

WHICH CAME FIRST? CONTRIBUTION DYNAMICS IN ONLINE PRODUCTION COMMUNITIES

Completed Research Paper

Gerald C. Kane

Boston College
Chestnut Hill, MA, USA
gerald.kane@bc.edu

Sam Ransbotham

Boston College
Chestnut Hill, MA, USA
sam.ransbotham@bc.edu

Abstract

While considerable research investigates collaboration in online production communities, particularly how and why people join these communities, little research considers the dynamics of the collaborative behavior. This paper explores one such dynamic, the relationship between viewing and contributing. Building on established theories of community involvement, this paper argues that a recursive relationship exists, resulting in a mutually reinforcing cycle where more contributors lead to more viewers and, in turn, more viewers lead to more contributors. We also analyze the effect of time and anonymity within this dynamic relationship. This paper offers guidance for research into online production communities that builds on the large behavioral data these communities generate.

Keywords: knowledge management, user-generated content, wikis

Introduction

Recent years have witnessed considerable interest in online communities, as they are increasingly becoming a valuable source of information and knowledge for companies (Bateman et al. 2011; Faraj and Johnson 2011; Forman et al. 2008). These communities may develop as independent virtual organizations that exist for the purposes of developing free or open source content, such as open source software communities (Grewal et al. 2006; Oh and Jeon 2007; von Krogh and von Hippel 2006) or Wikipedia (Ransbotham and Kane 2011; Zhang and Zhu 2011). Other online production communities are intentionally sponsored by organizations to allow customers to collaborate and generate valuable insight for companies, such as Dell's Ideastorm (DiGangi et al. 2010) or MyStarbucksIdea.com (Gallaughar and Ransbotham 2010). Still others are sponsored within large organizations, such as IBM Community Groups (Alavi et al. 2005) and knowledge sharing communities in consulting firms (Wasko and Faraj 2005), that allow employees to connect and collaborate with one another to share knowledge.

Considerable research has investigated how and why online communities produce valuable knowledge (e.g., Bateman et al. 2011; Faraj and Johnson 2011; Forman et al. 2008; Garg et al. 2011; Leimeister et al. 2005; Ma and Agarwal 2007), but researchers have recently lamented the failure of this substantial previous literature to investigate the dynamic behavior of online communities (Faraj et al. 2011). For instance, people often have complex and conflicting motivations for participating in online that may interact with one another in unexpected ways (Roberts et al. 2006). People not only join online communities but they also leave, requiring the community to compensate for this membership turnover in their development process (Ransbotham and Kane 2011). Communities that focus on creating information artifacts, such as open source software communities and knowledge creation communities like Wikipedia, often experience qualitatively different stages of interaction as the production process evolves (Kane et al. 2009b; O'Mahony and Ferraro 2007). Most of the extant research on online communities tends to investigate a cross-sectional snapshot of these communities at a single point in time or aggregate data in ways that obfuscates the interesting dynamics occurring over time in the community.

As companies increasingly develop and use online communities for creating knowledge (Faraj et al. 2011), understanding the dynamics of online communities can provide valuable insight for how to manage these new communities in the future. The data provided by online communities facilitates the ability to study these dynamics in ways that were not possible even a few years ago. Most online communities preserve voluminous data about its members' contributions and activities, providing considerable opportunity to analyze the collaboration that occurs within online communities. As the cost of bandwidth and storage continues to decrease, communities collect increasing amounts of data and preserve this data for an extended time. Furthermore, not only have many online communities grown considerably in membership but have now also been in existence for many years. This longevity allows robust longitudinal analysis of online that would not have been possible before now — although online communities generate copious amounts of data, technology cannot speed up the time over which collaboration occurs.

This paper attempts to address this gap in the literature by examining contribution dynamics in online production communities. Extending previous theories of online community membership (Lave and Wenger 1991; Preece and Schneiderman 2009), we explore whether a mutually reinforcing relationship exists between the contributors to a production community and the audience for the content it creates. Such a dynamic is easy to conceptualize. The creation of content by a production community is likely to attract an audience for that content. That audience is likely to generate new contributors, who will then help create new content or revise existing content, further increasing its audience. Thus, a virtuous cycle may develop in which better content attracts more contributors, who then help to create better content. We further hypothesize that this recursive relationship attenuates over time as the content matures, such that more mature content will continue to attract increased viewers but that it will not continue to attract new contributors. Finally, we hypothesize that anonymity will facilitate the transition from viewer to contributor, particularly for people who may hold perspectives or knowledge that differs somewhat from existing community members that are particularly valuable for collaboration.

We test our hypotheses by analyzing Wikipedia's Medicine Wikiproject and examining how network characteristics affect the market value of user-generated content. Our results provide evidence of a virtuous cycle between viewers and contributors. Simultaneous equation models show that not only does

a larger audience lead to more contributors but also that more contributors lead to a larger audience. Furthermore, this reinforcing dynamic attenuates over time, such that the audience of a community's content gets larger over time but the ability to attract additional contributors decreases. Anonymity positively reinforces this cycle.

Theoretical Development

A key challenge facing online production communities is its ability to attract enough contributors to sustain production (Butler 2001; Ma and Agarwal 2007; Stewart and Gosain 2006). Some argue that this challenge will become exacerbated over time as online communities proliferate and the growth of Internet users slows, leading online communities to increasingly compete with one another for members' time and attention (Wang et al. 2013). Thus, understanding how people are attracted to an online community and become more integrally involved is important for developing vibrant online communities. This need to understand the process of membership attraction and retention is particularly salient for companies seeking to entice customer feedback, reviews, or new product ideas (Dellarocas et al. 2010; Kiron et al. 2012). It is also important for organizations seeking to adopt online communities for internal knowledge management initiatives, because employees have often been reluctant to voluntarily contribute their knowledge to these efforts (Griffith et al. 2003).

