Collabio: Extracting Latent Information in Social Networks to Tag People

ABSTRACT
Applications such as personalization, ad targeting, and online expert matching benefit from having a rich understanding of a user's history, opinions, personality, interests, and expertise. While current techniques rely on behavioral observation, automated document mining, or even self-labeling, we assert that social networks contain vast amounts of latent information about people that could be leveraged for this purpose. In this paper, we present Collabio, an application embedded in a social network that extracts this latent information by encouraging friends to tag each other with descriptive terms. Collabio is a simple game that leverages properties of the social network such as competition and social accountability in order to elicit a wide range of useful tags. To evaluate the efficacy of the approach, we examine usage statistics, ask users to verify the quality of their tags, and explore how Collabio tags augment ones that could have been generated through traditional online scraping techniques.

Author Keywords
Social computing, human computation, tagging.

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

INTRODUCTION
In a world of increasingly personalized computing, there is large opportunity for systems to move beyond treating all users uniformly and to begin building a rich understanding of individuals in order to better serve their needs. This personalized behavior may rely on knowledge about topics such as user history, opinions, personality, interests, and expertise. For example, news sites could filter irrelevant items to alleviate the information deluge, a search engine could offer up the "CHI" conference website more prominently than pages about the Chinese concept of spiritual energy, and online merchants could recommend appropriate products. Additionally, with such understanding, applications may assist in connecting people with complementary expertise, interests, or desires in order to facilitate task completion and social interaction.

There are several approaches to gathering user information to build rich user models. One could solicit the information directly from users themselves (e.g., Facebook's profile model). Unfortunately, this approach tends to be tedious and many users are not willing to expend the effort required when the benefit is not immediate, even when they are able. One could also try to automatically infer the information by mining the web, personal communication, or other documents (e.g., [10, 12]). However, data mining is non-trivial and often requires semantic interpretation that automated processes cannot yet provide. Alternatively, one could use collaborative filtering to derive likely preferences for individual users (e.g., Amazon.com or Netflix recommendations). These methods assume some amount of homogeneity between users, or at least presuppose that the system can build an adequate model for extracting the right information for each user, and suffer from bootstrapping issues.

A much less explored approach lies in leveraging latent knowledge that is distributed within social networks of

Figure 1. The Collabio tagging page. The user has guessed a series of tags on John Smith, including *band, ohio* and *vegas*.

friends and acquaintances. Whereas describing yourself can be a burdensome activity, describing friends can be engaging and socially enjoyable. Friends' impressions, when combined with self-report, have also been shown to paint a very accurate picture of an individual [17]. These tags provide access to information available with other approaches, but more importantly, to unique information that would be difficult to otherwise acquire. For example, friends within the network know your personality, your artistic and musical tastes, topics of importance to you, quirky habits, and so on, sometimes even better than you do yourself.

In this paper, we present Collabio (Figure 1), a game that elicits descriptive tags for individuals within the Facebook social network. The application leverages specific properties of the social network to motivate behavior and generate the information we desire. Specifically, Collabio motivates engagement through providing a feeling of social connectedness when used (people enjoy sending and receiving tags from friends), as well as fostering friendly competition through its scoring system. It also encourages accurate tags and discourages undesirable behavior by enforcing author attribution and leveraging social accountability. We evaluate the efficacy of the approach by examining usage statistics, having users verify the quality of their tags, and exploring how Collabio tags augment ones that might have been generated through traditional online scraping techniques. We find that the knowledge generated is relatively accurate and often unavailable through other means. To try Collabio, go to http://apps.facebook.com/collabio, log onto your Facebook account, and start playing.

RELATED WORK

Collabio aims to supplement existing methods of learning descriptive words and phrases about users. One common approach to this problem is to mine a user's personal information and communication such as e-mail, chat, personal web sites or publications (e.g., [1, 10, 12]). However, such automated systems suffer from false positives and possible difficulty disambiguating named references in documents. Other systems require the user to explicitly indicate interests or expertise, for example the New York Times' newsletter settings[1] and expert matching system K-Net [15]. Manual elicitation has obvious time and effort costs, however, and upkeep becomes difficult.

