Collabio: A Game for Annotating People within Social Networks

ABSTRACT

The operation of numerous computing applications can be personalized by gaining access to information about users' goals, interests, and personalities. We focus on the opportunity to endow systems with such information by developing methods for harnessing latent knowledge held by people about other people. We present Collabio, a game within an online social network that encourages friends to tag one another. To evaluate the efficacy of the approach, we examine usage statistics, review self-assessments of the quality of tags, and explore how data acquired via Collabio augments tags that could have been scraped online. We demonstrate how Collabio tags are useful in three proof-of-concept applications: an aggregate tag cloud visualization, a personalized RSS feed, and a question and answer system.

Author Keywords

Social computing, human computation, tagging.

ACM Classification Keywords

H5.3. Information interfaces and presentation (e.g., HCI): Web-based interaction.

INTRODUCTION

Researchers and developers have been pursuing opportunities to personalize the behavior of software applications and services by considering individuals' preferences, interests, and needs. News sites aim to harness details of users' interests to highlight stories likely to be relevant; search engines might rank the UIST conference website more prominently than the South Uist island when used by a computer science researcher; applications could facilitate task completion by connecting people with complementary expertise. Achieving this vision requires that systems gain access to knowledge such as a user's background, history of interactions, expertise, activities, and interests.

There are several possible approaches to collecting information to build user models. One could solicit the information directly from users (*e.g.,* use a question template, Facebook's profile page, or expert matching system K-Net [21]). Unfortunately, these approaches tend to be tedious and many users are not willing or able to expend the effort required. One could try to automatically infer the information by mining the web, personal communication, or other documents (e.g., [2, 15, 17]). However, data mining is non-trivial and not all facets of a user's life may be readily available through existing data sources. Alternatively, one could use collaborative filtering or other recommender system technologies to infer such attributes as the preferences of individual users (*e.g.,* Amazon.com or Netflix recommendations). While such systems can provide value, they can fail to be effective when there is limited data and/or great amounts of variation among individuals.

One of the two major hypotheses framing our work is that friends and associates are a rich source of information relevant to personalization. For example, friends within your social network often know your personality, expertise, artistic and musical tastes, topics of importance, quirky habits, and so on, sometimes even better than you might know yourself. Our second major hypothesis is that members of a social network can be motivated to share personal information about each other, and a conceptually new type of Game with a Purpose [27]—a social version—is an appropriate

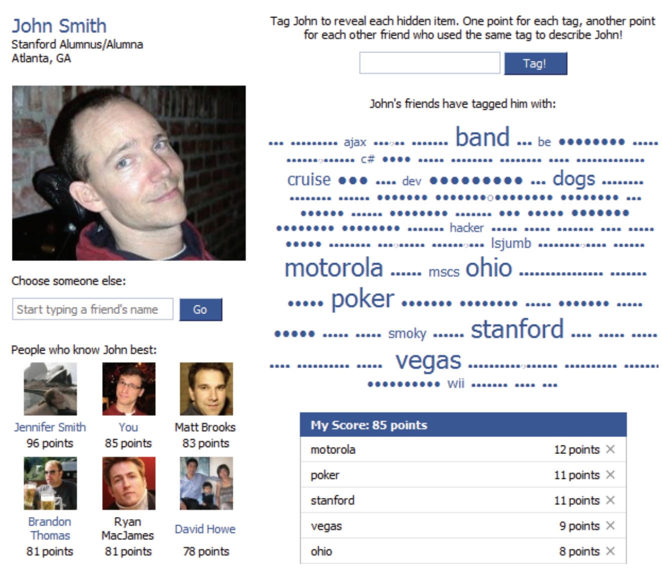

Figure 1. The user has guessed several tags for John Smith, including *band, ohio* and *vegas*. Tags guessed by John's other friends are hidden by dots until the user guesses them.

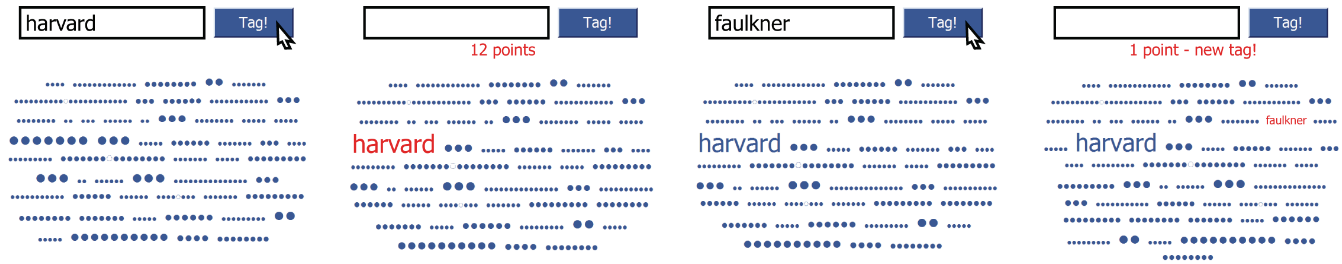

Figure 2. The tag cloud begins completely obscured. The player guesses *harvard*, receives 12 points for agreeing with eleven other players and reveals Harvard as a large tag. *Faulkner* is next; it does not match existing tags and is added to the cloud.

incentivizing mechanism. We believe that the information obtained in this manner can provide a basis for personalization that is not obtainable by other means.

The contributions of this paper are threefold:

1. We present Collabio[1] (for *Collab*orative *Bio*graphy), a game that elicits descriptive tags for individuals within the Facebook social network (Figure 1). The application leverages properties of the social network such as competition and social accountability to solve the tag motivation and accuracy problems completely within a social framework.
2. We evaluate the efficacy of our approach by examining usage statistics, investigating how tag quality is impacted by the number of friends generating the tag, and exploring how Collabio tags augment ones that might have been generated through traditional online scraping techniques. We find that the knowledge generated is accurate and often unavailable through other means.
3. We demonstrate that the collected tags are useful for applications by prototyping a trio of visualization and personalization applications: an aggregate tag cloud visualization (Collabio Clouds), an expert-finding question and answer system (Collabio QnA), and a personalized RSS feed (Collabio RSS).