Questions of membership involvement are somewhat complicated in online communities because online communities often exhibit considerable participation inequality. A small portion of the users typically performs most of the work and the vast majority of the users make relatively modest contributions (Faraj et al. 2011; Marwell and Oliver 1993; Oliver et al. 1985). Some research has suggested that this unequal distribution of participation is actually essential for healthy collaboration in online groups (Kuk 2006). Participation inequality allows the community to benefit from the contributions of more people without creating communication overload that would cause people to leave or fail to contribute the community (Butler 2001). Participation inequality factors prominently in existing theories of community involvement. The theory of legitimate peripheral participation posits that people become active members through a gradual process of escalating commitment (Lave and Wenger 1991). Members typically begin as passive observers during which they observe the activity of the community and learn its rules and norms. While some have referred to these passive participants as nefarious "lurkers," others simply refer to them as "readers" (Preece and Schneiderman 2009). Only after these peripheral members have learned how to interact with the community they slowly begin to contribute and become active members.

Researchers, however, have often distinguished between two types of online communities – production-focused communities and discussion-focused communities. Discussion communities largely exist for the purpose of conversation between its members, such as online discussion boards, and people typically participate in order to receive some form of social capita (Butler 2001; Ma and Agarwal 2007). Production communities, in contrast, largely exist to produce some sort of information artifact, such as open source software (Lee and Cole 2003), product ideas (Jeppesen and Frederiksen 2006), or new information and knowledge (Fleming and Waguespack 2007). Theories of community participation have largely been formulated in support communities, and we argue that contribution activity may differ somewhat in production communities because of the creation of an information artifact by the community. Most simply, the information artifact focuses the attention of outsiders to a single product, rather than requiring them to wade through the multiple conversations that occur in a discussion community. This focused attention may lead people to contribute, because they know a greater audience will view their contribution. Prior research demonstrates that people are more willing to contribute to online production communities if there is a larger audience for that contribution (Zhang and Zhu 2011).

Information artifacts generated by a production community also serve as boundary objects that coordinate activity between contributors and the audience (Carlile 2002; Levina 2005). Boundary objects quickly summarize the collective knowledge currently possessed by the community, providing opportunities for readers to contribute if they notice a limitation in that knowledge. For example, readers might identify errors of commission and choose to fix them. For example, the users of an open source software product can identify and highlight bugs in the system, turning the user of the software into a contributor (Raymond 1999). Likewise, readers can also identify errors of omission. For example, readers of online reviews may be more motivated to add to information missing from other reviews

(Dellarocas et al. 2010). Boundary objects call attention to skills that may be missing from a community. For instance, research in wiki communities identify the unique role of “shaper,” people whose contribution is simply to improve the organization of existing knowledge rather than contribute substantive content to it (Majchrzak et al. 2013). Thus, because the boundary object facilitates transition between reader and contributor by allowing people to contribute without the need for long periods of “lurking” behavior, we hypothesize that the number of viewers of a production community’s information artifact will be positively related to the number of contributors it enjoys.

Hypothesis 1: The number of viewers of an information artifact created by an online production community will be positively related to the number people who contribute to it.

At the same time, the number of contributors to an online production community will also influence the development of that information artifact. Contributors represent valuable resources for online production communities, as they represent the time, energy, and knowledge that are the inputs of production. Online production communities with a larger number of contributors are likely to develop more valuable content. Using the Delphi technique (Jolson and Rossow 1971) and prediction markets (Foutz and Jank 2009), the value of aggregate information improves when more opinions are taken into account. In virtual teams, the number of ideas generated increases with team size (Martins et al. 2004). Research on the voice of the customer (Griffin and Hauser 1993) suggests that interviews with more customers increases the likelihood of identifying all customer needs. In knowledge sharing communities, people are more likely to find valuable information when they get more responses to a query (1996). Thus, a greater number of contributors are likely to generate more valuable information artifacts. Content that is more valuable is likely to attract more viewers. Individuals are more likely to seek out those are perceived to possess valuable information (Borgatti and Cross 2003; Perry-Smith and Shalley 2003). The marketing literature also connects viewership with perceived content value (Miller 2009; Ransbotham et al. 2012).

The impact of contributors on the boundary object, however, may not be universally positive. Considerable research suggests a curvilinear relationship between the number of contributors and the effectiveness of their work (Asvanund et al. 2004; Butler 2001; Hansen and Haas 2001; Oh and Jeon 2007; Ransbotham and Kane 2011). More contributors increase coordination cost; these costs grow exponentially as the number of possible interactions between contributors increases exponentially with the addition of each new member (Espinosa et al. 2007). Social media platforms often support thousands of users, often attempting to contribute to the community at the same time. It is not uncommon for communities to support thousands of contributors to the same production process at the same time (Kane 2011). Additional contributors increase coordination costs by making it more difficult for the community to organize their work, potentially limiting the marginal value of additional contributors after a certain point. Similar dynamics occur in software development teams that need sufficient resources to accomplish their goals, but adding more new members to a troubled or delayed project compounds delays by making coordination more difficult (Brooks 1975). People suffer from information overload when they try to make sense of and respond to others’ contributions, which reduces a group’s ability to organize information effectively (Hiltz and Turoff 1985). Thus, we expect a curvilinear relationship between the number of contributors and the value of an information artifact. We hypothesize:

Hypothesis 2: The number of contributors to an online production community has a curvilinear relationship with viewership, such that information artifacts with a moderate number of contributors attract the most viewers.

Online communities often move through distinct phases of collaboration, and later collaboration is distinct from that which occurred earlier (c.f., Gersick 1988; Tuckman 1965). Early collaboration is likely dedicated to brainstorming and generating new content, because no content exists; middle phases involve organizing disparate ideas from a generated critical mass; and later phases entail maintaining and defending the content amidst ongoing collaboration, after the group has developed the organized whole (Kane et al. 2009b). These characteristics suggest that this recursive dynamic between contributions and viewership should be stronger for newer than for older content. In earlier stages of development, the information artifact may be substantive enough to generate some viewership, but still inchoate enough that most of these viewers will recognize that they possess knowledge that can improve it. Furthermore, these early contributions are likely to yield the greatest improvement to the artifact, which will attract more viewers. In later stages of development, the artifact has become more developed and the opportunity to improve it will not be as obvious to new viewers, as it will contain fewer errors of omission and

commission in the information artifact. It may have already been “shaped” by other contributors (Majchrzak et al. 2013) and packaged for better consumption (Markus 2001). Thus, additional viewers may be less likely to become contributors, simply because there is less need for them to do so. If this recursive dynamic weakens over time, it suggests that age of the community is positively associated with viewership but negatively associated with contributors. This leads to the following hypotheses.