Social tagging applications allow users to use tags to describe their friends. Since most of these applications are deployed primarily for entertainment, they have been tailored to maximize ease of use and engagement rather than quality of tags. iDescribe[2] and Compare People[3] allow users to place pre-defined descriptors on their friends. Unfortunately, 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 could lead to undesirable behavior since there is no real motivation to 'play nice'.

Other social people-tagging applications such as Fringe Tagging [4] 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. Collabio distinguishes itself by its integration into a mainstream social network and its attempt to solve the tag motivation and tag accuracy problems within the social framework.

We believe that our work occupies a unique point in the social tagging design space. Collabio is the first application we know of to explicitly integrate game elements and to tactically hide unguessed tags from users in a bid to get pseudo-independent verification of the tags' accuracy. To further encourage desirable user behavior in generating useful tags, we also intentionally create social accountability by attributing all tags to authors, and making this information available to the person being tagged. Additionally, this work is the first we know to directly evaluate the quality of tags accumulated through social means.

Our work shares much in common with Human Computation [18], which aims to obtain useful information for computers by enticing users to share it willingly. However, we extend the design principles of these Games with a Purpose (e.g., [19]). 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 [6]. Rather than pair random players to prevent cheating, we explicitly target users who are part of the same social groups to contribute data. We hypothesized that this would alleviate collusion and cheating, because social motivators such as social accountability and the desire to maintain a pleasing public profile cause users to counterbalance these forces. 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. IBM's dogear social bookmarking game shares several of these characteristics [3].

We motivate our approach through recent work exploring social networking profiles and impressions. Facebook users are confident that their profiles portray them in accurate and positive ways [9], and outside observers who view these profiles or personal web sites do tend to form clear and ac-

[1] http://www.nytimes.com/mem/email.html#nl
[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

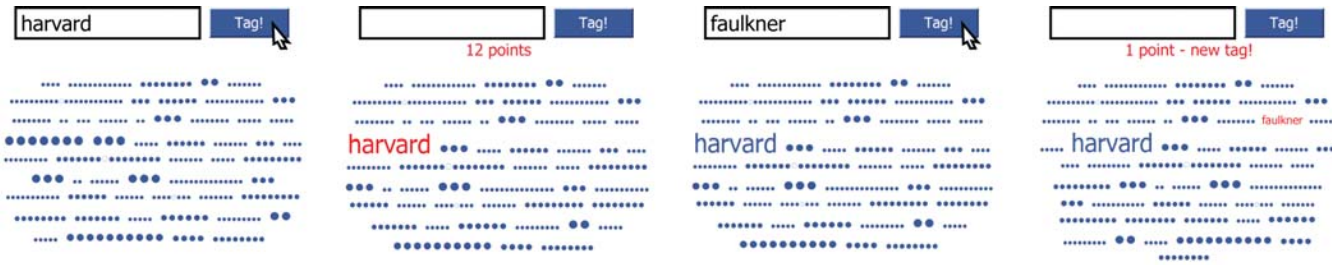


Figure 2. The tag cloud begins completely obscured. The player guesses *harvard*, receives 12 points for agreeing with eleven other players and reveals Harvard as a large tag. *Faulkner* is next; it does not match existing tags and is added to the cloud.

curate impressions of the author's personality even when given an extremely small subset of information [5, 16, 17]. 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 [17], which suggests that the integration of friends' impressions into our profiles may lead to more accurate portrayals. Moving beyond personality and into hobbies and interests, however, the above results may suffer. Further, there is data scarcity: in a dataset of 30,773 Facebook profiles, Lampe et al. found that on average only 59% of profile fields are completed [9].

Studies of contribution in collaborative enterprises such as Collabio have received much attention. One danger is *social loafing:* users exhibiting little effort on a collective task because they believe others will participate instead [8]. 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 [11, 13]. Thus, we may motivate a game challenging individuals' (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 [7], 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 [14], motivating our Facebook notifications upon tagging a friend.

COLLABIO

Collabio is currently embedded in the Facebook network (Figure 1). In this section, we describe the application as it stands and discuss design decisions we made, as well as the implications associated with each of them. We divide the main description into the three top level tabs that existed within the interface: 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 begins by finding a friend to tag, either by choosing one of Collabio's suggested friends, manually selecting a friend, or allowing Collabio to find a friend at random. Because we only peripherally explored priming behaviors based on how we recommended friends to tag, studying this in more detail remains future work.

