Bootstrapping Wikis: Developing Critical Mass in a Fledgling Community by Seeding Content

Jacob Solomon
Michigan State University
solomo93@msu.edu

Rick Wash
Michigan State University
wash@msu.edu

ABSTRACT

Online communities depend on content contributed by their members. However, new communities have not yet achieved critical mass and are vulnerable to inadequate contribution. To encourage contribution, many fledgling communities seed the site with data from 3rd parties. We study the effectiveness of such seeding by looking at how people react to different types of seeded content. We found that people make larger contributions when there is no seeded content. But when there is seeded content, users learn from that content and contribute similar types of content. Therefore, if websites prefer specific types of contributions, seeding that type of contribution can be a valuable way to elicit appropriate contributions.

Author Keywords

online contribution, bootstrapping, critical mass

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Miscellaneous

General Terms

Human Factors

INTRODUCTION

System designers frequently seed content into newly created social media sites in an attempt to establish expectations and elicit participation. This seeded content presents the appearance of site activity for first-time visitors, and can be interpreted by visitors as the behavior of other community members. This is an attempt at *bootstrapping* participation on the site.

The perception of site activity by first-time visitors can establish expectations of the community's utility and the behavior which is afforded on the site. These expectations can then influence a user's motivation to contribute to the community. Motivations to contribute to online communities can be diverse and complex. Motivations can vary across dimensions of function or sociality [1]. They can also be related to expressions of values or ideology [6] or simply a desire to provide information [4]. If newcomers perceive the affordances

or utility of a site to be incompatible with their motivations to contribute, the site could have difficulty in developing a critical mass of participation.

In this study, we hope to inform the design of online communities by examining the influence of bootstrapping on the behavior of contributors. We investigate how seeding content to a site can be used to influence user contributions. This influence may be the result of a perception on the part of the users that the existing content on a site is a representation of the behavior of other community members. The content on the site can serve as a model from which first-time users can learn. Social cognitive theory is the basis for this argument. Social cognitive theory asserts that people learn through the observation of models [2]. First-time users of a site can use the seeded content as a model for their own behavior.

It is important that the model of behavior provided through seeded content be congruent with the goals of the site. As an obvious example, seeding spam would be counter-productive to the goals of most sites. The behavior modeled in seeded content must be consistent with visitor's motivations to contribute. Inconsistencies in incentives and motivations for a particular behavior can be de-motivating [7]. Because there is considerable variance in motivation to contribute to online communities, it can be difficult to model behavior through seeded content which can reliably motivate first-time visitors to contribute.

For this reason, it is important to understand the contexts in which bootstrapping can effectively be used in congruence with the goals of an online community, as well as the contexts in which it is best avoided. Seeded content provides a model of behavior for first-time contributors. But which characteristics of contributions are detected by users as being the behavior of community members which can be modeled?

We have looked at the type of information which has been contributed as a marker of the behavior of community members. We have examined variance in information type across two dimensions. First, contributions vary in their subjectivity, ranging from factual to opinionated. Second, some contributions can provide structure to a community, such as new pages, headings, threads, or sections. The type of seeded content can be interpreted by newcomers as the behavior of existing community members, thus influencing newcomers' behavior.

RQ1. When users are given a behavioral model in the form of seeded content, will their contributions be different than if they are given no model?

Table 1. Specific Hypotheses

H1.1: Users in the blank group will contribute more structural information than users in the seeded groups.

H1.2: Users in the blank group will contribute less factual information than users in the seeded groups.

H1.3: Users in the blank group will contribute less opinionated information than users in the seeded groups.

H1.4: Users in the blank group will contribute less total information than users in the seeded groups.

H2.1: The percentage of sentences containing facts will be greatest for users in the factual group than for users in the opinion or structure groups.

H2.2: The percentage of sentences containing opinions will be greatest for users in the opinion group than for users in the factual or structure groups.

H2.3: The percentage of structural sentences will be greatest for users in the structural group than for users in the factual or opinion groups.

