Identifying Competition and Mutualism between Online Groups

Nathan TeBlunthuis[1 2], **Benjamin Mako Hill**[1]

[1] Department of Communication, University of Washington,
[2] Department of Communication Studies, Nothwestern University
nathante@uw.edu

Abstract

Platforms often host multiple online groups with overlapping topics and members. How can researchers and designers understand how related groups affect each other? Inspired by population ecology, prior research in social computing and human-computer interaction has studied related groups by correlating group size with degrees of overlap in content and membership, but has produced puzzling results: overlap is associated with competition in some contexts but with mutualism in others. We suggest that this inconsistency results from the aggregation of intergroup relationships into an overall environmental effect that obscures the diversity of competition and mutualism among related groups. Drawing on the framework of community ecology, we introduce a time series method for inferring competition and mutualism. We then use this framework to inform a large-scale analysis of clusters of subreddits that all have high user overlap. We find that mutualism is more common than competition.

Introduction

Online groups do not exist in isolation.[1] Recent research has sought to quantify how online groups share users or topics (Datta, Phelan, and Adar 2017), and how such interactions relate to outcomes such as the emergence of new groups (Tan 2018), the spread of hate speech (Chandrasekharan et al. 2017) and contributions to peer-produced knowledge (Vincent, Johnson, and Hecht 2018). This work has demonstrated that intergroup interactions matter, but very little intergroup research has tackled questions of group success—i.e., why some online groups succeed in maintaining active and long-lived participation while most do not (Kraut and Fiore 2014; Resnick et al. 2012). Can intergroup relationships explain whether online groups will grow or decline?

We seek to answer this question by following prior ecological studies in social computing (Wang, Butler, and Ren 2012; Zhu, Kraut, and Kittur 2014; Zhu et al. 2014). We take inspiration from organizational ecology (Hannan and Freeman 1989; Baum and Shipilov 2006), an influential body of theory in sociology, and analyze *competition* and *mutualism* between online groups. Prior ecological 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 (Zhu et al. 2014), but Usenet groups with overlapping memberships failed more quickly (Wang, Butler, and Ren 2012).

We propose that limitations of the *population ecology* framework used by these studies give rise to these inconsistencies. Therefore, we introduce an alternative framework inspired by *community ecology* that seeks to directly study competitive and mutualistic interactions between groups. Population ecology models how overlapping resources among groups affect their subsequent growth, decline, or survival (Freeman and Audia 2006; Astley 1985), but it does not directly study interactions. By contrast, community ecology models related groups as an "ecological community" structured by a network of competitive and mutualistic relationships.

We introduce our community ecology approach and compare it to the population ecology approach from prior work in a two-part empirical study of 641 clusters of online groups among the 10,000 communities on Reddit with the most contributors. Study A illustrates the population ecology approach in order to provide a basis to compare it with the community ecology analysis of Study B. Study A demonstrates a prototypical population ecology analysis by testing density dependence theory. Its findings suggest that competition is strongest when user overlap is high and that mutualism is weakest when overlap is low. Prior studies would interpret the results of this analysis as suggesting that high degrees of user overlap are associated with competition.

In Study B, we introduce our method for inferring networks of ecological relationships among related online groups based on clustering analysis and vector autoregression (VAR) models of group size over time (Ives et al. 2003). VAR models are used in biological ecology to make inferences about competitive or mutualistic interactions between species. We illustrate the method in four case studies and present a large-scale analysis showing that mutualistic interactions are far more common than competitive ones. To validate our approach, we show that including ecological interactions in our VAR models improves time series forecast-

[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.

ing.

Our findings illuminate the different contributions of population ecology and community ecology. While Study A suggests that competition is strongest when user overlap is high, Study B finds widespread mutualism among groups with highly overlapping memberships. Although these findings might seem contradictory, we argue that population and community ecology analyses provide complementary views. Population ecology points to favorable or unfavorable conditions for building online groups—conditions that may or may not involve competition and mutualism. A community ecology analysis can infer local networks of competition and mutualism to explain how specific ecological relationships contribute to growth or decline. By demonstrating that ecological relationships within clusters of highly related groups are important—and by describing how to measure them—this paper lays the groundwork for future investigations into interdependent online groups and designs that support ecological communities.

Related Work

Online groups are important sites for social support (De Choudhury and De 2014), entertainment (Ducheneaut et al. 2006), information sharing (Benkler 2006), and political mobilization of disinformation campaigns and protest movements (Choudhury et al. 2016; Benkler et al. 2013; Krafft and Donovan 2020). Although an online group's ability to achieve its goals depends on attracting and retaining contributors, few develop a sizable group of participants (Kraut, Resnick, and Kiesler 2012). Many attempts to explain the growth and decline of online groups look to properties of individual groups such as characteristics of founders and designs for regulating behavior (Kraut, Resnick, and Kiesler 2012; Halfaker et al. 2013; TeBlunthuis, Shaw, and Hill 2018).

By contrast, recent research shows the importance of interdependence among online groups (Kairam, Wang, and Leskovec 2012; Tan 2018; Waller and Anderson 2019). For example, banning hate subreddits reduced hate speech in related subreddits (Chandrasekharan et al. 2017), Reddit and Stack Overflow receive substantial benefits from activity on Wikipedia (Vincent, Johnson, and Hecht 2018), and editors make valuable and qualitatively different contributions across different languages of Wikipedia (Hale 2015). Our work contributes to this literature by providing a new conceptual lens and statistical method for studying intergroup connections.