Hypothesis 3: The age of an online production community will be positively related to the number of viewers and negatively related to the number of contributors.

Furthermore, online production communities often allow people to contribute without revealing their identity; considerable research investigates the influence of anonymity on collaboration (Connolly et al. 1990; Siegel et al. 1986; Valacich et al. 1992). Much of this research shows that anonymity improves collaboration in online groups. Because people do not need to worry about being identified, they may share their ideas that diverge from existing knowledge or norms, which may contain unique and novel information not currently available to the production process. Anonymity may be a key factor associated with the dynamic interaction between viewership and contribution, and research has documented the importance of anonymity in online production communities (Kane 2011). As people read the information artifact created by the community, people may be more willing to share their knowledge or critique the contributions of others if comments are not connected with identity. If this knowledge proves valuable for the production process, it will also make the information artifact more valuable and attract additional viewers. Thus, we expect anonymity to positively reinforce both sides of the production process.

We also note that anonymity has been shown to be detrimental to collaboration in online groups in some cases. Anonymity lowers the level of social presence in a group – the degree to which people establish personal connections in a communication setting – often leading people to behave in disruptive ways that they otherwise would not (Short et al. 1976; Sia et al. 2002). Anonymity increases disruptive behavior and polarizing dynamics in collaborative environments (Jessup et al. 1990; Sia et al. 2002). Yet, we do not expect this polarizing dynamic affects the dynamics in our context, as we are exploring the dynamics associated with the earliest stages of community involvement – viewership and contribution. The polarizing dynamics observed in the other studies of anonymity occurred among the sustained work of the most committed members, such as the ongoing collaborators and community leaders (Preece and Schneiderman 2009). In short, anonymity may have detrimental effects in other aspects of online production communities, but not likely in the early contribution dynamics we study here.

Hypothesis 4: The number of anonymous contributors will be positively related to the number of viewers and contributors to an information artifact.

Research Method and Setting

We test our hypotheses by analyzing the relationship between viewership and the number of contributors to 16,068 Wikipedia articles in the Medicine Wikiproject (i.e., all articles in this project during the study period). To confirm the robustness of our findings, we use a holdout sample, in which parameter estimates from the first ten months of data predict viewing for the remaining nine months. As an additional test, we use parameter estimates from the Medicine Wikiproject to estimate viewership for the fashion and auto Wikiprojects, and compare predicted to actual viewership for these additional samples.

Research Setting

Drawn from the Hawaiian word meaning “quick,” a wiki is a Web site that anyone can edit. Wikipedia, established in 2001, uses a wiki platform to host an open-source encyclopedia. Users of the English version of Wikipedia have generated more than 4 million separate articles, and an additional 20 million articles are available in the 270 other languages in which Wikipedia is published. The hundreds of Wikiprojects on Wikipedia are dedicated to a wide range of topics, from the mainstream to the obscure. Considerable research has investigated collaboration on Wikipedia (Denning et al. 2005; Kittur and Kraut 2008; Kriplean et al. 2008; Ransbotham and Kane 2011). Focusing on all articles of a particular type allows us to account for competitive dynamics between communities, since they are likely to compete for the time and attention of a limited group of contributors. Moreover, studying articles dedicated to a

particular WikiProject limits potentially confounding factors. Because of their common subject matter these articles are more likely to share contributors such that we obtain a relatively smaller, clearly defined, cluster of articles and contributors than we would with a wider, unconstrained, sample.

We focus on medical information, which represents an early and prominent use of online sources (Ferguson and Frydman 2004). A recent Pew study reveals that Internet users increasingly turn to user-generated medical information online, and nearly 60% of Internet users have relied on Wikipedia as a source of health information (Fox and Jones 2009). The healthcare industry also draws on user-generated content to promote lifestyle changes, encourage collaboration among physicians, develop collaborative patient support networks, and provide a valuable resources to patients and providers (Kane et al. 2009a). Previous studies affirm the quality of medical information on Wikipedia (Clauson et al. 2008; Devgan et al. 2007; Laurent and Vickers 2009), which also has considerable economic value. Healthcare in the United States is a \$2.3 trillion industry, and by 2011, online pharmaceutical advertising expenditures are expected to reach \$2.2 billion, or 5% of Internet advertising (Phillips 2007). Wikipedia does not accept formal advertising, but other online providers of medical content (e.g., WebMD, HealthCentral) do, and these sites increasingly use user-generated content, such as blog communities and user forums.

Data Collection

We downloaded the full text history of 2,677,397 revisions of 16,298 articles by 40,0479 unique contributors in the Medicine WikiProject as of August 2011, which resulted in a 105 GB data set of raw data. We employed a 114-node Linux cluster to allow for simultaneous downloads and processing of these extensive data. For each contribution, we record the contributor's identity, the changes made, a description of the change, and the time of the change. Wikipedia offers an API for exporting XML versions of article revisions. The data elements mentioned (contributor identity, change made, change description, and time of the change) were all distinct XML elements.

To ensure that our analysis was based on the behavior of people, rather than computers, we excluded edits made by automated software programs (i.e., bots). Wikipedia's site policy requires that all bots be approved and registered; we obtained a list of active and previously active bots from Wikipedia. Bots on this list made 1.6% of the changes in our sample (43,309 revisions) and we excluded their edits from the analysis. A manual check of 75 random articles similarly revealed that 2.13% of the edits were bot activity. We also manually checked the userpages of the 100 most prolific contributors and found no unknown bots. Bot activity in other areas could be greater, as Wikipedia's own statistics show that most automated edits occur in non-English-language Wikipedias and reflect particular types of edits (e.g., formatting dates, deleting article stubs). (See <http://stats.wikimedia.org/EN/BotActivityMatrix.htm>.) Thus, though we may have missed some edits by unregistered bots in our sample, we excluded most automated activity. From the remaining full-revision history, we constructed a 89,406-observation monthly panel.

Dependent Variables

We operationalize viewership as the number of times a Wikipedia article is viewed in a given month. Viewership reflects the value that the market ascribes to particular content, and advertisers focus on content that delivers more viewers (Miller 2009). For each article, we collected the number of views each day from December 2007 until August 2011. These data are not available for the entire history of Wikipedia; although Wikipedia started in January 2001, viewership data is only available since December 2007. Therefore, our analysis does not reflect the early years of Wikipedia; data is uniformly missing for all topics for the early years. We summarized the view counts by month; then scaled the monthly article views by the number of days in the month so that months with fewer than 31 days were comparable with months with 31 days. Article views are integer counts; we transform the variable by taking the natural log.

