A Community Ecology Approach for Identifying Competitive and Mutualistic Relationships Between Online Communities

other members of
size reputation of
has
measures of group health

ANONYMOUS AUTHOR(S)

Online groups affect each other's ~~sizes~~ lives as people, content, and ideas flow between groups. How can researchers and designers understand the relationship between these inter-group interactions and ~~group sizes~~ over time? Inspired by population ecology, prior HCI research correlated group size with overlap in content and membership between all other groups. The resulting body of evidence is puzzling and suggests that overlap will sometimes help and sometimes hurt. We suggest that this confusion results from aggregating inter-group relationships into an overall environmental effect while ignoring the network of varying inter-group relationships. To capture these dynamics, we propose a theoretical framework based on community ecology and a method for inferring competitive and mutualistic relationships from time-series participation data. We compare population and community ecology analyses of online community growth by analyzing subreddit clusters with high user overlap but varying degrees of competitive and mutualistic network structures.

ACM Reference Format:
Anonymous Author(s). 2018. A Community Ecology Approach for Identifying Competitive and Mutualistic Relationships Between Online Communities. 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, a growing body of HCI scholars have shown how online groups, such as wikis, discussion forums, and mailing lists spawn new groups and wage conflicts against, compete with, and help each other [17, 60, 67, 72]. This emerging body of work has demonstrated that a full understanding of online groups involves understanding how they relate to one another. Prior investigations of interdependence between online groups draw from ecology

[1]We use the term "online group" instead of "online community" to help avoid confusion with our analytical term "community ecology."

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

Although designed to study biological systems,

[67, 71, 72]. Ecology has been fruitfully applied to human organizations ranging from commercial industries to social movements for over 40 years [4, 29]. However, there remain very few explicitly ecological studies of online groups and those that exist yield a puzzling body of results. This research has shown that, for example wikis whose memberships overlap with other wikis survived longer [71], but Usenet groups with overlapping memberships failed more quickly [67]. Groups may also overlap with respect to topic, and in some settings, the relationship between topical overlaps and growth or survival follows theoretical predictions that moderate levels of overlap should lead to the best outcomes [71], while studies in other settings find weak or theoretically inconsistent relationships [60, 63].

We argue that a path to increased clarity involves recognizing that there are two distinct strands of ecological theory ~~that may be applicable to social computing~~ but that these strands—called *population ecology* and *community ecology*—have different levels of analysis and make distinct theoretical predictions [2]. Population ecology focuses on how environmental factors shape the dynamics within a population, ~~such as~~ originally a biological species. The earliest and most influential works of ecology as applied to human organizations used the population ecology approach to answer questions such as those about how organizational forms become established or decline [1]. This approach requires treating a population as consisting of entities with similar resource needs while ignoring the distinct roles that individual groups play. Population ecology has framed every ecological analysis of online groups published in social computing venues that we are aware of.

Community ecology is a relational approach that analyzes *ecological communities* comprised of heterogeneous and interdependent groups [2]. In biology, this might be different populations of organisms inhabiting a lake or valley. In organization science, this might be the network of technology developers, manufacturers, and suppliers in the semiconductor industry [54, 65]. Community ecology is also of considerable influence in organization science [e.g. 2, 3, 45, 47, 54, 57], but to our knowledge, community ecology has never been applied in ~~HCI or~~ social computing research.

~~However,~~ while ecological studies of online groups like those by Wang et al. [67], Zhu et al. [71], and Zhu et al. [72] are consistently framed in terms of *population ecology* theory, their empirical analyses all deviate from classical population ecological analysis in important ways. Ecological studies in social computing have focused on overlapping resources like topics and participants. However, theories of population ecology were developed under the assumption of homogenous populations and are thus not designed to explain these phenomena. The ultimate result, we believe, is the confusion that ecological studies in social computing are mired in today.

Population ecology analyses of resource overlaps in HCI are based on the degree to which a group~~s~~'s participants or topics overlap (with every other group) is correlated with that groups growth or survival. Thus, they treat competition and mutualism as environmental forces that emerge from relationships between groups. Yet, through aggregation, the relationships themselves are obscured and no inference is made about whether any specific pair of groups ~~two communities~~ are competitors or mutualists. Our intervention is to directly infer ecological relationships in networks of related communities through the framework of community ecology. In doing so we make both theoretical and methodological contributions to social computing scholarship.

right?

First, in §3, we contribute to theory by providing a framework that supports a conceptual shift from treating individual online groups as indistinct parts of an ecological environment to treating them as belonging to a ecological community of related entities. Where prior approaches aggregate individual relationships between groups, our approach makes it possible to answer critical questions like "Which are a given online group's strongest mutualists or competitors?" Our second contribution, in §4.3 is to introduce and apply vector autoregression (VAR) models of community size over time for inferring networks of competitive and mutualistic relationships from online group participation activity data [11, 36, 58]. This approach has been widely used in biological ecology to make inferences about competitive or mutualistic relationships between species in shared environments. We apply it to 1996 3 networks of online communities with overlapping participants hosted on the platform Reddit and reveal that mutualistic relationships are more common than competitive ones. We show in §5.3.3 that including ecological relationships in ~~this~~ our VAR model s of community size might improves their predictive forecasting performance. Third, we present a typology of ecological networks and illustrate it using four case studies in §5.2. Fourth, we briefly test the central population ecology theory of density dependence in order to show how findings from population ecology about environmental competition or mutualism ~~may not~~ can fail to reflect typical relationships between highly overlapping communities. While this result suggests that ~~the~~ denser environments are more competitive, mutualistic relationships are more common than competitive ones in our community ecology analysis of Reddit. ~~Discussing this seemingly contrasting finding motivates future investigations into how competitive or mutualistic ecological~~

~~communities form and why some environments for online groups are competitive or mutualistic.~~

2 BACKGROUND

Our contributions build on a body of existing theoretical work that theorizes online groups as depending on several distinct types of resources as well as a smaller body of ecological research in social computing conducted over the last ~~1~~ decade. We briefly summarize both before introducing our conceptual approach.

2.1 Online groups depend on resources

Online groups are important social structures built to serve a wide range of goals [7, 19, 43]. While many online groups are formed, only an extreme minority develop a sizable group of participants [37, 42]. Understanding why some groups grow large while others do not is important because an online group's ability to achieve its goals depends on attracting and retaining contributors [43]. Ecological approaches to explaining the growth of online groups build on a theoretical approach that treats resources contributed by participants (such as content, information, attention, or social interactions) as playing a role in motivating participation by subsequent participants [9, 67, 70]. Although this dynamic implies a positive feedback loop, groups clearly do not grow forever and increasing costs of participation will eventually limit further growth [9]. While factors such as leadership, organizational practices, network structure, and design decisions can lower costs and increase benefits of participation [9, 42, 64], tactics intended to improve quality appear to limit growth. For example, the end of English Wikipedia's growth in early 2007 can be explained by barriers to newcomer participation erected during its growth phase [27] ~~in a pattern observed in other collaborative knowledge production projects [62]~~.

Of course, the nature of resources is clearly important to resource-based theories. Online group research has been particularly concerned with *rival* resources that become unavailable for use by others when used by any one group [7, 44, 50, 56]. Interdependence between online groups can be important to explaining outcomes like growth, decline and survival because important rival resources like the time, attention, and efforts of participants are ~~are~~ all subject to competition. Although opportunity costs of participation mean that people will not participate in a group if they prefer alternatives like sleep, entertainment, or work [6, 10], participation in a related online group providing the same benefits at lesser costs might be a compelling alternative. In this way, declines in online participation can sometimes be explained in terms of competition over rival resources like participant's time [67].

On the other hand, not all resources important to online groups will be subject to competition. Some groups produce connective and communal public goods like opportunities to communicate and information [25] which can be non-rival or even "anti-rival" when their usefulness increases as a result of others using them [44, 68]. For example, the usefulness of a communication network increases as more people join it [25, 40]. Similarly, the usefulness of an information good can increase as more people come to know and depend upon it [44, 68]. For example, awareness of an audience can motivate participation in online groups [70]. If multiple online groups help build the same connective or communal public goods, they may form mutualistic relationships with one another such that contributions to one group may "spill over" and motivate participation in related groups [72].