Ecological Interdependence

Ecological approaches to online groups see online groups as depending on resources. Our conceptual approach, like prior ecological research in social computing and information systems, builds on resource dependence theory (RDT) (Butler 2001; Wang, Butler, and Ren 2012). According to RDT, members of online groups contribute resources such as content, information, attention or social interactions that sustain the group.

Ecological approaches observe that interrelated online groups may share resources with one another and affect each other's growth and survival as a result. *Rival resources* like participants' time, attention, and efforts become unavailable when used by one group (Benkler 2006; Romer 1990), and competition over important rival resources can explain declines in participation (Wang, Butler, and Ren 2012). On the other hand, the value of a *nonrival* resource does not decrease (and may even increase) when it is used. One example is a network effect, when the usefulness of a communication network increases as more people join it (Fulk et al. 1996). Similarly, the usefulness of an information good can increase as more people come to know, refer to and depend upon it. Nonrival resources that "spill over" can result in mutualism that promotes growth in related groups (Zhu, Kraut, and Kittur 2014).

Population Ecology, Density Dependence and Overlapping Resources

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 (Hannan and Freeman 1989). Organizational ecology has inspired at least three high-quality empirical studies of how resources shared by online groups shape their growth, decline, or survival (Wang, Butler, and Ren 2012; Zhu, Kraut, and Kittur 2014; Zhu et al. 2014). All three studies draw from the *population ecology* strand of organizational ecology, and specifically enage with *density dependence theory* (DDT).

DDT concieves of competitive or mutualistic forces as a function of population *density*. In the earliest and most influential studies of DDT, density is simply the size of a population, a homogeneous set of organizations of groups facing the same competitive and mutualistic pressures (Aldrich and Ruef 2006). However, online groups sharing a platform have diverse topics (Kairam, Wang, and Leskovec 2012), norms (Chandrasekharan et al. 2018), and user bases (Tan 2018; Tan and Lee 2015). To account for this diversity, ecological studies of online groups have modeled density dependence based on the concept of *overlap density* (Baum and Shipilov 2006; Wang, Butler, and Ren 2012; Zhu, Kraut, and Kittur 2014; Zhu et al. 2014). Overlap density measures the extent to which one group's members or topics overlap with all other groups'. Overlap density thus characterizes a group's *niche* or local *resource environment* defined by its distinctive topic and membership.

DDT proposes a model for the growth of organizational populations in which mutualism drives a virtuous cycle of population growth (Carroll and Hannan 1989; Hannan and Freeman 1989). For example, a population of online groups, such as those sharing a platform, may grow in size as their platform gains in popularity, as established groups spin off new ones, and as useful knowledge develops that can be shared between groups (Tan 2018; Zhu, Kraut, and Kittur 2014). On the other hand, when density is high, competition among population members over rival resources limits growth (Hannan and Freeman 1989). DDT thus proposes a trade-off in which low density reflects limited opportunities for mutualistic contributions of nonrival resources, while high density reflects competition over rival resources. Therefore, DDT predicts that the relationship between density and positive outcomes such as growth or survival is ∩-shaped (inverse-U-shaped) (Baum and Shipilov 2006; Carroll and Hannan 1989).

Tests of DDT in populations of online groups yield inconsistent results. Wang, Butler, and Ren (2012), find that user overlaps among Usenet newsgroups are associated with decreasing numbers of participants. Similarly, TeBlunthuis, Shaw, and Hill (2017) finds that topical overlaps between online petitions are negatively associated with participation. By contrast, Zhu, Kraut, and Kittur (2014) find that membership overlap is positively associated with increasing survival of new Wikia wikis. Only Zhu et al. (2014) find support for the ∩-shaped relationship predicted by DDT in their study of an enterprise social media platform.

In Study A, we test DDT using data from Reddit. The classical logic of DDT appears reasonable in the context of Reddit because low overlap density likely reflects an impoverished environment lacking in nonrival resources such as the skills and knowledge of experienced users, while a group with high overlap density likely faces competition over its members (Zhu et al. 2014; Zhu, Kraut, and Kittur 2014): ***(H1)** The relationship between overlap density and the growth of online groups is ∩-shaped.*

Introducing Community Ecology

The fundamental distinction between population ecology and community ecology theories is where they locate ecological dynamics like competition and mutualism. In population ecology, competition and mutualism are properties of an environmental niche. In community ecology, they are relations in networks of interdependent groups called *ecological communities* (Freeman and Audia 2006; Aldrich and Ruef 2006; Astley 1985). While most community ecology studies of classical organizations analyze ecological communities of different organizational forms, some, like our study, analyze communities of related organizations (Freeman and Audia 2006; Powell et al. 2005; Margolin et al. 2012).

Community ecology focuses on *ecological interactions* (Aldrich and Ruef 2006). Mutualism is an ecological interaction where one group has a positive influence on a second such that growth in the first group leads to growth in the second. Competition is when one group has a negative effect on the second such that growth in the first group leads to decline in the second. Mutualism (or competition) from one group to another group may (or may not) be returned. Moreover, ecological interactions can be mutualistic in one direction and competitive in the other. As a result, these relationships are modeled as the edges of a directed network. The goal of many community ecology analyses in both biology and organization science is to infer and analyze the *community matrix*, which quantifies this network competitive and mutualistic interactions (Ives et al. 2003; Aldrich and Ruef 2006).

