Journal of Computer-Mediated Communication, 2023, zmad002
https://doi.org/10.1093/jcmc/zmad002
Full-length Research Article

International Communication Association

OXFORD

Communication networks do not predict success in attempts at peer production

Jeremy Foote [1]*, **Aaron Shaw** [2], **Benjamin Mako Hill** [3]

[1]Brian Lamb School of Communication, Purdue University, USA
[2]Department of Communication Studies, Northwestern University, USA
[3]Department of Communication, University of Washington, USA

*Corresponding author: Jeremy Foote, 100 North University Avenue, West Lafayette, IN 47907-2050, USA. Email: jdfoote@purdue.edu

Abstract

Although peer production has created valuable information goods like Wikipedia, the GNU/Linux operating system, and Reddit, the majority of attempts at peer production achieve very little. In work groups and teams, coordination and social integration—manifested via dense, integrative communication networks—predict success. We hypothesize that the conditions in which new peer production communities operate make communication problems common and make coordination and integration more difficult, and that variation in the structure of project communication networks will predict project success. In this article, we measure communication networks for 999 early-stage peer production wikis. We assess whether communities displaying network markers of coordination and social integration are more productive and long-lasting. Contrary to our expectations, we find a very weak relationship between communication structure and collaborative performance. We propose that technology may serve as a partial substitute for communication in coordinating work and integrating newcomers in peer production.

Lay Summary

Most attempts at online collaborative production start small and never gain more than a few contributors. One possible explanation is that new communities struggle to coordinate their work and integrate their members. When coordination and social integration are working well in offline groups, a group's communication networks are inclusive and dense. We look for evidence of the importance of coordination and integration processes by testing whether communication network structures in 999 early-stage wikis help to predict whether the wiki is productive and how long it stays active. We find little relationship between network structures and performance, and we discuss how technology may partially substitute for communication in coordinating and integrating members.

Keywords: group formation, social network analysis, peer production, group coordination

Introduction

In January 2001, Larry Sanger and Jimmy Wales had a problem. They had worked to recruit experts to build a free, peer-reviewed online encyclopedia called *Nupedia*, but the project had stalled. They decided to create a publicly editable version of their encyclopedia to seek feeder articles which would then go through expert review. Soon, this side project—which they called *Wikipedia*—grew to tens of thousands of contributors and is now one of the most important information repositories in the world.

Wikipedia is built through *peer production*: a term coined by Benkler (2002) to describe a form of Internet-driven collaborative production that involves the aggregation of contributions from many diversely motivated individuals. While it is most frequently associated with knowledge bases like free/libre open-source software, Wikipedia, and OpenStreetMap, peer production also drives collaborative filtering sites like Reddit and many of the most popular websites on the Internet (Benkler, 2016).

There are thousands of new attempts to begin peer production communities every day on platforms like Fandom or Github. The majority of these attempts at peer production never get off the ground. Why can some encourage people to contribute their time and talents while most remain small and short-lived? For one, building new organizations is notoriously difficult. Most firms fail to survive long term and reams of research have been written on the reasons they fail (e.g., Aldrich & Ruef, 2006; Ruef et al., 2003). Attempts at peer production also struggle to grow or last (Schweik & English, 2012), but these have attracted less scholarly inquiry. The extremely low survival rate of new peer production projects suggests that there may be something particularly difficult about the early stages of these efforts. In their early days, groups need to establish initial rules and norms (Fiesler et al., 2018; Tuckman, 1965), a task which is complicated by the inherent lack of hierarchical control structures in peer production (Faraj et al., 2011). Offline groups with "integrative" (i.e., dense, inclusive, and nonhierarchical) communication structures are better able to motivate their members and meet group goals (Balkundi & Harrison, 2006; Cummings & Cross, 2003). Do similar communication network structures support peer production?