Even though online groups can share and compete over resources, studying interdependence between online groups adds enormous complexity to research projects [34]. Despite the challenges, a growing body of empirical research in social computing has sought to quantify how online groups share users or topics [17, 18, 32, 61] and how interactions between groups relate to outcomes like the emergence of new groups [60], contributions to peer produced knowledge [66], and the spread of hate speech [15]. However, few general theories or theoretical frameworks exist with which to understand how forms of interdependence between online groups shape their growth or survival.

2.2 Ecological research in social computing

Our theoretical approach draws from ecology. While our work focuses on the ecological study of online groups, other social computing scholars like Nardi and O'Day [48] have used the term "ecology" (and related concepts like "ecoystem" and "environment") as metaphors denoting assemblages of sites, technologies, or platforms. We use the term in a more narrow sense to refer 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 ~~interrelatedness~~ relationships between firms, organizational ecology was inspired by, and has drawn closely from, ecological studies in biology [29].

Because online groups bear many similarities to traditional organizations, organizational ecology provides a compelling theoretical framework for understanding interdependence between online groups and has been adopted by at least three empirical studies of how ~~the~~ resources shared ~~between online groups~~ shape its ~~the~~ growth, decline, or survival ~~of~~ online groups [67, 71, 72]. All three take up propositions of density dependence theory which sees competitive or mutualistic forces in a population of groups as a function of density. In the earliest and most influential strands of organizational ecology, density is simply the number of members in a population [1].

Density dependence theory proposes a trade-off between positive and negative consequences of density such that low levels of density are associated with resource-scarce environments unable to support a large population while high levels of density lead to competition. Low levels of density reflect limited opportunities for mutualistic relationships that contribute to shared non-rival resources like legitimacy, connectivity, and ~~or~~ knowledge. When density is high, it becomes difficult to avoid competitive relationships.

Models of density dependence theory in the population ecology of ~~or~~ organizations treat growth trajectories of populations with a logic akin to the models of resource dependence described in §2.1. Just as online group growth is thought to be driven by the attraction of new members to the contributions of prior participants, so growth in an organizational population is driven by positive feedback as successful organizations legitimate a business model causing new organizations to enter the market [13, 30]. In an analogous way, a population of online groups may grow as their platform gains in popularity, as existing groups spin off new ones, and as useful knowledge develops that can be shared between groups [60, 72]. An individual online group's growth may be limited by the ability of their social structures to scale to include more members [9] or due to competition with other groups over members [30]. Similarly, a population of online groups may decline when its members ~~if they~~ compete with one another over participants.

Organizational ecologists often ~~predict—and~~ find ~~—~~that the relationship between density and positive outcomes like growth or survival is inverse-U-shaped ~~(~~∩-shaped~~)~~ [4, 13]. This is because many environments present a trade-off between mutualism and competition; ~~in which~~ mutualistic forces are stronger when density is low and competitive forces are stronger when density is higher. The intuition is that Low-density environments reflect poor environmental conditions for success—if conditions were good then they would attract more growing communities and ~~hence~~ be more dense. On the other hand, high-density environments are thought to become crowded and competitive [30].

While ~~the~~ foundational studies of density dependence in organizational research measured ~~study~~ density and growth at the population level, ecological studies of online groups model a different notion of density dependence based on the concept of *overlap density*.[2] Rather than the number of groups that exist in a population, overlap density measures the extent to which an individual group's members or topics overlap with all other groups. Overlap density is not a property of a population of groups, but a property of the resource environment particular a group faces.

In contrast to biological species, this approach does not assume a population is homogeneous. Instead individuals have different *niches* corresponding to their

[2]Although it is less common in organizational research, overlap density has also been used by some organizational ecologists [e.g. 20].

resource needs [1]. This makes sense for online groups sharing a platform with diverse topics [39], norms [16, 21], and user bases [61]. Yet results from studies of overlap density in populations of online groups are inconsistent both with each other and with theoretical predictions from density dependence theory. ~~For example,~~ In Wang et al. [67], user overlaps in Usenet newsgroups are associated with decreasing numbers of participants. Similarly, TeBlunthuis et al. [63] find that topical overlaps between online petitions are negatively associated with participation. By contrast Zhu et al. [72] find that membership overlap is positively associated with increasing survival of new Wikia wikis. Only Zhu et al. [71] find support for the ∩-shaped relationship predicted by density dependence theory.

The classical logic of density dependence theory appears reasonable when applied to online groups on a platform like Reddit (see ~~similar~~ arguments in Zhu et al. [71] and [72]). Tradeoffs between commitment to a subgroup and commitment to the broader platform provides a plausible mechanism for density dependence. Kraut et al. [x] claim that people's commitment to subgroups substitutes ~~complements~~ ? ~~their~~ for commitment to ~~the whole~~ a platform, so subreddits with low user overlap density may have participants with low commitment to Reddit. On the other hand, a subreddit with high user overlap density may receive little commitment from its participants who seem more involved in Reddit overall than in any particular community ~~and the different communities may compete over their time.~~ If both are true, Intermediate levels of density ~~will be the~~ could reflect a "sweet spot" where participants ~~have particular commitment~~ then is maximized ~~to the subgroup while their participation in other communities increases their~~

~~commitment to the platform~~. We test the following hypothesis ~~both to lend evidence to~~ from density dependence theory both because it allows us to capture an approach to processual in social computing and to contrast the environmental approach of population ecology to community ecology's relational lens:

H1: The relationship between user overlap density and growth is inverse-U-~~shaped~~ (∩-shaped). no bold

3 A COMMUNITY ECOLOGY APPROACH FOR SOCIAL COMPUTING

Density dependence theory sees competition ~~or~~ and mutualism as an environmental property of an online group's niche. In community ecology, by contrast, competition and mutualism are properties of relationships between communities. Community ecology focuses on studying ecological communities of online groups related to each another in networks of competitive or mutualistic relationships. Doing so makes visible the distinctive roles that particular groups play in their ecological ~~communities~~ contexts. While varying conceptions of community ecology are found in the literature on organizational ecology [22], our approach follows that of Aldrich and Ruef [1] and Hawley [31].

Overlap density has been useful for advancing ecological theory and the empirical analysis of interrelated online groups because it provides a way to characterize the environment that an online group faces. As noted in §2.1, empirical studies of online groups find that higher levels of overlap density are associated with decreasing group sizes in some contexts but increasing growth or survival in others.

We propose that results can be explained by disaggregating density and taking a closer look at commensal relationships between communities. Population ecologists draw conclusions about the environments groups face, find that increasing

Table 1. The five possible commensal relationships between two online groups. Values in the column "A ⟶ B" represent the sign of hypothetical group A's effect on group B. Based on table 11.1 from Aldrich and Ruef [1].

A ⟶ B ($\Phi_{A,B}$)	B ⟶ A ($\Phi_{B,A}$)	Relationship type
+	+	Full mutualism
+	·	Partial mutualism
+	−	Predation
−	·	Partial competition
−	−	Full competition
·	·	Neutrality

Counterintuitively, this dynamic

density decreases growth, and ~~from this~~ conclude that the environment is competitive. ~~However, this~~ does not imply that typical relationships between groups sharing many users are ~~likely to be~~ competitive. ~~Drawing conclusions about relationships between individual-level variables from relationships between their aggregates is known as~~ an example of the "ecological fallacy" [53, 55]. The term ecological fallacy does not refer to theories of population or community ecology, but rather to "ecological correlations," ~~meaning~~ where correlations involving aggregates. ~~Overlap density aggregates a group's many relationships into a single property of the group's environment and correlates this with individual-level growth or survival. But~~ finding the density of an individual group's environment does not tell us which overlapping communities are competitors and which are mutualists.

footnote

?

While population ecology analyzes populations, community ecology aims to understand how different groups shape each other's growth, survival, and evolution [1, 2]. In organizational science, this can mean studying relationships between different organizational populations [e.g. 46, 59], or networks of relationships between organizations [e.g 45, 54]. The community ecology approach for social computing that we propose, in an analogous way, seeks to theorize the relationships between different online groups. It follows community ecology in biology and organization science by focusing on networks of commensal relationships [1]. A commensal relationship is a way groups affect one another through changes in group size. Commensal relationships can be mutualistic ~~if~~ when one group has a positive influence on the second such that growth (decline) in the first group leads to growth (decline) in the second. They can also be competitive if one group has a negative effect on the second such that growth (decline) in the first group leads to decline (growth) in the second.

