Assessing the Value of Contributions in Tagging Systems

Elizeu Santos-Neto[1], Flavio Figueiredo[2], Jussara Almeida[2], Miranda Mowbray[3], Marcos Gonçalves[2], Matei Ripeanu[1]

Elec. & Computer Engineering Department[1]
University of British Columbia
Vancouver, BC, Canada
{elizeus,matei}@ece.ubc.ca

Dep. de Ciência da Computação[2]
Univ. Federal de Minas Gerais
Belo Horizonte, MG, Brazil
{flaviovdf,jussara,mgoncalv}@dcc.ufmg.br

Hewlett-Packard Labs[3]
Bristol, United Kingdom
miranda.mowbray@hp.com

Abstract — **Assessing the value of individual users' contributions in peer-production systems is paramount to the design of mechanisms that support collaboration and improve users' experience. For instance, to incentivize contributions, file-sharing systems based on the BitTorrent protocol equate value with volume of contributed content and use a prioritization mechanism to reward users who contribute more. This approach and similar techniques used in resource-sharing systems rely on the fact that the physical resources shared among users are easily quantifiable.**
In contrast, information-sharing systems, like social tagging systems, lack the notion of a physical resource unit (e.g., content size, bandwidth) that facilitates the task of evaluating user contributions. For this reason, the issue of estimating the value of user contributions in information sharing systems remains largely unexplored. This paper introduces this problem and takes the first steps towards a solution. More precisely, it presents a framework to design algorithms that estimate the value of user contributions in tagging systems, proposes three complementary success criteria for potential solutions, and outlines the methodological evaluation challenges.

Keywords - tagging systems; information value; entropy; web;

I. INTRODUCTION

The wide adoption of blogs, wikis, and tagging systems has transformed online information production from a centralized, proprietary, and hierarchical editorial model to a decentralized, non-proprietary, and collaborative model. Benkler [1] defines systems with the above characteristics as *commons-based peer-production systems* (or, simply, peer production systems).

Quantifying the value of individual user contributions in peer-production systems is instrumental for a number of mechanisms that enable their efficient functioning and even their long-term survival. One direct application of methods to quantify the value of individual user contributions is to support incentive mechanisms to boost participation and collaboration. For instance, in *offline* peer-production systems like car pooling, drivers have an incentive to share their cars (i.e., give rides to other people), as cars with a higher occupancy are allowed to use a dedicated faster lane [1]. Similarly, in *online* peer-production systems, like BitTorrent [6], for instance, users who contribute more (i.e., they upload more) have higher download priority. Similarly, the value of contributions can also be used to deter malicious and opportunistic users by marginalizing those users who do not contribute at all.

In peer-production systems designed for the shared use of a single type of resource, which embeds an easily measurable physical quantity, quantifying users' contributions is generally straightforward. In *Folding@Home*[1], for instance, users donate CPU cycles, thus user contributions can simply be calculated by estimating the number of (normalized) CPU/hours donated by a user. Analogously, in BitTorrent-based systems contributions can be evaluated by estimating the volume of uploaded traffic. Thus, to a great extent, quantifying contributions in these systems reduces to counting the units of donated resources. Even in peer-production systems that deal with multiple types of resources, accounting for the amounts of physical resources produced by one user and consumed by others is at the core of techniques that quantify the value of contributions [2].

However, some systems lack a clear mapping between the amount of physical resources donated and the value of a contribution. This is the case of peer-production systems designed to support the production of information goods, such as social tagging systems (e.g., *del.icio.us* or *CiteULike.org*) and wikis (e.g., *Wikipedia.org*). In particular, the lack of a quantifiable resource unit creates a new challenge: the value of each user's contributions cannot anymore be directly linked to some amount of resource used to produce them. For example, whereas a user may produce a large number of tags, only a few of them may indeed help their peers (i.e., other users of the system) in particular tasks such as navigating the list of items or organizing their item collections. Moreover, the value of information is contextual to the user. Some users may find the tag 'agneta' valuable when searching for information about a specific item in the system (e.g., Pedro Juan Gutierrez's novel Tropical Animal where Agneta is one of the characters), while others may

This work was supported in part by IDRC/AUCC (LACREG 2009/10), INCT-Web (MCT/CNPq grant 57.3871/2008–6), the Institute for Computing, Information and Cognitive Systems (ICICS) at UBC, and the authors' individual grants from BC Innovation Council (BCIC), UBC, CNPq, and FAPEMIG.

[1] http://folding.stanford.edu

find the tag ‘Latin American novel’ more useful to find the same relevant item.

To take a first step towards estimating the value of individual user contributions in online peer-production systems designed to support the production of information goods, this paper focuses on a popular instance of such systems, namely, social tagging systems, where users produce and share metadata by *annotating* content *items* (e.g., URLs, photos) with *tags* (i.e., free-form words).