In the right half of the "Tag!" page, the user sees the tag cloud others have created by tagging the selected friend (Figure 1). 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 black placeholder circle, so the tag *HCI*, for example, appears as ●●●. Spaces in obscured tags are 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 to help users guess at tags, terms in the tag cloud are alphabetically ordered. Items in the tag cloud are scaled in size by the number of people who have used the tag to describe the friend. The largest tags are the most popular descriptors of the friend.

If the user is the first to tag this person, 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.

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. Early pilot tests showed that users sometimes made mistakes, or tentative guesses they wanted to be able to retract, so we provide a simple interface for deleting a tag that they have used.

Users receive points for the number of total people who have applied a tag they have guessed. If they are the only person to have guessed that tag, then they get 1 point; if there are 14 others, they get 15 points. These points continue to accumulate as more people apply the tag. Points are presented in a table below the cloud so players can track their points as they grow. There is no ordering dependency in this system; in other words, the point score for the first

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

user to enter a tag rises as each subsequent user enters the same tag.

We had briefly explored providing additional incentive for being the first to tag someone by providing extra points, but we replaced this with the tag cloud seeding because it was not immediately successful at motivating users to do so. In fact we prototyped several point schemes, including an inverted point scheme that richly rewarded tags that only one or two other players had suggested and gave fewer points to popular tags, a scheme whereby a new tag received zero points instead of one point, and one that required users to expend points to reveal their own tag clouds, thus causing users to tag others in order to see their own descriptions. Systematically examining the effects of the different point schemes remains future work.

To expose one's score to others, and to further stimulate competition, each friend has a "People who know [this friend] best" pane, which lists friends who have earned the largest number of points from tagging the friend. If the user knows these people, they can click on them to tag them.

Managing "My Tags"

The "My Tags" interface allows users to inspect and manage tags their friends have placed on them. This 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. Because author attribution is completely exposed to the person being tagged, they can also follow-up with friends as appropriate.

Tagging anonymity is an ongoing debate. Some users report wanting anonymous tags so they can be more honest. However, given our goals (tags useful for interactive applications), the likely classes of anonymous tags (negative or hurtful), and the social upsides of accountability (e.g., inside jokes and messaging, as well as more engaging notification messages), we chose to make tag sources available to the tagged individual. In future work, we aim to compare

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.

the types and quality of tags that we attain with anonymous and attributed systems.

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). We hope to explore effects of these boards in future work.

Collabio Propagates itself through Social Spread

Collabio relies mainly on social mechanisms to spread to new users and retain existing ones. Some of these mechanisms are available through Facebook and others are built into the design of the application.

The application's design language encourages 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 to view them.

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:

> John Smith tagged Tom Anderson, Linda Williams, and Elizabeth Jones using Collabio. 2:41pm
>
> Tom has 8 new tags thanks to Douglas. Play to find out what they are!

Furthermore, users can place the occluded version of the tag cloud onto their Facebook profile. This serves as an advertisement of the application to anyone that visits and demonstrates to visitors the number of tags the individual has acquired from friends. It serves as a hook and provides a link from which new users can install and play.

Dealing with Cheating and Abuse

Many games and social applications suffer from undesirable gaming behavior such as cheating, collusion, or abuse of the system to perform malicious actions. We aimed to design Collabio in such a way so as to mitigate the need for explicit system controls to prevent this behavior. Instead we rely exclusively on the mechanics of the social network. This is reasonable mainly because Collabio users can only operate on friends that have mutually accepted the social connection.

There are several possible ways friends could potentially conspire to increase their score. For example, they could ask the person they are tagging for the answers, or ask other friends for the tags they have used. They could also reverse engineer tags using a binary 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 would simultaneously increase their scores as well.

Another way to artificially increase one's score is 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, as more friends play Collabio, this strategy quickly deteriorates in effectiveness because one point becomes worth less and less. Furthermore, the tagged individual is likely to disapprove of such activities on his or her tag cloud and can apply social pressures to discourage the behavior. The tagged individual can also easily delete the tags on the "My Tags" page, thereby eliminating the ill-gotten points associated with them.

