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

ANONYMOUS AUTHOR(S)

Research in social computing shows how online groups affect each other's group sizes 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 demonstrate our approach and its benefits for social computing research through analyses of simulated data and of three real networks of groups hosted by Reddit.

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, 17 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 spawn new groups and wage conflicts against each other, compete with, and help each other [19, 56, 67]. 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 [63, 66, 67]. Ecology has been fruitfully applied to human organizations ranging from commercial industries to social movements for over 40 years [5, 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 wikis that groups whose memberships overlap with other groups survived longer [66], but such Usenet groups failed earlier [63]. In some settings, the relationship between topical overlaps and performance follows theoretical predictions that moderate overlap should lead to the best outcomes [66], while studies in other settings find weak or theoretically inconsistent relationships [56, 59].

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 *community ecology* and *population ecology*—have different levels of analysis and make distinct theoretical predictions [2]. Population ecology considers the dynamics

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

within a population such as a biological species. The earliest and most influential works of ecology as applied to human organizations have 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 distinct approach that adopts a higher level of analysis to consider dynamics within an *ecological community* comprised of heterogeneous and interdependent populations [2]. In biology, this might be those organisms inhabiting a lake or valley. In organization science, this might be the network of technology developers, manufacturers, and suppliers in the semiconductor industry [51, 61]. Community ecology is also of considerable influence in organization science [e.g. 2, 4, 45, 47, 51, 53]. 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. [63], Zhu et al. [66], and Zhu et al. [67] are consistently framed in terms *population ecology* theory, their empirical analyses deviate from population ecological analysis in important ways. Although ecological studies in social computing have focused on overlapping resources like topics and participants, theories of population ecology focus on homogenous populations and are thus simply 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 studies of resource overlaps in HCI measure the degree to which a groups's participants or topics overlap with every other group is correlated with growth or survival. Thus, they treat competition and mutualism as environmental forces. While these forces may emerge from relationships between groups, the relationships themselves are obscured through aggregation. In this paper, we describe and demonstrate a novel community ecology approach to studying online groups that emphasizes how different online groups have heterogeneous roles and corresponding resource needs which give rise to competitive or mutualistic relationships between groups. In doing so, we make distinct theoretical and methodological contributions to HCI and social computing scholarship.

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?" In the process, our theoretical work brings clarity to a confusing set of empirical results in prior research. Our second contribution, in §4.2 is to introduce a statistical method for inferring networks of competitive and mutualistic relationships from online group participation activity data using vector autoregression models [13, 37, 54]. We validate our method using simulated data to show that it can identify a full range of ecological relationships and conduct a series of three case studies of groups hosted on the platform Reddit in §6. Although limited, these case studies make a third contribution in the form of empirical findings that suggest that specific patterns of relationships vary substantially across networks of groups and that mutualism appears to be much more common than competition.

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 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 [8, 21, 43]. While many online groups are formed, only an extreme minority develop a sizable group of participants [38, 42]. Understanding why some groups take off 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 [10, 63, 65]. Although this dynamic implies a feedback loop, groups clearly do not grow forever and increasing costs of participation will eventually limit further growth [10]. While factors such as leadership, organizational practices, network structure, and design decisions can lower costs and benefits of participation [10, 42, 60], institutions 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 [28] in a pattern observed in other collaborative knowledge production projects [58].

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 [8, 44, 50, 52]. Interdependence between online groups can be important to explaining growth, decline and survival because important rival resources like the time, attention, and efforts of participants are are 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 [7, 11], 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 [63].

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 [26] which can be non-rival or even "anti-rival" when their usefulness increases as a result of others using them [44, 64]. For example, the usefulness of a communication network increases as more people join it [26, 40]. Similarly, the usefulness of an information good can increase as more people come to know and depend upon it [44, 64]. For example, awareness of an audience can motivate participation in online groups [65]. 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 [67].

Even though online groups can share and compete over resources, studying interdependence between online groups adds enormous complexity to research projects [35]. Despite the challenges, a growing body of empirical research in social computing has sought to quantify how online groups share users or topics [19, 20, 33, 57] and how interactions between groups relate to outcomes like the emergence of new groups [56], contributions to peer produced knowledge [62], and the spread of hate speech [17]. 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 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 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 the growth, decline, or survival of online groups [63, 66, 67]. 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, or knowledge. When density is high, it becomes difficult to avoid competitive relationships.

