Dissertation Proposal: Ecology of Online Communities

Nathan TeBlunthuis
nathante@uw.edu *

March 23, 2019

Introduction

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 one another over content or contributors? Is the evolving world of online communities better understood as a competitive strugle 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., Halfaker, Geiger, Morgan, & Riedl, 2013; Kraut, Resnick, & Kiesler, 2012; Schweik & English, 2012; Shaw & Hill, 2014; TeBlunthuis, Shaw, & Hill, 2018). These introspective, approaches based in social psychology and engineering typically account for only a small amount of variation in communities' growth, longevity, and performance. Ecology, "an attempt

*Some of the text and ideas presented here also appear in a grant proposal submitted to the NSF that I drafted in the Fall of 2018 and to which Jeremy Foote, Aaron Shaw, and Benjamin Mako Hill contributed.

to deal holistically with the phenomenon of organization" (Hawley, 1986, p. 7) provides an 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 (Hannan & Freeman, 1989). From an ecological view, the individual organism or online community cannot be fully understood except through in its interdepence relations with others; its function and role in the broader ecosystem (Worster, 1994).

Ecology promises implications for the design and management of online communities. Analyses of ecological factors in the life sciences enable effective wildlife management, pest control, and sustainable utilization of renewable resources. In sociology, organizational ecology provides compelling explanations for the life-cycles of industries, organizational specialization, and patterns of collaborative partnerships. Recent research from the social computing literature on interdependence between online communities suggests that ecological analyses can provide not only novel scientific understandings but also viable community management strategies (Chandrasekharan et al., 2017, CSCW; Datta, Phelan, & Adar, 2017; Graeff, Stempeck, & Zuckerman, 2014; Kiene, Shaw, & Hill, 2018, CSCW; McMahon, Johnson, & Hecht, 2017; Tan, 2018; TeBlunthuis, Shaw, & Hill, 2017; Vincent, Johnson, & Hecht, 2018; Wang, Butler, & Ren, 2013; Zhu, Kraut, & Kittur, 2014).

An ecological understanding of online communities can identify environmental conditions favoring projects in given ecological niches to suggest the creation of new communities to fill them; predict how regulatory decisions that platforms make through policy or code will affect the health of particular individual communities; simply by better explaining the success and failure of communities in terms of forces beyond those communities' direct control, the ecological approach can help us know whether we should attribute the success or failure of given communities to their own choices or to external forces which may drive their fates. In statistical terms, ecological forces can confound observational studies of online communities and we need new conceptual and analytical approaches to account for them.

My goal is to develop and explicate a partial ecological theory of online communities through three linked projects answering fundamental ecological questions: (1) How does the growth and survival of online communities depend on their ecological communities?[1] (2) How do ecosystems of online

[1]The word "community" is used by researchers of online communities to to refer to

groups themselves develop over time? (3) How do the choices of individual participants in online communities lead to ecological dynamics? These three questions relate three different levels of analysis: ecological communities, the interdependent online groups that comprise them, and the individuals who contribute to the groups.

I propose to apply this approach to two popular online platforms for online communities: Wikia and Reddit.[2] For each, I 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 variations in external resources in the form of users and underlying interest in topics. I will test these predictions empirically using large scale statistical analysis of longitudinal digital trace data drawn from both platforms. Throughout this project I will iterate and alternate between theoretical and empirical analysis to synthesize and formalize a partial theory of ecology for online communities. In the course of this work I will also design, build, and publish software and unique research datasets for other researchers to use.

Background

Online Communities and Peer Production

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 (Benkler, Shaw, & Hill, 2015). Other online communities like Reddit sometimes engage in peer production, and often provide information, social support, and entertainment to millions of people. Online community platforms support millions of attempts to build communities but only a tiny percentage manage to mobilize participants and to sustain collaboration (Healy & Schussman, 2003; Hill & Shaw, 2019; Schweik & English, 2012;

---

groups of individuals having regular interactions, typically at a given virtual place. It is also used by ecologists to describe interrelated species or organizational forms coexisting in a shared environment. In this projects framework, online communities occupy the "species" position and interrelated online communities are analogous to "ecological communities." I'll use the terms "online community" (or "online group") and "ecological community" for clarity and I'll avoid using the term "community" without making it clear whether I'm referring to individual online groups or to and ecological community comprised of those groups.

[2] Wikia has re-branded as "Fandom"

Shirky, 2008). The significance of these accomplishments should not be understated. Provisioning new categories of global public goods like free software and encyclopedias is a rare achievement for an organizational form that places peer production in an elite class shared with governments, markets, social movements, and universities.

Despite the success of peer production, the decline in active contributors to Wikipedia motivates renewed attention to the challenges of maintaining a pool of active contributors to an online community (Suh, Convertino, Chi, & Pirolli, 2009). Most prior studies of the growth, survival, and success of online communities have focused almost exclusively on communities' internal features with an emphasis on attracting and retaining participants (Hill & Shaw, 2019; Kraut et al., 2012). Studies focusing on growth and survival outcomes often—and many times implicitly—use some version of critical mass theory (Marwell & Oliver, 1993) and may make recommendations for technological or organizational designs intended to help attract and retain a critical mass of participants.

This is typified by Kraut et al. (2012), who, after reviewing a great deal of HCI and social psychological research relevant to online community success, say that people will contribute to online communities when the benefits to them exceed the costs. Butler, 2001; Kraut et al., 2012. Benefits of participation, such as intrinsic or extrinsic motivations for contributing, and identity and bonds-based commitments of participants to the group, depend on activity in the community, such that unless a *critical mass* of contributors is maintained, the community will become inactive (Kraut et al., 2012).

On the other hand, the size of communities is apparently limited by factors including the social structures and technical tools that communities create to manage quality and regulate behavior (Halfaker et al., 2013; TeBlunthuis et al., 2017). An increasing group size leads to increasing costs to participation and so communities will grow to an equilibrium size at their *carrying capacity*. (Butler, 2001) considers costs of communication in larger groups. the decline of Wikipedia suggest that that systems for maintaining the established order and content quality impose high costs that differentially impact inexperienced and under-represented participants in the online community (Halfaker et al., 2013). These mechanisms likely extend to other online knowledge production projects (TeBlunthuis et al., 2018).

Yet evidence supporting such lifecycle-based accounts of online community success may be confounded by external factors that can drive participation in online communities. Because this approach does not formally account for the fact that contributor time and energy are finite, it cannot rule out alternative ecological explanations including the presence of competitors, an exhausted pool of potential participants, and so on (Suh et al., 2009). For example, the contemporaneous rise of Facebook and other social networking sites remains a possible alternative explanation for Wikipedia's transition from rise to decline. As a result, we still do not understand how the growth of online communities is limited by these types of external resources.

This project proposes to extend prior research on online communities by building on established approaches to the ecological study of biological populations (Verhoef & Morin, 2010; Worster, 1994) and organizations (Hannan & Freeman, 1989; Ruef, 2000). No competent wildlife biologist predicting the survival of an animal population would only consider the species' internals and physiology in isolation. Instead, they would consider the availability of necessary resources like food and shelter, the presence or absence of other organisms, and competitive dynamics that might give the species an advantage. This project's 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.

Ecological studies of online communities

This approach to online communities is not altogether new. Indeed a handful studies published in top venues for scientific research into online communities and peer production systems have already taken up the framework (Tan, 2018; Tan & Lee, 2015; Wang et al., 2013; Zhu, Chen, et al., 2014; Zhu, Kraut, & Kittur, 2014). Most notable are a series of of three related studies, Wang et al. (2013), Zhu, Chen, et al. (2014) and Zhu, Kraut, and Kittur (2014), that apply density dependence theory from organizational ecology in studies of three different community ecosystems. One of the most striking findings from this work is that newly formed Wikia wikis have greater longevity when they have many early contributors who also participate in more established wikis (Zhu, Kraut, & Kittur, 2014).

These studies are limited in the degree to which they adapt organizational ecology to theorize how ecological dynamics depend on distinctive aspects of online communities compared to conventional forms of organization, such as the fact that human and information resources can move almost freely between online communities. Moreover, many central features of ecological analysis—like ecological relationships and the creation and discovery of new ecological niches—remain entirely unexplored in the context of online communities. Prior studies have not attempted to model online community environments in terms of dyadic interactions—a first step in community ecology analysis (Verhoef & Morin, 2010).

