Guess Who? Enriching the Social Graph through a Crowdsourcing Game

Ido Guy[*], **Adam Perer**[**], **Tal Daniel**[*], **Ohad Greenshpan**[***], **Itai Turbahn**[****]

[*]IBM Research-Haifa, Mt. Carmel, Haifa 31905, Israel   {ido, taldan}@il.ibm.com
[**]IBM Research, One Rogers Street, Cambridge, MA 02142, USA adam.perer@us.ibm.com
[***] School of Computer Science, Tel-Aviv University, Tel Aviv 69978, Israel ohadgree@tau.ac.il
[****] MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA   itai@mit.edu

ABSTRACT
Despite the tremendous popularity of social network sites both on the web and within enterprises, the relationship information they contain may be often incomplete or outdated. We suggest a novel crowdsourcing approach that uses a game to help enrich and expand the social network topology. The game prompts players to provide the names of people who have a relationship with individuals they know. The game was deployed for a one-month period within a large global organization. We provide an analysis of the data collected through this deployment, in comparison with the data from the organization's social network site. Our results indicate that the game rapidly collects large volumes of valid information that can be used to enrich and reinforce an existing social network site's data. We point out other aspects and benefits of using a crowdsourcing game to harvest social network information.

Author Keywords
Crowdsourcing, Games, Social Networks, Human Computation

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Experimentation, Human Factors, Measurement

INTRODUCTION
As social network sites (SNSs) gain popularity, more people are connecting to their colleagues, family, and friends in these online spaces. Such relationships allow users to stay up-to-date with their peers and perhaps exploit their social network for many tasks, including finding a job

[17], ideation [8], getting answers to questions [29], and more. Data mining and information retrieval algorithms can also use social network information to enhance their functionality (e.g., [26]). However, there are several drawbacks to relying on online social network information for these human and machine information discovery tasks.

Relationships on social network sites are often incomplete. Many populations are not active in SNSs for various reasons, including privacy concerns, uneasiness with the technology, limited access to the technology, or simply a lack of time. Relationships may also be outdated or frozen [6]. Often, people connect online with new people who come into their lives, but tend not to disconnect online even when a relationship no longer exists offline. Finally, online social network relationships are typically unweighted [16]. This means all online relationships are treated equally, even though a variety of strengths exist across all connections.

Incomplete, outdated, or unweighted information suggests that discovery tasks performed by both people and computational algorithms may suffer from this imperfect data. In this paper, we propose using crowdsourcing games to help fill the gaps in social network information and improve the representation of the overall social network.

We introduce GuessWho, a crowdsourcing game for the enterprise, where users enter knowledge about their peers to enrich the organizational social network. The advantage of using a game, rather than a tool, is that it rewards people for contributing valuable information with a fun experience. We show that inviting 'the crowd' to compete in entering social network information can produce high-quality data.

While crowdsourcing games have been used in the past to gather tags about peers [4], our game focuses on collecting data about relationships. As past research indicates that the crowd can generate tags that enhance user profiles that may be incomplete or frozen, GuessWho also supports people tagging by the crowd. This also allows us to compare the relationship information collected through the game with previously studied tag information. However, this paper's main novelty lies in the collection of social relationships through a crowdsourcing game. These relationships can

affect the social graph in a more global way, bridging previously disconnected clusters of people, which can lead to a dramatic improvement in tasks and algorithms for recommender systems, search engines, or expertise locators.

Existing literature offers a variety of motivations and potential uses for the kind of information collected by GuessWho. Social relationship information can be used for a wide range of enterprise applications, such as recommending people and content [21,22], finding paths to experts [11], producing dynamic contact lists [18], and personalizing search results [9]. Other studies have focused on the use of tags in the enterprise, for example, to infer expertise and interests [12] and build communities [13].

We allow users to play GuessWho on both explicit and implicit ties. We refer to explicit ties as ones that have been articulated in a formal way, such as a connection within an SNS or a manager-employee relationship within the organizational chart. We infer implicit ties from social media activity, such as people who co-edited the same wiki page or shared a file with one another. Previous research [19] has shown that these mined relationships indeed reflect people with whom the user is familiar. Thus, our approach extends beyond existing "friendsourcing" techniques [4], since we allow players to input information about implicit ties, for which they may have knowledge that can be useful as part of the crowd's input.

We evaluated our results by comparing the data generated by GuessWho to the relationships and tags on the organization's existing widely-used enterprise SNS, which supports relationships (via "friending") and people tagging. Our analysis suggests that data generated via the GuessWho game is of significantly higher quality than data already existing within the SNS, and thus can be used to enrich, expand, and validate it.

The main contributions of this work are (1) suggesting a novel approach to elicit social relationship information using a human computation game, by having players input people's names as opposed to tags, (2) providing a large-scale evaluation in the enterprise, which includes a comparison to a widely used system that does not apply a gaming approach, and (3) introducing a novel scoring scheme that aims to balance the validity and diversity of the input, and can serve as a subject for future research.

Our paper begins with a reflection on related work, where we discuss crowdsourcing and human computation games. We then briefly describe the enterprise SNS we used as a benchmark, and its usage within the organization. We follow with a detailed description of GuessWho. Next, we describe the deployment of GuessWho in a large organization where almost 2,000 players generated over 80,000 items of social network data. We then describe our evaluation of 132 participants, where we compared data generated by the game to data generated by the enterprise SNS. The evaluation suggests that data generated by the

GuessWho game is of significantly higher quality than the SNS data. Finally, we discuss different aspects of our work, and then conclude.

