Benjamin Mako Hill — CAREER: New Approaches to Managing Lifecycles of Digital Knowledge Commons

Digital knowledge commons like Wikipedia, open source software, and collaborative filtering systems like Reddit produce enormous social and economic value and serve as critical information infrastructure. These online communities rely on "peer production" to aggregate contributions from Internet users into vast knowledge bases that are then made freely available. Decades after many of the most important peer produced knowledge commons were launched, many are under attack by vandalism, disinformation campaigns, and a range of special interests. At the same time, many of the largely volunteer-based groups who sustain mature communities have been stable or shrinking for years.

A body of research suggests that these patterns of decline are due—at least in part—to commons becoming increasingly closed to contributions. Why do peer produced knowledge commons increasingly reject the work of volunteers necessary for their long-term survival? How general are these observed lifecycle dynamics? How should communities structure themselves to better manage growth? How should they balance the competing goals of remaining open to contributions while protecting the value they have produced? Integrating and building on a body of social computing and social scientific research, the proposed work will seek to answer these questions by developing and validating a general theory of knowledge commons' lifecycles and identifying a set of strategies to help structure and govern peer produced commons effectively as they grow.

In four parts, the project will attempt to (A) develop a theoretical framework to explain why online communities follow regular patterns of growth and decline and (B) conduct a series of empirical studies of wikis, open source software, and collaborative filtering sites. Using insights from the first two parts, the work will seek to (C) identify a set of strategies for the effective management of lifecycles in knowledge commons. Finally, the work will (D) create tools and datasets to help researchers and practitioners manage online community lifecycles. This work will be conducted in close collaboration with community managers and disseminated through a series of outreach-focused meetings, workshops, and information resources as well as through scholarly publications and university classes.

The **intellectual merit** of this proposal lies in three features of the work: first, it will develop an explanation for why knowledge commons become increasingly closed and how these changes drive lifecycle patterns; second, it will contribute detailed empirical insights into the sociotechnical dynamics of knowledge commons; finally, it will develop a range of new computational methods for studying online activity. The **broader impacts** of this work stem from the fact that it will provide insight into how to better support the development and the maintenance of knowledge commons. In doing so, this work can impact the millions of people who contribute to these knowledge commons and the billions who rely on their products in their business and personal lives.

1. Introduction & Project Goals

Wikipedia is among the most popular websites on the Internet and is visited by hundreds of millions of people each month [82]. In many ways, these facts fail to capture Wikipedia's enormous impact. Wikipedia is a primary source of information returned by virtual assistants like Amazon's Alexa and Apple's Siri as well as the information highlighted at the top of a Google's search [81]. In other realms, Wikipedia has provided the raw material for tens of thousands of academic papers in fields as varied as information retrieval, natural language processing, sociology, and law [57, 41]. It provides the raw material used to train machine learning algorithms for machine translation, predictive text input, and other algorithms relied upon by millions [e.g., 75, 91].

Wikipedia is constructed through *peer production*, a term coined by Yochai Benkler [2] to describe a model of production that relies on the mass aggregation of many small contributions from large numbers of diversely motivated individuals working together over the Internet. Although it is one of the most visible examples, English Wikipedia is only the tip of the peer production iceberg. There are Wikipedias for more

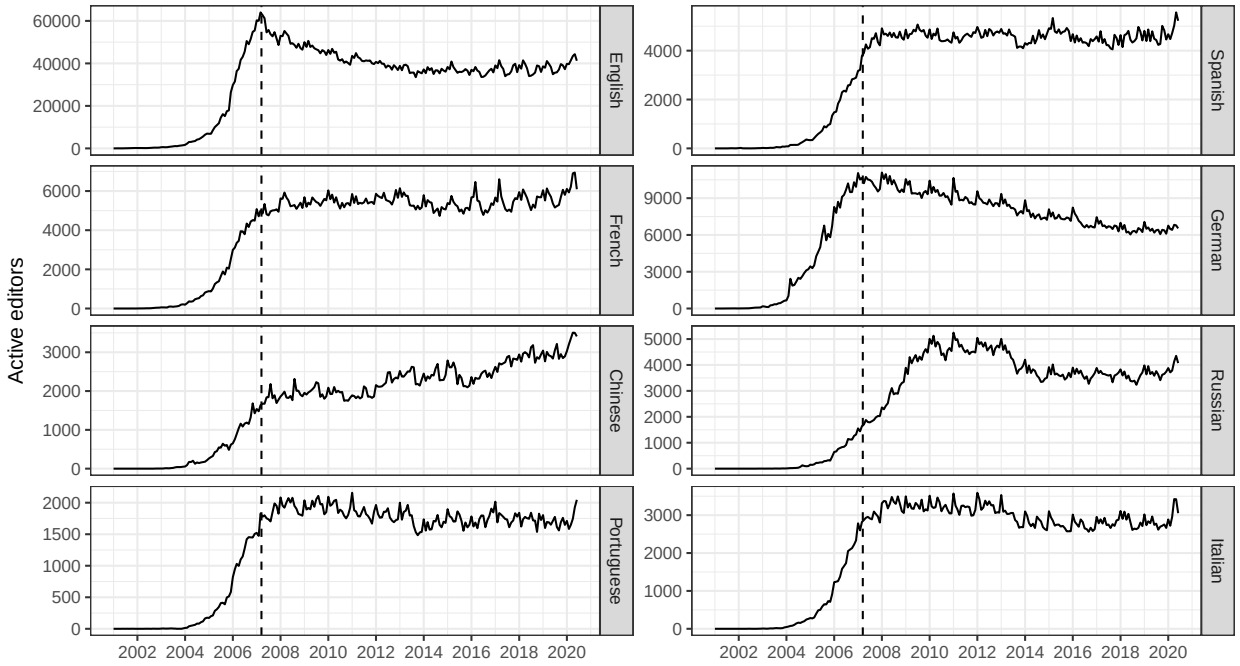

Figure 1: Number of users making at least five edits per month to different Wikipedia language editions over time for the eight largest Wikipedia editions as measured by the number of edits. Dashed lines reflect the peak of English Wikipedia editing activity in March 2007.

than 300 languages (15 with more than a million articles) and hundreds of thousands of other peer production communities including wikis, mapping systems like Open Street Map, and collaborative filtering systems like Reddit. Free/libre open source software (FLOSS) may be the most impactful example of the organizational form. It is simply not possible to use the Internet today without relying on infrastructure produced by masses of individuals working together. Peer production is among the most important organizational innovations made possible by the Internet [4].

Although most research on peer production has celebrated its achievements, large peer production projects face stark challenges that are typically much less visible [4]. One of the most important and underappreciated challenges is community lifecycle dynamics: most mature peer production projects face the prospect of maintaining rapidly increasing knowledge bases with stable or shrinking numbers of volunteers. Suh et al. [83] pointed out that English Wikipedia's meteoric growth in active contributors came transitioned into decline in early 2007. The top left panel of Figure 1 shows that the pattern of decline continued for several years before stabilizing at a lower level.

The pattern of growth and decline that Suh et al. noticed in English Wikipedia seems to be a general feature of peer production. The eight panels of Figure 1 show the eight largest Wikipedia editions by the number of edits and suggest that the pattern shown in English Wikipedia in the top left is remarkably general across many large Wikipedia language editions. This dynamic is visible in Figure 2 from one of my papers [87] which shows activity across the largest 740 wikis hosted by the largest commercial peer production wiki hosting firm Wikia. After total activity is standardized along the y-axis and after wikis—founded at many different periods of time—are sorted in terms of age, a pattern of rise and decline over a period of about 4–5 years becomes clear in the data. Similar lifecycles have been noted by Schweik and English's [74] study of FLOSS and are the implicit motivation for the Linux Foundation's Community Health Analytics for Open Source Software (CHAOSS) Evolution Working group—formerly known as the "Growth, Maturity and Decline Working Group" [50].

Figures 1 and 2 provide some evidence against several common explanations. For example, the vastly different scales on the y-axes in Figure 1 (obscured by standardization in Figure 2) suggest that the "ceiling" for different Wikipedias varies enormously. Because different language Wikipedia versions have the same scope but began to decline after having written very different numbers of articles, this suggests that participation in English Wikipedia has not declined simply because English Wikipedia is complete. Nor can the decline be explained by broader temporal trends—the dashed lines in the figures correspond to English Wikipedia's peak and suggest that although several other Wikipedia editions peaked at around the same time as English Wikipedia, communities like Russian Wikipedia seem to have followed a very similar pattern to English but at a several-year delay.

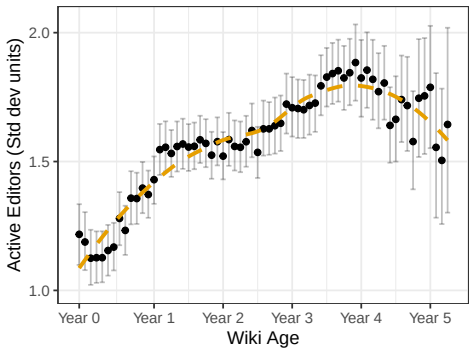

Figure 2: Figure is adapted from Figure 1 in TeBlunthuis et al. [87].

Halfaker et al. [24] argue that English Wikipedia's decline was a function of increased difficulties for newcomers leading to decreased newcomer retention. This work is supported by a range of studies that suggest that Wikipedia has become more formalized and bureaucratic [e.g., 6, 52, 71]. Halfaker et al. [24] argue that English Wikipedia is in decline not because newcomers have stopped showing up. Instead, they claim that English Wikipedia is in decline because it is becoming less open. In two related followup studies, I have shown that this institutional pattern is repeated across the largest peer production wikis [77, 87].

Although a number of scholars have pointed to increased wariness toward newcomers as a problem to be solved, it also presents a major empirical puzzle: **Why do mature knowledge commons become increasingly closed in ways that reduce the influx of participants that sustain them over time?**

2. Research Plan

The work I propose will address this puzzle through four overlapping initiatives. I will develop a theoretical framework that will answer this question and guide the rest of this work (§2.1), I will conduct a series of empirical analyses that will inform and serve to validate my theory (§2.2), I will identify a set of novel and effective strategies for governing knowledge commons over their lifecycles (§2.3), and I will produce datasets and tools that are useful to both academics and practitioners (§2.4).

