

Crafting The Initial User Experience to Achieve Community Goals

Sara Drenner, Shilad Sen, Loren Terveen
Department of Computer Science and Engineering
University of Minnesota
{mosch, ssen, terveen}@cs.umn.edu

ABSTRACT

Recommender systems try to address the “new user problem” by quickly and painlessly learning user preferences so that users can begin receiving recommendations as soon as possible. We take an expanded perspective on the new user experience, seeing it as an opportunity to elicit valuable contributions to the community and shape subsequent user behavior. We conducted a field experiment in MovieLens where we imposed additional work on new users: not only did they have to rate movies, they also had to enter varying numbers of tags. While requiring more work led to fewer users completing the entry process, the benefits were significant: the remaining users produced a large volume of tags initially, and continued to enter tags at a much higher rate than a control group. Further, their rating behavior was not depressed. Our results suggest that careful design of the initial user experience can lead to significant benefits for an online community.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces

General Terms

Design, Experimentation

Keywords

recommender system, entry barrier, user experience, online communities, tagging, user interfaces, new user problem

1. INTRODUCTION

First impressions matter. In online communities, potential members quickly evaluate the community and decide if they want to participate in it. And community designers want to present potential members an accurate picture of the community, including what is expected from members.

For recommender systems, the initial user experience has focused on the “new user problem” [15, 18]. The system must learn

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

RecSys’08, October 23–25, 2008, Lausanne, Switzerland.
Copyright 2008 ACM 978-1-60558-093-7/08/10 ...\$5.00.

enough about a user’s preferences to make accurate recommendations. The goal has been to do this with minimal user effort, thus preventing users from getting discouraged and giving up.

We suggest a different perspective: rather than a new user *problem*, we see a new user *opportunity*. By shaping new users’ experience, a system may influence the amount and type of contributions users make to the community, thus advancing community goals. Our work has three pillars:

- **Consider community goals.** We are interested in two main types of goals: (a) getting desired work accomplished (e.g., eliciting ratings for movies that don’t have many), (b) nurturing desirable user behavior (e.g., welcoming new users).
- **Manage the entry process.** We were inspired by two strands of social psychology research. Research on group “entry barriers” or initiation procedures – ranging from paying membership fees to serving a probationary period – suggests that more “severe” initiations may lead to more committed members [1]. Research on goal setting [13] shows that setting specific, challenging goals stimulates high performance. This work liberated us to take a different perspective on new users. Instead of just minimizing their effort, we explore whether asking them to do *more* leads to more valuable work being accomplished for the community and to more committed community members.
- **Diversify types of contributions.** Recommender systems typically operate within a larger context. This is true in the commercial world: Amazon.com makes extensive use of recommendations, along with product reviews, discussions, customer lists of favorites, tags, etc. It’s also true in research-oriented sites: MovieLens has added features like discussion forums [8] and tagging [17]. Thus, it is important to elicit different types of contributions from users: ratings, tags, forum posts, etc.

We applied these ideas in MovieLens (MovieLens.org). We redesigned the initial user experience, measuring the effect of our changes on work accomplished and user behavior patterns. Specifically, we modified the existing process – which required new users to rate a number of movies – to also require new users to tag some movies. We experimented with several different amounts of tags, thus constructing relatively higher and lower entry barriers. We then measured the proportion of users who made it through the entry barrier in each condition (as well as a control group), and the amount and types of contributions (ratings and tags) they subsequently made to MovieLens.

We organized our research around four main questions. We summarize them and briefly preview our results.

1. *How many new users do we lose by imposing higher entry barriers?* Not surprisingly, the more work we require, the higher the attrition rate. 90% of users completed the easiest condition, while 70% completed the most difficult.
2. *How much work (of the requested type) will users do?* Most users completed even the most difficult entry condition. In this condition, the number of tags entered just by new users represents an increase *by a factor of 5* over the number of tags currently entered by all MovieLens users. Further, the quality of tags entered did not suffer.
3. *To what extent can we “shape” ongoing user behavior?* Requiring users to enter tags during the entry process significantly altered subsequent behavior. After completing the entry process, users in these conditions were 4 times more likely to tag, and entered an order of magnitude more tags on average.
4. *Is user activity a “zero sum game”?* No. Gains in tagging were not purchased with losses in ratings. This is crucial, as ratings are the foundation of a recommender system.

Thus, while asking new users to do more work leads some to “drop out”, we found that the benefits outweighed the costs: the fewer users who completed the entry process did much more work and became more active members of the community.

In the remainder of the paper we survey related work, describe our research platform and methods for quantifying user contributions, outline our experimental design, present our findings and discuss their implications, and close with a brief summary.