Models of density dependence theory in population ecology have a similar form 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. In an analagous 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 [56, 67]. An individual online group's growth may be limited by the ability of their social structures to scale to include more members [10] or due to competition with other groups over members [30]. Similarly, a population of online groups may decline if they compete with one another over participants.

While the foundational studies of density dependence in organizational research 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 an individual group faces.

In contrast to biological species, this approach does not assume a population is homogeneous. Instead individuals have different *niches* corresponding to their resource needs [1]. This is plausible in the case of online groups sharing a platform with diverse topics [39], norms [18, 23], and user bases [57]. 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, Wang et al. [63] study user overlaps in Usenet newsgroups and find they are associated with decreasing numbers of participants. Similarly, TeBlunthuis et al. [59] find that topical overlaps between online petitions are negatively associated with participation. By contrast Zhu et al. [67] find that membership overlap is positively associated with increasing survival of new Wikia wikis. Only Zhu et al. [66] find support for the curvilinear relationship predicted by density dependence theory.

[2] Although it is less common in organizational research, overlap density has also been used by some organizational ecologists [e.g. 22].

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	B → A	Relationship type
+	+	Full mutualism
+	·	Partial mutualism
+	−	Predation
−	·	Partial competition
−	−	Full competition

3 A COMMUNITY ECOLOGY APPROACH FOR SOCIAL COMPUTING

We propose a community ecology approach to the study of online groups defined by networks of dyadic relationships between online groups as a focal object of study. Our focus is on making visible the distinctive roles that particular groups play in a broader ecological community. Overlap density has been useful for advancing ecological theory and the empirical analysis of online groups because it provides a way to characterize the environment that an online group faces. As noted in §2.1, empirical studies online groups find that higher levels of of overlap density are associated with decreasing group sizes in some contexts but increasing growth or survival in others. We propose that these puzzling results are due to how overlap density aggregates a group's many relationships into a single property of the group's environment.

One possible reasons for these inconsistent findings is that factors other than topic or user overlaps can lead to competitive and mutualistic forces. Zhu et al.'s [67] Wikia was a growing platform and they found increased survival among new communities with overlapping members from established groups. Perhaps the growth of Wikia wikis 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.' [63] study period and it may not have been limited in this way. In general, competition over overlapping resources will have no affect 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 [61]. Further, resources that limit growth may be unobserved. Our community ecology approach begins by relaxing the assumption that competition and mutualism are driven by observable resource overlaps.

A second 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 [67] context but competitive but in Wang et al.' [63]. Also, differences in communication modalities between discussion groups and Wikis may be associated with different resource needs and thus different potential relationships between user overlaps and competition or mutualism.

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 up to investigation. While varying conceptions of community ecology are found in the literature on organizational ecology [24], our approach follows Aldrich and Ruef [1] and Hawley [31].

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, 55], or networks of relationships between organizations [e.g 45, 51]. We propose a community ecology approach for social computing that, in an analogous way, seeks to theorize the relationships between different online groups. Our approach follows community ecology in biology and organization science by seeking to understand commensal relationships between groups [1]. A commensal relationship is a way groups affect one another through changes in group size. A group may have a commensal influence on another group in three ways. Mutualism is 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. Competition, is where 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 five 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 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. With commensalism, we can seek to explain the puzzling results of resource overlap studies by asking questions like "what kinds of resource overlap between two groups are associated with mutualism or competition?" or "do groups that have more overlap have more (or stronger) mutualistic or competitive relationships?"

Ecological dynamics play out through the network of such relationships over time as represented by a *community matrix*. Analysis of the community matrix can reveal indirect relationships between groups and properties of an ecological community like stability [37]. Finally, 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 and social media platforms [1, 2].

4 METHODS

4.1 Vector Autoregression Models

The goal of our methodology is to infer the community matrix of commensal relationships between groups which is a central analytical object in community ecology in biological and organizational ecology [1, 49, 61]. Because of data and statistical challenges, organizational scientists have rarely attemptted to estimate the full community matrix [e.g. 53, 55]. We take advantage of advances in tools for statistical inference and of granular behavioral trace data from online groups to use vector autoregression (VAR) models, an established approach in biological ecology, to infer ecological networks of commensal relationships. 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 [10] resource dependence framework from social computing [37]. Even in the presence of unmodeled nonlinearities, VAR(1) models can reliably identify competition or mutualism between species in empirically realistic scenarios [16]. They have also been widely adopted in the social sciences, particularly in political science and in macroeconomics [9]. VAR models are flexible enough to model a wide range of systems so long as sufficiently long time-series data are available [54]. 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 [13]. A VAR(1) model has the form:

$$Y_t = B_0 + \mathbf{B}X_t + \Phi Y_{t-1} + \epsilon_t \quad (1)$$

Where Y_t is a vector of of length m. The parameter B_0 is a vector of intercept terms. Φ is a matrix where the diagonal elements are intrinsic growth rates for each element of Y and the off-diagonal elements are the influence between different variables. Additional predictors can be included in X_t, and ϵ_t is the error term.