In particular, *we propose a framework for solutions that aim to assess the value of user contributions in tagging systems.* The context considered is *an exploratory search mechanism* such as the “Explore Tags” feature provided by *del.icio.us*, or *MrTaggy* [3]. The intuition behind the proposed method is that the *value of tags produced by a user*, from the perspective of another user, is proportional to their ability to lead that user to content relevant to her; while *the value of items published by a user* is linked to their usefulness to other users, which, in turn, could be quantified by their usage frequency and/or the ratings reported by users (as in *Flickr.com*, where users declare photos as favorites; and, *CiteULike.org*, where users declare the importance of each item they annotate).

We instantiate a key component of the proposed framework by defining a function that estimates *the value of the tags produced by one user from the perspective of another user*. The proposed function focuses on aspects that affect the value of information, in general, and the quality of tags, in particular (Section III). The definition of other components is left as future work. III

In summary, the main contributions of this work are:

- a formalization of the problem of assessing the value of user contributions in tagging systems (Section IV);
- a framework for solutions that aim to quantify the value of contributions, and a function to estimate the value of tags (Sections V and VI);
- an evaluation methodology and the criteria to evaluate possible solutions (Section VII); and,
- a roadmap for future investigations on assessing the value of items produced by a user in tagging systems as well as the accuracy and the robustness of the proposed methods (Sections VIII).

II. MOTIVATION

This section describes solutions to assess the value of user contributions in (generic) online peer-production systems; highlights the lack of techniques for systems focused on information production and sharing; and discusses possible uses for the assessed contribution value.

A. *Quantifying Contributions in Peer Production Systems*

Online peer production systems can be categorized into systems where users produce and share: (1) resources; and, (2) information.

In the former category, as we have already mentioned, quantifying the value of user contribution is largely based on counting the amount of resource units one user produces and donates to other users (and implicitly to the system). For example, in P2P grids (e.g. Folding@Home and OurGrid [4]), contributions are quantified as CPU hours, whereas in P2P content sharing (e.g. Tribler [5], BitTorrent, SopCast.com and PPLive.com) the value of contributions is estimated by the volume of content a peer donates to the other.

Valuing contributions in these resource-sharing peer-production systems relies on: first, the fact that the amount of resources donated are easily quantifiable, second, the assumption that contribution value can be directly linked to the resources consumed to deliver a service, and third, on the simplifying assumption that a unit of contributed resources has uniform perceived value across all users of the system (e.g., in Folding@Home one-hour of normalized CPU time has the same value regardless of when and to whom it is delivered).

In contrast, none of these assumptions hold for systems that support production/sharing of information. First, it is impossible to directly quantify the ‘effort’ that has led to the production of a specific piece of information; and, second, value of information (e.g., items in tagging systems) is highly subjective to users’ opinions and interests.

To address this latter issue of contextual value, some peer production systems, such as *Yahoo! Answers*[2] , *Flickr.com* and blogs in general, allow users to *rate* content items (e.g., answers, photos, or blog posts and comments respectively). The rating given by a user expresses how much she liked that particular item, and can be interpreted as an estimate of the value of that contribution from her perspective. Although this approach generates rich feedback about what users like (or sometimes dislike), it has two limitations. First, rating information is generally sparse (i.e., the majority of users do not express their preferences via ratings); and second, in tagging systems, item rating does little to address the problem of assessing the value of tags.

To fill these gaps a method that takes into account both the value of items and tags is necessary. Clearly, both explicit (e.g., ratings) and implicit (e.g., usage statistics) feedback are useful and could be used as part of a solution.

B. *Harnessing the Value of Contributions*

Let us assume, for the sake of exposition, that we have already devised a method to assess the value of user contributions in tagging systems. This section highlights mechanisms that could benefit from the output from such method. We note that, while each one of the cited mechanisms can be implemented independently, *the concept of contribution value we propose can serve as a unified currency* to support all these mechanisms.

[2] http://answers.yahoo.com

Once users are ranked according to the value of their contributions to their peers such ranking can be used:

- *To create incentives for participation*: by rewarding heavy contributors with, for example, tangible rewards such as more resources to store their own content[3], access to exclusive content[4], or even intangible rewards such as stars [7,8];
- *To prioritize access to scarce resources*: requests to the system could be prioritized according to the rank of the user who is requesting it. This can be at the same time a natural incentive mechanism in Q&A portals such as *vark.com* [9].
- *To tame spamming* [10], for instance, by filtering from search results those users who has low value contributions;
- *To highlight trends and prioritize content display*: the *content* produced by users who have higher value of contributions could be used to personalize the final search result;
- *To help recommending experts in a particular topic*: the value contributed by a user can be a useful signal to distinguish between users that are knowledgeable about a given field.