There are six possible commensal relationships as described in Table 1. Note that commensal relationships can be reciprocal (as in full mutualism and competition) or not (as in partial mutualism and competition). In our framework "predation" (also called parasitism) refers to cases where a relationship is positive in one direction but negative in the other. It is also ~~of course~~ possible that growth or decline in the first group has no effect on the second group, and visa-versa, a situation termed "neutrality." Within an ecological community, the network of commensal relationships can be quantified using a *community matrix* the entries of which correspond to (commensal) ties in the network.

A potential reason for inconsistent findings in prior work is that the overall amount of user and topic overlap for a particular group is a poor proxy for the degree to which its size is limited by a competitive or mutualistic environment. It may simply be that user overlaps indicate that two groups are related. Perhaps most relationships were mutualistic in Zhu et al.'s [72] context but competitive but in Wang et al.'s [67]. Also, differences in communication modalities between discussion groups and wikis may be associated with different resource needs and thus different potential relationships between ~~their measure of~~ user overlap and growth or survival. ~~Overlap density takes an indirect approach to understanding competition or mutualism between online groups. Critically, it does not infer whether or not two groups are competitors or mutualists.~~ The community ecology approach we propose opens ~~such~~ relationships between online groups up to investigation. We explore the types of ecological communities that can be found in our research question:

RQ1: What types of ecological communities can be found? no bold

Moreover, ~~Similarly~~, overlap density does not account for how factors other than topic or user overlaps can lead to competitive and mutualistic forces which might help explain ~~these~~ the inconsistent findings. Zhu et al.'s [72] Wikia was a growing platform and they found increased survival among new communities with overlapping members ~~from established groups~~. Perhaps ~~the~~ because growth of ~~Wikia wikis~~ groups during ~~that study~~'s data collection period was limited by knowledge of how to build a wiki and this knowledge was provided by more experienced users. Usenet was in decline during Wang et al.'s [67] study period and it may not have been limited in this way. In general, competition over overlapping resources will have no effect on group growth if growth is limited by something other than the resource subject to competition. Ecologists of biological organisms understand that different populations might consume the same resource (e.g. eating the same food, sharing the same habitat), but not compete if the resource is abundant [65]. Further, resources that limit growth may be unobserved. P Our community ecology approach begins by relaxing the assumption that competition and mutualism are driven by user overlap density. We argue for the importance of this conceptual shift by testing two hypotheses, the first ~~of which~~ tests whether commensal relationships provide additional information compared to overlap density for explaining variation in the growth of online groups:

H2: Variation in growth explainable by commensal relationships is not reducible to user overlap density.

~~While explaining additional variation in a regression model will support conclusions that commensal relations are not reducible to user overlaps, this~~ Support for H2 may be a relatively low bar for assessing whether commensal relationships are important factors shaping online community growth. Confounding moderators or mediators related to commensal relationships might be more important than the relationships themselves. Therefore, we also test a hypothesis that modeling commensal relationships is useful for making predictive forecasts of online community size.

H3: The addition of commensal relationships improves the predictive forecasting performance of a time series model.

4 DATA, MEASURES, AND METHODS

We analyze the publicly available Pushshift archive of Reddit submissions and comments which we obtained from December 5th 2005 to April 13th 2020 [5], including the top 10,000 subreddits by number of comments. We exclud~~ing~~e subreddits where a majority of submissions are marked “NSFW.” The top 10,000 subreddits include smaller communities and provide a sufficiently large number of ecological communities for our statistical analysis.

4.1 Group size and growth

Group size, measured as the number of distinct commenting users in a subreddit, is the dependent variable ~~of~~ for our vector autoregression models~~, which we~~ use d to infer commensal relationships ~~and~~ as describe ~~below~~ in §4.4. Group size quantifies the number of people who participated in the subreddit during a week~~, and is thus a reasonable measure of group size.~~ fix Typical of social media participation data, this variable is highly skewed and therefore we transform it by adding 1 and taking the natural logarithm. Log-transformation is also common in applications of vector autoregression in biological ecology [14, 36].

Subreddit *growth* is the dependent variable in our models testing **H1** and **H2** No bold and is measured as the change in the (log-transformed) size of a subreddit over the final 24 weeks of our data, from to November 4th 2019 to April 13th 2020.

plural?

4.2 User overlap and density

We measure user overlap~~s~~ between subreddits to construct clusters of related groups for our analysis of ecological communities in **RQ1** and in our test of **H1**. ~~While~~ Zhu et al. [72] and Wang et al. [67] measure overlaps between pairs of communities in terms of users who contribute to both communities at least once and ~~(~~excluding users who appear in more than 10 communities~~)~~. W~~w~~e found that this measure had poor face validity for subreddit similarity and our clustering algorithm converged to a solution placing over 80% of subreddits in a single cluster ~~(~~even without the exclusion condition~~)~~. Issues with this measure may result from how Reddit users often peripherally participate in many communities, while participating heavily in few [28, 61, 69]. Therefore, we adopt a measure of user overlap based on the amount each user participates in two communities.

To measures user overlap between subreddits, we first count the number of times each user comments in each subreddit. Then, we normalize comment-counts within each subreddit by the maximum number of comments left by a single author to prevent giving undue weight to subreddits with higher overall activity levels. This step yields our intermediate measure of *user frequency*.

$$\text{user frequency}_{\text{user,subreddit}} = F_{u,j} = \frac{N_{\text{u}}}{max_{u \in \text{j}} N_u} \tag{1}$$

This measure of user frequency ~~follows~~ is drawn from Datta et al. [17]'s study of subreddits with overlapping users and topics. Unlike Datta et al. [17], we do not divide *user frequency* by the number of subreddits where the user appears because we do not wish to assume that users who comment in many subreddits are less ecologically important.

We obtain our measure of *user overlap* by taking the cosine similarities between the user frequencies of each pair of subreddits.

$$\text{user overlap}_{i,j} = O_{i,j} = \frac{\sum_{\text{u} \in i \cup j} F_{u,j} F_{u,i}}{\|i\| \, \|j\|} \tag{2}$$

Where $\|i\| = \sqrt{\sum_{u \in i} F_{u,i}^2}$ ~~is~~ as the euclidean norm of i.

We measure *user overlap density* in order to test **H1**, that subreddit growth has a ∩-shaped dependence on density and to test **H2**, which compares density dependent and commensal models of growth. We obtain measures of user overlap density for a group by taking the average of its user overlaps over all M of the other the groups in our dataset.

$$\text{user overlap density}_i = D_i = \frac{1}{M} \sum_{j=1;j!=i}^{M} O_{i,j} \tag{3}$$

4.3 Identifying ecological communities

In order to test H2 and answer RQ1, we ~~infer~~ estimate the community matrix of commensal relationships between selected communities of online groups. The community matrix is a central analytical object in community ecology in both biological and organizational ecology [1, 49, 65]. Organizational scientists have rarely attempted to estimate the full community matrix in the analysis of large ecological communities due to statistical and data limitations [e.g. 57, 59]. For instance, 100,000,000 possible commensal relationships exist within our set of 10,000 communities. ~~and~~ Attempting to infer them all raises considerable computational and statistical challenges. Therefore, we use a heuristic approach to find ecological communities of online groups that all have high user overlap and that prior work suggests are most likely to ~~have~~ exhibit commensal ~~relationships~~ ism.

To identify ecological communities between related subreddits, we use affinity propagation clustering on ~~our~~ the user overlap measure described above [23]. Affinity propagation is appropriate for clustering given a measure of similarity between data points. In developing our analysis, we experimented with other clustering techniques, including spectral clustering and k-means. We found that affinity clustering provided greater face validity which, along with the distribution of cluster sizes and the number of found clusters, was robust to the choice of parameters compared to alternative algorithms. We use the implementation of affinity propagation clustering provided in `scikit-learn` with the *damping* parameter set to 0.85 (to improve convergence speed) and the default *preference* parameter ~~which is~~ median of the input similarities [51]. We find 1208 clusters and 788 isolated subreddits ~~and we proceed with the subreddits which are assigned to clusters~~. The median cluster has 4 subreddits and the largest cluster has 156.