We operationalize contributor activity as the number of unique people who make an edit to an article in a given month. (Activity on comment and talk pages are not included.) To test our hypothesized recursive relationship between viewership and content, we use simultaneous equation models. In our setting, we have two distinct, but interrelated, equations that model viewing activity and contribution activity. Importantly, the dependent variable in each equation is also an independent variable in the other equation. Therefore, to accurately estimate the multiple equations that have inherently correlated error

terms, we employ a three-stage least squares (3SLS) regression.

An important specification in simultaneous equation models is the need to identify the system of equations through a variable that will affect one set of equations and not the other. For this purpose, we chose the variable *protected*, a variable that indicates whether Wikipedia administrators have protected the article from additional contributions during the month of analysis. This protection is often used to prevent vandalism to the article and all instance of protection are logged by the system. Protection reduces the number of contributors to an article since fewer people can contribute; however, protection does not affect the ability to view the article. Thus, protection status affects editing activity but not viewing. Thus, for article i during period t , we use 3SLS to estimate the following equations simultaneously:

$$\ln(\text{views}_{i,t}) = \beta_1 X_{i,t} + \beta \ln(\text{views}_{i,t-1}) + \beta \text{authors}_{i,t} + \varepsilon_1 \quad (1)$$

$$\text{authors}_{i,t} = \beta_2 X_{i,t} + \beta \text{authors}_{i,t-1} + \beta \ln(\text{views}_{i,t}) + \beta \text{protected}_{i,t} + \varepsilon_2 \quad (2)$$

3.4.4. Content age. Age is the time in days since the article first appeared in Wikipedia; we use the natural log of the number of days. Article ages range from one day to 8.1 years, with an average of 2.9 years.

3.5.3. Anonymity of contributors. People can contribute to an article, whether they log in and identify themselves in the Wikipedia system or not. If a contributor is not logged in, his or her identity is recorded as an anonymous IP address. Because the raw number of anonymous contributors is highly correlated with the total number of contributors, we used the percentage of anonymous contributors, calculated as the total number of anonymous contributors divided by the total number of contributors to an article. On average, anonymous contributors made 29% of the contributions per article.

Control Variables

If the information artifact facilitates the transition from reader to contributor, it may be important control for aspects of this information artifact to account for these effects. To control for factors other than the recursive relationship between viewership and contribution, we include length, reading complexity, amount of multimedia content, information presentation, external references, and internal links as covariates. In Table 1, we present the descriptive statistics and in Table 2 the correlations of the variables.

Table 1: Descriptive Statistics

Variable	Minimum	Maximum	Mean	Std. Dev.
Age (days)	1.000	3,778.000	1,829.127	804.718
Length (characters/1000)	0.000	1,094.011	22.487	19.432
Complexity (ARI/1000)	0.001	209.86	20.262	3.811
Section Depth	1.000	6.000	2.880	0.785
External References	0.000	443.000	30.763	41.420
Internal Links	0.000	3,508.000	110.949	95.920
Multimedia Content	0.000	35.000	0.051	0.298
Anonymity (percentage)	0.000	0.782	0.318	0.154
Distinct Contributors	1.000	2,233.00	136.931	232.309
Monthly Article Views (/1000)	0.001	2,715.388	24.943	48.585

3.5.1. Length. Although Wikipedia has length guidelines (Wikipedia 2010) one group of active Wikipedians argues that, because it is not bound by the confines of traditional printed encyclopedias, an article should contain all possible relevant information about a particular topic (McAfee 2007). In short, an article may be perceived as more valuable by its audience simply because it has more; not better, information. To control for this possibility we include the length of each article, expressed in thousands of

characters of text (for scaling purposes), which ranges from 0 (for stub articles) to 1,094,010 characters. We use the natural log of article length in the statistical models.

3.5.2. Reading complexity. Articles may be more perceived as more valuable by its audience if written in a more sophisticated style. Complexity may also be an important signal to potential shapers (Majchrzak et al. 2013), as it may indicate how well developed content is. That is, articles may be perceived as containing valuable information if they sound authoritative, whether they actually are or not. Alternatively, articles may be incomprehensible if they are difficult to read. We control for the reading complexity of each article using the automated readability index (Smith and Senter 1967). (We applied models using the Coleman-Liau index and found similar results.) The ARI equals $ARI = (4.71 \times letters/words) + (0.5 \times words/sentences) - 21.43$, and estimates the U.S. school grade required to understand the article. For our analysis, relative values are more important than absolute values; the structure of Wikipedia articles results in relatively high ARI scores. For scaling purposes, we divide reading complexity by 1000.

3.5.4. Information presentation. Because multimedia content may influence the value of information (Schlosser 2003), we control for the total number of multimedia files (*multimedia content*). Similarly, an article’s organization may influence its market value, because well-organized information should be more accessible to readers. Articles in Wikipedia contain up to six levels of nested sections. To control for this effect, we include the maximum section level reached in the article, which we refer to as *section depth*.

Table 2: Variable Correlations

	1	2	3	4	5	6	7	8	9
1. Age (ln, days)	1.00								
2. Length (ln, chars)	0.33	1.00							
3. Complexity (ARI)	0.08	0.17	1.00						
4. Section Depth	0.26	0.48	0.18	1.00					
5. External References	0.18	0.82	0.19	0.34	1.00				
6. Internal Links	0.46	0.73	0.27	0.43	0.54	1.00			
7. Multimedia Content	0.02	0.06	-0.01	0.06	0.04	0.10	1.00		
8. Anonymity (%)	0.56	0.10	-0.02	0.17	-0.07	0.25	-0.01	1.00	
9. Contributors	0.53	0.46	0.09	0.29	0.34	0.59	0.04	0.45	1.00
10. Article Views (ln)	0.43	0.34	0.08	0.22	0.25	0.43	0.04	0.33	0.67

Correlations for the 91,907 Wikipedia Medicine monthly panel observations from December 2007 to July 2011. All correlations greater than 0.01 are significant.