In this article, we test whether early-stage peer production communities benefit from the same sorts of communication network structures as offline groups, using a dataset of 999 wiki communities gathered from Fandom in 2010. We create a network based on communication between members of each wiki and examine how well the structure of these networks predicts (1) how productive community members are in adding content to the wiki and (2) how long the community survives. We find that neither integrative nor hierarchical structures predict either productivity or longevity of early-stage wikis. We

Associate Editor: Nicole Ellison
Received: 28 June 2022. **Revised:** 17 January 2023. **Accepted:** 30 January 2023

identify unique features of the empirical setting to propose theoretical explanations for these results. These include the role of "stigmergic" coordination—coordination that happens through modifying the environment rather than through communication—as well as differences in the composition and motivations of participants.

Overall, our findings provide surprising evidence that successful peer production communities can sustain long-lasting, productive cooperation without the communication network structures that predict effective group performance in previous research.

Coordination and social integration in organizations

What determines when group processes work well? A large body of research has sought to answer this question by studying work groups. In work groups, employees have professional and economic incentives to contribute to group goals, but group outcomes remain heterogeneous. This research has identified that two of the most important explanations of the variance in group outcomes are a group's ability to coordinate its activity and its ability to create a sense of group identity via socially integrating its members.

Coordination involves linking together different actors toward collective tasks (Van De Ven et al., 1976). The amount and type of coordination needed depends on the tasks a group is engaged in. For example, more interdependent processes require more coordination (Tushman, 1979; Van De Ven et al., 1976). One type of coordination occurs at the organization level, allowing teams to build a shared understanding of their goals and what needs to be done (Mathieu et al., 2000). Another type of coordination occurs at the resource level, as team members learn who has various resources or capabilities and how to access these resources (Kotlarsky et al., 2015; Wegner, 1987).

Of course, effective groups require more than just coordination. Virtual teams, work groups, and other collaborative groups are social organizations whose ability to process inputs relates to their degree of social cohesion. Communication is especially important in helping to create social integration (Gibbs et al., 2016). Through communication, a new member learns about group norms and expectations and starts to identify as a member of the group (Scott, 2007). When group members see group membership as a key part of their identity they are more willing to sacrifice their own goals in order to contribute to the group's goals (Ellemers et al., 2004; Hogg & Terry, 2000; Van Knippenberg, 2000). Thus, social integration tends to support effective and long-lasting collaboration.

One approach for identifying successful social integration and coordination is to analyze networks of communication between group members. Group processes like coordination and social integration are enabled by and reflected in communication networks (Monge & Contractor, 2003). Both social integration and coordination are theorized to be enabled by what Cummings and Cross (2003) call "integrative" networks (Katz et al., 2004). These networks are characterized by a large number of connections between members, few people on the periphery, and low hierarchy and centralization. Empirical research by communication, organization, and psychology scholars has studied the relationship between communication network measures and group performance in firm-based work groups and distributed or virtual teams. Overall, this research finds that "integrative" communication structures correlate with better task performance (Balkundi & Harrison, 2006; Cummings & Cross, 2003; Hinds & Kiesler, 2002; Reagans et al., 2004).[1]

However, a group's composition, goals, incentives, and communication needs may alter the value of integrative communication structures. For example, early experiments showed that when performing very simple tasks, hierarchical structures are more effective than integrative structures (Leavitt, 1951). While we understand the role of social structures in work groups, very little research has studied how communication structures of less constrained groups relate to their ability to meet their goals. Indeed, Cummings and Cross (2003) call for a better understanding of "informal organizations where there are performance goals, yet there are not constraints from formal hierarchy or other cultural norms that impose structures" (p. 209). Our work takes up that call.

Communication networks in peer production communities

Peer production communities—described as "commons-based peer production" communities in Benkler's (2002) original formulation—differ from firm-based work groups along a number of important dimensions (Benkler, 2016; Benkler et al., 2015). The most salient differences for this study are organizational structure, membership composition, communication tools, and the nature of group outputs (Benkler, 2016). In terms of organizational structure, peer production communities rarely have formal governance or management hierarchies when they begin. While a small number of users have technological powers (e.g., for banning vandals), these users almost never assign tasks or organize other participants. In terms of membership, participants are primarily volunteers who come or go at will, rather than paid employees.

Communication in peer production projects typically happens through text-based, computer-mediated channels rather than face-to-face, which makes coordination both more important and more difficult (DeSanctis & Monge, 1998; Gibson & Gibbs, 2006; Hinds & Bailey, 2003; Kotlarsky et al., 2015; Wiesenfeld et al., 1998).