We note that most of the above mechanisms that harness the value of contributions in information-sharing peer-production systems have an equivalent in resource-sharing systems. For example, some BitTorrent-based systems, where contribution value is estimated as proportional to the content volume uploaded over the life of a peer (i.e., the *sharing-ratio enforcement mechanism* [11] is an example of prioritizing the user requests, such as publication of new torrents, based on the value of user contributions).

III. BACKGROUND AND RELATED WORK

This section provides the background for the design of future methods to quantify the value of user contributions in tagging systems. First, it reviews past work on the general problem of assessing the *value of information goods*. Next, it discusses previous efforts towards evaluating the *quality* tags and their usage in *social information foraging* tasks.

A. The Value of Information Goods

Assessing the value of peer-produced information in tagging systems is an instance of the problem of assessing the value of information goods. In this topic, Hirschleifer [12] describes five aspects that affect the value of information:

Certainty – the value of information goods depends on the certainty the information provide about the outcome of a particular process. For example, given a user in a tag-based navigation system, the value of a tag should be proportional to the increase of certainty that the outcome of the navigation process is a set of relevant items.

Diffusion – the availability of information goods across the user population may affect their value, for example when few users possess certain information. In the context of a tagging system, one may think of particular items or tags that are shared with selected people versus the popular ones.

Applicability – information goods can be of general or particular applicability or interest. That is, the information may serve a general audience or only a small fraction of the user population. For instance, some tags can be generic enough (e.g., networks) to be of interest to several users. Conversely, other tags (e.g., agneta) are only relevant to a restricted subset of the user population.

Content – naturally, the value of a piece of information may be affected by the characteristics of its contents. Hirschleifer points out two common subclasses of this aspect in markets, where content conveys information about the environment or individuals' behavior. The content dimension of peer-produced information maps to its semantic. For example, a tag may express how the user intends to use an item (e.g., 'to-read'); or, which topic the item belongs to.

Decision-relevance – this dimension captures the importance of the information for a decision problem. For instance, the information that a friend is reading a particular book may be valuable to a person to decide whether or not to buy that book.

Stigler [13] and Bates [14] complement the list presented by Hirschleifer [12] by stating that the value of information is only fully determined by its use. Thus, to estimate value of tags and items, this work must consider the context in which they are used.

B. Characterizing the Qualitiy of Tags

Several previous studies focused on assessing the quality of tags, and indirectly target some aspects discussed by Hirschleifer. Chi and Myticowcz [15] use information theory to evaluate the efficiency of tag-based faceted search: in particular, they use the entropy of the set of items conditional on tags as a measure of a tag's effectiveness to reduce the search space, which is a way to quantify the certainty aspect of tags. However, Chi and Myticowcz [15] do not account for the relevance of items retrieved by the tags when evaluating tag-based search efficiency.

Along a similar path, Heymann and Garcia-Molina [16] investigate whether tags help users to categorize content by analogy with well established classification tools deployed by library management systems. They use a qualitative analysis approach to evaluate the power of tags to build classification systems rather than a user-centric quantitative approach to assess value.

Similarly, Pirolli [17,18] studies the process of locating information of interest, using the metaphor of animals foraging for food to analyze the information seeking process. Pirolli explores the intuition that social search is more

[3] Flickr limits the number of high quality photos a normal user can upload. This limit is lifted once the user pays the service subscription.

[4] In CiteULike, one could think of free limited access to articles or online books that are only available on portals that require subscription.

efficient (i.e., when information seekers obtain 'clues' from other users, as opposed to discovering by themselves). The models proposed by Pirolli assume that users find towards information of interest by following 'foraging clues' left by other users. In tagging systems, we can view tag-based navigation as an information foraging process: tags are clues left by some users that may help others to find items of interest. The idea that tags represent clues used in the information search process [17,19,20] inspires this work. More specifically, this work proposes to quantify the value of tags by measuring its impact on information seeking tasks.

The social information foraging models concentrate on the process (and its efficiency) that users follow to find information of interest. Assessing the value of tags and items (i.e., the value of user contributions) may serve as the *utility function* that guides the user in the foraging process.

Other studies focus on the content aspect of peer-produced information. Suchanek et al. [21] studies the quality of tags by determining the descriptive power of a tag (i.e., its efficiency in describing an item). In the same spirit, Dubinko et al. [22] propose 'interestingness' as a metric to estimate the quality of items and tags, as perceived by the users. The metric is harnessed by mechanisms in *Flick.com* to rank photos and tags during the navigation process.