The parameter Φ in a VAR(1) model encodes the "community matrix" [49]. VAR models also capture "net effects," which include indirect relationships between groups. 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 $\phi_{A,B}$ and $\phi_{B,C}$ and $\phi_{A,C}$ will be nearly zero.

However, this does not mean that groups A and C are independent because an exogenous increase in A will cause a decrease in B and thereby an eventual increase in C. Indirect relationships like these can be inferred by using impulse response functions (IRFs) to interpret a VAR model [9]. In large VAR models, the great number of parameters can mean that few specific elements of Φ will be statistically significant, while the cumulation of many weak relationships might yield additional statistically significant IRFs [13].

4.2 Inferring the community matrix for social computing systems

Most VAR models in macroeconomics and biological ecology are fit using ordinary least squares, an approach which relies on assumptions that will be difficult to sustain in the online group settings of our interest. The number of parameters in the model increases quadratically with the number of variables in the system which can lead to over-fitting, estimation difficulties, and false discoveries arising from the testing of a large number of hypotheses. For models fit using OLS, non-normal errors can lead to bias. A Bayesian vector autoregression approach can overcome such limitations and allows the use of hierarchical priors that pull estimates towards 0 and thereby correct for multiple hypothesis tests [3, 12, 27].

We extend Equation 1 to include a Poisson link function and a hierarchical Bayesian prior structure. We use a Poisson link function in order to model count data for groups with smaller numbers of participants. In our model, Y_t is a count of participants distributed according to Poisson parameter λ_t which has a multivariate normal distribution evolving over time according to a VAR process. The parameter Φ corresponds to the community matrix as described in §4.1. We use X_t to account for seasonality as described in our case studies.

We use impulse response functions to quantify how much a group's size will change in response to a sudden increase in the size of another group represented by Θ_0, which is an identity matrix so our impulses represent a log-unit increase of 1. Our models have a latent VAR process which we transform into a count variable through a Poisson link. To interpret IRFs in terms of the number of members of a group, we transform them using a baseline of the median number of participants in a group over the study period, $(\widetilde{Y}_i)$ and exponentiating to obtain Θ^*_i as shown in Equation 2.

$$\Theta_i = \Theta_{i-1}\Phi, i = 1, 2, \ldots \quad (2)$$

$$\Theta^*_i = e^{\Theta_i + \log(\widetilde{Y}_i)}$$

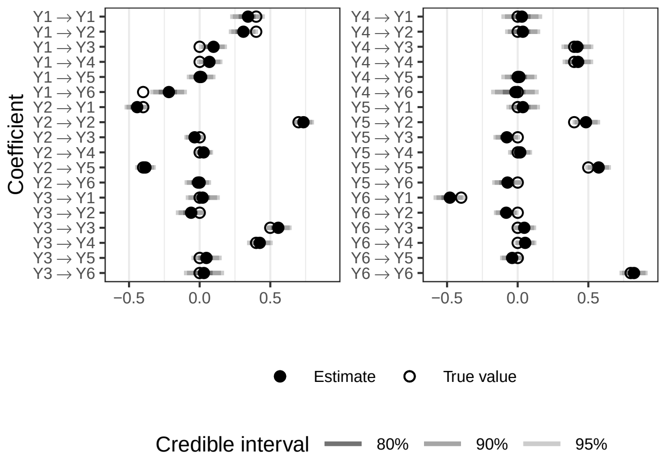


Fig. 1. Results of our simulation study demonstrating that our model can recover true values of Φ with estimated medians and 95% credible intervals (CI). True simulated values are shown as open circles and the median and 95% credible intervals (CI) from our fitted model are solid.