Since social cognitive theory explains that the behavioral models perceived by users will influence their behavior, it is important to understand what users will do when they have no such models. Because the blank wiki contains *no* structure, we hypothesize that the blank group will be compelled to add a greater percentage of structure than the seeded groups. However, we hypothesize that users will contribute more non-structural content, as well as more total content, when given a wiki seeded with content. Hypotheses 1.1–1.4 (in Table 1) operationalize this research question.

RQ2. Will users follow the model provided in seeded content by contributing the same type of information?

Answering this question is of value to designers who want to direct the behavior of contributors through bootstrapping. In contexts where certain types of information are more desirable to the goals of a site than other types of information, designers may consider bootstrapping so that contributors follow a carefully selected model of behavior. The extent to which users follow the model of behavior in contributing, as well as the strength and boundaries of this effect, is important in determining how to bootstrap. Likewise, a test of information type as a marker of the behavior of community members can help designers determine whether or not to vary information type in order to elicit the desired perceptions of behavior by first-time visitors. These questions have been examined in Hypotheses 2.1, 2.2, and 2.3.

DATA

Students in introductory communications courses were invited to participate in the study. Students were told that participation in the study would lead to the creation of a wiki which would be an open, community-generated resource for information about communications related courses at the university. Additionally, students who agreed to spend at least

20 minutes (in their own time outside of class) making contributions to this wiki and who would complete a survey about their contribution were offered a small amount of extra-credit in their course (approximately 1% of their final grade; all subjects received the same amount of credit regardless of contribution.). 96 students fulfilled this request. In this experimental stage of the study, a new wiki (using MediaWiki software) was created for each participant based on one of four randomly assigned templates (rather than taking all participants to the same wiki).

The *factual* template listed six courses along with strictly factual information taken from the university registrar's course descriptions (i.e. instructor name, number of credits, basic description). The *opinion* template listed the same courses, but included opinionated statements about the course. These statements were paraphrased (to best fit the experimental context) from actual course reviews on publicly available websites in which students share information about courses and instructors. The *blank* template gave only a title and no other information. The *structural* template listed some categories of classes as section headings (i.e. Advertising, Retailing etc.) and listed links to pages for a few courses. These pages were blank, and the structural template contained no other content.

After the experiment, the text contributed by participants was separated from template text and divided into sentences. Each contributed sentence was manually coded. A sentence was coded as *factual* if it contained any factual information and no opinionated information. An example of a factual sentence would be "The class had four tests." A sentence was coded as *opinion* if it contained opinionated information with no strictly factual information. "This class was boring." Some sentences were coded as containing *both* factual and opinionated information. "The class had two tests, but they were very easy." The amount of factual information contributed was measured as the percentage of sentences coded as *factual* or *both* from the entire contribution. Opinionated information was measured the same way, and *structural* information was the percentage of sentences that only provide structure (headings or new page links) to the wiki from the entire contribution.

METHODS

Two-tailed t-tests were used to compare the mean information ratio of the blank group to all other groups for H1.1, H1.2, and H1.3. Post-hoc OLS regression models were used to clarify the results and provide more precise comparisons between the blank group and each other group. The total number of sentences contributed was found to be a covariate with the type of information contributed and was included in these models. A two-tailed t-test was also used to compare the total amount of information contributed (total number of sentences) between the blank group and the non-blank groups (H1.4). OLS regression provided an additional test of this hypothesis.

OLS regression was used to test H2.1, H2.2, and H2.3. A model was built for each sub-hypothesis which regressed the experimental condition (structural, factual, or opinionated) and the total number of sentences contributed on the ratio of structural, factual, or opinionated information. In each model,

Table 2. T Test statistics for hypotheses tests of Hypotheses 1.1-1.4.

	# Sentences	Facts	Opinions	Structure
Blank	19.7	70%	38%	8%
Non-Blank	10.0	59%	17%	25%
Difference	9.7	11%	21%	-17%
P-Value	0.053	0.107	0.008	0.000

Table 3. OLS regressions for Hypotheses 2.1-2.3. Each column is a separate regression using either the Fact, Opinion, or Structure Ratio as the dependent variable. The empty level for the condition IV was used as the baseline in each regression and corresponds to the type of content that matches the DV.