Finally, peer production projects produce outputs that differ from many traditional organizations. Peer production is built around the creation of information public goods that are shared openly. Traditional public goods, like clean air or national defense, are expected to be underprovisioned when not provided by the state because people can get the benefits of the good whether or not they contribute (Olson, 1965). Fulk et al. (1996) describe the knowledge bases at the heart of peer production as "communal informational public goods" which suffer from an even stronger incentive problem: Contributions benefit everyone *except* the contributor. For example, because the author of a new Wikipedia article already knows the information in the article, the article will likely be of very little use to her.

And yet, many peer production projects have been wildly successful. Benkler (2006) suggests that one reason for this success is that these projects are able to harness diverse motivations—particularly noneconomic motivations. A bevy of empirical research on Wikipedia and open-source software confirms that people participate for many reasons, many of them social (Cheshire & Antin, 2008; Lakhani & von Hippel, 2003; Nov, 2007; von Krogh et al., 2012). The importance of social motivations, combined with peer production projects' technological and structural openness, suggests that successful efforts will have integrative communication networks—the sort of egalitarian structures characteristic of effective work groups and virtual teams. However, prior research indicates that many of the largest peer production communities possess extremely unequal levels of participation with a group of core contributors who make most of the contributions, supplemented by a "long tail" of more peripheral contributors who each add very little (Matei & Britt, 2017; Schweik & English, 2012). Similar patterns characterize the communication networks of mature peer production projects, with sparse networks which have many contributors on the periphery (Crowston et al., 2006) and high levels of hierarchy (Crowston & Howison, 2006). These patterns likely stem from the incredibly low barriers to entry and exit in peer production; this allows for very low-commitment contributions, a dynamic which simply doesn't exist for firm-based groups (Faraj et al., 2016).

These descriptive studies provide evidence that peer production communities are quite unequal and generally much less integrative than work groups, with much more inequality in participation. However, that does not mean that *relatively* less integrative projects are more or less successful. In other words, while all peer production projects have many peripheral contributors, it may be that successful projects have fewer peripheral members and are more effective at integrating these members. Some work has attempted to identify relationships between community structures and outcomes across projects, with mixed results. Overall, this research finds that equality of participation and network density typically have a weak but negative relationship with performance and community growth (Hinds & Lee, 2009; Qin et al., 2015; Roth et al., 2008).

However, these papers typically focus on relatively mature communities that have already produced substantial resources. As Katz et al. (2004) point out, we need to be careful when looking only at successful groups, as their network structures may be a result of (rather than a cause of) the group's success. For example, if peer production communities follow an "adoption curve" model (Rogers, 1962), then we would expect that successful groups would have many more peripheral, less-interested members as they grow—and therefore less integrative communication networks. Popularity would be the cause of the nonintegrative network structure rather than the structure causing popularity.

One approach to partially solve this problem is to look only at communities before they become popular. While some research includes smaller communities (e.g., Kittur & Kraut, 2010; Roth et al., 2008; Schweik & English, 2012) or considers early stages of peer production projects directly (e.g., Foote et al., 2017), we could only find one small-scale study that explicitly examines the network structure of early communities (Hinds & Lee, 2009). The typical focus on popular, mature projects may explain the prevalence of hierarchical, sparse, and unequal networks in the peer production literature. As we discuss in the next section, there are a number of reasons to believe that early-stage peer production communities differ from large-scale projects in ways that predict a greater need for integrative networks.

Communication needs in nascent peer production communities

While the largest peer production communities boast tens or even hundreds of thousands of participants, most attempts never grow larger than a few members (Schweik & English, 2012). If we want to understand when peer production works—and when it does not—we should focus much more attention on early projects. By studying only the relatively rare large and successful projects, much previous research selects on the dependent variable. Focusing on projects at the same stage allows us to more plausibly identify the actual relationship between network structure and community outcomes.

