Niche Dynamics in Complex Online Community Ecosystems

Nathan TeBlunthuis

University of Texas at Austin
School of Information
nathante@utexas.edu

Abstract

Online communities are important organizational forms where members socialize and share information. Curiously, different online communities often overlap considerably in topic and membership. Recent research has investigated competition and mutualism among overlapping online communities through the lens of organizational ecology; however, it has not accounted for how the nonlinear dynamics of online attention may lead to episodic competition and mutualism. Neither has it explored the origins of competition and mutualism in the processes by which online communities select or adapt to their niches. This paper presents a large-scale study of 8,806 Reddit communities belonging to 1,919 clusters of high user overlap over a 5-year period. The method uses nonlinear time series methods to infer bursty, often short-lived ecological dynamics. Results reveal that mutualism episodes are longer lived and slightly more frequent than competition episodes. Next, it tests whether online communities find their niches by specializing to avoid competition using panel regression models. It finds that competitive ecological interactions lead to decreasing topic and user overlaps; however, changes that decrease such niche overlaps do not lead to mutualism. The discussion considers future designs for online community ecosystem management.

Introduction

Online communities are increasingly important sites where people organize to meet their members' various needs to produce and consume information goods, provide social support (De Choudhury and De 2014), engage in collective and political action (Benkler et al. 2013; Choudhury et al. 2016; Krafft and Donovan 2020), make sense of the world, and connect with each other (Benkler 2006; Lampe et al. 2010). Individuals often belong to overlapping online communities that have suprisingly similar members and topics (Datta, Phelan, and Adar 2017; Tan and Lee 2015; TeBlunthuis et al. 2022). Why are there so many overlapping online communities? How do they come to have the topics that they do? Why do overlapping sets of users so often participate in communities about superficially similar topics (Datta, Phelan, and Adar 2017; TeBlunthuis et al. 2022)?

Interdisciplinary scholarship has explored these questions through the lens of organizational ecology (TeBlunthuis and Hill 2022; Wang, Butler, and Ren 2012; Zhu et al. 2014; Zhu, Kraut, and Kittur 2014). Organizational ecology considers the role of environmental and relational forces in the development of organizations and industries (Aldrich and Ruef 2006; Hannan and Freeman 1989; Hawley 1986; McPherson 1983; van de Ven and Poole 1995). Early studies in this vein used density dependence theory to relate the overlaps in community membership to growth and survival (Wang, Butler, and Ren 2012; Zhu et al. 2014; Zhu, Kraut, and Kittur 2014). Recently, TeBlunthuis and Hill (2022) introduced an alternative approach using time-series analysis to directly infer competitive and mutualistic relationships; however, their analysis is based on an unrealistic assumption in the online community context. They modeled intercommunity ecological relationships as static over time. However, the dynamics of attention online are often bursty (Ratkiewicz et al. 2010).

This study addresses this limitation using nonlinear time-series methods to infer competitive and mutualistic interactions between online communities that can vary over time. Specifically, it uses the S-Map model developed by ecologists of biological systems for this purpose (Cenci, Sugihara, and Saavedra 2019; Sugihara et al. 1994; Sugihara and May 1990). TeBlunthuis and Hill (2022) found that mutualism was much more common than competition; other ecological studies of online communities have similarly pointed to specialization and mutualism as crucial for online community success (TeBlunthuis et al. 2022; Zhu, Kraut, and Kittur 2014). Therefore, this study tests the hypothesis that *mutualism will be more common than competitive*.

This study also tests a model inspired by theories of specialization and adaptation in organizational ecology to explain how online communities find their *niches* and how highly overlapping online communities develop (Baum and Shipilov 2006; Dobrev, Kim, and Carroll 2002). Competition and mutualism might result from niche overlap, but might also cause communities to shift their niches. This study investigates such feedback between niche and ecological dynamics by combining the S-Map's measures of ecological interactions with longitudinal measures of user and topic similarity in a panel data analysis of 48,484 relationships between 8,806 online communities on Reddit over a 5 year time span. If online communities adapt in response to competition, *then more competitive ecological interactions* *should predict decreases in overlap*. If specialization helps communities avoid competition, then *decreases in overlap should predict increases in mutualism*.

This work contributes to the social science of online communities by (1) inferring time-varying nonlinear ecological dynamics, (2) shedding new light on findings from prior work, (3) engaging the organizational ecology concepts of structural intertia and competitive exclusion to theorize the role of adaptation in producing the widespread mutualism observed in prior studies and, (4) testing the resulting hypotheses to (5) synthesize insights for leaders and designers of online communities. The nonlinear time-series analysis reveals that ecological relationships tend to happen in bursty episodes and that mutualism has similar strength to competition; however mutualism happens more often and for longer durations. The panel models of niche adaptation found that online communities in competition tend to reduce their overlaps; however such changes lead to increases, not decreases in competition. The latter result is consistent with previous research findings that online communities tend to become oligarchical and rigid (Hill and Shaw 2019; TeBlunthuis, Shaw, and Hill 2018), and that organizational change often has negative outcomes (Hannan and Freeman 1989). Online community designers and leaders can attempt to manage online ecosystems by filling ecological roles. Feed or recommendation algorithms can be designed to adapt existing communities' niches by purposfully nudging individuals. That said, given the challenges of organizational change, a more pragmatic approach is likely to building new communities as spin-offs or offshoots from existing ones.

Background

Interactions Between Changing Online Communities

Online communities are a dynamic, growing, and increasingly important form of social organization. Peer production communities such as Wikipedia and open-source software projects have produced tens of billions of dollars' worth of software and encyclopedic documentation (Benkler, Shaw, and Hill 2015). Millions of members of smaller communities on platforms such as Reddit, Facebook, and Discord organize political mobilizations including disinformation campaigns and protest movements (Benkler et al. 2013; Choudhury et al. 2016; Krafft and Donovan 2020) and exchange social support (De Choudhury and De 2014), entertainment, and information (Benkler 2006). Online community platforms support millions of attempts to build communities (Hill and Shaw 2019; Schweik and English 2012); however, only a tiny percentage manages to mobilize participants and sustain collaboration (Schweik and English 2012).

Early research into the growth, survival, and success of online communities focused almost exclusively on the perspective of managers of a single community to find effective practices for motivating and sustaining productive participation (Kraut and Resnick 2012). Explanatory variables included the characteristics of founders (Kairam and Foote 2024; Kraut and Fiore 2014), language use (Danescu-Niculescu-Mizil et al. 2013), turnover (Dabbish et al. 2012), and designs for regulating behavior (Halfaker et al. 2013; TeBlunthuis, Shaw, and Hill 2018). Analyses from this "focal organization perspective" (Hannan and Freeman 1989) have typically accounted for only a small amount of variation in communities' growth, longevity, and performance and have limited ability to predict community growth or long-term participation (Cunha et al. 2019).

Recent scholarship has demonstrated that interdependence among online communities is widespread, important to explaining success and failure, and likely to provide new insights for designing and managing communities (Chandrasekharan et al. 2017; Cunha et al. 2019; Kairam, Wang, and Leskovec 2012; Mitts, Pisharody, and Shapiro 2022; Tan 2018; Tan and Lee 2015; TeBlunthuis et al. 2022; Vincent, Johnson, and Hecht 2018). Most pertinent are studies that adopt the theoretical lens of *organizational ecology* to understand *competition* and *mutualism* between overlapping online communities (TeBlunthuis and Hill 2022; Wang, Butler, and Ren 2012; Zhu et al. 2014; Zhu, Kraut, and Kittur 2014). The initial work in this line emphasizes competition and finds that overlapping Usenet groups are likely to compete (Wang, Butler, and Ren 2012). Similarly, in chapter 6 of the prominent textbook *Building Successful Online Communities*, Resnick et al. (2012) recommend that new online communities "carve out a useful and defendable niche in the ecology of competing communities."

