CHS: Small: Collaborative Research: Modeling the Ecological Dynamics of Online Organizations

Would Wikipedia be one of the most visited websites in the world if other online collaborative encyclopedia projects had been more established when it was founded? Or was Wikipedia helped by the fact that its predecessors had engaged and trained hundreds of its future contributors? Do new discussion communities on Reddit compete with existing communities for contributors? Is the evolving world of online communities better understood as a competitive struggle for resources or as symbiotic relationships that support a web of interdependent communities? How does the environment of existing online communities shape the growth, performance, and impact of new groups?

Answering these questions requires an *ecological understanding* of online communities that accounts for the complex dynamic interactions between communities and their environments. Established approaches to the comparative study of online community success have almost exclusively looked inside communities [e.g., 14, 27, 36, 37, 46]. These introspective approaches typically account for only a small amount of variation in communities' growth, longevity, and performance. Ecology provides a compelling alternative approach. In biology and organization studies, ecological approaches have shown that success is largely—and sometimes overwhelming—a function of what others in an individual's environment are doing [15, 51]. Analyses of ecological factors in the life sciences enable effective wildlife management, pest control, and sustainable utilization of renewable resources. In sociology, they provide compelling explanations for industrial life-cycles, organizational specialization, and patterns of collaborative partnerships.

Recent research from the social computing literature on interdependence between online communities [26, 43, 45, 50, 52] suggests that ecological analyses can provide not only novel scientific understandings but also viable community management strategies. Our goal is to develop an ecological theory of online communities through three linked projects answering fundamental ecological questions: (1) How is the growth and survival of individual online communities shaped by activity within their environments? (2) How do communities interact with each other in ways that are beneficial or harmful to their success? (3) How can communities strategically discover and build on opportunities for growth and productivity in their environments?

We propose to apply our approach to three popular online platforms for voluntary organization and open collaboration: Wikia, Reddit, and Stack Exchange. For each platform, we will draw on ecological theory to derive predictions about how a community's success relates to the presence, absence, and behavior of other communities and to variation in external resources in the form of users and underlying interest in topics. We will test these predictions empirically using large scale statistical analysis of longitudinal digital trace data drawn from our platforms. Finally, we will synthesize these results into a cohesive theory of ecology for online communities. In the course of this work we will also design, build, and publish software and unique research datasets for other researchers to use.

1. Project Goals

Online communities are a dynamic, growing, and increasingly important form of organization. Through peer production, the Wikipedia community has produced the largest collaborative effort and most important reference work in human history. Free/libre open source software (FLOSS) communities have produced tens of billions of dollars worth of software made freely available online [1]. Other online communities like Reddit provide information, social support, and entertainment to millions of people. Online community platforms support millions of attempts to build communities [23, 36], but only a tiny percentage manage to mobilize participants and to sustain collaboration [36, 38].

Prior studies of the growth, survival, and success of online communities have focused almost exclusively on communities' internal features [27] and have largely neglected environmental factors. In contrast, this project proposes extending prior research on online communities by drawing from the ecological study of biological populations [51] and organizations [15]. No competent wildlife biologist predicting the survival of an animal population would only consider the species' physiology in isolation. Instead, they would

1910202

consider the presence of necessary resources like food and shelter, the presence or absence of other organisms, and competitive dynamics that might give the species an advantage. We argue that these same sorts of ecological dynamics are important drivers of online community outcomes, and that neglecting them is one reason that researchers have had difficulty in understanding or predicting outcomes. As researchers of online communities, we believe that more rigorous steps in toward ecological analysis are overdue.

Our overarching goal is to transform current understandings of the conditions for successful online organization through an ecological analysis of the dynamic interactions between communities and their environments. We propose to adapt conceptual models and analytic methods developed by biologists and sociologists in order to study the growth and survival of online groups. We will carefully evaluate the ways that the ecological dynamics that drive online communities' successes may diverge from ecological dynamics in other domains. In doing so, we hope to provide new conceptual, analytical, and computational tools to produce actionable insights for community leaders and platform designers.

2. Ecological Studies of Online Organizations

Ecology is a scientific approach to studying how interdependence between individuals, collectives, and environments shapes the world [51]. Although first developed to understand biological ecosystems, ecology's analytic theories and methods influenced the development of human ecology, and later of organizational ecology [15, 29, 32]. Organizational ecology has been used to explain the success and failure of newspapers, microbreweries, and social movements by modeling the environments in which organizations exist [3, 4, 40].

Recent research analyzing interdependence between online communities illustrates the promise of ecological approaches in this domain [5, 43, 49]. Community outcomes such as growth and survival depend on membership overlaps between communities [50, 52], but the nature of these relationships remains unclear. Wang et al. [50] found that participant overlaps between Usenet groups were associated with competition and decreased participation in both communities. However, Zhu et al. [52] found evidence that membership overlap between wikis is associated with *mutualism* and benefits for both communities. Such contradictory findings point to the need for deeper, more precise theories of how ecological dynamics play out in online communities.

Moreover, many central features of ecological analysis—like ecological relationships and the creation and discovery of ecological niches—remain entirely unexplored in the context of online communities. For example, *commensal* relationships (where one party is helped by a second, but the second is unaffected by the first) and *amensal* relationships (where one party is hurt by an unaffected second) are common in biological systems [48] and it seems likely that they will arise in online communities as well. However, these dynamics have never been studied. Prior studies have not attempted to model online community environments in terms of the resource environment or of dyadic interactions—a first step in most ecological analyses. As a result, the tension between the results of Wang et al. and Zhu et al. suggests that whether participant overlaps coincide with competition or with mutualism may depend on a number of unknown factors. Confusion about whether participant overlaps are a signal of competition or mutualism may be reconcilable with more advanced methods.

We will address these shortcomings by first engaging deeply with ecological research in both biology and organization science from which we will borrow concepts and methods. In the context of online communities, we will define an ecological *population* as the set of communities that share a set of *resources*. In the context of online community research, resources include the labor and intellects of participants, content that they appropriate and produce, as well as the technological and social systems that communities develop to structure themselves like norms, rules, and technologies [2]. In an ecological model, a community must find a *niche*—i.e., a set of resources that it can utilize comparatively better than other communities—in order to survive.

3. Empirical Settings

We will draw ecological populations of online communties from three platforms: Wikia, Reddit, and Stack Exchange. All three platforms host large numbers of communities that engage in collaborative production of information. Each community within the platforms has some freedom to create their own internal structures. Additionally, communities on these platforms have overlapping users and overlapping content—preconditions for ecological processes. That said, the three platforms also have significant technological and cultural differences. Studying all three will allow us to evaluate whether the presence or importance of different ecological dynamics will generalize beyond the context of any individual platform.