In Study B, we demonstrate community ecology by inferring networks of ecological interactions in ecological communities on Reddit to determine whether mutualism or competition among subreddits is more common. We then present case studies to illustrate different types of ecological communities. Finally, we evaluate whether modeling ecological interactions is useful for making time series forecasts of participation in online groups: (***H2**) The addition of ecological interactions to a baseline time series model improves forecasting performance.*

Materials & Methods

We analyze data from the publicly available Pushshift archive of Reddit submissions and comments from December 5th 2005 to April 13th 2020 (Baumgartner et al. 2020). Within this dataset, we limit our analysis to submissions and comments from the 10,000 subreddits with the highest number of comments. The top 10,000 subreddits provide a sufficiently large number of ecological communities for our statistical analysis. There are 702 subreddits larger than the smallest subreddit included in our dataset that have a majority of submissions marked “NSFW,” which often indicates pornographic material. As others have done in large-scale studies of Reddit (e.g., Datta, Phelan, and Adar 2017), we exclude these subreddits to avoid asking members of our research team to inspect clusters including pornography.

Study A: Density Dependence Theory

In Study A, we illustrate population ecology by testing H1 using a log-linear regression model to predict subreddit *growth* as a quadratic function of our measure of *overlap density* which we construct by aggregating our measure of *user overlap*.

User overlap $o_{i,j}$ quantifies the degree to which two subreddits (i and j) share users. Zhu, Kraut, and Kittur (2014) and Wang, Butler, and Ren (2012) both measure the user overlap between two groups by counting the number of users contributing to both groups at least once and exclude users who appear in more than 10 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 groups while participating heavily in few (Zhang et al. 2017). Therefore, our measure of user overlap follows Datta, Phelan, and Adar (2017) by using the number of comments each user makes in each pair of groups.

To measure user overlap between subreddits, we first build user frequency vectors by counting the number of times each user comments in each subreddit. 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 latent semantic analysis (LSA) to reduce the dimensionality of the user frequency vectors:

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

where $\mathbf{F}$ is the matrix where columns are author frequency vectors F_j and $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ 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 $\widetilde{F_j}$, their approximation in the k-dimensional space. Our measure of *user overlap* ($o_{i,j}$) is the cosine similarity between these vectors:

$$o_{i,j} = \frac{\widetilde{F_j} \cdot \widetilde{F_i}}{\|\widetilde{F_i}\|\|\widetilde{F_j}\|} \tag{3}$$

where $\|\widetilde{F_i}\|$ is the Euclidean norm of the transformed user frequencies for subreddit i.

Growth Y_i, the dependent variable in our density dependence model testing H1 is measured as the change in the (log-transformed) size of a subreddit during the last 24 weeks of our data, from November 4$^{\text{th}}$ 2019 to April 13$^{\text{th}}$ 2020.

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

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

where M is the set of groups in our dataset.

Regression model for H1 To test H1, we fit Model 1:

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

where Y_i is the growth of subreddit i and d_i is its overlap density. The model has first and second-order terms for overlap density to allow for a curvilinear relationship between *overlap density* and *growth*.

Study B: Introducing Community Ecology

In Study B, we present our method for studying competition and mutualism in ecological communities. We first cluster subreddits having overlapping users in order to identify ecological communities on which we can fit VAR models predicting *group size*. We visualize competition-mutualism networks in example ecological communities based on interpreting the VAR models using impulse response functions (IRFs). To present a broader view of the types of ecological communities found on Reddit, we quantify the overall degree of competition and mutualism in each ecological community as *average ecological interaction* and quantify the overall intensity of ecological interactions as *ecological interaction strength*. Finally, we test H2 in terms of the root-mean-square-error (RMSE) and continuuous ranked probability score (CRPS) forecasting metrics.

Clustering to identify ecological communities Analyzing networks of ecological interactions is the key difference between community ecology and population ecology. To identify ecological communities of related subreddits, we use a clustering procedure based on user overlap. We selected a clustering model using grid search to obtain a high silhouette coefficient. The silhouette coefficient captures the degree to which a clustering creates groups of subreddits with high within-cluster similarity.

We ran the affinity propagation, HDBSCAN and *k*-means clustering algorithms and selected the algorithm, hyperparameters, and LSA dimensions k that resulted in the clustering with the greatest 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 hyperparameters for the HDBSCAN algorithm could improve the silhouette coefficient, but at the cost of greatly increasing numbers of isolated subreddits. Choosing a 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 isolated subreddits from our analysis. More details about our clustering selection process can be found in the online supplement.

We evaluate the external validity of the chosen clustering using the purity evaluation criterion. To do so, an undergraduate research assistant examined a random sample of 100 clusters including 744 subreddits. By visiting the subreddits and using their own judgment, the assistant flagged subreddits that did not seem like a good fit for their assigned cluster. Using these labels and excluding 25 subreddits that have been deleted, made private, or banned, we calculated the purity of our clustering as 0.92. In other words, we estimate that 92% of subreddits belong to their assigned cluster.

Group size is the dependent variable of the models that we use to infer ecological interactions. Measured as the number of distinct users commenting 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 so we transform it by adding 1 and taking the natural logarithm.

Inferring ecological interactions The community matrix $\mathbf{\Phi}$ 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 with the Gompertz of population growth, a common theoretical model in ecology (Ives et al. 2003).

VAR(1) models generalize auto-regressive AR(1) models in time series analysis. Where AR(1) models predict the state of a single time series as a function of its previous values, VAR(1) models predict several time series as a function of each other's previous values (Ives et al. 2003):

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

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 $\mathbf{\Phi}$, 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; 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 introduce a counter trend during the period prior to the creation of subreddits so that each group's growth trend begins in the period the group is created. For each group j created at time t_j^0 we fill X_j with the sequence $[1, 2, 3, \dots, t_j^0 - 1, 0, 0, 0, \dots]$. In other words, X_j adds a counter-trend only during the period prior to the first comment in subreddit j. We fix the elements $a_{j,i}$ of A_k equal to 0 unless $i = j$, so the counter trend only influences subreddit j. This effectively sets $a_{j,j}$ approximately equal to $-b_{1,j}$.