4.4 Vector Autoregression Models

To infer ecological commensal relationships within the clusters, we use vector autoregression (VAR) models, an established approach in biological ecology [36]. 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 Gompertz models of population growth, which are themselves similar to the Lotka-Volterra self-limiting growth models used in Butler's [9] resource dependence framework from social computing [36]. Even in the presence of unmodeled nonlinearities, VAR(1) models can reliably identify competition or mutualism ~~between species~~ in empirically realistic scenarios [14]. They have also been widely adopted in the social sciences, particularly in political science and in macroeconomics [8]. VAR models are flexible enough to model a wide range of systems so long as sufficiently long time-series data are available [58]. They can be intuitively understood as a generalization of a one-dimensional auto-regressive (AR) models in time series analysis. But while AR models predict the state of a single time-series variable as a function of its previous values, VAR models simultaneously predict multiple time series variables as a function of the values of every other variable in the system [11, 36]. We include a vector of intercept terms (to account for different equilibrium community sizes) and a vector of trends (to account for long-run endogenous growth) because we found that including these terms greatly improved the fit of our models to the data. Our VAR(1) models have this form in vector notation:

$$Y_t = B_0 + B_1 t + \sum_{k=1}^{K} A_k X_{t,k} + \sum_{j=1}^{M} \Phi_j Y_{t-1,j} + \epsilon_t \quad (4)$$

different approach

Where Y_t is a vector containing the sizes of M online groups at time t. B_0 is the vector of intercept terms and B_1 is the linear time trend for each community. Φ_j represents the influence of 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 to inter-group influences. Positive coefficients in the community matrix represent mutualism while negative coefficients represent competition.

Additional time-dependent predictors can be included in the vectors $X_{t,k}$ with coefficients A_k, and ϵ_t is the vector of error terms. Because subreddits are created at different times, growth trends must begin only after the subreddit is created. We use $X_{k,t}$ to remove the trends during the period prior to the creation of subreddits. For community j created at time t_r we fill $X_{j,t}$ with the sequence $[1, 2, 3, \ldots, r-1, 0, 0, 0, \ldots]$. In other words, $X_{j,t}$ adds a trend only during the period prior to the first comment in subreddit j. This effectively sets A_j approximately equal to $-B_j$. ?

This isn't reused? is it?

For each cluster found in §4.3, we fit VAR models using ordinary least squares as implemented in the `vars` R package to predict the number of *distinct commenting users* each week using data on the entire history of each subreddit [52].

is this the same as before?

4.5 Forecasting with VAR models

In our test of **H3**, we test whether modeling commensal relationships improves overall performance in forecasting future participation in online groups by comparing the model in Equation 4 to a baseline model with the off-diagonal elements of Φ fixed to 0. ~~Therefore, the~~ This baseline model is equivalent to our VAR model, but does not account for commensal relationships. 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 281 subreddits with less than 156 weeks of activity from the VAR analysis.

We test **H3** using 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 has important limitations for aggregating errors from forecasts on different scales. It does not take the scale of the predicted variable or forecast uncertainty into account. Thus it may place excessive weight on forecasts of larger subreddits, where errors may have greater magnitude simply because the absolute magnitite of the variance is greater. The CRPS is a proper scoring rule for evaluating probabilistic forecasts [26]. 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. Our CRPS calculations assume that the predictive forecast

is this a tech term?

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 [38].

4.6 Characterizing ecological communities

The independent variable ~~our~~ for test of **H2** measures the *average subreddit commensalism*, or the average influence of other subreddits in the ecological community on a given subreddit j, which we calculate by taking the mean of off-diagonal elements of row j of the community matrix *Phi*. We use the mean instead of the sum because different ecologial communities have different numbers of subreddits.

$$\text{average subreddit commensalism} = AC_j = \frac{1}{M-1}\sum_{i=1,i!=j}^{M} \Phi_{i,j} \tag{5}$$

To explore the types of ecological communities we find on Reddit in **RQ1**, we additionally construct two measures describing (1) the degree to which commensal relationships in the cluster are mutualistic or competitive and (2) the overall strength of commensal interactions. *Average commensalism* (AC) measures the extent to which an overall ecological community is mutualistic or competitive, ~~we~~ by taking the mean point estimate of off-diagonal coefficients of Φ:

$$\text{average commensalism} = AC = \frac{1}{M}\sum_{i=1}^{M} C_i \tag{6}$$

Finally, *Average absolute commensalism* quantifies the overall strength of commensal relationships in an ecological community as the mean point estimate of off-diagonal coefficients of Φ.

$$\text{average absolute commensalism} = AAC = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1;j!=i}^{M}\left|\Phi_{i,j}\right| \tag{7}$$

4.7 Networks of commensalism

In our exploration of **RQ1**, we examine the Φ parameter of the V~~ar~~ AR(1) model which encodes the "community matrix" and can be interpreted as a network of commensal relationships [36]. While the coefficients of Φ correspond to direct associations between group sizes [49], commensal relationships can also be indirect (e.g., ~~like~~ if a predator kills an herbivore and thereby helps fauna consumed by the herbivore). Consider relationships between three groups (A, B, C) such that A partially competes with B and B partially competes with C but A and C have no direct relationship. A VAR(1) model inferring these relationships will have negative coefficients for $\phi_{A,B}$ and $\phi_{B,C}$ but $\phi_{A,C}$ will be nearly zero.

However, this does not 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. Fortunately, such indirect relationships can be analyzed using impulse response functions (IRFs) to interpret a VAR model [8]. In large VAR models containing many groups, the great number of parameters can mean that few specific elements of Φ will be statistically significant, even as many weak relationships can combine into statistically significant IRFs [11]. We present networks of commensal relationships between subreddits where the 95% confidence interval of the IRF does not include 0. We bootstrap these confidence intervals with 1000 samples.

4.8 Density dependence and commensalism

To

~~In our~~ test ~~of~~ **H1** we fit **Model 1** in Equation 8, the overlap-density dependence model with first and second-order terms for overlap density to allow for a curvilinear relationship between *overlap density* and *growth*. We ~~then~~ test **H2**, which hypotheses that competition and mutualism between online groups is not reducible to resource overlaps ~~because resources like participants and information on which online groups depend are not strongly rival,~~ using liklihood ratio tests to compare ~~**Model 1** and~~ **Model 2** in Equation 9 ~~which~~ includes *average subreddit commensalism* (AC_i) as a predictor, and **Model 3** in Equation 10, which includes both sets of predictors.

rewrite models

How do we test H3

$$\textbf{Model 1} \quad \text{group growth}_i = B_0 + B_1 \log(D_i + 1) + B_2 \log^2(D_i + 1) \tag{8}$$

$$\textbf{Model 2} \quad \text{group growth}_i = B_0 + B_3 AC_i \tag{9}$$

$$\textbf{Model 3} \quad \text{group growth}_i = B_0 + B_1 \log(D_i + 1) + B_2 \log^2(D_i + 1) + B_3 AC_i \tag{10}$$

5 RESULTS

First

by constructing

We answer **RQ1**~~, which asks what kinds of ecological communities can be found through~~ a typology of ecological communities along ~~in terms of~~ two dimensions: (1) the degree to which a community is mutualistic or competitive, and (2) the overall strength of commensal interactions between the communities member groups. In §5.1 we apply ~~develop~~ this typology to our sample of 10,000 subreddits and find that ~~typical~~ ecological communities are usually mutualistic, and ~~and~~ in §5.2 we illustrate the typology using 4 example ecological communities. Then, in §5.3 test **H1** and **H2** and find evidence of a curvilinear, ∩-shaped

ours feels off from predictions

relationship between overlap density and growth~~, and that~~ We also show ~~how~~ that adding average subreddit commensalism to the regression improves model fit. Finally, in §5.3.3 we report on **H3** and show that including commensal terms improves forecasting performance in our time series models.

5.1 Characterizing ecological communities

Ecological communities of subreddits with overlapping users vary in both the overall strength of commensalism and in overall degree of mutualism and competition between member groups. If an ecological ~~communities'~~ community's average commensalism is positive, ~~then~~ the community is mutualistic, ~~but~~ but if its average commensalism is negative, it is competitive. However, Average commensalism can be close to 0 ~~within an ecological communities of subreddits~~ in two ways. First, significant competitive and mutualistic relationships ~~are~~ might be both present ~~and they~~ but cancel one another out ~~when aggregated~~. Such an ecological community will have high average absolute commensalism. ~~Second~~ Alternatively, there may be little-to-no commensalism between members of the ecological community. ~~and~~ In this case, average absolute commensalism will be low. In this way, The role ~~of~~ that average absolute commensalism plays ~~in~~ our framework is ~~thus~~ to characterize the overall magnitude of ecological relationships within the ecological community in order to ~~tease apart~~ separate communities with a mixture of competitive and mutualistic relationships from those where commensal relationships are relatively weak.

