Using Co-occurence of Tags and Resources to Identify Spammers

Ralf Krestel and Ling Chen

L3S Research Center
Universität Hannover, Germany
krestel|lchen@L3S.de

Abstract. Today, more and more social networking websites support *collaborative tagging*, which allows users to annotate resources (e.g., video clips, blog posts, and bookmarks) on the web. Due to its increasing popularity, however, spammers started to target this new type of service and generate misleading tags either to increase the visibility of some resources or simply to confuse users. Consequently, the performance of applications built upon tag data, such as recovery and discovery of web resources, can be limited. In this paper, we propose an algorithm to identify spammers from the collaborating systems by employing a *spam score propagating* technique. The three dimensional relationship among users, tags and web resources is firstly represented by a graph structure. A set of seed nodes, where each node represents a user, are then selected and assigned values to indicate whether the corresponding users are spammers or not. The initial values are propagated through the graph to infer the status of the remaining users. Our experimental results demonstrate the effectiveness of this approach in identify tag spammers.

1 Introduction

With the recent rise of Web 2.0 technologies, many social media applications like *Flickr*, *Del.ici.ous*, and *Last.fm* provide features which allow users to assign tags [1] to a piece of information such as a picture, blog entry, video clip etc. Web users from different backgrounds tag (annotate) resources on the Web at an incredible speed, which results in large volume of tag data obtainable from the Web today. The hidden value of tag data has been explored in many applications. For example, Tso-Sutter et al [2] incorporated tags into collaborative filtering algorithms to enhance recommendation accuracy. In [3], the authors discussed using tags to lighten the limitation of the amount and quality of anchor text to improve enterprise search. The usage of tags in Web search has also been investigated in Bao et al [4].

One notable reason which supports the increasing popularity of collaborative tagging is that users are permitted to enter tags without any constraints. Consequently, spammers can easily take advantage of this new service to generate

misleading tags to increase the visibility of some resources or simply to confuse users. *Therefore, identifying spammers from collaborating systems is an important problem so that top-quality tag data can be generated by removing those supplied by spammers.* Some research effort has been exerted to target this problem. For example, Koutrika et al [5] proposed to combat tag spam by ranking the results returned from a query tag, based on the co-occurrence frequency between the tag and each resource. Their approach is specially designed for tag based search, while our research objective is more general so that the results can be used in not only tag based search but also other applications of collaborative systems.

In our approach, we firstly construct a graph which models users as nodes and three types of relationship between users as edges. Particularly, we consider the following types of relationship between users: common tags supplied by users, common resources annotated by users and common tag-resource pairs used by users. We then select a set of seed nodes whose corresponding users are manually assessed as spammers or not. The identity of the remaining nodes/users are computed by propagating the status of seed nodes through the graph. The effectiveness of our approach is demonstrated on the bibsonomy data set[1].

The rest of this paper is organized as follows. We discuss the background knowledge by reviewing related work in Section 2. In Section 3, we describe the approach which propagates the identity of seed users through the graph. The evaluation results conducted on the bibsonomy data set are presented and analyzed in Section 4. Finally, Section 5 concludes this paper with some summary remarks and future work discussions.

2 Related Work

In this section, we review related work in two areas, collaborative tagging systems and spam detection.

A collaborative tagging system allows users of a web site to freely attach to a particular resource arbitrary tags which, in the opinion of the user, are somehow associated with the resource in question. The commonly noted structure of collaborative filtering systems is a tripartite model consisting of users, tags and resources. This model is developed as a theoretical extension of the bipartite structure of ontologies with an added "social dimension" in [6]. The dynamics of collaborative systems are examined in [7] using the tag data at the bookmarking site Del.ici.ous. According to this work, tag distributions tend to stabilize over time. Halpin et al. confirm these results in [8] and show additionally that tags follow a power law distribution. Considering the structure and stable dynamics of collaborative tagging systems, it seems likely that tag data would be a reliable source of semantic information reflecting the cultural consensus of a particular system's users. As a result, various applications of tag data have been researched. Mika [6] investigates the automatic extraction of ontological relationships from