2. RELATED WORK

The New User Problem. Until a new user’s preferences are discovered, it is difficult for a recommender system to make personalized recommendations. There are two main ways a system can learn new user preferences: *explicit* and *implicit*. Explicit techniques require users to state preferences for items by rating them on some scale (1 to 5 stars, thumbs-up vs. thumbs-down, etc.). Previous research [15, 18] has focused on clever methods for selecting items for a user to rate. Factors considered include: how likely a user is to be able to rate an item, how much rating a particular item will tell the system about the user’s taste, and how useful a rating would be for the community. However, while community utility has been identified as a factor of interest, algorithms have not yet incorporated it.

Implicit techniques infer user preferences from observed user behavior. Amazon.com treats the items users purchase and view as indicators of interest. Claypool, et. al [4] identified a set of implicit interest indicators, including mouse clicks and movement, time spent, and scrolling behavior, and correlated these with explicit ratings to assess their accuracy. They found that time spent on a page and scrolling behaviors correlated strongly with explicit ratings.

To emphasize, we are not trying to solve the new user problem. Instead of streamlining the process of learning new user preferences, we are exploring what happens when we ask new users to do *more* work.

The Group Entry Process. Off-line groups traditionally have ways to control group membership. They create entry barriers to filter out people who “don’t belong” and enhance the desirability of

the group. Entry barriers take many forms, including required demographic characteristics (American Association of Retired People: age 50 and older), intelligence test scores (Mensa), or income level (costly membership fees). They can require more or less commitment from potential members, ranging from undergoing college fraternity hazing to simply signing up for a campus student group.

Online communities and other web sites have entry barriers, too. Content sites may require users to pay a fee (Angie’s List, Flickr Pro), register for a free account (New York Times), or view advertisements (Salon Premium). Many social networking sites like Facebook and LinkedIn make new users complete a profile containing personal information, preferences, and a list of friends or colleagues.

Social psychology research has shown the power of entry barriers. Aronson and Mills [1] conducted a lab experiment that found a correlation between the difficulty of an entry barrier and people’s level of commitment to a group once they passed the barrier. Intuitively, if you “go through hell” to get into a group, you’re likely to convince yourself that it must be worth it. While this research inspired our idea to experiment with a more challenging entry process, we must be careful in how we apply it to online communities. In particular, subjects in a lab study are unlikely to leave the experiment, and they know it is just an experiment. However, people are used to checking out various web sites and quickly abandoning them if they’re not of interest. Thus, we did not expect Aronson and Mill’s results to transfer directly.

Motivating Contributions to Open Content Communities. Open Content Communities rely on their users to produce valuable content. Wikipedia is the most well-known example, but other sites like Slashdot [12] and Flickr also follow this approach. Recommender systems like MovieLens always have relied on users to enter ratings, and they have moved even further in the direction of Open Content. For example, Cosley [5, 6] experimented with letting MovieLens users enter movies into the database and edit movie information (actors, director, etc.).

When content comes from users, getting users to produce enough – and high quality – content is a key challenge. Cosley developed different algorithms to match movies that needed editing with users who might be interested in doing the editing. He found that intelligent matching techniques led to significantly more editing work being done than random matching. Cosley also showed these ideas could be extended to Wikipedia [7]. Beenen et al [2] took another approach, showing that setting specific, challenging goals (e.g., numbers of movies to rate) gets users to do more work than simply asking them to “do their best”.

Other work has explored at a general level what motivates people to participate in voluntary activities. Building on seminal studies of off-line volunteerism [3] and motivations for participating in open source software projects [10], Nov [14] surveyed Wikipedia editors about why they edited the encyclopedia. The most common responses were because editing was fun and because editors believed that information should be free.

In summary, where prior work on the initial user experience in recommender systems saw a “new user problem”, we see an opportunity. Research on entry barriers and goal setting leads us to consider whether asking more of potential users can increase the amount of valuable work they accomplish and shape their future behavior along useful paths.

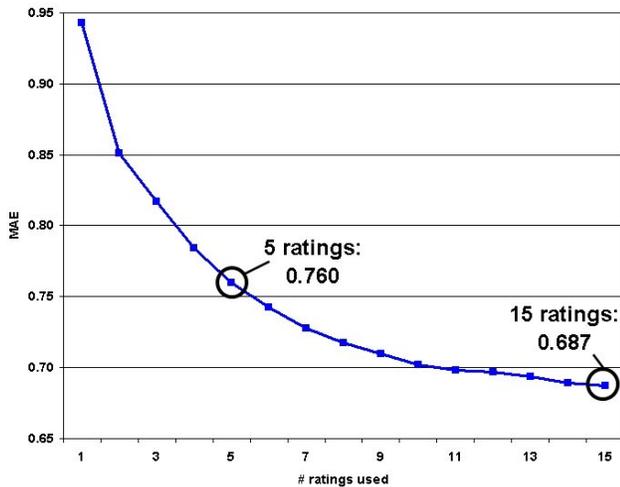


Figure 1: Prediction accuracy using initial ratings to predict subsequent 50 ratings

3. REDESIGNING THE MOVIELENS ENTRY PROCESS

We conducted our research in the MovieLens film recommendation web site, run by the GroupLens research group at the University of Minnesota. MovieLens enables users rate movies and receive recommendations, tag movies with descriptive keywords, and participate in movie-oriented discussions.