However, most empirical findings in ecological studies of online communities point to *mutualism* and *specialization* as beneficial to the success of overlapping communities (TeBlunthuis and Hill 2022; TeBlunthuis et al. 2022; Zhu et al. 2014; Zhu, Kraut, and Kittur 2014). Recent interviews with members of overlapping subreddits revealed that each community often provided benefits that others did not such as audience, belonging, and specific content and information (TeBlunthuis et al. 2022). Studying overlapping fandom Wikis, Zhu, Kraut, and Kittur (2014) suggest that although communities might "compete over shared members' time and efforts," overlapping communities share knowledge, diverse perspectives, and opportunities to recruit new members. Although the researchers hypothesized that competitive and mutualistic forces would trade off as user overlap increases, they instead found that increasing overlap was associated with increasing survival of new communities. A related study found evidence of both competition and mutualism; however mutualism was stronger than competition for most communities (Zhu et al. 2014). TeBlunthuis and Hill inferred networks of competition and mutualism among overlapping subreddits using time-series analysis and found that mutualism is more common than competition (TeBlunthuis and Hill 2022).

Yet the method TeBlunthuis and Hill propose for inferring competition and mutualism depends on vector autoregression (VAR) models. Although these models have been used in ecology to infer competition and mutualism (Ives et al. 2003), they also bear assumptions that may be unrealistic in online communities. VAR models can only represent linear dynamics and competitive and mutualistic interactions that do not vary over time (Cenci and Saavedra 2019; Cenci, Sugihara, and Saavedra 2019); however, online communities inhabit dynamic environments and experience shocks such as influx of newcomers and attention (Kiene, Monroy-Hernández, and Hill 2016; Lin et al. 2017; Ratkiewicz et al. 2010; Zhang et al. 2019). For example, policy changes and bans can influence related communities (Chandrasekharan et al. 2017; Matias 2016; Ribeiro et al. 2021). Therefore, this study uses nonlinear time-series analysis to investigate how ecological relationships in clusters of overlapping communities vary over time. The first hypothesis tests whether the linearity assumptions would not invalidate previous findings by TeBlunthuis and Hill and that these reflect the average relationships over time. H1: *Mutualistic interactions will be more frequent and longer lasting than competitive interactions.*

Explaining Widespread Mutualism

How do systems of overlapping mutualistic online communities come to be? How do these communities avoid competing with each other? Measuring time-varying ecological interactions opens a path to test whether adaptation can explain why mutualism appears more common than competition. Prior work drawing on organizational ecology have suggested that online communities depend on types of resources with the potential to create both mutualism and competition (TeBlunthuis and Hill 2022; Wang, Butler, and Ren 2012; Zhu et al. 2014). Online communities might compete for participants' time and attention, which is a rival resource in that its usefulness decreases when it is used. On the other hand, overlapping online communities can be mutualists as they produce non-rival or "anti-rival" goods, such as information, that increase in value with use and that other communities can utilize. If resources are the whole story, then widespread mutualism suggests that the non-rival resources shared among communities are more important, a view resonant with early optimism about online collaboration (Benkler 2006). However, an explanation based soley on resources does not account for how online community builders are knowledgeable agents with specific goals. Online communities do not spring from the ether; they are built by their leaders and participants.

This section proposes that online communities' participants construct systems of mutualistic communities that provide a range of complementary benefits as they collectively adapt existing communities to fill available niches. The competitive exclusion principle—one of the most important concepts in ecology (Armstrong and McGehee 1980; Hardin 1960)—proposes that natural selection disfavors competitive relationships. Strong enough competition between two groups will lead to the death of at least one of the groups. Coexistence is possible only through specialization (Levin 1970). Recall recent qualitative work that finds that overlapping communities provide different types of benefits to their users (TeBlunthuis et al. 2022). Ecological theory suggests that such specialization will reduce competition.

A two-stage process may occur in which competitive dynamics first lead to specialization, and this specialization subsequently decreases competition. Classical organizations facing competition may attempt to change and differentiate from competitors (McPherson and Ranger-Moore 1991). However, organizations often fail to affect such change (Hannan and Freeman 1984) because forces of institutional isomorphism (DiMaggio and Powell 1983), contagion (Greve 1995), or organizational learning (Dobrev, Kim, and Carroll 2003) may be more important causes of organizational change than competition avoidance. Furthermore, forces termed "structural inertia" also hinder change efforts, including organizational culture, internal patronage networks, conflicts among stakeholders, and routines (Hannan and Freeman 1984; van de Ven and Poole 1995). Such inertial forces can also increase failure rates following periods of change (Hannan and Freeman 1989).

Although structural inertia is central to organizational ecology, prior ecological studies of online communities have barely touched on it. However, empirical findings strongly suggest that online communities also experience structural inertia. Online communities tend to change their policies less frequently over time (Halfaker et al. 2013; TeBlunthuis, Shaw, and Hill 2018) and have little turnover in their leaders who often resist change (Shaw and Hill 2014). Similar to classical organizations, online communities have roles (Arazy et al. 2017, 2015), routines (Keegan, Lev, and Arazy 2016) and concertive control (Gibbs, Rice, and Kirkwood 2021), all of which can lead to structural inertia. Moreover, online communities typically lack capacities to coerce or incentivize change available to classical organizations such as firms and governments. Thus, endogenous structural changes in established online communities probably occur in bottom-up processes driven by the choices of individual members or small groups of them (Steinsson 2023).

Even without capacities for top-down structural change, members may tend to contribute to communities in ways that lead to specialization (Van Koevering, Ye, and Kleinberg 2024). Online communities are typically "open" organizations where individuals can freely participate in multiple communities at once and share or repost the same content in different communities (Butler and Wang 2011). When some members become unhappy, an online community characterized by structural inertia is unlikely to change policies or leadership to satisfy them. However, in a system of overlapping open communities, unhappy members can "exit" and migrate to an overlapping community and bring their contributions along (Frey and Sumner 2018; Hirschman 1970).

In ecological terms, such migrations correspond a competition episode between the communities because a decrease in participation in one group coincides with an increase in participation in the other. Migration may also decrease the overlap of the two communities' niches, as recent empirical work on Reddit suggests (Van Koevering, Ye, and Kleinberg 2024). If individuals participate more in communities where they have the greatest expectation of finding a specific type of benefit, their participation can reinforce the community's ability to provide these specific benefits. Their contributions can increase the supply of content, attention, and social interaction, and they may reward others who make similar contributions with thanks, awards, votes and other signals of approval. The size of the group of individuals desiring these specific benefits may subsequently increase in the second community and decrease in the first. The aggregated actions of individuals with aligned interests can collectively result in a positive feedback loop that increases and stabilizes the second community's provision of these benefits while decreasing the niche overlap between the two communities in the process (McPherson and Rotolo 1996; Schelling 1978).

Previous ecological studies of online communities have measured niches by quantifying user and topic overlaps (TeBlunthuis and Hill 2022; Wang, Butler, and Ren 2012; Zhu et al. 2014; Zhu, Kraut, and Kittur 2014). Logically, specialization might happen along either dimension or both. Therefore, the following two hypotheses test whether online communities specialize by reducing niche overlap.