2.1. Part A: Developing a Theoretical Framework

At the core of this proposed work is a theoretical framework I will develop to answer the puzzle posed above. My answer will involve an articulation of a theory in substantive terms, a formal analytic model, and a series of agent-based simulations. Although I will illustrate my theory using examples from Wikipedia where evidence is more readily available, I will explain how I plan to establish generalizability in §2.2.4.

2.1.1. Background: Theories of Peer Production There are two broad sets of social scientific theories that have been used to understand the organization of peer produced knowledge commons: (a) theories of collective action and public good provision, and (b) theories of common-pool resources and commons. The first set of explanations draws from classic work on public good provision by Mancur Olson [63] and is largely concerned with free ridership, managing "selective incentives," and issues related to achieving critical mass [e.g., 53]. These approaches seem like a good fit for peer produced knowledge because collective action is defined as the provision of non-excludable and non-rival public goods [63, 65]. The most obvious outputs of peer production—useful knowledge like an encyclopedia article or piece of code—are archetypes of pure public goods. By implicitly invoking a theory of selective incentives and explaining peer production in terms of lowered transaction cost, Yochai Benkler took this perspective when he theorized about peer production [2]. The majority of organizational theorizing about peer production has followed Benkler's lead.

The second set of explanations draws from the work of Elinor Ostrom and conceives of knowledge commons as common-pool resources (CPRs). Most of Ostrom's work focused on understanding the governance of natural CPRs like irrigation systems, fisheries, and forests [64]. Because CPRs are distinguished from public goods in terms of rivalness or subtractability (e.g., a fish in a fishery can only be removed once),[1] work on CPRs by Ostrom and others has primarily been focused on managing appropriation and identifying ways that CPRs avoid the "tragedy of the commons" when participants acting in self-interest take too much for themselves [25].[2] Attracted by the the metaphor of a commons as a way to understand knowledge bases [5], and because Ostrom's approach to thinking about governance is broader than just CPRs, a small number of scholars of sociotechnical systems have built on Ostrom's work. These studies have tended to focus on questions of institutional design and governance or have attempted to apply Ostrom's analytic framework to the contexts of peer produced knowledge goods [e.g., 5, 20, 21, 30, 79, 80]. Because knowledge goods like software and encyclopedia articles can be consumed without depletion, issues of appropriation have received relatively little discussion in social computing theories of peer production projects.

In general, institutional "openness" is often cited as the key feature in peer production and digital organization [3, 4, 78]. It is such a central concept that it exists in the name of peer production phenomena like open source, open hardware, and so on. Although openness remains undertheorized—something I hope to address in this work—the term is consistently used by both scholars and practitioners to refer to the relative absence of formal organizational boundaries and institutional structures. Surprisingly, the two bodies of theory described above seem to have very different things to say about open institutions. In Benkler's collective action framing, rules and boundaries reflect costs which will limit contributions from lightly motivated individuals. In Ostrom's approach, rules and boundaries are critical to preventing appropriation—"clearly defined boundaries" is the very first item in Ostrom's famous list of eight principles for effective CPR governance. Institutional openness is a prerequisite for effective peer production of knowledge in Benkler's framework. In Ostrom's CPRs, it places a commons at risk of ruin.

2.1.2. Sketch of a Theoretical Model My theoretical approach seeks to untangle this contradiction by conceiving of knowledge commons as having two distinct goals. The first goal is *building a stock of value*. Although the specific types of value will vary enormously between communities, all attempts at peer produced knowledge commons have this first goal. Open institutions are good for building valuable knowledge commons for reasons Benkler [2] articulates well. The second goal involves *protecting the stock of value from appropriation*. This second goal is only salient in situations where a community succeeds in their first goal and develops something to protect. Increased openness makes this second goal more difficult. Pursuing these goals in tandem requires managing trade-offs related to openness.

A more structured sketch of this explanation assumes that every knowledge commons has some stock of value and some level of openness to new contributions. During any particular time period, communities will receive some number of contributions which will increase their stock. The model also assumes that knowledge commons may be subject to damage that detracts from the stock. This is a counterintuitive claim that I unpack in detail in §2.1.3. The model also assumes that both good contributions and damaging ones are more likely if communities are open. Openness in this sense acts as an imperfect filter. Communities faced with an influx of bad contributions become less open in order to prevent future damage. Critically, my model also assumes that the opposite dynamics play out as a function of stock. Communities experience a virtuous cycle where they tend to receive more good contributions as their stock grows. However, the model also assumes a vicious cycle where swelling stocks invite increased damaging attempts at appropriation.

A very simple worked version of this model over three steps is shown in Figure 3. In the first period, a very open community receives 20 stock-increasing contributions. Because its stock has grown, this promotes a

[1] Work breaking down of the difference between public goods and common-pool resources relies can be found in Ostrom and Ostrom [65].

[2] Ostrom distinguishes between provision problems and appropriation problems in CPRs but is almost exclusively focused on the latter. When Ostrom discusses provision, she is focused on how participants in a CPR provide institutions to manage the underlying resource. In Ostrom's work, underlying resources are rarely if ever produced by individuals.

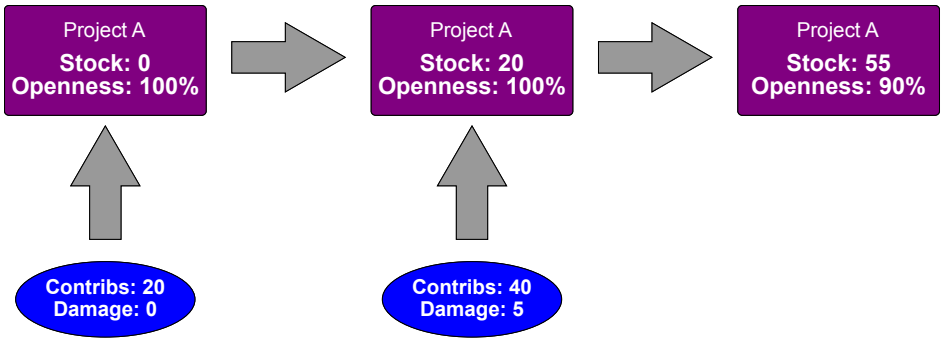

Figure 3: A simple example to illustrate the basic dynamics at the core of the theoretical model. The example is narrated in the text.

virtuous cycle where the stock grows by an even larger amount in the next period. However, the swelling stock also attracts damaging attempts at appropriation. To prevent further damage, the community becomes slightly less open in the third stage. In the terms of the theories discussed in §2.1.1, communities begin as pure collective action problems and then—to the extent that they are successful—become CPRs that must be concerned with appropriation.

2.1.3. Appropriation in Peer Production The explanation I have sketched out above requires conceiving of stocks of value as being at least partially appropriable. As I explain in §2.1.1, the knowledge goods produced by peer production are not typically thought of as producing appropriable information goods. For example, once a person contributes a sentence documenting a fact to Wikipedia, that text and fact is non-rival and non-excludable. However, my theorical argument suggests that peer production projects also create a range of forms of appropriable value. Examples include goodwill, reputation, and trust that the public has in a platform. Behaviors like Wikipedia vandalism spreading disinformation brings benefits like enjoyment to vandals while reducing the value the public has in Wikipedia [7].

In a well documented example of Wikipedia vandalism, an article about the journalist and political figure John Seigenthaler was created by an anonymous user—later identified as Brian Chase—that included false statements that Seigenthaler had been a suspect in the assassinations of John F. and Robert F. Kennedy [43]. Six months later, Seigenthaler published op-ed in *USA Today* describing Wikipedia as a vehicle for "Internet character assassination" [76]. The incident received wide international press coverage and reflects one of the greatest crises in Wikipedia's history [14, 27, 43]. Chase claimed the misinformation he added to the Seigenthaler article was intended as a joke [58]. Chase's contribution can be understood as appropriation because the benefit or enjoyment he derived from making the joke came at the enormous cost of damage to the public goodwill Wikipedia had built over time.

In the aftermath of Seigenthaler's op-ed, Wikipedia instituted new rules that made it more difficult to contribute [27]. It banned all new users from creating articles. It created a policy that allowed biographical articles of living people without references to be deleted without discussion. It created new technical features that allowed administrators and Wikipedia staff to delete content without discussion. Of course, most articles created by new users had been unproblematic. The likely loss of some of these good contributions was part of the cost Wikipedia chose to pay in prevent another incident. Despite these changes, hoaxes and disinformation remains common in Wikipedia [49, 72]. My framework suggests that more people want to vandalize or spread disinformation on Wikipedia because its enormous audience makes it an attractive target.

The social computing literature have documented similar dynamics in a range of other knowledge commons. I have shown that large wikis tends to become increasingly closed as they grow [77, 87]. Other research—by myself and others—suggests that massive influxes of users in Reddit lead moderator teams to institute policies that increasingly restrict participation, ban users, and remove content [47, 51]. These

studies show both how appropriation of some types of peer produced value in a knowledge commons is possible—and how it leads to increasingly closed institutions.

2.1.4. Formal Modeling Informed by the dynamics sketched out in §2.1.2, I will iteratively develop a series of analytic and simulation-based formal models. Although I cannot describe the specifics of the model I will develop, I have conducted work to construct a preliminary dynamical system model. This model treats communities in terms of some stock of value over time (V), and some openness to new contributions (P). It includes two production functions—$g(V,P)$ for "good" value-increasing contributions and $b(V,P)$ for "bad" damaging ones—where both are positively related to stock and openness. These production functions are based on Cobb-Douglas functions from economics that model production as a combination of capital and labor. Exponents in a Cobb-Douglas are elasticity parameters while K is described as baseline efficiency. In this model, openness can be thought of as a classifier that filters for good contributions so that the K parameters capturing the precision and recall of the filter. If one thinks of openness as a classifier for good contributions, K_g reflects the sensitivity of the classifier or the true-positive rate (TPR). Similarly, K_b reflects the false-positive rate (FPR) or one minus the specificity of the classifier. S captures the speed of change in stock and openness so that S_V can be interpreted as a measure of how much harm a damaging contribution will do relative to the good done by a value-increasing one while S_P captures how quickly communities close based on incoming damage. M captures minimum openness. The model includes the following equations:

$$g(V,P) = V^{g_V} P^{g_P} K_g \tag{1}$$

$$b(V,P) = V^{b_V} P^{b_P} K_b \tag{2}$$

$$\frac{dV}{dt} = g(V,P) - b(V,P)S_V \tag{3}$$

$$\frac{dP}{dt} = \frac{M - b(V,P)}{g(V,P)} P S_P \tag{4}$$

This preliminary formal model captures all the dynamics laid out in §2.1.2. A visualization of simulations with a set of reasonable-appearing parameter estimates is shown in Figure 4.[3] The top line in the upper panel of the model seems to reproduce the rise and decline activity dynamics shown in Figures 1 and 2. The bottom two panels capture the basic dynamics of decreased openness as a community builds a larger stock of value.

