It’s All ‘About You’ – Diversity in Online Profiles

Casey Dugan, Werner Geyer, Michael Muller, Joan DiMicco, Beth Brownholtz, David R. Millen
IBM T.J. Watson Research
One Rogers Street, Cambridge, MA, USA 02142
+1 617 693 0188
{cadugan, werner.geyer, michael_muller, joan.dimicco, beth_brownholtz, david_r_millen}@us.ibm.com

ABSTRACT
User profiles on today's social networking sites support only a small set of predefined questions. We report on an alternative way for users to richly describe themselves, by entering not only responses, but their own questions as well. Data from 10 months of usage shows that users of a social networking site created thousands of diverse questions and reused existing questions from other users. Our findings suggest that those with highly diverse user profiles have a higher number of friends.

ACM Classification Keywords
H5.3. Group and Organization Interfaces: Collaborative computing, Evaluation/methodology, Web-based interaction

General Terms
Human Factors, Measurement, Design.

Author Keywords
User profiles, social networking, self description, impression formation, self expression, reuse

INTRODUCTION
Could an application designer have anticipated that Jenn divides the world into two kinds of people: those who call it “soda” and those who call it “pop”? Probably not. If Jenn wants to express which kind of person she is, but is only given fields for describing her “Hobbies” and “Skills,” she has lost some control over the ability to express herself. However, user profiles on popular social networking sites support only a small set of predefined questions or topics.

Online user profiles play a vital role in the formation of personality impressions [3,8] and research has shown that the presence of this kind of information may aid in establishing online connections [5]. Previous work has also studied in detail the role and importance of individual fields in online user profiles [2,6], as well as offering a conceptual framework for signaling in large social networks [1]. Liu [7] analyzed interest tokens in MySpace, a system with predefined interest categories. Profiles can be considered the result of “front stage” actions which, according to Goffman [5], are actions that are publicly visible and intended to influence other people’s perceptions of an organizational actor. Despite the importance of user profiles, their current capabilities allow only limited self-presentation.

We propose an alternative way for users to richly express information about themselves, by entering not only responses but contributing their own questions or topics as well. We implemented these free-form ‘About You’ descriptions as a feature on an enterprise social networking site [4]. This paper describes the ‘About You’ feature and presents the results from 10 months of usage by 5,694 users. The goal of this research is to examine the benefits and drawbacks of allowing an infinite number of free-form profile fields. One possibility in such a system is that the questions of users will still be largely homogeneous (as in other social networking sites). However, we hypothesized that this freedom of describing oneself would create very diverse user profiles. We wanted to understand the potential value of this diversity, for example, in supporting impression management or connection formation. We also anticipated value in the potential of other users to be inspired by these diverse questions for their own profiles.

‘ABOUT YOU’ PROFILE ENTRIES
Profiles in the social networking system used [4] include user-created content, a social network of a user (“friends”), comments, and user activity. User-created content includes the free-form ‘About You’ question and answer pairs (called ‘About You’ entries) described in this paper as well as photos and lists. Note that we will use the term question for the remainder of the paper to also denote more topical titles. Figure 1 shows a detail of an ‘About You’ section, with real entries taken from multiple users.

Users have two ways to create new entries. First, they can add them directly on their profile page. When users add a new entry, they are given two text fields to provide their own question as well as their response to it. We also suggest five possible example questions to users. Entries
are added incrementally and the total number that a user can create is not constrained. Users are able to edit, reorder, and delete entries.

The second way of creating new ‘About You’ entries is to copy existing questions authored by other users to your profile but provide your own answer. This “reuse” feature was added 4 months after the site was launched to better support observed duplication behavior. Users are able to find questions to reuse on other users’ profiles and on pages listing all users’ questions organized by popularity, recency, network, and alphabetically.

**Figure 1. Example ‘About You’ section.**

**EVALUATION**

**Method**
We analyzed our database and the server access logs for a 10 month period between May 31, 2007 and March 31, 2008. Additionally, we conducted interviews with 8 users. Note that our analysis focuses on the questions (since this is the novelty of our system) and not the answers. Arguably, diversity in profiles can also be expressed through answers.

