Towards A Framework for Detecting Tag Spam

Eric Chang
Yale University

Abstract

Tagging system is the killer application of Web2.0 services that can help users to share, tag and discover those interesting resources. In recent years, as tagging systems are gaining in popularity, many studies also indicated that tagging services are susceptible to tag spam: misleading tags that are generated in order to increase the popularity of some resources or simply to confuse users in the networks. In this report, we summarize both the representative tagging systems and the existing methods of defending against tag spam in the application. Through analyzing the existing tagging systems, we will propose the conclusions and future works at the end of the report.

1. Introduction

The World Wide Web has had a tremendous impact on society and business in recent years by making information available in an instant and ubiquitous way. Especially, the rapidly growing popularity of Web 2.0 social applications originates in their ease of operating for inexperienced users and suitable mechanisms for supporting collaboration of shared good material, for instance, images in Flickr [17], bookmarks in del.icio.us [16], etc. In all of Web 2.0 social web sites, tagging system harvests high popularity mainly for its simplicity to create labels and add tags. In 2003 Joshua Schachter launched del.icio.us – the first social bookmarking service web site. Due to being inspired by the new opus, many kinds of social tagging sites have exploded recently, such as Flickr, Fringe, CiteULike [15], Blogmarks and Shadows. To illustrate the problem of spam tags in tagging system, we describe a general and ideal tagging system, similar as social bookmarking system, which can be seen as the archetype of the most tagging system for Web2.0 applications. The description will drive our discussion of spam-fighting techniques [30, 31, 32, 33, 34, 35].

General Social Tagging System

As users browse the Web, they often locally save visited URLs as bookmarks for later retrieval. Users of a tagging system, however, post them online, optionally with descriptive tags. They can then access these bookmarks from many locations as well as share them with others. They can also query the shared pool of bookmarks for certain statistics, such as the most popular and most recent sets of bookmarks. Fig. 1.1 shows the example of a tagging system.

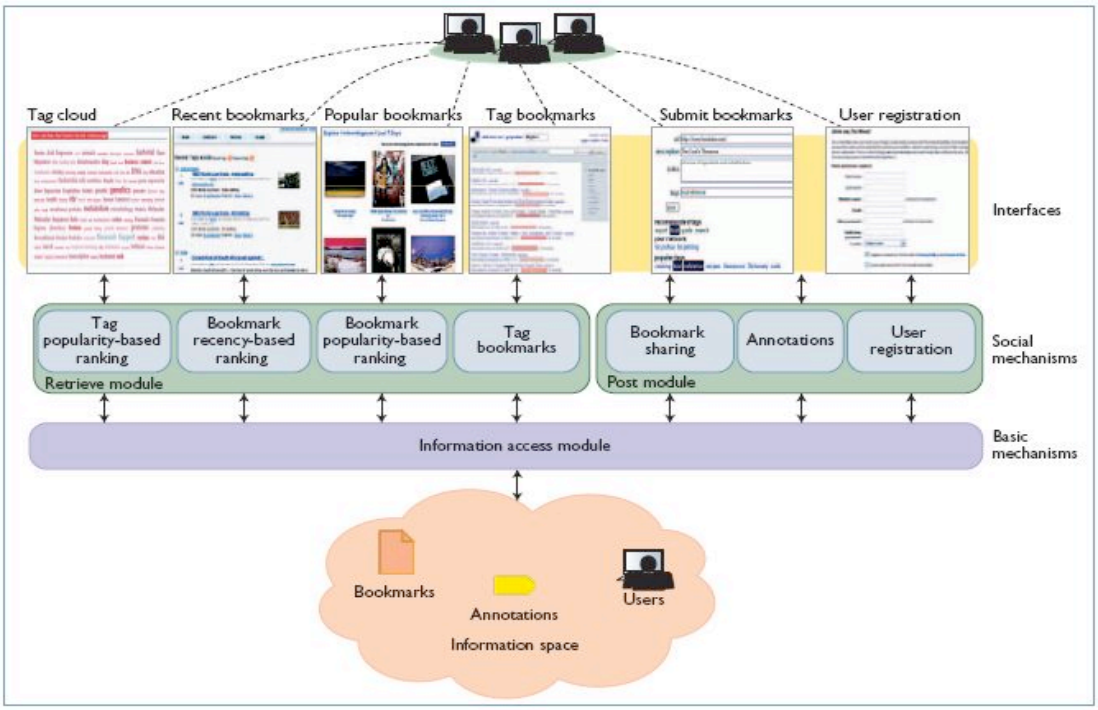

Fig. 1.1 The Example of Tagging System

In this system, there is only one interaction for content creation, consisting of a single action: a user may post a bookmark. A bookmark consists of a URL and optionally, a list of tags describing the URL. This information is then displayed through one or more interfaces:

