds The Bombay Club. In the review page, H.T. selects the "decor" facet. The "decor" facet page has a space for free-form text and as well as a list of tags. Pressed for time, H.T. selects the "sophisticated" tag, then backs out to the Bombay Club main page. Unfortunately, nobody has created a facet for "Lamb Rogan Josh" so H.T. selects "add dish". H.T. types in "Rogan Josh", and then tags dish with **taste:poor**. Delighted by this display of technology, every perspective student enters their own opinions and accepts an invitation to study Computer Science at Harvard.

4 System Design Goals

The remainder of this paper focuses on the underlying algorithms needed to enable tag-based online communities. An online community such as gourmetvillage.org requires a dynamic ranking of users and tags. Before introducing these algorithms **UserRank** and **TagRank** respectively, we discuss the core capabilities and benefits of the system as well as some problems to be addressed.

Given a ranking of users, we can provide a ranking of any set of entities in the system. For example, the top bookmarks in del.icio.us are those tagged by the highest ranked users. The interpretation of this sort

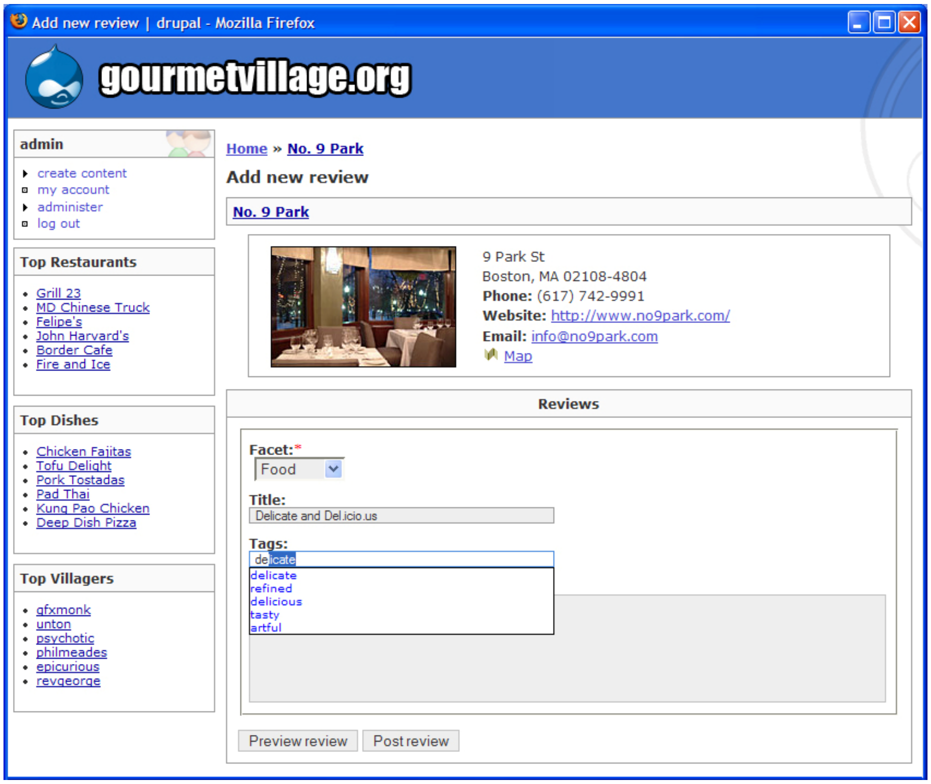

Figure 1: **Reviewing a restaurant at gourmetvillage.org.**

of ranking requries caution, however. Several top users in gourmetvillage.org might tag a restaurant with unfavorable tags. This does not mean, of course, that the restaurant is the best. It does indicate, that this restaurant has been reviewed by the top users, and that the reviews aught to be taken seriously.

