Identifying Competition and Mutualism Between Online Groups

NATHAN TEBLUNTHUIS, University of Washington
BENJAMIN MAKO HILL, University of Washington

Platform~~s hosting online group~~s often host ~~have~~ multiple online groups with highly overlapping topics and members. How can researchers and designers understand how interactions between ~~these~~ related groups affect ~~measures of~~ group health ~~over time~~? Inspired by population ecology, prior social computing research has studied competition and mutualism among related groups by correlating group size with degrees of overlap in content and membership. The resulting body of evidence is puzzling as overlaps seem sometimes to help and other times to hurt. We suggest that this confusion results from aggregating inter-group relationships into an overall environmental effect instead of focusing on networks of competition and mutualism among groups. We propose a theoretical framework based on community ecology and a method for inferring competitive and mutualistic interactions from time series participation data. We compare population and community ecology analyses of online community growth by analyzing clusters of subreddits with high user overlap but varying degrees of competiti~~ve~~on and mutualis~~tic~~m ~~network structures~~.

ACM Reference Format:
Nathan TeBlunthuis and Benjamin Mako Hill. 2018. Identifying Competition and Mutualism Between Online Groups. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 46 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Although the fact is frequently ignored in social computing scholarship, online groups do not exist in isolation.[1] Indeed, ~~even~~ al though studying interdependence between online groups ~~adds enormous~~ is difficult and complex~~ity to research projects~~ [45], ~~a growing body of empirical~~ research in social computing has sought to quantify how online groups share users or topics [23, 25, 44, 80], and how such interactions relate to outcomes like the emergence of new groups [79], contributions to

[1]We use the term "online group" instead of "online community" to help avoid confusion with our term "community ecology" which plays an important conceptual and analytic role in our paper.

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

peer produced knowledge [84], and the spread of hate speech [17]. Although this work has demonstrated that inter group interactions matter, very little intercommunity research has tackled questions of community success—i.e., why some online groups succeed in maintaining active and long-lived communities while most do not. Can intercommunity relationships explain whether online groups will grow or decline?

Several ~~A series of~~ studies in social computing have drawn from organizational ecology to answer this question [71, 85, 91, 92]. Inspired by the ecological study of biological systems, organizational ecology is an influential body of theory in sociology that studies competition and mutualism among human organizations [5, 40]. Although ecological studies of firms and social movements have developed a clear and established body of theory with strong empirical support [5], similar studies of online groups have yielded inconsistent results that differ both from one context to another and from theoretical predictions. For example, wikis whose memberships overlap with other wikis survived longer [91], but Usenet groups with overlapping memberships failed more quickly [85].

We argue that these confusing results are the result of a conflation of concepts and measures from two distinct strands of theory in organizational ecology: *population ecology* and *community ecology*. Both define competition as a form of interdependence that *decreases* growth and mutualism as one that *increases* growth. However, population ecology focuses on modeling the how overlapping resources among groups affect their subsequent growth, decline, or survival [3, 5, 27]. It does not attempt to directly study competitive and mutualistic interactions. On the other hand, community ecology recognizes that groups often exist within "ecological communities," or clusters of highly related entities ~~groups~~, and provides an approach for inferring competitive and mutualistic interactions among these ~~groups~~. Although the stated goal of ecological research in social computing has been to understand how groups influence each others' ability to sustain participation, ecological research in social computing has relied exclusively on concepts and measures from population ecology. This paper seeks to explain the puzzling set of findings in ecological social computing research by introducing community ecology ~~to social computing research and resolving the mismatch between goals and analytic approach~~.

We do so ~~pursue this goal~~ in a three-part empirical study using a dataset drawn from the ~~top~~ 10,000 communities on Reddit with the most contributors subs? to analyze 641 clusters of online groups with overlapping participants. In Study A, we conduct the most important typed a population ecology analysis a ~~by~~ testing and find support for the theory of what is called density dependence theory ~~and find that online group growth is greatest at intermediate levels of density as predicted~~. This analysis suggests that high degrees of uses overlap are associated with competition. ~~Next,~~ Then in Study B, we introduce our method for community ecology analysis that infers networks of competitive and mutualistic interactions by using clustering analysis and vector autoregression (VAR) models of community sizes over time [12, 47, 77]. We illustrate the method in four case studies and present a large-scale computational analysis showing that mutualistic interactions are far more common than competitive ones. Finally, in Study C, we bring Study A and Study B together to compare population ecology and community ecology by ~~E~~extending the density dependence models from Study A with a variable accounting for competition and mutualism.

While we find that adding this variable does not help predict growth, ~~However,~~ including ecological interactions in our VAR models improves time series forecasting.

We discuss how these findings illuminate differences between population ecology and community ecology and show how the two perspectives are complementary. While Study A suggests that competition is strongest when user overlap is high, ~~but~~ Study B finds widespread mutualism among groups with overlapping membership. Although these findings might seem contradictory, they reflect how population ecology studies overlapping resources related to favorable or unfavorable environmental conditions, while ~~but~~ community ecology studies competitive and mutualistic interactions playing out in local networks of specific groups. ~~Previous research into overlapping online groups had not considered~~ mutualistic and competitive ~~interactions~~ within clusters of highly related groups. By demonstrating that ~~these~~ interactions are important and by showing how to measure them, this paper lays the groundwork for future research to investigate and ~~ultimately~~ design for interdependence between online groups that supports their growth and success.

2 RELATED WORK

Online groups are important sites for social support [24], entertainment [28], information sharing [7], and political mobilization of disinformation campaigns and protest movements [8, 19, 55]. Although ~~An~~ online group's ability to achieve its goals depends on attracting and retaining contributors, ~~but of the many founded online groups~~, few ~~only an extreme minority~~ develop a sizable group of participants [7, 26, 48, 54, 56]. Many attempts to explain the success and growth of online groups look to properties of individual groups like characteristics of founders [56], language use [22], turnover [21], and designs for regulating behavior [38, 81].

Recent research suggests that interdependence among online groups is also important to explaining ~~their~~ successes and failures [20, 50, 79, 80]. For example, Banning hate subreddits reduced hate-speech in related subreddits [17]. In a very different context, there is evidence that Reddit and Stack Overflow receive substantial benefits from ~~information~~ activity on Wikipedia, ~~but this drives little reciprocal editing to Wikipedia~~ [84]. ~~and editors make valuable and qualitatively different contributions across different languages of Wikipedia [37]. In addition, growth trajectories of online groups initially about similar topics can diverge [89]~~. Our work contributes to this literature by providing a new conceptual lens and statistical method for studying competition and mutualism between online groups.

Like prior ecological research in social computing and information systems, we build on resource dependence theory (RDT) [10, 85], ~~which theorizes how online groups depend on distinct types of resources.~~ As we discuss in §2.1, the nature of these resources makes possible conditions for mutualism or competition. In §2.2, we explain how prior ecological studies of online groups extended RDT to consider how overlapping resources between communities can drive competition and mutualism and propose our first hypothesis which replicates part of these studies in Reddit, our empirical context. Finally, in §2.3, we draw anew from biology and organizational ecology to present our community ecology approach and propose hypotheses to validate its usefulness for predicting the growth of online groups.

Unnecessary recapping

2.1 Online groups depend on resources

Butler [10] introduces RDT to ~~propose~~ agree that growth in online groups is driven by positive feedback as participants contribute resources such as content, information, attention, or social interactions, which motivate further contributions by subsequent participants. That said, online groups do not grow forever and RDT explains that growth is self-limiting because costs of participation increase in larger groups [10, 11].

Ecological approaches recognize that interrelated online groups may share resources with one another in ways that constrain their growth and survival. *Rival* resources like participant's time, attention, and efforts raise the possibility of competition because they become unavailable to others when used by one group [7, 59, 67, 73]. RDT suggests that declines in online participation can be explained in terms of competition over important rival resources [85]. ~~So online groups that provide similar benefits may be the most likely competitors because once someone has obtained satisfying benefits from one group they may go offline or switch to another activity instead of seeking similar benefits from competitor groups~~.[2]

On the other hand, online connective and communal groups also rely on *nonrival* resources. They can even produce public goods like opportunities to communicate and collections of information [35] which can ~~even~~ be "antirival" when their usefulness increases as a result of others using them [59, 87]. For example, the usefulness of a communication network increases as more people join it [35, 52]. Similarly, the usefulness of an information good can increase as more people come to know, refer to, and depend upon it [59, 87] ~~as when awareness that an online group provides an audience~~

[2]Economists refer to these as "substitutes.'

~~motivates participation [90]~~. If multiple online groups help build the same connective or communal public goods, they may form mutualistic interactions where contributions to one group may "spill over" and motivate participation in mutualist groups [92]. Ecological approaches ~~promise~~ seek to understand how different types of resources will limit or promote growth ~~in online groups~~.

2.2 Density Dependence and Overlapping Resources

While ~~our work~~ this paper focuses on the ecological study of online groups, other social computing and HCI scholars have used the term "ecology" (and related concepts like "ecoystem" and "environment") to denote assemblage of sites, devices, or platforms [65, 86]. We use the term more narrowly in reference to the conceptual and mathematical models of ecological dynamics. In particular, our work builds on a tradition rooted in *organizational ecology*. First developed in the late 1970s by sociologists studying interactions between firms, organizational ecology was inspired by, and has drawn closely from, ecological studies in biology [40].

Because online groups bear similarities to traditional organizations, organizational ecology provides a compelling theoretical framework for understanding interdependence among online groups. It has inspired at least three high-quality empirical studies of how resources shared by online groups shared shape their growth, decline, or survival [85, 91, 92]. These studies draw from the *population ecology* strand of organizational ecology that studies ecological dynamics within a population of groups. In organizational ecology, populations have been defined as sets of organizations sharing an organizational industry or business model [41]. In social computing, populations have been defined as online groups sharing a given social media platform [85, 91, 92].