Current Entry Process. MovieLens learns user preferences by requiring each new user to rate 15 movies. Once users have done this, they are free to use the system as they choose.

Redesigning the Entry Process. Recall that the aim of our research is to investigate how different sizes of entry barriers affect how much “work” users do during the entry process and what their subsequent usage is like. Entering movie ratings already is an entry barrier, so an obvious path would be to manipulate the number of ratings new users enter.

However, we chose not to do this because – as detailed below – the number of initial ratings users enter affects the accuracy of the recommendations they receive. Since MovieLens is a real web site that supports an active user community (over 150 active users on any given day), we want all users to receive equally accurate recommendations. Therefore, we instead chose to introduce an additional entry barrier: entering tags for movies. Further, we didn’t want to make the entry process too onerous, since this could result in too many new users getting frustrated and abandoning the site. Therefore, we sought to minimize the number of ratings required while still preserving recommendations accuracy.

To find a lower number of ratings that would retain an acceptable level of predictive accuracy, we “replayed” the history of a sample of users from the MovieLens database, while asking the question: how accurately could we recommend to this user if he/she only rated 1 (2, 3, 4, ..., 15) movies? We selected a random sample of 5000 users with more than 65 ratings (about 30% of all users). For each user, we created a chronological history of ratings. We constructed an initial test set beginning with the user’s first rating, then measured the difference between predicted and actual rating for the next 50 movies in the user’s rating history. This was repeated for all 5000 users to determine the mean absolute error (MAE). The first point in Figure 1 shows the result of the first iteration. The



Figure 2: Screen shot of search results with tags



Figure 3: Thumbs-up, thumbs-down tag rating interface.

remaining points represent the MAE as the test set size increased from 2 to 15 ratings.

From these results, we decided that asking new users to enter just 5 ratings would be acceptable. It is low enough that we felt comfortable experimenting with a relatively high tagging requirement. Yet the prediction accuracy is high enough that we did not believe that user satisfaction with the system would suffer. MAE increased by 0.073, from 0.687 for 15 ratings to 0.760 for 5 ratings.

Overview of tagging in MovieLens. We chose tagging as the additional entry task for several reasons. Tags are familiar to most users, being featured in popular social websites like Flickr and del.icio.us, as well as major e-commerce sites like Amazon.com. Tags are useful: they help users evaluate movies, aid navigation, and support browsing to discover movies with common themes. Therefore, we believed that users would be willing and able to tag, and that tags they entered would be a valuable asset to the system. Finally, each “unit” of work – tagging one movie with one tag – is easy to accomplish. Therefore, we could create a very low entry barrier – tagging just a few movies – as well as an arbitrarily high one – simply asking users to tag many movies. We detail the barriers below.

Tags are shown in several places throughout MovieLens. When a user does a search, the results page shows movies and associated information including title, year of release genre, and tags (Figure 2). Tags are divided into those added by the user (“Your tags”) and popular tags added by other users. Tags also appear on movie “Details” pages; Figure 3 shows the tags applied to the movie *Juno*.

4. QUANTIFYING USER CONTRIBUTIONS

In order to measure the effectiveness of entry barriers, we must establish metrics that reflect the quantity and quality of work completed by a user. We first focused on examining *what* types of behavior should be included in our metrics. In MovieLens, the possi-

ble types of contributions include ratings, tags, discussion forum posts, and movie information edits. We chose to include ratings because they were critical to recommender system performance. We also included tags to measure the effects of entry barrier conditions. We considered counting forum posts and movie edits, but concluded that the level of these activities was too low to see significant differences¹.

Second, we explored *how* we should count tags and ratings. While a simple count is useful, it isn't sufficient, since not all contributions are equal. Therefore, we introduced a notion of *value*. For instance, if a user rates a movie with 8000 ratings, this may have less value to the community than a rating for a movie with 10 ratings [15]. Furthermore, if a user enters a purely personal tag like "Want to rent", this might not be as valuable as an objective tag such as "Oscar-winner".

We distinguish between two aspects of value: intrinsic *quality* and *influence* on the community.