- ***user bookmarks*** — for a given user *u*, a list of the most recent URLs that user posted;
- ***tag bookmarks*** — for a given tag *t*, a list of the most recently posted URLs annotated with target;
- ***recently list*** — a list of the most recent URLs posted to the system by any user;
- ***popular list*** — a list of the most popular URLs ordered as a function of number of postings and the amount of recent activity for each URL.

Previous works about tagging systems focused mostly on understanding tag usage and evolution [1, 2, 3, 4, 5, 10, 12, 13]. Also, the rank algorithms to enhance Web search were studied by [6, 7]. As we know, the scheme for defending web spam, such as create mislead content to result, is a well-known challenge for search engines. Not only search engines but also social tagging systems have become an attraction for malicious users. This report, we focus a problem for tagging systems that tag spam problem: misleading tags that are generated to make it more likely that some resources are seen by users, or generated simply to confuse users [11]. For instance, in a web page bookmarking system, malicious users may annotate a web page link to one famous scientist with the tag "farmer", so that users searching for that word will see the blog of scientist. Similarly, malicious users may annotate many photos with the tag "evil empire" so that this tag appears as one of the most popular tags.

Previous researches on malicious attacks of tagging systems mainly focused on spam tags in several studies, illustrating references [8, 9, 10, 11, 14] showed several solution schemes. [11] reported existing solution methods about attack of spam tags in tagging systems which

are detection-based, rank-based and prevention-based.

According to the analysis above, we present several models for user behaviors including good behaviors and a few mechanisms of spam attacks in Section 2. In Section 3, we will discuss the related works on combating spam tags and analyze both advantage and lacks of their works. Finally, the conclusions and future work will be presented in Section 4.

2. Analysis of Tagging System and User Behaviors

In this section, we will present two ideal models of behaviors of users that good and malicious. Before define several ideal models, a model for tagging system must be presented at first. As follow, Fig. 2.1 shows an ideal model for tagging system.

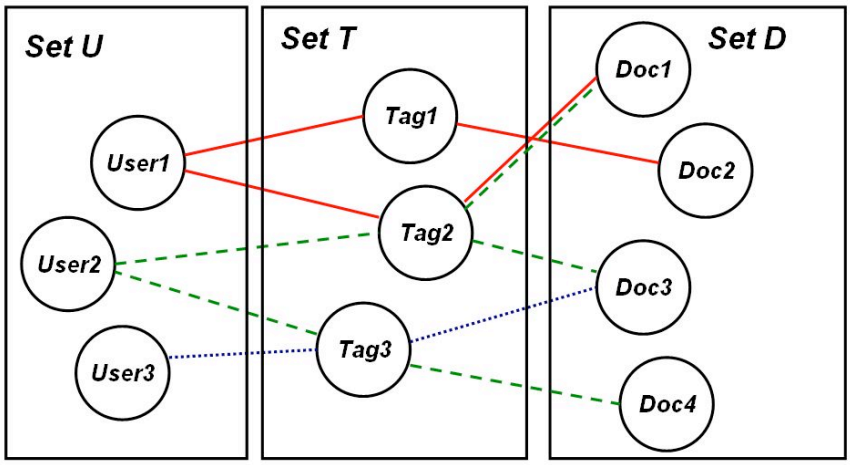

Fig. 2.1 An Ideal Model for Tagging System

As we can see in Fig. 2.1, a tagging system is comprised of a set U of users, who participates in the tagging system by annotated tags to resources, a set R of resources, and a set T of tags which constitute the tagging system annotations and a posting behavior set P which keeps the associations between tags, resources and users. For this ideal environment, we need not to know both the concrete content and the type of the resources nor the text associated with each tag. Relationship of three entities and behavior of users should be taken into consideration. For each resource $r \in R$ has a set $G(r) \subseteq T$ of tags that means correctly annotate the resource. For instance, a web page of a scientist, "person" and "scientist" both are the right tags, so they belong to the set $G(r)$. But other tags (e.g., "farmer", "athlete") are incorrect and are not in set $G(r)$, we call them spam tags, and they belong to the set $S(r) \subseteq T$. So we have $r \in R$, $G(r) \cup S(r) = T$, $G(r) \cap S(r) = \phi$ and if there is scene that $\exists r \in R$, $\forall t \in P(r)$, $t \in S(r)$ we call all the tags of this resource are spam. Now, we will present several ideal models of both good and malicious user behaviors.