	Fact Ratio		Opinion Ratio		Structure Ratio	
(Intercept)	70%	***	22%	***	45%	***
Factual			-14%	*	-27%	***
Opinion	-26%	***			-10%	
Structure	-33%	***	-10%			
# Sentences	+1.0%	**	+0.2%		-0.8%	*
R^2	0.29		0.07		0.22	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

the base condition corresponded to the congruent dependent variable. This allows us to see most clearly how contributions differed between condition in which the type of information of the dependent variable was seeded, versus the conditions in which it was not seeded.

RESULTS

Table 2 describes the differences between the non-blank groups and the blank group, as well as the t-tests. H1.1 was disconfirmed. Contrary to our prediction, the non-blank groups added *more* structural information than the blank group. H1.2 was not confirmed. Participants in the blank group actually contributed a higher ratio of facts than participants in the non-blank groups. However the difference was not statistically significant. The regression analysis determined that only the structural group contributed less factual information than the blank group. H1.3 was disconfirmed. Participants in the blank group contributed a *higher* ratio of opinionated information than participants in the non-blank groups. The post-hoc OLS regression analysis revealed that each group with seeded content contributed less opinionated information than the blank group. H1.4 was disconfirmed, as participants in the blank group contributed *more* sentences than participants in the non-blank groups. The t-test approached significance, and the regression analysis determined that the difference between the blank group and each of the other groups was significant.

Table 3 shows the results related to RQ2. H2.1 was confirmed, as participants who saw factual information on the wiki added a greater percentage of factual information than participants who saw either structural or opinionated information. H2.2 was partially confirmed. Seeing opinionated information led participants to contribute more opinions than those who saw factual information. The opinion group also added more opinions than the structure group, but the difference was not significant. H2.3 was confirmed. Participants who saw only structural information contributed more structural information than those who saw factual or opinionated information.

DISCUSSION

RQ1. When users are given a behavioral model in the form of seeded content, will their contributions be different than if they are given no model?

Participants who were shown a blank wiki contributed far more than those who saw seeded content, suggesting that a blank wiki was more motivating to participants than seeded content. It could be that a blank slate gives users more freedom, which they took up. Self-determination theory [3] ar-

gues that autonomy is an important characteristic of intrinsic motivation, particularly in creative behavior [7]. The presence of seeded content and the cognition of social expectations may have inhibited users' perception that complete autonomy of contributions is afforded.

Ling et al. [5] found that users contribute more when they feel their contributions are unique. A similar effect may have occurred in this study, as users contributing to a blank wiki may have been cognizant that their contributions will be salient and important, leading to greater motivation to contribute.

The stark contrast between the level of contribution in the blank group and contribution in the structure group is important to note. The structure template essentially offered participants a coloring book approach to editing, with blank sections and pages that needed filling. Like the blank template, there was no real content. Seeding a community with only structure could be harmful to eliciting contribution because it may detract from user's autonomy and provide an insufficient model from which expectations can be established. Since users in the blank group contributed very little structure, we conclude that users are likely least intrinsically motivated to add structure compared to other types of content.

The results of H1.1 and H1.2 can be explained by the significant portion of sentences in blank wikis which were coded as containing both factual and opinionated information. Users in the blank condition appeared more likely to add opinionated clauses to factual statements. For example, "This is a photography class that is pretty fun." The factual statement that the class is about photography is qualified with an opinion that it is fun. A post-hoc analysis using OLS regression revealed that the blank group had significantly more statements containing both fact and opinion (17% of all sentences) than any of the seeded groups. If separating factual information from opinionated information is an important goal for the design of a site, designers must consider ways to do this without demotivating users through structure.

RQ2. Will users follow the model provided in seeded content by contributing the same type of information?