RELATED WORK

Crowdsourcing [23] is an emerging paradigm that harnesses masses of users to construct knowledge bases, perform various types of tasks and solve problems [7,24,1]. Crowdsourcing applications have been recognized as a powerful tool for taking advantage of the "wisdom of the crowd" [30] in various domains, e.g., distribution of programming tasks [3], enhancement of search engines [14], or machine translation [1].

Luis von Ahn introduced Games with a Purpose (GWAP) [31,32], a framework for motivating crowdsourcing through games. Participants play a game, have fun, compete with others on the web, and at the same time help computers solve problems through the data they enter as part of the game. The ESP game [33] was the first game of this kind, soliciting players to label images with descriptive tags. The ESP game is typically played by two players simultaneously. They both receive an image as input and need to "agree" on as many tags as possible that describe the given image. These tags are used to enhance image search, as they enrich images' textual metadata. Following the ESP game, many other games have been developed by Von Ahn et al., harnessing the power of *human computation* (thus often referred to as *human computation games*). Most of these games focus on annotating non-textual media, such as images, music, or video.

Von Ahn et al. [31] classify the ESP game as an *output-agreement* game, in which both players are given the same input and each tries to produce the same output as their partner. GuessWho is also an output-agreement game, in which the input is a person and the output is either the name of another person or a tag. Other classes of games include *input-agreement* games, in which players are given inputs and are prompted to produce descriptive outputs, so their partners can asses whether their inputs are the same or different. An example for such a game is TagATune [27], in which players are given a musical pitch, asked to describe it with tags, and then prompted to assess, based on their partner's tags, whether they both received the same tune or different ones. In *inversion-problems* games, one player is the describer and the other is the guesser. The describer is given an input and is asked to provide outputs that help the guesser produce the original input. For example, in Peekaboom [35], a game for locating objects in images, one player is given an image along with a word related to it and must reveal parts of the image for the other player to guess the correct word. Verbosity [34] is a game for collecting common-sense facts. The describer provides short structured facts about a given input word (e.g., it is a kind of X, it is used for Y) so that the guesser can infer it.

Human computation games have also been studied by other authors in different domains. For example, Curator [36] is a

two-player game for creating collections of items that fit together. Given a set of items, the players need to group them into collections and are rewarded for matching collections. PageHunt [28], a single-player non-collaborative game, challenges the user to guess search queries that would yield a given web page among their top five results. The data received through the game is used for improving the search results of a commercial search engine. Arase et al. [2] describe a multi-player game for inferring image geographical relevance. Through an analysis of the game logs, they confirm that geographically relevant objects can be derived and that their game-based method can significantly improve image search results and enhance application in domains such as image location recognition. Finally, like GuessWho, the Dogear Game [10], is an enterprise single-player human computation game. The player is presented with a bookmarked web page and needs to guess the colleague to whom the bookmark belongs. Incorrect answers are used to produce human-sourced bookmark recommendations for users of an enterprise social bookmarking system. In all of these games, the input is an artifact such as an image or a web page, while in GuessWho, it is the name of an individual person.

The closest research project to GuessWho is Collabio [4], a Facebook application that encourages friends to tag one another with descriptive terms through a game. As in our case, the game application needs to make sure the player is familiar with the individuals who appear as input. Collabio does so by allowing people to tag only their Facebook friends. Ultimately, since the tag cloud for a given person is not created by the general public, but rather by that user's set of Facebook friends, Collabio is referred to as an example of *friendsourcing* rather than crowdsourcing [4]. Apart from being an enterprise game, GuessWho is different than Collabio, as it supports building the social network itself, rather than just annotating people within the social network. As opposed to Collabio and many other human computation games, the input players provide is not only in the form of tags, but also in the form of relationships. Other differences lie in the way the game is played, for example, how the people who appear as input are selected, and how scores are calculated. We discuss in detail the commonalities and differences between Collabio and GuessWho later in this work.

LOTUS CONNECTIONS

Lotus Connections (LC)[1] is a social software application suite for enterprises. It includes various enterprise social software applications, such as a blogging system, a wiki system, a social bookmarking system, a file sharing system, and more. LC's Profiles application is an enterprise version of a social network site (SNS). It allows employees to reciprocally connect to one another by sending and receiving invitations. This paper utilizes the deployment of

LC in a large, global IT company, where the system has been available for over three years. Overall, in this organization, 175,000 relationships have been formed between 80,000 employees. The basic profile of a user (name, job title, email address, etc.) is based on the corporate directory. In addition, employee profile pages are enriched with their activity within LC. In particular, LC includes an enterprise people tagging application that allows employees to tag one another with descriptive tags [12]. Those tags are presented as part of the profile page of each employee within LC. In the organization, 9,500 users have tagged 55,000 other individuals, with 170,000 tags.

GUESS WHO

GuessWho is a crowdsourcing game, in which users enter knowledge about their peers to enrich the social network with additional relationships and tags. In this section, we describe the game in detail, including the person selection process, the scoring calculation, and the user interface.

Person Filtering

In GuessWho, the *object* for which players are required to provide input is a person. In addition to soliciting person-to-person relationships, we wanted to examine how people tags are collected through the game; this has been previously inspected in a non-enterprise context [4]. Hence, for each round of the game, the input type of either relationships or tags, is randomly drawn. Each round lasts one minute, during which the player must specify as many relevant relationships or tags as possible, in order to gain points.