Figure 1 visualizes the distribution of average commensalism and average absolute commensalism over the N ecological communities in our sample. We observe ecological communities characterized by strong forms of both mutualism and competition, others

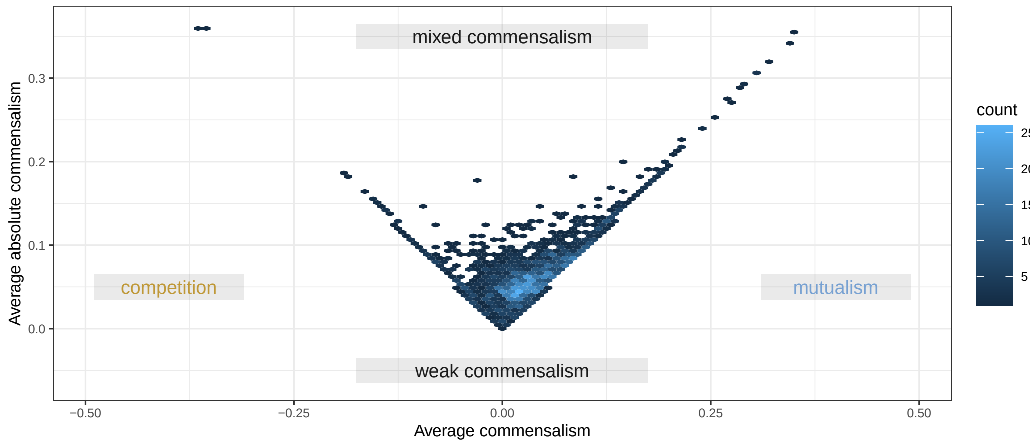


Fig. 1. Two-dimensional histogram showing ecological communities on Reddit in our typology. The X-axis shows the overall degree of mutualism or competition in clusters of subreddits with high user overlap based on the average commensalism. The Y-axis shows the average absolute commensalism to represent the overall magnitude of competitive or mutualistic interdependence.

having mixtures of different forms of commensalism and some with few significant commensal relationships. Mutualism is ~~typical~~ most common with the mean community having an average commensalism of 0.03 ($t = 22.5$, $p < 0.001$). ~~Indeed, mutualism is more common than competition as~~ 1339 clusters (78.3%) are ~~have positive~~ exhibit mutalism as measured by average commensalism. ~~Also,~~ We also find that communities with greater average commensalism ~~typically~~ have greater average absolute commensalism (Spearman's $\rho = 0.53$, $p < 0.001$) indicating that not only are most ecological communities mutualistic, but that mutualistic communities tend to have stronger commensal relationships. Note that our analysis, by construction, examines commensal relationships among subreddits having relatively high degrees of user overlap. Therefore, our community ecology analysis suggests that among groups with similar sets of users, ~~typical~~ commensal relationships in Reddit are typically mutualistic.

5.2 Example ecological communities

four

We turn now to ~~4~~ case studies which illustrate our typology of ecological communities of online groups ~~shown in~~ Figure 1. shows that we able to identify We ~~find~~ clusters of subreddits characterized by mutualism, competition, clusters with a mixture of mutualism and competition, and clusters without strong relationships ~~at all~~. To illustrate our typology and answer RQ1, We select one case of each of these four ~~4~~ types using our measures of average commensalism and average absolute commensalism.

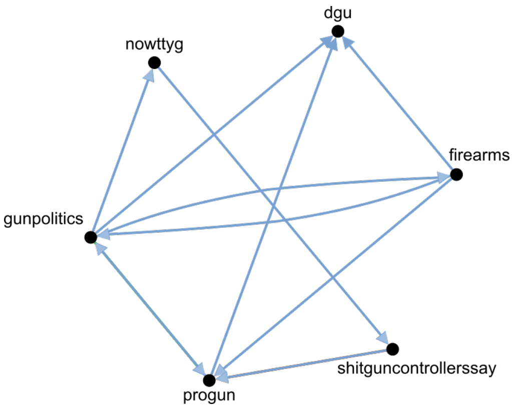
RQ

~~We select case studies from among the 1208 clusters found by the affinity propagation algorithm.~~ To allow for more interesting network structures, we draw ~~limit~~ our cases from ~~to~~ the 861 large clusters having at least four ~~4~~ subreddits.

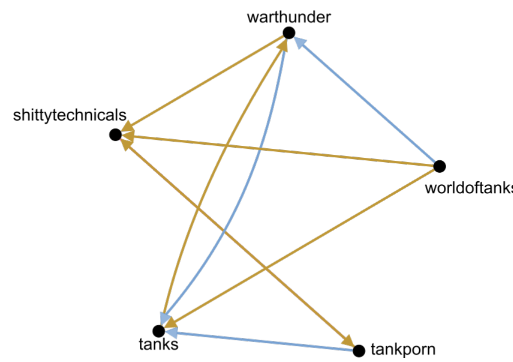
In Figure 2, we present network visualizations representing statistically significant impulse response functions as described in §4.7. In the course of our analysis, we also examined the terms of the vector autoregression parameter Φ, the impulse response functions, and model fits and forecasts, all 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.

5.2.1 Mutualism among pro-gun subreddits. To find a case characterized by mutualism, we selected the top 28 (95%) large clusters by average commensalism. From these, we arbitrarily chose one ~~an~~ ecological community ~~we found interesting~~, the **pro-gun** cluster, which includes 6 groups related to gun ownership and political issues. These generally discuss issues from a right-wing perspective, constitutive subreddits include ~~such as~~ /r/progun and /r/shitguncontrollerssay.

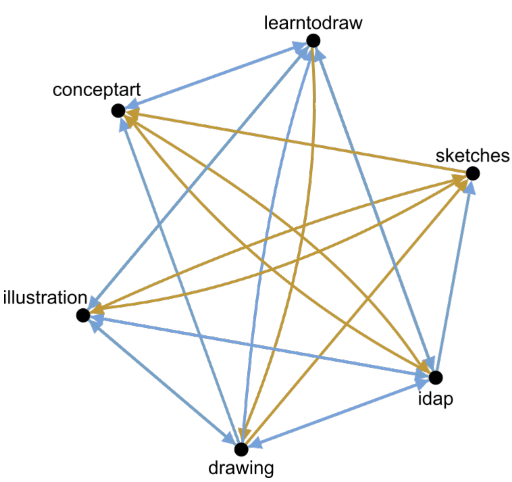
Italicize



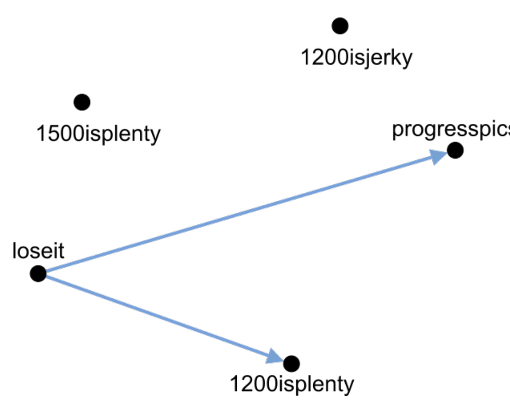
(a) An ecological community of pro-gun subreddits that is highly mutualistic. We detect fully mutualistic relationships between /r/firearms and /r/gunpolitics and between /r/gunpolitics and /r/progun as well as many partially mutualistic relationships.



(b) An ecological community of miliatry vehicle subreddits that ~~is~~ relatively competitive. We detect ~~that~~ /r/tanks ~~is~~ a predator of an relationship of /r/warthunder, a fully competitive relationship between /r/shittytechnicals and /r/tankporn, two partially mutualistic relationships and three partially competitive relationships.



(c) An ecological community of /r/ drawing subreddits characterized by a network dense with both mutualistic and competitive relationships. We detect that /r/learntodraw is a predator of /r/drawing, a number of fully competitive relationships including ~~that~~ between /r/illustration and /r/sketches, fully mutualistic relationships like those between /r/illustration, /r/idap, and /r/ drawing as well as several partially mutualistic or competitive relationships.



(d) An ecological community of weight loss subreddits characterized by relatively sparse commensalism. We detect only two partially mutualistic relationships from /r/loseit to /r/progresspics and to /r/1200isplenty.

Fig. 2. Network visualizations of commensal relationships in example ecological communities of subreddits with overlapping users.

