

The Effects of Group Size and Time on the Formation of Online Communities: Evidence From Reddit

Social Media + Society
October–December 2018: 1–13
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2056305118815908
journals.sagepub.com/home/sms


Elliot Panek, Connor Hollenbach, Jinjie Yang and Tyler Rhodes

Abstract

These studies examine the influence of group size and the passage of time on two characteristics of online communities: dispersion of participation in group discussions and active member turnover from month to month. We used multilevel analysis to examine the dynamics of user contributions to discussions on Reddit, a popular website that hosts large-group discussions, across 30 groups over 6 years. As groups grow in size, participation becomes more highly concentrated among fewer members while turnover decreases. As time passes, participation becomes more widely dispersed while group member turnover increases. An increase in group member turnover appears to be the result of both a maturation effect (as each group ages, turnover increases) as well as a cohort effect (groups formed at a later date have higher turnover than groups formed earlier). We can conclude from these results that as time progresses and groups become larger, they become less community-like, but in different ways.

Keywords

online community, virtual community, computer-mediated communication, group dynamics, online discussions, Reddit

Introduction

In the present research, we re-visit a question fundamental to the study of online social behavior: What is an online community? Foundational literature on the phenomenon of groups initiating sustained communication online, then referred to as “virtual communities” (e.g., Rheingold, 2000; Wellman & Gulia, 1999), predates the rise of “*Web 2.0*,” a term used to denote popular websites comprised of user-generated content and commentary (e.g., YouTube, Twitter, Reddit). Such websites host group interaction on a scale so massive (e.g., tens of millions of users) that it is not obvious whether users of each website should collectively be referred to as a single community, a collection of communities, or something else entirely. In light of these significant changes to the online landscape, scholarly definitions of “online community” grounded in this foundational literature require revisiting.

More recently, technology designers and companies have become increasingly apt to use the word “community” in descriptions of social media applications (Facebook1, 2018; Instagram, 2018; Reddit, 2018). The highest-profile shift toward the explicitly stated goal of online community-building came from Mark Zuckerberg, founder and CEO of Facebook, who, in response to criticism that Facebook and

other social media were enabling anti-social behavior, released a statement titled “Building Global Community” that invoked the word “community” 81 times (Facebook2, 2017). Although the word is increasingly central to the public-facing rhetoric of social media companies, it remains a goal without a clear definition.

We argue massively popular websites have produced three observable group dynamics, not all of which possess characteristics commonly associated with communities: stability and widespread active participation. In one dynamic, which we refer to as the “community dynamic,” many users actively contribute to discourse over a sustained period of time. In another dynamic, which we refer to as the “creator/audience dynamic,” a small number of users create most of the content for a far larger group of users to consume. In a third dynamic, which we refer to as the “crowd dynamic,” many users contribute for a relatively brief period of time before going their separate ways.

The University of Alabama, USA

Corresponding Author:

Elliot Panek, The University of Alabama, 490 Reese Phifer, Tuscaloosa, AL 35487, USA.

Email: epanek@ua.edu



Our goal in the present research is to re-visit some basic questions about online community in an era of mass online social interactivity: What defines online community? Under what conditions are communities likely to develop and persist online? We answer these questions by applying new methods of analyzing group discussion dynamics on Reddit, a website on which millions of users engage in conversation with one another. In our analyses of interaction patterns on Reddit over the past decade, we use an original means of assessing the “community-ness” of a group. In our first study, we examine the effects of time and group size—two factors previously found to influence group dynamics online—on groups’ abilities to achieve and maintain stability and widely dispersed discourse participation. In a follow-up study, we supplement a “sequential order” operationalization of time with one that takes into account the creation date of the group, allowing us to discern a maturation effect of time from a cohort effect of time.

Defining “Community”

We acknowledge the contentious history of attempts to define and measure the community-ness of online groups. Authors of some of the earliest scholarly work on online communities (e.g., Baym, 1998; Rheingold, 2000; Wellman & Gulia, 1999) were intent on establishing the plausibility of the existence of community online, often comparing online groups to offline counterparts (e.g., a corner bar, or a village) rather than exploring the differences among online groups. While some research included nearly all online group communication under the heading of “community” (e.g., Preece, 2000), other research (e.g., Porter, 2004) sought to establish non-hierarchical typologies of online groups that used observable characteristics to sort online groups into categories (e.g., asynchronous vs. synchronous). Preece (2001) offered a wide variety of objective measures by which the community-ness of an online group may be judged, including the number of active participants, the degree of user satisfaction, and the amounts of trust users have in one another. This was intended to provide future researchers with a “starting point” (Preece, 2001, p. 350) for establishing objective metrics by which they might judge the community-ness of a group online.

Subsequently, the majority of researchers analyzing online group behavior as an objectively measurable phenomenon (e.g., Brown, Broderick, & Lee, 2007) focused on measurements of the network structure of groups, eschewing Preece’s proposed metrics for determining the community-ness of a group. Indicative of this turn to social network analysis, Porter, Onnela, and Mucha (2009) define “community” as “one mesoscopic structure [that] consists of a group of nodes that are relatively densely connected to each other but sparsely connected to other dense groups in the network” (p. 1083). Such definitions and methods, which focus on exchanges among individuals (represented as nodes), are

ideally suited to analyzing online groups in which online group interaction is primarily person-to-person (e.g., employees communicating via email). Network analysis can also be used to identify frequent themes in group discourse (e.g., Park, Conway, & Chen, 2018) or to identify groups that share contributors with one another (e.g., Olson & Neal, 2015). These methods of mapping the structures of groups, although highly useful in identifying types of contributors as well as connections or commonalities among contributors, are not necessarily the only means by which online communities can be defined and measured.

For the purposes of defining community in terms of observable behavior on websites hosting discourse in which individuals do not directly communicate with other individuals, but instead contribute to a discourse that is available to any and all readers, we propose the following definition: a group in which *most members contribute equally* to a discourse (i.e., participation is widely *dispersed* among group members) and one in which *active participation is sustained* for a prolonged period of time. This definition is consistent with several previous definitions of community, such as that of Ridings, Gefen, and Arinze (2002) who, in their description of virtual communities, note that participants are expected to “communicate regularly and for some duration” (p. 273). The frequent recurrence of the words “interact” and “interaction” in definitions of virtual or online communities (e.g., Leimeister, Sidiras, & Krcmar, 2006) implies that groups must be comprised of individuals who *actively contribute* to discourse to be considered communities. Other researchers (e.g., Preece, 2001; Preece & Maloney-Krichmar, 2005) suggest using metrics of sociability similar to the one implemented in our study.