3.5.5. References and links. Wikipedia policy states that all contributions should be supported by an authoritative external reference. The Medicine WikiProject considers only peer-reviewed medical journals authoritative. Contributors may attempt to manipulate the market value of the article by including more references, or the number of references could indicate the popularity of a topic in the medical literature. For example, although lung cancer is the leading cause of U.S. cancer deaths, it is relatively underfunded and under-researched compared with other forms of cancer (Khullar and Colson 2009). Articles also often contain links to other Wikipedia articles that may be sources of views or a reflection of market value. Accordingly, we control for the *number of external references* and the *number of internal links*.

3.5.6. Lagged Variables. Finally, we include the lagged value of each dependent variable to control for endogeneity and unchanging article-specific features that we do not directly observe. For instance, the topic of the article likely influences the number viewers and contributors. More people may be motivated to read an article on cancer than on a relatively obscure condition. By controlling for previous period viewership and contribution activity, our analysis controls for these unobserved characteristics of the article or the community and isolate the contribution dynamics that occur within the observation window.

Table 3: Three Stage Least Squares Model of Article Views

	Equation 1: Article Views (ln/1000)				Equation 2: Contributors			
	0	1	2	3	0	1	2	3
Monthly Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Constant	708.18*** (6.61)	725.80*** (6.96)	768.43*** (7.45)	799.45*** (7.75)	-70.81*** (8.07)	-74.47*** (10.50)	-97.12*** (11.49)	-56.31*** (12.16)
Views (ln, lagged)	0.99*** (0.01)	0.98*** (0.01)	0.98*** (0.01)	0.97*** (0.01)				
Contributors (lagged)					1.01*** (6.61)	1.01*** (6.96)	1.01*** (7.45)	1.01*** (7.75)
Length (ln, characters)	11.35*** (1.52)	11.36*** (1.52)	10.15*** (1.52)	9.76*** (1.52)	16.75*** (6.61)	-0.95 (6.96)	-0.44 (7.45)	-0.76 (7.75)
Complexity (ARI)	-0.13 (1.14)	1.57 (1.14)	3.94*** (1.15)	4.66*** (1.15)	-13.88*** (6.61)	-11.03*** (6.96)	-12.14*** (7.45)	-11.24*** (7.75)
Section Depth	3.51*** (1.00)	2.74** (1.00)	2.97** (1.00)	3.20** (1.00)	8.42*** (6.61)	2.23 (6.96)	2.14 (7.45)	2.28 (7.75)
External References	-3.46*** (0.80)	-3.79*** (0.80)	-1.71* (0.81)	1.43 (0.83)	-17.26*** (6.61)	-9.70*** (6.96)	-10.62*** (7.45)	-7.49*** (7.75)
Internal Links	7.01*** (0.97)	0.80 (1.04)	-0.93 (1.05)	0.71 (1.05)	31.13*** (6.61)	18.83*** (6.96)	19.70*** (7.45)	20.90*** (7.75)
Multimedia Content	0.71 (0.61)	1.26* (0.61)	1.69** (0.61)	1.86** (0.61)	-0.81 (6.61)	-1.29 (6.96)	-1.52 (7.45)	-1.20 (7.75)
Contributors		0.13*** (0.01)	0.12*** (0.01)	0.08*** (0.01)				
Contributors ² (/1000)		-0.07*** (0.01)	-0.06*** (0.01)	-0.04*** (0.01)				
Article Protected?					-500.3*** (6.61)	-486.0*** (6.96)	-487.2*** (7.45)	-481.7*** (7.75)
Views (ln)						0.04*** (6.96)	0.04*** (7.45)	0.04*** (7.75)
Age (ln, years)			26.77*** (1.69)	20.76*** (1.74)			-12.30*** (7.45)	-18.57*** (7.75)
Anonymity (%)				17.13*** (1.17)				17.56*** (7.75)
Chi ² (x 10 ⁶)	6.11***	6.12***	6.14***	6.15***	4.67***	4.75***	4.75***	4.76***

80,806 monthly observations, standard errors in parentheses; significance *p<0.05, **p<0.01, ***p<0.001

Results and Discussion

We perform two stages of data analysis to test our hypotheses. First, we first analyze the effects of network structure on article views for the entire sample. Second, we use two alternative samples that demonstrate the robustness of our findings beyond the Medicine WikiProject. Table 3 summarizes the full results of a simultaneous equation 3SLS regression on the natural log of article views (scaled by 1000 for this presentation) using 89,406 monthly observations of articles from December 2007 until August 2011. 3SLS explicitly estimates interdependent endogenous effects.

We find support for our first hypothesis, the number of people who view an information artifact created by a production community positively related ($\beta = 0.035$, $p < .001$) to the number of people who choose to contribute to it. This suggests that the information artifact can serve as a boundary object that can facilitate the transition from passive lurkers to active contributors. We also find support for our second hypothesis; the number of contributors is significantly and curvilinearly related to how often that content is viewed by others, such that a moderate number of contributors create an information artifact that attracts the most viewers. The linear ($\beta = 0.077$, $p < .001$) and the squared term ($\beta = -0.035$, $p < .001$) indicate a concave relationship. These results confirm that more contributors to an information artifact attract more viewership but only up to a point. Taken together, these hypotheses confirm a recursive relationship between viewership and contribution to an online production community in which both aspects of participation reinforce each other.

Additionally, we also find support for hypothesis 3. Article age is positively related to the number of viewers ($\beta = 20.764$, $p < .001$), but negatively related to the number of contributors ($\beta = -18.570$, $p < .001$). The recursive relationship between viewership and content attenuates over time. As the content matures, it attracts a greater number of viewers but these viewers are less likely to develop into contributors. These viewers likely perceive the information artifact to be more complete and less in need of their contributions, making them less likely to contribute to it because they recognize there is little they can add. Thus, the production environment evolves toward a more traditional media environment that emphasizes the viewing of produced content, rather than high levels of sustained ongoing collaboration.

Finally, we find support our hypothesis on the influence of anonymity. Although previous research documents some conflicting effects of anonymity on collaboration, here the effect is uniformly positive. The percentage of anonymous contributors is positively related to both the number of viewers ($\beta = 17.131$, $p < .001$) and the number of contributors ($B = 17.562$, $p < .001$). The ability for users to remain anonymous facilitates their transition from viewer to contributor, and since anonymity may particularly encourage people to contribute knowledge that may diverge from the existing knowledge base, may be more likely to improve the information artifact and attract viewers. The protection variable, which affects contributions but not viewing, is significant and has a substantial impact on reducing the number of contributors ($\beta = -481.713$, $p < 0.01$). When a production community enacts technological barriers to prevent certain types of people from contributing, these methods lower the number of contributors.