Stationarity is a common assumption in time series analysis [9]. For a VAR(1) model, stationarity means assuming that the eigenvalues of Φ are all less than 1. Practically speaking, this assumes that groups will not grow infinitely nor that the probabilities of activity in them will go to zero. In developing our model, we found that enforcing stationarity through the Heaps prior improved forecast accuracy and helped us fit larger VAR models including more online groups [32]. More information on our models, including our prior specifications is available in Appendix A. We developed our model using the Stan probabilistic programming language building off of code published by Heaps [32] to add support for intercept terms, account for seasonality, and count data [14]. In our simulation and in all our empirical case studies we report results from Stan models fit using 4 chains that pass Stan's diagnostic checks. We have released all code and data necessary to reproduce these analysis in the Harvard Dataverse.

5 SIMULATING AN ECOLOGICAL COMMUNITY

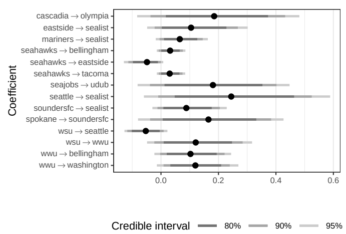
Before turning to our empirical analysis, we first present a brief simulation study to demonstrate that our our model can correctly infer all 6 possible commensal relationships included in Table 1. We simulated an ecological community consisting of 6 hypothetical groups with every possible commensal relationship over 320 weeks, a length of time similar to that available in our empirical data. Results from the simulation are shown in Figure 1. For each element of Φ we show the true simulated value along with the median and 95% credible intervals (CI) from our fitted model. In all but 1 case the true parameter value is within the 95% CI and even in this case the correct sign is inferred. We conclude from our simulation that our model can infer commensal relationships from data on online groups.

6 EMPIRICAL CASE STUDIES: RELATED SUBREDDITS

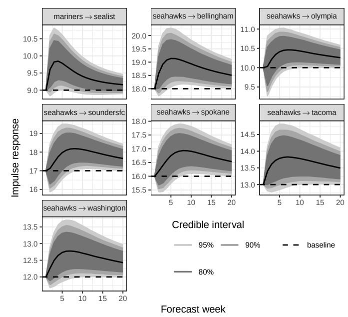
Confident that our approach can infer commensal relationships between groups, we turn to the analysis of three sets of related groups on Reddit. These case studies serve to demonstrate our method on real online groups and to surface empirical findings that motivate future directions for research. In each case study we analyze time series data from a set

Table 2. Summary of data collected for case studies.

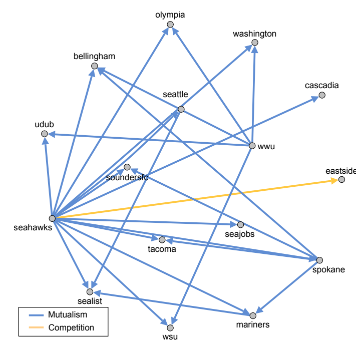
Case study	# Groups	Date Range	Time span
Seattle-area	16	2011-06-06—2015-11-30	235 weeks
Wallpaper	15	2014-11-03—2019-10-28	261 weeks
Design humor	10	2014-09-15—2018-04-02	186 weeks



(a) Estimated VAR(1) model for seattle subreddits. Only coefficients where the 80% CI does not contain 0 are shown.



(b) Impulse response functions for relationships between Seattle-area subreddits. Arrows above each subplot show the direction of the represented relationship. Relationships where the 95% CI does not include the baseline are shown.



(c) Network diagram of commensal relationships among Seattle-area subreddits. Edges represent impulse response functions where the 90% CIs does not include the baseline.

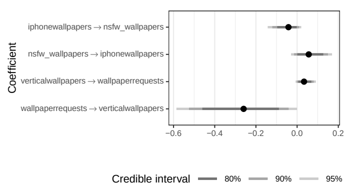
Fig. 2. Results of VAR(1) model for Seattle-area subreddits. Plot (a) shows model coefficients, (b) shows impulse response functions, and (c) shows the network of commensal relationships derived from impulse responses. The community ecology of Seattle subreddits is relatively dense with several mutualistic relationships.

of related online groups: subreddits related to the Seattle area, subreddits for sharing wallpaper (i.e., desktop or mobile phone backgrounds), and design and craft humor subreddits. All three studies draw data from the Pushshift archive of Reddit submissions [6]. Table 2 shows the number of groups and time ranges for each study. In all three case studies our analytic variable is the number of unique posters or accounts making submissions to a given subreddit each week. Compared to alternatives like the number of commenting accounts or the amount of activity in a subreddit, submitting content is a relatively costly form of participation and indicates a relatively high degree of involvement.

