Critical Mass of What? Exploring Community Growth in WikiProjects

Abstract

Fledgling online communities often hope to achieve *critical mass* so that the community becomes sustainable. This concept however is not well understood. At what point does a community achieve critical mass, and how does the community know this? Furthermore, online communities become sustainable when they achieve a mass of what? We explore this question by analyzing growth in a large number of online communities on Wikipedia. We find that individual communities often have different patterns of growth of membership from its pattern of growth of contribution or production. We also find that in the early stages of community development, building membership has a greater impact on community production and activity in later periods than accumulating many contributions early on, and this is especially true when there is more diversity in the early participants in a community. We also show that participation from a community's "power users" in its early stage is not as valuable to sustainability as the collective contributions of those who make only small contributions. We argue that critical mass is established by developing a diverse set of community members with heterogeneous interests and resources, and not purely by accumulating content.

Introduction

Online communities often struggle to get off the ground and grow large enough to sustain activity and achieve their goals. This problem is frequently characterized as a problem of *critical mass*: the community needs sufficient *something* to be successful. What exactly is critical mass in terms of an online community? Or, put another way, an online community needs a critical mass of *what*?

This is a difficult question to answer because online communities have a number of different components. Online communities have people, they have content, they have activity, and they have participation. For an online community to achieve a critical mass, does it need to have grown to a certain number of distinct members? Or does it need a certain number of posts, comments, pictures, or other content regardless of how and from whom they got there? Furthermore, how does an online community know when it has achieved a critical mass?

Online communities are so numerous and diverse that they cannot easily be studied as a population. Therefore, most research on online communities takes a case study approach, where a single community is studied in depth. This research has led to countless important insights and advances in understanding of online communities. However, this approach has limits in the generalizability of results to online communities at large.

In this paper, we conduct an exploratory analysis of over 1000 communities and analyze how they grow in order to better understand critical mass and how it develops. We do this by analyzing WikiProjects, a large number of online communities within Wikipedia.

Background

Theory

Oliver and Marwell (Oliver, Marwell, and Teixeira 1985; Oliver and Marwell 1988; Marwell, Oliver, and Prahl 1988) have applied the concept of critical mass to "collective action." They tried to use the concept to understand group-level behaviors. When does a collection of individuals become a group and start acting like a group, with group goals? To explain this, they use a metaphor of critical mass, and developed a theory of how critical mass forms in social groups.

This Critical Mass Theory identifies a production function for each group, which is the *relationship* between individual use and the group-level goals. Production functions are characterized by their shape. Critical Mass Theory (Oliver, Marwell, and Teixeira 1985) describe three primary shapes: accelerating, linear, and decelerating production functions. In an accelerating function, each subsequent contribution to the group has a greater value or greater return; for decelerating functions the opposite is true. It is not clear what production functions look like for social media systems; do early contributors have to do the hard work of setting social norms, and later contributors have it easy (decelerating production function), or do early contributors see few benefits because there is no one to read their contributions, and later contributors have the most impact because of their broad readership (accelerating production function)?

The production function determines the relationship between individual behavior and group outcomes. For online communities, it is certainly not clear if there is a group-level outcome. A "community" — a coherent feeling of belonging to the group —isn't always one of the desired outcomes. Many people contribute to Wikipedia or Facebook without feeling the need or desire to become part of the "Wikipedia community." And on Facebook, there isn't necessarily a common group output that everyone is working towards. As a result, it is not clear if the critical mass theory of Oliver and Marwell applies to these situations, even though many researchers studying social media have invoked the concept.

Markus (1987) points out that critical mass theory is *sequential*. It assumes that each person is influenced by the people who came before them. However, in reality influence runs both directions. Individuals can also be influenced by the people who will come after them through the setting of expectations. She suggests that people are frequently aware of production functions and that knowledge can be used to make participation decisions.

Critical Mass Theory proposes that one factor influencing the shape of a production function for a group is the heterogeneity of that group in terms of access to resources or information. It argues that groups with heterogeneous access to resources or information are more likely to have accelerating production functions. For example, a group in which one person has sandwiches, another has salad, another has a frisbee, and someone else has drinks is more likely to produce a picnic than a group where four people have frisbees. Diversity allows for contributions to complement each other and increase the value of the collective good, as opposed to contributions replacing each other and generating no new value for the group as a whole.

We believe that a community's production function can be estimated by examining growth over time because of its sequential nature. By exploring the growth patterns of a large number of online communities, we hope to gain insight into the nature of the production functions for online communities and inform communities regarding how to develop critical mass.

Critical Mass in Online Communities

Community Growth and Turnover. Much existing research on online communities has looked at how people join and leave communities and their motivations for doing so. Nov (2007) has identified several motivations offered by Wikipedians for their participation, such as expressing values, having fun, and to socialize. Large communities may be attractive to potential new contributors because there is more assurance that there will be people with whom to communicate or who will find value in one's contributions. Butler (2001) shows that more people join communities which have a larger volume of communication and more existing members than smaller communities. We can infer from this that community sustainability is a matter of growing large enough so that people develop positive expectations about their participation.