The results of this study suggest that users do look at the nature of previously contributed information as a model for their own behavior on the site. There are several possible mechanisms which might explain this behavior. This may be an

example of social cognition [2] whereby users learn about the functionality, utility, or expectations of a site by observing how others behave. The seeded content on the site may also have primed users to think about certain types of content when considering their own contributions. Regardless of the mechanism, the results suggest that seeding content at an early stage of development can direct future contributions.

The results suggest that the nature of seeded content must be carefully thought through and must be consistent with the goals of the community. A community which seeks to elicit specific types of contributions may benefit from bootstrapping. A community seeking diversity in contributions may find that goal inhibited by bootstrapping, and may prefer to provide users a blank slate for contribution. These results have application to communities at more mature stages of development as well. An established community which finds an imbalance in the types of contributions it receives may find that seeding content can lead to more balanced contribution. For example, a wiki with considerable information but lacking structure can seed some structure into the wiki and make those contributions salient. These data suggest that users would react to such contributions by increasing the ratio of structural contributions.

LIMITATIONS
As an experiment, this research falls between a true laboratory experiment (with high internal validity) and a true field experiment (with high external validity). Since the subjects did the experiment on their home computers, we cannot control the environment the subject is in and cannot ensure that subjects actually worked on the wiki for 20 minutes. It is possible that some subjects (such as those in the blank condition) spent less time reading than others, and therefore spend more time contributing. This is a possible explanation for the blank condition results. It is also possible that subjects were trying to do the minimum work to get the extra credit. However, since that should be true across conditions, this should not be a confound for the study. These issues, however, potentially limit the external validity of the study. We cannot be sure that extra credit motivates users in similar ways to whatever motivates contributions to wikis online. However, this principle of least effort would seem to apply to other types of motivations, whether it be to earn reputation or merely help a friend. People are not likely to do more than what is required to satisfy their motivation. So the differences between conditions do still suggest that bootstrapping can influence behavior. Also, the errors for the OLS models predicting structure ratio and opinion ratio were not normally distributed, violating an assumption of OLS regression making the significance of those models somewhat questionable.

CONCLUSION
Many new social media systems use bootstrapping — seeding the site with third party content — as a way to encourage users to contribute. We found that users tend to contribute more content, and more unstructured content, when they are given a blank slate. This suggests that bootstrapping is not always a positive. However, users tend to contribute content roughly similar to any seeded content. Bootstrapping can be

used to direct user effort toward contributing specific types of content. Site designers should carefully consider the goals of the site when seeding content.

Future work can look for markers of behavior other than information type which provide models for newcomers. Some possibilities include the style of language, chronemic characteristics, the size of individual contributions, variations in the structure of contributions, and other characteristics. For example, the discussion surrounding a wiki article may be useful to newcomers by providing a model of behavior which encourages direct communication and collaboration. The extent to which these characteristics can motivate newcomers to contribute or direct their future behavior on a site can inform designers in how they can best bootstrap a fledgling site. Evaluating these characteristics would then be an important extension of the current work.

REFERENCES
1. M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, San Jose, California, USA, 2007. ACM.

2. A. Bandura. *Social foundations of thought and action: a social cognitive theory*. Prentice-Hall series in social learning theory. Prentice-Hall, Englewood Cliffs, N.J., 1986. xiii, 617 p.

3. E. L. Deci and R. M. Ryan. *Intrinsic motivation and self-determination in human behavior*. Springer, Aug. 1985.

4. C. Lampe, R. Wash, A. Velasquez, and E. Ozkaya. Motivations to participate in online communities. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1927–1936, Atlanta, Georgia, USA, 2010. ACM.

5. K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, X. Lee, D. Cosley, D. Frankowski, L. Terveen, A. M. Rashid, P. Resnick, and R. Kraut. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4), 2005.

6. O. Nov. What motivates wikipedians? *Commun. ACM*, 50(11):60–64, 2007.

7. D. H. Pink. *Drive: the surprising truth about what motivates us*. Riverhead Books, New York, NY, 2009. 242 p.