Because the object in each round is a person, a filtering stage is required to make sure players are familiar, at least to some extent, with the person for whom they are entering information. To this end, we use SONAR [18], an enterprise social network aggregation system that collects and aggregates social network information from various sources across the organization; these sources include an enterprise SNS, a file sharing system, a wiki system, a paper and patent database, the organizational chart, and more. SONAR extracts both explicit relationships, such as being connected on an SNS or having an employee-manager relationship by the organizational chart, and implicit relationships, such as co-editing a wiki page or co-authoring a paper. These relationships are all aggregated to create a ranked list of the people most familiar to a given user. More details on SONAR, its data sources, and score calculation can be found in [18,19,20].

At the beginning of the game, the *object list* for a given player, i.e., the list of people on whom they can play, consists of the top 100 related people as returned by SONAR for that player. For each round of the game, a person is drawn at random from this list, and, as mentioned before, the input type (relationships or tags) is also decided at random per round. To allow more options for playing, we dynamically update the object list of a player any time they add the name of a new person as input; we assume that

player / round	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
r1	0									
r2	12	9								
r3	25	19	14							
r4	50	38	28	21						
r5	100	75	56	42	32					
r6	50	38	28	21	16	12				
r7	25	19	14	11	8	6	4			
r8	12	9	7	5	4	3	2	2		
r9	6	5	3	3	2	1	1	1	1	
r10	3	2	1	1	1	1	1	1	1	1

Table 1. GuessWho scoring scheme for a given answer.

specifying a person as input indicates some level of familiarity between that person and the player. In this way, the object list grows dynamically as the player continues with the game and provides more people as input.

Score Calculation

GuessWho is played in a single-player mode, in which the player's input is matched against previous input on the same person by other players. We opted for a single-player mode, as it is unlikely that two players will have enough people they both know to play simultaneously.

The GuessWho scoring function aims at encouraging original, yet valid, answers. Thus, an answer that has already been given by many other players would not yield as many points as an answer that has been given by only few players. We also apply a concept of *dividend points*. These are points accumulated **after** the input has been provided by the player, when more players note the same input—thereby corroborating the information. Most dividend points are granted for an original answer that was not given previously, but is later noted by many individuals.

Collabio [4] also uses an accumulating point mechanism, giving i points to the i-th player who mentions a tag, as well as to all preceding players who mentioned the same tag for the same person. This method rewards popular answers, while our scoring system aims at eliciting non-popular answers. The ESP game applies Taboo Words [33] as a means to disallow very popular answers. However, we felt that this method would not suffice, since it can only eliminate a limited number of very popular answers.

Table 1 describes the GuessWho scoring scheme for a specific answer. Each column shows the scores of the i-th player ($i \in [1,10]$), while the i-th row refers to the round in which the i-th player provided the answer. The top number in each column (bolded in blue) shows the immediate points the player gets, while the rest of the numbers in the column show the dividend points given to the player in case

other players mention the same answer later on. For example, when the first player ($p1$) gives the answer, s/he gets 0 points, as the answer is not valid yet, and we want to avoid encouraging arbitrary answers. When the second player mentions this same answer, s/he immediately gets 9 points, while $p1$ gets 12 dividend points. The highest amount of points is given when a fifth player provides the same answer. This player then immediately gets 32 points, granting the first player who provided the answer with 100 dividend points, the second with 75, the third with 56, and the fourth with 42. The eleventh player to specify an answer does not get any points, and does not grant dividend points to previous players. This method encourages diversity and prevents players from specifying the most popular answers.

We note that this scoring scheme has various parameters that control the way it behaves. Optimization of the scoring scheme is beyond the scope of this work and is left for future research.

User Interface

Figure 1 shows the GuessWho welcome page, presented once a player logs in with his organizational username and password. This page gives a short description of the game instructions and shows the leaderboard with the 15 players who scored the most points thus far. The bottom of the page indicates the number of people who played during the past week and prompts the player to join them. The player's overall points, ranking within the overall scoring list, and number of dividend points accumulated since the last login are also shown. Finally, a tip showing the number of points required to pass the next player on the list (i.e., improve the ranking by 1) is highlighted as another means to motivate players. By clicking "Start Game", the player begins a new round in GuessWho.

Figure 2 depicts the main user interface of GuessWho. Figure 2a shows a relationship round (required input is related people) and Figure 2b shows a tag round (required input is related tags). At the center, a person who is familiar
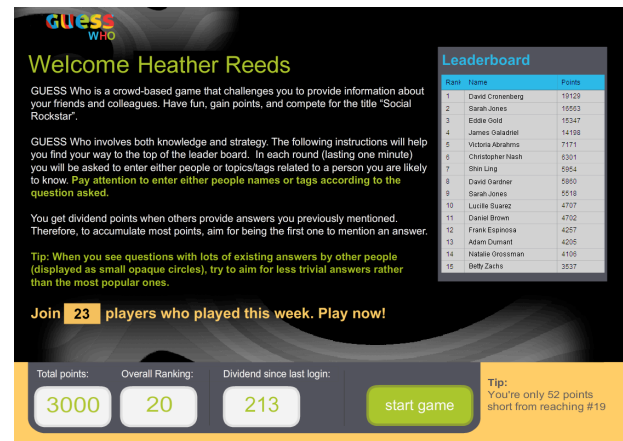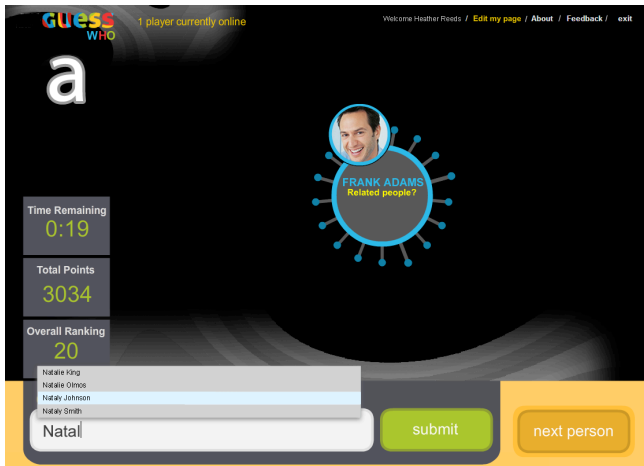