Wikia provides a platform for hundreds of thousands of wikis—i.e., communities that collaboratively produce knowledge bases about a specific topic. Wikia was created by Wikipedia founder Jimmy Wales and uses the same software infrastructure as Wikipedia. Many Wikia wikis focus on cultural topics, such as television series or books and operate under the brand Fandom. We have already built and published extensive datasets and software for analyzing Wikia wikis [e.g., 24] and have published several papers using these data [e.g., 8, 10, 37, 45].

The Reddit platform consists of more than a million sub-communities called "subreddits" where participants share, discuss, and create content on a wide variety of topics. Users can share links, write text-based posts, and comment on the submissions of others. Reddit communities almost universally use a collaborative social ranking system that aggregates participants' votes to surface and reward desirable posts and comments. Like Wikia, subreddits are sustained by the contributions of groups of pseudonymous strangers. We have published one paper using data from Reddit communities [25] and have collected comprehensive data on the population of subreddit communities that we have already used in a pilot project [6] described in §4.

Finally, Stack Exchange hosts question and answer (Q&A) communities. The oldest and largest of these communities is Stack Overflow, which is widely used by computer programmers to post, discuss, and resolve programming problems. Beginning in 2010, Stack Exchange allowed users to create new Q&A sites. Unlike Reddit and Wikia, these new communities must go through an elaborate design and vetting process intended to maximize the chances of community success. Compared to Reddit and Wikia, Stack Exchange communities have a relatively complex system for reputation that grants users more advanced capabilities as they make quality contributions. There are currently 171 active Stack Exchange communities.

All of the data that we need to conduct our planned analysis is archived and public. In preparation for the proposed work, we have collected data from Wikia and Reddit in the form of public archival "dumps" that have been published on the platform's websites and in the Internet Archive. We have not yet collected and analyzed data from Stack Exchange but have verified that it has been made available by the platform for research purposes.

4. Preparation and Pilot Work

Although this project is ambitious in both scope and scale, the project team has the experience, skills, and resources to execute all phases of the work. The two PIs have collaborated for nearly a decade on related research involving similar datasets drawn from Wikia and Wikipedia. With prior NSF IIS CHS support, we have built a cross-university research team (the Community Data Science Collective) that has a history of working together and has developed multiple innovative population-level studies of online communities. Our project team also holds the computational and social scientific skills needed to conduct the proposed analyses. Below, we describe three pilot studies which demonstrate the promise and feasibility of our proposal.

First, we have conducted a study that tested the ecological theory of "density dependence" (the concept is described in depth in (§5.1) in the online petitioning platform Change.org. This work has been presented as a peer-reviewed poster at CSCW '17 [45] and published as a masters thesis [44] advised by PI Hill. Although not focused on online communities, the project reflects direct engagement with key ecological theories in a digital context. More importantly, it develops methodological techniques we will build upon in our proposal. In particular, the project pioneers a way of using LDA topic models to establish semantic niches that allow us to measure the number of petitions being created about different topics. Our preliminary results (summarized in Figure 1) suggest that density dependence, and especially the benefits of mutualism, are an important driver of the number of signatures that petitions receive.

The pilot study also speaks to the potential explanatory power of ecological approaches in online contexts. The model predicting petition signatures relies almost exclusively on measures drawn from petition's environments instead of qualities of petitions and their creators. Nevertheless, the adjusted R^2 statistic for our model is 0.16 suggesting that environmental factors explain a meaningful amount of the variation in the signatures that petitions receive. We take this as a promising sign that an ecological approach might explain substantial variations in other kinds of community-level outcomes.

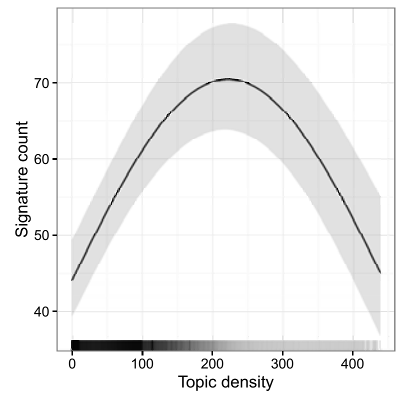

Figure 1: Preliminary results from our study of the online petitioning platform Change.org show evidence of a density dependence dynamic within petition topics. The lines at the bottom of the figure represent the location of petitions on the range of our density measure and show that, for the vast majority of petitions, our model predicts that an increase in density will lead to an increase in signatures.

A second pilot study revisits the extremely unequal distribution of participants across projects in online communities. We begin with a plausible individual-level decision-making theory suggested by prior literature [27]. Using an agent-based model that also takes environmental factors into account, we simulate individual decision-making in the presence of many overlapping communities. The results prove far more consistent with empirically observed distributions of membership in subreddits and imply that ecological forces can help online community research bridge the divide between micro- and macro-levels of explanation.

This work advances our preparation for the current project in two ways. First, work on this study has involved collecting and analyzing comprehensive population-level data from Reddit. Second, our analysis in the pilot project reflects an initial bridge between the large body of intra-community analyses conducted in social computing research and the types of ecological analyses we are proposing. Like previous social computing work [27], our pilot project models users' decisions to engage with individual online communities. However, our project extends this work by modeling how qualities of communities, like the number of existing participants, affect ecological competition between communities. In the last several months, this work was presented at the annual meeting of International Communication Association (ICA '18) [7] and at the International Conference for Computational Social Science (IC2S2 '18) [6].

In a final piece of pilot work, we have conducted exploratory analyses to understand whether different online encyclopedia wikis covering the same topics experience competition or mutualism. We identified the 25 largest English language encyclopedia communities hosted by Wikia. After identifying the topic of each encyclopedia, we then manually identified related categories in English Wikipedia which we used to identify related Wikipedia articles [47]. Figure 2 shows the volume of contributions to the Harry Potter encyclopedia on Wikia and the articles related to Harry Potter in English Wikipedia.