In addition, there are a number of theoretical and empirical reasons to believe that integrative networks should be particularly important in early-stage projects. First, organizational theories suggest that groups go through different stages with different communicative needs. Tuckman (1965) claims that newly created groups go through stages of orientation and norm creation before they can successfully cooperate. How, when, and whether these group formation stages happen in online contexts is not well understood, but empirical work suggests that peer production communities do engage in different types of work at different stages. For example, Schweik and English (2012) find that open-source software projects experience phases of growth, maturity, and decline. Distinct patterns of leadership, governance arrangements, and collective behavior characterize each phase and explain variation in project outcomes. Related research has found that as they grow communities become more structured (Kittur & Kraut, 2010; Shaw & Hill, 2014; TeBlunthuis et al., 2018). This theoretical and empirical evidence suggests that early-stage communities have different sets of needs such as norm-formation and goal-setting, which should require inclusive, broad conversations among group members.

Second, early-stage projects are likely composed of contributors who differ from later users. Both theories of communal public goods and theories about adoption curves suggest that the most interested contributors will join first (Fulk et al., 1996; Marwell & Oliver, 1993; Rogers, 1962). Indeed, empirical work provides some evidence that contributors to new peer production communities are more active and experienced than others (Foote & Contractor, 2018). If this initial group of dedicated members can successfully cohere, research shows the importance of having a core group like this to coordinate work and integrate newcomers (Kittur & Kraut, 2010).

These theories suggest that members of nascent peer production communities are a more dedicated set of participants engaged in work that requires more interdependence than mature projects. Therefore, we expect early-stage peer production projects to benefit from integrative communication structures similar to those found in effective work groups.

Success in peer production communities

Scholars have measured success in peer production in different ways, from membership growth to artifact quality to contribution amount (Howison et al., 2006; Roth et al., 2008; Schweik & English, 2012; Zhu et al., 2012). We focus on two outcomes which are commonly used as success measures and which wiki founders report as most important to them (Foote et al., 2017): the creation of information (*productivity*) and building a long-lasting community (*longevity*). Productivity is similar to measures of success used in studies of work groups. Longevity is not often used in that context, as work groups are more likely to be created and disbanded independent of their performance. In peer production, however, contributors are all volunteers and the barriers to leaving are incredibly low. Therefore, a community's continued existence is evidence that its members see it as valuable (Ren et al., 2012).

Taken together, theories of coordination and social integration, combined with theories of norm formation and the role of social motivations in early-stage peer production communities, and bolstered by empirical evidence of the importance of integrative networks in work groups and virtual teams, we predict that both productivity and longevity will be associated with integrative communication network structures. Put simply, community members who are part of integrative networks are more likely to be dedicated to the group. They should be better able to collaboratively overcome challenges, should be more willing to contribute to group goals, and should be more effective at welcoming and retaining newcomers. These should lead to communities which are both productive and long-lived. We use edit-level data from nearly 1,000 nascent online wikis to measure communication networks and test the following two hypotheses:

> H1a. Early-stage peer production communities with more integrative communication networks will be more productive.

> H1b. Early-stage peer production communities with more integrative communication networks will be more long-lasting.

Data and measures

Our data come from a population of peer production wikis hosted by Fandom. The term "wiki" refers to both the type of software which facilitates the collaborative creation and distribution of information resources as well as the resource produced. While Wikipedia is the most well-known instance of a wiki, there are hundreds of thousands of other wiki communities, with varied goals, topics, and community structures.

Fandom was founded as "Wikia" in 2004 by Jimmy Wales, a founder of Wikipedia, and Angela Beesley, an active Wikipedia contributor. Because many of the wikis on the platform focused on fan culture, the majority of the platform was rebranded as "Fandom" beginning in 2016. Fandom holds the largest sample of publicly accessible and editable wikis, with widely divergent topics and participation rates.

Our dataset is drawn from a complete set of wikis collected from Fandom in April 2010. We use this relatively old dataset for two reasons. First, Fandom began deleting data on small, inactive wikis in mid-to-late 2010. Second, the company began experimenting with new technology to improve wiki-based interpersonal communication in late 2011 that may have both changed communication behavior and affected our ability to measure it consistently (Narayan et al., 2019). Using this 2010 dataset provides us with the most complete set of small Fandom wikis available and allows us to be confident that we measure communication consistently across the full sample.