The purpose of ranking tags is to help users find the most relevant tags for a particular entity and facet. del.icio.us provides only a global ranking of tags based purely on the number of times each tag is used. If such a listing is used in type-ahead interface, a community will experience a rapid convergence to a general and boring set of tags. Instead, **TagRank** promotes tags used by the top-ranked users, and provides rankings localized to particular entities and facets. Figure 2 shows the usage pattern of tags. In particular, a few tags are used most of the time. Put another way, **TagRank** is designed to help a community *navigate the tail* and so tag in a much more meaningful way than simply reusing the globally popular tags. Even with **Tag Rank** care must be taken to assure that the community does become stuck on a sub-optimal folksonomy.

The most direct use of **UserRank** is the surfacing of community experts. With such a ranking place, users can find the most valuable text reviews. In addition, **UserRank** can also be localized to find community experts on certain topics, such as a particular restaurant, or a certain facet across all restaurants. Our **UserRank** algorithm must have the following two properties. First, it must be *incentive compatible*. That is, users should not be able to artificially boost their ranking. The second requirement is that the algorithm can reward users for recent contributions so that the ranking does not become stale.

5 UserRank

As the name suggests, **UserRank** is a variant of the well-known **PageRank** algorithm used by the Google [**?**] search engine. **PageRank** may be formulated as in [**?**]:

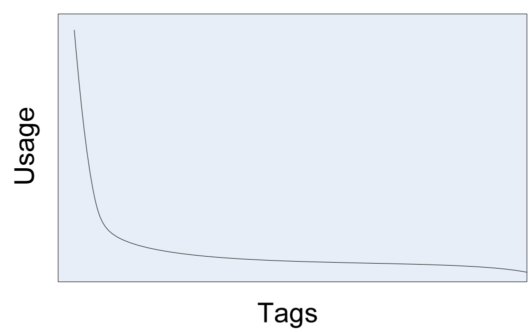

Figure 2: **The usage pattern of tags**

$$\forall i PageRank_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} PageRank_k(P_j)/N_{P_j} \qquad (1)$$

We start by defining pages and links for our pool of users. A page for user i denoted P_i is the set of all {facet,tag} pairs that user i has asserted. We define the order of such a pair for user i as $o(\{f,t\},P_i) =$ the number of users including i who have previously asserted the pair. $o(\{f,t\},P_i)$ is undefined if $o(\{f,t\} \ni P_i)$. We have a link from P_j to P_i if some pair $\{f,t\}$ exists in P_j and P_i, and $o(\{f,t\},P_j) > o(\{f,t\},P_i)$. That is, P_j links to P_i if j asserts a pair that i has already asserted. The set of all such P_j is B_{P_i}. The number our outgoing links from P_j is N_j. The strength of a link denoted $s(P_j,P_i)$ is used to scale j's contribution to i's **UserRank** so that

$$\forall i UserRank_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} s(P_j, P_i) UserRank_k(P_j) / S_{P_j} \quad (2)$$

Notice the location of the stength function in Equation 2. In addition, the contribution from P_j is scaled not by N_{P_j} as in Equation 1 but by S_{P_j}, the total strength of links eminating from P_j.

We now present two choices for the strength function $s(P_j, P_i)$. Notice that to compute the set B_{P_i}, we need not explicitly compute $o(\{f,t\}, P_j)$ and $o(\{f,t\}, P_i)$. We can answer $o(\{f,t\}, P_j) > o(\{f,t\}, P_i)$ simply by comparing the timestamps when the pairs were asserted. Thus, one version of the strength function is simply a constant multiple of the number of i's assertions that j agree's with. This function is stated formally in Equation 3.

Consider the implications of this approach for three users u, v, and w who all tag Bombay Club's decor with 'Eastern', in that order. Equation 3 implies that $s(v,u) = s(w,u) = s(w,v) = 1$. However, this equality is not desirable. Even though u benefits from links from both v and w, u should gain more per link because u made the assertion first. In addition, Both $s(w,u)$ and $s(w,v)$ should be penalized to avoid the "bandwagon" effect. That is, a repeated assertion should provide diminishing strength to the link it applies to. We can use the pair orders in the link to scale the strength appropriately.