[1] http://www.kde.cs.uni-kassel.de/ws/rsdc08/dataset.html

tag data and proposes the use of such emergent ontologies to improve currently existing ontologies which are less capable of responding to ontological evolution. Dmitriev et al. [3] explore the use of "annotations" for enterprise search to compensate for the lack of sufficient anchor text in intranet environments. In [4], tag data is exploited for the purpose of web search through the use of two tag based algorithms: one exploiting similarity between tag data and search queries, and the other utilizes tagging frequencies to determine the quality of web pages. Tso et al [2] incorporate the tag data into the collaborative filtering systems. Berendt and Hanser [9] demonstrate the benefits of using tag data for weblog classification by treating it as content instead of meta data. For searching and ranking within tagging systems, A. Hotho et al [10] propose the *FolkRank* algorithm which extends the seminal PageRank approach. In particular, they model the structure of the folksonomy as a graph, where nodes represent users, tags and resources, and edges represent the assignment relationship between users and tags, users and resources, tags and resources.

Everywhere in the internet where information is exchanged, malicious individuals try to take advantage of the information exchange structure and use it for their own benefit. The largest amount of spam and historically the first field where spam was generated is the electronic communication system (e-mail). Afterwards, various internet applications were attacked by spammers such as search engine spam, blog spam, wiki spam etc, which triggered numerous research efforts in spam combating. For example, TrustRank [11] separates spam pages from non-spam pages based on the intuition that trustworthy pages usually link to also trustworthy pages and so on. They select a seed set of highly trusted pages first and then propagate the trust score of seed pages by following the links from these pages through the Web. A survey of approaches fighting spam on social web sites can be found in [12]. Comparing to spam detection from other web applications, studies on detecting spam from collaborative tagging systems are very limited. Koutrika et al [5] propose to combat spam in the particular situation when users query for resources annotated with certain tags. Their method ranks a resource higher if more users annotated it with the queried tags, based on the assumption that tag spam may not be used by the majority. As mentioned before, our work is different in the way that our approach is not designed for a particular application. Consequently, the output of our algorithm — a set of identified tag spammers — can be used by any application based on tags. Xu et al [13] assign authority scores to users, and measure the goodness of each tag with respect to a resource by the sum of the authority scores of all users who have tagged the resource with the tag. Then, the authority scores of users are computed via an iterative algorithm similar to HITs [14]. Contrasting to their approach which iteratively computes authority scores for users and tag-resource pairs, we iteratively update scores for users only. Moreover, our approach is more flexible in the way that multiple relationship, such as co-tag, co-resource and co-tag-resource, can be taken into account, rather than considering only the tag-resource pairs shared by users.

3 Finding Malicious Users

Identifying malicious users (spammers) in a tagging environment with thousands of participants and millions of tag assignments can be done by exploiting the *wisdom of the crowds* [15]. If many known spammers use a certain tag for a certain resource, it might indicate that other users having the same tag assignment are also spammers. In our approach we use an algorithm similar to TrustRank [11] to propagate a spammer score through a graph with each node representing a user. As in TrustRank, we need a set of seed nodes which were manually assessed. For the competition, the training data was used as the seed set.

3.1 Problem Specification

Let $\mathcal{U}$ be a set of users of a collaborating system, $\mathcal{T}$ be a set of tags, and $\mathcal{R}$ be a set of resources. We define the functions $getT(u)$ and $getR(u)$ to retrieve the set of tags and resources assigned by user u respectively. In addition, we define the function $getTR(u)$ to return the set of tag-resource pairs used by user u. For example, $getTR(u) = \{t_m r_n\}$ indicates that the user u assigned the tag t_m to the resource r_n.

Our goal is to find a function $S(u_i), u_i \in \mathcal{U}$, which assigns a score to each user u_i such that the higher the value of $S(u_i)$, the higher the probability that u_i is a spammer. The value of $S(u_i)$ ranges in $[-1, 1]$ (the reason why negative values are involved will be explained later in Section **??**).

3.2 Tagging System Model

Given a set of data including users $\mathcal{U}$, tags $\mathcal{T}$ and resources $\mathcal{R}$, we model the data as a bidirected weighted graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is a set of vertices with each $v \in \mathcal{V}$ represents a $u \in \mathcal{U}$. $\mathcal{E}$ is a set of edges such that each edge (v_i, v_j) indicates that the two corresponding users u_i and u_j used at least one common tag or resource. That is, $|getR(u_i) \cap getR(u_j) \cup getT(u_i) \cap getT(u_j)| \geq 1$.