Table 2.1 Ideal Models of Both Good and Malicious User Behaviors

Good User Ideal Model

define set $GU \subseteq U$ *as the set of good users in systems*

for each $u \in GU$ *do*

  for each $r \in R(u)$ *do*

    $t \in T$ *and* $t \in P(r)$ *and* $t \in G(r)$ *;*

Spam Attacker Ideal Model

define set $BU = U - GU$ *as the set of malicious users in systems*

for each $u \in BU$ *do*

  for each $r \in R(u)$ *do*

    $t \in T$ *and* $t \in P(r)$ *and* $t \in S(r)$ *;*

Collude Attacker Ideal Model

define set $CU \subseteq BU$ *as the set of collude malicious users in systems*

define set $CAR \subseteq R$ *as the set of collude attack resources*

for all $u \in CU$ *do*

  for each $r \in CAR$ *do*

    $t \in T$ *and* $t \in P(r)$ *and* $t \in S(r)$ *;*

Disguise Attacker Ideal Model

define set $DU \subseteq BU$ *as the set of disguise malicious users in systems*

define set $DAR \subseteq R$ *as the set of disguise attack resources*

for each $u \in DU$ *do*

  for each $r \in DAR$ *do*

    $t \in T$ *and* $t \in P(r)$ *and* $t \in S(r)$

    $t' \in T$ *and* $t' \in P(r)$ *and* $t' \in G(r)$

In Table 2.1, we present all behaviors of users in tagging systems. The purpose of existing works is to defend these malicious spam attacks using a effective and reliable metric. In this report, we summarize the whole existing schemes to combat the spam tags in tagging system for Web 2.0.

3. Related Work Discussed

As our known, existing research works of detection spam tags in tagging systems are [8, 9, 10, 11, 14, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]. As the same framework, meth-

ods to detect comment spam and spam blogs have been proposed by [11]. A first reference to the spam detection problem in folksonomies is given in [10]. References [9, 10, 11] are the first to deal with spam in tagging systems explicitly. The authors proposed anti-spam strategies for tagging systems and evaluated the models for different tagging behavior.

Coincidence Model

Reference [9] is the most famous study in the spam detection for tagging systems area. In this report, we mainly discuss the advantages and shortcomings of this target model. The contributions of the papers are:

- The authors define an ideal tagging system that they believe is useful for comparing query answering schemes and we model user tagging behavior.
- The authors proposed a variety of query schemes and moderator strategies to counter tag spam.
- The authors define a metric for quantifying the spam impact on search results.
- The authors compare the various schemes under different models for malicious user behaviors.

In the first section of [9], authors introduce some interested questions: How many malicious users can a tagging system tolerate before results significantly degrade? What types of tagging systems are more prone to spam? What is the impact of encouraging users to tag documents already tagged? What can be done to reduce the impact of malicious users? Is there a way to use correlations to identify misused tags?

The questions above are all difficult to answer and many similar problems are easy to be confused, so the authors introduced an ideal tagging system model where good users, malicious tags and malicious user behaviors and so on are well defined. So the study of authors is done in this ideal model. Given that the authors admit using an ideal model, their results will not be useful for predicting how any one particular tagging system may perform, but the results of research can insights into the relative merits of the various protection schemes their study.

The authors studied the effect of users' tagging behavior for single-tag queries. Given a query containing a tag *t*, the system returns a ranked list of objects that contain *t*. To illustrate the availability of their novel metric, the authors present three search schemes, former two schemes are existence and the last is new:

- *Boolean model*, which randomly orders objects that match the query tag *t* in the results.
- *Occurrence-based model*, which ranks each object *o* by counting the number of tuples in posting with object *o* and tag *t*.
- Coincidence-based model, which takes into account users' reliability. The system can measure reliability in this way: user *i* is considered more reliable than user *j* if *i*'s tags more often coincide with other users' tags compared to *j*'s tags.