By contrast, Tan (2018)'s explores reproduction and inheritance in online communities and references ecology as an inspiration, but doesn't synthesize its empirical contributions into a general theory of community growth or survival. In general, such contextually focused empirical analyses of interdependence in online communities studies (e.g. Chandrasekharan et al., 2017, CSCW; Datta et al., 2017; Vincent et al., 2018) are valuable contributions making informative empirical contributions and demonstrate the potential relevance of ecological modeling to online communities, but more can be done to marshal the full power of ecology to explain why some online groups grow and survive while others do not.

Conversely, there is an opportunity to feed knowledge of ecological dynamics of online groups to the social science of organizations by contributing new ecological understandings of the diverse forms that human organization can take. Organizations are "the basic vehicles through which collective action occurs" (Aldrich & Ruef, 2006, p. 1), but most of what we know about the ecology of organizations derives from studies of publicly traded firms due to the convenience of data and career incentives. Therefore ecological studies of collective action in broader contexts such as social movements (e.g. Minkoff, 1997; Olzak & Uhrig, 2001; Soule & King, 2008) and voluntary organizations (e.g. McPherson, 1983) are important. The empirical context of online communities departs from the mainstream of organization studies. Conventional assumptions of the various ecological approaches, such as the clear definition of organizational boundaries and the importance of geographic locality, are violated in the cases considered here (Wang et al., 2013).

Prior attempts to apply organizational ecology to online communities that only translate theories from organizational contexts to online communities have encountered surprising results that suggest some specific open opportunities to study how ecological dynamics depend on distinctive features of the online environment. In two similar studies, Wang et al. (2013) observe evidence for competition between Usenet groups while Zhu, Kraut, and Kittur (2014) observe evidence for mutualism between Wikia wikis. These findings usefully illustrate how resource overlaps do not necessarily lead to a given type of ecological relationship (Verhoef & Morin, 2010), but importantly they also suggest that ecological dynamics in online ecosystems may depend on factors specific to the context and nature of online communities: the topic or identity associated with a community, the distinctive motivations of online community organizers, characteristics of platforms hosting communities, and the stage of the development of ecological communities (Wang et al., 2013).

We have evidence that 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. Explicating a theory of ecological dynamics in the online environment requires going beyond straightforward translation of theory from one domain into another. It requires new theoretical development with careful attention to how ecological dynamics that drive online communities' successes may diverge from those found in other domains. Ecology has potential to provide new conceptual, analytical, and computational tools to inform community leaders and platform designers. More theoretically and empirically rigorous steps toward an ecology of online communities are overdue.

Mathematical Formulation I will adopt from ecology a proven tool for theoretical development: mathematical formalization (e.g. Axelrod & Hamilton, 1981; Hannan & Freeman, 1977, 1989; McPherson & Ranger-Moore, 1991; McPherson & Rotolo, 1996; Shi, Dokshin, Genkin, & Brashears, 2017). Effective mathematical models represent theorized relationships in precise and clear ways, and afford generating predictions that are empirically falsifiable and non-intuitive (McPhee & Poole, 1981). This differs from the more widely adopted approach of statistical hypothesis testing, which relies on mathematical representations chosen not to better communicate or analyze theoretical ideas, but for empirical expedience. General and complex models, such as differential equations or agent based models, can more accurately and precisely represent theoretical ideas in a way that is useful for understanding the internal logic of ideas and for communicating them, but can be very difficult to fit to data. Ruef (2000)'s community ecology study of the emergence of organizational forms is a good illustration. He represents his theoretical ideas by extending Lotka-Volterra differential equations for density dependence to account for dyadic co-evolutionary interactions between organizations. Yet parameters of this model cannot be easily estimated from available data, so instead Ruef uses a Poisson regression to test predictions derived from the theory. I will use some simple formal models in the next section and I will use them extensively in Project C.

Which ecological approach to follow?

In this section I dive a bit deeper into the ecological literature to out sketch where I position my proposed work relative to prior work in organizational ecology and ecological analyses of online communities. Several quite different strands of ecological theory have a substantial influence in organization science. Often these are dichotomized as population ecology and community ecology (Astley, 1985). The distinction between these two tendencies can be defined as different levels of analysis: the population and the community. In Biology, population ecology focuses on interdependence internal to a population of a single species (Verhoef & Morin, 2010). The population ecology of organizations draws an analogy between a population of species of organisms and a set of organizations sharing an *organizational form*, which often corresponds to a business model (Ruef, 2000). Community ecology extends this analogy to analyze organizational communities which are systems of interdependent populations of organizations (Lee & Monge, 2011). Ruef (2000) treats this level of analysis as comparable to the notion of an *organizational field*.

Astley (1985) also uses the levels-of-analysis distinction between population and community ecology. But Astley emphasizes that moving to a higher level of analysis is not and end in itself, but a means to explanation of the origins of organizational diversity. Astley motivates the development of community ecology for organizations because "population ecology emphasizes forces that make organizations more uniform rather than more diverse" (Astley, 1985, p. 224) Indeed the population ecology approach to organizations was developed to explain organizational change under conditions of high degrees of structural inertia driven by institutional needs for reproducible organizational structure (Hannan & Freeman, 1984). In this environment a process of change by death and replacement was more realistic than a process of organizational adaptation. When a set of organizations can be seen as uniform, then they can be understood and modeled as a population, but different organizational forms by definition cannot be seen as uniform. So it is clear why studies at the level of organizational forms adopt a community ecology approach.

Theories of population ecology are exemplified by density dependence theory, which considers how opposing forces of legitimization and competition between members of a population govern the growth of the population. The following Lotka-Volterra equation for logistic growth represents a typical model in density dependence theory:

$$\frac{dN}{dt} = r(N - \frac{N^2}{K})$$

N is the number of active organizations in a population, $\frac{dN}{dt}$ is instantaneous change in the population size, K is the carrying capacity and r is the growth rate (Hannan & Freeman, 1989; Ruef, 2000). Forces of legitimization correspond to r and forces of competition correspond to $\frac{1}{K}$. The population will grow or decline until $N = K$ because $\frac{dN}{dt} = r(N - \frac{N^2}{N}) = 0$. In this model, growth of an organizational form is driven by legitimacy proportional to the current population size, but growth cannot continue forever because a second-order effect of population size is competition (observe the functional similarity to critical mass theories and Butler (2001)'s membership size model). This illustrates the population ecology approach: a population sharing a common organizational form is under consideration, and relationships to other forms are ignored.

But the online communities I am considering are not isomorphic, instead they are differentiated by their topics. This creates something of a puzzle: online communities are diverse, and so a population ecology approach emphasizing forces of uniformity doesn't seem to fit. The community ecology approach focuses on interdependent organizational forms, but I wish to study interdependence between communities, not between populations of communities. Zhu, Chen, et al. (2014), still operating in a population ecology framework, take a meaningful step by treating the density of an online community's niche as a function of overlapping resources. One might represent this idea in a Lotka-Volterra framework as below:

$$\frac{dN_i}{dt} = r_i(D_i - \frac{D_i^2}{K_i})$$

$$D_i = B_0 + B_1 T_i + B_2 U_i + B_3 L_i$$

Where D_i, the density of the niche of community i, is a linear function of three resource overlaps: T for topic, U for user and L for links. Forces of competition or mutualism acting on a given online community depend not on the total number of other online communities, but on the degrees to which a focal community shares resources with the other communities. Zhu, Chen, et al. (2014) assume that the relationship between these overlaps and density is homogeneous. They are studying density as an aggregate of resource overlaps, but they cannot show how these overlaps give rise to competitive or mutualistic relationships. They must believe that such relationships exist as the mechanism by which "density" leads to competition or mutualism, but they treat the relationships as latent, neither observing nor modeling them. My project builds on theirs by explicitly measuring and analyzing inter-community ecological relationships.