4.1 Tags: Quality

The quality of a tag is subjective: one user may consider "cheap Star Wars ripoff" a perfect description of a movie, while another vehemently disagrees. Some tags are useful only to the person who entered them, like "in netflix queue". We measured tag quality using collective opinion of the user community. We were able to do this because MovieLens includes an interface that lets users rate tags as "thumbs up" (good tag, I agree) or "thumbs down" (poor tag, I don't agree) for a tag [16]. Over 90% of tags in MovieLens have been rated. Tag ratings are independent from movie ratings, and currently have no impact on movie recommendations and predictions.

Figure 3 shows the tags associated with the movie *Juno*. Two icons appear to the right of each community tag. These icons let users rate tags thumbs up or thumbs down. A tag rated thumbs-up is moved to the top of the list of community tags, thus increasing its visibility. A tag rated thumbs-down is moved to the bottom of the list of community tags. More than 75% of tag ratings in MovieLens are thumbs-down². When enough users rate a tag thumbs-down, it will no longer be shown in the short list of tags in search results (Figure 2).

4.2 Ratings

For ratings in a recommender system, influence is easy to define (at least intuitively): how much does entering a rating for a movie affect how that movie and other movies are recommended in the future? Defining rating quality is harder.

4.2.1 Quality

The very point of a recommender system is to accommodate subjective preferences: one user may hate "No Country for Old Men", while another loves it. An algorithm finds the proper neighborhood for each user, and like-minded users in each neighborhood become a source of accurate personalized recommendations.

Although it seems counterintuitive to objectively measure the quality of a subjective rating, researchers who study malicious attacks on recommender systems address similar issues. *Shilling* is the practice of entering random or insincere ratings to try to manipulate the recommendations of a given item. Shilling typically appears in e-commerce settings, where a company might try to

¹To be safe, we did look for differences in our analysis. As we expected, we did not find any.

²We speculate that this is due to the high subjectivity of movie preferences and because the set of tags has not reached maturity yet: i.e., the good tags have not yet "crowded out" the poor ones.

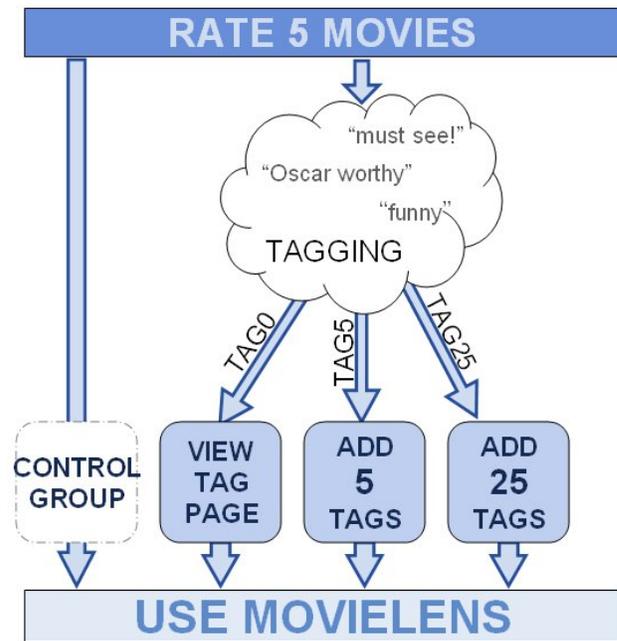


Figure 4: Overview of Experimental Design

get manipulate their products' ratings (up) and their competitors' products' ratings (down). Although this is an active area of research [11], there are no standard metrics for us to apply. Furthermore, it is not obvious that metrics developed for shilling detection would be appropriate for our purposes. Due to these considerations, we do not attempt to measure rating quality in this research.

4.2.2 Influence

In a collaborative filtering system, ratings are used to create predictions for other users or items. Thus, a rating's impact can be measured by how much it influences predictions.

To quantify the impact of rating a movie, we created a model from a subset of the MovieLens ratings data. We calculated the average change in prediction for the target movie and all other movies each time a rating was made. We binned movies by the number of ratings they had, and then calculated an average influence for each bin. Each time a user rated a movie, they were allocated a number of credits according to how many ratings the movie had. Users were given more credits for movies with fewer ratings, as rarely rated movies are greatly influenced by any ratings they receive. Movies with many ratings earned less credit, as rating a frequently rated movie has little impact on other movies, and even less impact on the target movie.

5. EXPERIMENTAL DESIGN

Figure 4 gives an overview of our experimental design. We assigned new MovieLens users to four groups, a control group and three experimental groups. All users were required to rate five movies. After doing this, the control group was immediately directed to MovieLens, while the experimental groups were taken to the tagging interface (Figure 5). Users were able to navigate through the list of movies in order to find movies they were able to tag. Users could tag any movie, whether or not they had rated it. Tags created by users in this experiment immediately became



Figure 5: Screen shot of the tagging interface shown during the entry barrier

“live” in MovieLens, so the experimental users created real value for the community.