Related to these previous studies, but focusing on the quality of tags for information retrieval tasks such as content classification and general search, Figueiredo et al. [23] and Bischoff et al. [24] evaluate the quality of information provided by tags in comparison to other textual features of the items available in the system. This work focuses on the value of tags that one user produces from the perspective of another user, instead of analyzing the quality of tags for a given task regardless of the user who is performing it.

IV. Notation, Assumptions and Problem Statement

This section introduces the notation and the assumptions about the user interaction model used in this work. It formally defines the problem of assessing the value of user contributions in social tagging systems.

A. Notation and Assumptions

Let $\mathbb{S} = \langle U, I, A\rangle$ be a social tagging system, where U represents the set of users in the system, I denotes the set of items, and A represents the set of annotations. An annotation is a tuple that specifies its author, the annotated item, the tags used, and the time it happened. Formally, $A = \{\langle u, i, L, t\rangle\}$, where $u \in U$, $i \in I$, L is a set of tags (i.e., free-form words selected by the user to annotate the item at time $\boldsymbol{t}$). Note that we group all tags a user annotates an item at one moment into a single annotation event.

The set of annotations A_u, where individual annotations can be simply distinguished by their timestamp, characterizes a particular user $u \in U$. More formally, $A_u = \{\langle q, i, L, t\rangle \in A | q = u\}$. From the set of annotations A_u, it is possible to derive the set of items I_u and the set of tags T_u annotated, and respectively, used by the user u. It follows that $I_u = \{i | \langle u, i, L, t\rangle \in A_u\}$, and $T_u = \{l | \langle u, i, L, t\rangle \in A_u, l \in L\}$. The set of tags assigned to a particular item i, T^i, and the set of items tagged with a particular tag l, I^l, are similarly defined. The set of all tags in the system is given by $T = \bigcup_{u \in U} T_u$.

Since one annotation may contain multiple tags, we define the set $\mathcal{L}_u$ as the set of tag annotations L_u produced by user u. More formally, $\mathcal{L}_u = \{L | \exists \langle u, i, L, t\rangle \in A_u\}$. Finally, for a tag $l \in T$, let $I^l = \{i \big| l \in T^i\}$ represent the set of items annotated with l. Unless stated otherwise, subscripts identify users and superscripts identify items (or tags).

B. Formal Problem Statement and Success Criteria

The problem of assessing the value of user contributions is formalized as follows: given two users s and u from a social tagging system $\mathbb{S}$, we want:

(1) To define a function $K(u, s)$ that quantifies the value provided by u to s, combining the value of both *tags* and *items* produced by u.

(2) To define a ranking method to order information producers from the perspective of each information seeker in the system according to $K(u, s)$[5].

Three aspects should be considered when evaluating solutions to compute functions (1) and (2) above. They are:

i) *feasibility* – given that social tagging systems can possibly deal with millions of users, the computational, storage and communication overheads involved in computing $K(u, s)$, ranking the information producers should not be prohibitive.

ii) *accuracy* – the estimation of user contributions' value should be as close as possible to the *true* value of user contributions; and,

iii) *robustness* – the method to compute $K(u, s)$ and ranking the information producers should be robust against malicious and opportunistic attempts to manipulate (inflate/deflate) the value of one's own or other users´ contributions.

V. A Framework to Assess the Value of User Contributions

Users in a tagging system are either information producers or information seekers, depending on the action they perform at a given moment. Information producers publish new items and/or annotate existing items. An information seeker navigates the set of items available in the system. To assess the value of a user's contribution in such system, one must combine the value of items and tags produced by the user.

[5] Note that this function is a building block and could be used to determine the value of one's contribution to the entire system -- $K(u,*)$.

More formally, let $v(l_u, s)$ and $r(i_u, s)$ be two functions that quantify the values of a tag l_u, and of an item i_u, respectively, produced by user u, from the perspective of an information seeker s. A function $K(u, s)$ should combine $v(l_u, s)$ and $r(i_u, s)$ for all tags and items produced by u.

In particular, the intuition behind computing $v(l_u, s)$ is that the value of a tag should be proportional to its ability to lead user s to items that are potentially relevant to her. One way to infer the items that are potentially relevant to a user is to assume that the user's past tagging activity provides a good approximation of her future information needs, and to apply a personalized item recommendation technique to infer both the set of items and their respective probability of relevance. One such technique is based on a personalized random walk approach on a tripartite graph that connects users to tags and tags to items, where the probability of relevance for the items is given by the stationary distribution of the random walk [25,26,27]. Note that there are other alternatives to infer the set of items relevant to a user, such as association rules mining [38].