**General Data**
Of the 21,998 members of our internal social networking system at IBM, 25.9% (5,694 users) created About You entries. This percentage is higher than most other user-created content types in the system (17.8% of users have lists, 19.2% left comments, and 22.0% have photos).

For those who created About You entries, the average number of entries was 3.33 (median 2, max 42). The total number created across all users was 18,979.

**Diversity of Questions**
In order to determine the overlap of questions, we applied various string manipulation techniques\(^1\) to fold the 18,979 About You entries into a set of distinct questions and count their uses. Based on this technique, we identified 2,822 distinct questions.

Though our system was deployed in a corporate environment, users created both personal questions, such as “What’s your desert island food?”, and professional questions, such as “Why is unit testing important to me?” Other examples can be seen in Figure 1 and Table 1.

<table>
<thead>
<tr>
<th>Question</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you do at IBM?</td>
<td>2653</td>
</tr>
<tr>
<td>Where did you go to school?</td>
<td>1386</td>
</tr>
<tr>
<td>What do I do with my spare time?</td>
<td>1229</td>
</tr>
<tr>
<td>When did I join IBM</td>
<td>1229</td>
</tr>
<tr>
<td>Where did I work before IBM</td>
<td>750</td>
</tr>
<tr>
<td>What was your first project at IBM?</td>
<td>749</td>
</tr>
<tr>
<td>Coffee Or Tea?</td>
<td>660</td>
</tr>
<tr>
<td>Where did I grow up?</td>
<td>521</td>
</tr>
<tr>
<td>Beer or wine?</td>
<td>451</td>
</tr>
<tr>
<td>Cats or Dogs?</td>
<td>347</td>
</tr>
<tr>
<td>What languages do you speak?</td>
<td>322</td>
</tr>
<tr>
<td>Do you call it &quot;pop,&quot; &quot;soda,&quot; or &quot;coke?&quot;</td>
<td>320</td>
</tr>
<tr>
<td>Children</td>
<td>287</td>
</tr>
<tr>
<td>What Day Were You Born?</td>
<td>220</td>
</tr>
<tr>
<td>How long have you been at IBM?</td>
<td>218</td>
</tr>
<tr>
<td>Where do you live?</td>
<td>198</td>
</tr>
<tr>
<td>What I am reading now?</td>
<td>154</td>
</tr>
<tr>
<td>Second Life?</td>
<td>143</td>
</tr>
<tr>
<td>What is your Myers-Briggs type?</td>
<td>135</td>
</tr>
<tr>
<td>Where did you come from?</td>
<td>126</td>
</tr>
</tbody>
</table>

Table 1. Questions ranked by frequency of uses in our system.

This freedom to describe oneself in a diverse way is limited in existing social networking sites and corporate directories which usually support at most 20 data fields. The corporate directory inside IBM, Bluepages, is set up in a similar way with predefined fields. One interviewee described what she saw as the difference between About You entries in our system versus Bluepages: “You know people on Beehive know they’re going to find out more about you as a person, not just an employee in the IBM internal directory.”

The number of distinct questions in our system shows that, if given the opportunity, users come up with diverse ways of describing themselves that systems with predefined fields would be unable to comprehensively cover.

**Diversity of Users’ Profiles**
The total number of distinct questions can not completely describe how diverse the *profiles* of our user population are, in particular given that there are still 16,157 duplicates of

---

\(^1\) Questions were considered equivalent after converting to lowercase, removing whitespace and stop words, trimming punctuation, and stemming.
Given that employees already share some common ground by being part of one company, an explanation could be that this profile diversity may simply help others in figuring out who someone is on a personal and professional level and thus increase the likelihood of connecting to them. Users spoke about this in interviews as well: “People I connect with […] are people who humanize themselves [I connect] if I find out something about who they are.” The same user referred to “stodgy boring” questions as the opposite of humanizing. Another user tried to appear interesting for others to increase the likelihood to connect: “I put in info people find interesting or might want to talk to me [about].”