Equation 4 defines a strength function that uses the orders of each element of the link. f is some diminishing function (monotonic in all derivatives) that assures repeated uses offer less agreement. g is some diminishing function (monotonic in all derivatives) that strengthens links when the order of the destination pair is low; early taggers receive more benefit. In practice, maintaining the orders of links is expensive and computing them on the fly while computing the **UserRank** on the whole graph is even worse. In addition, it is not obvious which approach provides more realistic results. Therefore, in our first implementation we will use the simple strength function in Equation 3.

Figure 3 shows a simple example of **UserRank** for some del.icio.us users. From other contributions, User A has rank 10. Based on the mlb.com tags, User A distributes $2/3$ of his Rank to user B and $1/3$ of his Rank to user C.

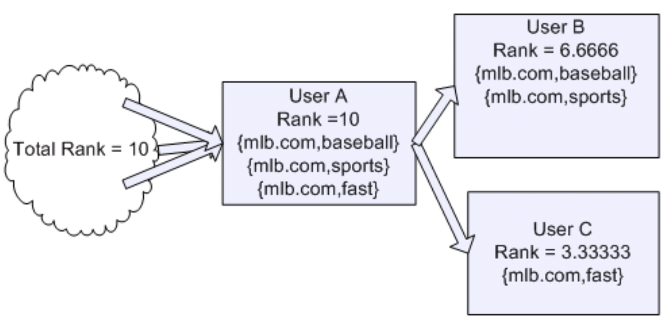

Figure 3: **A simple instance of UserRank**

Finally, we discuss some incentive compatibility concerns of **UserRank**. That is, we must ensure that users cannot take any action to artificially increase her rating. One deviant strategy would be to assert as many tags as possible so that any assertions made by other users have a high probability of improving the ranking. We can counteract this by imposing a small penalty of $\frac{\beta}{size(P_i)}$ to i's rank. We must select β to penalize haphazard tagging but that the penalty is easily overcome if a significant portion of user's page is followed. Equation 5 shows how the penalty fits in to the **UserRank**

$$\forall i Rank_{k+1}(P_i) = \frac{\beta}{size(P_i)} \sum_{P_j \in B_{P_i}} s(P_j, P_i) Rank_k(P_j) / S_{P_j} \quad (5)$$

A second problem is that users could create a fake user to assert agreeing tags. However, other users would have to agree with this user in order for it to have any rank to distribute. Even in these early stages, it seems that most strategies for rank inflation are provably doomed to failure.

5.1 TagRank

Given **UserRank** we may easily derive an algorithm for ranking tags. Defined formally in Equation 6, the rank of a tag is simply the sum of the **UserRank** of the tagger over all instances of the tag. This approach is based on the belief that the most relevant tags are those used by the best users.

$$\forall i TagRank(T_i) = \sum_{t \in inst(T_i)} UserRank(user(t)) \quad (6)$$

5.2 Localized UserRank and TagRank

Thus far, we have provided global ranking alogrithms for users and tags. As described in the System Design Goals section, community systems require rankings for a specific locality in a community. In one sense, these different rankings may be computed easily using **UserRank** and **TagRank**. For example, if we want to rank users in the context of The Bombay Club, we only consider facets of the Bombay Club or its dishes. If we want to find the best tags for describing 'decor', we would only sum the uses of tags on that particular facet across all restaurants. The difficulty with this approach is that a new ranking must be computed for each type of request. As the community grows, these rankings may take hours to compute. A focus of our future work will be the efficient computation and indexing of these localized rankings so that requests may be serviced fast enough for use in our user interface.