Similarly, the value $r(i_u, s)$ of an item i_u to an information seeker s should be proportional to its 'relevance' and 'usefulness' to user s. This can be estimated directly based on: (1) network analysis similar to that applied to the *citation graph* to find influential authors [28,29]; (2) direct user feedback such as ratings; or (3) indirect user feedback such as the frequency an item is (re)visited.

Figure 1 presents a block diagram that illustrates the process to assess the value of user contributions. The top part of the diagram presents the flow to calculate the value of tags produced by user u to an information seeker s as a function of the tags' ability to lead s to items relevant to her. The 'Tag Value Calculator' block combines the information seeker's set of relevant items Γ_s^W (produced by the relevant item set estimator, which can be based on an item recommendation engine) and the information producer's annotations to determine the value of tags $l \in T_u$ (i.e., the tags extracted from the annotations produced by u) to s.

The bottom part of the diagram presents the flow to calculate the value of items produced by u that are used by s. The 'Item Value Calculator' box combines the information seeker's item usage statistics, represented by F_s (output from the item usage monitor), and the set I_u of items originally published by u to estimate the value of these items. These usage statistics can be obtained via click traces, for example, that provide information about how often a user consumes a particular item.

Finally, the estimated values of tags and items are aggregated separately and then combined into the value of the contributions from u to s, $K(u, s)$.

It is important to highlight that the proposed framework (Figure 1) is generic. Each building block can be instantiated according to the specific characteristics of the system. For example, the availability of user activity data, such as records of tag assignments, click traces, item ratings, friendship links, or group co-membership information, can certainly drive the design of specific solutions for the value calculator and aggregator boxes.

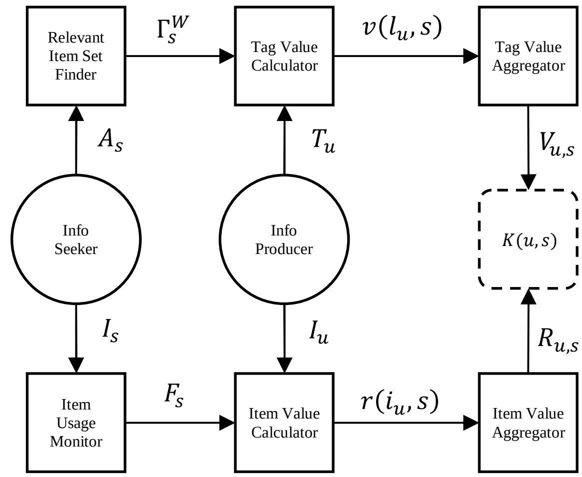

Figure 1. Components of a framework to quantify the value of user contributions.

The next section focuses on assessing the value of tags $v(l_u, s)$ assuming the availability of tag assignments traces (e.g., as collected and studied by Santos-Neto et al. [30]).

VI. ASSESSING THE VALUE OF TAGS

Our attempt to assess tag value builds on a formalization of the *certainty* and *applicability* aspects of information value (as categorized by Hirschleifer [12] – Section III.A) and the observation by Stigler that the value of information goods is only determined by its use [13].

More specifically, we consider the value of user contributions in the context of a faceted search mechanism such as the "Explore Tags" feature provided by *del.icio.us*, or MrTaggy [20]: the user interacts with the system by entering a set of keywords (tags) and the system retrieves the items that are annotated with *all* these keywords. The assumption is the set of items retrieved by the system is narrowed by adding new tags (that match a subset of items).

In the context of such navigation mechanism, the method described below assesses the value of a tag as directly related to the amount of uncertainty reduction about the set of relevant items, when an information seeker applies a given tag to navigate the system. Mutual information [31], reviewed below, formalizes this notion of uncertainty.

Given a set of items I, which are relevants to a given user, and a probability mass function $P(i)$ over I, which can be interpreted as the probability of an item being relevant to a given user, Shannon's entropy [31] quantifies the average information content (or uncertainty) in the set of items as follows: $H(I) = -\sum_{i \in I} P(i) \, log \, P(i)$.

Note that the maximum entropy occurs when the probability mass function is uniform, and it is zero if the set is either empty or the probability mass function is collapsed at a single element of the set (i.e., in both cases there is no uncertainty in the set). Thus, $0 \leq H(I) \leq log|I|$.

The conditional entropy is similarly defined by using the probability of items conditional on a set of tags. The conditional entropy is useful, for instance, to measure the remaining uncertainty in the set of items I, if a user uses a tag l to navigate. Conditional entropy is expressed as follows: $H(I|l) = -\sum_{i \in I} P(i,l) \log P(i|l)$, where $P(i,l)$ can be interpreted as the probability that item i is relevant and is annotated with tag l, in the context of an information seeker; and, $P(i|l)$ represents the probability that a particular item i is relevant to an information seeker, given that a tag l is used to navigate the system. Next, we use the subscript to indicate that the probabilities are contextual to an information seeker.