We acknowledge that individuals who do not contribute to a group’s discourse may still feel connected to one another based on shared passions or allegiances, and these feelings are linked to individuals’ senses of identity as members of a group. This kind of “passive belonging” is arguably no less important, to individuals or societies, than widespread, sustained participation in discourse. We only claim that community, as we have defined it, and passive belonging are distinct social phenomena and that determining when, where, and why community, as we have defined it, exists online can help scholars, practitioners, and users better understand group behavior online.

Our definition of community implies the existence of non-community groups in which participation in discourse is unequally distributed and/or group members participate for short periods of time. The former condition can be found on many websites that are primarily comprised of user-generated content, such as Yahoo! Answers (Shah, Oh, & Oh, 2008), Wikipedia (Ortega, Gonzalez-Barahona, & Robles, 2008; Peddibhotla & Subramani, 2007), and Digg.com (Lerman, 2007). It is also possible to conceive of instances in which individuals contribute equally to group discourse but do not do so for very long. Although the persistence of a high

number of active contributors may give the outward impression of stability, high turnover in membership within that group calls into question the extent to which members' experiences resemble those of individuals belonging to communities. Scholars seeking to define the characteristics of successful online communities note that groups must not only attract members to be successful but also retain them (e.g., Butler, 2001). Online, low turnover permits the development of the shared rules and norms commonly associated with successful communities (Lazar & Preece, 2002). From a resource-based perspective, a group with high turnover can be "sustainable" in the sense that it maintains a steady flow of resources in the form of actively contributing members. However, as individual members fail to stay engaged in the group for longer periods of time, it is highly likely that some of the benefits for individual members that take time to develop, such as the establishment of interpersonal trust and feelings of affiliation and companionship, would fail to accrue.

We posit that there are no clear borders separating groups into community, creator/audience, and crowd dynamics. A group does not suddenly become a community when interaction reaches a particular threshold nor when its membership remains consistent over a particular duration. Groups are rarely, if ever, "pure" in their embodiments of each of these aforementioned types. We submit that a group exists on continua, either closer to or further from the platonic ideals of each of these categories.

Factors Influencing Dispersed, Sustained Participation in Online Discourse

Assuming the existence of these online group dynamics, what might explain their variety? Many factors, including design elements (Chen, Harper, Konstan, & Li, 2010; Farzan, Dabbish, Kraut, & Postmes, 2011), qualities of discourse (Joyce & Kraut, 2006), and characteristics of the group members (Muller, 2012) likely influence whether a group becomes more crowd-like, audience-like, or community-like. So as to contribute to understanding of online group dynamics in as wide a variety of contexts as possible, we focus on two factors that pertain to all groups online: group size and time.

Group Size

Online, various attributes or tools are used to overcome the logistical challenges of large-group communication. Many discussion boards and comments sections allow users to reply to one another, and allow users to elect to be notified when someone replies. Posts on some discussion boards and comments sections are often arranged in a nested or branching structure, making it easy for small conversational

exchanges to occur within a larger group. Both of these attributes facilitate chance encounters among large numbers of interlocutors while allowing for meaningful interaction. In the absence of such tools, online group communication can often resemble chronologically ordered lists of statements relating to a particular topic rather than meaningful interactions among group members.

Even when the aforementioned tools for managing large-group communication are present, online groups often fail to sustain the aforementioned characteristics of community. Research on participation patterns on Usenet (Whittaker, Terveen, Hill, & Cherny, 2003) and Digg (Lerman, 2007) provides evidence of a lack of dispersed participation on websites or applications that host large groups. Several explanations of this phenomenon have been posited. Jones and Rafaeli (1999) reason that as online groups grow larger, the burden on each individual to contribute lessens. Hence, individuals are less motivated to participate frequently, if at all, as they may reap the benefits of group membership (e.g., reading insightful or entertaining posts) without exerting the effort required to generate quality content (a problem known as "social loafing").

Group size can also affect contributor turnover. Using a resource-based model of online group dynamics, Butler (2001) notes ways in which increases in the size of online communities can affect the sustainability of online groups. On one hand, additional group members increase potential pools of knowledge, entertainment, or support that make online communities alluring, thereby increasing their value to members. On the other hand, the logistical difficulty of large-group communication and increased likelihood of social loafing present impediments to the realization of such benefits for individual members, thereby increasing the likelihood that members won't participate in the future. The result is a community that attracts more new members than smaller communities but loses contributing members at a faster rate than would a smaller community (Butler, 2001). In the aggregate, the community is sustained and its total active membership grows, although it is comprised of an increasingly transient population.

Based on these theories, we propose the following hypotheses:

H1: Group size will be positively related to concentration of participation.

H2: Group size will be positively related to group member turnover.

Time

Another factor affecting online group dynamics is time. Both online groups (De Souza & Preece, 2004) and online group members (Kim, 2011) are said to pass through stages of development, that is, to change in predictable ways over time. To explain temporal changes in individual users'

behavior such as the progression from technology adoption to continued use (e.g., Karahanna, Straub, & Chervany, 1999; Kim, 2011), scholars have applied a range of theories such as the technology acceptance model (Hong, Thong, & Tam, 2006; Kim, 2011) and the theory of reasoned action (Karahanna et al., 1999). Temporal changes in online group dynamics have been documented in exploratory research (Schoberth, Heinzl, & Preece, 2006), but reasons for these changes remain undertheorized (Schoberth et al., 2006; Wang & Clay, 2012). We offer the following explanations, grounded in theories of motivation, for temporal changes in online group dynamics.

The ease with which individuals may join online groups is accompanied by an ease with which they may abandon them, a characteristic that could contribute to high turnover (Resnick & Kraut, 2011). Many online communities are free to join, reducing the incentive to stay merely to get one's money's worth. In addition, many online communities are anonymous or pseudonymous, so individuals considering leaving need not worry about lasting damage to their reputations upon joining or leaving a group.

Difficulty in sustaining group membership online is especially likely to be observed if we were to restrict our definition of "membership" to *active* membership, that is, those actively contributing to discourse rather than those passively belonging to online groups. While some consider contributing to online group discourse to be a social activity and/or an act of self-expression typically undertaken without expectations of remuneration, others consider it to be effortful (Kankanhalli, Tan, & Wei, 2005), more analogous to a form of labor (Ritzer & Jurgenson, 2010). As such, the effort necessary to sustain a steady flow of contributions to discourse, like the effort involved in other forms of labor, is subject to fatigue. In such cases, failure to sustain a steady level of contribution from group members may be a consequence not of the community failing to meet the needs of the contributor (Butler, Bateman, Gray, & Diamant, 2014) but rather a result of fatigue.