5.3 Age considerations

As mentioned in the System Design Goals section, the age of rankings is a serious concern. We want to assure that highly ranked users may be dethrowned by newcomers who are agreed with more recently. One possibile solution would be to modify the strength function to consider the age of the link. Links containing older agreements should be weaker.

Premature convergence on a sub-optimal folksonomy is possibly the greatest concern of the system. Without some assistance, new tags will never be surfaced in the user interface, and so will never be used to gain rank. To break this vicious cycle, we can give recent tags a boost in **TagRank** to give users a change to use them.

6 A Delicious Experiment

This section details an experiment with online communities and folksonomy development using our ranking methods. The purpose is to compare the traditional ranking methods used in folksonomy development in existing online communities with **UserRank** and its derivatives. We refer to our ranking methods as **URLRank**, **TagRank** and **UserRank** and the traditional methods as **URLCount**, **TagCount** and **UserCount** since current approaches are based primarily on usage totals.

We performed this experiment on a data set obtained from http://del.icio.us (henceforth referred to as Delicious). Delicious is a social bookmark manager where users can add their personal bookmarks, categorize them using keywords and share their bookmark collections with others in the community. The data was collected over a period of 70 hours through the RSS feed made available by Delicious. The data set encompassed 12,175 users, 54,561 bookmarks and 130,391

$$s(P_j, P_i) = \alpha\#(\{f,t\} \mid o(\{f,t\}, P_j) > o(\{f,t\}, P_i) \wedge \{f,t\} \in P_i \wedge \{f,t\} \in P_j) \quad (3)$$

$$s(P_j, P_i) = \sum_{(\{f,t\} \mid o(\{f,t\},P_j) > o(\{f,t\},P_i) \wedge \{f,t\} \in P_j \wedge \{f,t\} \in P_i)} f(o(\{f,t\}, P_j))g(o(\{f,t\}, P_i)) \quad (4)$$

taggings. We would like to note that since we did not have details of the ranking methodology used by Delicious, we used the ranking results generated by Scuttle. Scuttle is an open source implementation of a social bookmark manager that emulates Delicious very closely. We then proceeded to compare the ranking results for URLs, tags and users in both systems.

6.1 URLCount vs. URLRank

We first examined the top 25 URLs in both systems and found that not only the top URL matched but that there also was a 50% intersection in both rankings. At first, these two observations were very encouraging because it meant that totally different ranking algorithms on the same set yielded similar results. However, it brought the question whether our method was going to yield any significantly different or better results that the traditional methods.

As we looked closer into both top URL rankings we started to notice some interesting discrepancies such as the #2 URL in **URLCount** was only #21 in **URLRank**. Table 1 shows **URLRank** compared with **URLCount**. Many of these URLs have a very low count, but they were considered important because users using these URLs in their bookmark collections had a greater combined ranking than those using the URLs with a much larger usage. Clearly supporting our notion that usage count might not be the most accurate way of finding community consensus in online communities.

6.2 TagCount vs. TagRank

We then proceeded to compare the tag rankings by both methods. Tag rankings are extremely important because in some sense the highest ranked tags are the folksonomy. Technorati, Flickr and Delicious users have helped create these sites' respective folksonomies. For example, 'apple' according to the Delicous' folksonomy refers to URLs about the Apple company and its products as opposed to the forbidden fruit. After analyzing the top tags over time in both systems we discovered even more profound information regarding folksonomy development. Looking at Figure 4 one may see that not much variation occurs in the tag ranking, especially when contrasted to Figure 5. We believe this shows both how hard and easy it is for the folksonomy to evolve under the count methodology. It is hard because for something to go up a single position, it would take a relatively large shift in the community to stop using a very popular tag and increase another twofold for the lesser known to move up positions. But folksonomy changes are also easy to achieve, because it would be a simple matter of persistence from any given user to make one of their tags top ranked. On the other hand, **TagRank** easily derails those types of users because, its ranking not only depends on usage, but on **UserRank**. For example, the 'javascript' tag which made it to the top 10 in **TagRank** was used 726 times beating 'mac' tag in **TagCount** that was used 1015 times, that's almost a 40% difference. Another topic worth investigating further would be that of folksonomy convergence. As you can see from Figures **??** and 5 they both seem to converge on roughly on the same top 10 tags but at very different rates. It's still unanswered whether the **TagRank** convergence is stable in the long term. It can be said that **TagCount** tag convergence does not only stabilize quickly, but it is also a stable long term convergence, because according to Adam Mathes [**?**] Delicious most popular tags have not changed much since November 2004. Although, this also might be the case for **TagRank**, we have begun exploring the idea of