Users could also decide to tag an individual with an undesirable tag as a joke or punishment. Collabio controls this situation most directly by allowing a tagged individual to easily delete tags from their own clouds. Since the offending individual's name is associated with the tag, there may be social repercussions for such activity, such as retaliation in kind, or "defriending" on Facebook. Additionally, 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.

As regards misspelled or otherwise inaccurate tags, we rely on users' self-interest in maintaining a well-manicured public profile, similar to as reported by DiMicco [2]. We hypothesized that users would not want to risk players accidentally discovering that others had voted for incorrect or undesirable tags as they play. In practice, we observed that mistakes are probably the highest cause of deleting tags. A scheme in which multiple users have to agree before something appears on the cloud might alleviate this.

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. This is not unusual for Facebook applications of this form.

We tracked overall user activity by updating a "last activity" timestamp for each user on any visit to a Collabio tagging page. We also recorded each unique tag action in its own timestamped record. To preserve privacy, if a user chose to delete a tag, those tag records were deleted permanently from the database. We did not keep a permanent record of deletions. Tag cloud seeding is accomplished by using the Facebook API to query common profile fields. Returned fields are tokenized by commas (a common Facebook separator field) and filtered to only include tags of a reasonable length (12 characters or less).

EVALUATION

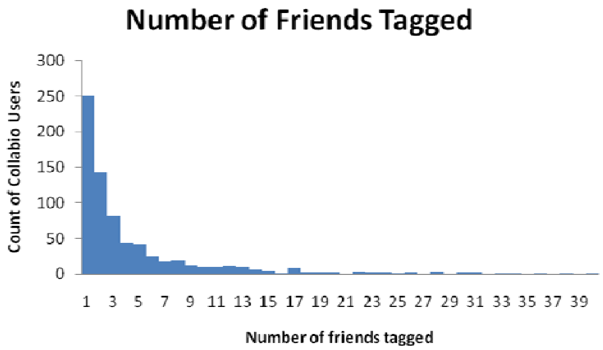
In this section we evaluate Collabio's success via three investigations. First, we report statistics of tags gathered and active usage over time. Second, we carry out a survey on active Collabio users to investigate tag accuracy and the effectiveness of the application's social motivators. Finally, we recruit expert raters who have not used Collabio and who are 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

Collabio has been publicly available to anyone with a Facebook account for eight weeks as of September 12, 2008. In that time, it has collected 6,991 unique tags on 2,909 individuals in 25,977 individual tagging events. These tags were generated by 736 different users.

Through the last eight weeks, Collabio has seen several stages of growth. Initially, we pilot tested the application internally in our research group. A week or so later, we began spreading it to outside contacts. About four weeks after launch, the application saw press on a major technology blog, which brought a new spike in usage. While there has been drop-off after each of these spikes, as we lost users who were not truly interested in sustained usage of the application, we continue to see a relatively steady set of users who visit and continue to tag daily.

Looking at the usage in more detail, each user in the system tagged 5.8 other friends ($\sigma = 13.6$) with 6.1 tags each ($\sigma = 7.3$), on average. Figure 5 presents two histograms reporting the number of tags provided and the number of individuals tagged by each user. The mean tag length is 8.3 characters ($\sigma = 5.2$). 5,263 tags (~75%) are a single word, and 1,728 tags (~25%) contain multiple words.



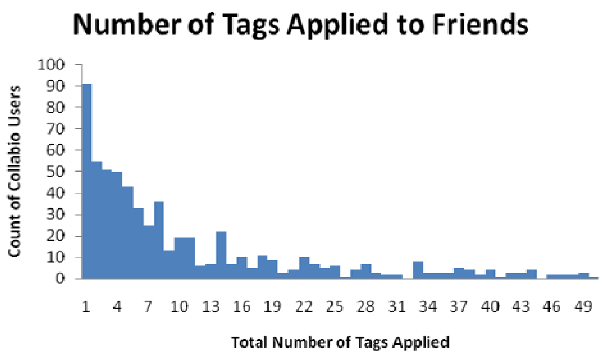


Figure 5. Histograms reporting Collabio tag trends. Top: the number of friends tagged by each user; Bottom: the number of tag instances produced by each user.