While these conditions make it difficult to retain group members, it does not necessarily follow that online groups will decline in size over time. Groups may maintain or even increase their size if they are able to replace outgoing members with new ones. However, members who join a group early on in its history ("pioneers") and those who enter the group later ("bandwagoners") may differ in ways that have important consequences for the community-ness of a group. In his examination of the diffusion of innovation, Rogers (2003) asserts that early adopters of innovative technologies are more inclined toward social participation than subsequent adopters (p. 258). Research on early adopters of the Internet bore out this assumption, as they were found to be more motivated to use the Internet for social interaction than later adopters (Stafford, 2003). It follows that online group pioneers may be motivated to spend the time and effort required to regularly engage in social participation by contributing to group discourse (Butler et al., 2014) while

bandwagoners may be less motivated in this respect, and thus less apt to exert this effort.

Due to the ease with which members may leave online groups, the effort needed to maintain a steady rate of contribution, and the likely disparity in motivation levels and types between pioneers and bandwagoners, it is hypothesized that turnover will increase over time.

H3: Group member turnover will increase over time.

High levels of motivation may manifest themselves not only in persistence but also in the frequency with which group members contribute to discourse. Those who are highly motivated to maintain a certain level of contribution to the group over time may also be the group members who post more frequently. Put another way, frequent contribution and future contribution may both be consequences of high levels of motivation.

This line of reasoning is supported by Wang and Clay's (2012) theoretical model of longitudinal contribution to online groups. The model, which uses Ryan and Deci's (2000) self-determination theory to elucidate the relationship between motivations and behaviors, depicts contribution to group discourse as part of a self-reinforcing cycle that, like any cycle, develops over time. According to the model, those who initially contribute more frequently to discourse obtain more of a reward in the form of feelings of relatedness, autonomy, and competence. Those feelings drive users' motivation to contribute to discourse in the future. Those who initially contribute less frequently to discourse do not feel the same degrees of relatedness, autonomy, and competence that provide the motivation that is necessary to sustain a steady level of contributions. Over time, a small, intrinsically motivated core group of frequent contributors develops and dominates discourse (Jackson, Yates, & Orlikowski, 2007; Peddibhotla & Subramani, 2007). This theoretical model leads us to expect an increase in the concentration of participation among the few as time progresses. A longitudinal analysis of participation in an online community hosted by a financial services provider is consistent with this expectation: as time progressed, a small group of participants accounted for an increasing proportion of the messages posted, even when controlling for group size (Schoberth et al., 2006).

Based on the aforementioned reasoning and extant research, we propose the following hypothesis:

H4: Concentration of participation will increase over time.

Study I

Method

Data. For the purposes of analyzing the relationships among time, group size, dispersion of participation, and turnover,

we obtained access to comments¹ posted between the years 2008 and 2017 to Reddit, a social news website that allows for asynchronous discussions among pseudonymous users on a variety of topics. This data set was chosen in part to allow us to observe temporal effects (e.g., dropout) that might not be observable in data sets covering shorter periods of time (e.g., the data set used in Schoberth et al.'s, 2006, study). The groups in the data set also vary in size, ranging from 5 to 264,268 unique monthly contributors. This variation allows us to observe effects of group size that might not be observable among smaller groups. The data set was also chosen for purposes of maximizing ecological validity. At present, Reddit is the sixth most popular website in the world (www.alexa.com). A significant portion of all online group interaction happens on Reddit in groups similar to the ones observed in this study.

Discussions on Reddit are initiated by registered users who post links to content on other websites, or statements or queries directed to the Reddit userbase (e.g., “What’s a short, clean joke that gets a laugh every time?”). These initial posts are organized in topic-based lists known as “subreddits” that are hosted on discrete webpages on the Reddit website. Although users may contribute comments to multiple subreddits, results of a study of posting behavior on Reddit (Buntain & Golbeck, 2014) suggested that users rarely significantly contributed to more than one community. In addition, subreddits exhibit a wide variety of growth patterns over time (Panek, Hollenbach, Yang, & Rhodes, 2017). This suggests that while subreddits are not mutually exclusive in terms of users, they could be treated as discrete online groups in an analysis.

Comments were entered into data files and made available in a publicly accessible database by a Reddit user (u/fhoffa) via Google BigQuery, an online tool that allows researchers to analyze large amounts of data. Included in the files were information specifying various characteristics of comments, such as the usernames of individuals who posted comments, the time at which comments are posted, and the subreddits to which they are posted. We used “comments posted in a given subreddit during a given month” as our unit of analysis. The month timeframe was chosen in part to facilitate efficient analysis (comment files in the database were divided into monthly increments, making it simplest to make comparisons across months rather than hours, days, or weeks) while allowing for sufficient variance in the “time” variable to observe change over time. By providing BigQuery with a series of commands or “queries,” we were able to obtain the frequency with which particular users posted in a given subreddit during a given month and the total number of users who commented in a given subreddit during a given month.

The structure of these data (months nested within subreddits) presents us with a distinct advantage in terms of the generalizability of our findings. Rather than conduct a case study of behavior within a single online group, as many

studies of online group dynamics have done (e.g., Schoberth et al., 2006), we can use multilevel modeling to analyze the relationships among our variables of interest across various online groups and determine whether those relationships remain consistent across groups.

As of May 2017, there are over 1 million subreddits containing over 1 billion comments. Given that the database querying processes is time- and resource intensive, we selected a sample of subreddits to analyze. The sample was drawn from the most popular subreddits as of May 2017 (www.redditlist.com). We chose popular subreddits to maximize variance in group size and to maximize external validity. As the purpose of the study is to observe change over time, we excluded subreddits that originated fewer than 6 years prior to the analysis, leaving us with a sample of 2,963 months from 30 subreddits over 9 years. A complete list of subreddits used in the analysis is included in the appendix.

Measures

Group Size. To determine the size of the group, we counted the number of unique commenters (UCs) during each month in each subreddit ($M=48,708.13$; $SD=87,336.88$).

Time (Order). We assigned each month a number based on its sequential order in the set of months collected for a given subreddit (e.g., comments from the r/news subreddit for January 2008 = 1, comments from the r/news subreddit for February 2008 = 2, etc.). For purposes of specificity, we subsequently refer to this variable as “order.”