Although not explicitly hypothesized it is also instructive to examine the effects of our control variables, which accounts for how the features of the article affect these contribution dynamics. On the viewership side, the content of the information artifacts drives viewership. Article length ($\beta = 9.757$, $p < 0.001$), section depth ($\beta = 3.195$, $p < 0.001$), writing complexity ($\beta = 4.658$, $p < 0.001$), and number of multimedia images ($\beta = 1.864$, $p < 0.01$) all significantly and positively relate to viewership. Taken together, these results indicate that features of content created by the production community have an important effect on whether people will ultimately view that content. Content that is longer, better organized, of a more sophisticated writing style, and utilizes multimedia attracts greater viewership.

Taken together, the results of our hypotheses and our control variables tell a consistent story. In online production communities, there is a recursive relationship between the number of viewers and the number of contributors. More viewers lead to more contributors. These contributors create a more robust information artifact, which then attracts more viewers. As the information artifact becomes fully developed, it continues to attract more viewers, but these viewers are less likely to become contributors. Thus, there is a virtuous cycle between viewers and contributors, but this cycle diminishes over time as the content becomes more mature and stable.

Alternative Sample

To assess the predictive validity of the models, we used both an internal holdout sample (to assess temporal predictive validity) and external samples (to assess predictive validity in alternate contexts). First, we generated coefficient estimates using only the first thirty months of data. We then used these estimators to predict monthly article viewing for the remaining eight months. The out-of-sample Mean Absolute Percentage Error for viewership was 1.5% and for contributors was 7.1%; this indicates the models yield accurate estimates of future viewership within the same context. Second, we used the estimators from the Medicine Wikiproject sample to generate predicted values of monthly article viewing for the fashion and auto Wikiprojects. These Wikiprojects are comparatively smaller (2,503 and 6,890 articles, respectively) but are interesting to study because of the importance of viewership to these industries. We collected the full text of 644,336 revisions in the Fashion Wikiproject and 1,026,892 revisions in the Auto Wikiproject, then analyzed monthly viewing over the same period (December 2007–August 2011). MAPEs for the alternative samples are slightly higher but remain low and confirm that the models can be generalized outside the context of our original sample.

Discussion

For this article, we studied the entire compendium of 16,298 Wikipedia articles in the Medicine Wikiproject to study collaborative dynamics of online production communities. We hypothesized a recursive relationship between viewership and contributors such that viewership would lead to more contributors and more contributors would lead to more viewers. We also hypothesized that this recursive relationship would stabilize over time, as the community matured, increasing viewership but decreasing the addition of new contributors. Finally, we explored the effect of anonymity on this dynamic relationship. In general, we found good support for our hypothesis. We do find evidence for a virtuous cycle in which viewers and content reinforce one another. Our control variables associated with the content produced by the community also suggests that this artifact does serve as a boundary object that attracts viewers and invites contributors. This reinforcing cycle, however, diminishes over time. While the community continues to attract a greater number of viewers, it becomes less likely to convert those viewers to contributors. Finally, anonymity does facilitate this cycle, encouraging people to contribute which creates a more attractive information artifact.

Theoretical Contributions

This article has several implications for theory. First, recent research has lamented the lack of research into the dynamic relationships of online communities (Faraj and Johnson 2011). In this paper, we respond to this call by examining the relationship between viewership and contribution behavior in online production communities. We expand on traditional theories of online community involvement that suggests people gradually progress from lurking observer to active contributor (Lave and Wenger 1991; Preece and Schneiderman 2009). We find evidence for the existence of a positive reinforcing relationship between these two dynamics, such that more viewers lead to more contributors and vice-versa. While we find that people to progress from lurker to contributor, we also find that for online production communities that create information artifacts, the work of the contributors can also attract additional viewers. Future research can build on our work to examine other types of dynamic interactions in online communities, particularly given the vast amounts of data now available over time. Further research should take advantage of the granularity of data to explore heterogeneity in these dynamics within subsets of the data.

Managerial Contributions

The findings of this article should be of particular interest to managers seeking to cultivate collaborative content. First, the curvilinear relationship between number of contributors and value of collaborative user-generated content suggests that managers should not necessarily pursue a more-is-better strategy toward the number of contributors. Although it is important to generate sufficient participation, once content attains a critical mass of contributors, it may be necessary to redirect new contributors to other

content—particularly if there is a virtuous cycle in which increased viewing leads to more contributors. Our data should not be used to predict the optimal number of contributors to a particular content source though, because the optimal number differs by cluster. Yet we argue that the search for contributors becomes unnecessary or even counterproductive after a point.

Second, we demonstrate the types of analyses becoming possible for online community managers. As online communities generate vast amounts of data over time, managers can use this data to assess the productivity of certain types of collaboration within employee groups or customer communities. As new types of communication and collaboration technology become ubiquitous, all work might become online collaboration to a certain degree. If collaboration platforms provide dashboards or network visualizations, organizations using these platforms for internal collaboration could more easily view the work of their of their employees (Smith and Fiore 2001). With this information, managers can identify possible information bottlenecks, find employees with considerable informal influence, and determine if particular groups who should be talking to one another are not. Managers further could plan interventions to improve collaboration activity and then use the dashboard to determine if the interventions were successful. Whereas previous generations of managers had to “walk around” to obtain imperfect information about the organization (Peters and Waterman 1982), managers whose employees communicate through online collaboration tools may gain access to a much more robust version of the same information, through their computers.

Limitations and Conclusion

This study has several limitations that should be considered when assessing the impact of this study. One limitation is that, although Wikipedia provides us with considerable data to analyze the unique contribution dynamics across thousands of projects over many years, the uniqueness of Wikipedia as a production community may limit its generalizability somewhat. Although we can think of no theoretical reason why the findings here would not apply to online production communities more generally, the findings here should be validated in other production communities. Furthermore, we use viewership to reflect article value but the measure is imperfect; additional research is needed to better understand the interrelationship between viewership and value. Another limitation is that we cannot determine why people are moving from viewer to contributor or why they viewed the content to begin with. These motivations may be important moderators of the contribution dynamics, yet we do not measure these motivations.