Research on “cognitive dissonance” [1, 9] predicts that requiring more work will make users who do complete the entry barrier value MovieLens more, and thus (we hope) contribute more. However, it isn’t clear what the proper size of the barrier is: too small, and the effect may be minimal; too large, and the added value produced by the users who complete the entry process may not make up for those who do not complete. Therefore, we experimented with several different “sizes” of entry barriers to try to find an effective setting.

The *high barrier* group was required to enter 25 tags; we refer to this group as TAG25 throughout the rest of this paper. The *low barrier* group was required to add five tags (TAG5). Note that the requirement was for tags, not movies. In other words, one TAG25 user could apply 25 tags to one movie, and another could apply one tag to each of 25 different movies.

We also created an *exposure only* group. This group was shown the tagging interface, but was not required to enter any tags; we refer to this group as TAG0. This condition let us distinguish two possible influences on user behavior: requiring them to do a task (e.g., enter tags) and simply exposing them to the possibility of doing that task. TAG0 users could use the tag rating interface as much as they wanted or simply click a link to go directly to MovieLens.

New MovieLens users were enrolled in the experiment over a period of 45 days, and were randomly assigned to a condition. In the first week, we experimented with several different values for the low and high entry barriers, and we only realized the need for the TAG0 condition after running the experiment for about a week. Therefore, the groups did not end up with equal numbers of users. Of the 583 users enrolled in the experiment, 177 were assigned to the control group, 117 were assigned to TAG0, 161 were assigned to TAG5, and 128 were assigned to TAG25. Our analysis all scale for the difference in group size, and the statistics we use are appropriate for groups of different sizes.

6. RESULTS

For each of our four research questions, we computed several metrics to answer that question.

1. *How many new users do we lose by imposing higher entry barriers?* To answer this question, we simply compute the proportion of users in each condition who complete the entry process.

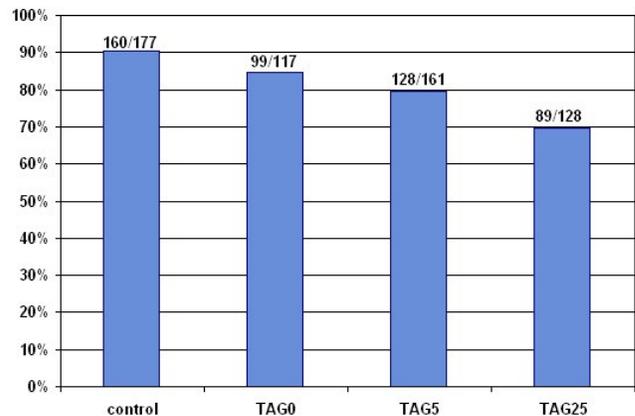


Figure 6: Percentage of users who completed the entry barrier

2. *How much work (of the requested type) will users do?* For this question, we first compute the number of tags entered by users in each condition, both during the entry process (“pre-entry”) and on subsequent visits to MovieLens (“post-entry”). We also measure tag quality.
3. *To what extent can we “shape” ongoing user behavior?* This question is related to the previous one; however, it focuses on the “steady state” behavior, that is, what users in each condition do after they have completed the entry process. Put simply, are users who are forced to enter tags (or exposed to the opportunity to tag) during the entry process more likely to enter tags when they return to MovieLens? For this question, we compare post-entry behavior across conditions, specifically, the number and quality of tags entered.
4. *Is user activity a “zero sum game”?* Suppose we can get users to enter more tags – do they make up for this by entering fewer ratings? For this question, we compare the number (and influence) of ratings entered by users in each of the conditions.

Who counts as “in” an experimental condition? There are two possible answers to this question: (1) all users who were assigned to a condition, and (2) all users who completed that condition’s entry process. Each answer is useful for some purposes. Generally, for pre-entry activities we prefer the first answer, and for post-entry activities we prefer the second answer. Put another way, the first approach is more useful for measuring total work accomplished, and the second is more useful for understanding the kind of user “created” by each entry condition. When we report results, we state clearly which of these two definitions we are using.

6.1 How many new users do we lose by imposing higher entry barriers?

Figure 6 shows the percentage of people who completed the entry process for each condition. As one would expect, the higher the entry barrier, the lower the percentage of users who completed the entry process. The control group had the highest completion rate at 90% (160/177). The completion rates for the experimental groups are as follows: TAG0 85% (99/117), TAG5 80% (128/161), and TAG25 69% (89/128). A likelihood ratio Chi-Squared test shows that the differences between conditions are statistically significant ($\chi^2=22.7$, $DF=3$, $p<.0001$).