H2a: *Two subreddits having greater competition (mutualism) will subsequently have greater decreases (increases) in **topic overlap**.*

H2b: *Two subreddits having greater competition (mutualism) will subsequently have greater decreases (increases) in **user overlap**.*

The competitive exclusion principle predicts that specialization reduces competition over rival resources (Armstrong and McGehee 1980; Hardin 1960). Resource partitioning theory and niche width theory in organizational ecology explain the relationship between competition and overlap (Carroll 1985; Dobrev, Kim, and Carroll 2003; Dobrev, Kim, and Hannan 2001). Sometimes, organizations can adapt to avoid competition as when automotive manufacturers with greater niche overlaps undergo greater transformations during periods of technological change (Dobrev, Kim, and Carroll 2003). Such change can be risky; however increasing specialization to avoid competition can be necessary for organizational survival (Baum and Shipilov 2006). Importantly, the ecological interactions between two online communities can involve a mixture of mutualistic and competitive components. For instance, two communities may compete over contributors' time and effort while simultaneously benefiting from sharing information or content. Even if the relationship between two communities is mutualistic overall, a decrease in niche overlap can reduce the competitive component of the relationship and increase the overall mutualism. Therefore, the final two hypotheses test whether increases (decreases) in niche overlaps will lead to subsequent (decreases) increases in (mutualism) competition.

H3a: *Two subreddits having decreasing (increasing) **topic overlap** will subsequently have greater mutualism (competition).*

H3b: *Two subreddits having decreasing (increasing) **user overlap** will subsequently have greater mutualism (competition).*

Data and Measures

Online Communities Hosted on Reddit.com

This study requires a dataset of communities that are all active over a sufficiently long period; it uses the publicly available Pushshift archive of Reddit submissions and comments with a study period of December 5$^{\text{th}}$ 2015 to April 13$^{\text{th}}$ 2020 (Baumgartner et al. 2020). The included subreddits had comments or submissions in at least 20% of weeks during the study period; however, they did not have more than 10% of submissions marked "not safe for work", which often indicates sexual content.

Overview

The analysis proceeds as follows and summarized in Figure 1: First, a clustering pipeline discovers groups of highly overlapping subreddits based on membership similarity. The S-Map algorithm then measures competition and mutualism episodes with which to test H1. Finally, the panel regression models test H2 and H3 using weekly similarity measures based on membership and content and competition and mutualism that the S-Map measured. The next section documents the measures and analytic plan, and the results follow.

Discovering Subreddit Clusters

A procedure similar to that of (TeBlunthuis and Hill 2022) constructs clusters of overlapping subreddits with overlapping users. The first step in this process is to measure *user overlap* $o_{i,j}$ between each pair of subreddits i and j. This is the number of contributions (posts or comments) made by each user account in each subreddit normalized by the maximum number of contributions to the subreddit by any account. Next, Latent Semantic Analysis (LSA) with 2,000 dimensions reduces the dimensionality of the frequency vectors. 2,000 dimensions for LSA gave the best clustering performance as described in the next paragraph. The cosine similarity of the resulting vectors for each pair of subreddits computes user overlap.

The HDBSCAN algorithm clusters subreddits into groups with high user overlap over the entire study period. TeBlunthuis and Hill tested different clustering algorithms and found that HDBSCAN worked the best in terms of the silhouette coefficient (Rousseeuw 1987), a measure of clustering performance that quantifies within-cluster similarity. A grid search finds HDBSCAN hyperparameters and the number of LSA dimensions that together yeild a high silhouette coefficient as long as fewer than 10,000 subreddits are not assigned to any cluster because removing subreddits increases the silhouette coefficient. The chosen solution had a silhouette coefficient of (0.53), 1,949 clusters and 9,833 isolated subreddits.

Inferring Dynamic Ecological Interactions

Time-series models of *group size* infer ecological interactions. Group size is the number of distinct contributors to the subreddit each week. As the number of groups increases, our method requires an exponentially longer time series (Cenci, Sugihara, and Saavedra 2019); thus 30 clusters that have more 28 communities were excluded. The final dataset has 8,806 subreddits in 1,919 clusters with 48,484 relationships measured 17,374,116 times over up to 758 weeks. Table 1 summarizes the activity levels in these communities.

As discussed above, adaptation to changing environmental conditions or exogenous shocks can drive changes in the interactions between online communities over time. This study measures these dynamic competitive and mutualistic interactions between online communities using the regularized S-Map, a nonlinear, nonparametric model that uses

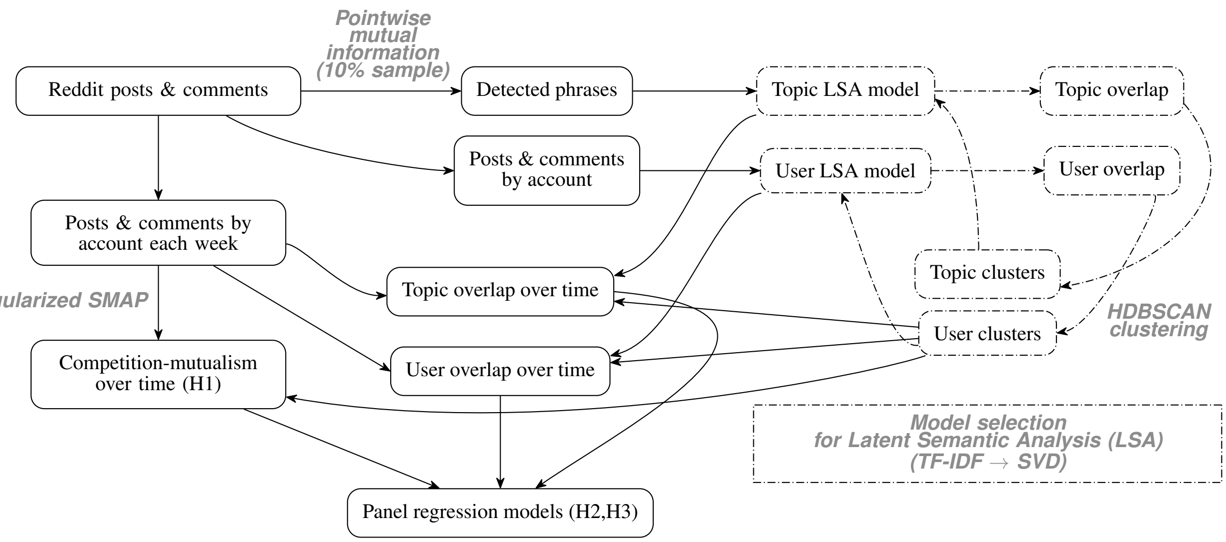

Figure 1: Flowchart representing the pipeline for building longitudinal measures of competition and mutualism as well as topic and author overlap and testing the hypotheses. Hyperparemters for LSA models are selected in a repeated process to find a good silhouette coefficient for the topic and user clusters.

Weekly Messages	Min	Mean	Med	Max	Std. Dev
Min	0.00	92	0	47548	858
Mean	0.43	378	40	78451	2032
Median	0.00	323	26	78090	1890
Maximum	3.00	1516	201	325882	7224
Std. Dev	0.45	238	35	64956	1191

Table 1: Statistics summarizing the distribution of contributions made in subreddits. The rows are statistics within each subreddit during the study period. The columns are statistics over the subreddits.

time-series data of population sizes to infer an evolving matrix of community interactions (Cenci et al. 2020; Cenci and Saavedra 2019; Cenci, Sugihara, and Saavedra 2019; Deyle et al. 2016). Intuitively, this model asks "when the system has been in a state similar to the present, what usually happens next?" and uses the answer to generate forecasts.

The S-map estimates a sequence of Jacobian matrices $\mathcal{C}^t$ quantifying community interactions at each time t; it models each observation as a linear regression of past observations, weighted by their similarity to the present. The weights are defined by a distance function and a scaling kernel. Following common choices in EDM research and S-Map applications, the distance function is the Euclidean norm, and the scaling kernel is the exponential kernel.