The dense mutualism among these subreddits, shown in Figure 2a, suggests that they have complementary roles in their ecological community. We observe a fully mutualistic relationship between the two subreddits that share the generally scope~~d~~ of ~~on topic of~~ "gun politics:" /r/progun and /r/gunpolitics. This means that an increase in the size of /r/progun predicts a subsequent increase in the size of /r/gunpolitics and~~, conversely,~~ that growth in /r/gunpolitics predicts growth in /r/progun. There is a second fully mutualistic relationship between /r/gunpolitics and /r/firearms, another generalist subreddit that contains political content~~, but is not solely focused on it~~.

All of the remaining subreddits have specialized focuses related to ~~on collecting specific categories of~~ pro-gun information~~, and~~ W~~w~~e observe many partially mutualistic relationships among them. /R/nowttyg is ~~titled as~~ an acronym for "no one wants to take your guns" and exists to detail "evidence contradicting the gun control movements claim that they merely seek moderate" proposals that don't involve the seizure of existing firearms. It has a partially mutualistic relationships with /r/shitguncontrollerssay, such that growth in /r/nowtyg predicts growth in /r/shitguncontrollerssay. /R/firearms, /r/progun, and /r/gunpolitics all have partially mutualistic relationships benefiting /r/dgu, which stands for "defensive gun use" ~~and according to~~ which ~~its posted description~~ is "dedicated to cataloging incidents in the United States where legally owned or legally possessed guns are used by civilians to deter or stop crime." While explaining why different online groups form mutualistic or competitive relationships is left to future research, the example of pro-gun subreddits shows how groups with related topics and overlapping participants can have mutualistic relationships ~~as~~ where growth in one predicts growth in the rest.

5.2.2 Competition among military vehicle subreddits. For a competitive case, we selected clusters with low average commensalism and high average absolute commensalism. Since mutualism is more common than competition, we first selected clusters from the bottom 27 (5%) ~~large~~ clusters by average commensalism. From these, we chose an ecological community that we label **tanks**, which includes ~~5~~ five subreddits about military vehicles and related video games. Among the ~~5~~ five subreddits in this cluster, /r/shittytechnicals, /r/tanks and /r/ tankporn all feature pictures of tanks or other military vehicles. /r/shittytechnicals is specifically for "improvised armed vehicles" which "~~are~~ are defined as makeshift/homemade vehicles that have been modified with weapons and armour," while /r/tanks and /r/tankporn have photography of more conventional modern and historical military vehicles. The other two subreddits in the group are about video games with realistic gameplay emphasizing military vehicles.

In contrast to the **progun** ecological community, the **tanks** cluster has a mixture of competitive and mutualistic ties~~, but mostly mutualistic ones~~ as visualized in Figure 2b. 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 ~~as discussed above in section §5.1~~. That said, we detect a fully competitive relationship between /r/shittytechnicals and /r/tankporn, meaning that an increase in the size of /r/shittytechnicals predicts a subsequent decrease in the size of /r/tankporn

vice versa
and ~~conversely, growth in /r/tankporn predicts decline in /r/shittytechnicals~~. We also observe a number of partially competitive relationships like that from `/r/worldoftanks` to `/r/shittytechnicals` indicating that an increase in the size of `/r/worldoftanks` predicts a decrease in the size of `/r/shittytechnicals`. We also observe that `/r/tanks` is a predator of `/r/warthunder` as growth in `/r/warthunder` predicts growth in `/r/tanks` while growth in `/r/tanks` predicts decline in `/r/warthunder`. ~~Contrasting the competitive **tanks** ecological community with the mutualistic **progun** community raises questions about when ecological communities will be mutualistic or when they will be competitive. It does not seem obvious that the **progun** ecological community is mutualistic while the **tanks** cluster is competitive.~~

5.2.3 Complex commensalism among art and drawing subreddits. Next, we turn to two examples of ecological communities with low average commensalism but different levels of average absolute commensalism. We begin by first selecting the bottom 54 (10%) large clusters ~~(those having at least 4 subreddits~~) with average commensalism closest to 0. To find an ecological community with a mixture of mutualism, we select from these the top 27 (50%) clusters by average absolute commensalism and chose the **drawing** cluster containing 6 groups where people share their sketches and illustrations. To find a case where commensal relationships are weak,
can't do this!
27 (50%)
we select the bottom ~~10 (20%)~~ by average absolute commensalism. From these we chose the **weight loss** cluster containing five ~~5~~ groups supporting people seeking to lose weight.

The commty includes ties
As shown in Figure 2c, the ecological community of drawing subreddits is dense with commensalism. ~~having~~ a number of fully mutualistic (e.g. between `/r/learntodraw`

and /r/idap, which stands for "I drew a picture") and fully competitive relationships (e.g., between /r/illustration and /r/sketches) as well as a predator y relationship (/r/learntodraw is a predator of /r/drawing) and numerous partially competitive and mutualistic relationships. Though average commensalism among these subreddits is near 0, our analysis reveals a complex ecological community with a mixture of ~~all types of~~ commensal relationships.

5.2.4 Sparse commensalism among weight loss subreddits. By contrast, the **weight loss** ecological community is sparse, having only two significant commensal relationships among its 5 member groups. ~~Among~~ In this ecological community ~~of weight-loss subreddits~~, /r/loseit is for "people of all sizes to discuss healthy and sustainable methods of weight loss." and has partially mutualistic relationships with /r/progresspics, where people share pictures from before and after they lost weight ~~and~~ as well as with /r/1200isplenty. When /r/loseit grows, ~~then~~ /r/progresspics and /r/1200isplenty are predicted to subsequently grow. /R/1200ispl advocates a 1200 calorie diet, a weight loss strategy considered relatively extreme by the more moderate /r/1500isplenty and /r/loseit communities. Based on its sidebar and top posts, /r/1200isjerky is another weight-loss support group that pokes fun at what many perceive to be unhealthy or unkind patterns associated with the /r/1200isplenty community.

The **drawing** and **weight loss** ecological communities illustrate how subreddits with overlapping users can have relatively strong or weak forms of ecological interdependence. Though both clusters are defined by having relatively high degrees of user overlap ~~among them compared to their overlaps with subreddits outside of~~

~~their ecological community~~, the **drawing** cluster is characterized by a dense commensal network reflecting complex ecological dynamics while the **weight loss** cluster has fewer ~~is sparse reflecting~~ between the ~~that the~~ growth trajectories of its member groups ~~among these communities~~ ~~lack~~ strong dependencies.

~~sOn the other shown in , between Seattle-area subreddits fMmsbut 210 very likely~~

5.3 Environmental and relational ecological models

5.3.1 Density dependence. ~~As discussed in §2.2, population ecology approaches in social computing propose that the relationship between overlap-density and growth/survival outcomes reflect an environment that may be competitive, mutualistic, or a mixture of both [67, 72]~~. In this section, we compare the environmental approach of population ecology with the relational approach of community ecology. We begin by testing the classical prediction of density dependence theory as formulated in **H1** using **Model 2** (Equation 8 in §4.8) which has first- and second-order terms for effect of overlap density on growth. As described in §2.2, **H1** hypothesizes ~~proposes~~ that overlap density will have a curvilinear ~~inverse-U-shaped~~ ~~(∩-shaped)~~ relationship with growth indicated by a positive first-order regression coefficient and a negative second-order coefficient.

As predicted by **H1**, we observe a ∩-shaped relationship between author overlap density and growth. Table 2 shows regression coefficients for **Models 1-3**. Figure 3 plots the marginal effects of author overlap density on growth for the median subreddit laid over the data on which the model is fit. For most subreddits, increasing author overlap density is associated with higher ~~increasing~~ growth rates ~~(or at least lesser~~

~~decline~~). The "sweet spot," or point where increasing density ceases to predict increasing growth and begins to predict decreasing growth is at the 53$^{\text{th}}$ percentile. Prototypical subreddits at this author overlap density declined by about -0.01 (95% CI:[-0.03,0.02]) distinct commenting users. The magnitude of this decline is less ~~on average that~~ than that experienced by subreddits at the lower and upper extremes of overlap density. Typical groups at the 20$^{\text{th}}$ percentile of overlap density decline by -0.2 (95% CI:[-0.22,-0.17]) distinct commenting users and typical groups at the 80$^{\text{th}}$ percentile decline by -0.15 (95% CI:[-0.2,-0.11]) distinct commenting users.

crosses zero..

While we find support for classical theoretical prediction of a ~~curvilinear,~~ (∩-shaped) relationship between user-overlap density and growth, this does not imply that relationships between highly overlapping communities are more competitive. ~~Rather we found in~~ Instead our results in §4.6 ~~that~~ show that relationships in ecological communities of subreddits with high user overlaps are ~~often~~ typically mutualistic.