Collabio is the first social tagging application we know of to explicitly integrate game elements and to tactically hide unguessed tags from users in a bid to acquire verification of the tags' accuracy. To encourage desirable user behavior in generating useful tags, we intentionally create social accountability by attributing all tags to authors, and make this information available to the person being tagged. This work also presents the most expansive evaluation of the quality and reproducibility of social people-tagging to date, and is the first to build personalized applications using the tags.

COLLABIO

Collabio is embedded in the Facebook social network. In the following sections, we describe Collabio's three top level interface tabs: the tab in which users can *Tag!* their friends, the one in which they can manage *My Tags*, and the one in which they can see the *Leaderboard*. We then discuss two other important issues: propagation through the social network and issues of cheating and abuse.

Tag! Friends

The main activity of Collabio is to tag a friend, so the focus of the user's experience is the tagging page (Figure 1). The user sees the tag cloud others have created by tagging the selected friend. When presenting this cloud, Collabio only displays tags that the user has already explicitly guessed (Figure 2). Tags not yet guessed are obscured by replacing each constituent letter with a solid circle; for example, the tag UIST appears as ●●●●. Whitespace in obscured tags is represented by clear circles such as ○. Thus, the length and makeup of the obscured tag provide hints as to the hidden text. As an additional hint, terms in the tag cloud are alphabetically ordered. The tags in the cloud are scaled so that the popular tags are larger.

As the user tags a friend, one of two things happens (Figure 2). If the tag is new and has not previously been placed on their friend, the tag is inserted into the cloud. If the tag exists, then it is revealed within the cloud. For each guess, users receive points equal to the total number of people who have applied a tag, including themselves. If they are the only person to have guessed that tag, then they get 1 point; if there are 11 others, they get 12 points. These points continue to accumulate as more people apply the tag, so earlier taggers' scores rise as well. A user can retract a tag by clicking on a small × by the tag.

To expose one's score to others, and to stimulate competition, each tagged friend has a "People who know [this friend] best" pane which lists friends who have earned the largest number of points from tagging that friend (Figure 1).

In the current system, if the user is the first to tag a friend, Collabio seeds the tag cloud with terms from the friend's public profile (such as network names, affiliations, or interests), ensuring that the tag cloud is never completely empty. These tags are attributed to the "Collabio Bot." We observed early on that users were typically unwilling to tag others who had not already added the application, so this tag seeding is helpful in overcoming reluctance to be the first to tag an individual. We had briefly explored various schemes in which we used bonus points to motivate being the first to tag someone. However, we found that this was not effective, and we replaced it with tag cloud seeding early in the application's life.

[1] http://apps.facebook.com/collabio

Figure 3. The *My Tags* page allows the user to view his or her tag cloud completely uncovered.

Most Tags from Friends	
You	1674
Jane Smith	484
John Smith	373
Tom Anderson	260
Margot Yang	245
Michael Johnson	235
Linda Williams	235
Elizabeth Jones	219
Sawako Roshi	214
Mark Davies	194

Most Friends Tagged	
Jen Brown	213
You	205
Marcos Brino	61
Grace Brown	55
Elizabeth Jones	55
Jane Smith	55
David White	45
Sarah Miller	40
John Smith	38
Michael Johnson	36

Figure 4. Collabio leaderboards feature the friends with the most tags and the friends who have tagged the most others.

Managing *My Tags*

The *My Tags* interface allows users to inspect and manage tags their friends have placed on them. The My Tags page contains three sections: a fully uncovered tag cloud (Figure 3), an expanded top scorers list, and a table explaining which friends tagged the user with which tags. In order to allow people to maintain control of tags placed on them, Collabio allows them to easily delete tags from their tag cloud by clicking on a small × by the tag.

The *Leaderboard*

The third Collabio tab is the Leaderboard. While the individual leaderboards on the *Tag!* tab encourage users to keep tagging a friend until they are listed as one of the Top Friends for that person, the global leaderboards encourage users to continue tagging activity within the application. We present two lists here, one of the friends that have the most unique tags placed on them, and the other of the friends who have tagged the most other friends (Figure 4).

Designing for Social Spread

Collabio relies on social mechanisms to spread to new users and retain existing ones. We crafted the application's design language to encourage increased activity and social spread. For example, the individual leaderboards are labeled "friends who know [this friend] best" to conflate closeness of friendship with score in the game, and notifications purposely do not share all the new tags to entice the user to visit the application (see below).

As with typical Facebook applications, users can explicitly invite others to play. More subtly, when a user tags a friend, the friend receives a Facebook notification of the action, whether or not that friend has previously played Collabio. This includes the user's name, the number of new tags, and a glimpse of the tags' contents:

> John Smith has tagged you with **cyclist** and 7 other tags using Collabio. Tag John back, or see what you've been tagged with. 2:41pm

A similar version appears on the tagger's wall feed and on the global Facebook news feed that all users see as their home page. Users can also place the occluded version of the tag cloud onto their Facebook profile. This demonstrates to visitors the number of tags the individual has acquired and serves as a hook for new users to install and play.

Dealing with Cheating and Abuse

Many games suffer from undesirable behavior such as cheating, collusion, or other malicious actions. Because Collabio activity can only occur between people with a mutually-established social connection, we rely exclusively on the mechanics of the social network to prevent this behavior.

There are several ways friends could conspire to increase their score. For example, they could ask the person whom they are tagging or their friends for the answers. They could also reverse engineer tags using a search strategy on the alphabetized cloud. This behavior does not do active harm to the tag cloud, as it simply reinforces already-existing tags. However, it does erode our premise that popular tags were generated by multiple independent sources. Fortunately, this is more work than just guessing at tags, and it is a poor method for drastically increasing one's score relative to everyone else's since mimicking friends' guesses simultaneously increases their scores as well. Another way to artificially increase one's score might be to tag a friend with a large number of nonsensical tags: e.g., *a*, *aa*, *aaa*, and so on. Each of these tags is worth one point. However, this strategy quickly deteriorates because this does not take advantage of the work others are doing to earn you points and one point becomes worth less and less as more users tag.