Mutual information combines the two entropy definitions above to quantify the reduction in uncertainty about a random variable given the knowledge about another random variable. In this study, mutual information provides a way to quantify the reduction in uncertainty about a set of items I by using a tag l to narrow down the navigation space. The formal definition of mutual information [31] is given by: $M(I;l) = H(I) - H(I|l)$. Normalizing it by $H(I)$, the normalized mutual information is expressed as follows:

$$\bar{M}(I;l) = \begin{cases} 1 - \frac{H(I|l)}{H(I)} & , H(I) > 0 \\ 0 & , H(I) = 0 \end{cases} \quad (1)$$

We use the normalized mutual information (Eq. 1) to estimate the value of a tag l, from the perspective of an information seeker s, as follows.

DEFINITION 1: *given an information seeker s, a sequence of tags W that expresses one of her information needs, and a set of items Γ_s^W that are relevant to the information seeker's specified information need,* ***the value of a tag l produced by a user u, with respect to the information need W from another user s****, is defined as follows*:

$$f(l_u, W, s) \stackrel{\text{def}}{=} \rho(l_u)\bar{M}(\Gamma_s^W; l_u) \quad (2)$$

where $\rho(l) = \frac{\sum_i P(i)}{|I^l|}, i \in I^l$ is the average probability of relevance of the items retrieved by tag l.

The rationale behind Eq. 2 is that if user u produces tag l_u that return only irrelevant items to s (i.e., $\rho(l_u) = 0$), these tags are useless to the information seeker, even though they may reduce the uncertainty about the set Γ_s^W.

On the other hand, if l_u leads the user to a subset of relevant items, the value of these tags is proportional to the reduction in the uncertainty about the set of relevant items and the relevance of the items retrieved by l_u, which is represented by the coefficient $\rho(l_u)$.

Note that the information needs of a particular information seeker s can be approximated by either the previously used query terms, or by the set of tags used to annotate items. Indeed, for more than 62% of URLs (in a sample collected from *del.icio.us*), at least 50% of tags associated to these URLs overlap with query terms used to locate them [32].

VII. EVALUATION: CHALLENGES AND METHODOLOGY

This section discusses the challenges and the possible solutions to evaluate the proposed method to assess the value of user contributions in tagging systems. In particular, it presents a high level view of the design decisions involved in the experiments and in estimating baselines to evaluate the accuracy of the proposed method.

A. *Estimating a Ground Truth*

Determining the accuracy of the estimated contribution value is challenging. Ideally, one should compare the value estimates to a ground truth. Obtaining a ground truth, however, is costly if not impossible.

The ground truth, however, can be itself *estimated* in at least three ways: (1) by conducting a controlled survey on a sample of the user population, where survey participants are asked to rate the annotations produced by other users; (2) by ranking users based on how frequently the tags they produce are actually used by their peers to find relevant items; and, (3) by ranking users according to their opportunity to produce valuable information given their position in the social network.

The first option, that is, the controlled survey, has traditionally been employed. This option has, however, two major shortcomings: first, it can be conducted only at a small scale compared to the scale of today's tagging systems; second, the unavoidable degree of subjectivity in human-produced ratings may lead to significant inconsistencies, which introduces an extra cost to the experiment [34,35].

The second option, estimating the value of annotations based on how often they are used to retrieve items is more practical. However, the feasibility of this approach is predicated on the availability of detailed activity records (search logs and click traces). To date, we do not have access to such detailed activity records from tagging systems.

Finally, another (yet even less precise) method may circumvent the above limitations. This method is based on two assumptions: first, that the opportunity to produce value/knowledge in a social network is linked to a participant's position in that social network, and, second, that the opportunity to produce value and the value actually produced are highly correlated.

Previous work by Burt [33] partially supports the first assumption above. Burt shows that the 'network constraint index' reflects well the *opportunity* for information brokerage in social networks, and consequently the opportunity to produce valuable information. In his experiments with the social network formed by employees of a supply chain company, Burt shows that the value of ideas was significantly *higher* for those employees with *low* network constraint index (i.e., in a position of the network that allowed them to bridge two highly connected clusters of people). We conjecture, admittedly at this point without strong support, that the social graph the can be constructed based on the explicit social links between users of an online social (for instance, in *del.icio.us* and *CiteULike* users declare others as their contacts, and in *Flickr* users declare friendship to others) can be used to infer the opportunity users have to produce value.

In more detail, our conjecture is that users of social tagging systems that have a *low* network constraint index are in a better position to produce *valuable* contributions, as information (items and tags) flow through the links in the social graph, which enable information producers to work like brokers between heterogeneous clusters of users. This brokerage position, in turn, enables the user to produce valuable information (both items and tags).