In time series analyses, we find a positive correlation between contribution volume in Wikia encyclopedias and Wikipedia articles on the same topics. This positive relationship holds even when we restrict the analysis to contributions from usernames present in both datasets (a group that accounts for more than a one third of all contributions to both wikis). These preliminary findings suggest the presence of mutualism: when individuals increase their contributions to a topic space on Wikipedia, they tend to increase their contributions to Wikia wikis on the same topics. We cannot rule out alternative explanations for these associations from these preliminary analyses. For example, underlying shifts in attention to a given topic

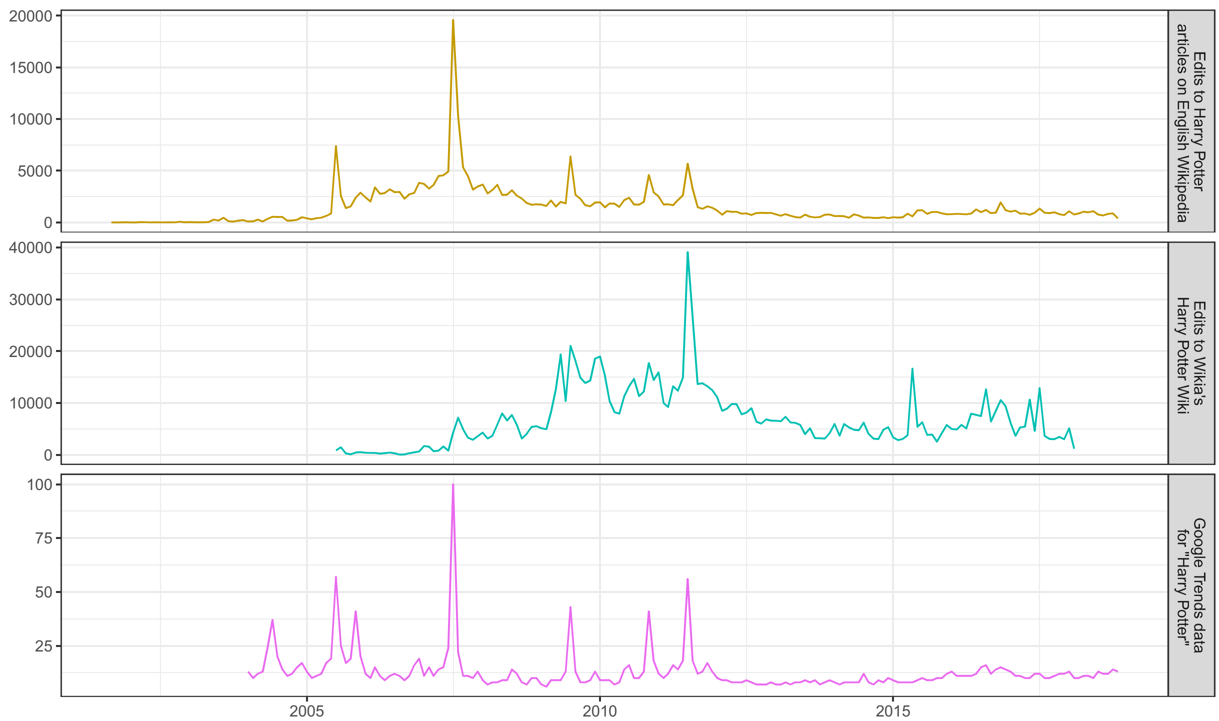

Figure 2: Three time series—each measured at the level of month. The top panel reflects the number of edits made to Harry Potter-related articles on English Wikipedia. The middle panel reflects the number of edits made each month to the largest Harry Potter encyclopedia hosted by Wikia. The bottom panel reflects general search engine volume for the term "harry potter" and is drawn from Google Trends (units are not meaningful).

might drive correlated editing activity. For example, the bottom of panel of Figure 2 shows underlying search engine volume for the term "harry potter" which appears to be strongly correlated to editing in English Wikipedia while diverging from Wikia in important ways. The work proposed below out a plan for disentangling such competing explanations.

5. Project Plan

Drawing on our understanding of ecology and of online communities research, we propose three projects to explain how (A) communities' environments, (B) communities' interactions with each other, and (C) communities' strategic approaches to finding or building a niche can affect their growth and survival. Each project is designed around a research question drawn from ecology:

> **Project A — Ecological environments:** How do online communities' environments shape their growth and survival?
>
> **Project B—Ecological relationships between communities:** How do online communities interact with other communities in competitive, mutalistic, amensal, and commensal dyads?
>
> **Project C—Ecological strategies:** How should online communities choose ecological niches?

Identifying ecological relationships from observational data about online communities presents a methodological challenge that we propose to solve using analytic techniques from prior work in biological ecology. In particular, we will use Sugihara et al.'s [41] convergent cross mapping algorithm to detect whether two communities are a part of the same dynamical system and will fit Lotka-Volterra equations, a family of differential equations that can be used to model and identify ecological dynamics. Combining these two methods will allow us to infer networks of ecological relationships between communities both within a given platform and across platforms. Across all three of the proposed projects, we will use multilevel longitudinal regression to evaluate empirical relationships in the populations of communities that we study [39]. We will employ structural topic modeling to measure topical relationships between communities [34].

We describe the research questions, designs, and anticipated findings for the studies we will carry out in detail in the corresponding subsections of the project plan below. In §5.4, we discuss how the findings from our empirical work will inform an integrated ecological theory of online communities. In §5.5, we elaborate how, in the process of doing this work, we will create and release software tools for building datasets that will help other researchers and practitioners engage in the kind of ecological studies we will conduct.

5.1. Project A: Ecological Environments

What limits the growth of online communities? A large body of research in social computing shows how online communities are able to elicit contributions based on the value of prior contributions and on the number of active contributors [2, 27]. On the other hand, the size of communities appears to be limited by various factors including the social structures and technical tools that communities create to manage quality and regulate behavior [14, 45]. However, because this second type of study does not formally account for the fact that contributor time and energy are finite, they cannot rule out alternative ecological explanations including the presence of competitors, an exhausted pool of potential participants, and so on [42]. As a result, we still do not understand how the growth of online communities is limited by these types of external resources.

Density dependence, one of the most well established theories from ecology offers an angle on this question by relating the success of individual members of a population to the density (i.e., crowding) of an environment. It suggests that the benefits of mutualism increase with density initially before these benefits are overwhelmed by the costs of competition [15, 48, 51]. In ecology, density dependence theory predicts an ∩-shaped relationship between density and success. In other words, the ideal conditions for community success will be at an intermediate level of density that maximizes the benefits of mutualism while minimizing the costs of competition.

Hypotheses Given the enormous amount of research on density dependence in biology and in organizational studies [48, 51], existing theory points to two hypotheses we hope to test:

> $H1_A$: The success of an online community will have an ∩-shaped relationship with the density of communities that share its resources.
>
> $H2_A$: The optimal level of density for the success of new communities will be greater when communities cover a more popular topic.