Despite these limitations, this study represents an important first step in studying the dynamics involved with online communities that have been under-researched (Faraj and Johnson 2011). The methods used here could be applied to study other types of dynamics in other types of online communities, such as product review, customer support, or knowledge management communities. These communities and their dynamics are likely to become more important as social media is adopted more widely within organizations for a variety of business purposes. Understanding how these communities function and why people participate in them, therefore, will be an important component of business analytics in the coming years. Further research in this important area is warranted.

References

- Asvanund, A., Clay, K., Krishnan, R., and Smith, M.D. 2004. "An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks," *Information Systems Research* (15:2), Jun, pp. 155-174.
- Borgatti, S.P., and Cross, R. 2003. "A Relational View of Information Seeking and Learning in Social Networks," *Management Science* (49:4), Apr, pp. 432-445.
- Brooks, F.P., Jr. 1975. *The Mythical Man-Month: Essays on Software Engineering*. Reading, MA: Addison-Wesley.
- Butler, B.S. 2001. "Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures," *Information Systems Research* (12:4), Dec, pp. 346-362.
- Carlile, P.R. 2002. "A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development," *Organization Science* (13:4), Jul-Aug, pp. 442-455.

- Clauson, K.A., Polen, H.H., Boulos, M.N.K., and Dzenowagis, J.H. 2008. "Scope, Completeness, and Accuracy of Drug Information in Wikipedia," *The Annals of pharmacotherapy* (42:12), pp. 1814-1821.
- Connolly, T., Jessup, M.L., and Valacich, J.S. 1990. "Effects of Anonymity and Evaluative Tone on Idea Generation in Computer Mediated Groups," *Management Science* (36:6), pp. 689-703.
- Constant, D., Sproull, L., and Kiesler, S. 1996. "The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice," *Organization Science* (7:2), pp. 119-135.
- Dellarocas, C., Gao, G.D., and Narayan, R. 2010. "Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?," *Journal of Management Information Systems* (27:2), Fal, pp. 127-157.
- Denning, P., Horning, J., Parnas, D., and Weinstein, L. 2005. "Wikipedia Risks," *Communications of the ACM* (48:12), p. 152.
- Devgan, L., Powe, N., Blakey, B., and Makary, M. 2007. "Wiki-Surgery? Internal Validity of Wikipedia as a Medical and Surgical Reference," *Journal of the American College of Surgeons* (205:3), pp. S76-S77.
- Espinosa, J.A., Slaughter, S.A., Kraut, R.E., and Herbsleb, J.D. 2007. "Team Knowledge and Coordination in Geographically Distributed Software Development," *Journal of Management Information Systems* (24:1), pp. 135-169.
- Faraj, S., Jarvenpaa, S.L., and Majchrzak, A. 2011. "Knowledge Collaboration in Online Communities," *Organization Science* (22), pp. 1224-1239.
- Faraj, S., and Johnson, S.L. 2011. "Network Exchange Patterns in Online Communities," *Organization Science* (22:6), Nov-Dec, pp. 1464-1480.
- Ferguson, T., and Frydman, G. 2004. "The First Generation of E-Patients," *British Medical Journal* (328:7449), pp. 1148-1149.
- Fleming, L., and Waguespack, D.M. 2007. "Brokerage, Boundary Spanning, and Leadership in Open Innovation Communities," *Organization Science* (18:2), Mar-Apr, pp. 165-180.
- Foutz, N.Z., and Jank, W. 2009. "Research Note--Prerelease Demand Forecasting for Motion Pictures Using Functional Shape Analysis of Virtual Stock Markets," *Marketing Science*, December, p. mksc.1090.0542.
- Fox, S., and Jones, S. 2009. "The Social Life of Health Information," Pew Research Center, Washington, DC.
- Gersick, C.J.G. 1988. "Time and Transition in Work Teams: Toward a New Model of Group Development," *Academy of Management Journal* (31:1), pp. 9-41.
- Griffin, A., and Hauser, J.R. 1993. "The Voice of the Customer," *Marketing Science* (12), Winter, pp. 1-27.
- Griffith, T.L., Sawyer, J.E., and Neale, M.A. 2003. "Virtualness and Knowledge in Teams: Managing the Love Triangle of Organizations, Individuals, and Information Technology," *Mis Quarterly* (27:2), Jun, pp. 265-287.
- Hansen, M.T., and Haas, M.R. 2001. "Competing for Attention in Knowledge Markets: Electronic Document Dissemination in a Management Consulting Company," *Administrative Science Quarterly* (46:1), pp. 1-28.
- Hiltz, S.R., and Turoff, M. 1985. "Structuring Computer-Mediated Communication Systems to Avoid Information Overload," *Communications of the ACM* (28:7), pp. 680-699.
- Jeppesen, L.B., and Frederiksen, L. 2006. "Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer-Controlled Music Instruments," *Organization Science* (17:1), Jan-Feb, pp. 45-63.
- Jessup, L.M., Connolly, T., and Galegher, J. 1990. "The Effects of Anonymity on Gdss Group-Process with an Idea-Generating Task," *Mis Quarterly* (14:3), Sep, pp. 313-321.
- Jolson, M.A., and Rossow, G.L. 1971. "The Delphi Process in Marketing Decision Making," *Journal of Marketing Research* (8:4), pp. 443-448.
- Kane, G.C. 2011. "A Multimethod Study of Information Quality in Wiki Collaboration," *ACM Transactions on Management Information Systems* (2:1).
- Kane, G.C., Fichman, R.G., Gallagher, J., and Glaser, J. 2009a. "Community Relations 2.0," *Harvard Business Review* (87:11), pp. 45-50.
- Kane, G.C., Majchrzak, A., Johnson, J., and Chen, G.L. 2009b. "A Lifecycle Model of Perspective Making and Perspective Taking in Fluid Online Collectives," in: *International Conference on Information Systems*. Phoenix, AZ.
- Khullar, O., and Colson, Y.L. 2009. "The Underfunding of Lung Cancer Research," *J. Thoracic*