While Population ecology ~~includes~~ involves several distinct theoretical propositions, ~~but~~ *density dependence theory* (DDT) is perhaps the most prominent and is the subject of all three prior ecological studies of online groups [85, 91, 92]. DDT models competitive or mutualistic forces in a population of groups as a function of *density* which, in the earliest and most influential studies of DDT, is simply the size of the population. ~~Therefore~~ In this way, ~~this theory models~~ DDT assumes that every group in the population ~~as~~ is facing the same competitive and mutualistic pressures [1]. However, online groups sharing a platform have diverse topics [50], norms [18, 31], and user bases [80] ~~and thus different resource needs~~. Because groups sharing few resources are unlikely to be strongly interdependent, ~~so~~ ecological studies of online groups have modeled ~~a different notion of~~ density dependence based on the concept of *overlap density* [5, 27, 85, 91, 92]. Rather than the number of groups that exist in a population, overlap density measures the extent to which an one group's members or topics overlap with all other groups'. Overlap density thus locates each group in its own *niche* or local *resource environment* defined by its distinctive topic and membership.

DDT proposes a model for the growth organizational populations that has a similar structure to the Butler's [10] RDT model for the growth of online groups. In DDT, mutualism is the engine of positive feedback driving population growth. Organizational ecologists show how successful organizations in an emerging industry develop nonrival resources like the legitimacy of a business model or industrial know-how ~~and thereby~~ that attract new organizations to enter the market [14, 41]. Similarly, a population of online groups, such as those sharing a platform, may grow establshd

in size as their platform gains in popularity, as ~~existing~~ groups spin off new ones, and as useful knowledge develops that can be shared between groups [79, 92].

In RDT, growth of online groups is self-limiting because of the challenges in managing large groups ~~such as regulating behavior and communication overhead~~ [10]. In DDT, competition among population members over rival resources limits ~~the population's~~ growth [41]. DDT thus proposes a trade-off in which low density reflects limited opportunities for mutualistic contributions of nonrival resources like legitimacy, connectivity, and knowledge, but high density reflects competition over rival resources. Therefore, DDT predicts that the relationship between density and positive outcomes like growth or survival is ∩-shaped (inverse-U-shaped) [5, 14].

Tests of DDT in populations of online groups yield inconsistent results. In Wang et al. [85], user overlap~~s~~ in Usenet newsgroups are associated with decreasing numbers of participants. Similarly, TeBlunthuis et al. [82] find that topical overlaps between online petitions are negatively associated with participation. By contrast, Zhu et al. [92] find that membership overlap is positively associated with increasing survival of new Wikia wikis. Only Zhu et al. [91] find support for the ∩-shaped relationship predicted by DDT in an enterprise social media platform.

using data from

In Study A, we provide a ~~novel~~ test of DDT ~~in the context of~~ Reddit ~~to provide an empirical basis for comparing the population ecology and community ecology models of~~ competition and ~~mutualism~~. The classical logic of DDT appears reasonable in the context of Reddit based on the arguments of Zhu et al. [91] and [92]. Furthermore, as Kraut et al. [57] argue, people's commitment to subgroups complements their commitment to a broader group. In addition, Tan and Lee [80] observe

makes no sense

that Reddit accounts are more likely to cease activity when they post in fewer communities. This suggests that subreddits with greater overlap density may receive mutualistic benefits~~ from their member's commitment to the Reddit platform~~. On the other hand, ~~members of a subreddit with high overlap density may have little commitment to it in particular. They seem more involved in Reddit overall than in the particular community. Therefore~~, trade-offs between commitment to a subgroup and commitment to the broader platform provide an additional plausible mechanism for density dependence: ***(H1)*** *The relationship between overlap density and the growth of online groups is ∩-shaped (inverse-U-shaped).*

DDT proposes that very high levels of density will decrease growth because of increasing forces of competition within a niche. However, to conclude that groups with the greatest membership overlaps are likely competitors would be to commit a well known statistical fallacy [69, 72]. The density of a group's environment suggests that it faces competition or mutualism, but it does not tell us which overlapping communities are competitors and which are mutualists.

~~DDT therefore relates resource overlaps to the growth of online groups, yet stops short of inferring competitive or mutualistic interactions among them. It does not provide a way of learning when and why groups are mutualists or competitors and this limits its ability to inform designs that take these interactions into account~~. Community ecology overcomes this limitation of DDT.

2.3 Introducing Community Ecology

Perhaps the most natural way to understand the distinction between the population ecology and community ecology is in ~~how~~ when they ~~conceive of~~ behave ecological dynamics like competition and mutualism play~~ing~~s out ~~in different levels~~ [3]. ~~Where~~ while population ecology locates competition and mutualism within an environmental niche, community ecology locates competition and mutualism in networks of interdependent groups called *ecological communities* [1]. In organizational ecology, this can mean studying interactions between different organizational populations [e.g. 63, 78], or networks of interactions between organizations [e.g 61, 70]. While varying conceptions of community ecology are found in the organizational ecology literature [32], our approach follows that of Aldrich and Ruef [1] and Hawley [42] ~~in seeing groups related by competitive and mutualistic interactions as important objects of analysis~~.

Community ecology focuses on *ecological interactions*~~, ways groups affect one another through changes in group size, between groups in ecological communities~~ [1]. Ecological interactions can be mutualistic when one group has a positive influence on the second such that growth in the first group leads to growth in the second. They can also be competitive if one group has a negative effect on the second such that growth in the first group leads to decline in the second. Ecological interactions can be reciprocated if mutualism (or competition) from one group to another group is returned in kind. An ecological interaction can also be mutualistic in one direction and competitive in the other. The competitive or mutualistic interactions in an ecological community are quantified by the *community matrix*, a central analytical object in community ecology in both biolog~~ical~~y and organiza-tion~~al~~ ~~ecology~~ science [1, 66, 83].

In Study B, we demonstrate community ecology by inferring networks of ecological interactions in ecological communities on Reddit. Because our understanding of community ecology theory does not suggest hypotheses about what we will find, we conduct an exploratory data analysis to find out whether mutualism or competition among subreddits is more ~~typical~~ common on Reddit and present case studies illustrating the types of ecological communities ~~found on Reddit~~ we identify.

In Study C we build upon our analyses from Study A and Study B ~~and seek to demonstrate that~~ by testing whether community ecology can explain~~s~~ the growth and decline of online groups in ~~a distinct~~ ways that ~~from~~ population ecology cannot. We do this by analyzing in two different ways whether accounting for ecological interactions helps predict future group sizes. ~~We expect it to do so because resource overlaps as modeled by DDT may be a poor proxy for the degree to which a group's environment is competitive or mutualistic.~~

In general, competition for overlapping resources will have no effect on group growth if something besides the overlapping resource limits growth [83]. For example, two wikis might share a large number of contributors (have high user overlap), but their growth might be limited by a lack of core contributors who perform important administrative tasks like policy making and software administration [92]. Community ecology relaxes the assumption that competition and mutualism are caused by user overlap density and instead seeks to infer ~~them~~ these relationships from data. We test the importance of this conceptual shift for predicting growth by testing two hypotheses. The first uses a model comparison approach to test if adding a measure of ecological interactions to the density dependence models from Study A ~~will better~~ improves predict ws of growth ~~in online groups~~: (***H2***) *A model with ecological interactions and density dependence predicts growth in online groups better than density dependence alone.*

Support for H2 may be a relatively low bar for assessing whether ecological interactions are important factors shaping the growth of online groups because of confounding moderator~~s~~ or mediator~~s~~ variables related to the occurrence of ecological interactions. Therefore, we also use a time series forecasting approach to test whether modeling ecological interactions is useful for predicting growth in online groups. While this does not directly compare population ecology and community ecology, it validates that ecological interactions are important: (***H3***) *The addition of ecological interactions to a baseline time series model improves the predictive forecasting performance.*

3 MATERIALS & METHODS

The presentation of our materials and methods is organized as follows: First we introduce the methods and measures for Study A, beginning with *user overlap* which is aggregated into *overlap density* to predict subreddit *growth* in a loglin~~ear regression model~~. Then, for Study B, we present our ~~clustering~~ procedure for identifying ecological communities on which we fit a series of ~~VAR~~ models predicting *group size*. To explore the types of ecological communities found on Reddit, we derive two measures from these models for each cluster: *average ecological interaction* which quantifies the degree of competition and mutualism in the ecological community and *ecological interaction strength* which quantifies its overall intensity of

~~ecological interactions. Next, we draw competition-mutualism networks in example ecological communities based on interpreting the VAR models using impulse response functions (IRFs) Then, in Study C, we test H2 to compare community ecology and density dependence theory by adding *subreddit average mutualism* to the model from Study A. Finally, we test H3 by evaluating whether including ecological interactions in the VAR models improves time series forecasting.~~

3.1 Data

Our data are drawn from ~~We analyze~~ the publicly available Pushshift archive of Reddit submissions and comments which we obtained from December 5th 2005 to April 13th 2020 [6]. From this dataset, we limit our analysis to ~~including~~ the ~~top~~ 10,000 subreddits ~~by~~ with the highest number of comments. There are 702 subreddits larger than the smallest subreddit included in our dataset having a majority of submissions marked "NSFW," which typically indicates pornographic material. As others have done in large-scale studies of Reddit [e.g., 23], we exclude these subreddits to avoid asking members of our research team to inspect~~ing~~ or validat~~ing~~ clusters including ~~them~~ pornography. The top 10,000 subreddits include ~~smaller communities and provide~~ a sufficiently large number of ecological communities for our statistical analysis.

3.2 Study A: Density Dependence Theory

3.2.1 User overlap $o_{i,j}$ quantifies the degree to which two subreddits (i and j) share users. Zhu et al. [92] and Wang et al. [85] both measure user overlap between two groups by counting the number of users contributing to both ~~communities~~ groups at least once and exclude users who appear in more than 10 ~~communities~~ groups. In our preliminary analysis, we found that this measure led to similarity measures and clusters with poor face validity. These issues may have stemmed from how Reddit users often peripherally participate in many communities~~,~~ while participating heavily in few [39, 80, 88]. Therefore, our measure of user overlap follows Datta et al. [23] by using number of comments each user makes in each pair of communities.