When organizational ecologists looked beyond trajectories of single organizational forms to consider how different organizational forms are related, they adopted community ecology, and replaced amorphous density with explicit treatment of ecological niches as a function of dyadic inter-organizational relationships (Monge et al., 2011). Community ecology doesn't ask "how does organizational growth and survival depend on density?" Instead it asks "how does growth and survival depend on the network of ecological relationships?" McPherson (1983), considering different forms of voluntary organizations, and Ruef (2000), studying the emergence of new organizational forms in the health care industry, both represent this in the Lotka-Volterra framework as below. As this model represents dynamics at the level of organizational forms, i and j index organizations of a given form, and a_{ij} represents the competition/mutualism coefficient between form i and form j.

$$\frac{dN_i}{dt} = r_i(N_i - \frac{N_i^2}{K_i} - \sum_{j=1,j\neq i}^{M} a_{ij}\frac{N_iN_j}{K_i})$$

The puzzling levels-of-analysis dichotomy between organizational and community ecology is bridged by a third strand of organization ecology, the one with the strongest influence on my proposed project. McPherson (1983)'s ecology of affiliation is similar to community ecology in that it doesn't assume that organizations are uniform, but, like population ecology, it theorizes at the organizational level. Through a long line of research McPherson and collaborators studied how dynamics pf competition over members who are distributed across demographic niches may drive evolution in communities of voluntary organizations.[3] McPherson (1983) calculates a matrix of

[3]Some of McPherson's studies analyze data at the level of organizational forms because they are limited by the use of survey samples. However, theorization is clearly at the organizational level. By McPherson and Rotolo (1996), McPherson and his colleagues completed a

competition coefficients between different organizational forms. McPherson and Ranger-Moore (1991) considers niche partitioning dynamics on the multi-dimensional space of demographics called "Blau space", and this is extended in (McPherson & Rotolo, 1996) to a turbulent social environment where organizational memberships that fall above or below an organization's long-run carrying capacity return to equilibrium by as organizations expand or shrink their niches. Following McPherson's analyses of ecological relationships between individual organizations, this project will analyze ecological relationships between different online communities.

Specifically, we can transpose the model above to the level of online communities. Each online community has a carrying capacity (K_i), and intrinsic growth rate (r_i). Ecological competition (mutualism) is captured by a_{ij}, which represents how much of organization i's potential niche is taken up (expanded) by online community j. The sign of a_{ij} indicates whether i is "helped" or "hurt" (in terms of membership size) by j, allowing us to study a set of qualitatively different types of ecological relationships: mutualism (both i and j are helped by the other), competition (both i and j are hurt by the other), amensalism (i is hurt but j is unaffected), commensalism (i is helped but j unaffected), and predation (i is hurt while j is helped) (Verhoef & Morin, 2010).

Drawing from McPherson's approach to study the ecological analysis of non-isomorphic online communities extends prior ecological work on online communities with a new focus on the kinds of relationships that are the central focus of community ecology (Verhoef & Morin, 2010). My dissertation will begin to develop a similar approach that accounts for the distinctive character of online communities. Utilizing granular behavioral trace data will also enable new applications of methods from bioecology to social scientific analysis. I will now turn to describe my plans for three sub-projects. The first focuses on the measurement of ecological relationships in online communities and their relationship to resource overlaps. The second builds on the first to consider co-evolutionary change processes at the level of ecological communities to test predictions about how ecological relationships will change over time and offer a test of community closure theory (Astley, 1985; Hawley, 1986). The third project (to take place contemporaneously with the first 2) will advance theoretical development using an agent based model to systematically consider how mechanisms at the level of individual give rise to

life-history survey that provided granular data on affiliations at the organizational level.

dynamics at the ecological and community-evolution levels. A short introduction and conclusion will complete the dissertation.

Data sources

I will study ecological populations of online communties on two platforms: Wikia and Reddit. Both 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 two platforms also have significant technological and cultural differences. Studying all both 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. I have already written software for working wiwith Wikia wikis and published a papers using it (TeBlunthuis et al., 2017).

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 wikis, subreddits are sustained by the contributions of groups of pseudonymous strangers.

I also seriously considered including analysis of stack exchange communities in this proposal. 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. Studying stack exchange could allow us to explore how ecological dynamics may differ under this system. However, adding a third dataset to this study will add too much complexity and work to fit within the scope of this project.

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.

Project A: Ecological relationships between communities

As discussed above, online communities frequently overlap in terms of users and content, but prior work offers competing conceptualizations or explanations of the effects of these overlaps (Datta et al., 2017; Tan, 2018; Wang et al., 2013; Zhu, Kraut, & Kittur, 2014). The first project's research question is "How does the growth and survival of online communities depend on their ecological communities?" I will answer it in terms of the ecological relationships between communities and of their overlapping human and topical resources using ecological framework identified above. This work will have practical significance to managers of online communities by helping them understand whether a community's success depends on environmental conditions.

Ecological relationships like mutualism and competition support long-run dynamics that influence community growth and survival. These relationships are necessarily mediated by resources like the time and effort of people working as moderators, creating new content, in other skilled and involved roles, as well as information inputs, legitimacy, and channels by which people discover the community including social networks and search engines (Wang et al., 2013; Zhu, Chen, et al., 2014). Specifically, mutualistic relationships depend on the production of resources that may be shared or spill over between communities. Concretely, this might mean attracting more users to a platform such as Reddit or popularizing a style of content or participation in a way that increases motivations for participation. On the other hand, competitive relationships are mediated by rival resources. For example, two communities might share overlapping contributors with limited time or attention for participation such that they tend to choose to participate in only one community but not the other.

Previous ecological studies of online communities in the population ecology framework focus on density-dependent growth and survival. Instead of attempting to identify ecological relationships between online communities, they treat resource overlaps as ecological forces on growth and survival. One way to see how this view is limited is to consider how overlaps of any kind of resource—whether rival or not—can only drive an ecological dynamic when the resource limits population growth (Verhoef & Morin, 2010). To see why, consider a simple hypothetical biological community with two species of rodent that eat the same food, occupy the same territory, and use the same kind of shelter. The only relevant difference between the two is that one is nocturnal while the other is not. If they are the only animals in the ecological community, they must enter into competition as their populations will eventually be limited by one or more of the scarce overlapping resources.

Now imagine that we introduce two predators (say an eagle and an owl) to the ecological community. One predator is nocturnal while the other is not. Say the predators are quite successful such that they limit the size of the population of both rodent species. Since the predators don't hunt the same food (the nocturnal predator only hunts the nocturnal prey and visa-versa), the rodents are not competing to escape the predators, and they are not competing over other resources either. Since their population sizes are limited by the predators, there is enough food, shelter, and territory to go around. Once the predators are introduced, resource overlaps no longer imply competition.

This is a general argument; it doesn't matter whether we are talking about rodents and predators or online communities. Identifying ecological relationships and relating them to overlapping resources will help us to find out which resources limit the growth of online communities. The relationships in the ecological community can drive what resources will limit a community's growth. Associations between categories of ecological relationships and resource overlaps will not only provide a grounded understanding of how "density" in resource overlaps appeared to lead to competition on Usenet and to mutualism on Wikia (Zhu, Chen, et al., 2014; Zhu, Kraut, & Kittur, 2014). Knowing what kinds of resource overlaps are most strongly associated with competition will suggest what kinds of resources are being competed over. Similarly, knowing what kinds of resources are most strongly associated with mutualism will suggest which of the resources that can limit growth can also spillover.

Hypotheses In this project I will test whether resources considered rival tend to mediate competitive relationships and inversely, whether resources considered non-rival tend to mediate mutualistic relationships. I will also test a prediction of the principle of competitive exclusion (Hawley, 1986).

> $\mathbf{H1_A}$: User overlap will be positively associated with competition.
>
> $\mathbf{H2_A}$: Content overlap will be positively associated with mutualism.
>
> $\mathbf{H3_A}$: Communities with high content overlap and high user overlap are even more likely to compete than communities with just high content overlap.

I will also test the relationship between ecological relationships and community longevity.

> $\mathbf{H4_A(a)}$: Communities in competitive relationships are less likely to survive.
>
> $\mathbf{H4_A(b)}$: Communities in mutualistic relationships are more likely to survive.

In addition to testing these theoretically driven hypotheses. This study will also provide a description of the two platform ecologies in terms of the relative frequencies of mutualistic, competitive, amensal, commensal, and "predetory" relationships.

Online communities depend on two types of resources: content and participation. Content is a nonrival resource because it is digital information that can be easily copied (Benkler et al., 2015). Therefore, we anticipate finding support for $\mathbf{H2_A}$ (overlapping content is likely to support mutualistic relationships). Similarly, $\mathbf{H1_A}$ (user overlap is likely to support competitive relationships) arises from ecological models of organizational resource constraint, wherein voluntary organizations operating in the same field seek members to engage in similar work and therefore compete for their time and effort (McCarthy & Zald, 1977, 2001; Minkoff, 1997; Soule & King, 2008). Similarly, scholars of online communities have used data from online discussion groups to show that discussion groups covering similar topics compete for participants (Wang et al., 2013).

There are reasons to doubt these hypotheses. Volunteer labor is different from many other resources in that organizations may generate interest in volunteering, thereby increasing the "supply" of volunteer effort (McCarthy

& Zald, 1977, 2001). Indeed, volunteer mobilization efforts by one organization may even generate positive "spillovers" in others (Marwell & Oliver, 1993; Meyer & Whittier, 1994). Such spillovers may be even more likely in the context of peer production because potential contributors can immediately utilize and re-purpose earlier contributions due to the systematic free licensing of work products in peer production that allow content reuse. If spillovers are widespread in the contexts we consider then we are likely to reject $\mathbf{H2_A}$ and $\mathbf{H3_A}$.