Although this model is offered as a proof-of-concept, it will serve as a starting point to guide theory development and empirical research. Work as part of this grant will identify more realistic production functions, derive useful ranges for parameters, and identify a more realistic function for openness—e.g., the function above cannot model communities becoming more open after closing. This limitation seems consistent with the studies of openness in knowledge commons [e.g., 24, 77, 87] but is not a logical or theoretical necessity. I also plan to explore moving away from modeling a single stock of value toward separately modeling the different types of value that knowledge commons produce—like knowledge artifacts and goodwill.

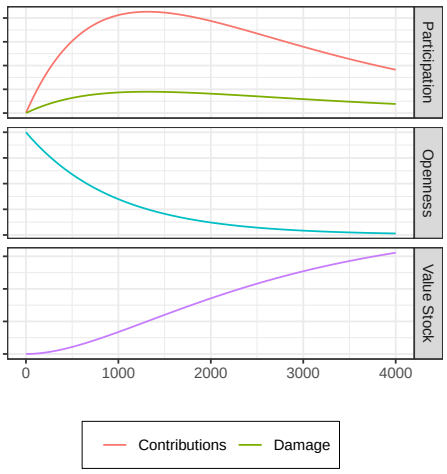

Figure 4: Visualization of a simulation from the preliminary model described above using a set of parameters described in footnote 3. At the end of the simulation, openness is zero and growth of the stock of value ceases completely.

2.1.5. Developing the Model As part of the process of iteratively developing this model, I will conduct two systematic literature reviews. The first review will focus on issues of value in

[3] This model holds all exponents to 0.5. K_g is set to 0.95 and K_b to 0.2. Conservatively, S_V is set to 0.02 suggesting that bad edits are relatively low cost and that S_P is 0.005 suggesting that damaging contributions have only a very small effect on openness. $M = 0$ meaning that communities can become totally closed.

peer production knowledge commons. Questions of value in sociotechnical systems have been an area of some interest in the last several years. For example, the P2PValue project—a mult-institution project funded by a European Commission grant and active in 2014–2016 [66]—has published a series of research projects related to cataloguing different types of value creation and production in online collaborative communities and has engaged in theoretical work to unpack value in these contexts conceptually [e.g., 13, 12]. I will also seek to conduct a systematic review on lifecyles in knowledge commons. Both of these reviews will be conducted following guidance on conducting systematic reviews in information systems from Okoli [62]. Results from the reviews will inform the way that value is operationalized in the model and will also drive decisions about measures and empirical strategies in Part B (§2.2).

The formal model presented in §2.1.4 will be extended to incorporate random chance and also to model interactions between different communities.[4] In doing so, I hope to be able to reproduce important features of peer production—like massive inequality in the number of participants across projects [11, 19, 23, 26, 44, 67]. Once defined, I plan to analyze the formal model and solve for important variables in order to identify important transitions and decision-points, to map the trade-offs implied by my theory, and to identify possible points of intervention. As the model becomes more complex and realistic—and especially as I begin to model interactions between communities—the model will likely become analytically intractable. When this happens, I will seek to understand the model using agent-based simulations.

I plan to present my theoretical framework, as well as the models themselves, in talks, workshops, and the working group meetings (described in §3.2) for feedback. Ultimately, I plan to submit one or more papers to journals in social computing, communication, or organizational science built around the theoretical argument in §2.1.1, presenting the formal models, and explaining how the apparently contradictory collective action and CPR-based explanations for peer production can be integrated into a coherent theoretical framework.

2.2. Part B: Empirical Studies

While the theoretical framework described in §2.1 will guide the research plan, the bulk of effort will be devoted to empirical analyses of online communities that seek to inform, explore, and validate the theory. Before I describe this empirical work, I first introduce the three empirical settings in which it will take place.

2.2.1. Empirical Settings My first empirical setting will be wikis hosted by from the Wikimedia Foundation (WMF) and Wikia. WMF is the non-profit organization that runs Wikipedia and more than 300 others projects including Wiktionary (a dictionary project), WikiQuote (a collection of quotes), Wikimedia Commons (a collection of freely licensed images), and WikiData (a structured database of facts). I have conducted many studies using WMF data [8, 36, 37, 40, 59, 89] and have served on the Wikimedia Foundation Advisory Board (2007–2018). Wikia is a firm hosting public peer production wikis that was founded by Jimmy Wales (Wikipedia's founder). Because many Wikia wikis are about fan culture including films, video games, and so on, Wikia rebranded many of its communities as "Fandom" in 2016. I have conducted several studies using Wikia data [39, 61, 87].

My second empirical setting will be free/libre open source software (FLOSS) development hosted on GitHub and GitLab. GitHub is software development platform owned by Microsoft. The platform is built around the distributed version control system Git and provides infrastructure for managing issues and bugs and workflow for proposing, reviewing, and accepting or rejecting code contributions. Although GitHub is used for private software development, it also hosts many of the largest freely licensed FLOSS software commons whose data are fully public. GitLab is a popular FLOSS clone of GitHub whose data are also largely public.

My third empirical setting will be collaborative filtering communities in Reddit. The term "collaborative filtering" stems from research on recommender systems and describes a process of multiple agents working together to identify relevant information [70, 88]. Collaborative filtering has been a major site for research

[4] I am co-PI on an NSF-funded project seeking to model interactions between online communities. This other work will inform this part of the model.

into social computing [28, 29, 69, 73]. On Reddit, collaborative filtering occurs when users rate content with positive or negative ratings (i.e., upvoting and downvoting). Like wikis and and FLOSS, Reddit is divided into distinct but overlapping communities in the form of more than two million "subreddits" that are focused around a vast range of topics and subcultures. I have conducted two studies using site-wide Reddit data [19, 45].

2.2.2. Inductive Qualitative Research In addition to the two systematic reviews described in §2.1.5, theory development will be guided by a series of qualitative ethnographic interviews focused on lifecycles in peer production. I will use a stratified non-representative sampling technique to conduct a theoretically-driven sample of diverse individuals from my three empirical contexts [90]. In particular, I will seek to recruit long-term contributors with experience over a full lifecycle, only in early stages, and only in later stages. I will prepare an interview protocol focused on questions related to types of value, questions of damaging behavior, questions about how community manage issues of openness, and questions about how these dynamics shift as communities mature and grow. I recognize that some of these questions have been explored in previous qualitative work conducted in all three of my empirical contexts [e.g. 1, 46, 92]. I will produce interview protocols and papers to complement and build on previous work. I will code and analyze qualitative data using grounded theory as described by Charmaz [10], as I have done in previous work [8, 31, 47, 48, 46].

As is common in inductive qualitative research, I cannot anticipate the results of this work. That said, I am committed to using results in two ways. First, they will inform further iteration of the theoretical model described in §2.1. Second, findings will play a useful role in shaping the quantitative empirical studies described in the subsequent sections.

2.2.3. Empirical Studies Testing Hypotheses Derived from the Model One strength of the formal modeling approach I am proposing is that it can generate hypotheses that can be tested empirically. For example, the sketch of the model presented in §2.1.4 and visualized in Figure 3 identifies several relationships that have not been studied in previous research. I plan to conduct at least three empirical analyses that test for relationships like these implied by the theory.

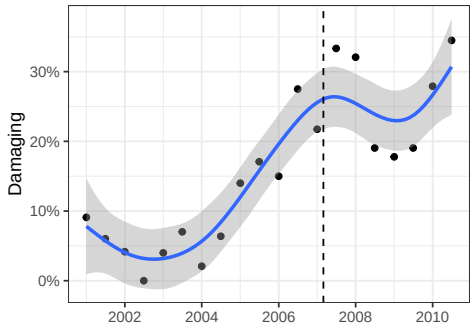

Figure 5: Estimated proportion of damaging edits to English Wikipedia from a hand-coded random sample of edits made within six month periods. New analysis of data originally collected and coded by Halfaker et al. [24].

For example, the model described in §2.1.4 suggests that the proportion of damaging contributions will increase as a knowledge commons' stock of value increases. To my knowledge, this will be the first time that this hypothesis will be tested. To do so as a proof-of-concept for this grant, I conducted a secondary analysis of a hand-coded random sample of edits to English Wikipedia made between 2001 and 2013 from Halfaker et al. [24]. Because Halfaker et al. were interested in the effect of reverts on high quality contributions, they coded edits as damaging or non-damaging. Figure 5 combines this data with data I collected on total contribution volume and then fits a LOESS smoother to estimate of the proportion of edits to Wikipedia that are damaging over time. As my theoretical model predicts, English Wikipedia appears to have experienced an enormous increase in rates of damaging contributions during the period of its rapid growth leading up to 2007.

As a second example, the model in §2.1.4 suggests that the communities should become more closed as vandalism increases. Although there are many ways of measuring openness, one measure in English Wikipedia might include the number of pages that are "protected" so that new users are not allowed to edit them. Using a dataset I produced in a prior study of page protection in Wikipedia [37], I produced Figure 6 which suggests that the number of newly protected pages increases with contributions and damage. Once

again, these results are in line with the model prediction.

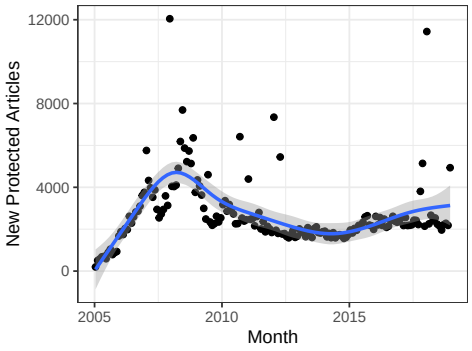

Figure 6: Number of "page protection" in English per month when pages are made uneditable by at least some subset of would-be contributors. New analysis of data from Hill and Shaw [37].