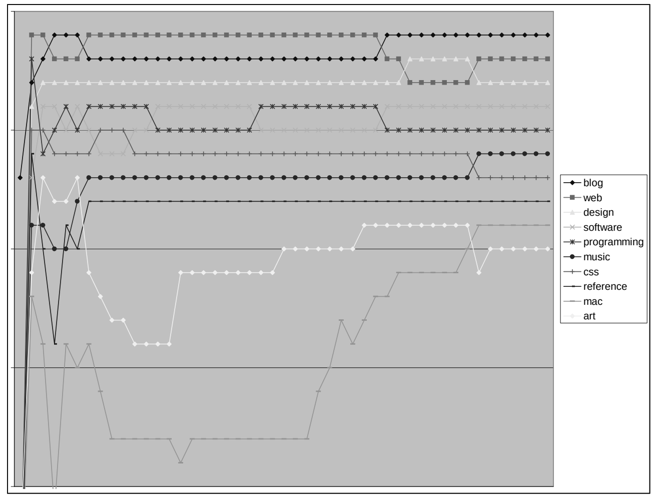

Figure 4: **Top Tags according to TagCount.**

extending **TagRank**'s strength function to include added strength for recent up and coming tags. The intent is to avoid a local maximum where the top tags never change and new ones do not have a chance of ever becoming mainstream in the folksonomy. This effort would be more easily understood as *navigating the tail* in what some have called a 'Broad Folksonomy' [4].

6.3 UserCount vs. UserRank

Our previous comparisons have thus far been satisfactory because two distinct methods arrive at similar, yet subtly different results. Suprisingly, however, the comparison of user rankings (the foundation for the other rankings) yielded profound diffrences. After comparing the user ranking results, we found no intersections in the top 10 users and after looking further, we only found a single intersection in the top 100. Even more intriguing was the numbers found when comparing the top users for each method. The **UserCount** top user had added 309 bookmarks to her collection with a total of 643 tags; this was definitely some busy user. On the other hand, **UserRank**'s top ranked user had only added a single bookmark during our experiment timeframe with just four tags. But the first thing we noticed is that this single bookmark was the top bookmark in our URL ranking comparison. The URL was tagged 749 times with at least one of the four tags used earlier by our **TagRank**'s top user. This was a surprising contrast to the almost non-existant following of **TagCount** 's top user. More importantly to support our consensus hypothesis, we found out that the top URL had been already used in prior taggings, but the tags were never really followed as opposed to **UserRank**'s top user's tags.

7 Conclusion

Online communities are now more than never taking an active role in the complex set of tasks associated with information management. Folksonomies and tags have recently been regarded as key paradigms for making the tasks more appealing to broader audiences. People with