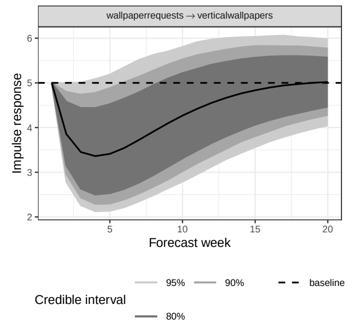
In each case study we report all off-diagonal elements of Φ where the 80% credible interval (CI) does not contain 0, impulse response functions where the 95% (CI) does not include the baseline, and a network of commensal relationships where edges represent impulse response functions where the 90% CI does not include the baseline. Because we are interested not in intrinsic growth, but rather in relationships between groups, we do not show diagonal elements of Φ or IRFs which correspond to loops the network. Plots showing our estimates for all our coeffients and all impulse response functions are available in the online supplement.

6.1 Seattle-area subreddits

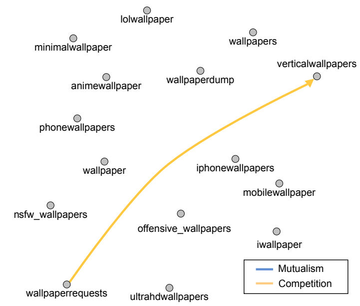
Our first case study looks at subreddits related to Seattle area. We collected a pool of 144 subreddits consisting of all subreddits found in the sidebars and Wikis of the /r/seattle subreddit as of July 20th 2020. In 2016, conflict in /r/seattle led much of the group to migrate to /r/seattlewa. Because this event constitutes a discontinuous shock that will violate



(a) Estimated VAR(1) model for wallpaper subreddits. Relationships where the 80% CI does not include the baseline.



(b) Impulse response functions for relationships between wallpaper subreddits. The only relationships where the 95% CI does not include the baseline is shown.



(c) Network diagram of commensal relationships among wallpaper sharing subreddits. Edges represent impulse response functions where the 90% CI does not contain zero.

Fig. 3. Results of VAR(1) model for wallpaper subreddits. Plot (a) shows model coefficients, and (c) shows the network of commensal relationships derived from impulse responses and (b) shows the impulse response functions where the 95% CI excludes the baseline. The ecological community of wallpaper subreddits is sparse with few commensal relationships between different groups.

assumptions of our model, we restrict our analysis to the period between June 6th 2011 and November 30th 2015, prior to the schism. We examined plots of the time series of the number of unique posters and excluded subreddits without posts in at least 30% of weeks in each year of the study.

We account for sports-related seasonality in */r/seahawks*, */r/soundersfc*, */r/mariners*, */r/udub* and */r/wsu* using a categorical variable with 5 levels: if a week is during a sports teams' (1) regular season, (2) post-season, (3) if the team has a pre-season game, (4) if they have an off-season friendly or it is the post season and they have already been eliminated from contention, and (5) if their sport is out of season. For the university groups we only account for NCAA American football which is the only sport which appears to drive significant activity. We obtained NFL data from the nflgame-redux python package,[3] MLB data from retrosheets.org, MLS data from the engsoccerdata in R[4] and from Wikipedia articles on SoundersFC seasons, and NCAA football from college football reference.[5] The subreddits */r/udub*, */r/wwu* and */r/wsu* also exhibit seasonality with the academic calendar so we include an additional variable indicating if the university is in session according to historical course calendars and registrar websites.

Our VAR analysis of Seattle-area subreddits reveals a large number of mutualistic relationships. Figure 2a shows median values and credible intervals for elements of the inferred community matrix where the 80% CI does not contain 0. While a handful of coefficients are negative at this credibility level, the vast majority are positive. Figure 2b shows several impulse response functions where the 95% CI indicates evidence of mutualism. For example, the top-left subplot shows that the commensalism from */r/Mariners* to */r/sealist* is strong enough that, for the median week, a sudden increase in the number of unique posters by about 8.6 in */r/Mariners* will most likely drive an increase in */r/sealist* of almost 1 unique poster per week.