Of course, the analyses in Figures 5 and 6 are presented only to illustrate the basic approach. In work funded by the grant, I will form more precise, nuanced, and theoretically informed hypotheses and will conduct multiple regression, time series, and experimental or quasi-experimental analyses to test them. Positive results from these analysis will serve as evidence in favor of the theory. Null results, surprising findings, or results that are contingent on other variables will all inform modifications that improve the theoretical model.

2.2.4. Empirical Studies Establishing Generalizability Most of the motivating examples and pilot work for this proposal are drawn from Wikipedia and other wikis. Although I believe that Wikipedia and wikis are important, the goal of this work is to to establish a general theory of knowledge commons lifecycles that extends to a range of peer produced goods and settings. Because my theory is at the level of communities, I will conduct population-level analyses in platforms [38]. Although this type of analysis is challenging, I have experience conducting population-level analyses [39, 61, 87, 77].

To establish generalizability, I will attempt to replicate my own findings across from the hypothesis tests described in §2.2.3 across WMF and Wikia, FLOSS projects from GitHub and GitLab, and Reddit. Although rare in HCI research [93], I have published two papers at CHI that conduct exactly this form of replication [55, 87]. I recognize that replication to establish generalizability will involve overcoming several challenges. First, I will need to identify ways to measure the key concepts in the theoretical model across the three settings. How should we measure value? How should we measure appropriation? I will rely on the literature review described in §2.1.5 and the inductive qualitative work in §2.2.2 to create measures of key concepts appropriate to each setting. For example, I might measure appropriation as reverted edits or the proportion of added words that removed over a period of time in wikis (both measures of damaging contributions drawn from earlier research). In GitHub, unmerged or rejected pull requests closed with a particular set of tags might serve as an analogous measure. In Reddit, deleted content or banned users might proxy for the same concept.

In work toward the end of the grant, I hope to fit a full version of a formal model to data using the Bayesian modeling software Stan which can now fit statistical models incorporating algebraic and ordinary differential equations [84]. In this way, I will be able to test not only specific hypotheses drawn from my model but the broader model itself. Of course, because wikis, FLOSS projects, and Reddit discussion forums reflect very different forms of peer produced knowledge commons, I do not expect model dynamics to play out identically in each setting. I hope to identify similarities and differences and to attempt to attribute differences in results to differences in sociotechnical features or affordances. I may also discover fundamental variation that requires updating the underlying theoretical framework.

2.3. Part C: Identify Strategies for Managing Lifecyles

In the final stage of the research, I will seek to propose and test a set of strategies that community managers can use in effectively governing knowledge commons over their lifecycles. I see this playing out in terms of both specific policy or technical decisions made in individual platforms as well as high-level strategies.

2.3.1. Evaluating Particular Interventions Although openness is presented as a single dimension in the model above, openness encompasses many policies and governance strategies. This work can provide insight into the wisdom of specific strategies designed to mitigate and prevent damage. Although I cannot articulate the specific strategies that I hope to evaluate in this way, I can point to a previous study that

demonstrates the type of work I will do.

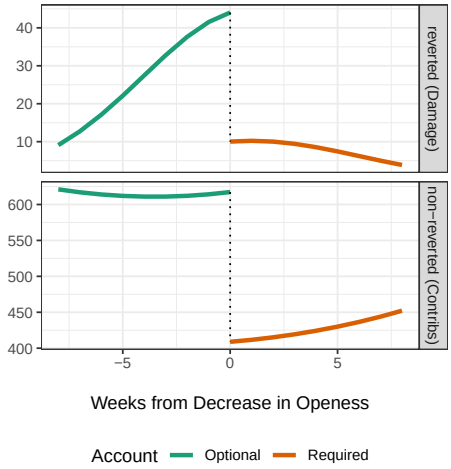

Figure 7: Model predicted average effects of the effect of blocking contributors by users without accounts in 136 wikis. Adapted from Figure 2 in Hill and Shaw [39].

In a recently published study, I examined a policy change conducted in 136 Wikia wikis that decreased openness by instituting a new requirement for would-be participants to create accounts before contributing. Using a quasi-experimental panel regression discontinuity design, I estimated that this decrease in openness was very effective at deterring damage (between 69–83% of bad contributions were prevented) but that it also blocked a significant proportion of good contributions as well (between 21–56%). These proportions mask the severity of the tradeoff: for every damaging contribution rejected, communities blocked more than 6 productive ones. Examples of the average estimated effects are shown in Figure 7 on non-reverted edits (a measure of valuable contributions) in the top panel and reverted edits (a measure of damage) in the bottom.

From this example, one can conclude that this policy would maximize value only when the damage caused by a single bad contribution is at least 6 times that of a single deterred valuable contribution—for example, if repairing a single act of vandalism requires the time needed to make 6 new edits. My analysis similarly suggests that communities can manage openness by reducing the cost of negative contributions from users without accounts. For example, the use of algorithmic triage systems are increasingly common in a range of social computing systems and might reduce the cost of a damaging edit by saving the time and labor [9, 46, 45].

Following this approach, I will evaluate at least one popular strategy in each of my empirical settings. Toward this end, I have already begun collecting examples of changes to design and governance contexts in wikis, FLOSS projects, and subreddits. I will evaluate each strategy *in situ* using quasi-experimental panel discontinuity designs similar to the ones I have used in two previous empirical analyses [39, 61].

In addition to providing specific advice for managers and moderators, analyses of specific strategies can help improve my theoretical model. For example, the analysis of the account requirement in Wikia can provide estimates for K_g and K_b: which capture how much damage a given change in openness will keep out and how much collateral damage will occur. In the simulation I present in Figure 4, the absence of data forces me to rely on intuition in setting these parameters. Given a change in openness within a community, I can compute the TPR and FPR and identify each of values for parameters empirically. In this way, I can update my models with various parameters drawn from real settings and can see how the model predictions are affected.

2.3.2. High-Level Lifecycle Strategy Previous work on organizational closure in peer produced knowledge commons has mostly pointed to increased newcomer rejection as a problem to be solved [e.g., 24]. An example of a high-level strategic takeaway from the preliminary formal mode articulated in §2.1.4 is the suggestion that beginning very open and becoming increasingly closed over time is an optimal strategy for maximizing a stock of value. One might conclude that if communities do not begin with porous boundaries, they will be unlikely to build a large enough stock to kick-off the virtuous cycle dynamics that drive later growth. The model would also suggest that very open communities that manage to grow large stocks will eventually want to reduce their openness in various ways—and at particular points in time—before they are overwhelmed by damage. Because the audience for this type of strategy is practitioners, I will write a practitioner-focused article summarizing these strategies which I will submit for publication in a venue like Harvard Business Review.

2.4. Part D: Creation of Datasets & Tools

In order to conduct the empirical studies above, I will require digital trace data from the three empirical contexts. As described in the data management plan, all of these data are public and distributed in ways that allow for research and reuse. Processing these data into forms that are useful for social computing and social scientific research requires substantial investments of engineering time as well as substantial computing power. I will produce software to process these data into datasets that can be analyzed using common statistical software including Python, R, and Stan.

I have a strong, demonstrated commitment to open science and the public sharing of research data and tools and was awarded the 2019 Research Symbiont Award from the Pacific Symposium of Biocomputing which is "given to a scientist working in any field who has shared data beyond the expectations of their field." To assist other researchers, I will release all the datasets and software used to conduct empirical analyses in replication datasets hosted in the Harvard Dataverse. In addition to replication data, I will produce fully documented dataset and tools that are useful for secondary analysis. All datasets will be released under CC0 permissive licenses and all software will be released under Free Software Foundation approved FLOSS licenses.

3. Education, Outreach, & Dissemination

3.1. Teaching & Mentorship

I will include methods and results from this research directly in the courses I teach at the University of Washington. I regularly teach both undergraduate and graduate classes about online communities as well as graduate methods courses on designing Internet research and conducting data scientific analyses of online communities. I will also design one new undergraduate course and one new graduate course as part of my department's Communication Leadership program—both focused on the topic of managing online community lifecycles. Although I believe these will be the first such classes on this topic anywhere, these courses will be structurally similar to courses about managing lifecycles in firms and non-profits which are routinely taught in management programs. As described in the departmental letter, courses in this area are at the center of the teaching goals of my department and lie at the core of one of the Department's "Technology and Society" area of emphasis. I will incorporate findings, tools, and data from this research into all of these courses.

Additionally, I will be assisted by a graduate student research assistant for the duration of the award. In this way, graduate students will be involved in the project as research assistants, collaborators, and advisees, and will interact with me in research activities and weekly team meetings. I intend to recruit a student to this project who will make part of the work described in this proposal part of their dissertation. I am committed to broadening participation in computing and will actively recruit qualified members of underrepresented groups.

My research lab at the University of Washington is part of an multi-institution research group called Community Data Science Collective (CDSC). I advise graduate students from UW's departments of Communication, Computer Science & Engineering, and Human-Centered Design & Engineering as well as undergraduate research assistants from these departments. Although I have not yet graduated a PhD student, my students have been productive and have received a number of grants, awards, and fellowships. The only postdoc I have mentored is now in a tenure track faculty position at the University of North Carolina. Members of the CDSC have provided inspiration and feedback for on this proposal and will provide a rich and supportive community for conducting the proposed work.

3.2. Community Lifecycle Working Group

To help ground this research in the experience of community managers, I will organize a set of five meetings to bring together the research team and a small group of community managers from the three target empirical settings. I will identify a number of interested parties who will actively engage in the research over time, shape its directions and put the results of this work into action. The working group will be

closely modeled on the MIT Innovation Lab which I helped coordinate as a graduate student at MIT. The goal of the working group will be to (a) disseminate findings from the research described in §2, (b) receive feedback and evaluation of the work as it progresses, and (c) cultivate relationships with communities to facilitate access and impact. I will work with members to identify novel community governance strategies that we can test in partnership.

I will recruit participants by reaching out to a range of individuals and organizations within my research group and collaborators' existing professional networks. Based on presentations I have given in the last year, I have identified members of research and product teams at Wikipedia, Fandom/Wikia, GitHub, as well as Reddit moderators who have all expressed interest in participating in meetings on community lifecycles. If necessary, I will advertise the working group on social media to recruit additional and more diverse participants. I hope to recruit individuals with background in industry, non-profit organizations, public and academia. I have budgeted funds to support a 1.5 day workshop at the University of Washington in each year of the grant. This budget will cover travel and accommodation for 10 non-UW participants.