Figure 1. After players log in to GuessWho, they are greeted with instructions, their current points and ranking, a leaderboard, and motivating tips to raise their ranking.

Figure 2. a) A GuessWho relationship round, where players input people related to the person in the center (Frank Adams). When a name is being entered, the auto-complete service suggests completion options. b) A GuessWho tag round, where players input tags related to the person in the center. In this example, a related tag has just been entered that was previously mentioned by 3 other players.

to the player is presented by their name and photo, and the player is asked to specify people or tags related to that person. Answers already given by the player for that person, if any exist, are highlighted in circles coming out of the central circle (as in Figure 2b). The small opaque circles denote answers given by other players for this person. The remaining time (starting from 60 seconds and decreasing to 0), current total points, and overall ranking, are indicated on the left. The bottom of the screen includes an input box where the player can enter the answers. Above the box, a text repeats the current requested input (e.g., "related tags to Frank Adams"). A representative icon to the left of this text also indicates whether the required input is a relationship or a tag. As long as the remaining time is greater than zero, the player can submit answers. The player can also choose to move to the next person before the allotted time expires, at the cost of 1 point.

If the required input is a related relationship, a name auto-complete service is used to facilitate identifying a person by their name. This service makes use of the corporate directory to suggest unique identities based on the prefix entered by the player. Figure 2a shows an example of suggestions made by this service, just above the input box. If the input is a tag, a spell-checking service is used. This service matches the answer submitted by the player with answers previously entered (by all players); it also uses the Yahoo! Spelling Suggestion API[2] to offer alternative spellings, when appropriate. Once the player enters an answer, a text line pops up indicating the number of players that already gave this answer. Then, another green box pops up, indicating how many points the player just received for the current answer. For example, Figure 2b indicates that the answer provided was already mentioned by three other

individuals, and that the player gained 21 points for this answer.

At the top of the page, the user can see how many other players are currently playing. In case dividend points are added due to answers given by other users playing at the same time, the player will receive a popup notification indicating dividend points were won in real-time. Users can also choose to edit the data entered on then. Clicking that option loads a profile editing page, as demonstrated in Figure 3. The page presents two separate lists of tags and people that were assigned to the user by players of the game, each sorted alphabetically. Each person or tag on these lists includes a count, indicating how many players provided it, without naming these players. The user can remove each of these people or tags, which would indicate to us that s/he does not want this data exposed to other

Figure 3. GuessWho profile editing page, where users can choose to remove sensitive or inaccurate information mentioned about them.

---

[2] http://developer.yahoo.com/search/web/V1/spellingSuggestion.html

services. The game, however, would not be affected by this action, i.e., in case another player enters this tag/relationship again, points will be given as if the tag/relationship has not been removed. A similar concept of editing your own tags has been applied in Collabio [4].

The GuessWho application sends weekly update emails to all people who ever provided input through the game. The email message includes the user's total points, current ranking, point missing for moving one ranking up, and the Top-15 leaderboard. In addition, the email lists up to 10 relationships/tags that yielded most dividend points to the user in the past week. For example, an item in this list looks like 'Frank Adams->"design": 112 points (3 new mentions)'. This way, we aim to take advantage of the dividend point mechanism to highlight for users their most successful answers and motivate them to engage in more rounds of GuessWho. The email message includes a link to the game and also a link that would unsubscribe the user from getting further GuessWho updates.

RESULTS

The GuessWho game was deployed in a large, global IT organization for a period of one month starting June 2010. Initially, 2000 active social media users spanning different countries and organizational divisions were invited to play the game. Further game participants were then recruited based on the social data played on by these initial users. In total, 1915 users from 52 different countries played the GuessWho game, generating data on 16,764 unique people in 61 countries across the company.

Relationships

During the game's deployment, 42,487 relationships between 13,957 people were specified by users playing the game. Of these, 22,741 were unique relationships. On average, each relationship was entered 1.87 times, with 14,614 relationships (64.26%) verified by at least two players.

Only 4,053 of the 22,741 unique relationships generated by GuessWho already existed on LC (17.82%). This shows that the game was able to generate many relationships previously missing from the enterprise social graph.