Measures and Analysis Because they are large enough to support statistical inference, we will test our hypotheses using our Reddit and Wikia population data. Following prior work, we will measure success both as the survival of communities over time [e.g., 35] and as the size of communities as measured by number of active participants [e.g., 37]. Within each population, we will measure shared resources as overlap in either topic or users [5]. We will measure topical overlap using a structural topic model [34] fit to the content produced by communities, a technique used in the pilot work on Change.org described in §4. In Wikia, topics come from the content of wiki articles. In Reddit, they will come from the text content of submissions and comments. Following previous work, will measure user overlap as shared membership between communities [2, 5, 43, 52]. Finally, we will measure topic popularity as both general interest in the topics of the community and available information on the topic. We will measure general interest by mapping the terms drawn from our topic models to Google Trends data (see Figure 2 for an example). We will measure available information by drawing on programmatically searchable databases of news media, Wikipedia, and magazines. We will use the results of these measures to build our analytic dataset which we will use to test our test our hypotheses using longitudinal panel regression models.

Anticipated Findings As discussed in §4, we have conducted a proof-of-concept version of our test of $H1_A$ using data from online petitions on Change.org. Although this work gives us some faith that will find support for $H1_A$, the evidence of the effect of competition in our pilot work is weak. We expect this result will also translate to Wikia and Reddit. As a result, we expect that resource constraints will have more influence on survival than on growth. We also anticipate finding support for $H2_A$ in that an increase in the availability of external resources will increase the number of communities and the number of participants a topic can support.

Taken together, we believe that our work will show that while competition for resources might have little effect on participation levels in online communities, external resources may be relevant to the overall survival of subpopulations of communities. Moreover, we anticipate finding that influence of resources acts not by limiting the growth of individual communities but by limiting the survival of new ones. Evidence for these results would imply that individuals starting new online communities should prioritize the consideration of opportunities for mutualism.

These anticipated results (plus the preliminary results from our pilot work) raise two new puzzles (1) Why does mutualism seem more important than competition on average? and (2) Why are populations not dominated by a single generalist community? i.e., Why and how do communities in overlapping topics divide up or "partition" users and contributors? We address these concerns in our second and third proposed projects.

5.2. Project B: Ecological relationships between communities

The density dependence model described in Project A (§5.1) illustrates how the growth and survival of online communities depends on the resources over which communities may compete or build mutalistic ties. Although fundamental to ecological research, the explanatory power of these types of ecological analyses is limited. They offer only indirect evidence of ecological dynamics at the level of the population and do not explicitly identify relationships between particular communities. A more complete explanation should also answer questions that depend on identifying and closely studying groups of specific communities that are linked in specific ecological relationships. For example, will the results for density dependence in Project A operate through a small number of stronger interactions or through a large number of weaker interactions? Online communities frequently overlap in terms of users and content but prior work offers competing explanations of the effects of these overlaps [50, 52]. We believe that an ecological approach that looks at the nature of specific relationships and that establishes differences in the ecological effects of overlap will be able to explain these conflicting results.

In our second project, we will develop and validate a method for detecting and analyzing competitive, mutualistic, amensal, and commensal ecological relationships between communities based on time series of communities' activities. This work will involve comparing communities both within our three platforms of Wikia, Reddit, and Stack Exchange and—inspired by recent work in social computing showing the importance of cross-platform dynamics [13, 28, 49]—comparing communities *across* our three platforms as well. Both parts of this analyses require a shift in frame from studying the effects of environments and environmental variables on online communities to the aggregate effects of dyadic relationships between the hundreds of thousands of communities in our platforms. Although this shift results in a massive increase in the scope and scale of our analysis, it can also offer us a window in the specific ecological mechanisms through which online communities succeed and fail.

Hypotheses As in Project A, we hope to test a series of hypotheses drawn from biological and organizational ecology. The first set of hypotheses concern the relative prevalence of types of dyadic ties within populations and follow findings in studies of populations of organisms and organizations:

> $H1_B$: In dyadic relationships, one-sided relationships will be more common than two-sided relationships; beneficial relationships will be more common than damaging relationships; and weak relationships will be more common than strong relationships.

$H2_B$: The relationship between density and the number and strength of beneficial ties will be curvilinear (∩-shaped).

Platforms constitute partially permeable boundaries that change the nature of ecological relationships. Although previous work has documented interactions across platforms [49], there remain technological and cultural differences between platforms and communities that we believe are likely to mitigate and reduce direct competition. Therefore:

$H3_B$ Cross-platform ecological relationships will be weaker in general and more likely to be mutualistic compared to relationships within a given platform.

Measures and Analysis We will test these hypotheses using data from all three of our populations: Wikia, Reddit, and Stack Exchange. Because this analysis only requires data on the time series of activity and number of active users, it is possible for us to observe ecological relationships between communities on different platforms. We will leverage this fact to test $H3_B$.

To evaluate these hypotheses, we will identify groups of communities (both within and across platforms) where we can observe patterns in temporal activity that provide causal evidence of ecological relationships. Toward this end, we will employ convergent cross mapping (CCM), a recent methodological innovation in ecology that can identify whether two variables may be part of the same dynamical system using observational time series data [41].

Using CCM, we can infer that two variables are linked but cannot identify what type of ecological relationship may exist between them. To test our hypotheses, we will turn to the Lotka-Volterra (LV) family of differential equation models. The parameters of LV models can be interpreted in terms of the sign and strength of relationships between communities, allowing us to identify strong and weak mutualist, competitive, commensalist, and amensalist relationships. For each group of connected communities that we are able to infer using CCM, we will fit a system of LV equations to the time series of the number of active participants and the amount of participation activity. Finally, we can count the occurrences of different dyadic relationships to test our hypotheses about the relative prevalence of these relationships.

Anticipated Findings $H1_B$ is drawn from the ecological literature that finds many ecosystems characterized by a greater number of weak, indirect, and one-sided relationships and a lesser number of strong, direct, and reciprocal relationships [48]. Results from our pilot studies, our intuition that the resources on which communities depend are weakly rival, and the principle of competitive exclusion in ecology which says that strongly competitive relationships are so damaging to survival that they will not last for long, all predict that mutualism will be more common than competition—especially so in the case of strong forms of competition and mutualism.

$H2_B$ tests whether density dependence operates at the level of ecological ties. We anticipate that as density increases, mutualistic ties will increase (but at a decreasing rate) and that competitive ties also will increase (but at an increasing rate). In other words, we expect mutualism between two communities to be greatest when communities have enough in common that resources attracted by one community are useful to the other, but not so much in common that synergy is impossible. Support for these hypotheses would help to explain results on density dependence dynamics in terms of the relationship between dyadic ties. Any set of non-null results will inform strategies for creating new communities—e.g., organizing around a topic in the optimal range of overlap with established communities.