Dispersion of Participation. Various methods have been used to measure dispersion, including using the Herfindahl-Hirschman Index (HHI) as a metric of dispersion and fitting curves to distributions (e.g., Peddibhotla & Subramani, 2007). We elected to use the Gini coefficient metric, a means of conveying dispersion that is popular among economists (e.g., Piketty, 2014). To calculate the coefficient, we divided the average difference in the number of comments posted by each possible pair of users who contributed to a given subreddit during a given month by the number of UCs for that month in that subreddit.

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n \sum_{i=1}^n x_i}$$

where x_i = number of comments by UC in a given month in a given subreddit and n = number of UCs in a given month in a given subreddit.

This formula yields a number ranging between 0 (perfectly even dispersion in which each UC accounts for an

Table 1. Time (Order) and Group Size Predicting Gini: Estimates of Fixed Effects.

	Estimate	SE	df	t	Confidence intervals	
					Lower bound	Upper bound
Intercept	.19	.13	29.04	1.47	-.08	.47
Time (order)	-.42	.08	28.80	-5.48**	-.58	-.26
Unique commenters (log transformed)	.68	.12	28.59	5.87**	.44	.92

All variables were standardized prior to analysis to facilitate comparison of estimates.

** $p < .001$.

equal proportion of discourse within a given subreddit during a given month) and 1 (perfectly uneven dispersion in which a single UC accounts for all comments in a given subreddit during a given month) ($M = .51$; $SD = .11$)

Turnover. Turnover was determined by comparing a list of UCs who contributed to discourse in a given subreddit during a given month to a list of UCs who contributed to discourse in that same subreddit during the subsequent month. Our turnover metric is an expression of the percentage of UCs that are not retained from one month to the next in a given subreddit.

$$1 - \frac{|\{UC_i \cap UC_{i+1}\}|}{|\{UC_i\}|}$$

where UC_i = UCs in month i and UC_{i+1} = UCs in the month after month i .

If all UCs who contributed to discourse in the first month also contributed in the second month, turnover was equal to 0. If no UCs who contributed to discourse in the first month also contributed in the second month, turnover was equal to 1 ($M = .65$; $SD = .13$).

Analysis and Results

If we are to assume that subreddits are discrete online groups, they must be treated as discrete groups in the analysis, accounting for possible inter-group differences and intra-group similarities in variables of interest and in the relationships among these variables. Although we did not enter into this study with a priori assumptions regarding how particular qualities of subreddits affect relationships among our variables of interest, this does not change the fact that the structure of the data is nested. It is possible that an as-yet-untheorized quality of subreddits changes the nature of the relationship among our variables of interest. By using multilevel analysis and treating subreddits as a random effect, we account for these differences and similarities. Multilevel analysis does not assume relationships between variables are consistent across groups, and quantifies the extent to which units of analyses within groups are related to one another relative to the extent to which they are related to

units outside of their groups (i.e., the Intraclass Correlation Coefficient, or ICC). We used SPSS statistical software to conduct the analysis.² To facilitate comparison of the relative contribution of our predictor variables to variance in our outcome variables, we standardized all of the variables used in our analyses ($M_s = 0$; $SD_s = 1$). Given the highly skewed (skewness: 3.68) and kurtotic (kurtosis = 17.06) distribution of UC within the sample, we used a log transformed version of this variable in the analysis.

Before proceeding with the analysis, we calculated an ICC for each of the two outcome variables: Gini coefficient and turnover. Doing so determines the amount of variance in each variable that is explained by the grouping structure of the data. In this case, data are nested in subreddits, so we consider subreddits to be the groups. The ICC for Gini coefficient was .69, indicating that 69% of the variance in Gini coefficient was due to the grouping structure of the data. The ICC for turnover was .53, indicating that 53% of the variance in turnover is accounted for by the grouping structure of the data. We interpret these coefficients as evidence that multi-level modeling is an appropriate analytic strategy.

To test the relative contribution of order and group size to variance in Gini coefficient, we entered two predictor variables—order and log transformed UC—into a linear regression equation with Gini coefficient as the outcome variable. Results (presented in Table 1) indicated that order is negatively related to Gini coefficient (Estimate = -.42; $t = -5.48$; $p < .001$) while log transformed UC is positively related to Gini coefficient (Estimate = .68; $t = 5.87$; $p < .001$). That is, as time progresses, dispersion of participation increased, and as groups grew larger, dispersion of participation decreased.

To test the relative contribution of order and group size to variance in turnover, we entered the aforementioned two predictor variables into a linear regression equation with turnover as the outcome variable. Results (presented in Table 2) indicated that order is positively related to turnover (Estimate = .45; $t = 5.05$; $p < .001$). Log transformed UC is not related to turnover (Estimate = .07; $t = .63$; $p = .48$).

Discussion

The results indicate that over time, groups on Reddit possess greater dispersion of participation while also possessing higher turnover. In other words, groups become more

Table 2. Time (Order) and Group Size Predicting Turnover: Estimates of Fixed Effects.

	Estimate	SE	df	t	Confidence intervals	
					Lower bound	Upper bound
Intercept	-.12	.13	28.99	-0.90	-.39	.15
Time (order)	.45	.09	27.48	5.02**	.27	.64
Unique commenters (log transformed)	.07	.12	28.02	0.63	-.17	.31

All variables were standardized prior to analysis to facilitate comparison of estimates.

** $p < .001$.

crowd-like over time. Conversely, groups tend to become more audience-like as they grow larger: as the number of contributors increases, so too does the concentration of participation among the few when controlling for the passage of time. The negative relationship between order and concentration of participation among fewer individuals is inconsistent with Schoberth and colleagues' (2006) observations of online group behavior on a website hosted by a financial services company. This inconsistency may be attributable to differences between the groups observed in this study and the one examined in Schoberth and colleagues' study, including differences in topics of discussion. Conversations about financial services may, over time, develop a dynamic in which novices defer to experts while the exchanges of jokes or life experiences that comprise much of the discourse on Reddit may not.

Due to the way time is operationalized in this study (as the order in a sequence of monthly data gathered from a particular subreddit), several important questions about the precise nature of the relationships between time and Gini, and time and turnover, remain unresolved. It is possible that the observed relationships are indicative of a maturation effect—as groups age, they become less community-like and more crowd-like. However, it is also possible that the observed effects are due to some exogenous factor such as a shift in online commenting behavior among all Internet users, that is, a cohort effect.