But people who join communities do not always stay, and people leaving communities may also effect growth. Butler also shows that more people tend to leave large communities even as more people are joining them. Velasquez, Wash, Lampe and Bjornrud (2013) found that people leave online communities because of policy changes to the site, personal life changes alter priorities, or because they can no longer fulfill their perceived role in the community. Socialization of new members into a community can also affect whether people remain in a community after an initial trial. Burke, Marlow and Rento (2009) found that when newcomers to a social media site receive a response from the community to their initial post, they are more likely to remain on the site and make future contributions.

It is clear that online communities experience significant turnover in their membership. Some experiments (Farzan et al. 2011; Dabbish et al. 2012) have shown that people are more likely to join and participate in a community where there is high turnover. This suggests that a community that not only accumulates a lot of new members, but also has members leaving as well, may actually be more sustainable in the long run than communities whose members never leave once they have joined. This counterintuitive finding is attributed to enhanced social presence on sites with high turnover. High turnover on a site offers the impression that those who are on a site are active and engaging, which increases the perception of the site as a social entity.

Distribution of Participation. Other research has taken different perspectives on critical mass. Peddibhotla and Subrahami (2007) argue that the majority of contributions are made by a minority of individuals, and that it is this small minority of contributors who constitute a community's "critical mass". From this perspective, the amount of contributions made in total to the community is what is important, and not necessarily who makes them or the number of different people making them. Raban and Rabin (2009) show that most internet data, particularly from online communities, are power-law distributed in that a large majority of content or contributions are made by a small minority of users. In fact, they show that without performing statistical transformations of data from Usenet communities, there may not be a meaningful relationship on average between the number of people participating and the total amount of participation.

The nature of this distribution has an important consequence for understanding critical mass. Do new members of a community really influence the productivity of that community? If new members do not show up, will existing "power users" make up the slack and make the necessary contributions to keep the community active and maintain its value? If not, it suggests that critical mass is not really about growing in terms of overall community membership, but in finding any way to accumulate valuable content that provides some value to the community. The notion that communities with high turnover increases participation supports the notion that it is the content on the site, and not necessarily the people themselves, that is more important.

Critical Mass Theory. Raban, Moldovan, and Jones (2010) have explicitly tested Critical Mass Theory on IRC chat channels by looking at community sustainability over time. They conduct survival analysis on the lifespan of a channel based on a number of factors related to interactions in the channel's earliest stages. They find that, unsurprisingly, channels survive longer when more messages are posted to the channel in its early stages. They also evaluate the production function of these communities by plotting the number of posters against the number of posts and define an accelerating production function as one in which the number of posts per poster increases over the life of the community. They found that increasing production functions lead to longer survival for communities. They also argue that group heterogeneity leads to longer community survival.

Raban et al. use some measures in their data that may not correspond well to concepts in Critical Mass Theory as proposed by Oliver et al. In particular, they define an accelerating production function as one in which the ratio of posts to posters grows over time, meaning that individuals are increasing their production. This measurement discounts the contributions of newcomers, but there is no evidence that newcomer contributions are less valuable. Raban et al. also define heterogeneity in terms of the number of distinct posters in a community. This assumes that all posters offer unique new resources to the community, which may not at all be true.

These issues highlight the difficulty of using Critical Mass Theory to characterize online community growth. The assumptions made by Raban et al. in their measurements are not unreasonable. However, their findings indicate that both accumulation of posts, and accumulation of new posters, lead to longer community survival. But since posts cannot be made without posters and vice versa, we cannot distinguish whether it matters to the sustainability of the community *who* is making contributions. This distinction is important for understanding critical mass. If critical mass is a function of content, a community may need to only find a small number of people with motivation to contribute at a high rate. And if it is a function of people, does it matter which people?

WikiProjects

WikiProjects are "groups of contributors who want to work together as a team to improve Wikipedia" (http://en.wikipedia.org/wiki/Wikipedia:WikiProject). This means that each WikiProject represents a distinct community of editors within Wikipedia. WikiProjects are typically organized around either a topic, such as Chinese History, or around some other task, such as monitoring copyright issues of Wikipedia content.

WikiProjects do not have explicit ownership of articles on Wikipedia, but rather they express an interest in maintaining information relevant to their topic across Wikipedia. For example, the Wikipedia article about Yankees Stadium might be monitored by several different WikiProjects, such as a New York Yankees project, a New York City project, a Major League Baseball project etc. Each of these projects might have an interest in some aspect of the Yankees Stadium article, and each project would work independently to ensure that the article meets their goals.