Table 1: Top 25 URLs according to **TagRank**

TagRank	TagCount	Count	URL
1	1	766	http://www.alvit.de/vf/en/essential-bookmarks-for-webdesigners-and-web- developers.html
2	-	10	http://www.openlaszlo.org/
3	3	318	http://searchsmb.techtarget.com/general/0,295582,sid44 _gci1076646,00.html
4	-	9	http://www.hoogerbrugge.com/
5	23	161	http://livsey.org/experiments/hoverhelp/
6	6	274	http://www.cs.virginia.edu/ robins/YouAndYourResearch.html
7	16	191	http://mboffin.com/post.aspx?id=1619
8	7	272	http://www.456bereastreet.com/archive/200504/fixed_or_fluid _width_elastic/
9	-	4	http://www.despair.com/demotivators/index.html
10	9	257	http://www.pui.ch/phred/archives/2005/04/tags-database-schemas.html
11	-	143	http://kupu.oscom.org/
12	24	153	http://www.urltrends.com/
13	-	6	http://kostas.yoink.com/past/2005/03/07/magnatune _improvements_wishlist/
14	11	218	http://www.ilovejackdaniels.com/php/php-cheat-sheet/
15	4	313	http://cssc.darkriftstudios.com/
16	-	3	http://www.vladstudio.com/wallpapers/
17	8	260	http://www.wirelessnetmanager.com/
18	10	250	http://www.cloudynights.com/item.php?item_id=1052
19	-	5	http://mpto.unistudios.com/xml/
20	-	146	http://www.squarefree.com/shell/
21	2	354	http://www.torrenttyphoon.com/
22	-	91	http://richard.jones.name/google-hacks/gmail-smime/gmail-smime.html
23	15	196	http://sqlzoo.net/
24	-	3	http://www.elephant-talk.com/
25	-	4	http://www.javalobby.org/articles/jython/

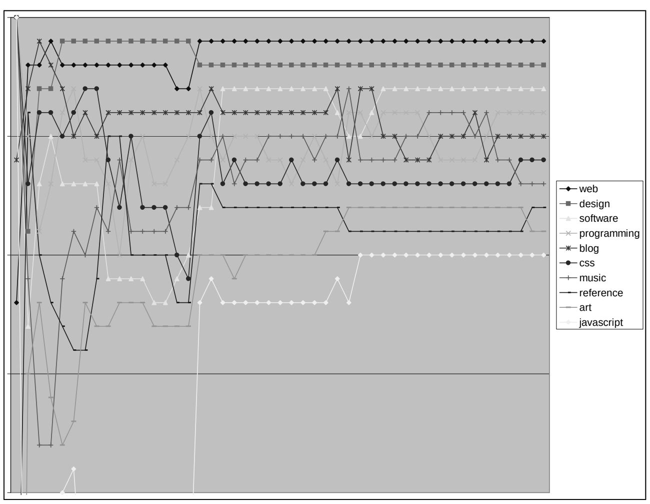

Figure 5: **Top Tags according to TagRank.**

the help of other people can be empowered to classify information for more than selfish reasons to provide others with the benefits of their work. However, this task is not to be taken lightly and very careful scrutiny must be given to the methodologies used in social software to achieve such classifications. It can be seen that people are already doing their part in the collective task by providing their tagging abilities in many different applications and information contexts. Therefore, it is our job to make sure that the ranking methodologies in these communities are worthy of trust and accuracy for both community members and external consumers of the information. Currently, we have more positive than negative references to tagging in online communities, but it will not be long before both advertisers and spaggers will start to dilute the collective efforts of online communities. We believe that it is early enough to make sure that folksonomies and tagging will be a part of online community development for future generations. Ranking should be of highest priority for any community interested in providing information access and archival features. We have shown how the consensus obtained through our series of ranking algorithms can be applied to virtually any online community and reflect more closely the communities conceptual model of their information.

References

[1] Folksonomy. http://en.wikipedia.org/wiki/Folksonomy.

[2] Ross Mayfield. Technorati takes tag global. http://www.corante.com/many/archives/2005/01/13/technorati_takes_tags_global.php, January 2005.

[3] Adam Mathes. Folksonomies - cooperative classification and communication through shared metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html, December 2004.

[4] Thomas Vander Wal. Explaining and showing broad and narrow folksonomies. http://www.vanderwal.net/random/entrysel.php?blog=1635, February 2005.