We fit VAR(1) models using ordinary least squares as implemented in the `vars` R package to predict the group size each week over the history of each subreddit prior to November 4th 2019. 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.

Characterizing ecological communities In Study B, we interpret the community matrix $\mathbf{\Phi}$ as a directed network of ecological interactions, a *competition-mutualism network* (Ives et al. 2003). Although the elements of $\mathbf{\Phi}$ correspond to direct associations between group sizes, 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. In large VAR models that contain many groups, the large 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 impulse response functions (IRFs) (Ives et al. 2003).

An IRF predicts how much each group's size would change in response to a sudden increase in the size of each other group:

$$\mathbf{\Theta_t} = \mathbf{\Theta_{t-1}} \mathbf{\Phi}, t = 1, 2, ... \quad (7)$$

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 $\theta_{i,j}^t$ corresponding to the response of group j to the impulse of group i.

We use IRFs of our VAR(1) models to make our visualizations of example competition-mutualism networks. We compute the IRFs with bootstrapped confidence intervals (CI) based on 1,000 samples using the `vars` R package. We draw an edge $i \to j$ in the competition-mutualism network if the 95% CI of $\theta_{i,j}^t$ does not include zero at any time $t \in (0, 10]$. If $\theta_{i,j}^t > 0$, the edge indicates mutualism and if $\theta_{i,j}^t < 0$ the edge indicates competition.

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 the off-diagonal coefficients of $\mathbf{\Phi}$:

$$\overline{m} = \frac{1}{|M| - 1} \sum_{i \in M} \sum_{j \in M; j \neq i} \phi_{i,j} \quad (8)$$

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 than mutualistic ones.

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 $\mathbf{\Phi}$:

$$\kappa = \frac{1}{|M| - 1} \sum_{i \in M} \sum_{j \in M; j \neq i} \left|\phi_{i,j}\right| \quad (9)$$

where $\left|\phi_{i,j}\right|$ is the absolute value of the coefficient $\phi_{i,j}$. The average ecological interaction can be close to 0 if the ecological interaction strength is low or if the ecological interaction strength is high and results from a mixture of competitive and mutualistic interactions that cancel one another out when averaged.

Forecasting growth To test H2, we evaluate whether the modeling of ecological interactions improves the time series forecasting of future participation in online groups by comparing the model in Equation 6 to a baseline model with off-diagonal elements of $\mathbf{\Phi}$ fixed at 0. This baseline model is equivalent to our VAR model, but excludes ecological interactions.

We compare our VAR model to the baseline in terms of two forecasting metrics with differing assumptions: the root mean square error (RMSE) and the continuous ranked probability score (CRPS). RMSE is commonly used, nonparametric, and intuitive, but does not take differing scales of the predicted variable or forecast uncertainty into account. Thus, it may place excessive weight on larger subreddits having greater variation in size. The CRPS accounts for the variance in the data and rewards the forecasts for both accuracy

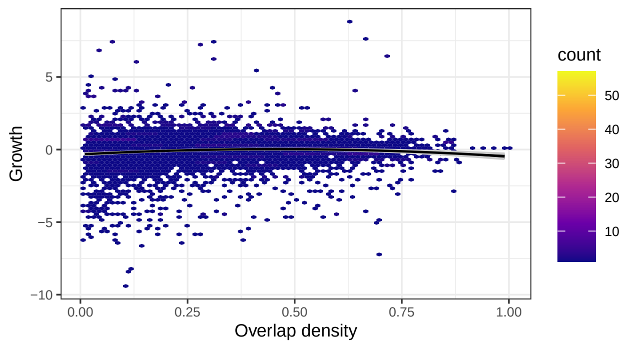


Figure 1: A 2D histogram of subreddits with overlap density (log-transformed) on the X-axis and the change in the logarithm of the number of distinct commenting users on the Y-axis. The black line shows the marginal effect of overlap density on growth as predicted by Model 2. The gray region shows the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

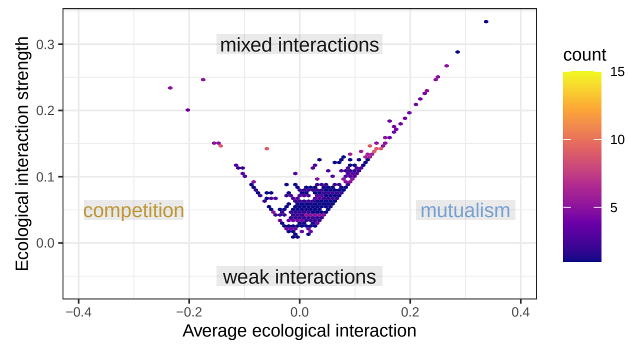


Figure 2: Two-dimensional histogram showing ecological communities on Reddit in our typology. The X-axis shows the overall degree of mutualism or competition in clusters of subreddits with high user overlap based on the average ecological interaction. The Y-axis shows the ecological interaction strength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

and precision and is thus a "proper scoring rule" for evaluating probabilistic forecasts (Gneiting and Raftery 2007). 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.

Results

Study A: Density Dependence Theory

We test density dependence theory as formulated in H1 using Model 1 (Equation 5) which has first- and second-order terms for the effect of overlap density on growth. H1 hypothesizes that overlap density will have a curvilinear ∩-shaped relationship with growth indicated by a negative second-order coefficient.