A WikiProject can be created by submitting a request to the WikiProjects Council. Once approved by the Council, a WikiProject is given its own namespace on Wikipedia. This namespace has the same components as regular Wikipedia articles. There are article pages, which frequently have some descriptive information about the project, and talk pages where editors discuss the project and its goals.

WikiProjects provide a useful dataset with which to conduct quantitative analysis of a large number of distinct online communities. All WikiProjects run on the same technology, with the same affordances, meaning that technological differences cannot account for any differences between communities. Likewise, WikiProjects are overseen by the WikiProjects council and by the larger Wikipedia governance structure, which means that all WikiProjects are subject to the same rules and policies. The large number and immense diversity of WikiProjects allows us to generalize the findings from our analyses to many types of online communities.

Likely for these reasons, WikiProjects have become a common subject of research on online communities. Choi, Alexander, Kraut, and Levine (2010) have studied socialization of new community members in WikiProjects, examining various tactics used to socialize newcomers and evaluating their effectiveness. Other research (Forte et al. 2012) shows how WikiProjects help Wikipedians coordinate their work on the encyclopedia efficiently, and that the opportunity to communicate and socialize with others was one motivating factor for participating in a WikiProject. Zhu, Kraut, and Kittur (2012) have looked at how WikiProjects formally and informally set goals for their community and how forms of goal setting influence community productivity.

Measuring WikiProjects. One issue with studying WikiProjects is that it can be difficult to consistently measure some important concepts such as community membership and community productivity. Some WikiProjects research (Chen, Ren, and Riedl 2010) has measured community membership by using the stated list of members present on the home pages of many WikiProjects. Many WikiProjects invite editors to officially declare their membership by adding their names to this list. However, this is not a universal convention in WikiProjects and so using it creates an additional selection criteria for any sample that is studied, and it is not clear whether this criteria would create a biased sample. Also, other research (Morgan et al. 2013) has demonstrated that many people who make contributions to a WikiProject's project page do not explicitly state their membership on the page even in communities that do use this convention. Furthermore, Morgan et al. found that there was very little qualitative difference in the contributions of those who officially declare membership and those who did not.

Chen et al. (2010) define production of a WikiProject by measuring the revisions made by the project's members to articles within the project's scope, as determined by Wikipedia's category structures. This measure may ignore contributions made within a WikiProject page that do not directly relate to a revision made in an article. For example, an editor who asks a clarifying question or makes a communication to set community goals may not directly relate to an edit in the Encyclopedia, but such revisions do make a contribution to the project.

	Year 1	Year 3	Year 5
Revisions	103 (119)	681 (498)	1405 (1005)
Editors	36 (27)	179 (116)	266 (167)

Table 1: Means (Standard Deviations) for WikiProjects growth

	Editors		
	Decelerating	Linear	Accelerating
Decelerating	197	8	0
Linear	66	23	8
Accelerating	400	144	223

Table 2: Editor Growth Curve Shape by Revision Growth Curve Shape (Number of Projects)

WikiProject Growth

For our study of critical mass in WikiProjects, we build on the existing WikiProjects literature by collecting the revision histories of all WikiProjects listed in the WikiProjects directory as of March 6, 2013. For each project we collected all revisions from the first five years of that project's existence. After removing projects less than five years old, we were left with 1,069 projects for analysis. These revision histories are limited to revisions made within the project namespace, including revisions made in the "Talk" pages of the namespace.

For each project, we took two measurements of the project's growth at weekly time intervals starting from the project's inception. We measured the number of revisions made to the project namespace in the given week, and the number of editors who made their first revision in the project during the given week.

These measures of community growth correlate to two distinct aspects of communities that are relevant for understanding critical mass. Revisions made to a project page represent community productivity. We argue that measuring revisions made within project pages is a better measure of community production than revisions made to articles on the encyclopedia. This is because no objective link can be made between an edit on an article and value to a WikiProject, meaning that any attempt to do so (such as Chen et al.) is surely imprecise. Most work done on articles for a project, however, is reported in some form or another to the WikiProject. Therefore, revisions made on the project namespace are a reflection of the work done in the encyclopedia. This measure is also imprecise, but it does not suffer from ignoring the valuable contributions made exclusively within the project pages, and we therefore feel it is the best objective measure of productivity for WikiProjects.

Likewise, we consider anyone who has made at least one revision to the project page namespace to be a member of the community. This definition is consistent with the findings of Morgan et al. that those who explicitly declare membership are overall not different from those who do not declare membership.

WikiProjects have an average growth rate of 5.70 revisions per week. They also add 1.08 new editors per week on average. Table 1 gives the average number of revisions and editors across all WikiProjects at years 1, 3, and 5.

We see that there is considerable variance in the sizes of projects after 5 years of existence. WikiProject editors on average make 5.29 revisions over 5 years to a given project. 55% of editors in a project only make 1 revision.

Modeling Growth We modeled the growth curve of each community over the first five years of its existence as an estimate of its production function. Critical Mass Theory suggests that these communities are likely to have either an accelerating, linear, or decelerating production function.