This kernel has a hyperparameter θ controlling the degree of locality (closeness in states of the system) in the weights. With $\theta \rightarrow 0$, the S-Map becomes a linear VAR model. Larger values of θ correspond to greater degrees of nonlinearity (Sugihara et al. 1994).

The S-map contains parameters for each pair of communities at each time point and thus is prone to overfitting. Therefore, elastic net regularization is used to improve numerical stability and reduce the risk of incorrectly interpreting the noise introduced by endogenous or unobserved factors as community interactions (Cenci, Sugihara, and Saavedra 2019).

For each cluster, the procedure given by Cenci and Saavedra selects a model. A grid search based on leave-one-out cross validation (Cenci and Saavedra 2019) finds the hyperparameters θ (locality), α (ratio of $l1$ to $l2$ penalization), and λ (total regularization penalty). Following previous applications of regularized S-map models, data on group sizes are log-transformed and standardized.

Mutualistic and Competitive Episodes

This study tested H1 by analyzing the prevalence and longevity of competitive and mutualistic interactions. For each tuple of communities (a, b) in a cluster, *episodes* of competition or mutualism are consecutive weeks when the element of the Jacobian $\mathcal{C}^t_{a,b}$ is positive or negative. Ecological interactions may be asymmetrical; therefore, for each pair of communities a, b, episodes of interaction from a to b ($\mathcal{C}^t_{a,b}$) may differ from b to a ($\mathcal{C}^t_{b,a}$). The study includes a total of 17,374,116 episodes.

Measuring Subreddit Similarity Over Time

Measuring *weekly user overlap* $U_{i,j,t}$ replicates the procedure for *user overlap* described above using the LSA model computed over the entire study period, but the contributions from each week.

Weekly topic overlap $T_{i,j,t}$ between subreddits is similar to weekly user overlap, but using token-frequency inverse-document-frequency (TF-IDF) vectors instead of author-frequency vectors. Pointwise mutual information (PWMI) detects phrases for inclusion in the TF-IDF vectors. Phrases of up to four terms were used as it was the maximum computationally tractable given the available resources. The number of possible phrases grows exponentially with the phrase length.

To be included, a phrase must have a PWMI score of at least 3, indicating that the phrase's probability is 20 times greater than the product of the probabilities of each constitutive term. In addition, the phrase count must be at least 3,500 within the 10% sample and must appear in at least 200 different subreddits. All single tokens appearing in at least 200 different subreddits also included.

Next, TF-IDF quantifies the ratio of the count of each token within each subreddit divided by the maximum count for the subreddit and by the log-count of the number of subreddits with the token. LSA reduces the dimensionality and sparsity of the resulting TF-IDF while preserving subreddit similarities (Dumais 2004). An HDBSCAN clustering procedure similar to that used to find user-based clusters chose 1,000 dimensions for the topic overlap LSA model. The best topic-based clustering had 1,216 clusters with 9,557 subreddits not assigned to any cluster. The resulting topic-based clusters are not used in subsequent analyses to follow TeBlunthuis and Hill's use of user-based clusters and because user-based clusters obtained a much greater silhouette coefficient (0.53) than topic-based clusters (0.34). As with the author similarities, an LSA model computed over the entire study period computes weekly topic overlap by taking the cosine similarities of transformed weekly TF-IDF vectors for each subreddit in the study period. Both similarity measures range from 0 (no overlap) to 1 (complete overlap).

Panel Regression Models

Fixed effect panel data estimators fit using ordinary least squares test H2 and H3 (Wooldridge 2011). The advantage of this method is that it is robust to confounding by time-invariant unobserved variables.

The units of analysis in all of the regression models are dyadic relationships between subreddits. This dyadic structure violates the assumption that observations are drawn independently from an identical distribution. When a subreddit's topic changes, for instance, resulting changes in its overlaps in all related subreddits will be correlated. Therefore, the analysis uses the "cluster–robust variance estimation for dyadic data" technique developed by Aronow, Samii, and Assenova and implemented in the `dyadRobust` R package[1] that provides consistent estimates of standard errors for dyadic data with repeated measures (Aronow, Samii, and Assenova 2015). The Hypotheses are tested via 95% confidence intervals calculated by multiplying the standard errors by 1.96. The dataset is sufficiently large to allow normal distribution approximation.

Results

The analysis supports H1, that mutualism episodes are both more frequent and longer lasting than competitive episodes. However, competitive episodes tended to be slightly stronger. It also supports H2, that competitive episodes lead to reductions in topic or user overlaps between communities. However, it does not find evidence of H3, that decreases in overlaps result in increasing mutualism.

[1]https://rdrr.io/github/jbisbee1/dyadRobust/

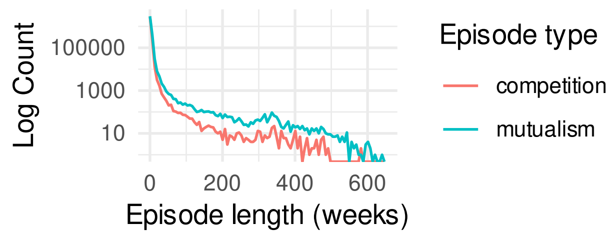

Figure 2: Frequency plot of the durations of competition and mutualism episodes. Mutualism tends to last longer than competition. The y-axis is log-transformed. The axes are truncated to omit outliers for visibility.

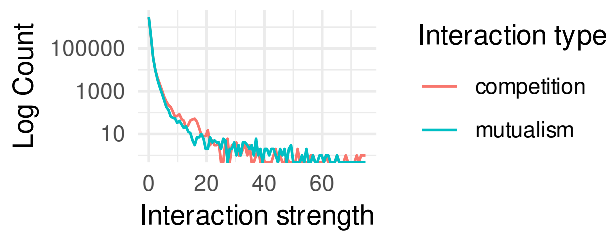

Figure 3: Frequency plot of average competition or mutualism strength for each episode. Competition and mutualism tend to have similar strengths. The y-axis is log-transformed.

Mutualistic Interactions Are More Frequent and Longer Than Competitive Ones

Although previous studies conceptualize mutualism and competition as static relationships, the present analysis finds that ecological interactions among overlapping online communities are bursty and relatively short-lived. The average episode of competition or mutualism was only 2.13 weeks.

Mutualism episodes are slightly more frequent, 52.1% of all episodes. They also last longer than competition episodes. The average mutualism episode lasts 2.42 weeks and the average duration of competition is 1.83. The average episode is weak mutualism with an average interaction strength of 0.04.

Mutualism is more common at any given moment because it lasts longer. Figure 2 shows the distributions of the durations of competition and mutualism episodes in weeks. Most episodes are short: 80% of competition episodes last 2 weeks or less, and 80% of mutualism episodes last 3 weeks or less. Among the longest episodes, most are mutualism. Of the top 5% longest episodes, 100% are mutualism. The top 1% of mutualism episodes last 13 weeks or more compared to the top 1% of competition episodes which last at least 8 weeks.