Additionally, we associate a weight to each edge so that the weight of an edge depends on the number of shared tags and resources of the end nodes of the edge: $W(v_i, v_j) = W(u_i, u_j) = (|getT(u_i) \cap getT(u_j)| \times W_t) + (|getR(u_i) \cap getR(u_j)| \times W_r) + (|getTR(u_i) \cap getTR(u_j)| \times W_{tr})$. $W_i, i \in \{r, t, tr\}$ represent static weighting factors to pay tribute to the different degrees of proximity depending on whether they are sharing the same tags t, resources r or even the same tag assignments tr.

In Figure 1 (a), we present a very simple tagging scenario: Suppose we have three users $\mathcal{U} = \{u_1, u_2, u_3\}$, three different tags $\mathcal{T} = \{t_1, t_2, t_3\}$ and two resources $\mathcal{R} = \{r_1, r_2\}$. Each user has annotated the resources with certain tags. For example, the leftmost link in Figure 1 (a) indicates that both users u_1 and u_2 have supplied the tag t_1 with the resource r_1. Based on the tag assignments in this figure, the corresponding data model can be created as Figure 1 (b). Three nodes, representing the three users, are connected with each other according

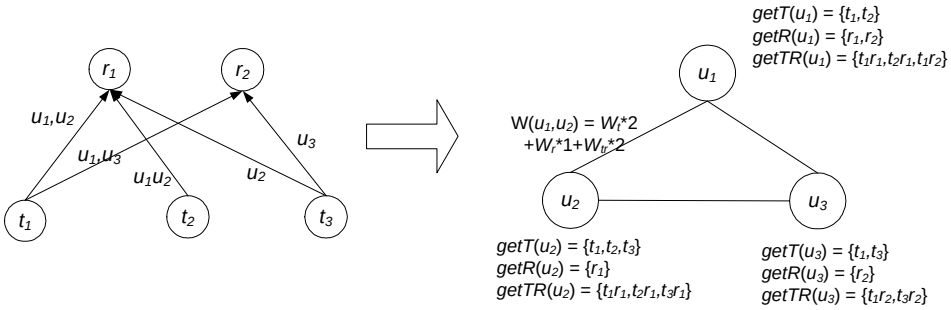

Fig. 1. A tagging scenario and its data model

to common tags/resources/tag-resources pairs. The results of the three functions related to a user, $getT(u_i), getR(u_i), getTR(u_i)$, are shown in the figure as well. Then, based on the tags, resources, and tag-resources used by a user, the weight of an edge connecting two users can be computed. For example, as shown in the figure, the weight of the edge between u_1 and u_3 is calculated as $W(u_1, u_2) = W_t * 2 + W_r * 1 + W_{tr} * 2$, since the two users shared two tags, one resource and two tag-resource pairs.

Based on this graph model, we introduce a right stochastic transition matrix T, which is defined as:

$$T(i,j) = \begin{cases} 0 & \text{if } (v_i, v_j) \notin \mathcal{E} \\ \frac{W(v_i,v_j)}{\sum_{v_k \in \mathcal{V}} W(v_i,v_k)} & \text{if } (v_i, v_j) \in \mathcal{E} \end{cases}$$

Suppose W_t, W_r and W_{tr} are set as 1. Figure 2 shows the adjacency matrix and the transition matrix for the example in Figure 1. Note that, the adjacency matrix is symmetric since the graph model is bidirected, while the transition matrix is asymmetric.

	v_1	v_2	v_3
v_1		5	3
v_2	5		2
v_3	3	2	

$$T = \begin{pmatrix} 0 & \frac{5}{8} & \frac{3}{8} \\ \frac{5}{7} & 0 & \frac{2}{7} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{pmatrix}$$

Fig. 2. Adjacency (left) and transition (right) matrixes of the example in Figure 1.