Users could also decide to tag an individual with an undesirable tag as a joke or punishment. Since a tag is not automatically revealed to other users until they guess it, the payoff for such a strategy is rather low and non-public, and we did not see much of this in practice. Furthermore, the tagged individual is likely to disapprove of and delete inappropriate tags on the *My Tags* page, thereby eliminating ill-gotten points or reward. We have also seen people apply social pressures to friends to discourage such behavior. As regards misspelled or otherwise inaccurate tags, we rely on users' self-interest in maintaining a well-manicured public profile [5]. In practice, we observed that mistakes are probably the most common reason for deleting tags.

RELATED WORK

Social tagging systems for organizing photos, bookmarks, files, and other items are widespread on the web [9, 18]. In the domain of people tagging, most applications have been tailored to maximize ease of use and entertainment rather than quality of tags. For example, iDescribe[2] and Compare People[3] allow users to place pre-defined descriptors on their friends. This assumes a small set of static tags and does not leverage the richness of knowledge in the network. Systems like Describe Me[4], Define Me[5], and Impressions[6] encourage users to author new tags. However, they allow authors to see and reapply existing tags, hence potentially skewing perception and reducing the actual richness of tags. They also keep author identities anonymous, which we believe leads to undesirable behavior since there is no real motivation to 'play nice'. Other social people-tagging applications such as Fringe Tagging [7] have targeted coworkers within an enterprise setting. Tagalag[7] is incorporated into e-mail clients to draw these advances into the Web 2.0 productivity space.

Our approach is inspired by prior work on Human Computation [27], which aims to obtain useful information for computers by enticing users to provide it. We believe that this work extends the design principles of Games with a Purpose (*e.g.*, the ESP Game [28]). Specifically, though Collabio utilizes game motivations such as point scores and leader boards, it leans just as heavily on social motivators such as social reciprocity, the practice of returning positive or negative actions in kind [10]. Rather than anonymously pairing random players to prevent cheating, we explicitly target users within established social groups to contribute data, relying on social accountability and profile management [5]. Finally, rather than gather information common to all web-enabled humans, we directly target information that is known and verifiable only by a small social group: information about a friend [23]. IBM's Dogear social bookmarking game shares several of these characteristics, though it is focused around web bookmarks [6].

We motivate our work by research exploring social networking profiles and impressions. Facebook users are confident that their profiles portray them in accurate and positive ways [14], and outside observers who view these profiles or personal web sites do form clear and accurate impressions of the author even with extremely small subsets of information [8, 22, 24]. Results show that combining personality ratings of outside observers with self-evaluations produces a more accurate picture than either the raters or the individual alone [24], suggesting that the integration of friends' impressions into our profiles may lead to more accurate portrayals. However, as Lampe *et al.* found, only about 59% of profile fields are completed [14].

Studies of contribution in online communities motivate several design decisions in Collabio. One danger is *social loafing:* users exhibiting little effort on a collective task because they believe others will participate instead [13]. However, individuals are likely to continue contributing when reminded of the uniqueness of their contributions, given specific, challenging goals, and helping groups similar to themselves [16, 19]. Thus, we may motivate a game challenging individuals' (potentially obscure) knowledge of members of their social group. Both active and loafing users can be motivated simply by comparing their activity to the median participation of the community [11], as in the kind of competition that Collabio has designed into its leaderboards. Loafing can also be overcome via opportunities for reciprocity toward other friends [20], motivating our Facebook notifications upon tagging a friend.

IMPLEMENTATION

The Collabio application interface is built as an AJAX-enabled ASP.NET web application, which calls a SQL Server-backed Windows Communication Foundation web service for data storage and querying.

EVALUATION

We evaluated Collabio's success in three ways. First, we gathered statistics of tags and application usage over time. Second, we conducted a survey on active Collabio users to investigate tag accuracy and the effectiveness of the application's social motivators. Finally, we recruited expert raters who had not used Collabio and who were outside the active Collabio users' social network to evaluate whether Collabio tags provided new knowledge beyond what was available on Facebook or the Internet.

Usage Statistics

We analyzed tag statistics collected between July 2008 and March 2009. In that time, Collabio has gathered 7,780 unique tags on 3,831 individuals in 29,307 tagging events. These tags were generated by 825 different users.

Each user in the system tagged an average of 5.8 other friends ($\sigma = 13.6$) with 6.1 tags each ($\sigma = 7.3$). Figure 5 presents two histograms reporting the number of tags provided and the number of individuals tagged by each user — these metrics follow a power law pattern common in studies of participation in online communities. The mean tag length is 8.3 characters ($\sigma = 5.2$). 5,867 tags (~75%) are a single word, and 1,913 tags (~25%) contain multiple words.

There is evidence that the social aspects of Collabio are important motivators of usage. Of the 396 Collabio users who both tagged friends and were tagged by friends, 244 (62%) tagged a friend only after having been tagged themselves first and thus receiving a notification about the tag and the application. Of the 244 users who became active contributors and tagged only after having been tagged by others, 179 (73%) of them began tagging after having been tagged by just one other person; 41 users did so after

[2] http://apps.facebook.com/idescribe
[3] http://apps.facebook.com/comparepeople
[4] http://apps.facebook.com/describeme
[5] http://apps.facebook.com/defineme
[6] http://apps.facebook.com/impression
[7] http://www.tagalag.com

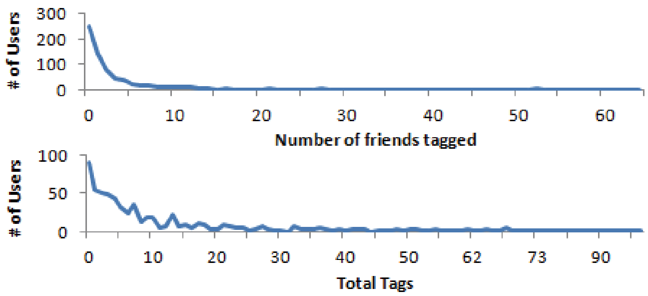

Figure 5. Histograms reporting Collabio tag trends. Top: the number of friends tagged by each user; Bottom: the number of tag instances produced by each user.

receiving tags from 2 friends, and 24 after 3 or more friends tagged them. Continued activity on Collabio also generally led to reciprocity. Only 43 (18%) of the above users failed to tag any of the friends that initially tagged them.