Competition and mutualism episodes are of similar strength on average, as Figures 4 and 3 visualize. The average

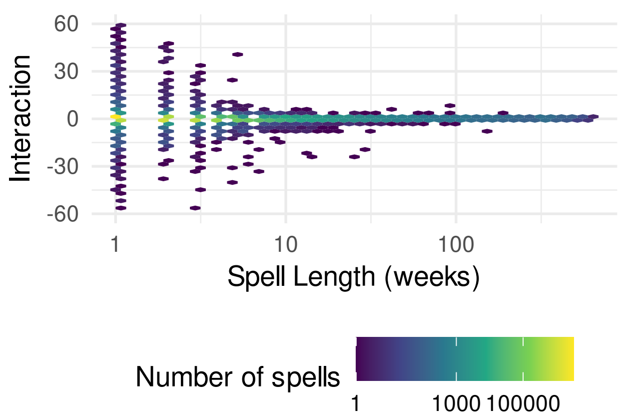

Figure 4: Episode lengths by the average competition and mutualism strength. Most episodes are short, but mutualism tends to last longer than competition. For clarity, the axes are truncated to remove extreme values.

competition strength is 0.2029 compared to 0.2039 for mutualism. As before, the extremes of the distribution are clarifying. Most interactions are relatively weak: 50% of mutualism episodes have an average $\mathcal{C}_{i,j}$ less than or equal to 0.12 and 50% of competition episodes have an average $\mathcal{C}_{i,j}$ greater than or equal to -0.11. However, the strongest episodes are competition more often than mutualism. Of the 5% strongest episodes, 51% are competition. The 5% strongest mutualism episodes have an average $\mathcal{C}_{i,j}$ greater than 0.62 and the top 5% competition episodes have an average $\mathcal{C}_{i,j}$ less than -0.66.

Online Communities Increase Specialization in Relatively Competitive Conditions

H2 proposed a positive (negative) relationship between mutualism (competition) and future topic or user overlap because online communities would specialize in response to competition. The evidence from the panel regression models supports this hypothesis for both types of overlap. Mutualism (competition) is positively associated with increasing (decreasing) topic overlap (B_1=0.024; CI=[0.0155, 0.0324]). The relationship is relatively subtle. A one-unit increase in mutualism $\mathcal{C}_{i,j}$ corresponds to an increase in term overlap by 0.024 standard deviations.

Similarly, online communities in mutualism (competition) are likely to have subsequent increases (decreases) in user overlap (B_1=0.02; CI=[0.0129, 0.0266]). An increase of one unit mutualism strength $\mathcal{C}_{i,j}$ corresponds to an increase in user overlap of 0.02 standard deviations; thus H2 is supported.

Reducing Overlaps Do Not Lead to Measurable Increases in Mutualism

Competition results in a decrease in niche overlap between subreddits. Does this specialization, in turn, decrease competition or increase mutualism, as H3 proposes? A negative (positive) relationship between previous niche overlaps and subsequent mutualism (competition) would support this hypothesis. However, the panel regression model instead has a positive relationship between previous user overlaps and ecological interactions (B_1=0.06; CI=[0.046, 0.074]), as well as between previous topic overlaps and subsequent mutualism (B_2=0.09; CI=[0.066, 0.113]), opposite in sign to that hypothesized. In sum, the analysis finds no support for H3.

Discussion

Online community leaders and participants can build overlapping communities to meet diverse and interconnected needs (TeBlunthuis et al. 2022; Zhu, Kraut, and Kittur 2014). This study investigates how they do so with the a goal of providing designers with insights into the relationships among overlapping communities. Prior work treated competition and mutualism as static relationships and found that mutualism was much more common than competition (TeBlunthuis and Hill 2022). However, more flexible nonlinear time series methods reveal that these relationships are not static, but usually take the form of episodic bursts. Findings from these models explain that mutualism is more common than competition, not primarily because mutualistic episodes have greater frequency or strength, but because they last longer. Competitive episodes are slightly less common than mutualistic ones and have similar strength on average; however mutualistic interactions last over half a week longer on average.

How do online communities find their niches? What gives rise to frequent mutualism? Theories of competitive exclusion and organizational ecology predict that high degrees of niche overlap lead to competition (Hannan and Freeman 1989). By implication, mutualistic online communities might be "made" via an adaptive process if communities shift their niches in ways that decrease competition and promote mutualism. Alternatively, communities "born" in a competitive niche may struggle to sustain activity, making those born in a mutualistic niche more likely to survive. This study investigated these processes in a series of panel regression models to test how ecological interactions might drive behavior patterns that shift communities' niches to decrease user and topic overlaps. Although this study finds evidence that competition tends to decrease niche overlaps, it does not find that such changes subsequently reduce competition. Thus, the adaptive "made" hypotheses seems inadequate to explain the preponderance of mutualism over competition.

In fact, the results suggest that change tends to be negative because it leads to increasing competition. These findings resonate with previous research suggestive of structural inertia in online communities (Halfaker et al. 2013; TeBlunthuis, Shaw, and Hill 2018). Such inertia may limit the ability of communities to adapt to competition. In some cases, online communities have significantly changed in response to the rise of competitors; this has depended on major structural changes implemented by leadership and has resulted in long-run participation declines (Hill 2011). Moreover, the burstiness of mutualism and competition may itself be a source of inertia. Even if a community perceives competition, any urgency for responsive adaptation may be lost when the episode ends.

An alternative ecological explanation for frequent mutualism is the variation, selection, and retention process (Campbell 1965; van de Ven and Poole 1995). Social scientists in the mid-20th century used this framework to explain sociocultural change in organizations and institutions, and it is still influential among organizational and cognitive scientists today (Aldrich and Ruef 2006; Hannan 2019; Heyes 2018; Monge and Poole 2008). A selection process can produce groups of overlapping, specialized, and frequently mutualistic online communities as follows: Online community founders generate *variation* in the form of newly created online communities. Contributors *select* communities that provide the distinctive benefits that gratify them. Nascent communities that successfully meet such members' needs will be *retained* as they bring together a critical mass of committed members. Future work should develop and test such a model, perhaps using agent-based models of how people join new online communities (Foote et al. 2023) or behavioral data on online communities' early development (Tan 2018).

Design and Management Considerations

Similar to ecologists caretaking natural environments, designers and leaders of online communities may seek to intervene to manage their ecosystems. They could introduce competitors of problematic communities or develop mutualistic communities to complement existing ones. Should they do so by building a new community or by adapting an existing one? Building new communities is notoriously difficult, prone to failure, and seems slow compared to the timescales of mutualism and competition (Schweik and English 2012); therefore, adaptation might seem more practicable. However, this study's findings suggest that communities face structural intertia and may struggle to adapt mutualistically; still, solutions are conceivable. When existing communities spawn new communities that have strong connections with their parents, the new communities can grow faster (Tan 2018). Organizing a spin-off or offshoot community may thus be a viable strategy to fill needed roles in online commuity ecosystems. The "parent" community should create a channel where members can collaborate to design the new community and coordinate their efforts to build it. Of course, doing so is only possible through the efforts of interested members.

Alternatively, the study's finding that online communities do not adapt to find mutualistic niches may reflect constraints that may be surmounted by future designs. Community leaders will not attempt a niche shift unless they see a connection between such change and their community's goals. An *ecological monitor* that detects and displays competition episodes and opportunities for mutualism can help them recognize such opportunities. Even if community leaders recognize an opportunity, platforms such as Reddit offer limited levers to execute change. They can alter their rules to differentiate themselves from related communities; however, crafting and enforcing rules requires time and effort, and change can be resisted or spark migration (Davies et al. 2021; Kiene 2024). Subtle changes to feeds or recommendations may nudge audiences and contributors to gradually shift community niches. An implementation challenge will be to ensure that such nudges do not violate users' expectations or cause communities' niches to migrate beyond the bounds of their topic and goals. Without such care, communities may resist such designs.