B. Experimental Design

This section discusses aspects of the experimental evaluation of potential solutions to assess the value of contributions in tagging systems. First, it presents the metrics used to evaluate the *accuracy*, *feasibility* and *robustness* of a proposed method. Second, it lists comparison baselines. Finally, it briefly comments on factors that may influence the accuracy and robustness of a method.

Metrics. We proposed three criteria to evaluate a method to quantify the value of users' contributions: *accuracy*, *feasibility* and *robustness*. To evaluate *accuracy*, we propose to use a metric to compare ranked lists such as Kendall's tau distance [36] or NDCG [37]. The idea is to rank users based on the value of their contributions and compare the rankings. *Feasibility* can be evaluated by the algorithm's time and space complexity together with some characterization of the system activity that may impact the cost of updating the value of users' contributions (e.g., the rate of annotations produced by users in the system). *Robustness* should be evaluated by a metric such as *spam factor* – i.e., the fraction of spammers (or malicious users) that gets promoted by a given attack strategy (e.g., imitation of genuine user).

Baseline. There are several potential baselines to compare the proposed method against. The comparison should highlight the tradeoffs between the techniques. For instance, one could use *activity similarity* to assign value to the tags/items produced by a user to her peers – while this may be efficient to compute, it may not be robust, as it is easy to an attacker to imitate the tagging behavior of a target user; another baseline could be *popularity* of items and tags.

Factors. At least two aspects may affect the efficiency of a method to quantify user contributions: the *search model* and the *aggregation methods*. The goal is to investigate the impact of these design decisions on the efficiency of the proposed solutions.

Scalability issues. The contextual nature of information value creates a challenge to techniques that quantify the value of user contributions. To take the context into account, the value of a tag or item produced by a user needs to be evaluated from the perspective of each user in the system separately. This requirement clearly poses a scalability issue, as it suggests that changes in the system state may imply the computation of value for the entire population too often.

VIII. Summary and Future Work

This work introduces the problem of assessing the value of user contributions in peer production systems and focuses on one particular class of systems: tagging systems. We propose three success criteria that should be considered when evaluating possible solutions: feasibility, accuracy, and robustness. Moreover, we propose a framework that describes a class of solutions to estimate contribution value in tagging systems, and we suggest how this framework can be instantiated for particular algorithms that make use of tag annotation traces and social networks. Our proposed solution to assess the value of the tags formalizes the idea that: tags produced by a user are valuable to another user, if they increase her ability to find relevant items.

We are aware that there is a long way before fully demonstrating the merits of the solution we propose. To get there, however, there are four paths to be pursued as future work: first, the design of a method to assess the value of items; second, the robustness evaluation of the proposed method; third, the evaluation of the impact of different search models and aggregation methods on the assessed value of user contributions; and, fourth, a full analysis of the time and space complexity of the algorithms proposed.

Acknowledgements

The authors thank Laks Lakshmanan for insightful discussions and guidance; Lauro Beltrão, Paulo Ditarso, Nazareno Andrade and Francisco Brasileiro for valuable comments.

References

[1] Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale Univ. Press, 2006.

[2] M. Mowbray, F. Brasileiro, N. Andrade, J. Santana, and W. Cirne, "A Reciprocation-Based Economy for Multiple Services in Peer-to-Peer Grids," In *Proceedings of the 6th IEEE international Conference on Peer-To-Peer Computing*, 2006.

[3] B. M. Evans and E. H. Chi, "Towards a model of understanding social search", In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 2008.

[4] W. Cirne , F. Brasileiro, N. Andrade, L. Costa, A. Andrade, R. Novaes, and M. Mowbray, "Labs of the World, Unite!!!", *Journal of Grid Computing*, vol. 4, 2006, pp. 225-246.

[5] J. A. Pouwelse, P. Garbacki, J. Wang, A. Bakker, J. Yang, A. Iosup, D.H. Epema, M. Reinders, M.R. Steen, H.J. Sips, "TRIBLER: a social-based peer-to-peer system," *Concurrency and Computation: Practice Experience*, vol. 20, 2008.

[6] B. Cohen, "Incentives Build Robustness in BitTorrent", In *Proceedings of the International Workshop on Peer-to-Peer Systems*, 2003.

[7] M. Ripeanu, M. Mowbray, N. Andrade, and A. Lima, "Gifting Technologies: a BitTorrent Case Study," *First Monday*, vol. 11, 2006.

[8] D.R. Raban, "The incentive structure in an online information market," *Journal of the American Society for Information Science and Technology*, vol. 59, 2008.

[9] D. Horowitz and S. D. Kamvar, "The anatomy of a large-scale social search engine", In *the Proceedings of the 19th International World Wide Web Conference*, 2010.