Furthermore, we believe that platforms create boundaries that make it easier for resources to flow between communities on a platform than between communities across different platforms. For example, users may have to create different accounts to participate in communities on different platforms and content may have to be reformatted or redesigned in order to be shared. As the different platforms that we consider have quite different technological affordances and are commonly understood to support different types of communities, we expect the ecological forces between communities on different platforms—and especially competition—to be more rare and less strong. Therefore, we expect to find support for $H3_B$.

5.3. Project C: Ecological strategies

How do you choose a topic or domain that can attract and support a new community? Our third project seeks to understand how new communities might adopt strategies to maximize growth and survival. Building on the results and methods developed in Projects A (§5.1) and B (§5.2), this project considers the means by which new communities can identify or create a niche. The results of this analysis will support the broader impact of this work by linking our analysis to strategic recommendations for the creators and managers of communities.

Previous work conducted by our research group has demonstrated that community founders often answer the question posed in the opening of this section based on intuitions that avoiding competition with existing communities is desirable [10]. Existing communities can also make consequential tactical shifts around community goals. For example, in previous work, we found evidence that the growth of Wikipedia caused other online encyclopedia projects to shift their own focus as they sought to build niches in a shifting resource space [17].

In one sense, Project C can be seen a synthesis of our first two projects. Like Project B (§5.2), it involves identifying different types of ecological ties between communities. It goes beyond Project B by measuring the relative preponderance of these ties in the topical areas of new communities. We will use this characterization of new communities' environments to predict their growth and survival. We will also explore the relationship between existing mutualistic relationships and survival. In this way, our final project attempts to test theories that, if empirically supported, will provide community mangers with a way to evaluate and identify promising niches for new communities.

Hypotheses Our hypotheses are based on intuitions drawn from the ecological theory of resource partitioning: (1) online community success depends on the ability of a community to find a niche; (2) communities that struggle to find a niche are likely to find themselves in competitive relationships; (3) one way to find a niche is to form mutualistic ties.

The first two of our hypotheses aim to connect our dyadic analysis from Project B to density dependence theory as in Project A:

> $H1_C$*:* Online community success will have a curvilinear, ∩-shaped relationship to the amount of mutualistic ties among incumbent communities that share resources with the new community.
>
> $H2_C$*:* Online community success will be negatively related to the number of competitive ties among incumbent communities that share resources with the new community.

Our third hypothesis proposes that communities that are able to innovate by finding novel beneficial connections will be more likely to survive:

> $H3_C$*:* Online community success will be positively related to the number of mutualistic ties to communities having few other mutualistic ties.

Measures and Analysis As in Project A, we will conduct this analysis using data from populations of communities on Wikia and Reddit. Our analytic approach will build on our datasets and measures constructed in the previous projects. Unlike the previous projects, $H1_C$ and $H2_C$ will each rely on a cohort study of focal communities that were founded after a selected point in time (t_0). This choice is driven by the fact that the CCM technique necessary for these analyses depends on long time series, making it difficult to observe ecological networks changing over time or to directly measure the presence of mutualistic or competitive ties in new projects. On the other hand, because it is important to understand the role of mutualistic ties, and not just resource overlaps, our test of $H3_C$ will model the success of incumbent communities at times following (t_0) using a longitudinal analysis.

Our dependent variable for $H1_C$ and $H2_C$ will be the growth and survival of our focal communities and for $H3_C$ it will be the growth and survival of the incumbent communities. Our independent variables will include measures of ecological ties among the incumbent communities that overlap with the focal community at t_0 using the methods described in Project B. We will detect topical overlaps using the methods described in Project A in order to identify incumbent communities that overlap (in topic and users) with the focal communities. Finally, we will test our hypotheses using panel and survival regression analyses.

Anticipated Findings We anticipate finding evidence to support all three of our hypotheses. As per $H1_C$, we expect to observe a curvilinear, ∩-shaped relationship between the number of mutualism ties and a new community's chances of survival because we expect that the presence of mutualism reflects that incumbent communities are effectively providing resources that new communities may be likely to share. We believe that the presence of mutualism will increases the chances that a new community will be able to survive and grow.

As per $H2_C$, we expect competitive ties will be most likely to form when the topical environment becomes crowded and that existing competitive ties will reflect increased chances that new communities will compete and fail. Finally, we expect to find support for $H3_C$ that forming novel mutualistic ties is a way for communities to construct new niches where they can thrive.

These anticipated results would support a clear set of strategic directives for the founders of online communities: seek out environments that are likely to be mutualistic ($H1_C$), choose environments that minimize competition ($H2_C$), and seek mutualistic relationships with other communities where these relationships provide a distinctive advantage ($H3_C$). In addition to general advice, our work would provide data and software that online community founders and managers could use to make better specific strategic decisions about where, when, and how to engage.

5.4. Articulating a Theory of Online Community Ecology

As a way of summarizing our findings, we will produce two additional review papers. The first is a summary of the implications of this work for community managers and founders in the business practitioner press (like *Harvard Business Review*). The second is an attempt to synthesize the results of this work with the broader literature on online community success. For this second effort, we plan to create a prediction model that runs widely-used within-community predictors of community success drawn from previous work [e.g., 27, 36], with the ecological measures created as part of this work. We anticipate showing that—as implied in §4—predictions of community growth and survival can be improved substantially by the incorporation of ecological measures and by the "fitness" of a community within its niche. We believe that this type of analysis will summarize and syntehsize our findings while underscoring the idea that online communities' ecologically interdependent natures reflects a critically overlooked driver of online community success.

5.5. Creating Research Tools and Datasets

In order to produce the empirical studies described above, we will collect raw digital trace data from the online communities that we will analyze. Next, we will build datasets from this raw data by cleaning it, validating it, and building our analytic variables. Some of the variables we will analyze will be derived from algorithmic techniques including topic models, convergent cross mapping, and Lotka-Volterra models. The software tools that we use to collect data and construct measures will be designed for those who may wish to repurpose, reproduce, or extend our analyses. We will reuse components of existing software for topic modeling, convergent cross mapping, and fitting Lotka-Volterra models. Our tools will ease the burden of applying these components to online community data. As we have done in previous research projects, we will release all of the tools we create under FLOSS licenses.

While developing and releasing software tools will help some researchers, it will not help those who lack substantial computational resources, such as the Hyak high performance computing system which we will use to build these variables. To empower others without these resources, we will release our datasets through open data repositories such as the Harvard Dataverse. Our research team has a strong history of releasing code and data from our research projects. As evidence of this commitment by our team, PI Hill will be 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."