To measure user overlap between subreddits, we first build user frequency vectors by counting the number of times each user comments in each subreddit. ~~Then,~~ we prevent giving undue weight to subreddits with higher overall activity levels by normalizing the comment-counts for each subreddit by the maximum number of comments by a single author in the subreddit:

$$f_{u,j} = \frac{n_{\mathrm{u,j}}}{max_{v \in \mathrm{J}} n_{v,j}} \tag{1}$$

where $n_{u,j}$, the user frequency, is the number of times that user u authors a comment in subreddit j.

This results in a user frequency vector F_j for each subreddit that is sparse and high-dimensional, having one element for each user account that comments in any subreddit in our dataset. Next, we use LSA to reduce the dimensionality of the user frequency vectors. LSA is based on the singular value decomposition and is common in natural language processing and information retrieval. LSA preserves subreddit similarities while removing noise and dealing with sparsity ~~by capturing correlations between Reddit users~~ [29]:

$$\mathbf{F} = \mathbf{U}\Sigma\mathbf{V}^T \tag{2}$$

$$\tilde{F}_j = \mathbf{U_k}^T F_j$$

F is the matrix where columns are author frequency vectors F_j and $\mathbf{U\Sigma V}^T$ is its singular value decomposition. Truncating the singular value decomposition to use only the first k left-singular vectors gives $\mathbf{U_k}$. Left-multiplying a subreddit's author frequency vector by $\mathbf{U_k}$ transforms the high-dimensional author frequencies into $\tilde{F}_j$, their approximation in the k-dimensional space.

We then obtain our measure of *user overlap* by taking the cosine similarities between the resulting vectors for a pair of subreddits:

$$o_{i,j} = \frac{\tilde{F}_j \cdot \tilde{F}_i}{||\tilde{F}_i||||\tilde{F}_j||} \tag{3}$$

where $||\tilde{F}_i|| = \sqrt{\sum_{x=1}^{k} \tilde{f}_{x,i}^{\ 2}}$ is the euclidean norm of the transformed user frequencies for subreddit i.

3.2.2 Growth is the dependent variable in our density dependence models testing H1 and is also used in our test of H2 as part of Study B. Growth is measured as the change in the (log-transformed) size of a subreddit over the final 24 weeks of our data, from to November 4$^{\text{th}}$ 2019 to April 13$^{\text{th}}$ 2020.

3.2.3 Overlap density d_i is the normalized average user overlap for a given subreddit. It is the independent variable in our density dependence model testing H1:

$$d_i^* = \frac{1}{|S| - 1} \sum_{j \in R; j \neq i} \mathrm{o}_{i,j}$$
$$d_i = \frac{d_i^*}{\max_j d_j^*} \tag{4}$$

where S is the set of groups in our dataset.

3.2.4 *Regression model for H1.*

To test H1, we fit Model 1 which has first and second-order terms for overlap density to allow for a curvilinear relationship between *overlap density* and *growth*.

$$\text{Model 1} \qquad Y_i = B_0 + B_1 d_i + B_2 d_i^2 \tag{5}$$

where Y_i is the growth of subreddit i and d_i is its overlap density.

3.3 Study B: Introducing Community Ecology

3.3.1 *Clustering to identify ecological communities.*

Analyzing networks of ecological interactions is the key difference between community ecology and population ecology. ~~In Study A, we set out to survey the types of ecological communities found on Reddit to provide a comparison with a large-scale population ecology analysis.~~ To identify ecological communities of related subreddits, we use a clustering procedure on the user overlap measure described above in §3.2.1. We selected a clustering model using grid-search to obtain a high silhouette coefficient [74]. The silhouette coefficient captures the degree to which a clustering creates groups of subreddits with high within-cluster similarity ~~relative to similarity with subreddits in other clusters~~. ¶ Our description of our measure for user overlap in §3.2.1 does not explain how we choose the number of LSA dimensions k ~~because we do so here~~.

To do so, We ran the affinity propagation [33], HDBSCAN [62] and k-means clustering algorithms and selected the algorithm, hyperparameters, and LSA dimensions k that resulted in the clustering with the best silhouette coefficient having less than 5,000 isolated subreddits, and at least 50 clusters. We limit the number of isolated subreddits because some choices of hyperparamters for the HDBSCAN algorithm could improve the silhouette coefficient, but at the cost of greatly increasing numbers of isolated subreddits. Choosing a relatively high limit to the number of isolates helps ensure that our clusters contain highly related communities. We chose an HDBSCAN clustering with 731 clusters, 4964 isolated subreddits, $k = 600$ LSI dimensions, and a silhouette score of 0.48. We exclude the isolated subreddits from further analysis. More details about our clustering selection process are found in the online supplement.

We ~~additionally~~ evaluate the external validity of the chosen clustering using the purity evaluation criterion [60]. To do so, An undergraduate research assistant examined a random sample of 100 clusters including 744 subreddits. By visiting the subreddits and using her own judgment, ~~she~~ the undergrad flagged subreddits that did not seem like a good fit for their assigned cluster. Using her labels and excluding 25 subreddits that have been deleted, made private, or banned, we calculated the purity of our clustering as 0.92. This means that we believe that 92% of subreddits belong to their assigned cluster. ~~Note that although we clustered subreddits based on user overlap, we obtain a high purity score based on a subjective evaluation of the subreddit's content~~.

3.3.2 Group size is the dependent variable of the models we use to infer ecological interactions. Measured as the number of distinct commenting users in a subreddit each week, group size quantifies the number of people who participate in a subreddit over time. Typical of social media participation data, group size is highly skewed; ~~and~~ Therefore, we transform it by adding 1 and taking the natural logarithm.

3.3.3 *Inferring ecological interactions using Vector Auto Regression.*

The community matrix Φ of ecological interactions can be inferred from time series data using vector autoregression models (VAR models). VAR models are a workhorse in biological ecology because VAR(1) models (i.e. VAR models with a single autoregressive term) have a close relationship to the Gompertz models of population growth which are widely used in ecology [47]. Even in the presence of unmodeled nonlinearities, VAR(1) models can reliably identify competition or mutualism in empirically realistic scenarios [16]. VAR models also been widely adopted in the social sciences, particularly in political science and in macroeconomics [9].

VAR(1) models can intuitively understood as a generalization of auto-regressive AR(1) models in time series analysis. But while AR(1) models predict the state of a single time series as a function of its previous value, VAR(1) models simultaneously predict multiple time series as a function of the values of every other variable in the system [12, 47]:

$$Y_t = B_0 + B_1 t + \sum_{k \in K} A_k x_{k,t} + \sum_{j \in M} \Phi_j y_{j,t-1} + \epsilon_t \tag{6}$$

where Y_t is a vector containing the sizes of a set of online groups (M) at time t. B_0 is the vector of intercept terms and B_1 is the vector of linear time trends ($b_{1,j}$) for each community (j). Φ_j represents the influence of $y_{j,t-1}$, the size of the j^{th} online group at time $t-1$ on Y_t. Φ_j is a column of Φ, a matrix of coefficients in which the diagonal elements correspond to intrinsic growth rates (marginal to the trend) for each online group and the off-diagonal elements are inter-group influences, and ϵ_t is the vector of error terms

Additional time-dependent predictors ($x_{k,t}$) can be included in the vectors X_k with coefficients a_k. Because subreddits are created at different times, growth trends must begin only after the subreddit is created. We use X_k to cancel out trends during the period prior to the creation of subreddits. For each community j created at time t_j^0 we fill X_j with the sequence $[1, 2, 3, \ldots, t_j^0 - 1, 0, 0, 0, \ldots]$. In other words, X_j adds a trend only during the period prior to the first comment in subreddit j. We fix elements $a_{j,i}$ of A_j equal to 0 unless $i = j$, so the counter trend only influences subreddit j. This effectively sets $a_{j,j}$ approximately equal to $-b_{1,j}$.

not after?

We fit VAR(1) models using ordinary least squares as implemented in the `vars` R package to predict the group size each week using over the history of each subreddit prior to November 4th 2019 [68]. We hold out 24 weeks of data for forecast evaluation and fit our models on the remainder. To ensure that sufficient data is available for fitting the models, we exclude 946 subreddits and 89 clusters having less than 156 weeks of activity.

3.3.4 Characterizing ecological communities. In Study B, we interpret the community matrix Φ as a directed network of ecological interactions, a *competition-mutualism network* [47]. Although the elements of Φ correspond to direct associations between group sizes [66], ecological interactions can also be indirect. Consider 3 one-directional interactions between three groups (a, b, c) such that growth in a predicts decreased growth in b ($\phi_{a,b} < 0$), growth in b predicts decreased growth in c ($\phi_{b,c} < 0$), but a and c do not directly interact $\phi_{a,c} \approx 0$.

This does not necessarily mean that groups A and C are independent. Rather, an exogenous increase in A predicts a decrease in B and thereby an eventual increase in C. Such indirect relationships are analyzed by using impulse response functions

(IRFs) to interpret a VAR model [9]. In large VAR models containing many groups, the great number of parameters can mean that few specific elements of $\mathbf{\Phi}$ will be statistically significant, even as many weak direct relationships can combine into statistically significant IRFs [12].

3.3.5 Average ecological interaction $\overline{m}$ measures the extent to which an overall ecological community is mutualistic or competitive by taking the mean point estimate of off-diagonal coefficients of $\mathbf{\Phi}$:

$$\overline{m} = \frac{1}{|M| - 1} \sum_{i \in M} \sum_{j \in M; j \neq i} \phi_{i,j} \tag{7}$$

if $\overline{m} > 0$ then mutualistic interactions within the ecological community are stronger than competitive ones, and if $\overline{m} < 0$ then competitive interactions are stronger then mutualistic ones.