Not all tagging actions on Collabio convinced the recipient to install the application and begin tagging. Of 736 people who tagged friends using Collabio, 340 (46%) were never tagged back by any of their friends. Ignored invitations were common: 87% of individuals who were tagged never reciprocated by tagging anyone. Collabio saw a number of lurkers as well, 35% of people who installed the application never tagged a friend. We believe that iterative experiments aimed at exploring the influence of subtle changes in the design and signaling of Collabio on the above statistics could lead to a variant that would more rapidly penetrate the population of users.

Globally, the tags applied to the most individuals in Collabio are generic descriptors like *kind* and *smart* as well as affiliations such as *stanford*. Tags which tended to get deleted were typos, tags in negative spirit (e.g., *boring*, *arrogant*, *witch*, *blab*, *unreliable*, and *retarded*), humorous tags (e.g., *married and unavailable Ha!*, *Croatian sensation*, and *Stanford Sucks*), sensitive information (e.g., *sexxxy*, *S&M,* and *smoke break*), inaccurate information, and tags used as temporary guesses. Some tags were deleted without clear connotation, for example: *world travel*, *observant*, *yacht*, *trekkie*, *gentleman*, and *reasonable*.

User Survey: Verifying Quality of Tags

Having collected a large number of tags using Collabio, we set out to understand whether we had obtained accurate and novel information about individuals.

Survey Participants and Method

Using Facebook's notification service, we invited Collabio's active users in September 2008 to fill out a survey about their experience. We defined an active user as one having tagged at least three friends, having been tagged by at least three friends, and having at least nine distinct tags.

Forty-nine users (24 female) responded to the survey, a 44% response rate. The median age was 27 ($\sigma = 4.1$). The respondents were skewed toward students and researchers with an interest in user interfaces. We offered a small gratuity for responding.

The survey solicited comments on three major sets of topics. The first set focused on social elements surrounding the use of Collabio, including the efficacy of the social motivators and the nature of the tag network built in Collabio. Were friends tagged in Collabio closer on average than the typical Facebook friend? Which of the notifications, news feed items, and other mechanisms were most effective in encouraging users to tag friends? How interested were users in the tags others placed on them? How did the attribution of tags affect tagging behavior? The second set were tagging strategy questions, which sought to find the most common tagging strategies, as well as detect and understand surprising ones. The third set was aimed at examining the perceived quality of tags, that is, whether or not individuals felt tags were accurate descriptors of themselves.

Each user was asked to rate nine of the tags in their tag cloud. These tags were drawn from three buckets: *Popular Tags*, the top three tags applied by the most friends; *Middling Tag***s**, tags drawn randomly from the set of tags that occurred at least twice but less often than the Popular Tags; *Unique Tags*, tags drawn randomly from ones applied by only a single friend. For users who did not have enough tags to fill the Middling Tags category, we instead presented a randomly-generated string as a control condition. Twenty-nine of the participants had Middling Tags, twenty did not.

For each tag presented, the user provided a rating of agreement on a 7-point Likert scale (1 for disagreement and 7 for agreement) for each of two questions: "This is a good tag for me," and "This tag is something I would expect lots of people to know about me." In addition, participants classified each tag into the following categories: school, workplace or group affiliation; professional or academic interest, expertise or title; recreational hobby, interest, or expertise; location; personality trait; physical description; name or nickname; another person in the participant's life; inside joke; don't know; or other.

Survey Results

When asked about reasons they tagged someone, an overwhelming 81.6% of respondents cited personal notifications that the friend had tagged them first, suggesting that social reciprocity is a strong motivator in social tagging. 44.9% reported responding to Collabio's suggestions of friends to tag, 44.9% used tagging as a way to invite someone to join Collabio, 40.8% tagged as a result of a Collabio event on the Facebook News Feed, and 36.7% did it to move up on the friend's high score board. Only 16.3% of respondents reported tagging someone so they were the first, and a small number explicitly started tagging in order to try and reveal things that they did not know about someone.

Tag strategies were fairly consistent with expectations. All participants reported using their own knowledge of the friend to tag. Of secondary strategies, 67.3% of respondents used the number of characters in popular strings to guess, and 36.7% additionally relied on the alphabetical ordering.

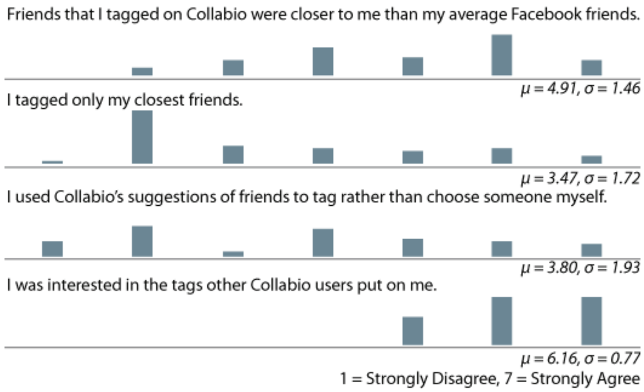

Figure 6. Collabio users were likely to tag their own friends, and were very interested in tags placed on them.

Surprisingly, no respondents reported colluding by asking other friends or asking the tagged individual for ideas.

Participants reported being very interested in the tags their friends placed on them (Figure 6). The highest number of respondents (34.7%) reported checking their tags multiple times per week, followed by once a week (20.4%), multiple times a month (14.3%), once a month (12.2%), then more frequently such as once a day (10.2%) and multiple times a day (8.2%).

Opinions were split on whether non-anonymous tagging affected the tags used, with 55% reporting that this did have an effect. We saw comments of the form: "Didn't want to hurt their feelings perhaps. Not that the tags were going to be nasty. But maybe just...true." or "I won't tag someone with something I wouldn't say to their face." and "I picked things that were inside jokes with my friends."

Relatively few participants (11.2%) reported deleting tags from their tag clouds. Those who did mainly reported removing spelling errors, though one did remove an ex-boyfriend's name.