$\mathbf{H3_A}$ follows from the principle of competitive exclusion, one of the most influential ideas in bioecology (Verhoef & Morin, 2010). Essentially, strong and direct competition is unlikely to exist in nature because of the serious negative implications for survival. To avoid direct competition, organisms (and organizations) that both depend on the same resource are likely to differentiate from one another in their utilization of other resources (Carroll, 1985). This leads the environment to become resource partitioned. I will study resource partitioning in greater depth in project B, but project A provides an opportunity to test the mechanism of a multiplicative association between multiple resource overlaps and competition. Concretely, online communities with overlapping users will be less likely to compete if they have different content, and therefore online communities with overlapping users and overlapping content will be even more likely to compete.

Measures For all the complicated Bayesian modeling below, this study has only three analytic measures: textual overlap, author overlap, and active participants and I plan to closely follow Datta et al. (2017)'s measures of author and textual similarity to construct these measures.

Textual overlap: The textual overlap measure is the widely used TF-IDF approach. Following Datta et al. (2017) I will use standard NLP techniques to clean text and compute a TF-IDF vector for each community (this is the ratio of the number of times a word or phrase is used in a community to number of communities in which it is used). The cosign these vectors provides a measure of overlap. For Reddit I will apply the measure to comments and for Wikia I will apply it to the Wiki mainspace.

User overlap: Again following Datta et al. (2017) I will use an author-TF-IDF vector for each community (this is the ratio of the number of times a user account participates in community to number of communities in which it participates). For Wikia I will measure participation as edit sessions and for

Reddit I will measure it as comment-sessions (Geiger & Halfaker, 2013).

Active participants: I will use the widely established measure for active participation on a Wiki of five edits per month and analogously I will use five comments per month as my measure for active participation on Reddit.

I also considered adding measures of mechanisms for textual overlap such as a comparison of cross posting and linking submissions on Reddit, as well as characteristics of contributors such as roles, the use of bots, and contributor diversity. Such ideas are appealing in order to show that ecological approach is a *meta-theory* capable of incorporating social-psychological mechanisms. However, constructing and validating such measures and incorporating them into the already complicated models in Projects A and B is best left for future projects that can better engage with substantive theory about such mechanisms.

Analysis We will test these hypotheses using data from Wikia and Reddit using a vector autoregression approach. Vector autoregression models are a workhorse in bioecology because VAR(1) models (with a single autoregressive term) have a close relationship to Gopertz models (a popular alternative to the logistic model used above) (Certain, Barraquand, & Gårdmark, 2018). Certain et al. (2018) demonstrating using a simulation that even in the presence of unmodeled nonlinearities, VAR(1) models can reliably identify competition or mutualism between species in empirically realistic scenarios.

Vector autoregression models have also been widely adopted in the social sciences, particularly in political science and in macroeconomics (Box-Steffensmeier, 2014). Sims (1980) advanced VAR modeling in macroeconomics to address a problem in the field as an alternative to structural equation modeling (SEM), which required detailed specification of a large number of theoretical assumptions to identify.[4] At the time macroeconomists needed strong assumptions that were likely to be false or at least driven by embedded theoretical assumptions (sometimes close to those under debate) just to fit their models (Sims, 1980). VAR models don't require as many theoretical assumptions to identify, but are flexible enough to allow them.

Vector autoregression can be intuitively understood as a generalization of a one-dimensional auto-regressive AR models. But instead of predicting the current state of a single time-series variable as a function of its lag, the model

[4] For his work advancing a VAR based approach to macroeconomics Sims was awarded the 2011 Nobel memorial prize in economic sciences.

predicts the state of a vector time series variables as a function of the lag of every other variable in the system. The simplest vector autoregression model is a VAR(1) model:

$$N_t = \beta_0 + \beta_1 N_{t-1} + \Sigma$$

In the setting of the ecology of online communities N_t is a vector of the number of active participants to each online community under study at time t; β is matrix and its entries can be interpreted as competition/mutualism coefficients between the communities. Now, as is common in other settings where VAR models are used, there might be different correlations between the different communities at different time scales. The model can be extended to consider p lags (Box-Steffensmeier, 2014).

$$N_t = \beta_0 + \sum_{i=1}^{p} \beta_i N_{t-i} + \Sigma$$

This model can be interpreted in terms of long-run competition and mutualism using the cumulative impulse response function, which shows how the model predicts an exogenous increase in participation in one community will cause changes in participation in every other community (Box-Steffensmeier, 2014). It can be fit using ordinary least squares.

I have already experimented with this model in some preparatory work. As is commonly done, I chose p using the Bayesian information criteria (BIC). I analyzed the interactions two different communities on Reddit. I analyze weekly time series of the count of active contributors (commenting >3 times per week) for entire participation history over which all the in the group communities is active. I compared the two different subreddits about Seattle: /r/Seattle and /r/SeattleWA.

The story of /r/SeattleWA and /r/Seattle is one of ecological competition between communities with (at this point I assume) a high degree of topical and user overlap. As a small-time participant in these communities I have first-hand knowledge of the rise of /r/SeattleWA in 2016.[5] The short story is that the /r/Seattle community rebelled against a heavy-handed moderator and most members of /r/SeattleWA defected to organize a new community

[5] Also see https://www.seattleweekly.com/news/seattles-reddit-community-is-big-active-and-at-war-with-itself/ for media coverage.

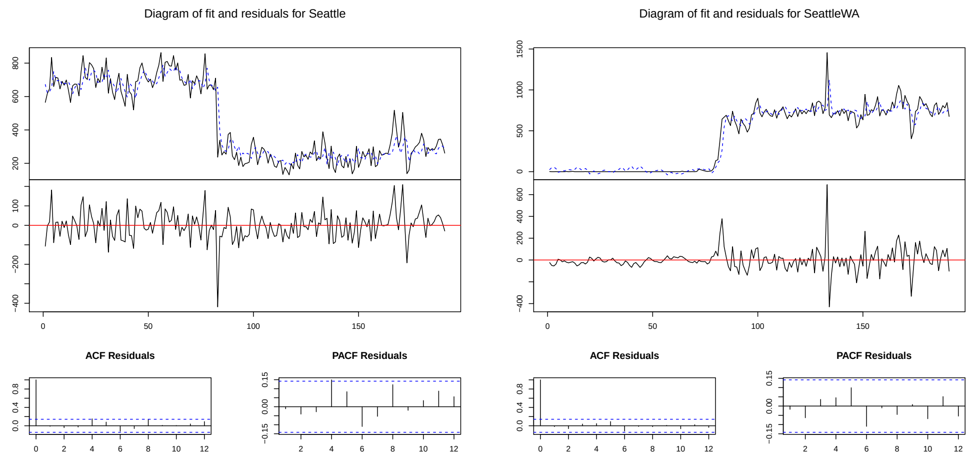


Figure 1: Fit of VAR model to data from /r/SeattleWA and /r/Seattle. The top plots show the data in black and the model fit in blue. The plot below that shows the residuals. The small plots at the bottom are diagnostic plots showing that the residuals are not serially correlated as consistent with model assumptions.

at /r/SeattleWA. Thereafter the two communities co-exist, with /r/Seattle having around 800 active weekly contributors and /r/Seattle around 300.

The impulse response function of this model indicates that the two communities are competitors. The model predicts that an exogenous shock to /r/Seattle causing a cumulative increase in 400 participants over 10 weeks is expected to cause a decrease of around 200 participants in /r/SeattleWA. Similarly, an exogenous shock to /r/SeattleWA causing an increase of around 500 participants over 10 weeks will cause a decrease of around 200 participants to /r/Seattle. This illustrates how I will identify competitive or mutualistic relationships.