We observe this predicted relationship between overlap density and growth. Figure 1 plots the marginal effects of overlap density on growth for the median subreddit laid over a scatterplot of the data. The point where increasing density ceases to predict increasing growth and begins to predict decreasing growth is at the 49^{th} 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 20^{th} percentile of overlap density decline by 1.1 members (95% CI:[-1.1,-1.15]) and typical groups at the 80^{th} percentile decline by 1.2 members (95% CI:[-1.1,-1.28]).

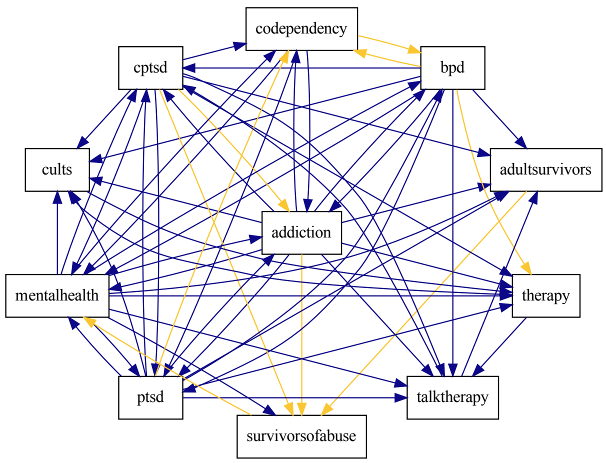
Study B: Introducing Community Ecology

Figure 2 visualizes the distribution of average ecological interaction and ecological interaction strength over the 641 ecological communities we 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 the ecological communities with greater mutualism have greater ecological interaction strength (Spearman's $\rho = 0.58$, $p < 0.001$). Therefore, our community ecology analysis suggests that among groups with similar users, mutualistic ecological interactions are more common than competitive ones.

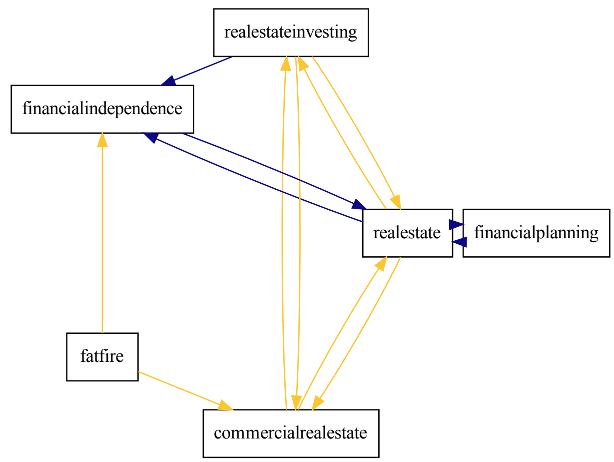
Example ecological communities We present four case studies to illustrate our typology of ecological communities of online groups. Using our measures of average ecological interaction ($\overline{m}$) and ecological interaction strength (κ), we select cases of subreddit clusters characterized by mutualism, competition, a mixture of mutualism and competition, and few ecological relationships at all. 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 that represent statistically significant impulse response functions for all relationships within our four case clusters. For each case, we examined the terms of the vector autoregression parameter $\mathbf{\Phi}$, the impulse response functions, and the model fits and forecasts, all of which are available in our online supplement. We also visited each subreddit in the clusters and read their sidebars and top posts to support our brief qualitative descriptions.

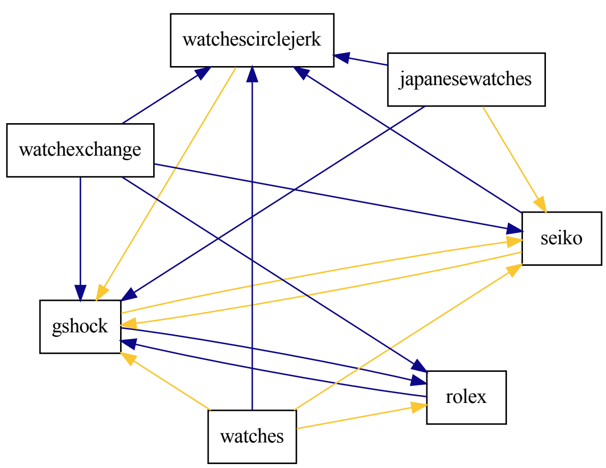
Mutualism among mental health subreddits To find a case characterized by mutualism, we selected the top 37 large clusters with the greatest average ecological interaction. From these, we arbitrarily chose one interesting ecological community, the *mental health* cluster, which includes 11 subreddits 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`



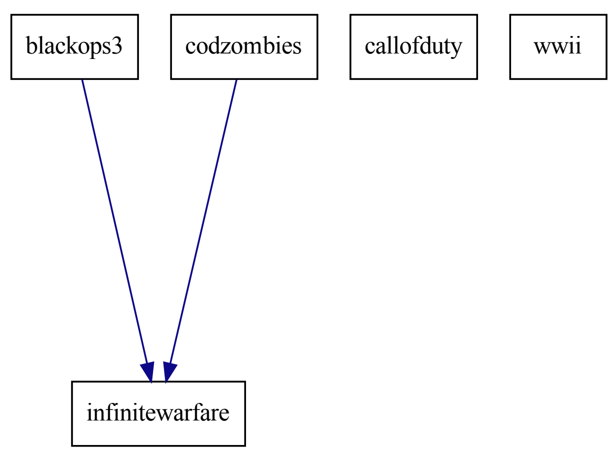
(a) The ecological community of subreddits for supporting mental health and survivors of abuse is dense with largely mutualistic interactions.