Popular tags were mainly reported to be affiliations, and Middling Tags and Uncommon Tags were more commonly reported to capture interests, expertise and hobbies (Table 1). Thus, a large percentage of Collabio's tags are affiliations, interests, expertise and hobbies. The Uncommon Tags were commonly categorized as Miscellaneous, including clothing choices, special abilities, and the name of a friend's dog.

Participants rated all three classes of tags as accurate descriptors of themselves, and all but Uncommon Tags as known by many people (Table 2). We ran a one-way ANOVA on these results and found significant effects of tag bucket on goodness of tag ($F_{2,384}$=34.5, $p<0.001$, η^2=.15) and expectation that others know the given facts ($F_{2,384}$=67.1, $p<0.001$, η^2=.26). Pairwise posthoc comparisons using Bonferonni correction indicate that Popular Tags are more accurate than Middling Tags and Uncommon Tags, Popular Tags are more well-known than Middling Tags and Uncommon Tags, and Middling Tags are more accurate and well-known than Uncommon Tags. The popularity result reinforces our expectation that the number of friends guessing a tag aligns with how many people actually know the fact about the user.

Tag Bucket	Definition	Most Popular Information
Popular Tags N = 147	Three most popular tags for the user	School, workplace or group affiliation (66.0%) Interests or expertise (16.3%)
Middling Tags N = 93	Less popular than Popular Tags, but occurring more than once	School, workplace or group affiliation (27.2%) Interests or expertise (23.9%) Hobbies (15.2%) Location (10.9%)
Uncommon Tags N = 147	Occurred only once	Interests or expertise (21.1%) Miscellaneous (15.6%) School, workplace or group affiliation (13.6%) Hobbies (12.9%)

Table 1. Affiliation and interest categories were the most popular among Collabio tags.

	Popular Tags	Middling Tags	Uncommon Tags
Accurate	μ = 6.42 σ = 0.92	μ = 5.83 σ = 1.39	μ = 5.13 σ = 1.61
Widely known	μ = 6.22 σ = 1.22	μ = 5.21 σ = 1.58	μ = 4.14 σ = 1.77

Table 2. User ratings of how accurate and widely known the tag buckets were, on 7-point Likert scale (1=very inaccurate / not widely known, 7 = very accurate / widely known).

Tag Novelty: Rating Exercise

Our survey results suggested that Collabio generated accurate tags that were reasonably ordered by importance. However, if these tags are available elsewhere we have not significantly advanced the state of the art. Could an algorithm or individual outside the social network just as easily create these tags by mining information available in users' Facebook profiles or the web? Could these methods also reproduce the relative ordering of tags?

Rating Study Method

We first wished to determine whether Collabio tags were already available on users' Facebook profiles or the web. We randomly selected twenty respondents from the twenty-nine who completed our previous survey and had three non-random Middling Tags. For each survey respondent we utilized the nine tags they had rated in the survey, as well as three *Fake Tags* that were false and thus should not appear anywhere associated with the individual. Fake Tags were chosen from the set of global Collabio tags: one from the top 5% most popular tags, one that occurred less than the 5% most popular tags but more than once, and one that occurred only once. Fake tags excluded any tags applied to the individual.

We recruited four native English speakers comfortable with Facebook and web search, but who had never used Collabio and did not know any Collabio users, to serve as raters. We gave them a brief demonstration of Collabio. The raters' task was to find evidence for each tag on the user's Facebook profile and on the web. For each target individual, raters were presented with the twelve tags in random order and asked to rate each on a 7-point Likert scale according to the following statement: "I can find strong evidence that the tag applies to this individual." Raters were trained to give a score of 7 if the tag appeared verbatim, a score of 1 if there was no evidence in support of the tag, and a score of 4 if moderate inference was required based on the available evidence (e.g., the tag was *Atlanta* but the only relevant evidence was that the person attended Georgia Tech). Raters were trained on example tags and profile sets until satisfactory agreement on the scoring scale was achieved. We randomized the order that raters viewed individuals.

Raters rated the set of tags under two scenarios: first using only the individual's Facebook profile available to friends, and second using only web search. In the web search scenario, raters were disallowed from concatenating the individual's name and the tag name into a search query (e.g., "john smith atlanta"), in order to better simulate a tag generation task with no prior knowledge of the tag. We believe this is a more difficult test for Collabio to pass than that undertaken by Farrell et al. [7], who performed string comparisons to test whether tags existed on profiles, because human raters perform semantic inferences.

We also wanted to investigate whether our raters could determine how popular a tag had been, as verified by our survey data. For each individual, we asked raters to place each tag into its original bucket: Popular Tags, Middling Tags, Unpopular Tags, and Fake Tags. They were told that three tags came from each bucket.

Rating Study Results
Raters evaluated tag evidence on Facebook and the web for a total of 480 tags across the twenty individuals. Cronbach's alpha was calculated to measure agreement between the raters, producing an overall agreement score of .82.

Experts found more supporting evidence for the more popular tag buckets, both on Facebook and the web (Table 3). A two-factor ANOVA comparing the effect of tag bucket (Popular vs. Middling vs. Uncommon vs. Fake) and evidence type (Facebook vs. Web) on rating found a main effect of tag bucket ($F_{3,1915} = 270.0$, $p<0.001$, η^2=.30), and pairwise Tukey posthoc comparisons (all significant $p<0.001$) suggest that the more popular a tag was, the higher rating it received and thus the easier it was to find evidence for. We found no significant effect of Evidence type, and inspection suggests that the scores between Facebook and the web are nearly identical.

Raters were the most reliable at identifying the extreme buckets: Popular Tags and Fake Tags (Table 4). Raters had the poorest performance on Middling Tags and Uncommon Tags, correctly recognizing only about 40% of them.

Overall, raters found evidence supporting Popular Tags, but moderate inference was required for Middling Tags and very little evidence was available for Uncommon Tags. Our

	Popular Tags	Middling Tags	Uncommon Tags	Fake Tags
Facebook Evidence	$\mu = 5.54$ $\sigma = 2.36$	$\mu = 4.20$ $\sigma = 2.68$	$\mu = 2.87$ $\sigma = 2.56$	$\mu = 1.56$ $\sigma = 1.76$
Web Evidence	$\mu = 5.72$ $\sigma = 2.29$	$\mu = 4.17$ $\sigma = 2.81$	$\mu = 3.04$ $\sigma = 2.65$	$\mu = 1.5$ $\sigma = 1.4$

Table 3. Mean ratings applied to tags, from 1 (no evidence to support tag) to 7 (tag appeared verbatim).