To measure how reliable a user is, the authors define the coincidence factor *c(u)* of a user *u* and rank of tags as follows:

$$c(u) = \sum_{o,t:\exists P(u,o,t)} \sum_{\substack{ui \in U \\ ui \neq u}} |P(u_i, o, t)|$$

$$rank(d,t) = \frac{\sum_{\forall u \in users(d,t)} c(u)}{C}$$

We can show an example to illustrate the calculating of above formulas:

Table 3.1 A Simple Example

user	document	Tag
1	d_1	*a*
2	d_1	*a*
3	d_1	*b*
4	d_1	*b*
5	d_1	*b*
3	d_2	*a*
3	d_2	*c*
4	d_2	*c*

In above table, we assume that correct tags for document d_1 and d_2 belong to the sets {*b, c*} and {*a, c*}, respectively. Different users may assign the same tag to the same document. For instance, *user3, 4* and *5* have all assigned tag *b* to document d_1. According to the coincidence-based metric, we can get: *c(1) = 1, c(2) = 1, c(3) = 3, c(4) = 3* and *c(5) = 2*.

$rank(d_1, a) = (c(1) + c(2)) / C = 2/10$

$rank(d_1, b) = (c(3) + c(4) + c(5)) /C = 8/10$

$rank(d_2, a) = c(3) / C = 3/10$

$rank(d_2, c) = (c(3) + c(4)) / C = 6/10$

In words, a document's importance with respect to a tag is reflected in the number and reliability of users that have associated t with d. A document is ranked high if it is tagged with *t* by many reliable taggers. Documents assigned a tag by few less reliable users will be ranked low.

After calculating the coincidence and rank, including us, the all person are interested in measuring the impact of tag spam on the result list. So the authors define a metric called *SpamFactor(t)* as follows:

$$SpamFactor(t) = \frac{\sum_{\forall di \in D_K} w(d_i) * \frac{1}{i}}{H_K}$$

where

$$w(d_i) = \begin{cases} 1 & bad \\ 0 & good \end{cases}$$

SpamFactor measures the spam in the result list introduced by bad documents. This is captured by the *w(d)* in the formula, which returns 1 if d is a bad document and 0 otherwise. SpamFactor is affected by both the number of bad documents and their position in the list. Higher SpamFactor represents greater spam in the search results.

In [9, 11], the authors show some results of experiments as follows:

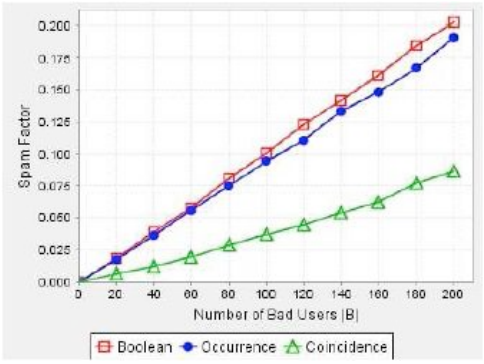

Fig. 3.1 Impact of the Number of Bad Users

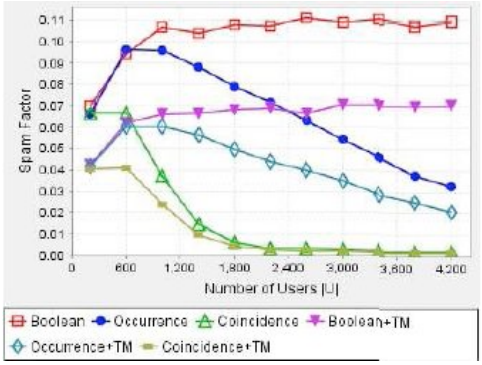

Fig. 3.2 Impact of the Number of Users

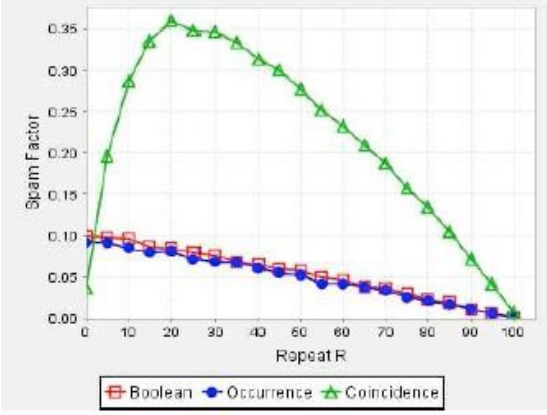

Fig. 3.3 Impact of Targeted Attacks

Comparing with the Boolean model and Occurrence model, the Coincidence model proves the power to defending the spam tag attack.