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. The remaining 152 began tagging before having been tagged. 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 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. Figure 6 demonstrates an example of this phenomenon, wherein a user tagged multiple friends but the tagging was never reciprocated.

Collabio saw a number of lurkers as well. 35% of people who installed the application never tagged a friend: this activity suggests either an interest in viewing one's own tags, or in simply exploring the application.

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. We hoped to move

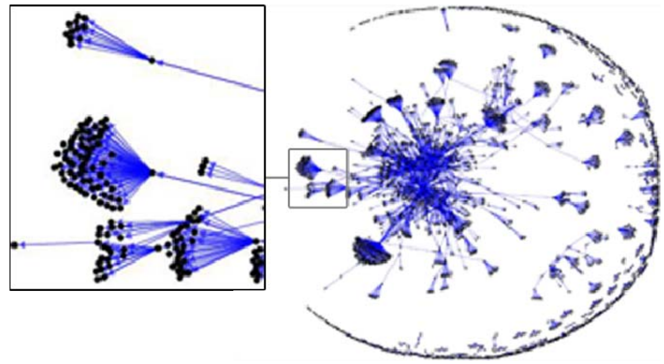


Figure 6. Right: the Collabio social network graph. Left: Callout focusing on Collabio users who tagged many friends but were rarely reciprocated.

beyond the anecdotal feedback we continued to receive and get slightly more formal measures of user opinion.

Survey Participants and Method

Using Facebook's notification service, we invited 112 of Collabio's most active users to fill out a survey on 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 in their cloud.

Forty-nine users (24 female) responded to the survey. The median age was 27 ($\sigma = 4.1$) and 33 of the participants had completed a graduate degree. This is a slightly older skew than the typical Facebook demographic. As an application early in stages of social spread, Collabio's user base was skewed toward a demographic similar to that of the authors: highly educated postgraduates 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 being 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 Tags*, 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 representing strong disagreement and 7, strong 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 specified the nature of each tag according to 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

The results provide mixed support for the hypothesis that Collabio users mainly tagged close friends. Participants were typically neutral or in support of a claim that friends tagged in Collabio were closer than average Facebook friends ($\mu = 4.91$, $\sigma = 1.46$), but dismissed a claim that they exclusively tagged their closest friends ($\mu = 3.47$, $\sigma = 1.72$).

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 factor in these networks. 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 stated 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. Surprisingly, no respondents reported colluding by asking other friends or asking the tagged individual for ideas. In freeform responses, a small number of participants reported trying to think about what other taggers would have used and tried reflecting tags that the friend had used on them.

Participants reported being very interested in the tags their friends placed on them ($\mu = 6.16$, $\sigma = 0.77$). In practice, participants varied in how often they checked in on their tags. 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 reported removing an ex-boyfriend's name.

Popular Tags (N=147) were mainly reported to be a school, workplace, or group affiliation (66.0%). Professional or academic interests, expertise or title were rated second (16.3%) and locations were rated third (8.8%). *Middling Tags* (N=93) seemed more diverse than the Popular Tags: School, workplace or group affiliation (27.2%) and professional or academic interests, expertise or title (23.9%) were the most frequent categories. Recreational hobbies (15.2%) and location (10.9%) were also present, as well as a scattering of other categories. *Uncommon Tags* (N=147) came from an even wider variety of categories. Professional or academic interest, expertise or title was the most frequent selection (21.1%), followed by the catch-all "other" category (15.6%, including such descriptions as "regular clothing choice," "website I frequent," "special ability," and "advisor's dog"), school, workplace or group affiliation (13.6%), recreational hobby, interest or expertise (12.9%), location (9.5%), inside joke (8.8%), personality trait (6.1%), another person in participants' lives (5.4%), and names or nicknames (5.4%). Every category had at least one vote.

When asked whether each tag was a good tag for the participant, Popular Tags were rated highly with a mean score of 6.42 out of 7 ($\sigma = 0.92$). Unsuprisingly, participants also rated Popular Tags as likely that many others would know about them ($\mu = 6.22$, $\sigma = 1.22$). Middling Tags' goodness was rated with a mean score of 5.83 ($\sigma = 1.39$) and likeliness that many friends would know about them was rated as 5.21 ($\sigma = 1.58$). Uncommon Tags' goodness was rated with a mean score of 5.13 out of 7 ($\sigma = 1.61$) and likeliness that many friends would know about them was rated near neutral at 4.14 ($\sigma = 1.77$).