An interesting trend in the data suggests that the more times a relationship was verified by GuessWho players, the more likely the relationship already existed in LC. As Figure 4a shows, the higher the verification, the greater the chance the relationship already exists on LC (e.g., there is a 58% chance the relationship exists on LC if it was verified 10 or more times).
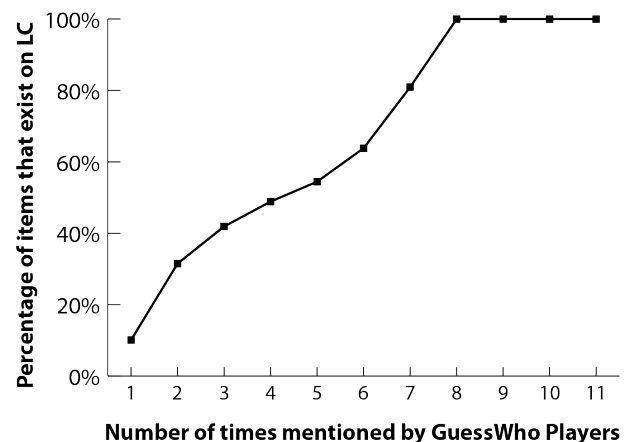
Tags

During deployment, GuessWho players entered a total of 38,768 tags about 7,499 peers. Of these, there were 35,657 unique person-tags. Only 2,221 of these person-tags were verified 2 or more times (6.23%), resulting in high tag diversity. On average, each person-tag was entered 1.09 times by players of the game.

Only 3,598 of the 35,657 unique person-tags generated by GuessWho already existed on LC (10.09%). This shows that the game was able to generate many new tags about people, resulting in a much richer overall data.

Similarly to the trend with relationships, the data suggests that the more times a tag was verified by GuessWho players, the more likely the tag already existed in LC. As Figure 4b shows, the higher the verification, the greater the chance the tag already exists on LC (e.g., there is a 55% chance the tag exists on LC if it was verified 5 or more times).
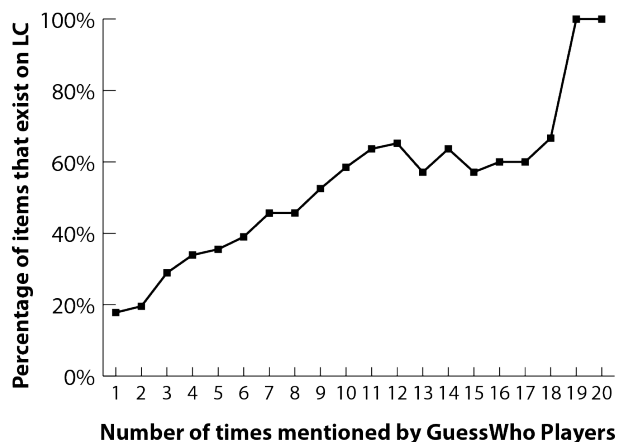
a. Relationships

b. Tags

Figure 4. A correlation exists between the number of times items were mentioned by GuessWho players and the probability that the item exists on the enterprise social network, LC.

Privacy

As users are not in control of the data being entered about them in GuessWho, users were able to remove any data about them that they might find undesirable or inaccurate. Only 455 items (0.006%) of the 81,255 user-generated tags and relationships were marked to be deleted by users. Of the hidden items, 295 were relationships and 160 were tags. 224 users out of all 1,915 GuessWho users did opt to hide at least 1 item mentioned about them.

EVALUATION

In order to better understand the quality of data produced by the GuessWho game, we invited 332 people to rate the data generated about them. These 332 people were chosen based on having a sufficient amount of data generated by the game, as well as having sufficient data on LC, to generate 10 relationships and 10 tags for evaluation. This is explained in the following paragraphs.

Each of the players were sent an email invitation to a within-subjects survey, in which they were asked to rate the quality of up to four randomly chosen tags and four randomly chosen relationships that were generated uniquely by the GuessWho game. In order to generate a baseline, four random tags and relationships were also extracted from LC. Participants were also asked to rate up to two tags and two relationships that existed on both systems. Of the 332 invitations, 132 people (39.76%) opted to take the survey.

Relationships

After a set of brief instructions, participants were shown a list of 10 people. Next to each person's name was a set of radio buttons representing an ascending Likert scale from 1 to 5. Participants were asked to rate the strength of the relationship between themselves and that person (where 1 indicates no relationship and 5 represents a strong relationship). The set of 10 people was chosen as follows: 2 of the people were listed as relationships on both GuessWho and LC, 4 of the people were only listed on GuessWho, and 4 were listed only on LC. Comparing relationships mentioned in GuessWho to those listed on LC allowed us to understand the quality of data entered in a crowdsourcing game. The 10 people were presented in a randomized order and participants were not told the sources of the data to avoid any biases.

As the first row of Table 2 shows, participants gave an average score of 3.84 to relationships that originated only from GuessWho. Relationships from LC only were given a less favorable rating, with an average score of 3.56. Finally, relationships that existed in both systems were given a rating of 4.25. A one-way ANOVA indicates that ratings across the three conditions are significantly different, $F_{(2,1230)}=28.297$, $p<0.001$. This suggests that data gathered by the GuessWho game is significantly better than friending data from the LC SNS, indicating the very real potential of our crowdsourcing technique.

	LC	GuessWho	Both
Relationships	3.56	3.84	4.25
Tags	4.14	4.35	4.78

Table 2. Participants rated tags and relationships generated in LC and GuessWho. Data generated in GuessWho was rated significantly higher than LC data, on an ascending Likert scale. Data that was present in both systems was rated even higher.

	Mentioned 1 time in GuessWho	Mentioned 2 or more times	Mentioned 3 or more times
Relationships	3.62	3.96	4.31
Tags	4.34	4.63	4.81

Table 3. The quality of data generated by GuessWho increases based upon the number of times items were mentioned by separate players.

There is also a correlation between the number of times a relationship was entered by multiple players in GuessWho and the quality of the data. As Table 3 shows, relationships that were entered only once got an average rating of 3.62, whereas relationships that were entered by two or more players got an average of 3.96. This rating rose to 4.31 when three or more players all verified the same relationship.