Generally

wrong section, ro?

5.2)?

5.3.2 Commensalism is not statistically reducible to overlaps. In §5.1 ~~above~~, we presented examples of diverse ecological communities among subreddits with overlapping members. ~~Yet, this~~ Hence the presence of this diversity does not ~~demonstrate~~ mean that commensal relationships ~~have significance for predicting~~ are related to the growth of online groups. We therefore ~~proposed in~~ hypothesized **H2** that commensal relationships will improve the fit of ~~the~~ a density dependence model ~~predicting growth~~.

the key advance of previous ecological studies

To test **H2**, we compare **Model 1**, our density dependence model with first and second order terms for user overlap density with **Model 3**, which also includes average subreddit commensalism as a predictor. We also examine **Model 2** which

	Model 1	Model 2	Model 3
User overlap density	0.45* (0.04)		0.44* (0.04)
User overlap density^2	−0.09* (0.01)		−0.09* (0.01)
Average subreddit commensalism		0.81* (0.17)	0.71* (0.17)
Constant	−0.58* (0.05)	−0.10* (0.01)	−0.58* (0.05)
Log Likelihood	-12310	-12354	-12302
Observations	9,307	9,307	9,307

Note: Loglinear *p< 0.01

Table 2. Regression predicting subreddit growth as a function of user overlap density. The model supports the prediction of density dependence theory of a ∩-shaped relationship between user overlap density and growth.

only contains average subreddit commensalism. Table 2 shows regression coefficients for our models. We observe that average subreddit commensalism is positively associated with growth ($B_3 = 0.71, SE = 0.17$)~~, which~~. This makes sense as subreddits with greater average subreddit commensalism benefit more from mutualism or are hurt less from ~~commensalism~~ competition (?). (I don't follow)

A likelihood ratio test comparing **Model 3** to **Model 1** supports **H2** and shows that including average subreddit commensalism improves the fit of the model to the data ($\chi^2 = 17$, $p < 0.001$). This shows that looking at a subreddit's commensal relationships explains additional variation compared to ~~first and second order~~ a quadratic specification of user overlap density. Comparing **Model 1** to **Model 2** shows that similarly, user overlap density explains variation that average subreddit commensalism does not

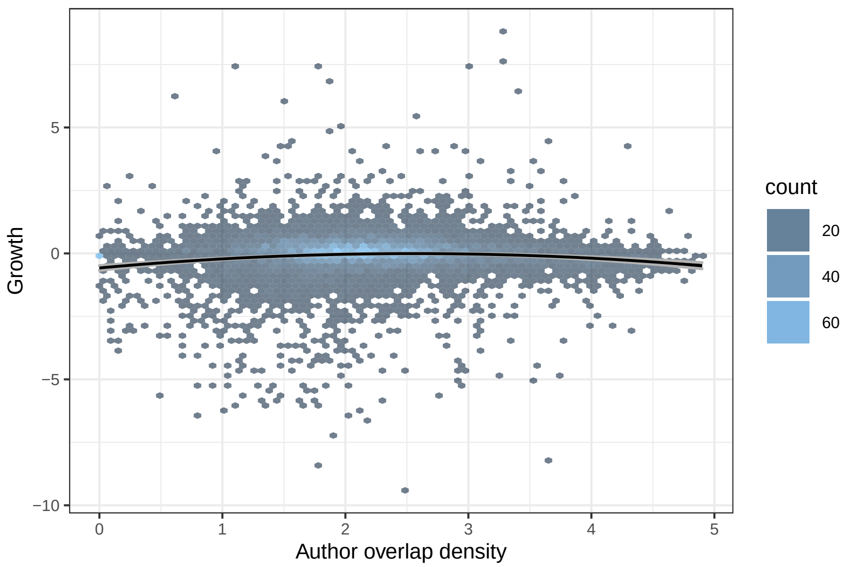


Fig. 3. Relationship between density and growth. A 2D histogram of subreddits with user overlap density (log-transformed) on the X-axis and the change in the logarithm of the number of distinct commenting users on the Y-axis. The black line shows the marginal effect of author overlap density on growth as predicted by **Model 2**. The gray region shows the 95% confidence interval of the marginal effect.

($\chi^2 = 100$, $p < 0.001$). This suggests that environmental and relational forms of competition and mutualism are complementary factors when it comes to explaining subreddit growth.

5.3.3 Forecasting accuracy. The likelihood ratio tests in §5.3.2 are limited because improvements in model fit may be due to unobserved factors predictive of growth that are correlated with commensalism. We ~~proposed~~ hypothesized in **H3** that ~~including~~ accounting for commensal relationships ~~in the model~~ will improve~~s~~ the predictive ~~forecasting~~ performance of our time series model in order to ~~directly~~ test whether the inter-group dependencies in ~~the~~ our VAR models ~~with which we infer~~ can ~~commensal relationships are useful for~~ predict~~ing~~ the future size of subreddits. As described above in §4.5 we test **H3** by comparing two forecasting metrics: the root-mean-squared-error (RMSE) and the continuous ranked probability score (CRPS).

VAR models including commensal terms have superior forecasting performance over 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 ~~into the future~~. The RMSE under the baseline model (0.8) is greater than the RMSE of the VAR models (0.74) and the CRPS of the baseline model (210178) is also greater the CRPS of the VAR models (209268). This ~~is~~ reflects a substantive improvement in forecast accuracy ~~which is~~ robust to the choice of forecasting metric.

Our ~~The~~ 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 we gain from modeling commensal relationships must ~~therefore~~ reflect associations not captured by overlap density, which is a static property of the environment.

6 THREATS TO VALIDITY

Our work is subject to several important threats to validity. The method we propose for identifying commensal relationships between online groups has limitations common to all time-series analysis of observational data. While our community ecology approach assumes that commensal relationships 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 causal 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 causality. ~~Including such groups in our models would potentially change our results.~~ 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 ~~for simplicity~~, ~~but~~ we hope future work ~~can~~ will ~~improve upon our approach and~~ model more complex dynamics with additional lags. Our results are offered as limited temporal associations consistent with inferred commensal relationships. Similarly, regression estimates in **Models 1-3** may be confounded by omitted variables and cannot support causal interpretations. ~~Rather, they test theoretically predicted correlations (in the case of **H1** and are used in likelihood ratio tests of model fit.~~

Like most other time series analysis~~, inferences made by~~ vector autoregression analyses assumes that error terms are stationary. This is difficult to evaluate empirically and may not be realistic [11]. Future work might relax these assumptions using more complex models with time-varying VAR parameters or state space models [8], but such approaches may require additional contextual knowledge and be difficult to scale to an analysis of hundreds of different ecological communities ~~such as ours~~. Such approaches may also be useful for investigating how commensal relationships form or change over time.

Additional threats to validity stem from our use of algorithmic clustering to identify ecological communities. While ~~we justify this on the basis that the algorithm~~ we ~~use finds~~ choose clusters ~~having~~ based on high degrees of user overlap, this is by no means the only possible heuristic ~~that can serve as a basis for clustering~~. Had we instead chosen ~~a heuristic~~ clusters based on topical similarity, we would likely have obtained different results. Furthermore, clustering algorithms like the one we use may not have unique solutions and different initial conditions ~~can~~ might lead to different clusters. While ~~clustering~~ these algorithms allow us to scale our analysis up ~~to a large number~~

~~of subreddits~~, future investigations should also consider qualitative approaches to constructing ecological communities.

7 DISCUSSION

Prior ecological studies in social computing use overlap in participants or topics to characterize the resource environment faced by ~~particular~~ online groups [67, 71, 72]. However, they yield a puzzling set of empirical results about the relationship between resource overlaps and outcomes like growth, decline and survival. Observing that the use of overlap density in these studies potentially aggregates many ~~varying~~ kinds of relationships, we propose a community ecology approach as a first step toward resolving this puzzle. By shifting the focus from environmental forces to commensal relationships between communities that can shape each other's growth, community ecology promises to explain *when* online environments will be competitive and *when* they will be mutualistic. Ital?