We ran a one-way ANOVA on these results and found significant differences in Popular Tags, Middling Tags, and Uncommon Tags both in goodness of tag ($F(2,384)=34.5$, $p<0.001$) and expectations that others know the given facts ($F(2,384)=67.1$, $p<0.001$). Pairwise posthoc comparisons using Bonferonni correction indicate that Popular Tags are reported as better significantly better than Middling Tags ($p<0.001$) and Uncommon Tags ($p<0.001$), and Middling Tags than Uncommon Tags ($p<0.001$).

Tag Novelty: Expert Rating

Our survey results suggested that Collabio generated accurate tags that were reasonably ordered by importance. However, we also wanted to know whether these tags are *novel*. Could an algorithm or individual outside the social network just as easily create these tags based on available information such as the users' Facebook profiles, mining the web,

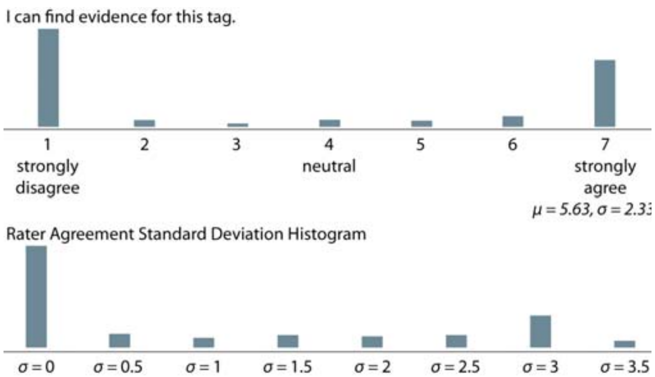


Figure 7. Top: Raters could usually find strong evidence for a tag, or none whatsoever. Bottom: On some tags one rater disagreed strongly, as evidenced by the spike in standard deviation at σ = 3.

or mining users' files? Could these methods also reproduce the relative ordering of tags generated by Collabio?

Rating Study Method

To determine whether Collabio tags are readily available through alternative sources, we recruited four native English speakers comfortable with Facebook and web search to serve as expert raters. We chose a subset of twenty respondents from the forty nine who completed our previous survey as test cases and used the nine tags these respondents had previously rated. In addition, for each survey respondent we randomly chose three additional *Fake Tags* from the global set of tags not applied to the individual: one tag from the top 5% most popular tags on all Collabio users, one tag that occurred less than the 5% most popular tags but more than once, and one tag that occurred only once.

For each individual, raters were given the twelve tags in random order and asked to rate each tag 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 before beginning the task, communicating vocally with each other and the experimenter until satisfactory agreement on the scoring scale was achieved. We randomized the order raters viewed individuals to compensate for learning effects on tag semantics.

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 Facebook scenario, we provided cached copies of the individual's Facebook wall tab, information tab, and photos tab. In the web search scenario, raters were allowed to use their favorite search engine to find information. They were, however, disallowed from concatenating the individual's name and the tag name into a search query (e.g., "john smith atlanta"). This restriction was imposed because we were interested in comparing Collabio to automated methods that generate tags (rather than verify them once given the term), and such methods would not be able to perform a concatenated web search without knowing the tag *a priori*. Raters were allowed to use any other information from the Facebook part of the rating task to help locate relevant web pages.

We believe this evaluation is a more difficult test for Collabio to pass than that undertaken by Farrell et al. [4], who performed string comparisons to test whether tags existed on profiles. Specifically, human raters perform semantic inferences and provide a metric which emulates what a "perfect AI system" might be able to do in the future. Literal string matching can be viewed as a lower bound on such a system's capabilities.

After completing these rating tasks for an individual, the raters were asked to attempt and reproduce the original buckets the tags were drawn from: Popular Tags, Middling Tags, Unpopular Tags, and Fake Tags. They were told that three tags came from each bucket, and their task was to reassign each tag to the correct bucket. The purpose of this exercise was to determine how well an algorithm or individual outside the social network could reproduce the relative ordering of tags as determined by Collabio and verified by our survey data.