		True Buckets			
		Popular	Middling	Uncommon	Fake
Rater Predictions	Popular	**151**	61	24	7
	Middling	63	**94**	50	30
	Uncommon	15	51	**103**	73
	Fake	11	34	63	**130**

Table 4. Confusion matrix of rater bucketing decisions. Raters were accurate at identifying Popular Tags and Fake Tags, but less so at Middling Tags and Singular Tags.

original survey respondents indicated that even Uncommon Tags were good descriptors of themselves, so we may conclude that Collabio is collecting accurate information with Middling and Uncommon Tags that would otherwise be difficult or impossible to acquire. Raters had considerable difficulty distinguishing Middling from Uncommon tags, and Uncommon from Fake Tags, so beyond the most obvious information it may also be difficult to recreate Collabio's tag ordering even coarsely.

PROOF-OF-CONCEPT APPLICATIONS

We believe that the Collabio tags enable a wide variety of interactive applications. In this section we describe three illustrative prototypes we built utilizing the Collabio database: a tag cloud aggregator for tag visualization and exploration, an expert-finding question answering system, and a personalized RSS feed.

Collabio Clouds

Our system has learned thousands of tag clouds for users, so it seems appropriate that we consider tools for making sense of the tag space. Collabio Clouds allows users to compare themselves and other users of the system.

Collabio Clouds (Figure 7) aggregates tag clouds based on user queries. The user can query his or her own tag cloud as well as the aggregated tag cloud of friends, Collabio users, users tagged with specific Collabio tags (like *tennis* or *Adobe*), or users in Facebook networks or groups. Collabio Clouds allows users to explore questions such as: What do the tag clouds of members of the Penn State network look like? What other tags show up on individuals tagged with *machine learning*? What tags are most popular amongst all my friends?

Collabio Clouds uses a comparison tag cloud technique developed by ManyEyes [25] to allow users to compare two groups. Thus, a user can compare his or her friends to all

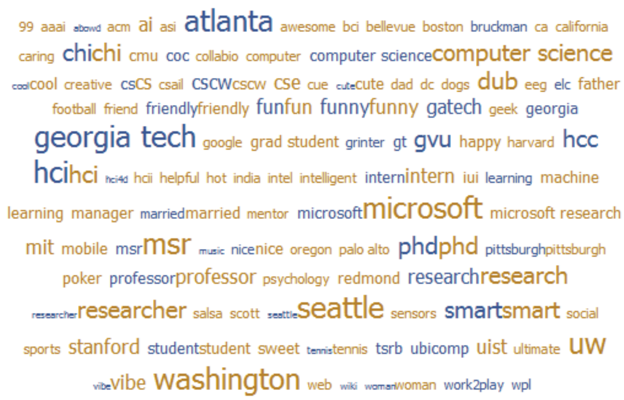

Figure 7. A tag cloud comparing users tagged with *washington* to users tagged with *georgia tech* in Collabio Clouds.

Collabio users, compare Michigan students to Michigan State students, compare people tagged with *football* to people tagged with *baseball*, or compare Stanford members of the ACM SIGCHI group to Carnegie Mellon members of the ACM SIGCHI group.

Tag clouds are aggregated by number of members of the group who have a tag, so larger tags are more common in the population. To improve privacy, only tags that are applied to more than one individual are shown in the aggregate tag cloud.

Collabio QnA

Once we can aggregate tags, it makes sense to match people to each other in an expert-finding system. Many Question and Answer (QnA) systems such as Yahoo! Answers[8] rely on a large community of answerers actively seeking out questions. Expert-finding algorithms can broaden the effectiveness of these tools by actively routing questions to users likely to know the answer. QnA systems with expert-finding components include Answer Garden [1] and Aardvark[9]; Farrel *et al* [7] suggested that tags could be used for people-ranking.

As we did with Collabio, we embedded the Collabio QnA system (Figure 8) in Facebook. Users ask questions, and Collabio QnA searches over the collected Collabio tags to identify friends and friends-of-friends who are most likely to be able to answer the question. The user can then choose which friends to send the question to, and Collabio QnA provides a comment board for the answer thread.

Collabio QnA's expert-finding algorithm utilizes the Lucene search engine[10]. Each user's tag cloud is translated into a document in the search engine with terms weighted by number of friends who applied the tag. The user's question is then fed as a query to the search engine, and the ranked results are restricted to the user's friends and friends-of-friends. Lucene's default scoring function prefers short documents – in this context, users with fewer tags – so

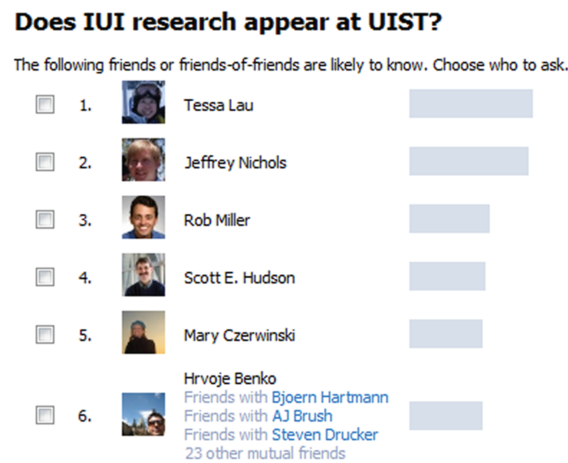

Figure 8. Collabio QnA is a question and answer system that uses Collabio tags to find friends and friends-of-friends who can answer your questions.

we utilize a length-independent scoring function to give all tag clouds equal scores regardless of size.

Collabio tags and the social network context provide the opportunity for our QnA system to route questions more highly relevant within the user's social network, such as *When is the next HCI group meeting?* Or *Who might be interested in starting an IM football team at Google?* These kinds of questions are difficult to answer using global QnA sites such as Yahoo! Answers.