3.3.6 Ecological interaction strength κ quantifies the overall strength of ecological interactions in an ecological community as the mean absolute value of the point estimates of the off-diagonal coefficients of $\mathbf{\Phi}$:

$$\kappa = \frac{1}{|M| - 1} \sum_{i \in M} \sum_{j \in M; j \neq i} \left|\phi_{i,j}\right| \tag{8}$$

where $\left|\phi_{i,j}\right|$ is the absolute value of the coefficient $\phi_{i,j}$.

Ecological communities of subreddits with overlapping users vary in both the overall strength of ecological interactions and in the overall degree of mutualism and competition between member groups. If an ecological community's average ecological interaction is positive, we say the ecological community is mutualistic. If it is negative, we say the ecological community is competitive. The average ecological interaction can be close to 0 in two ways. First, average ecological interaction strength can simply be low. Alternatively, the ecological community can have a mixture of competitive and mutualistic interactions that cancel one another out when averaged.

3.3.7 Impulse response functions (IRFs) of our VAR(1) models correspond to our visualizations of example competition-mutualism networks in §4.2.1. An IRF predicts how much each group's size would change in response to a sudden increase in the size of each other group [83]:

$$\mathbf{\Theta_t} = \mathbf{\Theta_{t-1}}\mathbf{\Phi}, t = 1, 2, \ldots \tag{9}$$

where $\mathbf{\Theta_t}$ is the impulse response function at time t. $\mathbf{\Theta_0}$ is an M-by-M identity matrix so our impulses represent a log-unit increase of 1 to each group. $\mathbf{\Theta_t}$ is a matrix with elements $\theta^t_{i,j}$ corresponding to the response of group j to the impulse of group i. We draw an edge $i \longrightarrow j$ in the competition-mutualism network if the 95% CI of $\theta^t_{i,j}$ does not include zero at any time $10 >= t > 0$. If $\theta^t_{i,j} > 0$, the edge indicates mutualism and if $\theta^t_{i,j} < 0$ the edge indicates competition.[3] We compute the IRFs with bootstrapped confidence intervals (CI) based on 1,000 samples using the `vars` R package.

3.4 Study C: Predicting growth

3.4.1 Average subreddit mutualism m_j is the independent variable for our test of H2 and measures the average influence of other subreddits in the ecological

[3]In higher-order VAR(p) models that use $p > 1$ past observations as predictors $\theta^t_{i,j}$ can be less than 0 for some t_a and greater than 0 for some t_b. However, this is not possible in the VAR(1) models we use.

community on a given subreddit j, which we calculate by taking the mean of off-diagonal elements of row j of the community matrix:

$$m_j = \frac{1}{|M| - 1} \sum_{i \in M; i \neq j} \phi_{i,j} \tag{10}$$

where M is the set of subreddits in the ecological community and $|M|$ is the number of subreddits in M. We use the mean instead of the sum because different ecologial communities have different numbers of subreddits.

3.4.2 *Regression models for H2.*

We test H2 by using likelihood ratio tests to compare Model 1 and Model 2 which adds *average subreddit mutualism* (c_j) as a predictor. We also fit Model 3 compare to Model 2 to test if overlap density explains variation that average subreddit mutualism does not.

$$\text{Model 2} \qquad Y_i = B_0 + B_1 d_i + B_2 d_i^2 + B_3 m_i \tag{11}$$

$$\text{Model 3} \qquad Y_i = B_0 + B_3 m_i \tag{12}$$

where Y_i is the growth of subreddit i, d_i is its overlap density, m_i is its average subreddit mutualism, and B_0, B_1, B_2, and B_3 are regression coefficients.

3.4.3 *Forecasting growth using ecological interactions.*

To test H3, we evaluate whether modeling ecological interactions improves time series forecasting of future participation in online groups by comparing the model in Equation 6 to a baseline model with the off-diagonal elements of Φ fixed to 0. This baseline model is equivalent to our VAR model, but excludes ecological interactions.

We use two forecasting metrics with differing assumptions: root-mean-squared-error (RMSE) and the continuous ranked probability score (CRPS). RMSE is commonly used, non-parametric, and intuitive, but does not take differing scales of the predicted variable or forecast uncertainty into account. Thus, in our setting it may place excessive weight on forecasts of larger subreddits where errors may have greater magnitude simply because the absolute magnitude of the variance is greater. By rewarding forecasts where the true value has high probability under the predictive distribution, the CRPS accounts for variance in the data and rewards forecasts for both accuracy and precision and is thus a "proper scoring rule" for evaluating probabilistic forecasts [36]. Our CRPS calculations assume that the predictive forecast distribution for each community is normal with standard deviations given by the 68.2% forecast confidence interval. We calculate CRPS using the `scoringRules` R package [49].

4 RESULTS

The organization of our results follows that of our methods. We begin with Study A in which we find, as predicted by H1, that the relationship between overlap density and growth is ∩-shaped relationship. Then, in Study B,we explore a typology of ecological communities along two dimensions: (1) the degree to which a community is mutualistic or competitive, and (2) the overall strength of ecological interactions between the communities member groups. In the 641 ecological communities analyzed in our VAR(1) analysis, we find that mutualistic relationships are much more common than competitive ones. Our case studies illustrate the typology using 4 example ecological communities. Finally, in Study C, we do

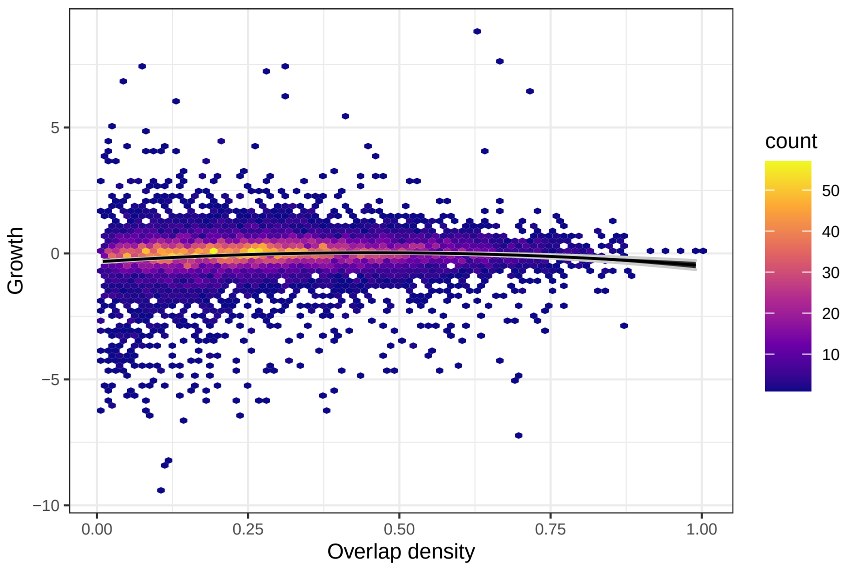


Fig. 1. Relationship between density and growth. A 2D histogram of subreddits with overlap density (log-transformed) on the X-axis and the change in the logarithm of the number of distinct commenting users on the Y-axis. The black line shows the marginal effect of overlap density on growth as predicted by Model 2. The gray region shows the 95% confidence interval of the marginal effect.

~~not find support for H2 as adding average subreddit mutualism to the density dependence model does not improve growth prediction. But we do find, in support of H3, that ecological interactions improve forecasting performance in our time series models.~~

4.1 Study A: Density Dependence Theory

We test the classical prediction of density dependence theory as formulated in H1 using Model 1 which has first- and second-order terms for effect of overlap density on growth. As described in §2.2, H1 hypothesizes that overlap density will have a curvilinear ∩-shaped (inverse-U-shaped) relationship with growth indicated by a positive first-order regression coefficient and a negative second-order coefficient.

what is it?

As predicted, we observe a ∩-shaped relationship between overlap density and growth. Figure 1 plots the marginal effects of overlap density on growth for the median subreddit laid over the data on which the model is fit ~~and~~ Table 1 shows regression coefficients for Models 1-3. For most subreddits, increasing overlap density is associated with higher growth rates. The ~~"sweet spot," or~~ point where increasing density ceases to predict increasing growth and begins to predict decreasing growth is at the 49th percentile. Prototypical subreddits at this overlap density grew slightly (95% CI:[0.001,0.06]). Yet subreddits at the lower and upper extremes of overlap density slightly declined on average. Typical groups at the 20th percentile of overlap density decline by 1.1 (95% CI:[-1.1,-1.15]) members and typical groups at the 80th percentile decline by 1.2 (95% CI:[-1.1,-1.28]) members. per week

While we find support for classical theoretical prediction of a curvilinear, (∩-shaped) relationship between overlap density and growth, this does not imply that relationships between highly overlapping communities are more competitive. ~~Instead our results below show that relationships in ecological communities of subreddits with high user overlaps are typically mutualistic.~~

this doesn't even make sense

4.2 Study B: Characterizing ecological communities

~~As described in §3.3.4, an ecological community can have positive or negative average ecological interaction §3.3.5 indicating if it is competitive or mutualistic and ecological interaction strength §3.3.6 provides a way to distinguish ecological communities with a mixture of competitive and mutualistic interactions from those where ecological interactions are weak.~~ Figure 2 visualizes the distribution of average ecological interaction and ecological interaction strength over the 641