Rating Study Results

Raters evaluated tag evidence on Facebook and the web for a total of 480 tags across the twenty individuals. Figure 7 presents a histogram of rater scores, highlighting raters' opinion that tags were either obviously present or conspicuously missing from the information available on the individuals; the latter was more often the case. Cronbach's alpha was calculated to measure agreement between the raters, producing an overall agreement score of .82.

Specifically investigating the source of discrepancies, a histogram of the standard deviation between the four raters on each tag (Figure 7) suggests that most tags elicited no variance in scores amongst raters. However, a subset of 77 tags produced standard deviations of 3.0, corresponding most commonly to instances when three raters agreed on one extreme score and one rater produced the far opposite score – e.g., three raters chose a score of 7 and one rater chose a score of 1.

Disagreements on scores fell into three major categories:

- Information seeking discrepancies: information was available on the page but one rater did not find it, or one rater interpreted evidence in support of the tag but others did not.
- Background knowledge: experts had different background information, for example that *Schaumburg* (a tag) is near Chicago (a city listed on the profile).
- Tag interpretation: experts interpreted ambiguous tags in different ways, for example *still alive* could either

be a song title (for which there is no evidence) or the state of being alive (for which there is considerable evidence).

More popular tags were generally rated as having more supporting evidence both on Facebook and the web (Table 1). 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 ($F(3,1912)=269.8$, $p<0.001$), suggesting that the more popular tags received higher ratings and were thus easier to find evidence for. We found no significant effect of Evidence type, and while we cannot draw any firm conclusions from this, inspection reveals the scores between using Facebook only and the web only are nearly identical.

Raters were the most reliable at identifying the extremes: Popular Tags and Fake Tags (Table 2). In Table 3, *precision*, the percentage of true Popular Tags that raters identified as Popular Tags rather than placing in other buckets, was 62.1%. *Recall*, the percentage of tags that raters bucketed as Popular Tags that were in fact Popular Tags rather than other tag types, was 62.9%. 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 and Middling Tags, but not much evidence supporting Uncommon Tags or Fake Tags. In fact, raters had considerable difficulty distinguishing Uncommon Tags and Fake Tags from each other, even though participants in the survey indicated that Uncommon Tags were good descriptors of themselves. This suggests that the tags were unlikely to have been generated by automated means, and are relatively unique to the social network approach.

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

Table 1. Mean ratings our expert raters applied to tags given Facebook and web search. The more popular the tag in Collabio, the more evidence our raters found.

		True Buckets			
		Popular	Middling	Uncom-com-mon	Fake
Rater Pre-dictions	Popular	**151**	61	24	7
	Middling	63	**94**	50	30
	Uncom-mon	15	51	**103**	73
	Fake	11	34	63	**130**

Table 2. Confusion matrix of rater bucketing decisions when compared to actual buckets. Correct answers are bolded. Raters were accurate at classifying Popular Tags and Fake Tags, and less so at identifying Middling Tags and Singular Tags.

	Popular	Middling	Uncom-mon	Fake
Precision	0.62	0.40	0.43	0.55
Recall	0.63	0.39	0.43	0.54

Table 3. Precision/recall values for rater bucketing, an alternate view onto Table 2. Precision represents what percentage of all tags in a bucket raters identified as so; recall represents what percentage of tags that raters identified as belonging to a

DISCUSSION

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.

More interestingly, 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 seems to generate tags that are not available to typical mining methods, and furthermore, these tags cannot even reliably be judged as accurate by individuals outside the social network.

Our expert raters informally corroborated this ability to understand only affiliations and high-level interests. One rater, when asked how well she knew each individual after looking at their Facebook profiles and information on the web, explained, "I get an idea of what their work is; I don't know who they are as a person."

Collabio has been very successful in spreading through the social network with a relatively small number of starting seeds. However, at time of writing, its use is far from widespread. We can attribute this result to several factors extrinsic to the design of Collabio. First, Facebook launched a new API and a design downplaying application presence in the middle of our rollout, causing some churn. Second, many users distrust Facebook applications as they fight off a constant deluge of irrelevant requests. We have observed that those not yet part of the application are hesitant to join until a critical mass of their friends have.