Threats to Validity

Ecologists designed the regularized S-Map to infer competition and mutualism in nonlinear systems, and it does so accurately over time up to a constant (Cenci and Saavedra 2019). However, this method has limitations. Foremost, it does not account for uncertainty in the estimation. In addition, it is possible that unobserved variables, such as other communities not included in the models, may confound these models. Finally, the method depends on a long time-series of activity, which requires that the dataset exclude small and short-lived subreddits whose ecological dynamics may differ from those included. Future work should investigate other approaches to nonlinear time series analysis that have different assumptions and can better account for uncertainty (Kantz and Schreiber 2003).

A second set of threats emerges from the panel regressions. Although fixed effect panel models are consistent even in the presence of time-constant omitted variables (Wooldridge 2011), time-varying unobserved variables, such as trends in topics' popularity, can potentially bias estimates. Therefore, readers should not draw causal conclusions from these models. In general, this study reveals the overall trends of competition, mutualism, and overlaps on Reddit at a large scale. However, the wide view and analytical approach obscure the particular mechanisms driving competition and mutualism among these communities and when they are likely to arise. Future fine-grained case studies and experiments should seek to uncover such mechanisms.

Additional threats may result from weaknesses in the correspondence between the study's measures and its scientific concepts. The measure of topic overlap depends on a relatively simple language model that cannot fully capture the meaning of what people are saying. These measures are relatively straightforward to understand; however, future work may use more advanced language models to create more sensitive and accurate measures of online community niches. All of these measures are based on observed activity within these communities; however, the analysis includes communities that sometimes have little activity. Future work should explore whether communities of different sizes tend to interact similarly or differently and develop measures of online community niches with greater precision in low-activity communities .

Furthermore, the inferences of competition and mutualism are based on group size; however, group size is an incomplete definition of online community success. Larger online communities are not necessarily more successful; however, they provide different types of benefits than small communities (Hwang and Foote 2021). Similarly, as is common in the literature on online community ecosystems, the measures of membership and participation lump together a range of contributions including both minor or repetitive contributions and intensive efforts. Similarly, the measure of user overlap is based only on the number of contributions and does not account for qualitative differences in participation. Future work should develop alternative measures of online community success, participation, and membership.

Finally, this study analyzes the ecological dynamics within clusters of overlapping communities on Reddit. The solution to the clustering algorithm may not be unique, and analyses based on different strategies for finding groups of overlapping subreddits may yield different results. Furthermore, although Reddit is among the largest platforms for online communities, this study's findings may not generalize to online communities on other platforms or during other time periods. Future research should investigate online community ecosystems on other online community platforms.

Conclusion

Recent studies investigated relationships between overlapping online communities using an ecological framework and found widespread mutualistic relationships using time-series analysis (TeBlunthuis and Hill 2022). This study assumed that ecological interactions between online communities are linear and static; it also provided little explanation of how these mutualistic interactions emerged. The present study uses nonlinear time-series analysis methods from ecological studies of biological organisms to reveal bursty ecological interactions that vary over time. It finds that mutualism is more common than competition because mutualistic interactions are more common and, more importantly, longer lasting. This study also tests a model for how online communities adapt to avoid competition and increase mutualism. Although user and topic overlaps decrease under relatively competitive conditions, evidence does not support a theory that such specialization effectively reduces competition or leads to mutualism. Future research should explain the emergence of mutualistic interactions between overlapping online communities in terms of how people organize and join communities early in their development. Future designs may enable online community ecosystem management by identifying open roles in the ecosystem and then filling them by purposfully adapting existing communities or organizing new ones.

Acknowledgments

Text from a draft of this article was included as part of the author's PhD dissertation at the Department of Communication at the University of Washington. He thanks his committee: Professors Benjamin Mako Hill, Kirsten Foot, Aaron Shaw, David McDonald and Emma Spiro, for their generous support, wise advice and insightful comments. Versions of this paper received helpful feedback at the International Communication Association's 2022 annual meeting and at the International Conference for Computational Social Science. Special thanks to Simone Cenci for advice on the regularized S-MAP. Thanks to the Community Data Science Collective for additional feedback and support.

Additional thanks to Jason Baumgartner and pushshift.io for providing the Reddit data archive as well as to the Reddit users and communities whose actions this study analyzes. Also, thanks to the peer reviewers and program committee members whose insightful comments improved the this article's quality. Any remaining errors and imperfections are the author's. This work was supported by NSF grants IIS-1908850 and IIS-1910202 and GRFP #2016220885 and was facilitated through the use of the advanced computational infrastructure provided by the Hyak supercomputer system at the University of Washington as well the Texas Advanced Computing Center at The University of Texas at Austin.

References

Aldrich, H.; and Ruef, M. 2006. *Organizations Evolving*. Thousand Oaks, CA: SAGE Publications, 2nd edition. ISBN 978-1-4129-1047-7.

Arazy, O.; Liifshitz-Assaf, H.; Nov, O.; Daxenberger, J.; Balestra, M.; and Cheshire, C. 2017. On the "How" and "Why" of Emergent Role Behaviors in Wikipedia. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, 2039–2051. New York, NY, USA: ACM. ISBN 978-1-4503-4335-0.

Arazy, O.; Ortega, F.; Nov, O.; Yeo, L.; and Balila, A. 2015. Functional Roles and Career Paths in Wikipedia. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, 1092–1105. New York, NY: ACM. ISBN 978-1-4503-2922-4.

Armstrong, R. A.; and McGehee, R. 1980. Competitive Exclusion. *The American Naturalist*, 115(2): 151–170.

Aronow, P. M.; Samii, C.; and Assenova, V. A. 2015. Cluster–Robust Variance Estimation for Dyadic Data. *Political Analysis*, 23(4): 564–577.

Baum, J. A. C.; and Shipilov, A. V. 2006. Ecological Approaches to Organizations. In *Sage Handbook for Organization Studies*, 55–110. Rochester, NY: Sage.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The Pushshift Reddit Dataset. *ICWSM*, 14: 830–839.

Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven London: Yale University Press. ISBN 978-0-300-12577-1.

Benkler, Y.; Roberts, H.; Faris, R.; Solow-Niederman, A.; and Etling, B. 2013. Social Mobilization and the Networked Public Sphere: Mapping the SOPA-PIPA Debate. SSRN Scholarly Paper ID 2295953, Social Science Research Network, Rochester, NY.

Benkler, Y.; Shaw, A.; and Hill, B. M. 2015. Peer Production: A Form of Collective Intelligence. In Malone, T. W.; and Bernstein, M. S., eds., *Handbook of Collective Intelligence*, 175–204. Cambridge, MA: MIT Press. ISBN 978-0-262-02981-0.

Butler, B. S.; and Wang, X. 2011. The Cross-Purposes of Cross-Posting: Boundary Reshaping Behavior in Online Discussion Communities. *Information Systems Research*, 23(3-part-2): 993–1010.

Campbell, D. T. 1965. Variation and Selective Retention in Socio-Cultural Evolution. In Barringer, H. R.; Blanksten, G. I.; and Mack, R. W., eds., *Social Change In Developing Areas*. Cambridge Mass.: Schenkman Publishing Company.

Carroll, G. R. 1985. Concentration and Specialization: Dynamics of Niche Width in Populations of Organizations. *American Journal of Sociology*, 90(6): 1262–1283.

Cenci, S.; Medeiros, L. P.; Sugihara, G.; and Saavedra, S. 2020. Assessing the Predictability of Nonlinear Dynamics under Smooth Parameter Changes. *Journal of The Royal Society Interface*, 17(162): 20190627.

Cenci, S.; and Saavedra, S. 2019. Non-Parametric Estimation of the Structural Stability of Non-Equilibrium Community Dynamics. *Nature Ecology & Evolution*, 3(6): 912–918.