Lampe et al. [6] showed that the use of profile fields in Facebook, which is using a limited number of predefined fields, is associated with the number of friends. They explain this finding with common ground theory. Also, they found that there is a weak correlation between friends and amount of information provided. Our work complements these findings in a different context. As compared to Facebook, our study was done in a corporate setting with users having an unlimited number of freeform profile fields. Our data suggest that users with less similarity in their profile questions are more connected. Given that employees already share some common ground by being part of one company, an explanation could be that this profile diversity may simply help others in figuring out who someone is on a personal and professional level and thus increase the likelihood of connecting to them. Users spoke about this in interviews as well: “People I connect with […] are people who humanize themselves [I connect] if I find out something about who they are.” The same user referred to “stodgy boring” questions as the opposite of humanizing. Another user tried to appear interesting for others to increase the likelihood to connect: “I put in info people find interesting or might want to talk to me [about].”

We also found that users with high diversity have more ‘About You’ entries than the average user, 4.35 versus 3.3. Note, again, that our diversity score is independent of the number of entries a user has.

**Duplication of Questions**

Free-form ‘About You’ entries allow diversity but for some users they may be too open-ended. These users may require guidance in coming up with questions. This need for guidance appears to be supported by the duplication behavior we observed during the first 4 months of our study period (duplicates of 34 distinct questions made up 60% of the total entries created). To facilitate this kind of duplication behavior, we added the reuse feature described earlier.

Table 2. Differences between three uniqueness groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>% users</td>
<td>50.1%</td>
<td>37.5%</td>
<td>12.4%</td>
</tr>
<tr>
<td>mean (median)</td>
<td>23.3 (14)</td>
<td>14.52 (9)</td>
<td>12.1 (7)</td>
</tr>
<tr>
<td>% reuse</td>
<td>54.7%</td>
<td>34.6%</td>
<td>18.6%</td>
</tr>
<tr>
<td>mean (median)</td>
<td>4.35 (3)</td>
<td>2.66 (2)</td>
<td>1.3 (1)</td>
</tr>
<tr>
<td>distinct</td>
<td>2560</td>
<td>497</td>
<td>6</td>
</tr>
<tr>
<td>questions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Differences between three uniqueness groups.

About 50% of our users are in the high diversity group, i.e. their similarity to other users and to one another is relatively low. Only 12.4% of our users are relatively similar to one another. This suggests that a free-form profile entry approach does not only lead to a large number of distinct questions but can also create a user population with pretty diverse, i.e. dissimilar, profile entries, allowing users to express their personality more freely. One user (from the high diversity group) said: “I’m not a follower […] so some questions I asked, are unique to me, very individual.” While the nature of profiles on the two systems is different, our diversity result is convergent with Liu’s finding that MySpace users tended to differentiate themselves from their friends [7].

Beyond the immediate benefit for profile readers of being able to better understand the various facets of a person, we were interested in other advantages of this diversity. We analyzed the number of friends users have in each group and were surprised to find that users in the high diversity group have 23.3 friends on average which is almost twice as much as the medium or low diversity users (T-test, p<.001). Note that the number of friends was not correlated with the number of About You entries (r = 0.2105) and the length of membership between high and low diversity groups was not a factor either.

The higher the frequency, the lower the number of entries a user has. Note that this measure is independent of the number of About You entries a user has.

In order to get a sense for how our 5,694 users are distributed across all entries, we grouped them into high (.667 < u < 1), medium (.334 < u <=.667), and low (0 <= u <=.334) diversity groups. Table 2 shows the differences between the three groups.

the 2,822 distinct questions. A straightforward diversity measure is the percentage of uniquely answered questions (answered by only 1 user) on each user’s profile page. This method, however, ignores the frequency of use of a question. For example, user A with 2 questions, 1 unique question and 1 question used only 2 times in total would have 50% uniqueness whereas user B with 2 questions, one unique and the other question used 1000 times in total would have the same 50% uniqueness. Clearly this is not a good measure because there is only 1 more user who is similar to user A but there are 999 other users similar to B.