- Cardiovascular Surgery* (138:2), p. 275.
- Kiron, D., Palmer, D., Phillips, A.N., and Kruschwitz, N. 2012. "Social Business: What Are Companies Really Doing?: 2012 Social Business Global Executive Study and Research Project.," *Sloan Management Review*).
- Kittur, A., and Kraut, R.E. 2008. "Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination," in: *Proceedings of the 2008 ACM conference on computer supported cooperative work*. San Diego: ACM, pp. 37-46.
- Kriplean, T., Beschastnikh, I., and McDonald, D.W. 2008. "Articulations of Wikiwork: Uncovering Valued Work in Wikipedia through Barnstars," in: *Proceedings of the 2008 ACM conference on computer supported cooperative work*. San Diego: ACM, pp. 47-56.
- Kuk, G. 2006. "Strategic Interaction and Knowledge Sharing in the Kde Developer Mailing List," *Management Science* (52:7), Jul, pp. 1031-1042.
- Laurent, M.R., and Vickers, T.J. 2009. "Seeking Health Information Online: Does Wikipedia Matter?," *Journal of the American Medical Informatics Association* (16:4), pp. 471-479.
- Lave, J., and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.
- Lee, G.K., and Cole, R.E. 2003. "From a Firm-Based to a Community-Based Model of Knowledge Creation: The Case of the Linux Kernel Development," *Organization Science* (14:6), Nov-Dec, pp. 633-649.
- Levina, N. 2005. "Collaborating on Multiparty Information Systems Development Projects: A Collective Reflection-in-Action View," *Information Systems Research* (16:2), Jun, pp. 109-130.
- Ma, M., and Agarwal, R. 2007. "Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities," *Information Systems Research* (18:1), Mar, pp. 42-67.
- Majchrzak, A., Wagner, C., and Yates, D. 2013. "The Impact of Shaping on Knowledge Reuse for Organizational Improvement with Wikis," *Mis Quarterly* (37:2), pp. 455-469.
- Markus, M.L. 2001. "Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success," *Journal of Management Information Systems* (18:1), Sum, pp. 57-93.
- Martins, L.L., Gilson, L.L., and Maynard, M.T. 2004. "Virtual Eeams: What Do We Know and Where Do We Go from Here?," *Journal of Management* (30:6), pp. 805-835.
- Marwell, G., and Oliver, P.E. 1993. *The Critical Mass in Collective Action: A Micro-Social Theory*. Cambridge University Press.
- McAfee, A.P. 2007. "Wikipedia (B)," in: *Harvard Business School Supplement* Cambridge, MA: Harvard Business School.
- Miller, C.C. 2009. "Ad Revenue on the Web? No Sure Bet," in: *New York Times*. p. B1.
- Oh, W., and Jeon, S. 2007. "Membership Herding and Network Stability in the Open Source Community: The Ising Perspective," *Management Science* (53:7), Jul, pp. 1086-1101.
- Oliver, P.E., Marwell, G., and Teixeira, R. 1985. "A Theory of the Critical Mass: Interdependence, Group Heterogeneity, and the Production of Collective Action," *American Journal of Sociology* (91), pp. 522-556.
- Perry-Smith, J.E., and Shalley, C.E. 2003. "The Social Side of Creativity: A Static and Dynamic Social Network Perspective," *Academy of Management Review* (28:1), Jan, pp. 89-106.
- Peters, T.J., and Waterman, R.H. 1982. *In Search of Excellence: Lessons from America's Best Run Companies*. New York: HarperCollins Publishers.
- Phillips, L.E. 2007. "Pharmaceutical Marketing Online: Stuck in Web 1.5," in: *eMarketer*.
- Preece, J., and Schneiderman, B. 2009. "The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation," *AIS Transactions on Human-Computer Interaction* (1:1), pp. 13-32.
- Ransbotham, S., and Kane, G.C. 2011. "Membership Turnover and Collaborative Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia," *Mis Quarterly* (35:3), pp. 613-627.
- Ransbotham, S., Kane, G.C., and Lurie, N. 2012. "Network Characteristics and the Value of Collaborative User-Generated Content," *Marketing Science* (31:3), pp. 387-405.
- Raymond, E.S. 1999. *The Cathedral & the Bazaar : Musings on Linux and Open Source by an Accidental Revolutionary*, (1st ed.). Cambridge, Mass.: O'Reilly.
- Schlosser, A.E. 2003. "Experiencing Products in the Virtual World: The Role of Goal and Imagery in Influencing Attitudes Versus Purchase Intentions," *Journal of Consumer Research* (30:September), pp. 184-198.

- Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*. New York: John Wiley.
- Sia, C.L., Tan, B.C.Y., and Wei, K.K. 2002. "Group Polarization and Computer-Mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity," *Information Systems Research* (13:1), Mar, pp. 70-90.
- Siegel, J., Dubrovsky, V., Kiesler, S.B., and McGuire, T.W. 1986. "Group Processes in Computer-Mediated Communication," *Organizational Behavior and Human Decision Processes* (37:2), pp. 157-187.
- Smith, E.A., and Senter, R.J. 1967. "Automated Readability Index." Wright-Patterson Airforce Base, OH: Aerospace Medical Research Laboratories, pp. 1-14.
- Smith, M.A., and Fiore, A.T. 2001. "Visualization Components for Persistent Conversations," in: *Proceedings of the SIGCHI conference on Human factors in computing systems*. Seattle, Washington, United States: ACM, pp. 136-143.
- Stewart, K.J., and Gosain, S. 2006. "The Impact of Ideology on Effectiveness in Open Source Software Development Teams," *Mis Quarterly* (30:2), Jun, pp. 291-314.
- Tuckman, B.W. 1965. "Developmental Sequence in Small Groups," *Psychological Bulletin* (63:6), pp. 384-399.
- Valacich, J.S., Dennis, A.R., and Nunamaker, J.F. 1992. "Group Size and Anonymity Effects on Computer-Mediated Idea Generation," *Small Group Research* (25:1), pp. 83-104.
- Wang, X., Butler, B.S., and Ren, Y. 2013. "The Impact of Membership Overlap on Growth: An Ecological Competition View of Online Groups," *Organization Science* (24:2), pp. 414-431.
- Wikipedia. 2010. "Wikipedia: Article Size."
- Zhang, X., and Zhu, F. 2011. "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia," *American Economic Review* (101:4), pp. 1601-1615.