As shown in Figure 2c, the commensal network of Seattle-area subreddits is dense with mutualistic relationships revealing a tightly connected ecological community characterized by positive spill-overs. From the network we can also

[3]https://pypi.org/project/nflgame-redux/
[4]https://cran.r-project.org/web/packages/engsoccerdata/index.html
[5]obtained from https://gamethread.redditcfb.com/gamedb.php

see how certain groups play outsized roles. The growth of */r/Seahawks* in particular is strongly associated with growth of a number of other subreddits, especially smaller regional subreddits like */r/Bellingham*, */r/Oympia* and */r/Spokane*. Though we do not measure the trajectories of individual accounts, the network suggests people may be drawn to Seattle-area subreddits by way of /r/Seahawks. For example, it may be relatively common for Seahawks fans from Bellingham to come to Reddit to discuss the Seahawks but then find that activity on the Bellingham subreddit is relavent to their interests. Influxes of sports fans from communities like */r/Seahawks* also appear to promote the growth of a wider array of sports subreddits like */r/SoundersFC*.

While the mutualistic relationships associated with sports fan groups are those most confidently inferred by our model, many additional mutualistic relationships are visible in the network in Figure 2c. One intersting example is */r/wwu*, associated with Western Washington University in the town of Bellingham. From Figure 2a, we see that a direct relationship from */r/wwu* to */r/Bellingham* is plausible. The network suggests that when posters become active in */r/wwu*, they may later become posters to */r/Bellingham*, the state-wide */r/Washington* subreddit, and more.

6.2 Wallpaper sharing subreddits

Our second case study looks at a set of subreddits for sharing wallpaper background images. The wallpaper sharing groups are interesting because many seem superficially very similar. For example */r/wallpaper* and */r/wallpapers* are both general purpose desktop wallpaper subreddits and */r/mobilewallpaper* and */r/phoneallpapers* are based around wallpapers for mobile devices. The high degree of similarity between the content shared in these groups suggests that they may compete.

Using Pushshift submissions data, we selected all 898 subreddits including "wallpaper" in their names. As with Seattle area subreddits, we excluded groups with lower levels of activity by requiring posts in at least 40% of weeks in each year of our analysis.[6] We analyze data from between June 6th 2011 and November 30th 2015.

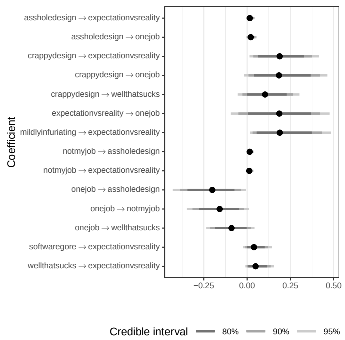
While we found many mutualistic relationships in the case of Seattle-area subreddits, our analysis of wallpaper-sharing subreddits reveals very few commensal relationships at all. Only four out of the 210 non-diagonal elements of Φ shown in Figure 3a have an 80% CI that does not contain 0. In only one case does the 95% CI of the impulse response functions not contain the baseline. Even though the content shared in these groups is similar, they have little influence on each other's membership size over time.

This shows that even among groups with apparently similar content, commensal relationships may not even exist. Even though this case study covered the longest time period and had fewer groups than our study of Seattle-area subreddits—resulting in more statistical power—we found evidence of only one competitive relationship and weak evidence of three other commensal relationships.

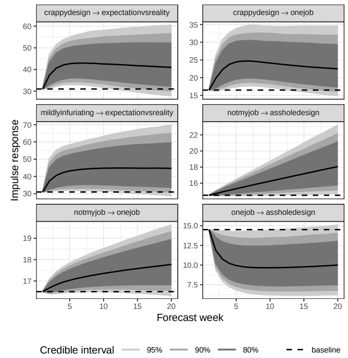
6.3 Design and craft humor subreddits

The final case study is a set of subreddits for humorous sharing of design and crafts. Subreddits typically place boundaries around the kind of content that is appropriate for their groups in rules posted in their sidebars or wikis [18, 23]. Often these rules suggest alternative subreddits for submitting content that is not a good fit for a particular subreddit. We are interested in how such boundaries and links may lead to commensal relationships between groups so we collected design and craft humor subreddits from a thread in */r/assholedesign* discussing a flowchart for deciding where to post

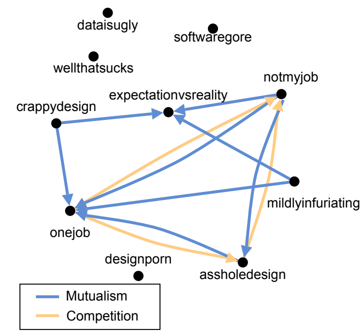
[6] We chose a lower level of activity as our threshold in our study of Seattle because we were interested in the role of some smaller subreddits like */r/seajobs* in the community ecology of Seattle area subreddits.