6. Dissemination Plan, Education, and Outreach

We will disseminate the outcomes of this project through scholarly communication channels including conferences, workshops, and journals. We have a history of publication in communication journals such as the *Journal of Communication*, *Communication Research*, and *Information, Communication, and Society* 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, we will release public and freely licensed versions of our research products. In the past, we have published in open access scientific journals, released work under open access licences, and published pre-print versions of papers on our websites. We also have published open datasets and freely licensed software tools which were used for our research. Under the current proposal, we will create and disseminate additional tools and datasets.

The PIs both plan to include methods and results from this research directly in the courses they teach at Northwestern University and the University of Washington. Both PI Shaw and Hill teach undergraduate courses in online communities and graduate courses in both online communities and Internet-based research design with a strong conceptual and practical engagement with online communities Both PIs will incorporate findings, tools, and data into their courses. At both institutions, graduate students will be involved in the project as research assistants, collaborators, and advisees, and will interact with both PIs through research activities and monthly team meetings over video conference. These experiences will provide them with a supportive learning environment incorporating personnel from both Northwestern and the University of Washington.

As both PIs have done in their previous research, the research team will work closely with organizations supporting online communities in ways that go beyond simply using these communities as a source of data. For example, both PIs are active contributors to Wikia and Wikipedia and PI Hill has been a member of the Wikimedia Foundation advisory board since 2007. PIs Shaw and Hill also deliver an annual talk on academic research at Wikipedia's yearly International conference, Wikimania. Both PIs have regularly given talks on their research at Wikimania, the Wikimedia Foundation, Wikia Inc., and WikiHow (another commercial wiki hosting firm). Hill is also a founding board member of the Wikimedia Cascadia User Group, and a regular attendee and organizer of events related to Wikipedia in the Seattle area. We will use these relationships with the peer production contributor and business communities to disseminate our research and to ensure that it is relevant, useful, and usable.

We will also disseminate the findings from this research more broadly. 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. We will engage in multiple outreach activities to ensure that managers, designers, and leaders of other systems pursuing large-scale online collaboration access our findings. PI Shaw serves on the Citizen Science Advisory Board of the Adler Planetarium in Chicago, where he assists in the design and assessment of the Zooniverse, another widely successful peer production platform that engages hundreds of thousands of volunteers in Citizen Science projects. PI Hill is a Board Member of the Free Software Foundation, a former Director of the Ubuntu Foundation, and an active developer in the Debian FLOSS community. Both PIs are Faculty Associates at the Berkman Klein Center for Internet and Society at Harvard University. In these capacities, we frequently

		Year 1				Year 2				Year 3			
Tasks	Preparation	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Project A (Ecological Environments)													
Study planning and design		■											
Measure creation and validation			■										
Analyze data and prepare manuscripts				■	■								
Publish and disseminate findings						■	■						
Project B (Ecological relationships between communities)													
Study planning and design					■								
Measure creation and validation						■							
Analyze data and prepare manuscripts							■	■					
Publish and disseminate findings									■	■			
Project C (Ecological strategies)													
Study planning and design								■					
Measure creation and validation									■				
Analyze data and prepare manuscripts										■	■		
Publish and disseminate findings												■	■
Paper Summarizing Theory and Findings													
Synthesize materials and draft analysis										■	■	■	
Publish and disseminate findings												■	■
Software and Tools													
Collect data from Wikia	Completed												
Collect data from Reddit	Completed												
Collect data from Stack Exchange					■								
Write prototypes of data processing tools for Wikia	Completed												
Write prototypes of data processing tools for Reddit		■	■										
Write prototypes of data processing tools for StackExchange		■	■										
Implement, interate, and improve software tools		■	■	■	■	■	■	■	■	■	■	■	
Release initial version of software tools									■				
Release final version of software tools and dataset													■

Figure 3: Project timeline broken up in the quarter system used by both Northwestern and UW. Q1 is the start of the academic year on the quarter system (i.e., Oct-Dec).

speak at industry events and public forums and engage in advisory activities with firms and non-profits. We will leverage these opportunities to disseminate findings from the research we propose here.

7. Project Timeline

The timeline in Figure 7 describes the staging of the research proposed in §5.

8. Intellectual Merit

The **intellectual merit** of this project is to advance scientific understanding of virtual organizations through the elaboration and empirical validation of a theory of ecological dynamics in large scale virtual organizations. We will contribute knowledge of the environmental preconditions of successful collaboration to the study of social computing systems. We will also contribute to organizational sociology and complex systems science by adapting and applying theories and models from these areas into a novel domain that is rich in detailed communication and behavioral data.

9. Broader Impacts

The proposed work will achieve **broader impacts** by paving the way to understanding the environmental conditions under which online communities grow and survive. Online communities play an increasingly important role in many fields including software production, knowledge management, cultural production, and education. Additionally, a growing number of organizations look to online communities as a source of innovation and customer support. Our work will provide actionable insights for initiators and managers of online communities that will support the design of more successful communities by allowing managers to effectively discover and construct niches in which their communities are likely to thrive. Finally, we will produce freely licensed and publicly available computational research systems and datasets that will enable reproducible research and the dissemination of our results. Through scholarly and public engagement activities, we will ensure wide dissemination of our findings.

10. PI Preparation

10.1. Personnel

Aaron Shaw, Ph.D. (PI), is Assistant Professor in the Department of Communication Studies at Northwestern University and a Faculty Associate of the Berkman Center for Internet and Society at Harvard University. At Northwestern, he is Director of Graduate Studies in the Ph.D. Program in Media, Technology and Society and is affiliated with the Sociology Department, the Institute for Policy Research, the Buffett Institute, and the Science of Networks in Communities (SONIC) Research Group. His work has been supported by FUSE Labs at Microsoft Research, the Ford Foundation, the Ewing and Marion Kaufman Foundation, the Berkman Klein Center for Internet and Society, the University of California Office of the President, and the United States Department of Education. Shaw holds graduate degrees in Humanities and Sociology from Stanford University and UC Berkeley.

Shaw studies collective action, collaboration, and mobilization in online communities and crowds, and has published widely on these topics. He has received awards from the American Political Science Association, American Sociological Association, the International Communication Association, and the Association for Computer Machinery Conference on Computer Supported Cooperative Work (CSCW) for several papers and his dissertation research. Shaw brings expertise in organizational research, peer production, quantitative and quasi-experimental methods, empirical analysis of online communities, and research design.