3.3. Public Outreach Workshops

Finally, I will integrate findings from these results into a set of two large-scale (150 participants) public computing workshops conducted during the last two years of the grant. The workshops will be closely modeled after a series of workshops called the Community Data Science Workshops (CDSW) that I have designed and run 6 times since 2014 for more than 500 participants—a majority of whom are women [35]. The CDSW curriculum is generally focused on the social media analytics skills necessary to use data from social media sites like Twitter. By contrast, the goal of these workshops will be to provide community managers with the skills to manage online community lifecycles.

Instead of the CDSW's curriculum based on studying Twitter and Yelp, the new workshops will teach skills related to three empirical settings described in §2.2.1. The curriculum will focus on providing participants with the ability to build time-series datasets of measures of activity so that participants can observe lifecycle patterns in their own communities—like those in Figure 1. It will also focus on helping communities quantitatively evaluate interventions, like those conducted using A/B tests with automated moderation tools and bots [46, 45, 56, 54]. As I have consistently done in the CDSW, complete curricular materials including example code, documentation, lecture recordings, and so on, will be made freely available online under free software and free culture licenses.

Like the CDSW, the lifecycle-focused workshops will be free of charge, informal, and open to the public. I plan to run these outreach workshops in Seattle once during the final two years of the grant. I will advertise the workshops on local meetup groups and mailing lists. Given my experience with the CDSW and the large network of enthusiastic CDSW alumni, I am confident the workshops will be oversubscribed. Because these workshop will be run by volunteers at UW, I will not need additional funding to conduct the workshops.

3.4. Dissemination

I will disseminate the outcomes of this project through scholarly communication channels including conferences, workshops, and journals. I have a history of publication in communication journals such as the *Journal of Communication* and *Communication Research* as well as computer science and human-computer interaction venues including *Computer-Supported Cooperative Work* (CSCW), *Human-Computer Interaction* (CHI), the *International Conference on Weblogs and Social Media* (ICWSM), and the *International Symposium on Open Collaboration* (OpenSym, formerly WikiSym). Whenever possible, I will release public and freely licensed versions of our research papers. Given the highly integrated nature of the ideas in this proposal and the strong role that a single theory plays in motivating the work, I will use support from this award to write a book to serve as capstone for the project. I plan to complete and submit a book proposal in the final year of the grant.

As I have done in my previous research, I will work closely with organizations supporting online commu-

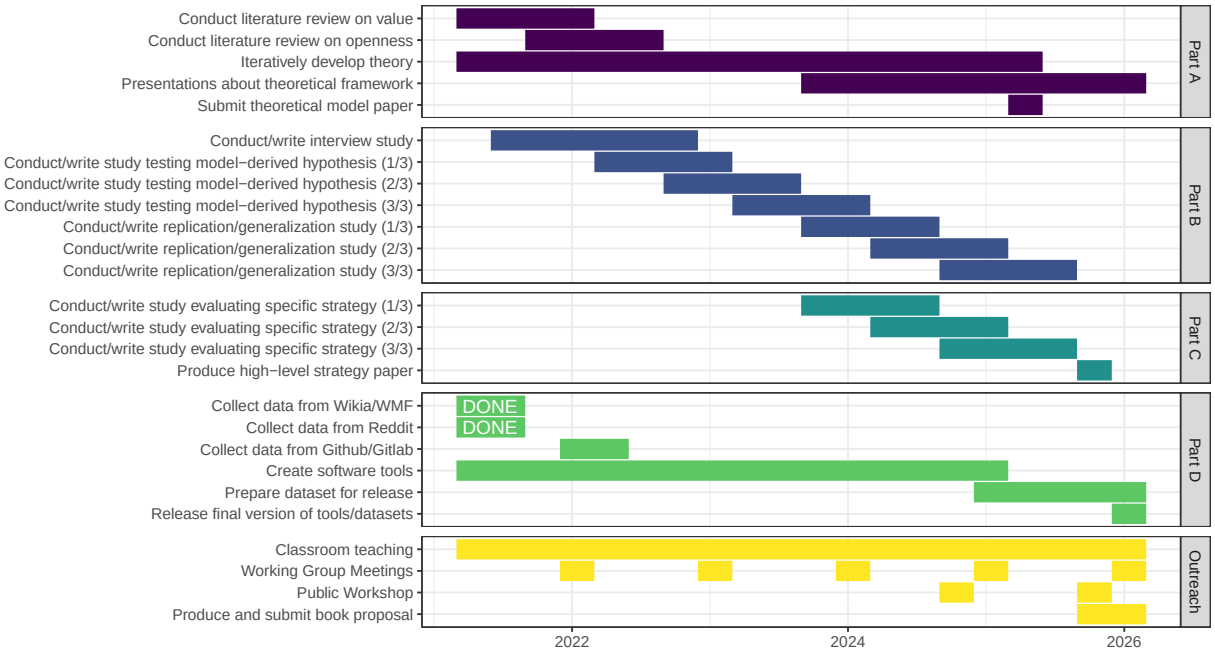

Figure 8: Project timeline assuming a spring 2021 starting date.

nities in ways that go beyond simply using these communities as a source of data. For example, I am an active contributor to Wikia and Wikipedia and was a member of the WMF advisory board during its entire period of activity. I have delivered an annual talk on academic research at Wikipedia's yearly International conference and co-chaired the research track at the conference in 2019. I have given talks on my research at WMF, Wikia, and GitHub. I am a founding board member of the Wikimedia Cascadia User Group, a regular attendee and organizer of events related to Wikipedia in the Seattle area, a former Board Member of the Free Software Foundation, and leader in the Debian and Ubuntu FLOSS communities. I have been a participant in Reddit since the year it was founded and have personal relationships with its co-founders. I will use my relationships with the peer production contributor and business communities to disseminate my research and to ensure that it is relevant and useful.

I will also disseminate the findings from this research to the public more broadly through my my research group's blog and social media channels and through public talks. The questions that motivate this project—how to mobilize organizations and communities to engage in effective collaboration, knowledge sharing, and teamwork—speak to a wide variety of contexts and publics. I will engage in multiple outreach activities to ensure that managers, designers, and leaders of other systems pursuing large-scale online collaboration have access to my findings. I am a Faculty Associate at the Berkman Klein Center for Internet and Society at Harvard University. In these capacities, I frequently speak at industry events and public forums and engage in advisory activities with firms and non-profits. I will leverage these opportunities to disseminate findings from the research I propose here.

4. Project Timeline

The timeline in Figure 8 describes the staging of the research and outreach activity.

5. Intellectual Merit

The intellectual merit of this work lies in three areas. First, the work will develop a theory of why knowledge commons become increasingly closed and how this drives lifecycle dynamics of rise and decline that occur frequently in these settings. In doing so, the work will advance our understanding of the relationship

between collective action, public goods, and common pool resource governance. Second, the work will contribute specific empirical insights into the sociotechnical dynamics of peer production projects. Third, the work will develop a range of new computational social scientific techniques, measures, and methods for understanding online community behavior, lifecycles, and governance.

6. Broader Impacts

The broader impacts of this work lie in the fact that peer produced knowledge commons make up critical information infrastructure. By providing insight into how to better support the development and the maintenance of knowledge commons like wikis, open source software, and collaborative filtering, this work seeks to directly impact millions of contributors to these knowledge commons and the firms and non-profit organizations that support them by identifying novel strategies for community governance. Managers of knowledge commons can use these strategies to navigate tradeoffs between openness and closure across their communities' lifecycles. By better supporting the work of peer production organizations, the broadest impacts of this work are the indirect effects it will have on nearly all Internet users who rely on peer produced software and information to conduct their business and personal lives.

7. PI Preparation

I am an Assistant Professor of Communication and an Adjunct Assistant Professor in the Departments of Computer Science & Engineering and Human Centered Design & Engineering at the University of Washington. I am also a Faculty Associate of the Berkman Klein Center for Internet and Society and an affiliate at the Institute for Quantitative Social Science at Harvard University. I hold a Masters degree from the MIT Media Lab and a Ph.D. from MIT in Management and Media Arts and Science from an interdepartmental program overseen by HCI faculty at the MIT Media Lab and social science faculty at the MIT Sloan School of Management. I have published numerous articles in peer reviewed journals and conference proceedings and have received awards from the International Communication Association, the ACM Conferences on Computer Supported Cooperative Work (CSCW) and Human Factors in Computing Systems (CHI), the Pacific Symposium on Biocomputing, MTV, and Cisco.

Prior to my graduate studies, I worked full time as a software engineer and received my masters degree from MIT for software development and HCI research. I have a background in technology management, data-driven statistical analyses of online communities, computational research, management science, and peer production. Additionally, I have been a leader, developer, and contributor to the free and open source software for more than a decade as part of the Debian and Ubuntu projects, two of the most popular Linux distributions with millions of users worldwide, and am the author of several best-selling technical books [32, 34, 68]. I have been a member of the Wikipedia community since 2005 in a series of leadership roles.

7.1. Training and Preparatory Work

I possess training in organizational research, statistical analysis, and quantitative methodology and have published peer reviewed research employing large-scale, empirical data analysis techniques needed to carry out this work. I have experience conducting systematic literature reviews, inductive qualitative theory building research, and quantitative studies including field experiments, quasi-experimental research designs, and correlational studies.

I have already collected most of the raw data necessary to complete the studies described in this proposal. Although analyzing these data in the way I have described offers a range of challenges, I am confident that I can complete this research with the resources requested. I have already parsed and compiled data on a population of wikis from Wikia which I have used in a number of research projects [39, 61, 77, 87] as well as in several others that are in preparation and under review. I have also fully collected and begun processing data from Reddit as part of work on an ongoing collaborative project related to inter-community ecological dynamic; I have published one study using these data [45] with one additional study under review [19]. I have conducted pilot analyses with a sample of GitHub data from from GHTorrent.