(b) The subreddits about real estate and finance are relatively competitive.



(c) Subreddits about watches are dense with both mutualistic and competitive interactions.



(d) The ecological community of subreddits about Call of Duty video games is characterized by relatively sparse ecological interactions.

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

(complex post traumatic stress disorder) while others like r/survivorsofabuse and r/adultsurvivors are support groups.

The interactions among these subreddits are dense and primarily 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 an increase in growth in r/cptsd, even as that growth in r/cptsd predicts a decrease in 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 Figure 3a suggests that different subreddits have complementary functions in this ecological community as people turn to different types of groups for help with 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 demonstrates how groups with related topics and overlapping participants can have mutualistic interactions.

Competition among real estate and finance subreddits To find competitive clusters, we selected an ecological community that we label *finance* from the 36 large clusters with the lowest average ecological interaction having six subreedits. Three of them: r/realestateinvesting, r/realestate and r/commercialrealestate, deal with different aspects of the real estate industry, while r/financialindependence and r/fatfire (the acronym "fire" means "financial independence/retire early") focus on building wealth and becoming financially independent, and r/financialplanning is a general financial advice subreddit.

Unlike the ecological community for mental health, the finance cluster has mostly competitive ties as visualized in Figure 3b. 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 also competitive.

Although this cluster is among the most competitive in our data, it contains mutualistic ties between the general finance subreddits (r/financialplanning and r/financialindependence) and r/realestate. This reflects just how prevalent mutualism is among subreddits with high degrees of user overlap.

Mixed interactions among timepiece subreddits Next, we turn to the *timepiece* ecological community of 7 subreddits about watches that has low average ecological interaction but high ecological interaction strength. We selected the *timepiece* subreddits from the 36 large clusters with average ecological interaction closest to 0 and then from the 15 clusters with the greatest ecological interaction strength.

As shown in Figure 3c, the network of timepiece subreddits is dense with ecological interactions (although not as dense as the mental health subreddits). We observe both reciprocated interactions, like the mutualism between r/rolex and r/gshock or the competition between r/gshock and r/seiko and unreciprocated interactions like the mutualism between r/watchexchange and r/watchcirclejerk[2] or the competition between r/japanesewatches and r/seiko. Although the average ecological interaction among these subreddits is near 0, our analysis reveals a complex ecological community with a mixture of competition and mutualism.

Sparse interactions among Call of Duty subreddits To find a case where ecological interactions are weak, we returned to the group of the 36 large clusters with average ecological interaction closest to 0 but selected 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 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 growth in r/blackops3 or r/codzombies predicts growth in r/infinitewarfare, but 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. Although both clusters are characterized by high degrees of user overlap and low average ecological interaction, the timepiece cluster has a dense competition-mutualism network, while the Call of Duty network is sparse.

Forecasting accuracy We test H2 using two metrics of whether we have improved the 24-week forecast performance for all subreddits which were assigned to clusters: root-mean-square-error (RMSE) and continuous ranked probability score (CRPS). We find that VAR models including ecological interactions have better forecasting performance than the baseline model in terms of both. 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 greater than the CRPS of the VAR models (72,669).

Threats to Validity

Our work is subject to several important threats to validity. First, we study only one platform hosting online groups, and our results may not generalize to other platforms or time periods. The method we propose for identifying ecological interactions between online groups has limitations common to all time series analysis of observational data. 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 conclusions using our method would depend on several untestable

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

assumptions. For example, groups we do not consider—including groups on other platforms—could affect ecological communities in ways unaccounted for in our models. Potential omitted variables may also include additional time lags of group size. We chose to use VAR(1) models with a single time lag for simplicity, but we hope future work will model more complex dynamics with additional lags.

Our vector autoregression models assume that the error terms are trend stationary. This is a common assumption in time series analysis and is difficult to evaluate empirically (Ives et al. 2003). Future work might relax these assumptions using sophisticated models or additional contextual knowledge of 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. While we choose clusters based on high degrees of user overlap and validate our clustering, we might have obtained different results if we had clustered in a different way. Additionally, our efforts to obtain clusters with a high silhouette coefficient led us to remove subreddits from our analysis. Thus, our results are not representative of Reddit in general, but only of those subreddits that were included in our analysis. Furthermore, clustering algorithms may not have unique solutions, and different initial conditions can lead to different results.

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. Ruef 2000; Sørensen 2004). For instance, there are nearly 100 million possible ecological interactions among 10,000 communities. Attempting to infer all of them 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 clustered communities according to user overlap in order to explore typical ecological communities on a platform, but future investigations should consider other quantitative or qualitative approaches to constructing ecological communities.

Discussion

In the final chapter of their book on *Building Successful Online Communities*, Kraut, Resnick, and Kiesler (2012) advise managers of online groups to select an effective niche and beware of competition. However, these recommendations are based on little direct evidence from studies of online groups and offer almost no concrete steps that a designer or group should take based on either piece of advice. Although further research into ecological interactions is needed to derive design principles, we provide a novel framework for online group managers to think about ecological constraints on group size. Intuition suggests that online group managers might seek mutualistic relationships and avoid competitive ones, but it is not clear whether another group is a competitor or mutualist. Our method provides a way for group managers to know.

We presented two studies with the purpose of introducing our community ecology framework and comparing it with previous work using population ecology. In Study A, we found support for H1 by showing—as predicted by density dependence theory—that overlap density has an $\cap$-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 less conducive to growth than medium-density environments.