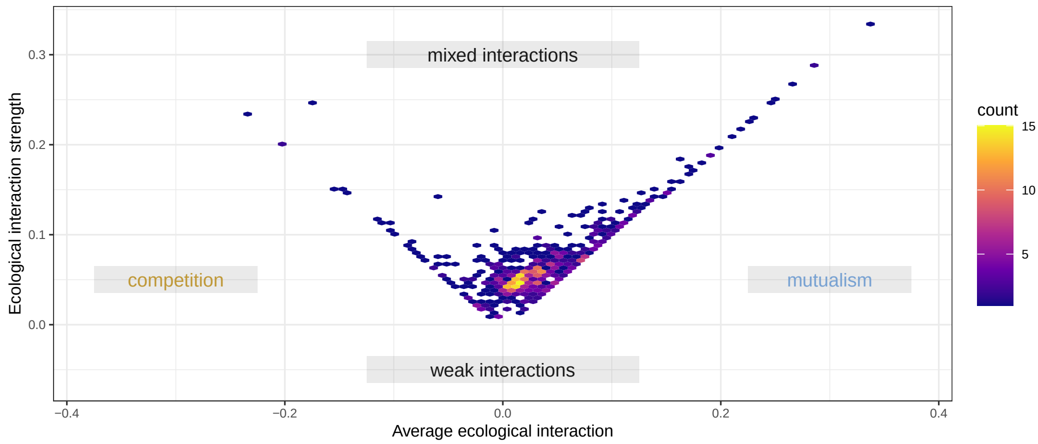


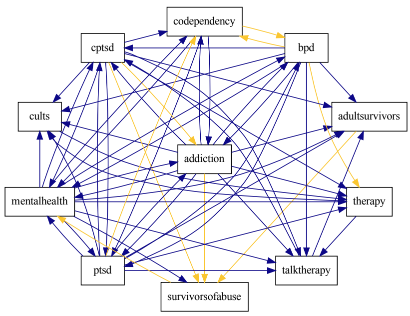
Fig. 2. Two-dimensional histogram showing ecological communities on Reddit in our typology. The X-axis shows the overall degree of mutualism or competition in clusters of subreddits with high user overlap based on the average ecological interaction. The Y-axis shows the ecological interaction strength representing the overall magnitude of competition or mutualism.

ecological communities we ~~analyze~~ identify. We observe ecological communities characterized by strong forms of both mutualism and competition, others having mixtures of the two, and some with few significant ecological interactions. Mutualism is more common than competition with the mean community having an average ecological interaction of 0.03 ($t = 14.5$, $p < 0.001$). We find that 524 clusters (81.7%) are mutualistic. Not only are most ecological communities mutualistic, but more mutualistic ecological communities have greater ecological interaction strength (Spearman's $\rho = 0.58$, $p < 0.001$). ~~Note that due to our clustering procedure, our analysis examines ecological interactions among subreddits with relatively high degrees of user overlap.~~ Therefore, our community ecology analysis suggests that among groups with similar users, mutualistic ecological interactions are more common than competitive ones.

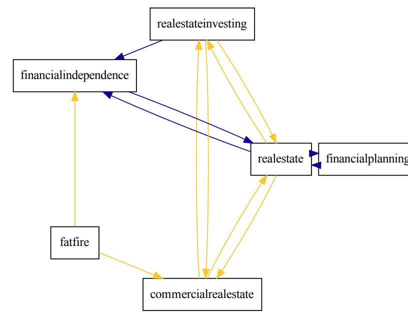
4.2.1 Example ecological communities. We present four case studies to illustrate our typology of ecological communities of online groups. Figure 2 shows that we find clusters of subreddits characterized by mutualism, competition, a mixture of mutualism and competition, and few ecological relationships at all. We select one case from each of these four types using our measures of average ecological interaction (§3.3.5) and ecological interaction strength (§3.3.6). To allow for more interesting network structures, we draw our cases from the 367 large clusters having at least five subreddits.

Figure 3, presents visualizations of competition-mutualism networks representing statistically significant impulse response functions as described in §3.3.7. In the course of our analysis, we also examined the terms of the vector autoregression parameter $\mathbf{\Phi}$, the impulse response functions, and model fits and forecasts, all of which are available in our online supplement. We also visited each community in the clusters and read their sidebars and top posts in order to validate each subreddit's membership in the ecological community and to support our brief qualitative descriptions.

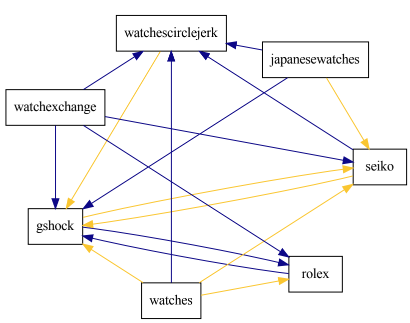
4.2.2 Mutualism among mental health subreddits. To find a case characterized by mutualism, we selected the top 37 large clusters with the greatest average ecological interaction ~~(§3.3.5)~~. From these, we arbitrarily chose one interesting ecological community, the *mental health* cluster, which includes 11 subreddits for supporting people in struggles with mental health, addiction, and surviving abuse. Constitutive subreddits include those focused on specific mental health diagnoses like `r/bpd` (bipolar disorder) and `r/cptsd` (complex post traumatic stress disorder) while others like `r/survivorsofabuse` and `r/adultsurvivors` are support groups.



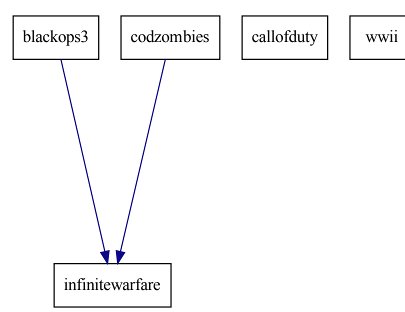
(a) The ecological community of subreddits for supporting mental health and survivors of abuse is dense with largely mutualistic interactions. Some interactions, like that between r/mentalhealth and r/survivorsofabuse are mutualistic in one direction but competitive in the other.



(b) The subreddits about real estate and finance are relatively competitive. We detect reciprocal competitive relationships among the real estate subreddits in the triad including r/realestateinvesting, r/realestate and r/commercialrealestate.



(c) Subreddits about watches are dense with both mutualistic and competitive interactions. There is a reciprocal competitive interaction between r/gshock and r/seiko, a reciprocal mutualistic interaction between r/gshock and r/rolex well as several unreciprocated mutualistic and competitive interactions.



(d) An ecological community of subreddits about Call of Duty video games characterized by relatively sparse ecological interactions. We detect only two mutualistic interactions from r/blackops3 to r/infinitewarfare from codzombies to r/infinitewarfare.

Fig. 3. Network visualizations of commensal relationships in example ecological communities of subreddits with overlapping users. Yellow indicates competition and purple indicates mutualism.

The interactions among these subreddits are dense and ~~largely~~ primary mutualistic as shown in Figure 3a. There are a handful of competitive interactions like the reciprocal competition detected between r/codedependence and r/bpd. We also observe some interactions that are mutualistic in one direction and competitive in the other. For example, growth in `r/addiction` predicts increasing growth in `r/cptsd` even as that growth in `r/cptsd` predicts decreasing growth in `r/addiction`. This suggests a pattern in which `r/cptsd` siphons members from `r/addiction`. That said, the density of mutualistic interactions shown in F~~f~~igure 3a suggests that different subreddits have complementary roles in this ecological community as people turn to different types of groups for help with ~~their~~ interrelated problems. While attempting to explain why different online groups form mutualistic or competitive interactions is left to future research, the example of mental health subreddits shows how groups with related topics and overlapping participants can have mutualistic interactions where ~~as~~ growth in one predicts growth in many of the rest.

4.2.3 *Competition among real estate and financial independence subreddits.*

To find competitive clusters we first selected clusters from the 36 large clusters with the lowest average ecological interaction (~~§3.3.5~~). From these, we chose an ecological community that we label *financial independence*. Among the 6 subreddits in this cluster, `r/realestateinvesting`, `r/realestate` and `r/commercialrealestate` all deal in different aspects of the real estate industry, while `r/financialindependence` and `r/fatfire` (the acronym "fire" means "financial independence/retire early") are focused on building wealth and becoming financially independent ~~and~~ `r/financialplan` is a general purpose subreddit for financial advice.

In contrast to the mental health ecological community, the finance cluster has Mostly ~~a mixture of~~ competitive ~~and mutualistic~~ ties as visualized in Figure 3b. The fact that even this cluster, among the most competitive in our data, contains a number of mutualistic ties reflects just how prevalent mutualism is among subreddits with high degrees of user overlap. That said, we detect three reciprocal competitive interactions among the three subreddits that focus on real estate. The edges from `r/fatfire` to `r/commercialrealestate` and `r/financialindependence` are competitive as well. Interestingly, all the interactions between the general finance subreddits (`r/financialplanning` and `r/financialindependence`) and `r/realestate` are mutualistic.

this is kinda out of context

4.2.4 Mixed interactions among timepiece subreddits. Next, we turn to an example of an ecological community with low average ecological interaction (~~§3.3.5~~) but high ecological interaction strength (~~§3.3.6~~). We first select the 36 ~~large~~ clusters with the average ecological interaction (~~§3.3.5~~) closest to 0. To find an ecological community with a mixture of mutualism and competition, we select from the 15 clusters with the greatest ecological interaction strength (~~§3.3.6~~) and chose the *timepiece* cluster containing 7 subreddits about watches.

from within this group

As shown in Figure 3c, the ecological community of timepiece subreddits is dense with ecological interactions (though not as dense as the mental health subreddits). We observe both reciprocated mutualistic interactions, like that between `r/rolex` and `r/gshock`, and competitive interactions like that between `r/gshock` and `r/seiko`. We also observe numerous unreciprocated competitive and mutualistic relationships like the mutualism between `r/watchexchange` and `r/watchcirclejerk`[4] and the competition between `r/japanesewatches` and `r/seiko`. Though the average ecological interaction among these subreddits is near 0, our analysis reveals a complex ecological community with a mixture of competition and mutualism.

[4]The suffix is widely understood on Reddit to signify a jokey, meme, or satirical subreddit.

4.2.5 Sparse interactions among Call of Duty subreddits. To find a case where ecological interactions are weak, we return to the group of the 36 large clusters with the average ecological interaction ~~(§3.3.5)~~ closest to 0 but select from the 15 clusters within this group with the lowest ecological interaction strength. From these we chose the *Call of Duty* cluster containing five groups about the popular military first person shooter series of video games.

The Call of Duty ecological community is sparse, having only two significant ecological interactions among its 5 member groups. This ecological community includes subreddits about different editions of the series such as `r/blackops3`, `r/infinitewarfar` and `r/wwii` as well as one about a popular spin-off zombie game `r/codzombies` and the more general `r/callofduty` subreddit. We find that the growth in `r/blackops3` or `r/codzombies` predicts growth in `r/infinitewarfare` and no other ecological interactions.