The next methodological question is how I will relate resource overlaps to ecological relationships. Now, the simplest approach might a 2-stage process where I first VAR model, next interrogate it to detect ecological relationships and finally use a regression model to test my hypotheses about mutualistic and competitive relationships. This approach has the limitation that the hypotheses tests will not account for uncertainty in the VAR model. A more advanced approach to solve this problem is to use a Bayesian hierarchical VAR model Canova (2011). A Bayesian approach makes it much easier to specify and fit this kind of model. The idea is to jointly estimate both stages in one step by nesting them. Here is a rough sketch of this kind of model to give you the

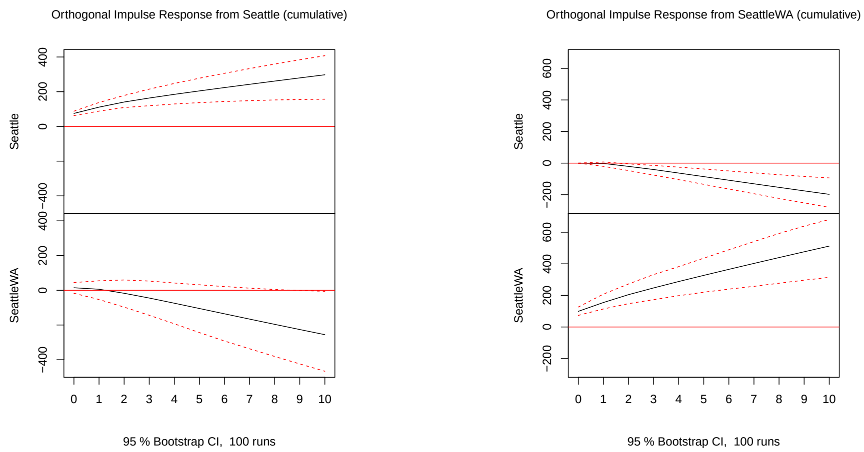


Figure 2: Cumulative impulse responses of a VAR model fit to data from /r/SeattleWA and /r/Seattle. The left plot shows how the model predicts the two communities would be effected by an exogenous increase in participation in /r/Seattle and the right shows the response from a comparable shock to /r/SeattleWA. Qualitatively, the plot shows evidence of competition between the two communities.

idea. If you are familiar with multilevel Bayesian regression the generalization to vector regressions is straightforward.

$$N_t = \beta_0 + \sum_{k=1}^{p} \beta_k N_{t-i} + e_t$$

The interesting part is the prior for β_k—I'm leaving out the others. Let b_{ij}^k be an element of β_k and let X_{ij}^k be predictors of b_{ij}^k.

$$\begin{gathered}
\text{if } i = j \quad b_{ij}^k \sim N(0, \sigma_b^k) \\
\text{if } i \neq j \quad b_{ij}^k = \alpha^k X_{ij}^k \\
\alpha^k \sim N(0, \sigma_\alpha^k)
\end{gathered}$$

The diagonal elements of β model the communities intrinsic growth and don't depend on any covariates. However, the off-diagonal element like b_{ij}^k are a linear function of X_{ij}^i, how much of i's resources overlap with j. Interpreting α^k, the coefficients for this relationship, will allow me to test the hypotheses about the average relationship between resource overlaps and ecological relationships. Of course, now the model is much more complex, and it might be difficult to fit. I think I will be able to fit this model (or a similar model) using Stan.

There is one more detail to point out. The number of parameters in the model grows quite rapidly—quadratically with respect to the number of online communities under study. This could make the model challenging to estimate well for a large number of communities. The hierarchical model above will help with this problem, but it will likely be necessary to use a shrinkage prior for b_{ij}^k instead of the Normal prior above. These priors assume that the number of nonzero elements of β is low. The Minnesota prior is a popular prior in macroeconomic VAR models (Canova, 2011).

A final point regards the test of $H4_A$. The hierarchical VAR model doesn't predict survival directly, but it will be straightforward to test for significant correlations between survival and ecological relationships it detects. Model uncertainty can be incorporated into the tests by bootstrapping using the posterior draws of β.

I will evaluate the model's predictive performance at out-of-sample forecasting. In addition to testing the theoretically motivated hypotheses above, this paper will discuss applications of the model for time series forecasting by predicting the futures (using posterior predictive intervals) of some interesting communities. I will also interpret β by visualizing it as a network of ecological relationships.

Project B: Evolutionary change processes in online community platforms

My second project continues developing an ecological theory of online communities by introducing analysis of change processes at the level of the ecological community (Aldrich & Ruef, 2006).[6] Prior work suggests two ways online communities respond to competition. The first is to select a niche with a favorable environment (Zhu, Chen, et al., 2014). Evidence from a survey of community founders shows that they often choose topics for new communities based on intuitions to avoid competition with existing communities (Foote, Gergle, & Shaw, 2017). The second strategy is adaptation. 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

[6] My thinking about Project A is much more advanced compared to the others

as they sought to build niches in a shifting resource space (Hill, 2013).

These two different strategies for operating in a competitive environment exemplify different types of change processes: selection and adaptation (Ven & Poole, 1995). Change through selection is Darwinian evolution while change through adaptation process is Lamarkian (Hannan & Freeman, 1989). The selection change process is driven by the chances that a community will survive in a given environment. If and the adaptation change process is driven by the ability of rational actors to shift the niche positions of their communities to more favorable local environments (McPherson & Rotolo, 1996).

The population ecology of organizations assumes structural inertia: organizations are not very adaptable (Hannan & Freeman, 1989). Similarly, population ecology in biology focuses on populations of individual species on timescales not historically believed to support evolution. Community ecology, on the other hand, is deeply interested in questions of how relationship between ecological relationships such as mutualism and competition drive the course of evolutionary change processes (Astley, 1985). It looks not only to explain the survival of species but also the origins of strong symbioses such as lichens or between acacia trees and ants (Verhoef & Morin, 2010).

Over time, mutualistic and competitive relationships in conditions of evolutionary change (by either process) will lead to changes in the environment and both lead to increasing specialization. Driven by the principle of competitive exclusion, competitive relationships will become weaker over time as competitors partition resources to avoid direct competition. On the other hand, mutualistic relationships create an environment beneficial to both species, and in an environment of scarcity may become essential for survival. Feedback between the success of the mutualists and the strength of their relationship leads to specialization wherein two communities become increasingly tightly coupled over time. Both these processes together stabilize the entire ecosystem, such that the ecological community reaches a state of *closure* where new ecological relationships are unlikely to form and established ones are unlikely to fade (Astley, 1985).

Hypotheses

The first hypothesis tests that an evolutionary change process is operating on ecological relationships, leading to resource partitioning and niche construction.

$\mathbf{H1_B(a)}$: Competitive relationships become weaker as an online community ecosystem ages.

$H1_B$(b) : Mutualistic relationships become stronger as an online community ecosystem ages.

The second hypothesis tests for evidence that adaptation is a mechanism for resource partitioning and niche construction.

$H2_B$(a) : Online communities with greater niche drift are likely to be hurt less by competitive relationships over time.

$H2_B$(b) : Online communities with greater niche drift are likely to benefit more from fewer mutualistic relationships over time.

The third hypotheses tests whether ecological communities "close", becoming stable over time.

$H3_B$: As online communities age, their ecological relationships are less likely to change.

Measures and Analysis

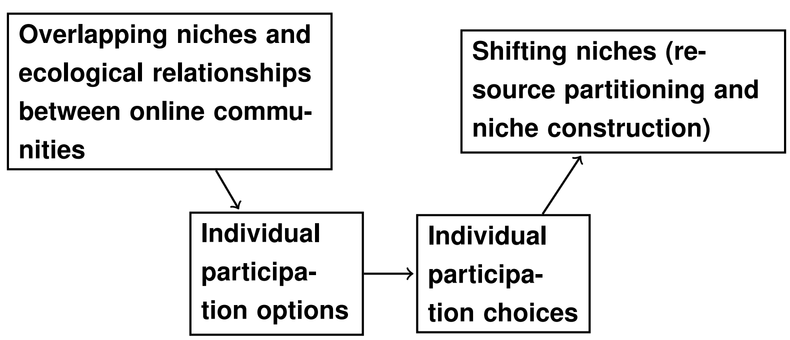
The analytic strategy for this project follows closely on that of project A. The major extension is to specify the model so as to allow competition/mutualism coefficients to vary over time. Time varying coefficients are an advanced technique, but have been around for a long time, and applied to large VAR models before (Box-Steffensmeier, 2014; Koop & Korobilis, 2013). I will operationalize *niche drift* by measuring change in user and topic similarity measures within communities over time.

Project C: An Agent Based Model of Online Community Ecology

So far this project has considered organization at the macro level. I have discussed how ecological theory applied to the context of online communities predicts that patterns of content and participant overlap between online communities can lead to ecological interdependence. This interdependence over time, in an environment of scarcity where communities may be able to adapt to shift their niches over time, can lead to predictable arrangements of increasing specialization and stability at the ecological level. Implicit in all of this is of course the decisions of human actors who choose in what communities (if any) in which they will participate, what content they will contribute, and whether they will strike out and attempt to form a new community. Working in a rational choice framework, this project will advance an agent based model to consider how a parsimonious description of an individual decision making process can lead to empirically observed macro level patterns.