For each week from the inception of the project, we calculated the percentage of the project's five year total that had been accumulated up to that point in time. We then fit a second-order polynomial for each project that estimates the percentage of the five year total at each week for the five years. In these models, $Y = \beta_0 + \beta_1 * weeks + \beta_2 * weeks^2$, the coefficient β_2 represents the degree of acceleration (positive or negative) for the project.

We found that the polynomial models fit the data very well. The median R^2 of these models was .98. This very high degree of fit suggests that the patterns of acceleration or deceleration described by Critical Mass theory are an accurate way to describe growth in WikiProjects.

We found that WikiProjects showed much variation in the shape of their growth curves for revisions. Figure 2 illustrates the distribution of the second-order coefficients from the models. Most projects have a positive coefficient for growth of revisions to the project. This means that the majority of WikiProjects show accelerating growth in the number of revisions contributed. However, WikiProjects showed a tendency to have decelerating growth in terms of the accumulation of editors.

We explored this further by explicitly classifying the shape of each growth curve into accelerating, linear, and decelerating categories. We classified projects whose second-order coefficients were not statistically significant ($p < .05$) as having linear growth. Projects with a positive coefficient are classified as having accelerating growth, and those with negative coefficients as having decelerating growth.

Table 2 is a cross-tabulation of the classification of projects' growth curves for revisions and editors. 37% of all projects have accelerating growth in revisions and decelerating growth in editors, whereas only 20% of projects had accelerating growth in both editors and revisions. No project showed decelerating growth in revisions but accelerating growth in editors.

Critical Mass of What?

Content vs. People

These results have implications for understanding critical mass in online communities. Projects are more likely to slow down in accumulating new members of the community, yet increase their rate of production of content over the same time. This suggests that the acceleration of production of revisions is due to existing community members increasing