But there are some shortcomings of the metric. They are as follows:

- The coincidence-based model can not defend the collude attack. From the Fig. 3.3, we can see the reliability of the metric to defend target attack, but if the model is under more than 50% collude attack, we find the model has been compromised.
- If there are several disguised malicious users in the systems, they posting both good and bad tags to the same documents. According to the metric, the system can not defend this attack.
- Due to lack of the user weighs and some critical values, many ticklish users can create some troubles for tagging system with coincidence metric.

4. Conclusions and Future Works

From the target model, we can find the main problems of the existing metric on defending the spam tags. Collusion is a very serious problem, especially, that the coverage of malicious users is more than 50%, the existing tagging systems cannot defend this headachy attack. Decoy attack is also a problem, especially in the tagging system. Such as file system or the other the same systems do not send two different remarks to one object, but in the tagging system every object can be tagged discretionarily. If we consider the tag as a remark of object, there will be a scene that one object can be remarked by both good and bad two different labels and this scene is absence in the other systems. So the problems above will be the future work focused.

References

[1] Golder S A and Huberman B A. *The Structure of Collaborative Tagging Systems*. In Journal of Information Science, 2006.

[2] Wu X, Zhang L and Yu Y. *Exploring Social Annotations for the Semantic Web*. In Proceedings of the 15th World Wide Web Conference, 2006, 417-426.

[3] Wu H, Zubair M and Maly K. *Harvesting Social Knowledge from Folksonomies*. In Proceedings of ACM Hypertext 2006 Conference, 2006, 111-114.

[4] Zhang L, Wu X and Yu Y. *Emergent Semantics from Folksonomies: A Quantitative Study*. In Journal of Data Semantics VI, 2006, 168-186.

[5] Marlow C, Naaman M, Boyd D and Davis M. *HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read*. In Proceedings of ACM Hypertext 2006 Conference, 2006, 31-40.

[6] Yanbe Y, Jatowt A, Nakamura S and Tanaka K. *Can Social Bookmarking Enhance Search in the Web*. In Proceedings of the 2007 Conference on Digital Libraries, 2007, 107-116.

[7] Hotho A, Jaschke R, Schmitz C, and Stumme G. *Information retrieval in folksonomies: Search and ranking*. Lecture Notes in Computer Science, 2006, 411-426.

[8] Wetzker R, Zimmermann C and Baukhage C. Analyzing social bookmarking systems: A del.icio.us cookbook. In Proceedings of Mining Social Data, 2008.

[9] Koutrika G, Effendi F A, Gyongyi Z, Heymann P and Garcia-Molina H. *Combating Spam in Tagging Systems*. In Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web, 2007, 57-64.

[10] Xu Z, Fu Y, Mao J and Su D. *Towards the semantic web: Collaborative tag suggestions*. In Collaborative Web Tagging Workshop at WWW2006.

[11] Heymann P, Koutrika G, and Garcia-Molina H. *Fighting spam on social web sites: A survey of approaches and future challenges.* IEEE Internet Computing, 2007, 11(6): 36-45.

[12] Kumar R, Novak J, and Tomkins A. *Structure and evolution of online social networks*. In Proceedings of the 12th ACM SIGKDD, 2006, 611-617.

[13] Sen S, Lam S, Rashid A, Cosley D, Frankowski D, Osterhouse J, Harper F M, and Riedl J. *Tagging, communities, vocabulary, evolution*. In Proceedings of the 20th CSCW'06, 2006, 181-190.

[14] Krause B, Schmitz C, Hotho A, Stumme G. *The Anti-Social Tagger: Detecting Spam in Social Bookmarking Systems.* In Proceedings of the 4th AIRWeb'08, 2008.

[15] CiteUlike. http://www.citeulike.org.