(a) Estimated VAR(1) model for design and craft humor subreddits. Only coefficients where the 80% CI does not include 0 are shown.



(b) Impulse response functions for design and craft humor subreddits. Relationships where the 95% CI does not include the baseline are shown.



(c) Network diagram of commensal relationships among humorous design and craft humor subreddits. Edges represent IRFs where the 90% CI does not include the baseline.

Fig. 4. Results of VAR(1) model for design and craft humor subreddits. Plot (a) shows model coefficients, (b) shows impulse response functions, and (c) shows the network of commensal relationships derived from impulse responses. This ecological community is characterized by a mix of mutualistic and competitive relationships.

content.[7] We extended our search for related subreddits by reading the wikis and sidebars of */r/assholedesign* and */r/crappydesign*. This search yielded 13 subreddits. We excluded 3 subreddits that were created much later than the rest and analyze the period from September 15th 2014 to April 2th 2018 during which all the remaining subreddits had submissions.

Similar to our analysis of Seattle-area subreddits, our analysis of design and craft humor subreddits reveals a number of commensal ties. However, the portrait of the ecological community that emerges is quite different from both other cases. A shown in Figure 4a, we estimate a number of negative coefficients indicating the existence of competitive relationships. Figure 4b shows impulse response functions for relationships inferred by our model where the 95% CI does not include the baseline. The bottom-right plot in Figure 4b shows how, according to our model a sudden large increase of about 98 unique posters in */r/onejob* can be expected to drive a decrease of about 5 unique posters in */r/assholedesign* in a median week.

Figure 4c shows the network of commensal relationships between the 10 design and craft humor subreddits. As with the Seattle-area subreddits, we find relatively few reciprocal competitive or mutualistic relationships where one group is helped (harmed) by another and the second is helped (harmed) by the first. Interestingly, while the network contains more mutualistic ties than competitive ones, it contains three predatory relationships where one group benefits from an increase in the size of the other, but the other group is hurt by an increase in the first. The subreddit */r/onejob* is a predator of */r/assholedesign* as increases in the size of */r/onejob* predict subsequent decreases in the size of */r/assholedesign* even as increases in the size of */r/assholedesign* predict increases in the size of */r/onejob*. Similarly, */r/onejob* is predator of */r/notmyjob* and also of */r/assholedesign*.

It is interesting that we do not find commensal relationships between the three groups that were most visibly differentiated in guidelines about what kind of content belongs in each group (*/r/assholedesign*, */r/mildlyinfuriating*, and */r/crappydesign*). That we do not find competition between these communities is consistent with the prediction of

[7] https://www.reddit.com/r/assholedesign/comments/a02ezp/meta_is_it_asshole_design_a_handy_flowchart/

ecological theory that partitioning of resources (here topics or styles of content) between groups may reduce competition [1, 15].

7 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 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. Potential omitted variables might also include additional time lags of group size. We chose to use VAR(1) models with only 1 time lag for simplicity but future work can 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.

Like most other time series analysis, our analysis assumes stationarity. Doing so implies that dynamics in the community ecologies of online groups that we study have a single equilibrium. This is difficult to evaluate empirically and may not be realistic [13]. We selected case studies and restricted our analysis to time spans in ways that sought to avoid obvious violations of these assumptions that are common in populations of online groups. Future work might relax these assumptions using more complex models with time-varying VAR parameters or state space models [9]. These types of approaches may also be useful for investigating how commensal relationships change over time. We hope to explore these techniques in future work.

Finally, our three cases studies are limited in that they can offer only a proof-of-concept analysis and an enticing hint at more comprehensive future analyses with more rigorously defined populations of online groups. Although we found varying results in the three ecological communities we selected, these case studies can provide little explanation for when one should expect to find different forms of commensalism in online groups. Our hope is that these initial results can point in new directions for research. As is true in all case study research, there is little reason to expect findings from any one of our case studies to generalize to any specific other set of contexts.

8 DISCUSSION

Prior ecological studies in social computing use overlap in participants or topics to characterize the resource environment faced by particular online groups [63, 66, 67]. However, they yield a puzzling set of empirical results about the relationship between resource overlaps and 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. We set out to infer networks of commensal relationships between groups directly from time series of membership sizes. Applying our method in three case studies reveals three qualitatively different ecological communities. In an ecological community focused on Seattle, we find a great number of mutualistic relationships. Among groups for sharing wallpaper images, we found almost no commensal relationships. In design and craft humor subreddits we find many mutualistic ties, but also predation, and no commensal ties between those groups where visibly posted norms clearly distinguished between content appropriate for sharing in each group.