8. Results from Prior NSF Support

Although still early in my career, I have served as a PI or Co-PI on several previous NSF funded research projects. The most closely related award is an active NSF award for a collaborative project with Dr. Aaron Shaw at Northwestern University: "CHS: Small: Collaborative Research: Pathways to Community Success: Advancing a Comparative Science of Online Collaborative Organization" (IIS-1617468 for $194,325 at Northwestern & IIS-1617129 for $305,359 at UW). As part of the work, I created a joint research group called the "Community Data Science Collective" which has, over the period of the award, become a premier research group working on computational studies of online communities. The previous award supported a series of empirical research projects testing three major theories of peer production growth in a population of wikis drawn from Wikia. This included one paper that documented the lifecycles dynamics that form the central puzzle this proposal seeks to answer [87]. This proposal is a direct result of the work in the previous award. The prior award was granted a second one-year no-cost extension and is scheduled to end several months after the proposed award would begin.

8.1. Intellectual Merit

The previous award advanced scientific understanding of collaborative organization by testing three of the most important theories of peer production and validation of this theory through large-scale longitudinal comparison of many peer production systems. The previous work has contributed to knowledge at the intersection of social computing and human collaboration by using organizational theory to draw inference about factors that shape the growth and effectiveness of peer production systems. Near the end of fourth year of the award (the first NCE), the award has resulted in nine peer reviewed papers [15, 17, 22, 39, 46, 48, 59, 61, 87], two peer reviewed poster presentations and short papers [45, 85], three book chapters [41, 33, 38], four datasets [16, 18, 60, 86], one piece of research software [42], and more than a dozen other talks and conference presentations. This work has also resulted in multiple awards.

8.2. Broader Impacts

The broader impacts of the previous award are two-fold. First, it has contributed to actionable insights and novel theoretical approaches that communities, system designers, organizations, and movements engaged in online collaboration can use to achieve their collaborative goals at different stages of their projects. For example, I have shared my work with researchers and managers at companies and non-profit organizations running large online communities. Additionally, I have generated a set of freely licensed and publicly available computational research systems and datasets which other researchers have used in their projects. In both ways, my work has contributed to the design of more effective and more collaborative organizations in online communities, in business, and in society.

References

[1] Judd Antin. "My Kind of People?: Perceptions about Wikipedia Contributors and Their Motivations." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: ACM, 2011, pp. 3411–3420. ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979451. URL: http://doi.acm.org/10.1145/1978942.1979451 (visited on 09/15/2016).

[2] Yochai Benkler. "Coase's Penguin, or, Linux and "The Nature of the Firm"." In: *The Yale Law Journal* 112.3 (Dec. 2002), p. 369. ISSN: 00440094. DOI: 10.2307/1562247. JSTOR: 1562247?origin=crossref.

[3] Yochai Benkler. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press, 2006. 528 pp. ISBN: 0-300-12577-1.

[4] Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill. "Peer Production: A Form of Collective Intelligence." In: *Handbook of Collective Intelligence*. Ed. by Thomas W. Malone and Michael S. Bernstein. Cambridge, MA: MIT Press, 2015, pp. 175–204. ISBN: 978-0-262-02981-0.

[5] David Bollier and Silke Helfrich, eds. *The Wealth of the Commons: A World beyond Market and State*. Amherst, MA: Levellers Press, May 23, 2014. 699 pp. ISBN: 978-1-937146-14-6. Google Books: vcOgAwAAQBAJ.

[6] Brian S. Butler, Elisabeth Joyce, and Jacqueline Pike. "Don't Look Now, but We've Created a Bureaucracy: The Nature and Roles of Policies and Rules in Wikipedia." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. New York, NY, USA: ACM, 2008, pp. 1101–1110. ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357227. URL: http://doi.acm.org/10.1145/1357054.1357227 (visited on 05/09/2015).

[7] Kaylea Champion. "Characterizing Online Vandalism: A Rational Choice Perspective." In: *International Conference on Social Media and Society*. SMSociety'20. Toronto, ON, Canada: Association for Computing Machinery, July 22, 2020, pp. 47–57. ISBN: 978-1-4503-7688-4. DOI: 10.1145/3400806.3400813. URL: https://doi.org/10.1145/3400806.3400813 (visited on 07/02/2020).

[8] Kaylea Champion, Nora McDonald, Stephanie Bankes, Joseph Zhang, Rachel Greenstadt, Andrea Forte, and Benjamin Mako Hill. "A Forensic Qualitative Analysis of Contributions to Wikipedia from Anonymity Seeking Users." In: *Proceedings of the ACM: Human-Computer Interaction* 3 (CSCW Nov. 2019), 531:1–53:26. DOI: 10.1145/3359155.

[9] Eshwar Chandrasekharan, Chaitrali Gandhi, Matthew Wortley Mustelier, and Eric Gilbert. "Crossmod: A Cross-Community Learning-Based System to Assist Reddit Moderators." In: *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW Nov. 7, 2019), 174:1–174:30. DOI: 10.1145/3359276. URL: http://doi.org/10.1145/3359276 (visited on 02/09/2020).

[10] Kathy Charmaz. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. 2nd ed. Thousand Oaks, California: SAGE, 2015. ISBN: 0-7619-7352-4.

[11] Kevin Crowston, Kangning Wei, James Howison, and Andrea Wiggins. "Free/Libre Open-Source Software Development: What We Know and What We Do Not Know." In: *ACM Computing Surveys* 44.2 (Mar. 2008), 7:1–7:35. ISSN: 0360-0300. DOI: 10.1145/2089125.2089127. URL: http://doi.acm.org/10.1145/2089125.2089127 (visited on 02/01/2016).

[12] Primavera De Filippi. "Translating Commons-Based Peer Production Values into Metrics: Toward Commons-Based Cryptocurrencies." In: *Handbook of Digital Currency*. Ed. by David Lee Kuo Chuen. San Diego, CA: Academic Press, Jan. 1, 2015, pp. 463–483. ISBN: 978-0-12-802117-0. DOI: 10.1016/B978-0-12-802117-0.00023-0. URL: http://www.sciencedirect.com/science/article/pii/B9780128021170000230 (visited on 08/05/2020).

[13] Primavera De Filippi and Samer Hassan. "Measuring Value in Commons-Based Ecosystem: Bridging the Gap between the Commons and the Market." In: *The MoneyLab Reader*. Ed. by Geert Lovink and Nathaniel Tkacz. Warwick, UK: Institute of Network Cultures, University of Warwick, 2015. URL: http://papers.ssrn.com/abstract=2725399 (visited on 08/05/2020).

[14] Paul Duguid. "Limits of Self-Organization: Peer Production and "Laws of Quality"." In: *First Monday* 11.10 (Oct. 2, 2006). ISSN: 13960466. URL: http://ojphi.org/ojs/index.php/fm/article/view/1405 (visited on 07/07/2013).

[15] Jeremy Foote and Noshir Contractor. "The Behavior and Network Position of Peer Production Founders." In: *iConference 2018: Transforming Digital Worlds*. Ed. by Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett. Lecture Notes in Computer Science. Springer, 2018, pp. 99–106. ISBN: 978-3-319-78105-1. DOI: 10.1007/978-3-319-78105-1_12.

[16] Jeremy Foote, Darren Gergle, and Aaron Shaw. *Replication Data for: Starting Online Communities: Motivations and Goals of Wiki Founders*. May 12, 2017. DOI: 10.7910/DVN/YG9IID. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/YG9IID (visited on 11/11/2018).

[17] Jeremy Foote, Darren Gergle, and Aaron Shaw. "Starting Online Communities: Motivations and Goals of Wiki Founders." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. New York, NY: ACM, 2017, pp. 6376–6380. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025639. URL: http://doi.acm.org/10.1145/3025453.3025639 (visited on 05/15/2017).

[18] Jeremy Foote, Aaron Shaw, and Benjamin Mako Hill. *Replication Data for: A Computational Analysis of Social Media Scholarship*. Jan. 22, 2018. DOI: 10.7910/DVN/W31PH5. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/W31PH5 (visited on 11/11/2018).

[19] Jeremy Foote, Nathan TeBlunthuis, Benjamin Mako Hill, and Aaron Shaw. "How Individual Behaviors Drive Inequality in Online Community Sizes: An Agent-Based Simulation." In: (June 4, 2020). arXiv: 2006.03119 [cs]. URL: http://arxiv.org/abs/2006.03119 (visited on 06/09/2020).

[20] Seth Frey, P. M. Krafft, and Brian C. Keegan. ""This Place Does What It Was Built for": Designing Digital Institutions for Participatory Change." In: *Proc. ACM Hum.-Comput. Interact.* 3 (CSCW Nov. 2019), 32:1–32:31. ISSN: 2573-0142. DOI: 10.1145/3359134. URL: http://doi.acm.org/10.1145/3359134 (visited on 11/13/2019).

[21] Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg, eds. *Governing Knowledge Commons*. 1 edition. Oxford ; New York: Oxford University Press, Sept. 2, 2014. 520 pp. ISBN: 978-0-19-022582-7.

[22] Emilia F. Gan, Benjamin Mako Hill, and Sayamindu Dasgupta. "Gender, Feedback, and Learners' Decisions to Share Their Creative Computing Projects." In: *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW 2018), 54:1–54:23. DOI: 10.1145/3274323.

[23] John David Nicholas Graves. "Open Source Software Development as a Complex System." Thesis. Auckland University of Technology, 2013. URL: http://aut.researchgateway.ac.nz/handle/10292/5729 (visited on 11/01/2017).

[24] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. "The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline." In: *American Behavioral Scientist* 57.5 (May 1, 2013), pp. 664–688. ISSN: 0002-7642. DOI: 10.1177/0002764212469365. URL: https://doi.org/10.1177/0002764212469365.

[25] Garrett Hardin. "The Tragedy of the Commons." In: *Science* 162.3859 (Dec. 13, 1968), pp. 1243–1248. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.162.3859.1243. pmid: 5699198. URL: https://science.sciencemag.org/content/162/3859/1243 (visited on 08/03/2020).

[26] Kieran Healy and Alan Schussman. "The Ecology of Open-Source Software Development." Working Paper. Working Paper. 2003. URL: http://kieranhealy.org/files/drafts/oss-activity.pdf (visited on 10/28/2016).

[27] Burt Helm. "Wikipedia: "A Work in Progress"." In: *Business Week Online* (Dec. 13, 2005). URL: http://www.businessweek.com/stories/2005-12-13/wikipedia-a-work-in-progress (visited on 10/16/2013).