Hence, a better indicator of the diversity of a user’s profile is the frequency of use of the questions they chose for their profile. The lower the frequency is, the lower the number of entries a user has. In order to make this operational, we compute the average frequency of use, AFU, as the sum of the frequency of use FU of all entries ni of a given user i as follows:

\[ \text{AFU}_i = \frac{\sum_{j=1}^{n_i} \text{FU}_j}{n_i} \]

For example, user A has an AFU of 1.5 = (2 + 1)/2, whereas user B has an AFU of 50.5. Across all 5,694 users with About You entries, the AFUmin is 1 and AFUmax is 2656. Given AFU, we compute a user’s uniqueness score as follows:

\[ u_i = \frac{\text{AFU}_\text{max} - \text{AFU}_i}{\text{AFU}_\text{max} - \text{AFU}_\text{min}}, u_i \in [0,1], \text{AFU}_i \in [1,2656] \]

The higher ui is, the more diverse, or less similar, user i is compared to other users. Note that this measure is independent of the number of About You entries a user has.
Reused entries now make up 47% of the total created, though independent duplication still took place after the feature existed and those duplications make up for 37% of the total. Of the 5,694 users with entries, 36% made use of the feature and 21% of the users’ profiles contained only reused questions. In the top 50 users ranked by number of ‘About You’ entries, 66% have 100% reuse. Clearly, our feature was well adopted. When comparing reusers with non-reusers, we found that reusers created more than twice as many ‘About You’ entries on average, 5.33 versus 2.22 (T-test, p<.001). When we compared users of the first 4 months (no reuse feature) with users who joined afterwards (reuse feature) we found that the average number of ‘About You’ entries per user stayed the same, despite a drop in contributions of all other content types after our site was opened up to the general IBM audience.

It is also interesting to mention that some of the duplication resulted from the examples we provided in the ‘About You’ creation form. These examples were: “What do you do at IBM?”, “What was your first project at IBM?”, “Where did you go to school?”, “Where did you work before IBM?”, and “What do you do with your free time?”. As shown in Table 1, these questions make up 5 of the top 6 most used questions and 35.2% of all ‘About You’ entries.

Reuse and Diversity
We also studied the relationship between reuse and the diversity of our users. Before the reuse feature was introduced, 80.45% of questions were answered by only 1 person. After the feature was introduced, the percentage of questions answered by only one person was 80.66%. The percentage of uniquely answered questions has not decreased, i.e. the question base is growing and the reuse feature did not have a negative impact on the diversity of user profiles. On the contrary, looking at the high diversity group in Table 2, we can see that, despite the high number of 2,560 distinct questions, 54.7% of all ‘About You’ entries were created through the reuse feature. One would expect that the reuse feature would be most used by the more similar users in the medium or low diversity groups. It seems that the highly diverse users benefited from the feature and it helped them pick and choose the questions that differentiated them from other users. Many of the low diversity users on the other hand created ‘About You’ entries following the examples we suggested which explains the low reuse in that group, further confirmed by the number of distinct questions answered by that group.

Interviews also suggest that users were able to reuse questions and still feel they were creative or unique: “I started looking at different profiles, seeing what people had on their profiles, picked and chose [the] ones I thought were interesting [and] hadn’t thought of.” And “I picked questions that were funny or I could give a funny answer to, my guide [was] mostly humor.”

CONCLUSION
Thousands of distinct questions were created in our system during the 10 month period, showing that, if given the opportunity, users come up with diverse and creative ways of describing themselves. We have developed a diversity measure based on the frequency of use of questions that allowed us to measure how uniquely users are described compared to each other. A major finding of our work is that diversity seems to be associated with the number of friends a user has, suggesting that it might help users to make connections. It would be difficult, if not impossible, to capture all of these rich and unique questions using any predefined set. The free-form approach also supports an evolving set of questions, making the system robust to new trends, such as the appearance of the “Second life?” question.

While freeform self descriptions afford creativity, they do not provide any guidance to users on how to describe themselves. This lack of guidance might be the reason why the average number of entries for those with questions was only 3.33 and why only 25% of all 21,998 members had ‘About You’ entries. Therefore we believe that it is necessary to provide additional mechanisms that assist users in creating more questions and answers. Our system included two such features: a static set of example questions and an explicit reuse feature. Those entries now account for a large proportion of the entries in the system, supporting the need for exemplars. However, more needs to be done. One possibility is to provide personalized recommendations to the user during the sign-up process and as they use the site.

REFERENCES