The timepiece and Call of Duty ecological communities illustrate how subreddits with overlapping users can have relatively strong or weak forms of ecological interdependence. Though both clusters are characterized ~~defined~~ by ~~having relatively~~ high degrees of user overlap~~,~~ and ~~have~~ low average ecological interaction, the timepiece cluster has a dense competition-mutualism network while the call of duty network is sparse.

4.3 Study C: Comparing Density Dependence Theory and Community Ecology

We now compare the environmental approach of population ecology with the relational approach of community ecology. In Study B above, we presented examples

	Model 1	Model 2	Model 3
Overlap density	1.50* (0.26)	1.50* (0.26)	
Overlap density^2	−2.08* (0.41)	−2.09* (0.41)	
Average subreddit commensalism		0.12 (0.26)	0.11 (0.26)
Constant	−0.23* (0.03)	−0.23* (0.04)	−0.04* (0.01)
Log Likelihood	-4970	-4970	-4986
Observations	4,090	4,090	4,090
Note:			*p< 0.01

Table 1. Loglinear regression predicting subreddit growth as a function of overlap density. The model supports the prediction of density dependence theory of a ∩-shaped relationship between overlap density and growth.

of diverse ecological communities among subreddits with overlapping members. However, the presence of this diversity this does not mean that ecological interactions are related to the growth of online groups, the key outcome of previous ecological studies. We therefore ~~proposed~~ hypothesize in H2 that ecological interactions will improve the predictive performance of a density dependence model.

4.3.1 *Ecological interactions do not improve growth prediction.*

To test H2, we compare Model 1, our density dependence model having first and second order terms for overlap density, with Model 2, which also includes average subreddit mutualism (§3.4.1) as a predictor. We also examine Model 3, in which the only predictor is average subreddit mutualism. Table 1 shows regression coefficients for our models.

We do not observe a statistically significant association between average subreddit mutualism and growth ($B_3 = 0.12, SE = 0.26$).

Remove JK

~~A~~ likelihood ratio test comparing Model 1 and Model 2 does not support H2 as Model 2 does not predict subreddit growth better than Model 1 ($\chi^2 = 0.23$, $p > 0.05$). ~~Therefore, average subreddit mutualism does not help predict growth compared to the density dependence model alone.~~ Comparing Model 2 to Model 3 shows that overlap density explains variation that average subreddit mutualism does not ($\chi^2 = 33$, $p < 0.001$). Overlap density helps explain a group's future growth, but the overall degree of mutualism or competition a group faces in its ecological community does not. ~~As~~ we discuss in §6, this how reflects ~~how~~ reflect that overlap density captures the hospitality of a group's environment and may be independent of mutualism and competition within its ecological community.

4.3.2 Forecasting accuracy. The likelihood ratio tests in §4.3.1 are limited because improvements in predictive performance (or lack thereof) may be due to unobserved factors predictive of growth that are correlated with average subreddit mutualism. We hypothesized in H3 that the inter-group dependencies in our VAR models can better forecast the size of subreddits compared to baseline time series models that do not account for ecological interactions. As described ~~above~~ in §3.4.3, we test H3 by comparing two forecasting metrics: the root-mean-squared-error (RMSE) and the continuous ranked probability score (CRPS).

VAR models including ecological interactions have forecasting performance superior to the baseline model in terms of both RMSE and CRPS. We evaluate forecast performance for all subreddits which were assigned to clusters for 24 weeks. The RMSE under the baseline model (0.84) is greater than the RMSE of the VAR models (0.75) and the CRPS of the baseline model (72,853) is also the greater than the CRPS of the VAR models (72,669). This reflects a substantive improvement in forecast accuracy robust to the choice of forecasting metric.

Our baseline model contains a constant term and a trend term for each group and therefore accounts for all time-invariant within-community variation. ~~Therefore,~~ the improvement in forecasting performance comes from modeling ecological interactions in ways not captured by to measure overlap density, which is a subreddit-level variable that does not vary over time.

Because we know that that we cannot meaningfully address

5 THREATS TO VALIDITY

Our work is subject to several important threats to validity. Firstly, We study ecological communities on only one platform hosting online groups and ~~our~~ results ~~may~~ will not generalize to other platforms or time periods. cannot know if our The method we propose for identifying ecological interactions between online groups has limitations common to all time series analysis of observational data. Additionally While our community ecology approach assumes that ecological interactions drive dynamics in the size of groups over time and cause groups to grow or decline, drawing causal inference using our method would depend on several untestable assumptions. For example, our ability to infer causal relationships might be limited if groups we do not consider—including groups on other platforms—play a role in an ecological community. Therefore, we refrain from claiming that the relationships we infer are causal.

Potential omitted variables might also include additional time lags of group size. Although we chose to use VAR(1) models with only 1 time lag, we hope future work can improve upon our approach and model more complex dynamics with additional lags. ~~Similarly,~~ R ~~r~~egression estimates in Models 1-3 may be confounded by omitted variables and cannot support causal interpretations.

Like most other time series analysis, vector autoregression assume that error terms are stationary. This is difficult to evaluate empirically and may not be realistic [12]. Future work might relax these assumptions using more complex models with time-varying parameters, state space models [9], nonlinear time series models [15, 51], or stationarity-enforcing priors [43]. Such approaches may require additional contextual knowledge and be difficult to scale to an analysis of hundreds of different ecological communities, but may prove fruitful in future work focusing on ecological dynamics within ecological communities of interest. Such models may also be useful in future work investigating how ecological interactions change over time.

Additional threats to validity stem from our use of algorithmic clustering to identify ecological communities. Organizational ecologists have rarely attempted to estimate the full community matrix for an entire population containing a large number of groups because of data and statistical limitations [e.g. 75, 78]. For instance, 100~~,000,000~~ million possible ecological interactions exist within a set of 10,000 communities. Attempting to infer them all raises considerable computational and statistical challenges. This ~~makes it necessary to narrow the scope to the ecological communities of interest in ways appropriate to the research question~~. We chose to use a clustering analysis to explore the typical ecological communities on a platform. However, clustering algorithms are limited and principled definitions of an ecological community based on qualitative contextual knowledge may be more appropriate for focused studies of particular ecological communities.

While we choose clusters based on high degrees of user overlap, and validate our clustering in terms of the silhouette coefficient and purity criteria, ~~had we instead clustered based on topical similarity~~, we ~~may~~ might have obtained different results. due had clustered using a different model Additionally Our efforts to obtain clusters with a high silhouette coefficient lead use to remove a large number of subreddits from our analysis. Thus, our results are not representative of Reddit overall, but only of those subreddits that were included in our analysis. Furthermore, clustering algorithms like the one we use may not have unique solutions and different initial conditions might lead to different ~~clusters.~~ results. While these algorithms allow us to scale up our analysis, future investigations should also consider qualitative approaches to constructing ecological communities.

6 DISCUSSION

To introduce community ecology and compare it to population ecology, we presented three studies. In Study A, we found support for H1 showing—as predicted by density dependence theory—that overlap density has an ∩-shaped association with subreddit growth. Subreddits with moderate overlap density in our data declined less than subreddits with either very low or very high overlap density. According to population ecology theory, this suggests that high-density environments are competitive and ~~therefore~~ less conducive to online community growth than medium-density environments.

Surprisingly, this contrasts with our results in Study B, where we studied the diversity of ecological communities using vector autoregression models of group sizes over time to infer networks of ecological interactions. We find ecological communities that are mutualistic or competitive, that mix the two, or that have few significant ecological interactions at a ~~.~~ ~~Yet~~ overall however, ecological communities of subreddits are typically mutualistic and mutualistic interactions are stronger on average than competitive ones. Although we find evidence of density dependence, density-dependent competition does not necessarily reflect typical relationships in ecological communities of highly overlapping subreddits.

Our results in Study C show that the size of the other members of an ecological community improve time series forecasts of participation in online groups. Here, ~~Yet~~ average subreddit mutualism did not help predict growth ~~in our model comparison approach~~. This suggests that population ecology and community ecology offer complementary environmental and relational perspectives. Population ecology's focus on environmental factors such as niche and resource overlaps is useful for predicting growth, but ~~it does~~ not provide a way to study networks of mutualism and competition ~~because measures like overlap density are limited because of how it aggregates many relationships into a single quantity~~. Community ecology provides a way to unpack density ~~through direct analysis of ecological interactions in competition-mutualism networks~~ and ~~can thus~~ provide insights about the ~~different kinds of~~ specific relationships between groups. While Modeling these interactions helps forecast participation levels in groups, ~~but their~~ the existence of these interactions may be independent of future growth. For example, If mutualistic relationships are common in declining ecological communities, that would explain our result for H2.

The complementary nature of the two ecologies is seen in the coincidence of our findings in Study A and Study B. Indeed, these results can help explain the puzzling set of empirical results about the relationship between resource overlaps and outcomes like growth, decline and survival ~~in~~ eg [85, 91, 92]. Studies of density dependence theory in social computing measure the density of an online group's niche in terms of its overlaps in participants or topics. ~~However,~~ Our analysis clearly shows that resource overlaps between two groups might have little to do with whether they are mutualists or competitors. Instead, overlaps may simply reflect the hospitality of an environment to groups with overlapping topics or user-bases. As a result, differences in Wikis and Usenet groups in general might explain why user overlap was associated with the survival of Wikis [92] but with the decline of Usenet groups [85]. Wikia was a young and growing platform during Zhu et al.'s [92] data collection period ~~and they suggest that~~ when the growth of groups ~~was~~ may have been limited by knowledge of how to build a wiki. ~~and~~ This knowledge was provided by ~~more~~ overlapping experienced users. Usenet was in decline during Wang et al.'s [85] study period ~~and it may not have been limited in this way.~~ ~~Instead,~~ As we suggested in §2.3, users of groups with high overlap density ~~within a~~ in a Matter platform like Reddit or Usenet might have greater commitment to the platform than to any particular group. Such groups may face greater challenges in sustaining participation when the platform goes into decline.