Surprisingly, this seems to contrast with our results in Study B. When we studied ecological communities using vector autoregression models of group size over time to infer networks of ecological interactions, we found that ecological communities of subreddits are typically mutualistic and that these mutualistic interactions are stronger on average than competitive ones. These findings corroborate recent qualitative studies arguing that multiple online communities about the same topic exist because they provide different and complementary benefits (TeBlunthuis et al. 2022), such as those provided by combinations of small and large subreddits (Hwang and Foote 2021). Moreover, we found support for H2 by showing that the ecological interactions in these models are useful for forecasting group size. This validates our inferences of ecological interactions and provides compelling evidence that ecological interactions are an important factor affecting the development of online communities.

We also found a diversity of ecological dynamics among clusters of subreddits with high degrees of user overlap. We found ecological communities that are mutualistic, competitive, that mix the two or that have few significant ecological interactions at all. This explains the puzzling set of empirical results from previous work on the relationship between overlap density and outcomes like growth, decline and survival (Wang, Butler, and Ren 2012; Zhu, Kraut, and Kittur 2014; Zhu et al. 2014). These studies have measured the density of an online group's niche in terms of its overlap in participants or topics. Although we find support for the relationship between density and growth predicted by density dependence theory, our analysis of ecological communities suggests that degrees of overlap may have little to do with whether two groups are mutualists or competitors.

By correlating an online communities' aggregated resource overlaps to their growth or survival, tests of density dependence theory obscure complex networks of ecological relationships. As a result, interpreting these correlations as evidence that pairs of groups with the greatest resource overlaps are most likely to compete commits an ecological fallacy (Piantadosi, Byar, and Green 1988; Robinson 1950). Our method provides a path for future work to understand the relationship between resource overlaps and ecological interactions by directly inferring when groups are competitors or mutualists instead of relying on aggregations.

That said, we believe that density dependence models can usefully reflect environmental conditions. Density is associated with growth when a platform provides a hospitable environment to build online communities that share certain topics or membership bases. Yet, when conditions change and these topics lose popularity or membership bases migrate off a platform (Fiesler and Dym 2020), density can become associated with decline. For example, the differing environmental conditions of Wikia wikis and Usenet groups might explain why user overlap was associated with the survival of wikis (Zhu, Kraut, and Kittur 2014) but with the decline of Usenet groups (Wang, Butler, and Ren 2012). Wikia was a young and growing platform during Zhu, Kraut, and Kittur's data collection period when the growth of groups may have been limited by knowledge of how organize and build a wiki; perhaps this knowledge was provided by overlapping experienced members. Usenet was in decline during Wang, Butler, and Ren's study period and this may have created competition over increasingly scarce members.

Future work should seek to explain when two online groups will be mutualists or competitors. Long-held understandings of ecological interactions in evolutionary theory suggest that, as we find, mutualism will be more common than competition (Kropotkin 2012). Competition is unlikely to persist because it decreases survival; but mutualistic relationships are likely to endure because they increase it. In this line of theory, groups might avoid competition by adopting specialized functions in their ecological communities, a dynamic known as resource partitioning (Carroll 1985). For example, the competition among the real estate subreddits observed in Figure 3b may occur due to insufficient specialization. By contrast, mental health support groups such as those observed in Figure 3b appear to have specialized purposes or functions.

Online groups may use multiple platforms with distinctive affordances for different purposes (Kiene, Jiang, and Hill 2019). Since our VAR method relies only on time series data to infer ecological interactions, it can be applied to study ecological communities spanning social media platforms. While we focus on relationships between groups sharing a platform, one can apply our concepts and methods to understand how higher levels of social organization emerge from interdependent systems of technologies and users on social media platforms.

Conclusion

An ecological explanation for the success of online groups looks beyond internal mechanisms to understand how different groups influence each other's growth or decline. Prior research has investigated competition and mutualism among online groups with overlapping users and topics using the population ecology framework (Wang, Butler, and Ren 2012; Zhu, Kraut, and Kittur 2014; Zhu et al. 2014), yet has not provided a way to infer competitive or mutualistic interactions among related groups. We introduce the community ecology framework as a complementary perspective to population ecology. By inferring competition-mutualism networks directly from time series data, our community ecology approach helps resolve the empirical tensions raised by prior work and reveals that most interactions within clusters of highly overlapping subreddits are mutualistic. Our methods provide a foundation for future work investigating related online groups.

Ethics Statement

The intended broader impact of this work is to improve the design and management of online communities by advancing scientific understanding of overlapping communities. It is conceivable that this work could contribute to potential harms, such as if it is used to organize socially harmful online communities. We hope and believe that any negative consequences of this work will be outweighed by positive and productive ones. We are sensitive to the ethical concerns about large-scale analysis of publicly available social media communication and behavior, such as ours, in that the individuals whose traces we analyze are unlikely to anticipate their data will be used in such a way. That said, because our analysis aggregates these activities to such a degree that no individual is exposed to scrutiny, we believe that the resulting harms are minimal. We have no competing financial interests in this work.

Code and Data Availability

Supplementary material, as well as the code and data to replicate this analysis, is available via the Harvard Dataverse at https://doi.org/10.7910/DVN/KLGHKY.