Benjamin Mako Hill, Ph.D. (PI), is Assistant Professor of Communication and Adjunct Assistant Professor in the Department of Human Centered Design and Engineering at the University of Washington. He is also a Faculty Associate of the Berkman Klein Center for Internet and Society and an affiliate at the Institute for Quantitative Social Science—both at Harvard University. Hill holds 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. He has published numerous articles in peer reviewed journals and conference proceedings, and has 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 his graduate studies, PI Hill worked full time as a software engineer and has received his masters degree from MIT for software development and HCI research. Hill has a background in technology management, data-driven statistical analyses of online communities, computational research, management science, and peer production. Additionally, Hill has been a leader, developer, and contributor to the free and open source software community 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 is the author of several best-selling technical books [16, 19, 33]. He is a member of the Free Software Foundation board of directors and has served on the Advisory Board of the Wikimedia Foundation since 2007.

In addition to the two PIs, two graduate researchers have committed to play an important role in carrying out these work. **Nate TeBlunthuis** is a Ph.D. candidate at the University of Washington in the Department of Communication. His Ph.D. research involves ecological dynamics of online communities and he has lectured on the subject. He has led the design and implementation of the pilot study on e-petitions using data from Change.org described in §4. **Jeremy D. Foote** is a Ph.D. candidate Northwestern University in the Department of Communication Studies. His Ph.D. research focuses on network analysis in population studies of online communities, studies of inter-community dynamics, and experiences of early-stage online communities. We have budgeted to involve both students in regular team meetings if they graduate and move to other institutions during the period of the grant.

10.2. Related Experience and Prior Work

Both PIs possess advanced training in organizational research, statistical analysis, and quantitative methodology and have published peer reviewed research employing large-scale, empirical data analysis techniques. Over the past decade, the PIs have been working collaboratively on research projects and pilot studies to build the experience and skills to successfully complete these projects. As collaborators, we have undertaken multiple research projects on peer production, resulting in eight co-authored peer reviewed publications to date [20, 21, 22, 26, 30, 37, 46], three book chapters [1, 11, 23] and many additional working papers and conference presentations. In a book chapter that will published in an edited volume by Oxford University Press, we lay out out the promise and the need for the kind of ecological research described in this proposal [23].

We have already collected much of the raw data necessary to complete the studies described in this proposal. Although analyzing these data in the way we have described is challenging, we have the preparation to complete this research with the resources requested. We have already parsed and compiled data on a population of wikis from from Wikia which have used in several published research projects [37, 46] and several other that in prepartion and under review. We have also collected data from Reddit which have begun to explore in some of the analyses published here. We are confident that we can incorporate data from Stack Overflow as well.

11. Results from Prior NSF Support

PIs Shaw and Hill each have an active NSF award as part of a collaborative project: "CHS: Small: Collaborative Research: Pathways to Community Success: Advancing a Comparative Science of Online Collaborative Organization" (UW IIS-1617468 for $194,325 at Northwestern & IIS-1617129 for $305,359 at UW). As part of the work, the two PIs have built an exceptionally strong and productive collaboration. The PIs have created a joint group between their two institutions called the "Community Data Science Collective" which has, over the last two years, become a premier research group working on computational studies of online communities. The PIs coordinate with each other on a daily basis, serve on committees of each others' advisees, hold bi-weekly virtual lab meetings with people from both universities year-round, meet in person several times a year, and hold face-to-face lab meetings in both Chicago or Seattle once each year.

The previous award supported a series of empirical research projects comparing the effects of design or policy changes to online communities in a population of wikis drawn from Wikia. An unanticipated insight of this work was a recognition of the salience of the fact that communities within populations interact and effect each other [23, 26]. This proposal is a direct result of these observations. The period of the prior award is scheduled to end before the proposed award would begin.

11.1. Intellectual Merit

The previous award advanced scientific understanding of collaborative organization through the elaboration of a longitudinal theory of pathways to effective peer production and validatation 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 inferences about the factors that shape the growth and effectiveness of peer production systems. Just over two years into the three-year award, the award has already resulted in six peer reviewed papers [8, 10, 12, 26, 30, 46] one peer reviewed poster presentation [45], two book chapters [18, 23] two datasets [9, 31], one piece of research software [24], and more than a dozen other talks and conference presentations. This work has also resulted in multiple awards. A number of other papers, datasets, and tools are still in preparation and under review.

11.2. 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, we have shared our work with researchers and managers at companies and non-profit organizations running large communities. Additionally, we have generated a set of freely licensed and publicly available computational research systems and datasets which other researchers have used in their own projects. In both ways, our work has contributed to the design of more effective and more collaborative organizations in online communities, in business, and in society.

References

[1] Benkler, Y., Shaw, A., and Hill, B. M. (2015). Peer production: a form of collective intelligence. In Malone, T. W. and Bernstein, M. S., editors, *Handbook of Collective Intelligence*, pages 175–204. MIT Press, Cambridge, MA.

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

[3] Carroll, G. R. (1985). Concentration and specialization: Dynamics of niche width in populations of organizations. *American Journal of Sociology*, 90(6):1262–1283.

[4] Carroll, G. R. and Swaminathan, A. (2000). Why the microbrewery movement? Organizational dynamics of resource partitioning in the U.S. brewing industry. *American Journal of Sociology*, 106(3):715–762.

[5] Datta, S., Phelan, C., and Adar, E. (2017). Identifying misaligned inter-group links and communities. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–23.

[6] Foote, J. (2018a). An Agent-Based Model of Online Community Joining.

[7] Foote, J. (2018b). An Agent-Based Model of Online Community Joining.

[8] Foote, J. and Contractor, N. (2018). The behavior and network position of peer production founders. In Chowdhury, G., McLeod, J., Gillet, V., and Willett, P., editors, *iConference 2018: Transforming Digital Worlds*, Lecture Notes in Computer Science, pages 99–106. Springer.

[9] Foote, J., Gergle, D., and Shaw, A. (2017a). Replication data for: Starting online communities: motivations and goals of wiki founders. *Harvard Dataverse*. type: dataset.

[10] Foote, J., Gergle, D., and Shaw, A. (2017b). Starting online communities: motivations and goals of wiki founders. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, pages 6376–6380, New York, NY. ACM.

[11] Foote, J., Shaw, A., and Hill, B. M. (2017c). A computational analysis of social media scholarship. In Burgess, J., Marwick, A., and Poell, T., editors, *The Sage Handbook of Social Media*. Sage, Thousand Oaks, CA, USA.

[12] Gan, E. F., Hill, B. M., and Dasgupta, S. (2018). Gender, feedback, and learners' decisions to share their creative computing projects. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):54:1–54:23.