To resolve this ambiguity, we took advantage of existing temporal variation in our sample and conducted a post hoc analysis. The subreddits in our sample varied in terms of when they were created. The oldest subreddits in our sample, *r/science*, *r/AskReddit*, *r/philosophy*, *r/pics*, *r/politics*, and *r/sports*, included data from January 2008, while data from the newest subreddit in our sample, *r/personalfinance*, commenced October 2010. This variation allowed us to calculate a variable—"creation date"—the effect of which we may compare to the effect of the previously calculated time variable ("order").

Researchers advocate centering group-level variables on the grand mean (the mean value for the entire sample) in multilevel analysis (Enders & Tofighi, 2007). To calculate a grand-mean-centered creation date variable, we first calculated the overall midpoint date for all data in the dataset (the grand mean date). We then calculated the difference between this grand mean date and the midpoint date for each subreddit. Older subreddits had lower values on this variable (e.g.,

r/science = -133 days); newer subreddits had higher values on this variable (e.g., *r/personalfinance* = 354 days). We then calculated a standardized version of this variable and used it in the analysis.

If creation date accounted for more variance in the outcome variables than order, this would suggest that an exogenous factor or factors is/are responsible for changes in the levels of outcome variables. If order accounted for a greater amount of variance in our outcome variables than creation date, this would suggest that group maturation, independent of exogenous factors, is responsible for changes in levels of our outcome variables.

To test these premises, we conducted post hoc tests of two multilevel models in which log transformed UC, order, and creation date were predictor variables and, in the first model, Gini coefficient was the outcome variable and, in the second model, turnover was the outcome variable. Results of the first model provided evidence that order was a better predictor of Gini coefficients (Estimate = -.41; $t = -5.56$; $p < .001$) than creation date (Estimate = -.24; $t = -2.02$; $p = .05$). Results from the second model provide evidence that both order and creation date significantly contribute to variance in turnover. As predicted, turnover increases in subreddits over time, independent of creation date (Estimate = .45; $t = 5.29$; $p < .001$). There was also an effect of creation date, such that the more recently the subreddit was created, the higher the level of turnover (Estimate = .41; $t = 3.94$; $p < .01$).

While this sheds some light on the contribution of order and creation date to changes in Gini coefficient and turnover, the sample used for this study is not ideal for the purposes of comparing these effects. The sample was not created with this analysis in mind, and thus overrepresented subreddits with an older creation date. To address this shortcoming, we performed a follow-up analysis using a stratified sample of subreddits.

Study 2

Method

Data. As in Study 1, we analyzed comments made available for analysis by Reddit user *u/fhoffa* on Google Query. To facilitate comparison across creation date, a stratified sampling strategy was implemented in which subreddits were

Table 3. Order, Creation Date, and Group Size Predicting Gini: Estimates of Fixed Effects.

	Estimate	SE	df	t	Confidence intervals	
					Lower bound	Upper bound
Level 1						
Intercept	.12	.16	25.42	0.74	−.21	.45
Order	−.27	.06	26.78	−4.35**	−.39	−.14
Unique commenters (log transformed)	.62	.10	28.29	6.16**	.41	.18
Level 2						
Creation date	−.13	.15	27.52	−.87	−.45	.18

All variables were standardized prior to analysis to facilitate comparison of estimates.

** $p < .001$.

Table 4. Order, Creation Date, and Group Size Predicting Turnover: Estimates of Fixed Effects.

	Estimate	SE	df	t	Confidence intervals	
					Lower bound	Upper bound
Level 1						
Intercept	−.16	.13	27.20	−1.23	−.44	0.11
Order	.30	.07	25.01	4.41**	.16	0.44
Unique commenters (log transformed)	−.24	.11	27.39	−2.19*	−.48	−0.02
Level 2						
Creation date	.81	.11	26.59	7.27**	.58	1.04

All variables were standardized prior to analysis to facilitate comparison of estimates.

** $p < .001$; * $p < .05$.

selected for analysis based on their creation dates. We sampled five popular subreddits created in each of the following years: 2008, 2009, 2010, 2011, 2012, and 2013. Our Study 1 sample was primarily drawn from subreddits that were created in 2008 and 2009. From the Study 1 data set, we randomly selected five subreddits created in 2008 and five subreddits created in 2009. To fill out the rest of our Study 2 sample, we returned to a list of popular subreddits and selected the five most popular subreddits that were created in 2010, 2011, 2012, and 2013. We chose to analyze the first 50 months of comments from each of these 30 subreddits ($N=1,500$ months). This choice was made because the most recently created subreddits in the sample (those created in 2013) have existed for roughly 50 months, and because we wanted to maintain equal sample size among the subreddits in our sample.

Measures. We calculated UC ($M=16,532.17$; $SD=29,401.14$), order, Gini ($M=.46$; $SD=.12$), and turnover ($M=.68$; $SD=.13$) in the same manner as was used in Study 1. As in our Study 1 post hoc analysis, we calculated a grand mean centered creation date variable. In this sample, the creation date variable ranged from -994 to 923 ($M=1$; $SD=652.05$).

Analysis and Results

We created two multilevel models, the first of which included Gini coefficient as the outcome variable and the second of

which included turnover as the outcome variable. In both models, log transformed UC, order, and creation date were included as predictor variables. As in Study 1, all variables were standardized before analysis.

In the model predicting Gini coefficients (Table 3), log transformed UC was positively related to Gini (Estimate = .62; $t=6.16$; $p<.001$) and order was negatively related to Gini (Estimate = -.27; $t=-4.35$; $p<.001$). Creation date was not related to Gini. In the model predicting turnover (Table 4), log transformed UC was negatively related to turnover (Estimate = -.24; $t=-2.19$, $p<.01$), order was positively related to turnover (Estimate = .30; $t=4.41$, $p<.001$), and creation date was positively related to turnover (Estimate = .81; $t=7.27$, $p<.001$).

Discussion

While prior longitudinal studies of online group dynamics (e.g., Schoberth et al., 2006) established an effect of time on those dynamics, this is the first such study to discern maturation effects from cohort effects. In regard to the maturation effect on dispersion of participation, the results of Study 2 confirm the findings from Study 1: As a group ages, participation in discourse tends to become less highly concentrated among the few when controlling for the size of the group and creation date. Study 2 confirmed another finding from Study 1: Group turnover also increases as groups mature. Results

from Study 2 confirmed the results of the post hoc test in Study 1 regarding a cohort effect: Groups with more recent creation tended to have higher turnover, independent of their size. As in Study 1, there was no compelling evidence of a relationship between creation date and dispersion of participation. The results also confirmed the positive association observed in Study 1 between group size and Gini: As groups grow, participation in discourse tends to become more highly concentrated among the few. Study 2 also provided evidence that as their sizes increase, groups tend to retain a higher percentage of their active participants.