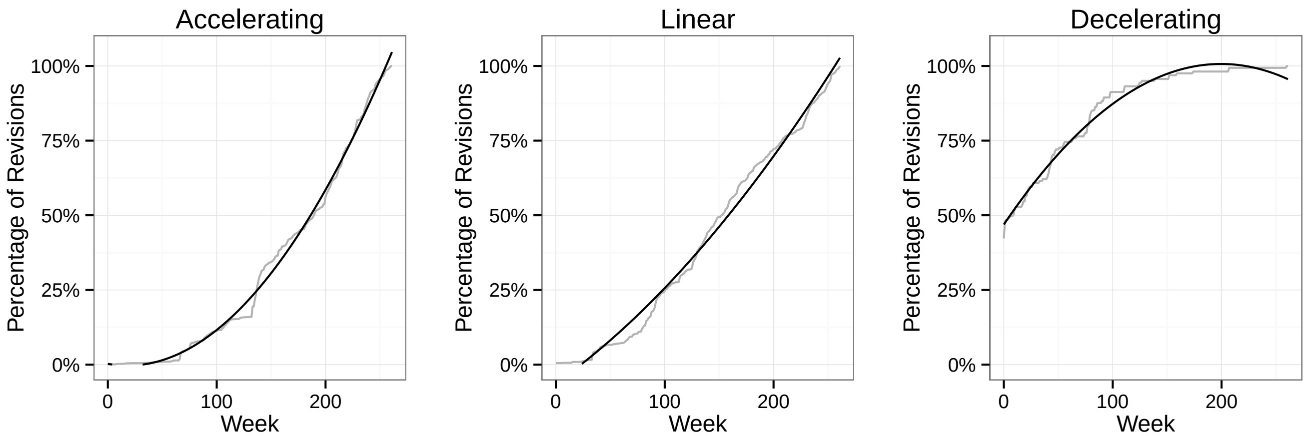

Figure 1: Examples of Accelerating, Linear, and Decelerating projects

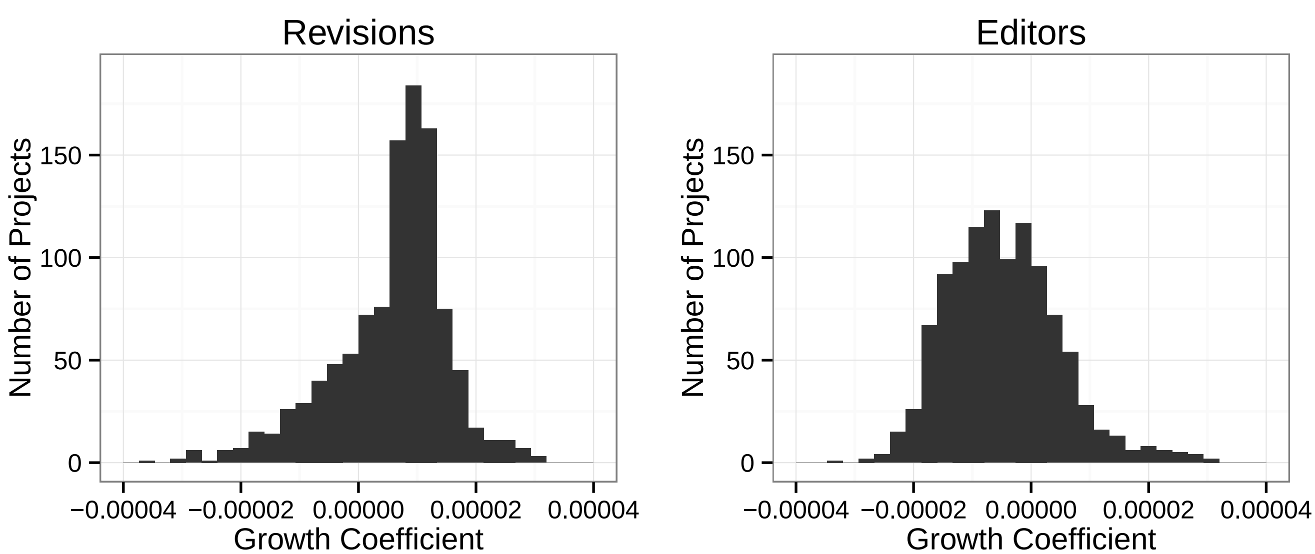

Figure 2: Distributions of second-order coefficients

	Revisions after 5 years	
	OLS	*negative binomial*
Intercept	670.66***	6.57***
	(55.68)	(0.05)
Revisions after 1 year	0.68	0.0002
	(0.49)	(0.0004)
Editors after 1 year	22.40***	0.020***
	(1.68)	(0.001)
Revisions:Editors	−0.016***	−0.00001***
	(0.003)	(0.00000)
Observations	1,069	1,069
R^2	0.219	
AIC		17,450

*p<0.1; **p<0.05; ***p<0.01

Table 3: Effect of one year accumulation of editors and revisions on 5 years accumulation of revisions

their individual production more than increasing production by adding new contributors.

This discrepancy suggests an important question for online communities. What exactly is this critical mass that will help sustain online communities? Do communities need activity or content, do they need people, or some combination in order to become self-sustaining and have an accelerating production function?

We explored this question further by fitting a model to the WikiProjects data that compares the effects of growth of content (in the form of revisions) and the growth of community membership. In this model, we estimated the number of revisions that a project has accumulated after five years using two independent variables: the number of revisions accumulated after one year and the number of editors accumulated after one year. We also included an interaction term between these variables. Table 3 describes this model. This model suffers from some violations of OLS regression assumptions because the of the nature of the distribution of the dependent variable. To correct this, we fit a second model using negative binomial regression. Negative binomial regression can be used to model count data such as these where the dependent variable cannot be less than 0 and the data approximates a negative binomial distribution. The distribution of revisions after 5 years in our data are overdispersed, meaning that the variance of the distribution is greater than the mean. Negative binomial regression provides reliable estimates of standard errors and statistical significance when data are overdispersed, meaning that the direction (positive or negative) of the modeled effects on revision accumulation from this model are reliable. The effect sizes of this model are more difficult to interpret than OLS however, and therefore we have included both models here.

These models show large and statistically significant effects of the number of editors at year one on the number of

	Value	Std. Error	t value
Revisions	-0.01	0.00	-7.61
Editors	0.03	0.00	6.48
Decelerating|Linear	-1.19	0.13	-9.09
Linear|Accelerating	-0.65	0.13	-5.13

Table 4: Ordered Logit Model. DV is the shape of the revision growth curve (Decelerating, Linear, or Accelerating

revisions at year 5, holding the number of revisions at year 1 constant. However, there is little effect of the number of revisions at year 1 on the number of revisions at year 5, holding editors constant.

According to model 2, projects that have 100 revisions after one year from 25 different editors would be expected to have 992 revisions after 5 years. Conversely, projects that have 100 revisions from 50 different editors after one year would be expected to have 1520 revisions after 5 years.

We can illustrate this model with one example taken from the data. The WikiProject for Astronomical Objects received 111 revisions after 1 year from 17 editors. The WikiProject for Astrology also received 111 revisions, but from 54 different editors. Both projects had accelerating growth curves over 5 years, but the Astrology project had 2242 revisions after 5 years, compared to 592 for the Astronomical Objects project.

We also tested the effect of accumulating editors on the probability of having an accelerating growth curve. We fit an ordered logit model that estimates the odds of a project being classified as accelerating versus either linear or decelerating with the number of editors and the number of revisions after one year. Table 4 describes this model. Based on this model, each additional editor accumulated after one year, when holding the number of revisions constant, increases the odds of having accelerating growth by approximately 3%. Consequently, getting 20 more editors for a project nearly doubles the odds of an accelerating growth curve. Also in this model, receiving more revisions *reduces* the odds of project acceleration when holding the number of editors constant.