[16] Del.icio.us. http://del.icio.us.

[17] Flickr. http://www.flickr.com.

[18] Yonggang Wang, Ennan Zhai, Jian-bin Hu, and Zhong Chen. Claper: Recommend classical papers to beginners. In FSKD, August 2010.

[19] Ennan Zhai, Ruichuan Chen, Zhuhua Cai, Long Zhang, Huiping Sun, Eng Keong Lua, Sihan Qing, Liyong Tang, and Zhong Chen. Sorcery: Could we make P2P content sharing systems robust to de- ceivers? In 9th P2P, September 2009.

[20] Ennan Zhai, Ruichuan Chen, David Isaac Wolinsky, and Bryan Ford. An untold story of redundant clouds: Making your service deployment truly reliable. In 9th HotDep, November 2013.

[21] Ennan Zhai, Ruichuan Chen, David Isaac Wolinsky, and Bryan Ford. Heading off correlated fail- ures through Independence-as-a-service. In 11th OSDI, October 2014.

[22] Ennan Zhai, Liang Gu, and Yumei Hai. A risk- evaluation assisted system for service selection. In ICWS, July 2015.

[23] Ennan Zhai, Huiping Sun, Sihan Qing, and Zhong Chen. Sorcery: Overcoming deceptive votes in P2P content sharing systems. Peer-to-Peer Networking and Applications, 4(2):178–191, 2011.

[24] Ennan Zhai, David Isaac Wolinsky, Hongda Xiao, Hongqiang Liu, Xueyuan Su, and Bryan Ford. Auditing the Structural Reliability of the Clouds. Technical Report YALEU/DCS/TR-1479, Department of Computer Science, Yale University, 2013. Available at http://www.cs.yale. edu/homes/zhai-ennan/sra.pdf.

[25] Bo Liu, Ennan Zhai, Huiping Sun, Yelu Chen and Zhong Chen. Filtering spam in social tagging system with dynamic behavior analysis. In ASONAM, Jul 2009.

[26] Ennan Zhai, David Isaac Wolinsky, Ruichuan Chen, Ewa Syta, Chao Teng and Bryan Ford. AnonRep: Towards Tracking-Resistant Anonymous Reputation. In NSDI, Mar 2016.

[27] Thierry Titcheu Chekam, Ennan Zhai, Zhenhua Li, Yong Cui, and Kui Ren. On the synchronization bottleneck of openstack swift-like cloud storage. In INFOCOM, 2016.

[28] Jianchun Jiang, Liping Ding, Ennan Zhai, and Ting Yu. VRank: A context-aware approach to vulnerability scoring and ranking in SOA. In SERE, 2012.

[29] Yonggang Wang, Ennan Zhai, Eng Keong Lua, Jian-bin Hu, Zhong Chen. iSac: Intimacy based access control for social network sites. In UIC/ATC, 2012.

[30] Ennan Zhai, Qingni Shen, Yonggang Wang, Tao Yang, Liping Ding, Sihan Qing. SecGuard: Secure and practical integrity protection model for operating systems. In APWeb, 2011.
[31] Cong Sun, Ennan Zhai, Zhong Chen, and Jianfeng Ma. A multi-compositional enforcement on information flow security. In ICICS, 2011.
[32] Ennan Zhai, Liping Ding, and Sihan Qing. Towards a reliable spam-proof tagging system. In SSIRI, 2011.
[33] Yonggang Wang, Ennan Zhai, Cui Cao, Yongqiang Xie, Zhaojun Wang, Jian-bin Hu, and Zhong Chen. DSpam: Defending against spam in tagging systems via users' reliability. In ICPADS, 2010. [58] Ennan Zhai, Huiping Sun, Sihan Qing, and Zhong Chen. SpamClean: Towards spam-free tagging systems. In CSE(4), 2009.
[34] Ennan Zhai, Zhenhua Li, Zhenyu Li, Fan Wu, and Guihai Chen. Resisting tag spam by leveraging implicit user behaviors. In VLDB, 2017.
[35] Ennan Zhai, Ruichuan Chen, Eng Keong Lua, Long Zhang, Huiping Sun, Zhuhua Cai, Sihan Qing, and Zhong Chen. SpamResist: making peer-to-peer tagging systems robust to spam. In GlobalCom, 2009.