While our case studies reveal variation in networks of commensal relationships between related groups, case studies alone cannot resolve the puzzle of why resource overlaps appear related to competition in some circumstances [59, 63] but in others appear related to mutualism [66, 67]. However our results provide strong support for the claim that relationships between similar communities might be competitive or mutualistic depending on factors other than content or topical overlaps. We believe that a search for conditions leading to the emergence of different forms of commensalism is a good next step toward resolving the puzzle.

Within a large platform like Reddit there are a great number of sets of related groups that can be studied. This makes it possible to eventually construct and test generalizable theories about when and how different types of ecological communities are constructed. However, interest groups often use platforms with distinctive affordances for different purposes [41]. Since our method relies only on time series data, it can be applied to study ecological communities spanning different social media platforms.

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 [25, 36]. The existence of a sufficient diversity of alternative institutions is likely to lead to competition, but might also make the expression of voice more compelling to moderators [36]. Future studies should investigate such aspects of the dynamics between governance and commensalism.

Finally, we propose that future work investigates higher-level properties of ecological communities. The motivation for community ecology is not only to better describe commensal relationships between online groups, but also to better understand how groups shape one another as well as higher order social structures like social media platforms and technological cultures.

8.1 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 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 related community is a competitor or mutualist. Our method provides a way for group mangers to 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, 15, 43, 51]. For example, the growth of Wikipedia caused other online encyclopedia projects to shift their focus [34]. Our method provides a tool by which 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.

Groups looking to increase activity should desire to seek out mutualistic relationships. 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.

9 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 [28]. Did competition with emerging groups 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 step to resolve empirical puzzles raised by prior work. By narrowing the focus to the dyadic commensal relationships between groups that matter, the community ecology framework we present also raises a host of new directions for social computing researche. 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] Marta Bańbura, Domenico Giannone, and Lucrezia Reichlin. 2010. Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics* 25, 1 (2010), 71–92.

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

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

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

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

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

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

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

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

[12] Fabio Canova. 2007. Bayesian VARs. In *Methods for Applied Macroeconomic Research*. Princeton University Press, 373–417.

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

[14] Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2016. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 20, 2 (2016), 1–37.

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

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

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

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

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

[26] 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.

[27] Andrew Gelman, Jennifer Hill, and Masanao Yajima. 2012. Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* 5, 2 (April 2012), 189–211.

[28] 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.

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

[33] 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

[34] 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.

[35] 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.

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

[37] 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.

[38] 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.

[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] 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.

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

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

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

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

[56] 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.

[57] 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.

[58] 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.

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

[60] 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.

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

[62] 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.

[63] 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.

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

[65] 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.

[66] 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.

[67] 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.

A BAYESIAN POISSON VECTOR AUTOREGRESSION MODEL SPECIFICATION

Below is the full prior specification of our model. Y_t is Poisson distributed with parameter λ_t having a multivariate normal distribution which evolves over time according to a VAR process. More information on our models, including our prior specifications is available in the online supplement. We use a weakly informative hierarchical prior for the elements of B_0.

In developing our model we found that enforcing stationarity through the Heaps prior improved forecast accuracy and helped us fit larger VAR models including more online groups [32] We use the Heaps exchangeable prior over stationary values of Φ , which places separate hierarchical normal priors on the diagonal and off-diagonal elements of an unconstrained matrix that is homeomorphic to Φ. Our choice of hyperparameters follows the recommendations of [32], but we chose a modestly more diffuse hyperprior for the means to increase the sensitivity of our model.

$$
\begin{aligned}
Y_t &\sim \text{Poisson}(e^{\lambda_t}) & \lambda_t &\sim \text{MVN}(\mu_t, \Sigma) \\
\mu_t &= B_0 + \Phi\mu_{t-1} & \Phi &\sim \text{Heaps}(1, \sqrt{3}, 0, \sqrt{3}, 2.1, \frac{1}{3}) \\
\Sigma &\sim IW(m+3, I(m)) & B_0 &\sim \text{N}(\mu_{B_0}, \sigma_{B_0}) \\
\mu_{B_0} &\sim \text{N}(0, 7) & \sigma_{B_0} &\sim \Gamma(4, 3) \\
\mathbf{B} &\sim N(0, 2)
\end{aligned} \tag{3}
$$