[28] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. "Explaining Collaborative Filtering Recommendations." In: *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*. CSCW '00. New York, NY, USA: Association for Computing Machinery, Dec. 1, 2000, pp. 241–250. ISBN: 978-1-58113-222-9. DOI: 10.1145/358916.358995. URL: http://doi.org/10.1145/358916.358995 (visited on 08/05/2020).

[29] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. "Evaluating Collaborative Filtering Recommender Systems." In: *ACM Transactions on Information Systems* 22.1 (Jan. 1, 2004), pp. 5–53. ISSN: 1046-8188. DOI: 10.1145/963770.963772. URL: http://doi.org/10.1145/963770.963772 (visited on 08/06/2020).

[30] Charlotte Hess and Elinor Ostrom, eds. *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, MA: The MIT Press, 2011. 382 pp. ISBN: 0-262-51603-9.

[31] Benjamin Mako Hill. "Almost Wikipedia: What Eight Early Online Collaborative Encyclopedia Projects Reveal about the Mechanisms of Collective Action." In: *Essays on Volunteer Mobilization in Peer Production*. Cambridge, Massachusetts: Massachusetts Institute of Technology, 2013.

[32] Benjamin Mako Hill. *Debian GNU/Linux 3.1 Bible*. In collab. with David B Harris. Indianapolis, Ind: Wiley Pub, 2005.

[33] Benjamin Mako Hill. "Whither Peer Production." In: *Decentralizing the Commons*. Ed. by Samer Hassan and Primavera De Felippi. Amsterdam, The Netherlands: Institute for Network Culture, 2018.

[34] Benjamin Mako Hill, Corey Burger, Jonathan Jesse, and Jono Bacon. *Official Ubuntu Book*. 3rd ed. Prentice Hall, 2008. ISBN: 0-13-713668-4.

[35] Benjamin Mako Hill, Dharma Dailey, Richard T. Guy, Ben Lewis, Mika Matsuzaki, and Jonathan T. Morgan. "Democratizing Data Science: The Community Data Science Workshops and Classes." In: *Big Data Factories: Collaborative Approaches*. Computational Social Sciences. Berlin, Germany: Springer Nature, 2017, pp. 115–135. ISBN: 978-3-319-59185-8. DOI: 10.1007/978-3-319-59186-5_9. URL: https://link.springer.com/chapter/10.1007/978-3-319-59186-5_9 (visited on 06/24/2018).

[36] Benjamin Mako Hill and Aaron Shaw. "Consider the Redirect: A Missing Dimension of Wikipedia Research." In: *Proceedings of The International Symposium on Open Collaboration*. OpenSym '14. New York, NY, USA: ACM, 2014, 28:1–28:4. ISBN: 978-1-4503-3016-9. DOI: 10.1145/2641580.2641616. URL: http://doi.acm.org/10.1145/2641580.2641616 (visited on 12/28/2014).

[37] Benjamin Mako Hill and Aaron Shaw. "Page Protection: Another Missing Dimension of Wikipedia Research." In: *Proceedings of the 11th International Symposium on Open Collaboration*. OpenSym '15. New York, NY, USA: ACM, 2015, 15:1–15:4. ISBN: 978-1-4503-3666-6. DOI: 10.1145/2788993.2789846. URL: http://doi.acm.org/10.1145/2788993.2789846 (visited on 10/02/2015).

[38] Benjamin Mako Hill and Aaron Shaw. "Studying Populations of Online Communities." In: *The Oxford Handbook of Networked Communication*. Ed. by Brooke Foucault Welles and Sandra González-Bailón. Oxford, UK: Oxford University Press, Sept. 2019, pp. 173–193. ISBN: 978-0-19-046051-8. URL: https://doi.org/10.1093/oxfordhb/9780190460518.001.0001.

[39] Benjamin Mako Hill and Aaron Shaw. "The Hidden Costs of Requiring Accounts: Quasi-Experimental Evidence from Peer Production." In: *Communication Research* (May 26, 2020), p. 0093650220910345. ISSN: 0093-6502. DOI: 10.1177/0093650220910345. URL: https://doi.org/10.1177/0093650220910345 (visited on 06/25/2020).

[40] Benjamin Mako Hill and Aaron Shaw. "The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation." In: *PLoS ONE* 8.6 (June 26, 2013), e65782. DOI: 10.1371/journal.pone.0065782. URL: http://dx.doi.org/10.1371/journal.pone.0065782 (visited on 07/09/2013).

[41] Benjamin Mako Hill and Aaron D. Shaw. "The Most Important Laboratory for Social Scientific and Computing Research in History." In: *Wikipedia @ 20: Stories of an Incomplete Revolution*. Ed. by Joseph M. Jr. Reagle and Jackie L. Koerner. Cambridge, Massachusetts: MIT Press, Oct. 2020. ISBN: 978-0-262-53817-6.

[42] Benjamin Mako Hill and Nathan TeBlunthuis. *Mediawiki Dump Tools*. Version a4e60a9f. Sept. 3, 2018. URL: https://code.communitydata.cc/mediawiki_dump_tools.git.

[43] Dariusz Jemielniak. *Common Knowledge?: An Ethnography of Wikipedia*. Stanford, CA: Stanford University, 2014. 312 pp. ISBN: 978-0-8047-8944-8.

[44] Steven L. Johnson, Samer Faraj, and Srinivas Kudaravalli. "Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment." In: *Management Information Systems Quarterly* 38.3 (2014), pp. 795–808. URL: http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3193&context=misq (visited on 04/26/2017).

[45] Charles Kiene and Benjamin Mako Hill. "Who Uses Bots? A Statistical Analysis of Bot Usage in Moderation Teams." In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, Apr. 25, 2020, pp. 1–8. ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3382960. URL: http://doi.org/10.1145/3334480.3382960 (visited on 05/26/2020).

[46] Charles Kiene, Jialun "Aaron" Jiang, and Benjamin Mako Hill. "Technological Frames and User Innovation: Exploring Technological Change in Community Moderation Teams." In: *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW Nov. 7, 2019), 44:1–44:23. DOI: 10.1145/3359146. URL: http://doi.org/10.1145/3359146 (visited on 01/16/2020).

[47] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. "Surviving an "Eternal September": How an Online Community Managed a Surge of Newcomers." In: *Proceedings of the 2016 ACM Conference on Human Factors in Computing Systems (CHI '16)*. New York, NY: ACM, 2016, pp. 1152–1156. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858356. URL: http://doi.acm.org/10.1145/2858036.2858356 (visited on 07/05/2016).

[48] Charles Kiene, Aaron Shaw, and Benjamin Mako Hill. "Managing Organizational Culture in Online Group Mergers." In: *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW 2018), 89:1-89–21. DOI: 10.1145/3274358.

[49] Srijan Kumar, Robert West, and Jure Leskovec. "Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes." In: *Proceedings of the 25th International Conference on World Wide Web*. WWW '16. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, Apr. 11, 2016, pp. 591–602. ISBN: 978-1-4503-4143-1. DOI: 10.1145/2872427.2883085. URL: http://doi.org/10.1145/2872427.2883085 (visited on 08/05/2020).

[50] Carter Landis. *CHAOSS Evolution Working Group*. July 1, 2020. URL: https://github.com/chaoss/wg-evolution (visited on 08/05/2020).

[51] Zhiyuan Lin, Niloufar Salehi, Bowen Yao, Yiqi Chen, and Michael S. Bernstein. "Better When It Was Smaller? Community Content and Behavior after Massive Growth." In: *ICWSM*. 2017, pp. 132–141. URL: http://itsmrlin.com/papers/2017_icwsm_eternal_september.pdf (visited on 08/02/2017).

[52] Max Loubser. "Organisational Mechanisms in Peer Production: The Case of Wikipedia." DPhil. Oxford, United Kingdom: Oxford Internet Institute, Oxford University, July 2010.

[53] Gerald Marwell and Pamela Oliver. *The Critical Mass in Collective Action: A Micro-Social Theory*. Cambridge, UK: Cambridge University Press, 1993. ISBN: 978-0-521-30839-7.

[54] J. Nathan Matias. "Preventing Harassment and Increasing Group Participation through Social Norms in 2,190 Online Science Discussions." In: *Proceedings of the National Academy of Sciences* 116.20 (May 14, 2019), pp. 9785–9789. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1813486116. pmid: 31036646. URL: https://www.pnas.org/content/116/20/9785 (visited on 07/30/2019).

[55] J. Nathan Matias, Sayamindu Dasgupta, and Benjamin Mako Hill. "Skill Progression in Scratch Revisited." In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. New York, NY, USA: ACM, 2016, pp. 1486–1490. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858349. URL: http://doi.acm.org/10.1145/2858036.2858349 (visited on 07/25/2017).

[56] J. Nathan Matias and Merry Mou. "CivilServant: Community-Led Experiments in Platform Governance." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: Association for Computing Machinery, Apr. 19, 2018, pp. 1–13. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173583. URL: http://doi.org/10.1145/3173574.3173583 (visited on 08/06/2020).

[57] Mohamad Mehdi, Chitu Okoli, Mostafa Mesgari, Finn Årup Nielsen, and Arto Lanamäki. "Excavating the Mother Lode of Human-Generated Text: A Systematic Review of Research That Uses the Wikipedia Corpus." In: *Information Processing & Management* 53.2 (Mar. 1, 2017), pp. 505–529. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2016.07.003. URL: http://www.sciencedirect.com/science/article/pii/S0306457316303004 (visited on 07/18/2018).

[58] Natalia Mielczarek. "Fake Online Biography Created as as a 'Joke'." In: *The Tennessean* (Dec. 11, 2005). URL: http://www.tennessean.com/apps/pbcs.dll/article?AID=/20051211/NEWS01/512110366/1006/NEWS.

[59] Sneha Narayan, Jake Orlowitz, Jonathan Morgan, Benjamin Mako Hill, and Aaron Shaw. "The Wikipedia Adventure: Field Evaluation of an Interactive Tutorial for New Users." In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. CSCW '17. New York, NY, USA: ACM, 2017, pp. 1785–1799. ISBN: 978-1-4503-4335-0. DOI: 10.1145/2998181.2998307. URL: http://doi.acm.org/10.1145/2998181.2998307 (visited on 03/21/2017).