[10] G. Koutrika, F. A. Effendi, Z. Gyöngyi, P. Heymann, and H. G. Molina, "Combating spam in tagging systems: An evaluation", In the *ACM Transactions on the Web*, vol. 2, 2008.

[11] N. Andrade, E. Santos-Neto, F. Brasileiro, M. Ripeanu. "Resource demand and supply in BitTorrent content-sharing communities", *Computer Networks*, vol. 53, 2009.

[12] J. Hirshleifer, "Where Are We in the Theory of Information?", *The American Economic Review*, vol. 63, 1973.

[13] G. J. Stigler, *The Organization of Industry*, University Of Chicago Press, 1983.

[14] B. D. Ruben, *Mediation, information, and communication*, New Brunswick, NJ [u.a.]: Transaction Publ., 1990.

[15] E. H. Chi and T. Mytkowicz, "Understanding the efficiency of social tagging systems using information theory", *In the Proceedings of the 19th ACM Conference on Hypertext and Hypermedia*, 2008.

[16] P. Heymann, A. Paepcke, and H.G. Molina, "Tagging human knowledge", *In the Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, 2010.

[17] P. L. Pirolli, *Information Foraging Theory: Adaptive Interaction with Information*, 2007.

[18] E. Chi, P. Pirolli, and S. Lam, "Aspects of Augmented Social Cognition: Social Information Foraging and Social Search," *Online Communities and Social Computing*, 2007.

[19] P. Pirolli, "Powers of 10: Modeling Complex Information-Seeking Systems at Multiple Scales", *Computer*, vol. 42, 2009.

[20] Y. Kammerer, R. Nairn, P. Pirolli, and E. H. Chi, "Signpost from the masses: learning effects in an exploratory social tag search browser", *In the Proceedings of the 27th International Conference on Human Factors in Computing Systems*, 2009.

[21] F. M. Suchanek, M. Vojnovic, and D. Gunawardena, "Social tags: meaning and suggestions", *in the Proceeding of the 17th ACM conference on Information and Knowledge Management*, 2008.

[22] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins, "Visualizing tags over time", *In the Proceedings of the 15th international conference on World Wide Web*, 2006.

[23] F. Figueiredo, F. Belém, H. Pinto, J. Almeida, M. Gonçalves, D. Fernandes, E. Moura, and M. Critso, "Evidence of Quality of Textual Features on the Web 2.0", *in the Proceedings of the 18th ACM conference on Information and Knowledge Management*, 2009.

[24] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu, "Can all tags be used for search?", *In The Proceedings of the 17th ACM conference on Information and Knowledge Management*, 2008.

[25] H. Cheng, P. Tan, J. Sticklen, and W. F. Punch, "Recommendation via Query Centered Random Walk on K-Partite Graph", *In the Proceedings of the 7th IEEE International Conference on Data Mining*, 2007.

[26] M. Clements, A. P. de Vries, and M. J. Reinders, "The task dependent effect of tags and ratings on social media access", In *ACM Transactions on Information Systems*, 2010.

[27] I. Konstas, V. Stathopoulos, and J.M. Jose, "On social networks and collaborative recommendation", *In the Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 2009.

[28] J. Weng, E. P. Lim, J. Jiang, and Q. He, "TwitterRank: finding topic-sensitive influential twitterers", *In the Proceedings of the 3rd ACM international conference on Web search and data mining*, 2010.

[29] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, and H.A. Makse, "Identifying influential spreaders in complex networks," *http://arxiv.org/abs/1001.5285*, 2010.

[30] E. Santos-Neto, D. Condon, N. Andrade, A. Iamnitchi, and M. Ripeanu, "Individual and social behavior in tagging systems", *In the Proceedings of the 20th ACM Conference on Hypertext and Hypermedia Hypertext*, 2009.

[31] T.M. Cover and J.A. Thomas, *Elements of Information Theory 2nd Edition*, Wiley-Interscience, 2006.

[32] M. J. Carman, M. Baillie, R. Gwadera, and F. Crestani, "A statistical comparison of tag and query logs", *In the Proceedings of the 32nd international ACM SIGIR conference on Research and Development in Information Retrieval*, 2009.

[33] R. S. Burt, "Structural Holes and Good Ideas," *The American Journal of Sociology*, vol. 110, 2004.

[34] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and evaluating choices in a virtual community of use", *In the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1995.

[35] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems", In the *ACM Transactions on Information Systems*, vol. 22, 2004.

[36] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists", In the *Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics*, 2003.

[37] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques", In the *ACM Transactions on Information Systems*, vol. 20, 2002.

[38] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules”, *In the Proceedings of the 20th International Conference on Very Large Databases*, 1994.