Cenci, S.; Sugihara, G.; and Saavedra, S. 2019. Regularized S-map for Inference and Forecasting with Noisy Ecological Time Series. *Methods in Ecology and Evolution*, 10(5): 650–660.

Chandrasekharan, E.; Pavalanathan, U.; Srinivasan, A.; Glynn, A.; Eisenstein, J.; and Gilbert, E. 2017. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined through Hate Speech. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW): 31:1–31:22.

Choudhury, M. D.; Jhaver, S.; Sugar, B.; and Weber, I. 2016. Social Media Participation in an Activist Movement for Racial Equality. In *Tenth International AAAI Conference on Web and Social Media*.

Cunha, T.; Jurgens, D.; Tan, C.; and Romero, D. 2019. Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities. In *The World Wide Web Conference*, WWW '19, 318–328. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6674-8.

Dabbish, L.; Farzan, R.; Kraut, R.; and Postmes, T. 2012. Fresh Faces in the Crowd: Turnover, Identity, and Commitment in Online Groups. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, 245–248. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-1086-4.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. In *Proceedings of the 22nd International Conference on World Wide Web - WWW '13*, 307–318. Rio de Janeiro, Brazil: ACM Press. ISBN 978-1-4503-2035-1.

Datta, S.; Phelan, C.; and Adar, E. 2017. Identifying Misaligned Inter-Group Links and Communities. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW): 37:1–37:23.

Davies, C.; Ashford, J.; Espinosa-Anke, L.; Preece, A.; Turner, L.; Whitaker, R.; Srivatsa, M.; and Felmlee, D. 2021. Multi-Scale User Migration on Reddit. In *Workshop on Cyber Social Threats at the 15th International AAAI Conference on Web and Social Media (ICWSM 2021)*. Virtual: AAAI.

De Choudhury, M.; and De, S. 2014. Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity. *ICWSM*, 8(1): 71–80.

Deyle, E. R.; May, R. M.; Munch, S. B.; and Sugihara, G. 2016. Tracking and Forecasting Ecosystem Interactions in Real Time. *Proc. R. Soc. B.*, 283(1822): 20152258.

DiMaggio, P. J.; and Powell, W. W. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2): 147–160.

Dobrev, S. D.; Kim, T.-Y.; and Carroll, G. R. 2002. The Evolution of Organizational Niches: U.S. Automobile Manufacturers, 1885–1981. *Administrative Science Quarterly*, 47(2): 233–264.

Dobrev, S. D.; Kim, T.-Y.; and Carroll, G. R. 2003. Shifting Gears, Shifting Niches: Organizational Inertia and Change in the Evolution of the U.S. Automobile Industry, 1885-1981. *Organization Science*, 14(3): 264–282.

Dobrev, S. D.; Kim, T.-Y.; and Hannan, M. T. 2001. Dynamics of Niche Width and Resource Partitioning. *American Journal of Sociology*, 106(5): 1299–1337.

Dumais, S. T. 2004. Latent Semantic Analysis. *Annual Review of Information Science and Technology*, 38(1): 188–230.

Foote, J.; Treem, J. W.; van den Hooff, B.; and Honcoop, M. 2023. Motivations to Use Multifunctional Public Goods in Organizations: Using Agent-Based Modeling to Explore Differential Uses of Enterprise Social Media. *Human Communication Research*, hqad029.

Frey, S.; and Sumner, R. W. 2018. Emergence of Complex Institutions in a Large Population of Self-Governing Communities. *arXiv:1804.10312 [cs]*.

Gibbs, J. L.; Rice, R. E.; and Kirkwood, G. L. 2021. Digital Discipline: Theorizing Concertive Control in Online Communities. *Communication Theory*.

Greve, H. R. 1995. Jumping Ship: The Diffusion of Strategy Abandonment. *Administrative Science Quarterly*, 40(3): 444–473.

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.

Hannan, M. T. 2019. *Concepts and Categories: Foundations for Sociological and Cultural Analysis*. New York, NY: Columbia University Press. ISBN 978-0-231-19272-9.

Hannan, M. T.; and Freeman, J. 1984. Structural Inertia and Organizational Change. *American Sociological Review*, 49(2): 149.

Hannan, M. T.; and Freeman, J. 1989. *Organizational Ecology*. Cambridge, MA: Harvard University Press, 1 edition.

Hardin, G. 1960. The Competitive Exclusion Principle. *Science*, 131(3409): 1292–1297.

Hawley, A. H. 1986. *Human Ecology: A Theoretical Essay*. Chicago; London: University of Chicago Press. ISBN 978-0-226-31983-4 978-0-226-31984-1.

Heyes, C. 2018. *Cognitive Gadgets The Cultural Evolution of Thinking*. Harvard University Press. ISBN 978-0-674-98515-5.

Hill, B. M. 2011. Almost Wikipedia: What Eight Collaborative Encyclopedia Projects Reveal About Mechanisms of Collective Action. In *Berkman Center for Internet and Society Luncheon Series Presentation*. Harvard University, Cambridge, MA.

Hill, B. M.; and Shaw, A. 2019. Studying Populations of Online Communities. In Foucault Welles, B.; and González-Bailón, S., eds., *The Oxford Handbook of Networked Communication*, 173–193. Oxford, UK: Oxford University Press. ISBN 978-0-19-046051-8.

Hirschman, A. O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press. ISBN 978-0-674-27660-4.

Hwang, S.; and Foote, J. D. 2021. Why Do People Participate in Small Online Communities? *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2): 1–25.

Ives, A. R.; Dennis, B.; Cottingham, K. L.; and Carpenter, S. R. 2003. Estimating Community Stability and Ecological Interactions from Time-Series Data. *Ecological Monographs*, 73(2): 301–330.

Kairam, S. R.; and Foote, J. 2024. How Founder Motivations, Goals, and Actions Influence Early Trajectories of Online Communities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, 1–11. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703300.

Kairam, S. R.; Wang, D. J.; and Leskovec, J. 2012. The Life and Death of Online Groups: Predicting Group Growth and Longevity. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, 673–682. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-0747-5.

Kantz, H.; and Schreiber, T. 2003. *Nonlinear Time Series Analysis*. Cambridge: Cambridge University Press, 2 edition. ISBN 978-0-521-52902-0.

Keegan, B.; Lev, S.; and Arazy, O. 2016. Analyzing Organizational Routines in Online Knowledge Collaborations: A Case for Sequence Analysis in CSCW. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, 1065–1079. New York, NY, USA: ACM. ISBN 978-1-4503-3592-8.

Kiene, C. 2024. *Identity, Legitimacy, and Voice: Understanding Rule Adoption, Compliance, and Evolution in Online Communities*. Ph.D. thesis, University of Washington, United States – Washington.

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 CHI Conference on Human Factors in Computing Systems*, 1152–1156. New York, NY: ACM. ISBN 978-1-4503-3362-7.

Krafft, P. M.; and Donovan, J. 2020. Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. *Political Communication*, 37(2): 194–214.

Kraut, R. E.; and Fiore, A. T. 2014. The Role of Founders in Building Online Groups. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, 722–732. Baltimore, Maryland, USA: ACM. ISBN 978-1-4503-2540-0.

Kraut, R. E.; and Resnick, P. 2012. *Building Successful Online Communities: Evidence-based Social Design*. Cambridge, MA: MIT Press. ISBN 978-0-262-29831-5.

Lampe, C.; Wash, R.; Velasquez, A.; and Ozkaya, E. 2010. Motivations to Participate in Online Communities. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, 1927–1936. Atlanta, Georgia, USA: ACM. ISBN 978-1-60558-929-9.