3.3 Spammer Score Propagation

In our approach, the spammer score for each user, $S(u)$, is computed similarly to TrustRank [11], which itself is based on PageRank [16]. The TrustRank employs the formula as follows:

$$\text{t-rank}_{i+1} = \alpha \cdot T \cdot \text{t-rank}_i + (1 - \alpha) \cdot \mathsf{d}, \tag{1}$$

with transition matrix T, a weighting factor α and the manually assessed seed vector d. We use this formula to propagate initial spammer scores of seed users through the graph. In addition to TrustRank which propagates only trust information, we adopt the distrust propagation idea described in [17] to allow the propagation of scores for not only good users but also explicitly bad users (spammers). Consequently, we extend the manual seed set assessment to include both good users and spammers. We populate the initial vector d with:

$$\mathsf{d}(u_i) = \begin{cases} \mathrm{O}(u_i) & \text{if } u_i \in SEED \\ 0 & \text{if } u_i \notin SEED \end{cases} \tag{2}$$

where $\mathrm{O}(u_i) \in \{-1, 0, 1\}$ is the oracle function which assigns initial score 1 to non-spammers, -1 to spammers and 0 to the rest. $SEED \subseteq \mathcal{U}$ is a set of seed nodes, which for the competition was the provided set users in the training data.

Consider the running example shown in Figures 1 and 2, the results of our approach (i.e. spammer score for each user) after 10 iterations are shown in Figure 3, where v_1 and v_3 are selected as seed nodes and the decay factor α is set as 0.5.

$$\text{spammer-score}_{i+1} = 0.5 \cdot \begin{pmatrix} 0 & \frac{5}{8} & \frac{3}{8} \\ \frac{5}{7} & 0 & \frac{2}{7} \\ \frac{3}{5} & \frac{2}{5} & 0 \end{pmatrix} \cdot \text{spammer-score}_i + (1 - 0.5) \cdot \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$i = 10$	v_1	v_2	v_3
spammer-score(v_x)	0.38621816	-0.42241633	0.03619808

Fig. 3. Spammer score computation and results for the example in Figure 1.

4 Evaluation

Evaluation was performed on the competition's data set. Due to the time constrains, we were not able to do extensive evaluation, investigating the results for different parameter settings, or do an in-depth analysis of the submitted results. Since we were only allowed to submit one run, we will try to provide more results for the final paper.

4.1 Data Set

The data set comes from bibsonomy[2] and was manually extended with spammer information[3]. Table 1 shows the properties of the provided training and test data set. The data consists mainly of tagged bookmarks rather than tagged bibtex entries (2%) and bookmark tag assignments are more likely to be spam compared to bibtex tag assignments (90.229% vs. 0.264% in the training data).

Table 1. Training and test data sets

	Training Data Set	**Test Data Set**
Users	31,715	7,205
Spammers	29,248	7,034
Tag Assignments	14,074,725	2,743,743
Tag Assignments from Spammers	13,258,759	2,612,634

4.2 Results

We evaluated different configurations. Firstly, we only considered co-occurence of tag-resource pairs between users. That means, only if two users assigned the same tag to a certain resource we created an edge in the graph for these two users. Secondly, we added resource co-occurence edges to the graph. Still ongoing are evaluations for other configurations like including tag co-occurence between users. The confusion matrices for the first two configurations can be seen in Table 2.

Table 2. Confusion matrices for different configurations

Only Tag-Resource Pairs Co-occurence			
True Positives:	**6085**	True Negatives	**13**
False Positives:	**158**	False Neagatives	**949**
Tag-Resource Pairs and Resources Co-occurence			
True Positives:	**6202**	True Negatives	**2**
False Positives:	**169**	False Neagatives	**832**

Table 3 shows *accuracy* and *ROC AUC* values. Since we only assigned boolean values to users the ROC curve is not very interesting and we ommit it here.

[2] `http://www.bibsonomy.org`
[3] http://www.kde.cs.uni-kassel.de/ws/rsdc08/dataset.html

Table 3. Accuracy and ROC AUC for different configurations

Strategy	Accuracy	ROC AUC
Tag-Resource Pairs	86.11%	0.4469
T-R Pairs, Resources	84.64%	0.4707

5 Conclusions and Future Work

In this paper, we mainly describe an approach to identify spammers from collaborative tagging systems. The basic idea follows the seminal PageRank approach. The specific feature which distinguishes our approach from existing work is the data structure we employ. In particular, we explicitly model users of collaborative systems as nodes in a graph, since our objective is to detect suspicious spammers. An edge is then created between two users if they co-used a resource, a tag and/or a tag-resource pair. After manually assessing a set of seed users, the scores indicating whether they are spammers or not are propagated through the graph. The intuitive is that nonspammers may annotate resources with similar tags, while spammers may have similar interests in particular resources and/or tags. Consequently, as another feature of our approach, we propagate the scores of not only nonspammers but also spammers. The experimental results on the challenge data demonstrate the effectiveness of our approach.