According to the sociologist James Coleman, the task of social theory is to explain change of higher-order macro phenomena like social systems (Coleman, 1990). Yet because society is made of people, a macro-level phenomena can never be a direct cause of another macro-level phenomena. Systemic change is always mediated at the micro level—through the choices of human actors (Opp, 2011). However, explicating theories that trace action at the macro level through the micro level is empirically difficult both because it requires large scale data about individual choices and also because of the difficulty of specifying mechanisms that connect the levels that go beyond straightforward aggregation (Opp, 2011). Coleman popularized diagrammatic representations of theory connecting macro level changes through a micro-level path known as "Coleman's boat" and these are useful for conceptualizing causal theories of systemic change.

Here's a Coleman's boat diagramming the relationship between ecological relationships and the nich locations of online communities:



This represents a theoretical idea that the locations of online communities in the resource environment (e.g. the pools of contributors they draw on and the scope of their content) depends on the choices that individual people make about how they will participate in online communities. These choices are constrained (not exclusively) by their "options for participation," the menu of online communities and ways of interacting in them that are available. People cannot participate however they like. They can only participate with the communities that actually exist, or they can start a new one. But the communities that already exist might have overlapping niches and these influence the menu of participation options that people will consider.

As a hypothetical example, Ada wants to connect with other Seattle residents online. She will have to choose whether to participate in /r/Seattle, /r/SeattleWA or both communities. Say she wants to start a discussion about being a bike commuter during snowy weather. The two communities' niches partially overlap so as to make this decision a matter of personal preference. On one hand /r/SeattleWA has a larger community and so Ada thinks she will be more likely to start a conversation about this topic in /r/SeattleWA compared to /r/Seattle. But threads on /r/SeattleWA tend to devolve into tangential heated political discussions. Ada might post in both communities, but at the cost of having to pay attention and participate in two discussions at once. So if Ada doesn't want to manage two discussions and wants to avoid the political debate she is likely to choose to participate in /r/Seattle instead of /r/SeattleWA. Notice that because Ada's decision was based on a prior difference between /r/Seattle and /r/SeattleWA her action will reinforce this difference. If Ada's experience is generalizable then other people who prefer to avoid politics will come to participate in /r/Seattle instead of /r/SeattleWA. This illustrates how the initial conditions of partial niche overlap between two communities influence the choices of individuals choosing how they will participate in online communities and in aggregate these choices can collectively lead to resource partitioning as communities with overlapping niches drift apart.

Such processes by which heterogeneous individual preferences lead to macro-level structural changes have successfully been studied through the use of agent based models (ABM). Perhaps most famously, Schelling (1971) developed a model (quite analogous the scenario above!) in which even a mild preference for racial homophily, if sufficiently widespread, can lead to stable patterns of racial segregation. Similarly, DellaPosta, Shi, and Macy (2015) argue that the great number of correlations between mundane personal preferences about music, coffee, or fashion and political stance should not be attributed to inherent political qualities of music, coffee, or fashion. Using an ABM they show that feedback between social influence and homophily can lead to clustering of preferences (including political preferences) in social networks. ABMs have also been used in ecological studies in the tradition of McPherson to model how organizations that isolate their members to avoid competition diminish their chances for recruitment and that to survive organizations should adopt strategies with a level of turnover appropriate to local environments (Shi et al., 2017).

For this project I will study agent based models and develop and analyze one tailored to the unique context of online communities, where individuals typically participate in multiple communities and can attempt to start new ones. I am already collaborating on Foote (2018), a project using agent based model to explain the unequal distribution of participants across online communities based on the idea that people who join communities that subsequently grow obtain increased benefits from their participation (Kraut et al., 2012). For Project C I will instead focus on how individuals with preferences make choices about which communities they will join. Here is a sketch of the components of this model:

Time: This is a discrete time model in which agents take turns making decisions.

Resource space: As (McPherson & Ranger-Moore, 1991) considers individuals as located in a high-dimensional demographic Blau space, I'll consider a high-dimensional space of topical and cultural qualities of online communities. However, agents don't have just one location in this space, instead they have multiple "interests" (i.e. they might want to contribute to a wiki about American Idol and a wiki about Pokemon). One interesting parameter of the model will control how these interests are distributed in resource space. Are they uniform? Are they clustered?

Online communities: Online communities have niches in the resource space. These niches are defined as a density over the locations of their participants in the resource space.

Agents: Each agent has a limited capacity for engaging in online communities. Agents have a decision making model where they can use their capacity to engage in existing online communities, create a new community, or do something else besides engaging with an online community. The decision making model follows a expectancy-value framework in which agents seek to maximize their personal enjoyment over time.

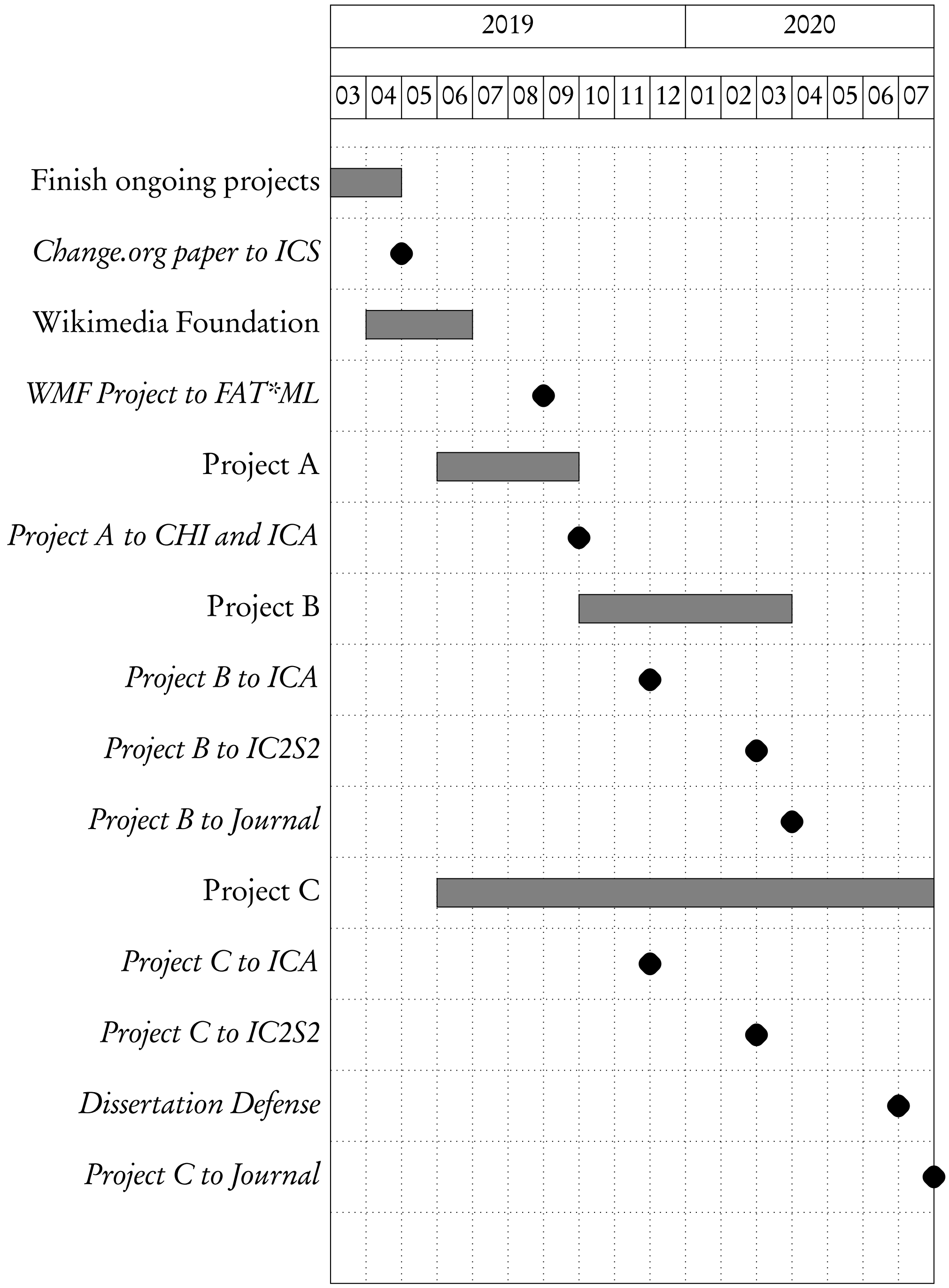
Engaging in existing communities: When agent chooses to engage in a community receive a benefit that increases logistically with group size and also increases as the mass of the community grows closer to the agent's nearest interest in the resource space.