[60] Sneha Narayan, Jake Orlowitz, Jonathan T. Morgan, Aaron D. Shaw, and Benjamin Mako Hill. *Replication Data for: The Wikipedia Adventure: Field Evaluation of an Interactive Tutorial for New Users*. June 7, 2017. DOI: 10.7910/DVN/6HPRIG. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6HPRIG (visited on 11/11/2018).

[61] Sneha Narayan, Nathan TeBlunthuis, Wm Salt Hale, Benjamin Mako Hill, and Aaron Shaw. "All Talk: How Increasing Interpersonal Communication on Wikis May Not Enhance Productivity." In: *Proceedings of the ACM: Human-Computer Interaction* 3 (CSCW Nov. 2019), 101:1–101:19. DOI: 10.1145/3359203.

[62] Chitu Okoli. "A Guide to Conducting a Standalone Systematic Literature Review." In: *Communications of the Association for Information Systems* 37 (2015). ISSN: 15293181. DOI: 10.17705/1CAIS.03743. URL: https://aisel.aisnet.org/cais/vol37/iss1/43/ (visited on 05/17/2019).

[63] Mancur Olson. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press, 1965.

[64] Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York, NY: Cambridge University Press, 1990.

[65] Vincent Ostrom and Elinor Ostrom. "Public Goods and Public Choices." In: *Alternatives For Delivering Public Services: Toward Improved Performance*. Ed. by Emanuel S. Savas. Boulder, CO: Westview Press, 1977, pp. 7–49.

[66] *P2Pvalue Blog*. Aug. 6, 2016. URL: https://p2pvalue.eu/ (visited on 08/05/2020).

[67] Elliot Panek, Connor Hollenbach, Jinjie Yang, and Tyler Rhodes. "The Effects of Group Size and Time on the Formation of Online Communities: Evidence from Reddit." In: *Social Media + Society* 4.4 (Oct. 1, 2018), p. 2056305118815908. ISSN: 2056-3051. DOI: 10.1177/2056305118815908. URL: https://doi.org/10.1177/2056305118815908 (visited on 06/01/2020).

[68] Kyle Rankin and Benjamin Mako Hill. *The Official Ubuntu Server Book*. Upper Saddle River, NJ: Prentice Hall, 2009. ISBN: 978-0-13-702118-5.

[69] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. "Grouplens: An Open Architecture for Collaborative Filtering of Netnews." In: *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*. CSCW '94. New York, NY, USA: ACM, 1994, pp. 175–186. ISBN: 978-0-89791-689-9. DOI: 10.1145/192844.192905. URL: http://doi.acm.org/10.1145/192844.192905 (visited on 07/19/2016).

[70] Francesco Ricci, Lior Rokach, and Bracha Shapira. "Introduction." In: *Recommender Systems Handbook*. Ed. by Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. Boston, MA: Springer US, 2011, pp. 1–35. ISBN: 978-0-387-85820-3. DOI: 10.1007/978-0-387-85820-3_1. URL: https://doi.org/10.1007/978-0-387-85820-3_1 (visited on 08/06/2020).

[71] Emiel Rijshouwer. *Organizing Democracy: Power Concentration and Self-Organization in the Evolution of Wikipedia*. Rotterdam: Erasmus University, Jan. 11, 2019. ISBN: 978-94-028-1371-5. URL: https://repub.eur.nl/pub/113937/ (visited on 08/31/2019).

[72] Diego Saez-Trumper. "Online Disinformation and the Role of Wikipedia." In: (Oct. 14, 2019). arXiv: 1910.12596 [cs]. URL: http://arxiv.org/abs/1910.12596 (visited on 08/06/2020).

[73] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative Filtering Recommender Systems." In: *The Adaptive Web: Methods and Strategies of Web Personalization*. Ed. by Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2007, pp. 291–324. ISBN: 978-3-540-72079-9. DOI: 10.1007/978-3-540-72079-9_9. URL: https://doi.org/10.1007/978-3-540-72079-9_9 (visited on 08/06/2020).

[74] Charles M. Schweik and Robert C. English. *Internet Success: A Study of Open-Source Software Commons*. Cambridge, MA: MIT Press, 2012. 351 pp. ISBN: 978-0-262-01725-1.

[75] Holger Schwenk, Vishrav Chaudhary, Shuo Sun, Hongyu Gong, and Francisco Guzmán. "WikiMatrix: Mining 135M Parallel Sentences in 1620 Language Pairs from Wikipedia." In: (July 15, 2019). arXiv: 1907.05791 [cs]. URL: http://arxiv.org/abs/1907.05791 (visited on 08/05/2020).

[76] John Seigenthaler. "A False Wikipedia 'Biography'." In: *USA Today* (Nov. 30, 2005). URL: https://usatoday30.usatoday.com/news/opinion/editorials/2005-11-29-wikipedia-edit_x.htm (visited on 08/31/2019).

[77] Aaron Shaw and Benjamin Mako Hill. "Laboratories of Oligarchy? How the Iron Law Extends to Peer Production." In: *Journal of Communication* 64.2 (2014), pp. 215–238. ISSN: 1460-2466. DOI: 10.1111/jcom.12082. URL: http://onlinelibrary.wiley.com/doi/10.1111/jcom.12082/abstract (visited on 05/09/2015).

[78] Clay. Shirky. *Here Comes Everybody : The Power of Organizing without Organizations*. New York, NY: Penguin Press, 2008. ISBN: 978-1-59420-153-0.

[79] Ben Shneiderman, Jennifer Preece, and Peter Pirolli. "Realizing the Value of Social Media Requires Innovative Computing Research." In: *Communications of the ACM* 54.9 (Sept. 1, 2011), pp. 34–37. ISSN: 0001-0782. DOI: 10.1145/1995376.1995389. URL: http://doi.org/10.1145/1995376.1995389 (visited on 08/06/2020).

[80] M. Six Silberman. "Reading Elinor Ostrom In Silicon Valley: Exploring Institutional Diversity on the Internet." In: *Proceedings of the 19th International Conference on Supporting Group Work*. GROUP '16. New York, NY, USA: Association for Computing Machinery, Nov. 13, 2016, pp. 363–368. ISBN: 978-1-4503-4276-6. DOI: 10.1145/2957276.2957311. URL: http://doi.org/10.1145/2957276.2957311 (visited on 08/06/2020).

[81] Tom Simonite. "Inside the Alexa-Friendly World of Wikidata." In: *Wired* (Feb. 18, 2019). ISSN: 1059-1028. URL: https://www.wired.com/story/inside-the-alexa-friendly-world-of-wikidata/ (visited on 08/05/2020).

[82] *Siteviews Analysis*. Aug. 5, 2020. URL: https://pageviews.toolforge.org/siteviews/?platform=all-sites&source=unique-devices&start=2019-08&end=2020-07&sites=en.wikipedia.org (visited on 08/05/2020).

[83] Bongwon Suh, Gregorio Convertino, Ed H. Chi, and Peter Pirolli. "The Singularity Is Not near: Slowing Growth of Wikipedia." In: *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. WikiSym '09. New York, NY: ACM, 2009, pp. 1–10. ISBN: 978-1-60558-730-1. DOI: 10.1145/1641309.1641322. URL: http://doi.acm.org/10.1145/1641309.1641322 (visited on 04/21/2016).

[84] Stan Development Team. *Algebraic Equation Solver*. 2020. URL: https://mc-stan.org/docs/2_24/functions-reference/index.html (visited on 08/06/2020).

[85] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. "Density Dependence without Resource Partitioning: Population Ecology on Change.Org." In: *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. CSCW '17 Companion. New York, NY, USA: ACM, 2017, pp. 323–326. ISBN: 978-1-4503-4688-7. DOI: 10.1145/3022198.3026358. URL: http://doi.acm.org/10.1145/3022198.3026358 (visited on 03/30/2018).

[86] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. *Replication Data for Revisiting 'The Rise and Decline' in a Population of Peer Production Projects"*. 2018. DOI: 10.7910/DVN/SG3LP1. URL: https://doi.org/10.7910/DVN/SG3LP1 (visited on 05/03/2020).

[87] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. "Revisiting "The Rise and Decline" in a Population of Peer Production Projects." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. New York, NY: ACM, 2018, 355:1–355:7. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173929. URL: http://doi.acm.org/10.1145/3173574.3173929 (visited on 07/01/2018).

[88] Loren Terveen and Will Hill. "Beyond Recommender Systems: Helping People Help Each Other." In: *HCI in the New Millennium*. Ed. by Jack Carroll. New York, NY: Addison-Wesley, 2001, pp. 487–509.

[89] Chau Tran, Kaylea Champion, Andrea Forte, Benjamin Mako Hill, and Rachel Greenstadt. "Tor Users Contributing to Wikipedia: Just Like Everybody Else?" In: *Conference Proceedings, IEEE Security and Privacy (Oakland 2020)*. IEEE Security and Privacy 2020. Vol. 1. 2020, pp. 974–990. DOI: https://doi.ieeecomputersociety.org/10.1109/SP40000.2020.00053.

[90] Jan E. Trost. "Statistically Nonrepresentative Stratified Sampling: A Sampling Technique for Qualitative Studies." In: *Qualitative Sociology* 9.1 (Mar. 1, 1986), pp. 54–57. ISSN: 1573-7837. DOI: 10.1007/BF00988249. URL: https://doi.org/10.1007/BF00988249 (visited on 03/18/2019).

[91] Burak Uzkent, Evan Sheehan, Chenlin Meng, Zhongyi Tang, Marshall Burke, David Lobell, and Stefano Ermon. "Learning to Interpret Satellite Images in Global Scale Using Wikipedia." In: (Aug. 11, 2019). arXiv: 1905.02506 [cs]. URL: http://arxiv.org/abs/1905.02506 (visited on 08/05/2020).

[92] Joel West and Siobhán O'Mahony. "The Role of Participation Architecture in Growing Sponsored Open Source Communities." In: *Industry and Innovation* 15.2 (Apr. 1, 2008), pp. 145–168. ISSN: 1366-2716. DOI: 10.1080/13662710801970142. URL: https://doi.org/10.1080/13662710801970142 (visited on 08/06/2020).

[93] Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. "RepliCHI - CHI Should Be Replicating and Validating Results More: Discuss." In: *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '11. New York, NY: ACM, 2011, pp. 463–466. ISBN: 978-1-4503-0268-5. DOI: 10.1145/1979742.1979491. URL: http://doi.acm.org/10.1145/1979742.1979491 (visited on 07/25/2017).