This suggests that the critical mass that allows projects to grow and thrive is more likely a critical mass of people, rather than a critical mass of content or contributions. Getting smaller contributions from more different people leads to larger long term community production than getting many early contributions from a smaller number of people.

Heterogeneity

Critical Mass Theory proposes that groups of people with greater heterogeneity of resources and information are more likely to produce an accelerating collective good, as it is more likely that someone among the interested parties will have a resource or information that is of greater value to the collective than the cost of contributing it. We examine this question in WikiProjects by looking at the participation of editors across multiple projects. In our data, the sum of all editors who are distinct within a given project is 377,146. However, the number of absolutely distinct editors (i.e. the

Table 5: Effect of editor heterogeneity and Power User Influence on 5 year growth

	Revisions after 5 years	
Constant	5.915***	7.046***
	(0.110)	(0.036)
Heterogeneity	0.001***	
	(0.0001)	
Editors after 1 year	0.010***	
	(0.001)	
Power User Revisions after 1 year		-0.001**
		(0.0003)
Non-Power User Revisions after 1 year		0.005***
		(0.0005)
Observations	1,069	940
AIC	17,439	15,509

*p<0.1; **p<0.05; ***p<0.01

number of distinct user names or ip addresses) within the entire sample is 58,705. This means that on average, WikiProjects editors participate in 6.4 different WikiProjects. We use participation in multiple projects as a proxy for the breadth of interests and resources that a given WikiProject member brings to any of his or her given projects. We measure the heterogeneity of a WikiProject as the total number of *other WikiProjects* that members of the given project have participated in. Since we are particularly interested in how this heterogeneity affects growth, we calculate the heterogeneity of editors within the first year of each project's existence. We only consider projects distinctly, meaning that if multiple editors of a project all contribute to a single other project, that project is only counted once.

We fit a negative binomial regression model estimating the number of revisions after five years with the heterogeneity of the project's editors in its first year. We included the number of editors in the first year as a covariate to further isolate the effect of heterogeneity, since projects with more editors are automatically more likely to have greater heterogeneity because each member has some probability of participating in another project. This model therefore estimates the effect of having greater access to different resources or information within a project not related to simply having more people.

This model is described in Table 5. The model shows a statistically significant positive effect of heterogeneity on the number of revisions accumulated after five years, even when holding the total number of editors who provide heterogeneity constant. We conclude that projects whose members participate in a greater variety of other WikiProjects grow faster as a result of this greater heterogeneity. These data support Critical Mass Theory, and suggest that obtaining a critical mass of people is most effective when a community develops a set of diverse people with greater reach to various information or other resources.

Power Users vs. Non-Power Users

This finding suggests that even though accumulating new community participants is the most likely way to develop a critical mass, that not all participants in a community make an equal contribution towards developing that mass. One of the characteristics of most online communities is that participation is highly unbalanced, with a few people providing the vast majority of contributions and many people individually providing only small contributions. These distributions tend to approximate a power law distribution (Raban and Rabin 2009). Because of the distinct imbalance, it is clear that there are two identifiable classes of participants in a WikiProject. One class sits at the head of the distribution where there are many people who have each made only a small number of contributions. The other class is the tail of the distribution, and consists of the few people who have participated most frequently in the project.

If critical mass is a critical mass of people, which people are actually more important? Does critical mass mean obtaining a small number of people who are willing to produce a lot for the community? Or is it developing a large number of people whose contributions and participation supplement the small number of "power users" to make the community sustainable? We analyze these questions by classifying users according to their location in the distribution of participation, and explore the relative importance of the contributions made by each class of editor.

To classify editors, for each project we fit a power law distribution to the data from the first year. Of the 1,069 projects, a power law distribution was a satisfactory fit for 940 projects according to a Kolmogorov Smirnov test. We limited further analysis to these 940 projects. From these fits, we divided the distribution at the "elbow" to separate the head from the tail of the distribution, with the distribution's x-value representing the number of revisions made by an editor and the y-value representing the estimated frequency that an editor would make x revisions. To determine the precise point of the elbow, we first identified a pivot point on a fit's graph with the minimum value of x (always 1) and the minimum value of y within the range of observed data. For example, if the most active editor made 100 revisions, the fitted power law estimate for 100 revisions was used as the minimum value of y and the y-value in the pivot point. We then found the minimum Euclidean distance from the pivot point to the fitted curve, and used the x-value of the point on the curve as the dividing point between the two parts of the distribution. Figure 3 provides a graphical representation of this method. This dividing elbow point for each project represents the number of revisions for a given editor that would put the editor into the category of a "power user." All editors in a project who made more revisions than this cutoff point were classified as power users. Figure describes the distribution of these cutoff points, which had a median value of 7 revisions.