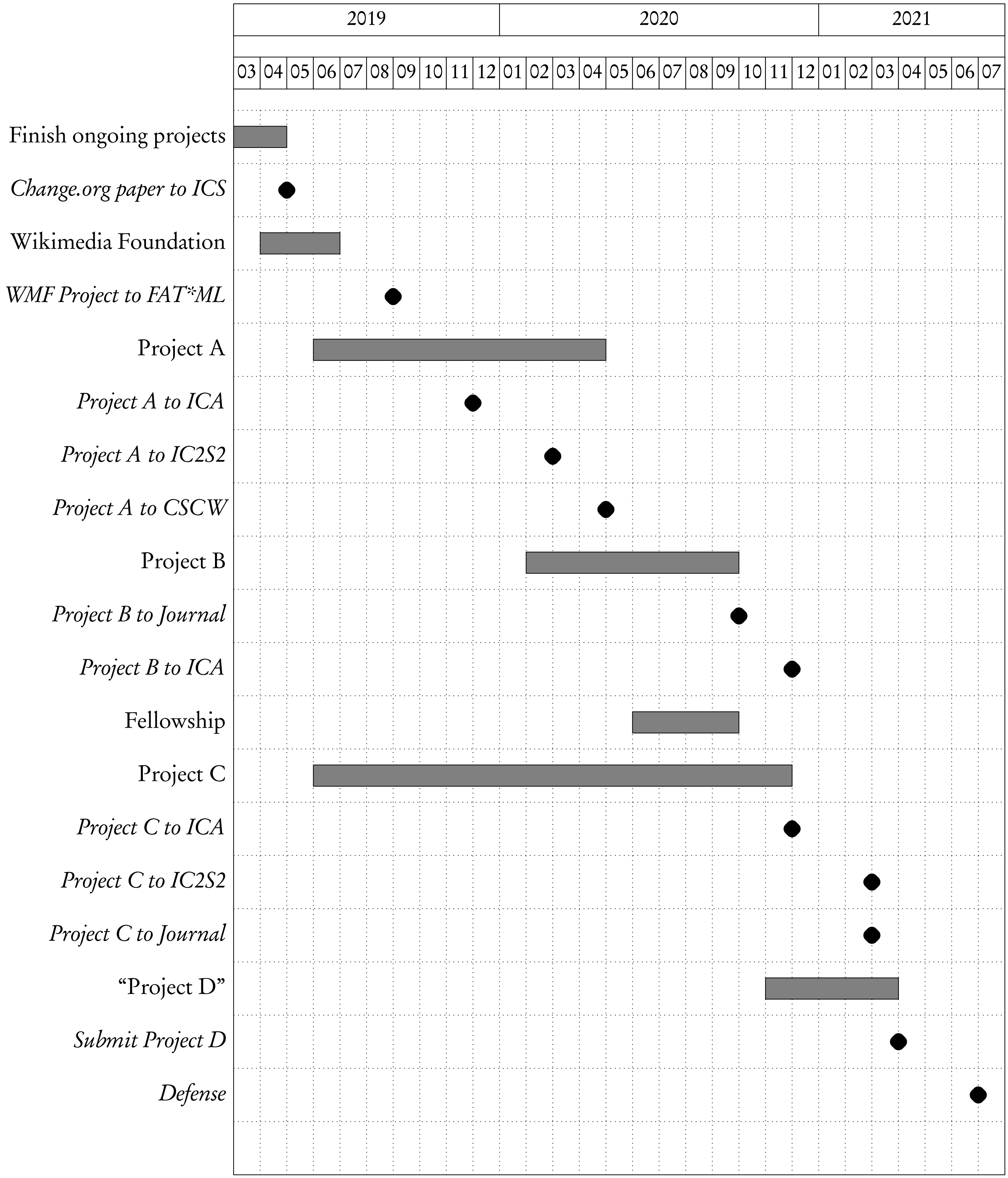
Creating a new community: Agents may choose to create a new community by paying a cost from the capacity for engagement. If they create a community they become its "leader" and receive a bonus benefit when others participate in the community. Agents will choose from among their locations in the resource space to create a new community.

This is just a rough sketch of the kind of model I will build. My goal is try to find a parsimonious model with interpretable parameters that is grounded in empirical research on the processes by which individuals choose to participate in online communities or create new ones. I will evaluate the model based on its ability to produce empirical patterns at high levels of analysis such as the participation levels, community longevity, and the number of communities in which individuals participate. Untested predictions of the model will provide motivations for future empirical research. I think of Project C as a theory paper that will provide more opportunity to review and synthesize a strong theoretical statement about ecology and online communities. Is there a mechanisms for mutualism here?

Timeline

I think I will work on Project C gradually on an ongoing basis throughout the next year. My goal is to submit Project A to CHI 2020, and Project B and C to Communication or other social science journals. There is a good chance that the following 16 month timeline will be too ambitious and that I will take a 6th year to complete the MA/PhD program. In that scenario I think it's likely that I'll do a fellowship during summer 2020 and add a Project D (maybe an empirical investigation of micro level mechanisms for community specialization). A 28 month timeline is on the next page.


2019
2020
03
04
05
06
07
08
09
10
11
12
01
02
03
04
05
06
07
Finish ongoing projects
Change.org paper to ICS
Wikimedia Foundation
*WMF Project to FAT*ML*
Project A
Project A to CHI and ICA
Project B
Project B to ICA
Project B to IC2S2
Project B to Journal
Project C
Project C to ICA
Project C to IC2S2
Dissertation Defense
Project C to Journal


2019
2020
2021
03
04
05
06
07
08
09
10
11
12
01
02
03
04
05
06
07
08
09
10
11
12
01
02
03
04
05
06
07
Finish ongoing projects
Change.org paper to ICS
Wikimedia Foundation
*WMF Project to FAT*ML*
Project A
Project A to ICA
Project A to IC2S2
Project A to CSCW
Project B
Project B to Journal
Project B to ICA
Fellowship
Project C
Project C to ICA
Project C to IC2S2
Project C to Journal
"Project D"
Submit Project D
Defense

References

Aldrich, H., & Ruef, M. [M.]. (2006). *Organizations Evolving* (2nd). Thousand Oaks, CA: SAGE Publications.

Astley, W. G. (1985). The Two Ecologies: Population and Community Perspectives on Organizational Evolution. *Administrative Science Quarterly*, *30*(2), 224–241. 00620. doi:10.2307/2393106

Axelrod, R., & Hamilton, W. D. (1981, March 27). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. doi:10.1126/science.7466396

Benkler, Y., Shaw, A., & Hill, B. M. (2015). Peer production: A form of collective intelligence. In T. W. Malone & M. S. Bernstein (Eds.), *Handbook of Collective Intelligence* (pp. 175–204). Cambridge, MA: MIT Press.

Box-Steffensmeier, J. M. (2014). *Time series analysis for the social sciences*. OCLC: 879601718.

Butler, B. S. (2001, December 1). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, *12*(4), 346–362. doi:10.1287/isre.12.4.346.9703

Canova, F. (2011, September 19). *Methods for Applied Macroeconomic Research*. Princeton University Press.

Carroll, G. R. (1985, May 1). Concentration and specialization: Dynamics of niche width in populations of organizations. *American Journal of Sociology*, *90*(6), 1262–1283. doi:10.1086/228210

Certain, G., Barraquand, F., & Gårdmark, A. (2018, September 1). How do MAR(1) models cope with hidden nonlinearities in ecological dynamics? *Methods in Ecology and Evolution*, *9*(9), 1975–1995. doi:10.1111/2041-210X.13021

Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017, December). You Can'T Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proc. ACM Hum.-Comput. Interact. 1*, 31:1–31:22. doi:10.1145/3134666

Coleman, J. S. (1990). *Foundations of social theory*. OCLC: 781975612. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Datta, S., Phelan, C., & Adar, E. (2017, December 6). Identifying misaligned inter-group links and communities. *Proceedings of the ACM on Human-Computer Interaction, CSCW*, *1*, 1–23. doi:10.1145/3134672

DellaPosta, D., Shi, Y., & Macy, M. (2015, March). Why Do Liberals Drink Lattes? *American Journal of Sociology*, *120*(5), 1473–1511. doi:10.1086/681254

Foote, J. (2018, July). *An Agent-Based Model of Online Community Joining*.