For future work we want to combine our link-based algorithm with a content-based approach. The benefits could be twofold: Firstly we could use the content-based approach to automatically generate the seed set, and secondly we could adjust the weights for propagation based on the content analysis. To improve our link-based algorithm we try to find more connections between users to minimize the number of unreachable partitions in the graph. The assignments of probabilities to users instead of boolean values could also comprise some potential for improvement. For real world applications, the question of seed set selection poses another interesting task which needs to be solved.

6 Acknowledgements

This work is supported by the EU project IST 45035 - Platform for searcH of Audiovisual Resources across Online Spaces (PHAROS).

References

1. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Ht06, tagging paper, taxonomy, flickr, academic article, to read. In Wiil, U.K., Nürnberg, P.J., Rubart, J., eds.: Hypertext, ACM (2006) 31–40
2. Tso-Sutter, K.H.L., Marinho, L.B., Schmidt-Thieme, L.: Tag-aware recommender systems by fusion of collaborative filtering algorithms. In Wainwright, R.L., Haddad, H., eds.: SAC, ACM (2008) 1995–1999

3. Dmitriev, P.A., Eiron, N., Fontoura, M., Shekita, E.J.: Using annotations in enterprise search. In Carr, L., Roure, D.D., Iyengar, A., Goble, C.A., Dahlin, M., eds.: WWW, ACM (2006) 811–817
4. Bao, S., Xue, G.R., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing web search using social annotations. [18] 501–510
5. Koutrika, G., Effendi, F., Gyöngyi, Z., Heymann, P., Garcia-Molina, H.: Combating spam in tagging systems. In: AIRWeb. (2007)
6. Mika, P.: Ontologies are us: A unified model of social networks and semantics. In Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A., eds.: International Semantic Web Conference. Volume 3729 of Lecture Notes in Computer Science., Springer (2005) 522–536
7. Golder, S.A., Huberman, B.A.: The structure of collaborative tagging systems. CoRR **abs/cs/0508082** (2005)
8. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. [18] 211–220
9. Berendt, B., Hanser, C.: Tags are not metadata, but just more content - to some people. In: ICWSM. (2007)
10. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In Sure, Y., Domingue, J., eds.: ESWC. Volume 4011 of Lecture Notes in Computer Science., Springer (2006) 411–426
11. Gyöngyi, Z., Garcia-Molina, H., Pedersen, J.O.: Combating web spam with trustrank. In Nascimento, M.A., Özsu, M.T., Kossmann, D., Miller, R.J., Blakeley, J.A., Schiefer, K.B., eds.: VLDB, Morgan Kaufmann (2004) 576–587
12. Heymann, P., Koutrika, G., Garcia-Molina, H.: Fighting spam on social web sites: A survey of approaches and future challenges. IEEE Internet Computing **11**(6) (2007) 36–45
13. Xu, Z., Fu, Y., Mao, J., Su, D.: Towards the semantic web: Collaborative tag suggestions. In: WWW2006: Proceedings of the Collaborative Web Tagging Workshop, Edinburgh, Scotland (2006)
14. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM **46**(5) (1999) 604–632
15. Surowiecki, J.: The Wisdom of Crowds. Anchor (August 2005)
16. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. techreport (1998)
17. Wu, B., Goel, V., Davison, B.D.: Propagating trust and distrust to demote web spam. In Finin, T., Kagal, L., Olmedilla, D., eds.: MTW. Volume 190 of CEUR Workshop Proceedings., CEUR-WS.org (2006)
18. Williamson, C.L., Zurko, M.E., Patel-Schneider, P.F., Shenoy, P.J., eds.: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007. In Williamson, C.L., Zurko, M.E., Patel-Schneider, P.F., Shenoy, P.J., eds.: WWW, ACM (2007)