We estimated the relative influence of power users to non-power users by again fitting a negative binomial regression model to the five-year revision totals. This model, described

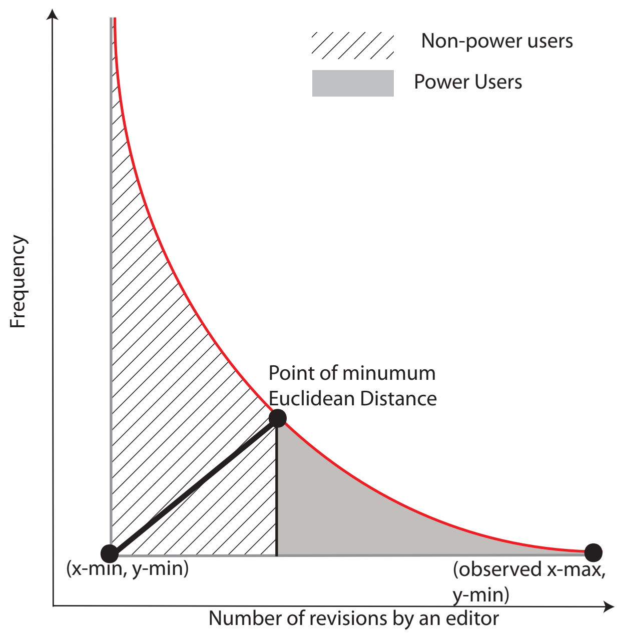

Figure 3: Separating power users from the distribution

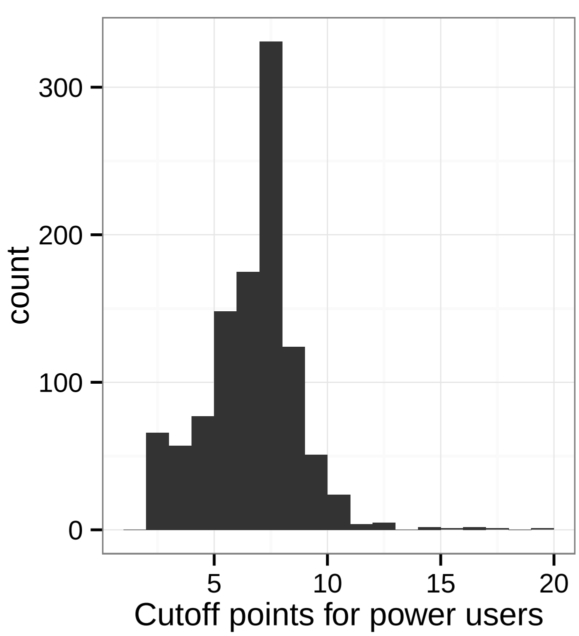

Figure 4: Distribution of power user cutoffs

in Table 5 uses the number of revisions made by power users and the number made by non-power users in the first year to estimate five year growth.

This model shows a statistically significant effect of the number of revisions made by both types of users. Of particular interest in this model is the direction of each effect. Revisions in the first year made by non-power users have a positive effect on five year growth, whereas revisions made by power-users have a negative effect. In our analysis above we showed that when controlling for the total number of editors, the total number of revisions in the first year had little effect. This model provides additional insight, showing that revisions made by those who infrequently participate do in fact help a project grow, but this can be offset by over-participation from a project's power users.

Discussion

Our analysis suggests that in the long term a community will be more productive by growing its membership in its early stages rather than increasing the contributions of existing members. It is better for a community to receive small contributions from many people than for it to receive large contributions from a few. The critical mass is a mass of community members. Once this mass is attained, contributions to the community become more valuable and therefore production accelerates even if growth of membership decelerates.

Solomon and Wash (2012) found that when wiki users were presented with a blank page, rather than a page that had been seeded with content, the users made more of a contribution to the wiki. This suggests that a sense of opportunity and need can drive contribution to a community. For WikiProjects, when users with an interest in a project see a small number of people rapidly making contributions early on, it may trigger a free-riding effect since it might appear that the goals of the project will be accomplished given the current rate of production. Conversely, when users see more sparse contributions, it leaves a more apparent opportunity to contribute. And once a large number of people have made some initial investment, this critical mass becomes aware of the potential for significant production and increases contributions.

This notion is further supported by the finding that the more participation by a project's power users in its early stages, the less that project will ultimately grow, whereas small amounts of participation by many other people lead to better long-term growth. However, some caution should be taken in interpreting this result as well. It is possible that power users of failing projects increase their participation in an effort to keep it afloat or attract new participation, and that such effort is simply futile. Both interpretations however support the conclusion that communities become more sustainable when they grow by adding new people, and particularly people with diversity in resources or interests. Power users who hope to see a community continue to grow may see their efforts best rewarded through recruiting others rather than contributing new content to a community.