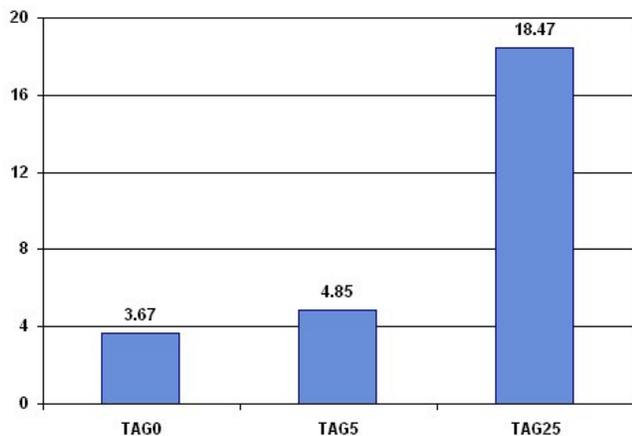


Figure 7: Average number of pre-entry tags, per user

6.2 How much work (of the requested type) will users do?

As we mentioned, there are two sources of tags that must be considered in answering this question. The first source is the entry process – when (some) users are required (or given the opportunity) to enter tags – and the second is subsequent normal system usage. We focus on the first source now, and deal with the second source next.

Tag Count. Figure 7 summarizes the number of tags entered during the entry process in each condition. We can draw several interesting conclusions from these data.

1. The entry barriers “worked”. Since most users completed the entry process for all conditions, the average number of tags entered per user assigned to each condition approached the requested number, with the shortfall due to users who dropped out before completing. TAG25 users entered 18.47 tags (on average), TAG5 users entered 4.85 tags, and TAG0 users entered 3.67 tags. A Wilcoxon test showed that TAG25 was significantly greater than both TAG5 and TAG0 ($\chi^2 = 117.2$, $DF = 2$, $p < 0.0001$).
2. Instituting any of the TAG conditions as the normal MovieLens entry process would completely transform the number of tags the system collects. Currently, MovieLens users enter 85 about tags per day on average, and about 25 new users join MovieLens everyday. If all 25 users were assigned to TAG25, the average number of tags entered per day would increase by a factor of 5 (from 85 to 461 ($= 25 * 18.47$)). If either TAG0 or TAG5 were used, the number of tags entered each day would approximately double.
3. TAG0 looks pretty competitive with TAG5: even though TAG0 users weren’t required to enter any tags, the number of tags they entered was statistically indistinguishable from TAG5 users.

There is one caution to our conclusion that the entry barriers worked: what if users simply rushed through the tagging process, entering poor quality tags to complete as quickly as possible? Our next two analyses investigate this issue.

Tag Quality. We measure the quality of a tag as the percentage of “thumbs up” votes the tag receives. Figure 8 shows average tag quality for each condition. We user-normalized tag averages by first calculating each user’s average rating for a tag, and then averaging

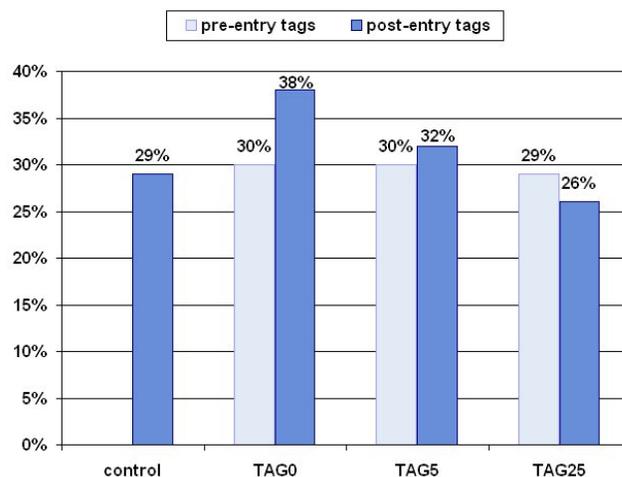


Figure 8: Tag quality by condition, measured by percentage of thumbs-up tag ratings

each user’s result. User-normalization reduced the effects of power users such as the user who rated the tag “comedy” thumbs-down on 206 different movies.

None of the differences between conditions were statistically significant. Further, the experimental observations (the tags themselves) are not truly independent since a single user creates multiple tags. This further reduces the apparent differences between conditions. Therefore, the potential concern that users in TAG5 and (especially) TAG25 entered poor quality tags during the entry process were unfounded. The vast increase in quantity was not purchased with a decrease in quality.

6.3 To what extent can we “shape” ongoing user behavior?

The TAG conditions led to a burst of tags during the entry process. However, a different issue was whether the initial exposure to tagging affected subsequent (post-entry) user behavior. Put simply, did we create “taggers”?

We did two analyses to answer this question. For each condition, we computed the percentage of users who tagged at all after the entry process and the average number of post-entry tags entered per user per day.