General Discussion

Results from these two studies indicate that groups tend to have trouble retaining contributing members over time, and that larger groups tend to become less equal in terms of participation. The challenge of active member retention is becoming more difficult for groups that were formed more recently. While dispersion of participation tends to increase somewhat with the passage of time and turnover tends to decrease somewhat as groups increase in size, the effects of the greatest magnitude work against groups' abilities to become more community-like.

One unexpected finding from both studies is that dispersion of participation in discourse tended to increase over time when controlling for group size, which is inconsistent with the findings of prior studies examining this relationship (Schoberth et al., 2006). One plausible explanation is that the self-reinforcing cycle of motivation and sustained, frequent contribution proposed by Wang and Clay (2012) has limits; over a sufficiently long period of time, even a highly motivated core group of contributors can fail to maintain a steady flow of contributions. Over time, the pioneers who were inclined to dominate discourse are displaced by a growing number of less-motivated bandwagoners. The lack of motivation differentiating bandwagoners from pioneers manifests itself in less frequent commenting (frequent commenting is required to dominate a conversation) and more transience (i.e., higher turnover).

Our observations confirmed our hypothesis that groups with later creation dates possess higher rates of turnover. Explanations for the observed effect fit into one of three categories.

Subreddit-Level Factors

It is possible that during the early stages of Reddit's evolution, subreddits were created for certain types of popular discussion topics (e.g., news, politics, music). As time went on, all of the obvious choices for discussion topics were more likely to have been in use, and all that remained for users to create were subreddits dedicated to highly specific, "niche" topics (e.g., r/AnimalsBeingJerks). Popular, general topics may be "evergreen" in the sense that novel conversations and

observations relating to these topics can be generated for months or even years, and this continuous renewal of novel content may keep users coming back for more. Less popular, highly specific topics may interest new users at first, but because fewer novel conversations and observations can be made about these topics, users leave once the content becomes too repetitive.

Reddit-Level Factors

It is also possible that observed cohort effects are due to differences in the types of people who joined Reddit and chose to contribute to its discourse in 2013 and those who joined and contributed in 2008. In her examination of Usenet, a collection of discussion groups that was, in many ways, an ancestor of Reddit, Wendy Grossman (1997) chronicled dramatic changes in discourse that resulted from an influx of new users. Like Usenet, Reddit experienced a significant influx of new users, from 2008 to 2013. This influx and other changes to the composition of the Reddit userbase may have been driven by changes to the site itself. Over the years, Reddit has changed the means by which posts on its home page are ranked, as well as the list of subreddits featured on new users' homepages by default. These changes to the design of the website could have attracted users who tended to be less likely to contribute to discourse in a sustained manner than those who discovered the website before the changes.

Societal-Level Factors

Finally, it is possible that Internet users *in general* are behaving less like communities and more like ephemeral crowds. Each year, the Internet offers more and more options, pulling users' attention in many different directions while the amount of time users may dedicate to participating in online group discussions remains finite. In response, many users may be increasingly likely to dabble in group participation rather than commit to a sustained pattern of contribution.

These are not necessarily three mutually exclusive explanations; it is possible that any combination of these three types of factors explain the observed relationships. Future research on online groups outside of the realm of Reddit, covering the same time period as was analyzed in our studies, could help to provide evidence of societal-level factors.

The Downside of Growth

Consistent with our initial suppositions and prior research on large online group communication, additional group members tend to be associated with a further concentration of participation among a smaller percentage of the group. This is consistent with social loafing/information overload views of online group participation (Butler, 2001). This observed pattern of dispersion of participation is not the result of a

difference between committed early adopters and flighty bandwagon jumpers; the passage of time partially mitigated the effect of additional group members on concentration of participation. Rather, it is something about the presence of many other group members that increases inequality of participation.

Here, it must be noted that Reddit allows users to “upvote” or “downvote” other users’ comments. Users are allowed to sort comments based on which comments obtained the highest vote score. In small groups, discourse can be navigated by users without sorting comments by their vote scores; it is relatively easy to read a smaller number of comments without having to use a sorting mechanism. But as the group grows, this strategy becomes less tenable. While sorting comments in this way makes large conversations more manageable for readers, it creates an environment in which votes become the difference between having one’s comments seen by others or “shouting into the void.”

Participation in large groups may resemble a popularity contest rather than a genuine attempt to connect with other users. Commenters who are unable to obtain sufficient votes may become discouraged while more popular commenters may be incentivized to maintain or grow their popularity by catering to popular tastes with their comments. Other social websites and applications (e.g., YouTube) use a similar means of indexing comments and thus may be susceptible to a similar inability to maintain equal participation among large numbers of group members.

Limitations

As with most studies introducing new methods of analyzing complex systems, our research possesses several important limitations. The samples used in both studies were drawn from the most popular subreddits on Reddit during early 2017. These samples differ from the vast majority of other subreddits in size, subject matter, and, potentially, the behavior of commenters. The decision to use non-random samples was made for practical purposes: to ensure that the subreddits would vary in size over time. The vast majority of subreddits which comprise the “long tail” of Reddit’s multitude of groups feature very few comments to be analyzed (McEwan, 2016). Future studies should test the generalizability of our findings by examining group dynamics within a random sample of subreddits and/or other online discussion-based groups.

It is possible that the deleted comments that did not appear in our data set, but did appear on Reddit for a time, might have influenced the relationships among our variables of interest had they been included in the analysis. At the very least, future researchers could determine whether or not this is likely to be the case by comparing relationships among variables of interest within subreddits that have many deleted comments with those relationships within subreddits that have very few. It is also important to note that we examined

commenting patterns outside of the context of the threaded conversations in which comments appear on the website. Reddit allows users to initiate comment threads by creating original posts, and it allows users to reply to one another either publicly within the thread or privately through direct messaging. Our focus can best be described as patterns of *contribution* to group discourse rather than interactivity per se. We acknowledge the importance of studying interactivity to the enterprise of understanding online group communication dynamics, and encourage future researchers to examine how time and group size affect aspects of interactivity such as the number of responses each comment receives (i.e., thread depth; McEwan, 2016).