Collabio RSS

Collabio QnA matches people to other people, but Collabio tags can also be used to match content to people. RSS (Really Simple Syndication) is a popular format allowing aggregation of web content, enabling users to subscribe to the feeds of web pages of interest. However, these feeds vary in relevance and can be overwhelming in number, making it difficult to identify the most relevant posts to read.

Collabio RSS is a personalized RSS feed of web content, powered by a user's Collabio tags. Collabio RSS builds on research in personalized web filtering (e.g., [3, 4]). It is unique from most content-based filtering algorithms in that its model is not implicitly learned; the tag knowledge base enables a simple information retrieval approach to filtering and enhances scrutability of its results [12, 26].

To produce the news feed, Collabio RSS indexes the title and text content of each feed item as a document in Lucene. When a user requests a personalized feed, it retrieves that user's Collabio tag cloud and performs a document-as-query search on the feed corpus: the weighted tag cloud is concatenated as an OR'ed search query and weighted by tag popularity. Tag weights are log-transformed to prevent the most popular tags from overwhelming the results. We filter the corpus using a sliding time window of the past day and a half to keep the feed's content fresh.

---

[8] http://answers.yahoo.com
[9] http://www.vark.com
[10] http://lucene.apache.org

To test the system, we crawled 2610 popular RSS feeds recommended as bundles by Google Reader[11], indexing 68,069 items posted over 36 hours. As an example, ten randomly-selected posts vary greatly in topic:

1. 2010 Pontiac Solstice GXP Coupe Test Drive: 28 MPG and Turbo Power, but Practicality Not So Much
2. The X-Files: Season 7: Disc 4
3. 26 Carriers Commit To Deploying LTE; Some Backers Look For Way To Make Voice Calls
4. 5 Reasons Why the PTR Sucks
5. 30 More Free Blog Icons, Website Icons, Symbol Icons
6. 2009 is the year of the comic book in Brussels
7. Superman Cartoons
8. 84th Precinct Crime Blotter
9. 1D-navigation using a scalar Kalman filter
10. 13 Tasteless Costumes Ever

However, Collabio RSS feed identifies items of much greater interest to one of the authors, containing items relevant to HCI, graduate school, and nerd culture in Boston:

1. Weekly Mashable Social Media & Web Event Guide
2. Job Offer: PhD Position in Information Visualization, Växjö University, Sweden, and TU Kaiserslautern, Germany
3. 6 Y Combinator Startups I Would Have Invested In Back Then
4. Mind Meld: Non-Genre Books for Genre Readers *[sci-fi books]*
5. Job: Postdoc in Visual Perception, Delft University
6. The Information School Phenomenon
7. Speaking of (and in) 2009 *[speaking schedule of HCI figure]*
8. Tonight: Video Game Orchestra at Berklee
9. Brain-Computer Interfaces: An international assessment of research and development trends
10. Exploring Siftables: the blocks that play back *[HCI research at author's university]*

The Collabio RSS feed has high relevance because Collabio collects so many tags related to professional and recreational interests. Affiliation-oriented tags, also popular, are responsible for returning news relevant to places and organizations the author has been associated with.

DISCUSSION AND FUTURE WORK

Collabio Collects Accurate and Novel Information

Tying together the survey and the rating exercise, we see that Popular Tags, which largely captured group affiliations, could in principle be generated by mining available information such as Facebook or the web, even though we know of no current system that can do this reliably. Middling Tags and Uncommon Tags, which users view as good descriptors of themselves, are difficult for others outside the social network to verify, and by extension, to generate. Thus, Collabio generates tags that are not available to typical mining methods and these tags cannot reliably be judged accurate by individuals outside the social network.

Collabio Succeeds In Tagging Active Users; Convincing New Users to Join Is Harder

We have learned that Collabio's social design is successful at tagging users who join the application. Users who tagged at least one other friend were well-tagged by their friends in return: the median such user received 11 unique tags. Social spread is a more difficult problem, and Collabio could be improved with knowledge gained through the design process. A large majority of tagged users never accepted the application invite, suggesting that the social incentives appropriate for notifying active users differ from those appropriate for invited users, that the Facebook population is largely wary of new applications following a period of intense application spam in 2007-2008, or both. The problem is exacerbated by active Collabio users who were hesitant to send invitations to their friends. One user reported: *"I'm reluctant to start tagging people that haven't added it. If they were already on Collabio, I'd probably tag [them]. Since they're not though, I feel like it'd be annoying if I started tagging them."*

We suggest that an exploration of designs for successful viral spread is thus an important avenue of future work. Neither the game incentive of extra points nor the social incentive of making individuals appear tagged was successful enough in Collabio's case. The result is that Collabio has stabilized at about 140 users per month. It is possible that publicly releasing our applications utilizing the tags would more strongly motivate users to participate.

Design Reflections

Social tagging games such as Collabio face design decisions that impact the number and nature of tags acquired. Figure 9 sketches our view of the design space and Collabio's place in it. In future work we hope to understand how the social processes and the data collected change as an effect of decisions such as tag anonymity and visibility.

Our experiences designing and iterating on Collabio may provide useful guidance to developers of future systems. Tagging activity attributed to the tagger leads taggers to avoid negative terms (e.g., *witch*) and encourages inside jokes. Social motivations such as notifications were most effective at encouraging players to begin tagging; game motivations such as points were most effective at motivating continued tagging once tagging had begun. Centering the game around static information meant that the tags did not become stale, but meant that players eventually ran out of ideas and then had less motivation to continue tagging.

Building Applications Using the Collabio Tags

Building Collabio Clouds, Collabio QnA, and Collabio RSS has given us some insight into techniques and challenges associated with mining the Collabio tags for end-user appli-

Tag anonymity	Anonymous	Public to tagged user only	Public to all users
Tag visibility	Never visible	Visible when guessed	Always visible
Game mechanics	None	Reward uncommon information	Reward common information
Bootstrapping untagged users	None	Game mechanism (e.g., extra points)	Social mechanism (e.g., pretending another user has tagged first)
Synchronity	Synchronous (e.g., ESP Game [28])		Asynchronous

Figure 9. The design space of social tagging applications. Collabio's design choices are highlighted in blue.