Recent research (Halfaker et al. 2013) has shown that Wikipedia has become calcified and perhaps even hostile to new editors, and consequently the number of new editors on Wikipedia has experienced a rapid decline for the last several years. We examined our data from WikiProjects to account for this, and we found that there is a tendency for more recent projects to have more decelerating production functions. Halfaker et al. suggest that this decline began around 2007. Since our data only include projects with at least five years of history, and were collected in 2013, the vast majority of our sample comes from before this decline began. However, our findings offer additional support that rejection of new participants is a likely cause for a decline in production, since we show that new contributors who make small contributions are actually more important for sustainability.

Panciera, Halfaker, and Terveen (2009) argue that power users are the lifeblood of Wikipedia, and show that on Wikipedia articles these power users produce higher quality edits and maintain consistent production over time. Our data argues that at least in the case of WikiProjects, these power users are insufficient to sustain a community and therefore do not constitute a critical mass. Rather, the larger mass of infrequent editors collectively produce a value that makes the community sustainable. There are a number of possible reasons for this, and we can only speculate at this point. The presence of a large number of shallow participants may be a proxy of the collective value of a project. More people willing to make at least some small contribution may mean the project is inherently more useful and therefore worthwhile to continue. These users may also set a higher expectation of a community, such that people interested in contributing may feel more assured that someone will interact or respond to them, or that there is some audience for their contribution. This is an important topic for future research.

These results can help those trying to start communities. Adding content or taking measures to make a site or community appear more active is not likely to work in the long run. Rather, technologies and policies should be crafted to encourage new community members and encourage smaller levels of participation from more different people. This approach is more likely to be successful for building critical mass and creating a sustainable and productive community.

References

Burke, M.; Marlow, C.; and Lento, T. 2009. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 945–954. ACM.

Butler, B. S. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research* 12(4):346–362.

Chen, J.; Ren, Y.; and Riedl, J. 2010. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 821–830. ACM.

Choi, B.; Alexander, K.; Kraut, R. E.; and Levine, J. M. 2010. Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 107–116. ACM.

Dabbish, L.; Farzan, R.; Kraut, R.; and Postmes, T. 2012. Fresh faces in the crowd: turnover, identity, and commitment in online groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 245–248. ACM.

Farzan, R.; Dabbish, L. A.; Kraut, R. E.; and Postmes, T. 2011. Increasing commitment to online communities by designing for social presence. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 321–330. ACM.

Forte, A.; Kittur, N.; Larco, V.; Zhu, H.; Bruckman, A.; and Kraut, R. E. 2012. Coordination and beyond: social functions of groups in open content production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 417–426. ACM.

Halfaker, A.; Geiger, R. S.; Morgan, J. T.; and Riedl, J. 2013. The rise and decline of an open collaboration system how wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist* 57(5):664–688.

http://en.wikipedia.org/wiki/Wikipedia:WikiProject. 2013.

Markus, M. L. 1987. Toward a "critical mass" theory of interactive media universal access, interdependence and diffusion. *Communication research* 14(5):491–511.

Marwell, G.; Oliver, P. E.; and Prahl, R. 1988. Social networks and collective action: A theory of the critical mass. iii. *American Journal of Sociology* 502–534.

Morgan, J. T.; Gilbert, M.; McDonald, D. W.; and Zachry, M. 2013. Project talk: coordination work and group membership in wikiprojects. In *Proceedings of the 9th International Symposium on Open Collaboration*, 3. ACM.

Nov, O. 2007. What motivates wikipedians? *Communications of the ACM* 50(11):60–64.

Oliver, P. E., and Marwell, G. 1988. The paradox of group size in collective action: A theory of the critical mass. ii. *American Sociological Review* 1–8.

Oliver, P.; Marwell, G.; and Teixeira, R. 1985. A theory of the critical mass. i. interdependence, group heterogeneity, and the production of collective action. *American journal of Sociology* 522–556.

Panciera, K.; Halfaker, A.; and Terveen, L. 2009. Wikipedians are born, not made: a study of power editors on wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*, 51–60. ACM.

Peddibhotla, N. B., and Subramani, M. R. 2007. Contributing to public document repositories: A critical mass theory perspective. *Organization Studies* 28(3):327–346.

Raban, D. R., and Rabin, E. 2009. Statistical inference from power law distributed web-based social interactions. *Internet Research* 19(3):266–278.

Raban, D. R.; Moldovan, M.; and Jones, Q. 2010. An empirical study of critical mass and online community survival. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 71–80. ACM.

Solomon, J., and Wash, R. 2012. Bootstrapping wikis. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, 261. New York, New York, USA: ACM Press.

Velasquez, A.; Wash, R.; Lampe, C.; and Bjornrud, T. 2013. Latent users in an online user-generated content community. *Computer Supported Cooperative Work (CSCW)* 1–30.

Zhu, H.; Kraut, R.; and Kittur, A. 2012. Organizing without formal organization: group identification, goal setting and social modeling in directing online production. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 935–944. ACM.