The widespread mutualism found in Study B resonates with long-held understandings of ecological interactions in evolutionary theory [58]. Competition is unlikely to persist because it decreases survival ~~chances~~. Because mutualism increases survival ~~chances~~, it will be favored by natural selection [2, 4]. Similarly, competition can be avoided if groups adopt specialized roles in their ecological community, a dynamic known as resource partitioning in organizational ecology

[13, 64, 76]. Resource partitioning theory suggests that competition among real estate subreddits observed in Figure 3b may be due to a lack of specialization. If specialization does not emerge over time, such groups of competing subreddits may have decreased chances of surviving. By contrast, mental health support groups like those observed in Figure §3b appear to have distinctive purposes or roles. Future work to test such mechanisms in ecological communities of online groups may reveal ways that online groups complement or cooperate with each other.

Within large platforms for online groups, the great number of ecological communities that can be studied should make it possible for future work to apply methods from network science to construct and test generalizable theories about the roles of different types of resources, design features of platforms, and governance institutions in these ecological interactions. Future work should also incorporate community ecology analysis in ~~deeper~~ case studies of important topics such ecological communities engaged in peer production, political mobilization, ~~spreading~~ misinformation, or mental health support.

Although we focused on online groups within a single platform, groups may use multiple platforms with distinctive affordances for different purposes [30, 53]. Since the VAR method relies only on time series data to infer ecological interactions, it can be applied to study ecological communities spanning social media platforms. Community ecology can thus provide a bridge between quantitative studies of participation in online groups and theories of interconnected information ecologies [65]. While we focus on relationships between groups sharing a platform, one can apply our concepts and methods to understand how interdependent systems of technologies and users give rise to higher levels of social organization on social media platforms [1, 3].

6.1 Implications for Design

In the final chapter of their book on *Building Successful Online Communities*, Kraut et al. [57] advise managers of online groups to select an effective niche and beware of competition. But these recommendations are based on little direct evidence from studies of online groups and offer almost no concrete advice on what a designer or group should do differently based on either piece of advice. Although further research into ecological interactions is needed before design principles can be derived, we provide a framework for online group managers to think about ecological constraints on group size ~~in terms of a competition–mutualism network~~. While intuition suggests that online community managers might seek out mutualistic relationships and avoid competitive ones, it is often not obvious whether another group ~~a community~~ with overlapping users is a competitor or mutualist. Our method provides a way for group managers to know.

Competitors have a negative impact on growth, but ecological theory suggests that specialization is an adaptive strategy in response to competition [1, 13, 57, 70]. Using our method, group managers might identify competitors limiting the growth of their groups. With knowledge from this analysis in hand, they might be able to escape a competitive dynamic by specializing ~~in distinctive designs, policies, or governance practices~~.

While competitive relationships are defined by how they decrease the size of groups, competition can also be important to the health of the broader ecological community. Exit to an alternative group can be an avenue for political change in response to grievances and poor governance [34, 46]. The threat of competition with other groups may make expressions of voice more persuasive to moderators or platforms [46].

Groups looking to increase activity should desire to seek out mutualistic relationships and we believe that designers of online platforms can help them do so. Features such as meta-groups, group search, recommendation engines, and practices like linking related groups may lower barriers ~~boundaries~~ between groups and ~~thereby~~ support mutualis~~tic relationships~~ism. ~~However,~~ that said, it is not obvious to what extent particular features will support competition, mutualism, or both. Using our method, managers and designers can test features intended to support mutualism ~~or competition.~~

7 CONCLUSION

While ~~E~~explanations for the rise or decline of online groups often look to internal mechanisms, ~~but~~ understanding the role of interdependence between online groups is increasingly important. While prior research has investigated competition and mutualism among online groups with overlapping users and topics using ~~in~~ the population ecology framework [85, 91, 92], this approach ~~framework~~ does not provide a way to infer competitive or mutualistic interactions among related groups. We introduce the community ecology framework as a complementary perspective to ~~the~~ population ecology ~~introduced in prior research~~. By inferring competition-mutualism networks directly from time-series data, our community ecology approach helps resolve empirical tensions raised by prior ecological work in social computing and reveal that most interactions within clusters of subreddits with highly overlapping users are mutualistic. Our methods provide a foundation for future work investigating related online groups.

REFERENCES

[1] H.E. Aldrich and M. Ruef. 2006. *Organizations Evolving* (second ed.). SAGE Publications, Thousand Oaks, CA.

[2] Robert A. Armstrong and Richard McGehee. 1980. Competitive Exclusion. *The American Naturalist* 115, 2 (Feb. 1980), 151–170.

[3] W. Graham Astley. 1985. The Two Ecologies: Population and Community Perspectives on Organizational Evolution. *Administrative Science Quarterly* 30, 2 (1985), 224–241.

[4] R. Axelrod and W. D. Hamilton. 1981. The Evolution of Cooperation. *Science* 211, 4489 (March 1981), 1390–1396.

[5] Joel A. C. Baum and Andrew V. Shipilov. 2006. Ecological Approaches to Organizations. In *Sage Handbook for Organization Studies*. Sage, Rochester, NY, 55–110.

[6] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 2020), 830–839.

[7] Yochai Benkler. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven, CT.

[8] Yochai Benkler, Hal Roberts, Robert Faris, Alicia Solow-Niederman, and Bruce Etling. 2013. *Social Mobilization and the Networked Public Sphere: Mapping the SOPA-PIPA Debate*. SSRN Scholarly Paper ID 2295953. Social Science Research Network, Rochester, NY.

[9] Janet M Box-Steffensmeier. 2014. *Time Series Analysis for the Social Sciences*.

[10] Brian S. Butler. 2001. Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. *Information Systems Research* 12, 4 (2001), 346–362.

[11] Brian S. Butler, Patrick J. Bateman, Peter H. Gray, and E. Ilana Diamant. 2014. An Attraction-Selection-Attrition Theory of Online Community Size and Resilience. *MIS Q.* 38, 3 (Sept. 2014), 699–728.

[12] Fabio Canova. 2007. VAR Models. In *Methods for Applied Macroeconomic Research*. Princeton University Press, 111–164.

[13] Glenn R. Carroll. 1985. Concentration and Specialization: Dynamics of Niche Width in Populations of Organizations. *Amer. J. Sociology* 90, 6 (May 1985), 1262–1283.

[14] Glenn R. Carroll and Michael T. Hannan. 1989. Density Dependence in the Evolution of Populations of Newspaper Organizations. *American Sociological Review* 54, 4 (Aug. 1989), 524.

[15] Simone Cenci, George Sugihara, and Serguei Saavedra. 2019. Regularized S-Map for Inference and Forecasting with Noisy Ecological Time Series. *Methods in Ecology and Evolution* 10, 5 (2019), 650–660.

[16] Grégoire Certain, Frédéric Barraquand, and Anna Gårdmark. 2018. How Do MAR(1) Models Cope with Hidden Nonlinearities in Ecological Dynamics? *Methods in Ecology and Evolution* 9, 9 (Sept. 2018), 1975–1995.

[17] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can’t Stay Here: The Efficacy of Reddit’s 2015 Ban Examined through Hate Speech. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 31:1–31:22.

[18] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet’s Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (2018), 32:1–32:25.

[19] Munmun De Choudhury, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber. 2016. Social Media Participation in an Activist Movement for Racial Equality. In *Tenth International AAAI Conference on Web and Social Media*.

[20] Tiago Cunha, David Jurgens, Chenhao Tan, and Daniel Romero. 2019. Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities. In *The World Wide Web Conference (WWW ’19)*. Association for Computing Machinery, New York, NY, USA, 318–328.

[21] Laura Dabbish, Rosta Farzan, Robert Kraut, and Tom Postmes. 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)*. Association for Computing Machinery, New York, NY, USA, 245–248.

[22] Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 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*. ACM Press, Rio de Janeiro, Brazil, 307–318.

[23] Srayan Datta, Chanda Phelan, and Eytan Adar. 2017. Identifying Misaligned Inter-Group Links and Communities. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 37:1–37:23.

[24] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity. *Proceedings of the International AAAI Conference on Web and Social Media* 8, 1 (May 2014), 71–80.

[25] Marco Del Tredici and Raquel Fernández. 2018. Semantic Variation in Online Communities of Practice. *arXiv:1806.05847 [cs]* (June 2018). arXiv:1806.05847 [cs]

[26] Paul DiMaggio, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. 2001. Social Implications of the Internet. *Annual Review of Sociology* 27, 1 (Aug. 2001), 307–336.

[27] Stanislav D. Dobrev, Tai-Young Kim, and Michael T. Hannan. 2001. Dynamics of Niche Width and Resource Partitioning. *Amer. J. Sociology* 106, 5 (2001), 1299–1337.

[28] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. 2006. "Alone Together?": Exploring the Social Dynamics of Massively Multiplayer Online Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 407–416.

[29] Susan T. Dumais. 2004. Latent Semantic Analysis. *Annual Review of Information Science and Technology* 38, 1 (2004), 188–230.

[30] Casey Fiesler and Brianna Dym. 2020. Moving Across Lands: Online Platform Migration in Fandom Communities. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 042:1–042:25.

[31] Casey Fiesler, Jialun" Aaron" Jiang, Joshua McCann, Kyle Frye, and Jed R. Brubaker. 2018. Reddit Rules! Characterizing an Ecosystem of Governance.. In *Proceedings of the AAAI International Conference on Web and Social Media*. AAAI, Stanford, CA, 72–81.

[32] John H. Freeman and Pino G. Audia. 2006. Community Ecology and the Sociology of Organizations. *Annual Review of Sociology* 32 (2006), 145–169.

[33] Brendan J. Frey and Delbert Dueck. 2007. Clustering by Passing Messages Between Data Points. *Science* 315, 5814 (Feb. 2007), 972–976.