Users repeatedly expressed an aversion to tagging someone who had not yet installed the application, hampering viral spread. Collabio also has a population of lurkers who wish to be tagged but do not tag themselves. In future work, we will explore methods of making the activity even more compelling, and also how we can translate the desires of lurkers into actions beneficial to the application.

Like many naturally-occurring phenomenon on the web, Collabio generates tag occurrences according to an approximate power distribution. Among the most popular tags are general positive descriptors such as *smart* (the single most popular tag in Collabio), *funny*, *fun, friendly*, and *awesome.*

At the far end of the distribution are unusual tags such as *space cheetah*, *haxx0r*, and *SUPER DUPER FUN!!!!* Such tags suggest that Collabio will also produce data which may not be of immediate use to applications, some because they are too general, and some because they are not understandable to those outside the individual's social group.

We imagine that we would have to target the goal of the tagging activity more carefully in order to generate tags relevant to a particular application. For example, if we are interested specifically in expertise matching or in personalizing search, we would want users to focus on generating specific kinds of tags. Human users are best situated to make these hard semantic decisions, and we would like to leverage this fact. Building an application that uses these tags directly and adjusting Collabio to generate just the right set of tags remains future work.

CONCLUSION

In this paper we have presented Collabio, a social network application that extracts latent information within the network by encouraging friends to tag each other with descriptive terms. We have described the specific design decisions that we have made as well as the implications many of them have on applications such as ours. We have also shown the relative success of the application in terms of spread and usage, but also in terms of the type and quality of the tags that the application has harvested. We believe that there is great potential in this approach and have presented future work that remains, largely in further exploring the systematic manipulations that can be made to elicit various social behaviors, but also in designing and building the applications that take advantage of the data generated.

ACKNOLEDGMENTS

We thank *anonymized*.

REFERENCES

1. 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.
2. DiMicco, J. M. & Millen, D. R. 2007. Identity Management: Multiple Presentations of Self in Facebook. In *Proc. GROUP '07*, 383-386.
3. Dugan, C., Muller, M., Millen, D. R., Geyer, W., Brownholtz, B., & Moore, M. 2007. The Dogear Game: A Social Bookmark Recommender System. In *Proc. GROUP '07*, 387-390.
4. Farrell, S., Lau, T., Nusser, S., Wilcox, E., & Muller, M. 2007. Socially Augmenting Employee Profiles with People-Tagging. In *Proc. UIST '07*, 91-100.
5. Gosling, S.D., Gaddis, S., & Vazire, S. 2007. Personality Impressions Based on Facebook Profiles. In *Proc. ICWSM 2007*.
6. Gouldner, A.W. 1960. The Norm of Reciprocity: A Preliminary Statement. *Sociological Review,* 25, 161-178.
7. 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.
8. 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.
9. 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.
10. Lin, C-Y., Ehrlich, K., Griffiths-Fisher, V., and Desforges, C. 2008. SmallBlue: People Mining for Expertise Search. *IEEE MultiMedia*, 15(1), 78-84.
11. Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., et al. 2005. Using Social Psychology to Motivate Contributions to Online Communities. *Journal of Computer-Mediated Communication*, 10(4), article 10.
12. Kautz, H., Selman, B., & Shah, M. 1997. Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, 40(3), 63-65.
13. Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., & Riedl, J. 2006. Motivating Participation by Displaying the Value of Contribution. In *Proc. CHI '06*, 955-958.
14. 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.
15. Shami, N. S., Yuan, Y. C., Cosley, D., Xia, L., & Gay, G. 2007. That's What Friends are For Facilitating 'Who Knows What' Across Group Boundaries. In *Proc. GROUP '07*, 379-382.
16. Stecher, K. & Counts, S. 2008. Thin Slices of Online Profile Attributes. In *Proc. ICWSM '08*.
17. Vazire, S., & Gosling, S.D. 2004. e-Perceptions: Personality Impressions based on Personal Websites. *Journal of Personality and Social Psychology*, 87(1), 123-132.
18. von Ahn, L. Games with a Purpose. 2006. *IEEE Computer*, 39(6), 92-94.
19. von Ahn, L. & Dabbish, L. 2004. Labeling Images with a Computer Game. In *Proc. CHI '04*, 319-326.