There are significant differences between the conditions in the number of users who tagged during the post-entry phase (Figure 9). Only 5.6% of the control group and 5.1% of the TAG0 group entered any tags after the entry process. Contrast this with users who were required to tag during the entry process: 19.5% of TAG5 users, and 24.7% of TAG25 users entered at least one tag during the post-entry phase. A likelihood ratio Chi-Squared test again reveals that TAG5 and TAG25 are significantly different from TAG0 and control ($n = 476$, $\chi^2 = 28.1$, $DF = 3$, $p < 0.0001$).

This analysis was done only for users who completed the entry barrier. However, we can also do it for all users assigned to each condition. Again, the differences between TAG5 (15.5%) and TAG25 (16.4%) on the one hand, and TAG0 (4.3%) and control (5.1%) on the other were significant (likelihood ratio chi-squared test, $n = 583$, $\chi^2 = 20.7$, $DF = 3$, $p < 0.0001$). In other words, being assigned to tag movies during a user’s initial experiences made him or her more likely to tag later. This is important, because it shows that the entry process can shape subsequent user

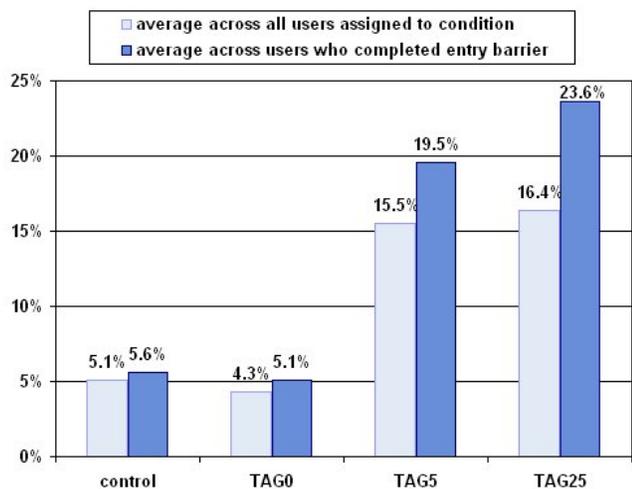


Figure 9: Percentage of users who added at least one tag post-entry

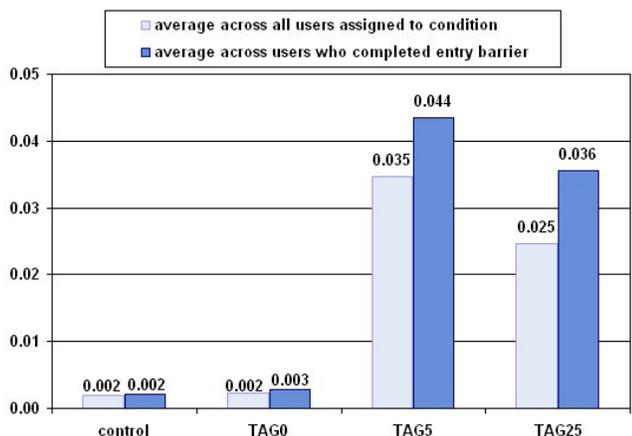


Figure 10: Average number of tags added post-entry, per user, per day

behavior, not just filter people who already were more likely to exhibit a particular behavior.

Figure 10 summarizes the amount of post-entry tagging for users in each condition. While all the numbers were small, TAG25 and TAG5 users entered an order of magnitude more tags than control and TAG0 users ($\chi^2 = 29.1$, $DF = 3$, $p < .0001$).

6.4 Is user activity a “zero sum game”?

While we have shown that we were able to induce new post-entry behavior in users, a final question is whether we did so at the cost of existing behavior. In particular, did users who tagged more rate less? Was it a zero-sum game?

Figure 11 shows the average number of post-entry ratings per user per day. It shows two averages per condition. The darker bars represent an average across only those users who completed their entry barrier, and the lighter bars show the average across all users who were assigned to that condition. In neither case were there any significant differences (for users who completed the entry barrier: Wilcoxon test, $\chi^2 = 2.1$, $DF = 3$, $p = .56$; for all users assigned to a condition: Wilcoxon test, $\chi^2 = 3.3$, $DF = 3$, $p = .34$).

We also applied the ratings influence metric to post-entry ratings.