Tags

Similarly, participants were also asked to rate how well 10 tags describe themselves using a Likert scale of 1-5, where 1 means 'not at all', and 5 means 'very well'. Tags were extracted in a similar manner to relationships (4 unique GuessWho tags, 4 unique LC tags, and 2 that were entered in both systems). As the second row of Table 2 shows, tags extracted from GuessWho resulted in an average rating of 4.35. Similar to relationships, tags from LC were rated lower than the GuessWho tags, with an average score of 4.14. Once again, tags that existed on both systems performed even better, with a rating of 4.78. A one-way ANOVA indicates that ratings across the three conditions are significantly different, $F_{(2,1238)}=29.209$, $p<0.001$. Thus, tags entered on GuessWho were significantly better than tags that existed on LC, suggesting that crowdsourcing games are indeed a good source for tag knowledge as well.

As with relationships, a correlation exists between the number of times a tag was mentioned by multiple players in GuessWho and the quality of the data. Table 3 shows that tags that were only entered once got an average rating of 4.34, whereas tags that were entered by two or more players received an average of 4.63. This rating increases to 4.81 when three or more players all verified the same tag. The Collabio study [4] showed that tags entered by more players were rated higher than tags entered by fewer players. Our

analysis reinforces these results for both tags and relationships in the enterprise.

DISCUSSION, LIMITATIONS, AND FUTURE WORK
Our evaluation indicates that harnessing the crowd to enrich the social network with relationships and tags can be highly beneficial. The information harvested through GuessWho contributed many new relationships and tags that did not exist before in LC. Ratings of these new relationships and tags indicate that their validity is higher than existing LC data, even if entered only once. Yet, the more mentions an answer gets, the higher its rating is likely to be. In addition to contributing plenty of new data, the GuessWho answers can be used to reinforce and validate existing LC data. Our survey indicates that answers that appear both on LC and GuessWho receive especially high ratings. Because relationships in LC (as in most other SNSs) are unweighted, GuessWho relationship data can be used to rank them, according to the number of people who mentioned these relationships. In practice, GuessWho can be tuned to give priority to collecting data for individuals whose information is sparse or suspected as being stale, by boosting these individuals within the user's object list.

Our scoring function aims at flattening the distribution of answers and yielding more original yet valid answers. Indeed, the diversity of received answers was high, as each item was only repeated an average of 1.39 times overall. In fact, only 55 answers were given more than 10 times (the area of no awarded points in our scoring function). On the other hand, the validity of the answers was also high, as reflected in the survey's results. This may imply that the scoring function indeed helps produce diverse yet valid results. Future work should explore how optimizing the scoring function can further improve the results; this should also consider factors such as the number of expected players, their level of engagement, the number of possible objects and possible valid answers per object, and more.

The fact that the game-based data was fresher than the LC data may offer one explanation for the higher rating the former received in our survey. This fact was pointed out in several of the comments we received from survey participants. For example:

"The one thing I noticed looking at both the relationships and tags was the staleness of some of them. In many cases, I once had strong relationships with the people that have waned over time. The same is true for tags where, because of, for example, job changes, the tags are no longer as relevant as they once were."

While the data in LC could be filtered by timestamp to make it fresher, it would also make it much sparser. Moreover, it is hard to infer which relationships and tags are stale and which have been maintained or developed even though entered a few years ago. GuessWho, or a variant of the game, can serve as a means to validate the freshness of the data in LC. As indicated in our results, the data that appeared on both GuessWho and LC was rated very positively by the users, with relationships rated 4.25 on average, and tags getting an extremely high average rating of 4.78.

Tag diversity was substantially higher than relationship diversity. On average, each tag-answer was repeated 1.09 times, as compared to 1.87 for a relationship-answer. The percentage of answers that were verified by two users or more is much smaller for tags than for relationships (6.2% vs. 64.3%). This can be attributed to the fact that the set of possible inputs for people is smaller than for tags. Moreover, relationship input is made through a unique identifier (given by the name autocomplete service), while tags may have different forms and synonyms. Hence, people-answers are easier to validate.

There was also a difference between relationships and tags with respect to the improvement of GuessWho answer ratings compared to LC data. The rating raise was higher for relationships than for tags (7.9% vs. 5.1%), indicating, again, the efficacy of crowd-based relationship extraction.

In summary, we observe several unique advantages of the relationship version of the game, compared to the tag version: (1) Since people in the enterprise have unique identifiers, there is no need for canonization, stemming, or synonym merging; thus validation of results is easier. (2) As opposed to people tags, which can be entered in LC by any other person, relationships can only be formed when one of the two parties initiates the connection. Our crowd-based approach allows others to point at a relationship between two other individuals and hence may reveal relationships between people who are not active in social media and will not initiate the connection. (3) The increase in average rating compared to LC data is higher for relationships than it is for tags (7.9% vs. 5.1%). (4) LC relationships are unweighted, while the crowd-based relationships can be weighted by the number of players who mentioned them. (5) The game revolves around three entities: the player, the object, and the input. The relationships among these entities are reflected throughout the course of the game: person filtering using SONAR determines the initial player-object relationships; the object list dynamically grows based on player-input relationships; and, the social graph is enriched through object-input relationships. Furthermore, this symmetry allows us to potentially infer relationships not only based on object-input relationships, but also based on player-input relationships. Future research should compare the data that can be extracted through these two types of relationships.