To demonstrate community ecology for online groups, we set out in ~~*RQ1*~~ to study networks of commensal relationships between groups which we infer directly using vector autoregression models of group sizes over time (RQ1). We applied this method ~~at a large scale~~ on hundreds of clusters of subreddits with overlapping users and found ecological communities that are dense with mutualism, competition, or mixtures of the two, as well as sparser clusters having few significant commensal relationships. In §5.2 we presented an example of each type to illustrate the variety of ecological communities ~~that can be found~~ we identified. Within a large platform like Reddit, the great number of ecological communities that can be studied should make it possible for future work to construct and test generalizable theories about when and how different types of ecological communities are constructed. Promising directions include studies of the ecological roles of different types of resources, design features of platforms, and governance institutions.

To illustrate the distinction between population and community ecologies, we tested one of the central predictions of the population ecology theory of density dependence in **H1**, and found that —as predicted by the theory— user overlap density has an ~~inverse U-shaped~~ (∩-shaped) association with subreddit growth such that subreddits with moderate overlap density in our data declined less ~~compared to~~ than subreddits with either very low or very high overlap density. Although Population ecology theory suggests that we should conclude that dense environments are more competitive, ~~Yet,~~ this does not mean that most relationships between subreddits with overlapping users are competitive. ~~Rather, in~~ On the contrary, our analysis of **RQ1** ~~we found~~ shows that most ecological communities of subreddits are mutualistic. Although contradicting These findings are not contradictory. ~~but rather~~ Instead, they show how the different levels of analysis of population ecology and community ecology correspond to different kinds of ecological dynamics. Future investigations comparing population and community ecology should study the mutual influence between environmental factors and commensal relationships.

To provide additional support for our claim~~s~~ that community ecology is an important lens for understanding interdependence among online groups, we tested **H2** ~~to~~ and showed that including subreddit average commensalism improves the fit of the density dependence model. Commensal relationships between online groups are significant factors in their growth or decline and are not reducible to resource overlaps or static environmental properties. Similarly, we found support for **H3** ~~to~~ showing that including commensalism in the vector autoregression (VAR) models substantially improves their forecasting ability. These results show that The dynamics of participation in online groups thus depend on the dynamics of other groups in their ecological communities.

Although. We focused on online groups within a single platform: ~~Reddit. However, inter-est~~ groups often use multiple platforms with distinctive affordances for different purposes [41]. Since the (VAR) method relies only on time series data to infer commensal relationships, it can be applied to study ecological communities spanning ~~different~~ social media platforms ~~once they have been identified~~. In this way Community ecology can provide a bridge between quantitative studies of participation in online groups and theories of interconnected information ecologies [48]. 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, 2].

7.1 Implications for Design

In their final chapter of their book on *Building Successful Online Communities*, Kraut et al. [43] 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. We provide a framework for online group managers to think about ecological constraints on group size in terms of a network dyadic relationships. While intuition suggests that online community managers might seek out mutualistic relationships and avoid competitive ones, it is often not obvious whether a community with overlapping users is a competitor or mutualist. Our method provides a way for group managers to know ~~solve this problem~~.

Competitors have a negative impact on growth, but ecological theory suggests that specialization is an adaptive strategy in response to competition [1, 12, 43, 54]. For example, the growth of Wikipedia caused other online encyclopedia projects to shift their focus [33]. 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, perhaps by differentiation through unique design, policies, or governance practices.

While competitive relationships are defined by how they decrease the size of groups, competition can also be important to the functioning of the broader ecological community. Exit to an alternative group can be an avenue for political change in response to grievances and poor governance [24, 35]. Although The existence of ~~a sufficient diversity of~~ alternative groups might ~~institutions is likely to~~ lead to competition, ~~but~~ it might also make the expression of voice more persuasive to moderators or platforms [35].

Groups looking to increase activity should desire to seek out mutualistic relationships and we believe ~~. We suggest~~ that designers of online platforms can help them do so. Features such as meta-groups, group search, and recommendation engines, and practices like linking related groups may lower boundaries between groups and thereby support mutualistic relationships. However, it is not obvious to what extent particular features will support competition, mutualism, or both. Using our method, community managers may be able to better understand how design features give rise to mutualism or competition.

8 CONCLUSION

Explanations for the rise or decline of online groups often look to internal mechanisms. For example, Wikipedia's decline appears to have been caused by structural barriers to newcomers [27]. Did competition with emerging platforms like Facebook also limit Wikipedia's growth? Ecological approaches provide a path toward answering such questions. Our work provides a novel approach for doing so and takes a step to resolve empirical puzzles raised by prior ecological work. By narrowing the focus to the dyadic commensal relationships between groups that matter, In social computing the community ecology framework we present also raises a host of new research directions ~~for social computing research~~. We look forward to building on this work and to building stronger and better online groups through in process.

REFERENCES

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

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

[3] William P. Barnett and Glenn R. Carroll. 1987. Competition and Mutualism among Early Telephone Companies. *Administrative Science Quarterly* 32, 3 (1987), 400–421.

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

[5] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The Pushshift Reddit Dataset. *arXiv:2001.08435 [cs]* (Jan. 2020). arXiv:2001.08435 [cs]

[6] Gary S. Becker. 1965. A Theory of the Allocation of Time. *The Economic Journal* 75, 299 (Sept. 1965), 493.

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

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

[9] 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.

[10] 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.

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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 (Nov. 2018), 32:1–32:25.

[17] 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.

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

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

[20] 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.

[21] 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, 72–81.

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

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

[24] 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.

[25] 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.

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

[27] 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.

[28] 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]

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

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

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

[32] 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

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

[34] 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.

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

[36] 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.

[37] 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.

[38] Alexander Jordan, Fabian Krüger, and Sebastian Lerch. 2019. Evaluating Probabilistic Forecasts with scoringRules. *Journal of Statistical Software* 90, 1 (Aug. 2019), 1–37.

[39] 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.

[40] Michael L. Katz and Carl Shapiro. 1985. Network Externalities, Competition, and Compatibility. *The American Economic Review* 75, 3 (1985), 424–440.

[41] 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.

[42] 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.

[43] Robert E. Kraut, Paul Resnick, and Sara Kiesler. 2012. *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press, Cambridge, MA.

[44] 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.

[45] 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).

[46] J. Miller McPherson. 1983. An Ecology of Affiliation. *American Sociological Review* 48, 4 (1983), 519–532.

[47] Peter Monge and Marshall Scott Poole. 2008. The Evolution of Organizational Communication. *Journal of Communication* 58, 4 (Dec. 2008), 679–692.

[48] Bonnie A. Nardi and Vicki L. O'Day. 1999. *Information Ecologies : Using Technology with Heart*. The MIT Press, Cambridge, Massachusetts.

[49] 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.

[50] 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.

[51] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830.

[52] Bernhard Pfaff. 2008. VAR, SVAR and SVEC Models: Implementation Within R Package Vars. *Journal of Statistical Software* 27, 1 (July 2008), 1–32.

[53] Steven Piantadosi, David P Byar, and Sylvan B Green. 1988. The Ecological Fallacy. *American Journal of Epidemiology* 127 (1988), 893–904.

[54] 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.

[55] W. S. Robinson. 1950. Ecological Correlations and the Behavior of Individuals. *American Sociological Review* 15, 3 (1950), 351–357.

[56] Paul M. Romer. 1990. Endogenous Technological Change. *Journal of Political Economy* 98, 5, Part 2 (Oct. 1990), S71–S102.

[57] Martin Ruef. 2000. The Emergence of Organizational Forms: A Community Ecology Approach. *Amer. J. Sociology* 106, 3 (Nov. 2000), 658–714.

[58] Christopher A. Sims. 1980. Macroeconomics and Reality. *Econometrica* 48, 1 (1980), 1–48.

[59] Jesper B. Sørensen. 2004. Recruitment-Based Competition between Industries: A Community Ecology. *Industrial and Corporate Change* 13, 1 (Feb. 2004), 149–170.

[60] 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.

[61] 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.

[62] 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.

[63] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2020. The Population Ecology of Online Collective Action.

[64] Sho Tsugawa and Sumaru Niida. 2019. The Impact of Social Network Structure on the Growth and Survival of Online Communities. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '19)*. Association for Computing Machinery, Vancouver, British Columbia, Canada, 1112–1119.

[65] Herman A Verhoef and Peter J Morin. 2010. *Community Ecology: Processes, Models, and Applications*. Oxford University Press, Oxford.

[66] 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.

[67] 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.

[68] Steven Weber. 2000. The Political Economy of Open Source Software. (June 2000).

[69] 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.

[70] 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.

[71] 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.

[72] 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)*. ACM, New York, NY, 281–290.