Levin, S. A. 1970. Community Equilibria and Stability, and an Extension of the Competitive Exclusion Principle. *The American Naturalist*, 104(939): 413–423.

Lin, Z.; Salehi, N.; Yao, B.; Chen, Y.; and Bernstein, M. S. 2017. Better When It Was Smaller? Community Content and Behavior after Massive Growth. In *Eleventh International AAAI Conference on Web and Social Media*, 132–141. Montreal, Canada: AAAI.

Matias, J. N. 2016. Going Dark: Social Factors in Collective Action against Platform Operators in the Reddit Blackout. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, 1138–1151. New York, NY: ACM. ISBN 978-1-4503-3362-7.

McPherson, J. M. 1983. An Ecology of Affiliation. *American Sociological Review*, 48(4): 519–532.

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

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

Mitts, T.; Pisharody, N.; and Shapiro, J. 2022. Removal of Anti-Vaccine Content Impacts Social Media Discourse. In *14th ACM Web Science Conference 2022*, 319–326. Barcelona Spain: ACM. ISBN 978-1-4503-9191-7.

Monge, P. R.; and Poole, M. S. 2008. The Evolution of Organizational Communication. *J Commun*, 58(4): 679–692.

Ratkiewicz, J.; Fortunato, S.; Flammini, A.; Menczer, F.; and Vespignani, A. 2010. Characterizing and Modeling the Dynamics of Online Popularity. *Phys. Rev. Lett.*, 105(15).

Resnick, P.; Konstan, J.; Chen, Y.; and Kraut, R. E. 2012. Starting New Online Communities. In *Building Successful Online Communities: Evidence-based Social Design*, 231–280. Cambridge, MA: MIT Press. ISBN 978-0-262-29831-5.

Ribeiro, M. H.; Jhaver, S.; Zannettou, S.; Blackburn, J.; Stringhini, G.; De Cristofaro, E.; and West, R. 2021. Do Platform Migrations Compromise Content Moderation? Evidence from r/The_donald and r/Incels. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2): 316:1–316:24.

Rousseeuw, P. J. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20: 53–65.

Schelling, T. C. 1978. *Micromotives and Macrobehavior*. WW Norton & Company.

Schweik, C. M.; and English, R. C. 2012. *Internet Success: A Study of Open-Source Software Commons*. Cambridge, MA: MIT Press. ISBN 978-0-262-01725-1.

Shaw, A.; and Hill, B. M. 2014. Laboratories of Oligarchy? How the Iron Law Extends to Peer Production. *J Commun*, 64(2): 215–238.

Steinsson, S. 2023. Rule Ambiguity, Institutional Clashes, and Population Loss: How Wikipedia Became the Last Good Place on the Internet. *American Political Science Review*, 1–17.

Sugihara, G.; Grenfell, B. T.; May, R. M.; and Tong, H. 1994. Nonlinear Forecasting for the Classification of Natural Time Series. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 348(1688): 477–495.

Sugihara, G.; and May, R. M. 1990. Nonlinear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series. *Nature*, 344(6268): 734–741.

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)*, 395–404. Palo Alto, California: AAAI.

Tan, C.; and 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*, WWW '15, 1056–1066. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-3469-3.

TeBlunthuis, N.; and Hill, B. M. 2022. Identifying Competition and Mutualism between Online Groups. In *International AAAI Conference on Web and Social Media (ICWSM 2022)*, volume 16, 993–1004. Atlanta, Georgia, USA: AAAI.

TeBlunthuis, N.; Kiene, C.; Brown, I.; Levi, L. A.; McGinnis, N.; and Hill, B. M. 2022. No Community Can Do Everything: Why People Participate in Similar Online Communities. *Proceedings of the ACM on Human-Computer Interaction: Computer Supported Cooperative Work*, 6: 1–25.

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*, 355:1–355:7. New York, NY: ACM. ISBN 978-1-4503-5620-6.

van de Ven, A. H.; and Poole, M. S. 1995. Explaining Development and Change in Organizations. *ACAD MANAGE REV*, 20(3): 510–540.

Van Koevering, K.; Ye, M.; and Kleinberg, J. 2024. What's in a Niche? Migration Patterns in Online Communities. arXiv:2407.11794.

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, 566:1–566:13. New York, NY: ACM. ISBN 978-1-4503-5620-6.

Wang, X.; Butler, B. S.; and Ren, Y. 2012. The Impact of Membership Overlap on Growth: An Ecological Competition View of Online Groups. *Organization Science*, 24(2): 414–431.

Wooldridge, J. M. 2011. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Mass.: MIT. ISBN 978-0-262-23258-6.

Zhang, A. F.; Wang, R.; Blohm, E.; Budak, C.; Jr, L. P. R.; and Romero, D. M. 2019. Participation of New Editors after Times of Shock on Wikipedia. *Proceedings of the International AAAI Conference on Web and Social Media*, 13: 560–571.

Zhu, H.; Chen, J.; Matthews, T.; Pal, A.; Badenes, H.; and Kraut, R. E. 2014. Selecting an Effective Niche: An Ecological View of the Success of Online Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, 301–310. New York, NY, USA: ACM. ISBN 978-1-4503-2473-1.

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, 281–290. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-2473-1.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes. We analyze public content in aggregate and do not reveal the identity of the users or communities included in this study.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? See introduction and data and measures
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? We do not observe any possible artifacts.
   (e) Did you describe the limitations of your work? Yes, see the limitations section
   (f) Did you discuss any potential negative societal impacts of your work? See ethics statement below.
   (g) Did you discuss any potential misuse of your work? See ethics statement below.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? See ethics statement below.

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes. We read and conform to the AAAI Code of Professional Ethics and Conduct.

2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? To the best of our ability. See Background.
   (b) Have you provided justifications for all theoretical results? We argue for our theoretical claims in the results section
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? See discussion
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? See discussion
   (e) Did you address potential biases or limitations in your theoretical framework? See background and discussion
   (f) Have you related your theoretical results to the existing literature in social science? See background and discussion
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? See discussion
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? No theoretical proofs
   (b) Did you include complete proofs of all theoretical results? No theoretical proofs
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Not a machine learning experiment.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Not a machine learning experiment.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Not a machine learning experiment.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Not a machine learning experiment.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Not a machine learning experiment.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Not a machine learning experiment.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? We site the pushshift reddit dataset
   (b) Did you mention the license of the assets? The dataset is unlicensed
   (c) Did you include any new assets in the supplemental material or as a URL? We do not
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Consent was not obtained, we do not discuss it.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? We mentioned that we remov some data likely to have nsfw material, but do not discuss this further. The data surely contains offensive content and PII, but we do not analyze these elements of it.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? Not releasing a new dataset
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? Not releasing a new dataset
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? Not crowdsourcing or human subjects research
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Not crowdsourcing or human subjects research
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Not crowdsourcing or human subjects research
   (d) Did you discuss how data is stored, shared, and deidentified? Not crowdsourcing or human subjects research

Ethical Statement

This work's mode of inquiry is to analyze large-scale data from social media that is publicly available. Some have concerns about the ethics of such analysis, such as the possibility that the people whose activities created the data we analyze would have been likely to expect their activity would be so used. The author is sympathetic to these serious concerns. That said, the nature of this analysis exposes no individual to scrutiny and the reporting obscures even the identities of the communities analyzed. As a result, the likely and potential harms of this work to individuals are low. The ultimate goal of this work is to improve online communities through a scientific understanding of how they develop and relate to each other. The negative consequences of this work if it is used to improve online communities that cause harm are easy to imagine. However, the author is optimistic in believing that the benefits of this knowledge to socially beneficial communities will ultimately outweigh the negative effects.