The measure of group dynamics used in these studies is not meant to be an exhaustive means by which to analyze every aspect of discourse, nor are the variables of time and size meant to be exhaustive possible influences on those dynamics. Non-verbal expressions of endorsement (e.g., upvotes, “likes,” shares) are common among social media applications as means of affecting discourse, as are implicit norms and explicit rules established and enforced by group moderators. In addition, differences in the content of users’ comments likely play an important role in establishing certain aspects of community, such as trust and feelings of closeness. The way in which Reddit user behavior is recorded and made available to researchers makes analyses of this kind of active behavior difficult at present. Given broader access to records of users’ behavior, studies of online community social dynamics could confirm that the current findings extend to those who vote but do not comment.

Our measure of turnover restricts us to a month-to-month view of member retention, leaving it unclear as to whether temporal changes in group dynamics were the result of a displacement of one type of user (e.g., pioneers) with another (e.g., bandwagoners). Such displacement effects may be better understood by using network analysis methods (e.g., the Barabasi-Albert model). By using a measure of turnover that identified and tracked a particular group of contributing members over time and assessed how long it took for them to drop out (i.e., conceptualizing turnover as attrition), the precise balance of pioneers and bandwagoners could be assessed, and the relationship between this ratio and dispersion of participation could be tested. Furthermore, we do not test group members’ levels or types of motivation. Doing so in future research would help to confirm or disconfirm the proposed explanations of temporal shifts in online group dynamics.

Finally, subreddit-level factors influencing the outcomes of interest are explored in only a preliminary fashion in this research. Our analysis makes clear that the relationships among our variables of interest are not consistent across subreddits, and that one subreddit-level variable—creation date—influences the outcomes. There are many other subreddit-level factors that may explain the variance observed

among subreddits. For example, commenters in subreddits relating to a certain type of topic (e.g., hobbies) may behave differently than commenters in other subreddits. It is also possible that aforementioned website-wide design changes (e.g., changes to the list of default subreddits) or changes to the moderation rules of particular subreddits (e.g., whether unsubscribed users are permitted to upvote or downvote comments) affected participation dispersion and turnover in a systematic way. Future research exploring these questions would do well to make such comparisons by analyzing a sample that is stratified based on subreddit topic type.

Conclusion

As online group communication continues to be a central part of social, civic, and political life, understanding its dynamics is integral to a variety of scholarly and professional disciplines. Online groups are astonishingly varied in terms of qualities of their participants, topics, and qualities of discourse, and, importantly, their sizes. The conflation of “groups” with “communities,” by scholars or social media companies, is potentially misleading when the majority of group members fail to participate in discourse and/or fail to remain part of groups for very long. By providing metrics by which the community-ness of online groups may be assessed and by studying the relationships between these metrics and other characteristics of groups in a variety of online contexts, we hope to develop a better understanding of how the trust, emotional investment, and social bonds indicative of true community are forged and maintained online.

Declaration of Conflicting Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Reddit users and discussion board moderators have the ability to delete comments. Evidence of such deletions are included in the dataset we analyzed (the date and subreddit on which the comment was posted remains, although the comment and the user are marked as “deleted”). Thus, it is possible to determine how many comments that were part of the original group discourse on Reddit have been excluded from our dataset due to user or moderator deletions. We estimate that roughly 10% to 20% of the original comments were deleted before the data were made available for analysis.
2. It has been demonstrated that, in multilevel analyses, SPSS produces results equivalent to results produced by other popular statistical software used for multilevel analyses such as Stata, HLM, R, and Mplus <https://stat.utexas.edu/images/SSC/documents/SoftwareTutorials/MultilevelModeling.pdf>

References

- Baym, N. (1998). The emergence of on-line community. In S. Jones (Ed.), *Cybersociety 2.0: Revisiting computer-mediated communication and community* (pp. 35–68). Thousand Oaks, CA: Sage Publications.
- Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, 21(3), 2–20.
- Buntain, C., & Golbeck, J. (2014). Identifying social roles in Reddit using network structure. In *Proceedings of the 23rd international conference on world wide web* (pp. 615–620). New York, NY: ACM.
- Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12, 346–362.
- Butler, B. S., Bateman, P. J., Gray, P. H., & Diamant, E. I. (2014). An attraction–selection–attrition theory of online community size and resilience. *MIS Quarterly*, 38, 699–728.
- Chen, Y., Harper, F. M., Konstan, J., & Li, X. (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *The American Economic Review*, 100, 1358–1398.
- De Souza, C. S., & Preece, J. (2004). A framework for analyzing and understanding online communities. *Interacting With Computers*, 16, 579–610.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121–138.
- Facebook1. (2018). *Bringing the world closer together*. Retrieved from <https://www.facebook.com/pg/facebook/about/>
- Facebook2. (2017). *Building global community*. Retrieved from <https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634>
- Farzan, R., Dabbish, L. A., Kraut, R. E., & Postmes, T. (2011). Increasing commitment to online communities by designing for social presence. In *Proceedings of the ACM 2011 conference on computer supported cooperative work* (pp. 321–330). New York, NY: ACM.
- Grossman, W. (1997). *Net. wars*. New York: NYU Press.
- Hong, S., Thong, J. Y., & Tam, K. Y. (2006). Understanding continued information technology usage behavior: A comparison of three models in the context of mobile internet. *Decision Support Systems*, 42, 1819–1834.
- Instagram. (2018). *About us*. Retrieved from <https://www.instagram.com/about/us/>
- Jackson, A., Yates, J., & Orlikowski, W. (2007). Corporate blogging: Building community through persistent digital talk. In *HICSS 40th annual Hawaii international conference on system sciences* (pp. 80–90). Los Alamitos, CA: IEEE Computer Society.
- Jones, Q., & Rafaeli, S. (1999). User population and user contributions to virtual publics: A systems model. In *Proceedings of the international ACM SIGGROUP conference on supporting group work* (pp. 239–248). New York, NY: ACM.
- Joyce, E., & Kraut, R. E. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11, 723–747.
- Kankanhalli, A., Tan, B. C., & Wei, K. K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, 29, 113–143.

- Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information technology adoption across time: A cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly*, 23, 183–213.
- Kim, B. (2011). Understanding antecedents of continuance intention in social-networking services. *Cyberpsychology, Behavior, and Social Networking*, 14, 199–205.
- Lazar, J., & Preece, J. (2002). Social considerations in online communities: Usability, sociability, and success factors. In H. Van Oostendorp (Ed.), *Cognition in the digital world* (pp. 112–132). Mahwah, NJ: Lawrence Erlbaum.
- Leimeister, J. M., Sidiras, P., & Krcmar, H. (2006). Exploring success factors of virtual communities: The perspectives of members and operators. *Journal of Organizational Computing and Electronic Commerce*, 16, 279–300.
- Lerman, K. (2007). User participation in social media: Digg study. In *Proceedings of the 2007 IEEE/WIC/ACM international conferences on web intelligence and intelligent agent technology-workshops* (pp. 255–258). Los Alamitos, CA: IEEE Computer Society.
- McEwan, B. (2016). Communication of communities: Linguistic signals of online groups. *Information, Communication & Society*, 19, 1233–1249.
- Muller, M. (2012). Lurking as personal trait or situational disposition: Lurking and contributing in enterprise social media. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 253–256). Seattle, WA: Association for Computing Machinery.
- Olson, R. S., & Neal, Z. P. (2015). Navigating the massive world of reddit: Using backbone networks to map user interests in social media. *PeerJ Computer Science*, 1, e4.
- Ortega, F., Gonzalez-Barahona, J. M., & Robles, G. (2008). On the inequality of contributions to Wikipedia. In *Proceedings of the 41st annual Hawaii international conference on system sciences* (pp. 304–304). Piscataway, NJ: IEEE Press.
- Panek, E., Hollenbach, C., Yang, J., & Rhodes, T. (2017, July 28–30). Growth and inequality of participation in online communities: A longitudinal analysis. *Proceedings of the 8th 2017 International Conference on Social Media & Society*, New York, NY.
- Park, A., Conway, M., & Chen, A. T. (2018). Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: A text mining and visualization approach. *Computers in Human Behavior*, 78, 98–112.
- Peddibhotla, N. B., & Subramani, M. R. (2007). Contributing to public document repositories: A critical mass theory perspective. *Organization Studies*, 28, 327–346.
- Piketty, T. (2014). *Capital in the 21st century*. Cambridge, MA: Harvard University Press.
- Porter, C. E. (2004). A typology of virtual communities: A multi-disciplinary foundation for future research. *Journal of Computer-Mediated Communication*, 10(1).
- Porter, M. A., Onnela, J. P., & Mucha, P. J. (2009). Communities in networks. *Notices of the American Mathematical Society*, 56, 1082–1097.
- Preece, J. (2000). *Online communities: Designing usability and supporting sociability*. Hoboken, NJ: John Wiley.
- Preece, J. (2001). Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology*, 20, 347–356.
- Preece, J., & Maloney-Krichmar, D. (2005). Online communities: Design, theory, and practice. *Journal of Computer-Mediated Communication*, 10(4).
- Reddit. (2018). The conversation starts on Reddit. Available from <https://about.reddit.com/>
- Resnick, P., & Kraut, R. E. (2011). Introduction. In R. E. Kraut & P. Resnick (Eds.), *Building successful online communities: Evidence-based social design* (pp. 1–20). Cambridge: The MIT Press.
- Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*. Cambridge: The MIT press.
- Ridings, C. M., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *The Journal of Strategic Information Systems*, 11, 271–295.
- Ritzer, G., & Jurgenson, N. (2010). Production, consumption, presumption: The nature of capitalism in the age of the digital “prosumer.” *Journal of Consumer Culture*, 10(1), 13–36.
- Rogers, E. M. (2003). *Diffusion of innovations*. New York, NY: The Free Press.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.
- Schoberth, T., Heinzl, A., & Preece, J. (2006). Exploring communication activities in online communities: A longitudinal analysis in the financial services industry. *Journal of Organizational Computing and Electronic Commerce*, 16, 247–265.
- Shah, C., Oh, J. S., & Oh, S. (2008). Exploring characteristics and effects of user participation in online social Q&A sites. *First Monday*, 13(9).
- Stafford, T. F. (2003). Differentiating between adopter categories in the uses and gratifications for Internet services. *IEEE Transactions on Engineering Management*, 50, 427–435.
- Wang, X., & Clay, P. F. (2012). Beyond adoption intention: Online communities and member motivation to contribute longitudinally. *Journal of Organizational Computing and Electronic Commerce*, 22, 215–236.
- Wellman, B., & Gulia, M. (1999). Net surfers don’t ride alone: Virtual communities as communities. In B. Wellman (Ed.), *Networks in the global village: Life in contemporary communities* (pp. 331–366). London, England: Westview Press.
- Whittaker, S., Terveen, L., Hill, W., & Cherny, L. (2003). The dynamics of mass interaction. In C. Lueg & D. Fisher (Eds.), *From Usenet to CoWebs: Interacting with social information spaces* (pp. 79–91). London, England: Springer.

Author Biographiess

Elliot Panek, PhD, is an assistant professor at The University of Alabama. His research interests include the effects of social media on individuals and society, media selection behavior, and the social dynamics of Reddit.

Connor Hollenbach (BS, MA) is a graduate of The University of Alabama. His research interests include machine learning and big data.

Jinjie Yang (MA) is a graduate student in the College of Communication & Information Sciences at The University of Alabama. His research interests include news frames, trust, and online social dynamics.

Tyler Rhodes is an undergraduate student of Mechanical Engineering at The University of Alabama. His research interests include shape morphing robotic structures.

Appendix. Subreddits Used in Study I Analysis.

Subreddit name	UC M (SD)
r/Art	4,544 (4,711)
r/AskReddit	307,654 (236,894)
r/askscience	9,851 (4,227)
r/aww	45,401 (32,968)
r/books	13,284 (11,770)
r/DIY	8,970 (7,619)
r/Documentaries	5,319 (6,306)
r/food	13,104 (11,218)
r/funny	146,967 (103,960)
r/gadgets	5,510 (4,664)
r/gaming	98,715 (55,133)
r/GetMotivated	6,430 (4,827)
r/gifs	57,376 (45,389)
r/history	4,785 (4,623)
r/IAmA	68,419 (35,862)
r/Jokes	11,288 (13,954)
r/movies	52,316 (40,306)
r/Music	35,712 (21,249)
r/news	38,832 (43,694)
r/personalfinance	14,127 (13,007)
r/philosophy	2,643 (1,527)
r/pics	140,296 (88,206)
r/politics	45,134 (32,765)
r/science	19,296 (7,859)
r/space	7,423 (7,095)
r/sports	8,710 (9,317)
r/television	14,565 (17,680)
r/todayilearned	82,946 (57,456)
r/videos	83,589 (65,481)
r/worldnews	56,527 (44,925)

UC, unique commenter; SD, standard deviation.