The volume of the data collected through the game was impressive. In the period of just one month, over 24% of the overall LC relationships and 22.5% of the LC tags were generated by the GuessWho game (LC has been deployed in the organization for over three years). In fact, over 50% of this data was produced in the first week of the game, while in subsequent weeks we witnessed a constant

decrease in data production. These findings indicate that a crowdsourcing game can be an effective means to extract large volumes of data in a very short period of time. Keeping up the pace of data production is a great challenge with such a game, as with other non-monetary incentive mechanisms [14]. Weekly emails with motivating personalized statistics are one method we applied to keep players engaged. Yet, as mentioned, a rapid decay of contribution was inevitable. Future research may examine new ideas (e.g., other incentives such as organizational rewards) and whether they can enhance longer-term participation in crowdsourcing games.

One limitation of the game approach and its evaluation here stems from this challenge of keeping the game's data fresh. It could be that the "novelty effect" of the game, leading many people to try it out early on, is the main reason for its data being of higher quality than the LC data. Ultimately, as the novelty wears off, these differences may fade. Future work should inspect whether the findings of our evaluation hold over a longer period of time.

After deploying the game, we received many comments by email or through the enterprise blogosphere. For example, one user wrote in his status update:

"Tried out GuessWho. Fun way to encourage tagging!"

And another wrote:

"For me watching the clock tick down definitely added pressure to the game! Whew! I'm exhausted...I think I need a long weekend!"

We also received quite a few emails expressing surprise that we asked people to play games inside the enterprise. This suggests that initiating a game culture inside the enterprise can be difficult, even when both the players and the community at large benefit. Showing the value games like GuessWho in enterprises can help create an organizational culture that is more receptive of such games.

Privacy concerns were also raised by a few of the game's participants. While crowd-based people-tagging in the enterprise was introduced a few years ago and found to be appropriately used [12], GuessWho introduces a paradigm in which a relationship between two individuals can be pointed out by others. As one user expressed:

"What if I don't want to be known for knowing or interacting with people because of a troubled project, secret project or different internal job offers?"

The profile editing enablement, which allows individuals to remove any relationship others have indicated, tries to mitigate privacy issues. Further means should be examined to help protect privacy, such as allow editing the profile in interfaces other than GuessWho or sending notifications whenever a new relationship or tag about the employee has been indicated by others. Our results indicate that relationship data might be more sensitive than tags data, since more relationships were removed than tags (295 vs.

160). Additional exploration of employee sensitivity to others making input about their relationship information should also be studied in more detail.

We also received many suggestions for enhancing the game, including dynamic updates of the player's object list with other players who mentioned him, rather than just with players he mentioned; allow players to correct an input they entered by mistake; extend the game to allow annotating relationships with labels indicating why two individuals are linked (e.g., boss of, worked on the same project, etc.); and allow players to indicate they do not know the person they received as the object in a specific round, so this person will not appear again in future rounds. This would also help us learn about relationships that do **not** exist, and potentially amend SONAR.

We note that aside from using the SONAR system to create the initial object list, GuessWho can also serve as a data source for SONAR, feeding back the relationships it collects to the SONAR aggregation system.

Implementing an equivalent version of GuessWho outside the firewall can be valuable in many ways, for example to maintain and enrich the social graph within an SNS or an online community, or even build a network from scratch within a social site. The challenge in producing high-quality and valid results is bigger in this case, as the accountability for one's ID is not as high as within the organization.

Crowdsourcing as a whole often raises a controversy around the quality of the data it produces. This controversy may even grow when the subject is as sensitive as unveiling relationships and areas of expertise. One of our users hypothesizes:

"My favorite relationships are those I select myself, rather than those others might choose for me."

While this may be true, we believe many latent relationships still exist that can be positively utilized. The results of this work show that the data stemming from a crowd-based game during one month is more representative than the user's own relationships within an SNS, accumulated over more than three years.

CONCLUSION

We introduced an approach that utilizes a crowdsourcing game to expand and enrich the enterprise social network. The game allows players to indicate relationships and tags related to a given person. We compared the data collected through a one-month deployment of the game with the data of a well-established enterprise SNS. Results indicate that the game can rapidly extract large amounts of relationships and tags that are both diverse and valid. Validity of the data rises as more people mention an answer, yet even an answer given just once is rated higher, on average, than a relationship/tag taken from the SNS. While tags have been previously used as input within crowdsourcing games, relationship input is novel. We point out several advantages

of a game that collects relationships as input in comparison to tags. In addition, the game introduces several unique mechanisms, such as a scoring function that includes dividend points and the dynamic update of the player's object list based on people s/he has previously mentioned. Future work should explore ways to exploit the data collected through the game; investigate refinements of the scoring function, means for longer-term engagement, and methods for privacy preservation; and examine whether this game-based approach can be applicable outside the firewall.

REFERENCES

1. Ambati, V., Vogel, S., & Carbonell J. 2010. Active learning and crowd-sourcing for machine translation. *Proc. LREC'10*.

2. Arase, Y., Xie, X., Duan, M., Hara, T., & Nishio, S. 2009. A game based approach to assign geographical relevance to web images. *Proc. WWW '09*, 811-820.

3. Archak, N. 2010. Money, glory and cheap talk: analyzing strategic behavior of contestants in simultaneous crowdsourcing contests on TopCoder.com. *Proc. WWW '10*, 21-30.

4. Bernstein, M., Tan, D., Smith, G., Czerwinski, M., & Horvitz, E. 2010. Personalization via friendsourcing. *ACM Trans. Comput.-Hum. Interact.* 17, 2 (May. 2010), 1-28.