[11] http://reader.google.com

cations. Information retrieval techniques such as tf-idf are important means for normalizing out common tags such as *kind*, *beautiful,* and *nice*. Tag sparsity issues may have been expected, but we found that Collabio users typically tried several different versions of a single idea when tagging (e.g., *computer science*, *CS, comp sci*), so in practice this was not a major issue. In addition, the stemming that Lucene applies to the tags often hashes together different conjugations of a tag. If sparsity becomes an issue for applications in the future, item-based collaborative filtering (people tagged with one tag were often tagged with another) could be used to implicitly add likely tags.

We cannot distinguish the semantics of any given tag, so we cannot know if a tag is appropriate for a given personalization purpose. In future work, we will explore targeting the goal of the tagging activity more carefully in order to generate tags relevant to a particular application. We believe human users are best situated to make these hard semantic decisions, and we would like to leverage this fact. In addition, new tagging tasks might help keep the application fresh.

CONCLUSION

We have presented Collabio, a social network application that extracts latent personalizing information by encouraging friends to tag each other with descriptive terms. Collabio is novel in its design as a social people-tagging application in game form. The system has been successful in motivating players to tag almost 4,000 people with tags that are both accurate and contain information not available elsewhere. We have demonstrated that Collabio tags are useful tools for building visualization and personalization applications, especially those requiring social knowledge.

ACKNOLEDGMENTS

We thank *anonymized*.

REFERENCES

1. Ackerman, M. S. & Malone, T. W. 1990. Answer Garden: a tool for growing organizational memory. In *Proc. SIGOIS ’90*, 31-39.
2. Becerra-Fernandez, I. 2001. Searching for Experts with Expertise-Locator Knowledge Management Systems. In *Proc. ACL ’01: Workshop on Human Language Technology and Knowledge Management 1*. 9-16.
3. Billsus, D., Pazzani, M. J., & Chen, J. 2000. A learning agent for wireless news access. In *Proc. IUI '00*, 33-36.
4. Brusilovsky, P., Kobsa, A., & Nejdl, W., Eds. 2007. *The Adaptive Web*. Springer.
5. DiMicco, J. M. & Millen, D. R. 2007. Identity Management: Multiple Presentations of Self in Facebook. In *Proc. GROUP ’07*, 383-386.
6. Dugan, C., Muller, M., Millen, D. R., et al. 2007. The Dogear Game: A Social Bookmark Recommender System. In *Proc. GROUP ’07*, 387-390.
7. Farrell, S., Lau, T., Nusser, S., et al. 2007. Socially Augmenting Employee Profiles with People-Tagging. In *Proc. UIST ’07*, 91-100.
8. Gosling, S.D., Gaddis, S., & Vazire, S. 2007. Personality Impressions Based on Facebook Profiles. In *Proc. ICWSM 2007*.
9. Golder, S., and Huberman, B. A. 2005. The Structure of Collaborative Tagging Systems. HP Labs technical report.
10. Gouldner, A.W. 1960. The Norm of Reciprocity: A Preliminary Statement. *Sociological Review,* 25, 161-178.
11. Harper, F.M., Li, S.X., Chen, Y., & Konstan, J. 2007. Social Comparisons to Motivate Contributions to an Online Community. In *Proc. Persuasive 2007*, 148-159.
12. Kammerer, Y., Nairn, R., Pirolli, P., and Chi, E. 2009. Signpost for the Masses: Learning Effects in an Exploratory Social Tag Search Browser. In *Proc. CHI ’09*.
13. Karau, S. J., & Williams, K.D. 1993. Social Loafing: A Meta-Analytic Review and Theoretical Integration. *Journal of Personality and Social Psychology*, 65, 681–706.
14. Lampe, C. A., Ellison, N., and Steinfield, C. 2007. A Familiar Face(book): Profile Elements as Signals in an Online Social Network. In *Proc. CHI ’07*, 435-444.
15. Lin, C-Y., Ehrlich, K., Griffiths-Fisher, V., and Desforges, C. 2008. SmallBlue: People Mining for Expertise Search. *IEEE MultiMedia*, 15(1), 78-84.
16. Ling, K., Beenen, G., Ludford, P., et al. 2005. Using Social Psychology to Motivate Contributions to Online Communities. *Journal of Computer-Mediated Communication*, 10(4).
17. Kautz, H., Selman, B., & Shah, M. 1997. Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, 40(3), 63-65.
18. Marlow, C., Naaman, M., Boyd, D., and Davis, M. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proc. Hypertext ’06*, 31-40.
19. Rashid, A. M., Ling, K., Tassone, R. D., et al. 2006. Motivating Participation by Displaying the Value of Contribution. In *Proc. CHI ’06*, 955-958.
20. Resnick, P. & Zeckhauser, R. 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of e-Bay’s Reputation System. *Advances in Applied Microeconomics*, 11, 127–157.
21. Shami, N. S., Yuan, Y. C., Cosley, D., *et al.* 2007. That's What Friends are For: Facilitating 'Who Knows What' Across Group Boundaries. In *Proc. GROUP ’07*, 379-382.
22. Stecher, K. & Counts, S. 2008. Thin Slices of Online Profile Attributes. In *Proc. ICWSM '08*.
23. Toomim, M., Zhang, X., Fogarty, J., & Landay, J. A. 2008. Access control by testing for shared knowledge. In *Proc. CHI ’08*, 193-196.
24. Vazire, S., & Gosling, S.D. 2004. e-Perceptions: Personality Impressions based on Personal Websites*. Journal of Personality and Social Psychology*, 87(1), 123-132.
25. Viegas, F.B., Wattenberg, M., van Ham, F., *et al.* ManyEyes: a Site for Visualization at Internet Scale. *IEEE Trans. Viz. and Computer Graphics,* 13 (6), 1121-1128.
26. Vig, J., Sen, S., & Riedl, J. 2008. Tagsplanations: explaining recommendations using tags. In *Proc. IUI '09*, 47-56.
27. von Ahn, L. Games with a Purpose. 2006. *IEEE Computer*, 39(6), 92-94.
28. von Ahn, L. & Dabbish, L. 2004. Labeling Images with a Computer Game. In *Proc. CHI ‘04*, 319-326.