[34] Seth Frey and Robert W. Sumner. 2019. Emergence of Integrated Institutions in a Large Population of Self-Governing Communities. *PLOS ONE* 14, 7 (July 2019), e0216335.

[35] Janet Fulk, Andrew J. Flanagin, Michael E. Kalman, Peter R. Monge, and Timothy Ryan. 1996. Connective and Communal Public Goods in Interactive Communication Systems. *Communication Theory* 6, 1 (1996), 60–87.

[36] Tilmann Gneiting and Adrian E. Raftery. 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Amer. Statist. Assoc.* 102, 477 (March 2007), 359–378.

[37] Scott A. Hale. 2015. Cross-Language Wikipedia Editing of Okinawa, Japan. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 183–192.

[38] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 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 (May 2013), 664–688.

[39] William L. Hamilton, Justine Zhang, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. Loyalty in Online Communities. *arXiv:1703.03386 [cs]* (May 2017). arXiv:1703.03386 [cs]

[40] Michael T. Hannan and John Freeman. 1977. The Population Ecology of Organizations. *Amer. J. Sociology* 82, 5 (1977), 929–964.

[41] Michael T. Hannan and John Freeman. 1989. *Organizational Ecology* (first ed.). Harvard University Press, Cambridge, MA.

[42] Amos Henry Hawley. 1986. *Human Ecology: A Theoretical Essay*. University of Chicago Press, Chicago; London.

[43] Sarah E. Heaps. 2020. Enforcing Stationarity through the Prior in Vector Autoregressions. *arXiv:2004.09455 [stat]* (April 2020). arXiv:2004.09455 [stat]

[44] Jack Hessel, Chenhao Tan, and Lillian Lee. 2016. Science, Askscience, and Badscience: On the Coexistence of Highly Related Communities. In *Tenth International AAAI Conference on Web and Social Media*. 11. arXiv:1612.07487

[45] Benjamin Mako Hill and Aaron Shaw. 2019. Studying Populations of Online Communities. In *The Oxford Handbook of Networked Communication*, Brooke Foucault Welles and Sandra González-Bailón (Eds.). Oxford University Press, Oxford, UK, 173–193.

[46] Albert O. Hirschman. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press.

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

[48] Steven L. Johnson, Samer Faraj, and Srinivas Kudaravalli. 2014. Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment. *Management Information Systems Quarterly* 38, 3 (2014), 795–808.

[49] Alexander Jordan, Fabian Krüger, and Sebastian Lerch. 2019. Evaluating Probabilistic Forecasts with scoringRules. *Journal of Statistical Software* 90, 1 (Aug. 2019), 1–37.
[50] Sanjay Ram Kairam, Dan J. Wang, and Jure Leskovec. 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)*. ACM, New York, NY, USA, 673–682.
[51] Holger Kantz and Thomas Schreiber. 2003. *Nonlinear Time Series Analysis* (second ed.). Cambridge University Press, Cambridge.
[52] Michael L. Katz and Carl Shapiro. 1985. Network Externalities, Competition, and Compatibility. *The American Economic Review* 75, 3 (1985), 424–440.
[53] Charles Kiene, Jialun "Aaron" Jiang, and Benjamin Mako Hill. 2019. Technological Frames and User Innovation: Exploring Technological Change in Community Moderation Teams. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 44:1–44:23.
[54] Joon Koh, Young-Gul Kim, Brian Butler, and Gee-Woo Bock. 2007. Encouraging Participation in Virtual Communities. *Commun. ACM* 50, 2 (Feb. 2007), 68–73.
[55] P. M. Krafft and Joan Donovan. 2020. Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. *Political Communication* 37, 2 (March 2020), 194–214.
[56] Robert E. Kraut and Andrew T. Fiore. 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)*. ACM, Baltimore, Maryland, USA, 722–732.
[57] Robert E. Kraut, Paul Resnick, and Sara Kiesler. 2012. *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press, Cambridge, MA.
[58] Peter Kropotkin. 2012. *Mutual Aid: A Factor of Evolution*. Courier Corporation.
[59] Ida Kubiszewski, Joshua Farley, and Robert Costanza. 2010. The Production and Allocation of Information as a Good That Is Enhanced with Increased Use. *Ecological Economics* 69, 6 (April 2010), 1344–1354.
[60] Christopher D Manning, Prabhakar Raghavan, Hinrich Schütze, and Cambridge University Press. 2018. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge.
[61] Drew B. Margolin, Cuihua Shen, Seungyoon Lee, Matthew S. Weber, Janet Fulk, and Peter Monge. 2012. Normative Influences on Network Structure in the Evolution of the Children's Rights NGO Network, 1977-2004:. *Communication Research* (Oct. 2012).
[62] Leland McInnes, John Healy, and Steve Astels. 2017. Hdbscan: Hierarchical Density Based Clustering. *The Journal of Open Source Software* 2, 11 (March 2017), 205.
[63] J. Miller McPherson. 1983. An Ecology of Affiliation. *American Sociological Review* 48, 4 (1983), 519–532.
[64] Bruce A. Menge. 1972. Competition for Food between Two Intertidal Starfish Species and Its Effect on Body Size and Feeding. *Ecology* 53, 4 (July 1972), 635–644.
[65] Bonnie A. Nardi and Vicki L. O'Day. 1999. *Information Ecologies : Using Technology with Heart*. The MIT Press, Cambridge, Massachusetts.
[66] Mark Novak, Justin D. Yeakel, Andrew E. Noble, Daniel F. Doak, Mark Emmerson, James A. Estes, Ute Jacob, M. Timothy Tinker, and J. Timothy Wootton. 2016. Characterizing Species Interactions to Understand Press Perturbations: What Is the Community Matrix? *Annual Review of Ecology, Evolution, and Systematics* 47, 1 (2016), 409–432.
[67] Vincent Ostrom and Elinor Ostrom. 1977. Public Goods and Public Choices. In *Alternatives For Delivering Public Services: Toward Improved Performance*, Emanuel S. Savas (Ed.). Westview Press, Boulder, CO, 7–49.
[68] Bernhard Pfaff. 2008. VAR, SVAR and SVEC Models: Implementation Within R Package Vars. *Journal of Statistical Software* 27, 1 (July 2008), 1–32.
[69] Steven Piantadosi, David P Byar, and Sylvan B Green. 1988. The Ecological Fallacy. *American Journal of Epidemiology* 127 (1988), 893–904.
[70] Walter W. Powell, Douglas R. White, Kenneth W. Koput, and Jason Owen-Smith. 2005. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *Amer. J. Sociology* 110, 4 (Jan. 2005), 1132–1205.
[71] Paul Resnick, Joseph Konstan, Yan Chen, and Robert E Kraut. 2012. Starting New Online Communities. In *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press, Cambridge, MA, 231–280.
[72] W. S. Robinson. 1950. Ecological Correlations and the Behavior of Individuals. *American Sociological Review* 15, 3 (1950), 351–357.
[73] Paul M. Romer. 1990. Endogenous Technological Change. *Journal of Political Economy* 98, 5, Part 2 (Oct. 1990), S71–S102.
[74] Peter J. Rousseeuw. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* 20 (Nov. 1987), 53–65.

[75] Martin Ruef. 2000. The Emergence of Organizational Forms: A Community Ecology Approach. *Amer. J. Sociology* 106, 3 (Nov. 2000), 658–714.
[76] Thomas W. Schoener. 1974. Resource Partitioning in Ecological Communities. *Science* 185, 4145 (1974), 27–39.
[77] Christopher A. Sims. 1980. Macroeconomics and Reality. *Econometrica* 48, 1 (1980), 1–48.
[78] Jesper B. Sørensen. 2004. Recruitment-Based Competition between Industries: A Community Ecology. *Industrial and Corporate Change* 13, 1 (Feb. 2004), 149–170.
[79] Chenhao Tan. 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)*. AAAI, Palo Alto, California, 395–404.
[80] Chenhao Tan and Lillian Lee. 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)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1056–1066.
[81] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 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)*. ACM, New York, NY, 355:1–355:7.
[82] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2020. The Population Ecology of Online Collective Action.
[83] Herman A Verhoef and Peter J Morin. 2010. *Community Ecology: Processes, Models, and Applications*. Oxford University Press, Oxford.
[84] Nicholas Vincent, Isaac Johnson, and Brent Hecht. 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)*. ACM, New York, NY, 566:1–566:13.
[85] Xiaoqing Wang, Brian S. Butler, and Yuqing Ren. 2012. The Impact of Membership Overlap on Growth: An Ecological Competition View of Online Groups. *Organization Science* 24, 2 (June 2012), 414–431.
[86] Yiran Wang, Melissa Niiya, Gloria Mark, Stephanie M. Reich, and Mark Warschauer. 2015. Coming of Age (Digitally): An Ecological View of Social Media Use among College Students. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 571–582.
[87] Steven Weber. 2000. The Political Economy of Open Source Software. (June 2000).
[88] Justine Zhang, William L. Hamilton, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. Community Identity and User Engagement in a Multi-Community Landscape. *Proceedings of the International AAAI Conference on Weblogs and Social Media. International AAAI Conference on Weblogs and Social Media* 2017 (May 2017), 377–386.
[89] Jason Shuo Zhang, Brian Keegan, Qin Lv, and Chenhao Tan. 2021. Understanding the Diverging User Trajectories in Highly-Related Online Communities During the Covid-19 Pandemic. *Proceedings of the International AAAI Conference on Web and Social Media* 5 (2021), 12. arXiv:2006.04816
[90] Xiaoquan (Michael) Zhang and Feng Zhu. 2011. Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review* 101, 4 (June 2011), 1601–1615.
[91] Haiyi Zhu, Jilin Chen, Tara Matthews, Aditya Pal, Hernan Badenes, and Robert E. Kraut. 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)*. ACM, New York, NY, USA, 301–310.
[92] Haiyi Zhu, Robert E. Kraut, and Aniket Kittur. 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)*. Association for Computing Machinery, New York, NY, USA, 281–290.