5. Bonhard, P. & Sasse, M. A. 2006. 'Knowing me, knowing you' -- Using profiles and social networking to improve recommender systems. *BT Technology Journal* 24, 3 (Jul. 2006), 84-98.

6. Boyd, D. & Heer, J. 2006. Profiles as Conversation: Networked Identity Performance on Friendster. *Proc. HICSS '06*.

7. Brabham, D.C. 2008. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence* 14, 1, 75–90.

8. Burt, R. S. 2004. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2), 349–399.

9. Carmel, D., Zwerdling, N., Guy I., Ofek-Koifman, S., Har'el N., Ronen, I., Uziel, E., Yogev, S., & Chernov, S. 2009. Personalized Social Search based on the User's Social Network. *Proc. CIKM '09*, 1227-1236.

10. Dugan, C., Muller, M., Millen, D. R., Geyer, W., Brownholtz, B., & Moore, M. 2007. The dogear game: a social bookmark recommender system. *Proc. GROUP* '07, 387-390.

11. Ehrlich, K., Lin, C. & Griffiths-Fisher, V. 2007. Searching for experts in the enterprise: combining text and social network analysis. *Proc. Group'07*, 117-126.

12. Farrell, S., Lau, T., Nusser, S., Wilcox, E., & Muller, M. 2007. Socially augmenting employee profiles with people-tagging. *Proc. UIST '07*, 91-100.

13. Farrell, S., Lau, T., & Nusser, S. 2007. Building communities with people-tags. *Proc. INTERACT '07*, 357-360.

14. Farzan, R., DiMicco, J., Millen, D. R., Dugan, C., Geyer, W., & Brownholtz, E. 2008. Results from deploying a participation incentive mechanism within the enterprise. *Proc. CHI '08*, 563-572.

15. Ganjisaffar, Y., Javanmardi, S., & Lopes, C. 2009. Leveraging crowdsourcing heuristics to improve search in Wikipedia. *Proc. WikiSym '09*.

16. Gilbert, E. & Karahalios, K. 2009. Predicting tie strength with social media. *Proc. CHI '09*, 211-220.

17. Granovetter, M. 1974. Getting a Job: A Study of Contacts and Careers. *University Of Chicago Press*.

18. Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V., & Farrell, S. 2008. Harvesting with SONAR: the value of aggregating social network information. *Proc. CHI '08*, 1017-1026.

19. Guy, I., Jacovi, M., Meshulam, N., Ronen, I., & Shahar, E. 2008. Public vs. private: comparing public social network information with email. *Proc. CSCW'08*, 393-402.

20. Guy, I., Jacovi, M., Perer, A., Ronen, I., and Uziel, E.: Same Places, Same Things, Same People? Mining User Similarity on Social Media. *Proc. CSCW'10*, 41-50.

21. Guy I., Ronen I., & Wilcox E. 2009. Do You Know? Recommending People to Invite into Your Social Network. *Proc. IUI'09*, 77-86.

22. Guy, I., Zwerdling, N., Carmel, D., Ronen, I., Uziel, E., Yogev, S., & Ofek-Koifman S. 2009. Personalized Recommendation of Social Software Items based on Social Relations. *Proc. RecSys '09*, 53-60.

23. Howe, J. 2006. The rise of crowdsourcing. *Wired Magazine* 14, 6 (June 2006).

24. Huberman, B. A., Romero, D. M., and Wu, F. 2009. Crowdsourcing, attention and productivity. *J. Inf. Sci.* 35, 6 (Dec. 2009), 758-765.

25. Kempe, D., Kleinberg, J., & Tardos, É. 2003. Maximizing the spread of influence through a social network. *Proc. KDD '03*, 137-146.

26. Kautz, H., Selman, B., & Shah, M. 1997. Referral Web: combining social networks and collaborative filtering. *Commun. ACM* 40, 3 (Mar. 1997), 63-65.

27. Law, E. & von Ahn, L. 2009. Input-agreement: a new mechanism for collecting data using human computation games. *Proc. CHI '09*, 1197-1206.

28. Ma, H., Chandrasekar, R., Quirk, C., & Gupta, A. 2009. Page hunt: improving search engines using human computation games. *Proc. SIGIR '09*, 746-747.

29. Morris, M. R., Teevan, J., & Panovich, K. 2010. What do people ask their social networks, and why?: a survey study of status message q&a behavior. *Proc. CHI '10*, 1739-1748.

30. Surowiecki, J. 2004. The wisdom of crowds: Why the many are smarter than the few.

31. Von Ahn, L. & Dabbish, L. 2008. Designing games with a purpose. *Commun. ACM* 51, 8 (Aug. 2008), 58-67.

32. Von Ahn, L. 2006. Games with a purpose, *IEEE Computer Magazine* 39, 6 (June 2006), 92-94.

33. Von Ahn, L. & Dabbish, L. 2004. Labeling images with a computer game. *Proc. CHI '04*, 319-326.

34. Von Ahn, L., Kedia, M., & Blum, M. 2006. Verbosity: a game for collecting common-sense facts. *Proc. CHI '06*, 75-78.

35. Von Ahn, L., Liu, R., & Blum, M. 2006. Peekaboom: a game for locating objects in images. *Proc. CHI '06*, 55-64.

36. Walsh, G. & Golbeck, J. 2010. Curator: a game with a purpose for collection recommendation. Proc. CHI '10, 2079-2082.

37. Yang, J., Adamic, L. A., & Ackerman, M. S. 2008. Crowdsourcing and knowledge sharing: strategic user behavior on taskcn. *Proc. EC '08*, 246-255