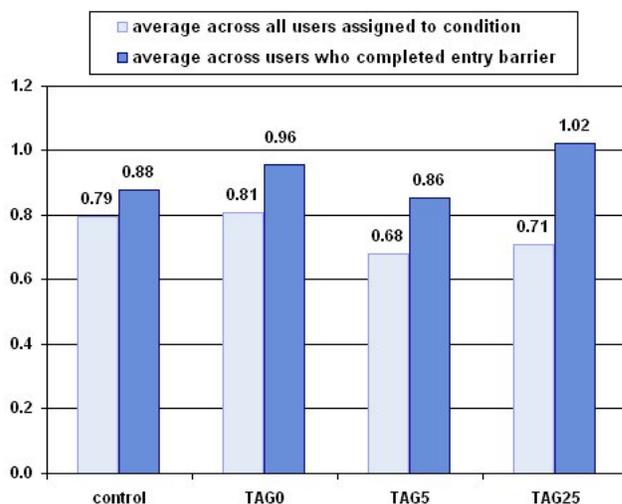


Figure 11: Average number of post-entry ratings per user, per day

Again, there were no significant differences between conditions. Thus, the increase in tags was not purchased with a decrease in (quantity or quality of) ratings.

7. SUMMARY

Our findings are encouraging. They suggest that designers of a recommender based community – and perhaps other types of online communities – can tailor the entry process to achieve community goals. We showed that introducing a tagging requirement to the MovieLens entry process resulted in a huge increase in the number of tags entered into the system with a relatively small increase in user attrition. Further, requiring users to tag during their initial experience significantly increased the probability they would tag in subsequent visits: thus, we appear to have shaped the long-term behavior of MovieLens users. Interestingly, this was true only when new users were required to tag, not when they were simply given the opportunity to do so. Finally, when users tagged more, they did not rate any less, so user activity did not appear to be a zero-sum game.

Our results suggest much interesting future research. First, trying our approach in other online communities would help establish generality of our results. Second, we would like to expand our focus beyond the entry process, seeking to shape user behavior throughout the user life cycle by suggesting tasks for users to perform. Finally, we would like to develop a unified metric for quantifying different types of user contribution to an online community.

8. ACKNOWLEDGMENTS

We gratefully acknowledge our experimental subjects and members of the GroupLens research group at the University of Minnesota. This research was supported by the National Science Foundation, grant IIS 03-2485.

9. REFERENCES

- [1] E. Aronson and J. Mills. The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 59:177–181, 1959.

- [2] G. Beenen, K. Ling, K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. E. Kraut. Using social psychology to motivate contributions to online communities. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 212–221, New York, NY, USA, 2004. ACM.
- [3] E. Clary, M. Snyder, R. Ridge, J. Copeland, A. Stukas, J. Haugen, and P. Miene. Understanding and assessing the motivations of volunteers: A functional approach. *Journal of Personality and Social Psychology*, 74:1516–1530, 1998.
- [4] M. Claypool, P. Le, M. Wased, and D. Brown. Implicit interest indicators. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pages 33–40, New York, NY, USA, 2001. ACM.
- [5] D. Cosley, D. Frankowski, S. Kiesler, L. Terveen, and J. Riedl. How oversight improves member-maintained communities. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 11–20, New York, NY, USA, 2005. ACM.
- [6] D. Cosley, D. Frankowski, L. Terveen, and J. Riedl. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1037–1046, New York, NY, USA, 2006. ACM.
- [7] D. Cosley, D. Frankowski, L. Terveen, and J. Riedl. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 32–41, New York, NY, USA, 2007. ACM.
- [8] S. Drenner, M. Harper, D. Frankowski, J. Riedl, and L. Terveen. Insert movie reference here: a system to bridge conversation and item-oriented web sites. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 951–954, New York, NY, USA, 2006. ACM.
- [9] L. Festinger and J. Carlsmith. Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*.
- [10] A. Hars and S. Ou. Working for free? - motivations of participating in open source projects. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 7*, page 7014, Washington, DC, USA, 2001. IEEE Computer Society.
- [11] S. K. Lam and J. Riedl. Shilling recommender systems for fun and profit. In *Proceedings of the 13th international conference on World Wide Web*, pages 393–402, New York, NY, USA, 2004. ACM.
- [12] C. A. Lampe, E. Johnston, E. Johnston, and P. Resnick. Follow the reader: filtering comments on slashdot. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1253–1262, New York, NY, USA, 2007. ACM.
- [13] E. A. Locke and G. P. Latham. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9):705–717, 2002.
- [14] O. Nov. What motivates wikipedians? *Commun. ACM*, 50(11):60–64, 2007.
- [15] A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl. Getting to know you: learning new user preferences in recommender systems. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 127–134, New York, NY, USA, 2002. ACM.
- [16] S. Sen, F. M. Harper, A. LaPitz, and J. Riedl. The quest for quality tags. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 361–370, New York, NY, USA, 2007. ACM.
- [17] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190, New York, NY, USA, 2006. ACM.
- [18] P. Viappiani, P. Pu, and B. Faltings. Conversational recommenders with adaptive suggestions. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 89–96, New York, NY, USA, 2007. ACM.