Acknowledgments

Text from a draft of this article was included as part of the first authors PhD dissertation and he thanks his committee—Professors Kirsten Foot, Aaron Shaw, David McDonald and Emma Spiro, for their generous support, wise advice and insightful comments. Versions of this paper received very helpful feedback at the International Communication Association's 2021 annual meeting. We are thankful to the Community Data Science Collective for additional feedback and Sohyeon Hwang, Jeremy Foote, Carl Colglazier, and Kaylea Champion in particular.

We owe special gratitude to Daryn McElroy for her work to externally validate our clusters. Thank you to Jason Baumgartner and pushshift.io for the Reddit data archive. Also, we thank to the peer reviewers whose insightful comments improved the quality of this article. Any remaining errors and imperfections are ours. This work was supported by NSF grants IIS-1908850 and IIS-1910202 and GRFP #2016220885 and was facilitated through the use of the advanced computational infrastructure provided by the Hyak supercomputer system at the University of Washington.

References

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

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

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

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14: 830–839.

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

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

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

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

Carroll, G. R.; and Hannan, M. T. 1989. Density Dependence in the Evolution of Populations of Newspaper Organizations. *American Sociological Review*, 54(4): 524.

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

Chandrasekharan, E.; Samory, M.; Jhaver, S.; Charvat, H.; Bruckman, A.; Lampe, C.; Eisenstein, J.; and Gilbert, E. 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): 32:1–32:25.

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

Datta, S.; Phelan, C.; and Adar, E. 2017. Identifying Misaligned Inter-Group Links and Communities. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW): 37:1–37:23.

De Choudhury, M.; and De, S. 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): 71–80.

Ducheneaut, N.; Yee, N.; Nickell, E.; and Moore, R. J. 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, 407–416. New York, NY, USA: ACM. ISBN 978-1-59593-372-0.

Fiesler, C.; and Dym, B. 2020. Moving Across Lands: Online Platform Migration in Fandom Communities. *Proc. ACM Hum.-Comput. Interact*, 4(CSCW1): 042:1–042:25.

Freeman, J. H.; and Audia, P. G. 2006. Community Ecology and the Sociology of Organizations. *Annual Review of Sociology*, 32: 145–169.

Fulk, J.; Flanagin, A. J.; Kalman, M. E.; Monge, P. R.; and Ryan, T. 1996. Connective and Communal Public Goods in Interactive Communication Systems. *Communication Theory*, 6(1): 60–87.

Gneiting, T.; and Raftery, A. E. 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477): 359–378.

Hale, S. A. 2015. Cross-Language Wikipedia Editing of Okinawa, Japan. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 183–192. New York, NY, USA: ACM. ISBN 978-1-4503-3145-6.

Halfaker, A.; Geiger, R. S.; Morgan, J. T.; and Riedl, J. 2013. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist*, 57(5): 664–688.

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

Hwang, S.; and Foote, J. D. 2021. Why Do People Participate in Small Online Communities? *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 462:1–462:25.

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

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

Kiene, C.; Jiang, J. A.; and Hill, B. M. 2019. Technological Frames and User Innovation: Exploring Technological Change in Community Moderation Teams. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 44:1–44:23.

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

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

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

Kropotkin, P. 2012. *Mutual Aid: A Factor of Evolution*. Courier Corporation. ISBN 978-0-486-12153-6.

Margolin, D. B.; Shen, C.; Lee, S.; Weber, M. S.; Fulk, J.; and Monge, P. 2012. Normative Influences on Network Structure in the Evolution of the Children's Rights NGO Network, 1977-2004:. *Communication Research*.

Piantadosi, S.; Byar, D. P.; and Green, S. B. 1988. The Ecological Fallacy. *American Journal of Epidemiology*, 127: 893–904.

Powell, W. W.; White, D. R.; Koput, K. W.; and Owen-Smith, J. 2005. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, 110(4): 1132–1205.

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

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

Romer, P. M. 1990. Endogenous Technological Change. *Journal of Political Economy*, 98(5, Part 2): S71–S102.

Ruef, M. 2000. The Emergence of Organizational Forms: A Community Ecology Approach. *American Journal of Sociology*, 106(3): 658–714.

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

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

Tan, C.; and Lee, L. 2015. All Who Wander: On the Prevalence and Characteristics of Multi-Community Engagement. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, 1056–1066. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-3469-3.

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

TeBlunthuis, N.; Shaw, A.; and Hill, B. M. 2017. Density Dependence without Resource Partitioning: Population Ecology on Change.Org. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17 Companion, 323–326. New York, NY, USA: ACM. ISBN 978-1-4503-4688-7.

TeBlunthuis, N.; Shaw, A.; and Hill, B. M. 2018. Revisiting "The Rise and Decline" in a Population of Peer Production Projects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 355:1–355:7. New York, NY: ACM. ISBN 978-1-4503-5620-6.

Vincent, N.; Johnson, I.; and Hecht, B. 2018. Examining Wikipedia with a Broader Lens: Quantifying the Value of Wikipedia's Relationships with Other Large-Scale Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 566:1–566:13. New York, NY: ACM. ISBN 978-1-4503-5620-6.

Waller, I.; and Anderson, A. 2019. Generalists and Specialists: Using Community Embeddings to Quantify Activity Diversity in Online Platforms. In *The World Wide Web Conference on - WWW '19*, 1954–1964. San Francisco, CA, USA: ACM Press. ISBN 978-1-4503-6674-8.

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

Zhang, J.; Hamilton, W. L.; Danescu-Niculescu-Mizil, C.; Jurafsky, D.; and Leskovec, J. 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: 377–386.

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

Zhu, H.; Kraut, R. E.; and Kittur, A. 2014. The Impact of Membership Overlap on the Survival of Online Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, 281–290. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-2473-1.