[13] Graeff, E., Stempeck, M., and Zuckerman, E. (2014). The battle for 'Trayvon Martin': Mapping a media controversy online and off-line. *First Monday*, 19(2).

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

[15] Hannan, M. T. and Freeman, J. (1989). *Organizational ecology*. Harvard University Press, Cambridge, MA, 1 edition.

[16] Hill, B. M. (2005). *Debian GNU/Linux 3.1 Bible*. Wiley Pub, Indianapolis, Ind.

[17] Hill, B. M. (2013). Almost Wikipedia: What eight early online collaborative encyclopedia projects reveal about the mechanisms of collective action. In *Essays on volunteer mobilization in peer production*. Massachusetts Institute of Technology, Cambridge, Massachusetts. PhD Dissertation.

[18] Hill, B. M. (2018). Whither peer production. In Hassan, S. and De Felippi, P., editors, *Decentralizing the Commons*. Institute for Network Culture, Amsterdam, The Netherlands. Forthcoming.

[19] Hill, B. M., Burger, C., Jesse, J., and Bacon, J. (2008). *Official Ubuntu book*. Prentice Hall, 3 edition.

[20] Hill, B. M. and Shaw, A. (2013). The Wikipedia gender gap revisited: characterizing survey response bias with propensity score estimation. *PLoS ONE*, 8(6):e65782.

[21] Hill, B. M. and Shaw, A. (2014). Consider the redirect: a missing dimension of Wikipedia research. In *Proceedings of The International Symposium on Open Collaboration*, OpenSym '14, pages 28:1–28:4, New York, NY, USA. ACM.

[22] Hill, B. M. and Shaw, A. (2015). Page protection: another missing dimension of Wikipedia research. In *Proceedings of the 11th International Symposium on Open Collaboration*, OpenSym '15, pages 15:1–15:4, New York, NY, USA. ACM.

[23] Hill, B. M. and Shaw, A. (2018). Studying populations of online communities. In González-Bailón, S. and Foucault Welles, B., editors, *Oxford Handbook of Networked Communication*. Oxford University Press, Oxford. Forthcoming.

[24] Hill, B. M. and TeBlunthuis, N. (2018). Mediawiki dump tools.

[25] Kiene, C., Monroy-Hernández, A., and Hill, B. M. (2016). 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)*, pages 1152–1156, New York, NY. ACM.

[26] Kiene, C., Shaw, A., and Hill, B. M. (2018). Managing organizational culture in online group mergers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):89:1–89–21.

[27] Kraut, R. E., Resnick, P., and Kiesler, S. (2012). *Building successful online communities: evidence-based social design*. MIT Press, Cambridge, MA.

[28] McMahon, C., Johnson, I. L., and Hecht, B. J. (2017). The substantial interdependence of Wikipedia and Google: A case study on the relationship between peer production communities and information technologies. In *International AAAI Conference on Web and Social Media (ICWSM 2017)*, pages 142–151, Palo Alto, California. AAAI.

[29] McPherson, M. (1983). An ecology of affiliation. *American Sociological Review*, 48(4):519–532.

[30] Narayan, S., Orlowitz, J., Morgan, J., Hill, B. M., and Shaw, A. (2017a). 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, pages 1785–1799, New York, NY, USA. ACM.

[31] Narayan, S., Orlowitz, J., Morgan, J. T., Shaw, A. D., and Hill, B. M. (2017b). Replication data for: The Wikipedia Adventure: field evaluation of an interactive tutorial for new users. *Harvard Dataverse*. type: dataset.

[32] Park, R. E. (1936). Human Ecology. *American Journal of Sociology*, 42(1):1–15.

[33] Rankin, K. and Hill, B. M. (2009). *The official Ubuntu server book*. Prentice Hall, Upper Saddle River, NJ. OCLC: 1001929364.

[34] Roberts, M. E., Tingley, D., Stewart, B. M., and Airoldi, E. M. (2013). The Structural Topic Model and Applied Social Science. In *2013 Workshop on Topic Models: Computation, Application, and Evaluation.*, page 4.

[35] Schoener, T. W. (1974). Resource Partitioning in Ecological Communities. *Science*, 185(4145):27–39.

[36] Schweik, C. M. and English, R. C. (2012). *Internet success: A study of open-source software commons*. MIT Press, Cambridge, MA.

[37] Shaw, A. and Hill, B. M. (2014). Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication*, 64(2):215–238.

[38] Shirky, C. (2008). *Here comes everybody : The power of organizing without organizations*. Penguin Press, New York, NY.

[39] Singer, J. D. and Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press, New York, NY.

[40] Soule, S. A. and King, B. G. (2008). Competition and resource partitioning in three social movement industries. *The American Journal of Sociology*, 113(6):1568–1610.

[41] Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., and Munch, S. (2012). Detecting Causality in Complex Ecosystems. *Science*, page 1227079.

[42] Suh, B., Convertino, G., Chi, E. H., and Pirolli, P. (2009). The singularity is not near: slowing growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, pages 1–10, New York, NY. ACM.

[43] Tan, C. (2018). Tracing Community Genealogy: How New Communities Emerge from the Old. In *Proceedings of the Twelfth International Conference on Web and Social Media (ICWSM '18)*, pages 395–404, Palo Alto, California. AAAI.

[44] TeBlunthuis, N. (2018). *Density dependence without resource partitioning on an online petitioning platform*. Thesis, University of Washington, Seattle, Washington.

[45] TeBlunthuis, N., Shaw, A., and Hill, B. M. (2017). 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, pages 323–326, New York, NY, USA. ACM.

[46] TeBlunthuis, N., Shaw, A., and Hill, B. M. (2018). 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)*, pages 355:1–355:7, New York, NY. ACM.

[47] Thornton, K. and McDonald, D. W. (2012). Tagging wikipedia: ollaboratively creating a category system. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, GROUP '12, pages 219–228, New York, NY. ACM.

[48] Verhoef, H. A. and Morin, P. J. (2010). *Community ecology: processes, models, and applications*. Oxford University Press, Oxford. OCLC: 876676566.

[49] Vincent, N., Johnson, I., and Hecht, B. (2018). Examining Wikipedia with a broader lens: quantifying the value of Wikipedia's relationships with other large-scale online communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 566:1–566:13, New York, NY. ACM.

[50] 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):414–431.

[51] Worster, D. (1994). *Nature's economy: a history of ecological ideas*. Cambridge University Press, Cambridge; New York, NY, USA. OCLC: 855524849.

[52] Zhu, H., Kraut, R. E., and Kittur, A. (2014). The impact of membership overlap on the survival of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, pages 281–290, New York, NY. ACM.