Foote, J., Gergle, D., & Shaw, A. (2017). Starting online communities: Motivations and goals of wiki founders. In *Proceedings of the 2017 CHI Con-*

ference on Human Factors in Computing Systems (CHI '17) (pp. 6376–6380). doi:10.1145/3025453.3025639

Geiger, R. S., & Halfaker, A. (2013). Using edit sessions to measure participation in Wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 861–870). ACM. Retrieved March 6, 2016, from http://dl.acm.org/citation.cfm?id=2441873

Graeff, E., Stempeck, M., & Zuckerman, E. (2014, January). The battle for 'Trayvon Martin': Mapping a media controversy online and off-line. *First Monday*, *19*(2). doi:10.5210/fm.v19i2.4947

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013, May 1). 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. doi:10.1177/0002764212469365

Hannan, M. T., & Freeman, J. (1977). The population ecology of organizations. *American Journal of Sociology*, *82*(5), 929–964. doi:10.2307/2777807

Hannan, M. T., & Freeman, J. (1984, April). Structural inertia and organizational change. *American Sociological Review*, *49*(2), 149. doi:10.2307/2095567

Hannan, M. T., & Freeman, J. (1989). *Organizational ecology* (1st ed.). Cambridge, MA: Harvard University Press.

Hawley, A. H. (1986). *Human ecology: A theoretical essay*. OCLC: 993363851. Chicago; London: University of Chicago Press.

Healy, K., & Schussman, A. (2003). *The ecology of open-source software development*. Working Paper. Retrieved October 28, 2016, from http://kieranhealy.org/files/drafts/oss-activity.pdf

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*. PhD Dissertation. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Hill, B. M., & Shaw, A. D. (2019). Studying Populations of Online Communities. In W. B. Foucault & S. González-Bailón (Eds.), *The Handbook of Networked Communication* (p. 25). New York, New York: Oxford University Press.

Kiene, C., Shaw, A., & Hill, B. M. (2018). Managing organizational culture in online group mergers. *Proceedings of the ACM on Human-Computer Interaction*, *2*, 89:1-89–21. doi:10.1145/3274358

Koop, G., & Korobilis, D. (2013, December 1). Large time-varying parameter VARs. *Journal of Econometrics*. Dynamic Econometric Modeling and Forecasting, *177*(2), 185–198. doi:10.1016/j.jeconom.2013.04.007

Kraut, R. E., Resnick, P., & Kiesler, S. (2012). *Building successful online communities: Evidence-based social design*. Cambridge, MA: MIT Press.

Lee, S., & Monge, P. (2011). The coevolution of multiplex communication networks in organizational communities. *Journal of Communication*, *61*(4), 758–779. doi:10.1111/j.1460-2466.2011.01566.x

Marwell, G., & Oliver, P. (1993). *The critical mass in collective action: A micro-social theory*. Cambridge, MA: Cambridge University Press.

McCarthy, J. D., & Zald, M. N. (1977). Resource mobilization and social movements: A partial theory. *The American Journal of Sociology*, *82*(6), 1212–1241.

McCarthy, J. D., & Zald, M. N. (2001). The enduring vitality of the resource mobilization theory of social movements. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 533–565). Handbooks of Sociology and Social Research. doi:10.1007/0-387-36274-6_25

McMahon, C., Johnson, I. L., & 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)* (pp. 142–151). Palo Alto, California: AAAI. Retrieved August 10, 2017, from http://brenthecht.com/publications/icwsm17_googlewikipedia.pdf

McPhee, R. D., & Poole, M. S. (1981, December 1). Mathematical Modeling in Communication Research: An Overview. *Annals of the International Communication Association*, *5*(1), 159–191. doi:10.1080/23808985.1981.11923844

McPherson, J. M. (1983). An ecology of affiliation. *American Sociological Review*, *48*(4), 519–532. doi:10.2307/2117719

McPherson, J. M., & Ranger-Moore, J. R. (1991, September). Evolution on a Dancing Landscape: Organizations and Networks in Dynamic Blau Space. *Social Forces*, *70*(1), 19–43. doi:10.2307/2580060

McPherson, J. M., & Rotolo, T. (1996). Testing a Dynamic Model of Social Composition: Diversity and Change in Voluntary Groups. *American Sociological Review*, *61*(2), 179–202. doi:10.2307/2096330

Meyer, D. S., & Whittier, N. (1994, May 1). Social Movement Spillover. *Social Problems*, *41*(2), 277–298.

Minkoff, D. C. (1997, October 1). The sequencing of social movements. *American Sociological Review*, *62*(5), 779–799. doi:10.2307/2657360

Monge, P. R., Lee, S., Fulk, J., Frank, L. B., Margolin, D., Schultz, C., … Weber, M. (2011). Evolutionary and ecological models for organizational communication. *Advancing research in organizational communication through qualitative methodology, Management Communication Quarterly*, *25*(1), 26–34.

Olzak, S., & Uhrig, S. C. N. (2001, October). The ecology of tactical overlap. *American Sociological Review*, *66*(5), 694. doi:10.2307/3088954

Opp, K.-D. (2011, January 25). Modeling Micro-Macro Relationships: Problems and Solutions. *The Journal of Mathematical Sociology*, *35*(1-3), 209–234. doi:10.1080/0022250X.2010.532257

Ruef, M. [Martin]. (2000, November 1). The emergence of organizational forms: A community ecology approach. *American Journal of Sociology*, *106*(3), 658–714. doi:10.1086/318963

Schelling, T. C. (1971, July 1). Dynamic models of segregation. *The Journal of Mathematical Sociology*, *1*(2), 143–186. doi:10.1080/0022250X.1971.9989794

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

Shaw, A., & Hill, B. M. (2014). Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication*, *64*(2), 215–238. doi:10.1111/jcom.12082

Shi, Y., Dokshin, F. A., Genkin, M., & Brashears, M. E. (2017, April 1). A Member Saved Is a Member Earned? The Recruitment-Retention Trade-Off and Organizational Strategies for Membership Growth. *American Sociological Review*, *82*(2), 407–434. doi:10.1177/0003122417693616

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

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, *48*(1), 1–48. doi:10.2307/1912017

Soule, S. A., & King, B. G. (2008, May). Competition and resource partitioning in three social movement industries. *The American Journal of Sociology*, *113*(6), 1568–1610. doi:10.1086/587152

Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009). The singularity is not near: Slowing growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (pp. 1–10). WikiSym '09. doi:10.1145/1641309.1641322

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)* (pp. 395–404). Palo Alto, California: AAAI. Retrieved from https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17811

Tan, C., & Lee, L. (2015). All Who Wander: On the Prevalence and Characteristics of Multi-community Engagement. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1056–1066). WWW '15. doi:10.1145/2736277.2741661

TeBlunthuis, N., Shaw, A., & Hill, B. M. (2017). Density dependence without resource partitioning: Population ecology on Change.org. In *Companion of the 2017 ACM Conference on Computer Supported Coopera-*

tive Work and Social Computing (pp. 323–326). CSCW '17 Companion. doi:10.1145/3022198.3026358

TeBlunthuis, N., Shaw, A., & 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)* (355:1–355:7). doi:10.1145/3173574.3173929

Ven, A. H. V. D., & Poole, M. S. (1995, January 7). Explaining Development and Change in Organizations. *Academy of Management Review*, *20*(3), 510–540. doi:10.5465/AMR.1995.9508080329

Verhoef, H. A., & Morin, P. J. (2010). *Community ecology: Processes, models, and applications*. OCLC: 876676566. Oxford: Oxford University Press.

Vincent, N., Johnson, I., & 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* (566:1–566:13). CHI '18. doi:10.1145/3173574.3174140

Wang, X., Butler, B. S., & Ren, Y. (2013). The impact of membership overlap on growth: An ecological competition view of online groups. *Organization Science*, *24*(2), 414–431. doi:10.1287/orsc.1120.0756

Worster, D. (1994). *Nature's economy: A history of ecological ideas*. OCLC: 855524849. Cambridge; New York, NY, USA: Cambridge University Press. Retrieved November 10, 2018, from http://site.ebrary.com/id/10740550

Zhu, H., Chen, J., Matthews, T., Pal, A., Badenes, H., & Kraut, R. E. (2014). Selecting an effective niche: An ecological view of the success of online communities. (pp. 301–310). 00001. doi:10.1145/2556288.2557348

Zhu, H., Kraut, R. E., & 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)* (pp. 281–290). doi:10.1145/2556288.2557213