In preparing the dataset for analysis, we applied several inclusion criteria. Before constructing our measures, we removed the edits that we could identify as automated.[2] While our focus is on new, small communities, in reality most communities never grow large enough to be considered communities at all. Of over 76,000 wikis that we have data for, only 1,264 had at least 700 nonbot edits to article pages as of April 2010.[3] At the point when 700 edits had been made, the median number of participants in the communication network is 10, a similar size to large work groups. We calculate our measures based on a data snapshot of all of the edits made to a community at the point it received its 700th nonbot, nonreverted article edit.[4]

Network definition

To construct measures of communication network structure, we must first create communication networks for each wiki. These networks attempt to capture interactions within the spaces dedicated to conversational activities. Fandom uses MediaWiki software—the same software used by Wikipedia. Like Wikipedia, sites are composed of "article" pages, which are the content of the site, and "talk" pages. These talk pages are connected to every article, user, or policy page on each wiki. Talk pages are used to discuss the page to which they are connected (see Figure 1). We base our networks on the interactions occurring on talk pages. These behavioral traces provide a more objective, behavioral measure of communicative actions than the survey-based self-reports typical in studies of communication structure in work groups (e.g., Cummings & Cross, 2003).

Specifically, we construct communication networks by creating a directed tie from each talk page editor to the five previous contributors to that page. This measure serves as a proxy for directed communication and is very similar to measures used in previous research (e.g., Crowston & Howison, 2005; Keegan et al., 2012; Qin et al., 2015). Determining who a user is actually replying to is difficult and often impossible in peer production contexts (Crowston & Howison, 2005). Fandom talk pages are particularly challenging because they are unstructured text and because users write in various languages and use different norms to structure responses. Although similar measures have often only created ties from the contributor to a single person who edited the page before them (e.g., Keegan et al., 2012; Qin et al., 2015), we consider five previous editors to reflect the reality that commenters—especially at the early stages when the talk pages are small—are frequently in conversation with more than just the previous editor.

Additionally, we create a directed tie when any contributor edits another's "User Talk" page. This is a public talk page connected to a user's account, typically used for interpersonal communication. To ensure that the networks capture communicative rather than purely coincidental interactions, we only include edges that represent at least two interactions. We create a network for each wiki at the point that it had 700 nonreverted article edits. Because this approach—and the choice of five editors in particular—is necessarily arbitrary, we performed a number of robustness checks, where we ran models with stricter conditions on when edges were created. These models are described in the *Limitations* section and their results are reported in the Supplementary Appendix.

We limit our analysis to communities which have at least four people in their communication network, dropping 202 communities. We do so because the network measures we apply (described below) are not meaningful for smaller networks. We also remove 63 wikis whose structure is so simple that our measures cannot be computed. This leaves us with a total of 999 wiki networks in our sample.

Devaronian? edit

I've seen it said elsewhere by fans that Maladi is a female Devaronian, but is there any source for this? Female Devaronian are not meant to be red, and they are meant to be covered in thick fur, ranging from brown to white, as per the Alien Anthology. This description doesn't fit Maladi. --Eyrezer 12:37, 13 June 2006 (UTC)

- What about Jeisel? Fleet Admiral J. Nebulax (Imperial Holovision) 12:39, 13 June 2006 (UTC)
  - I haven't read those comics. Is there a pic of her I can look at? --Eyrezer 12:52, 13 June 2006 (UTC)
    - It's on her article. Fleet Admiral J. Nebulax (Imperial Holovision) 19:04, 13 June 2006 (UTC)
      - 'Thick' doesn't necessarily have to be long -Nightcrawler from X-men is covered in blue fur, yet he also has a head of curly hair. I always figured the fur in these cases to be closer to the velvet on deer antlers; very fine & you don't notice it until you're up close. Tocneppil 02:09, 16 June 2006 (UTC)
        - <sighs> Let's not bring X-Men into this... Fleet Admiral J. Nebulax (Imperial Holovision) 12:18, 16 June 2006 (UTC)
          - =D Sorry, closest pop-cultural example I could come up with to convey the image. I always thought Jeisel looked better than the Alien Anthology's depiction.Tocneppil 16:53, 16 June 2006 (UTC)
            - As do I. Fleet Admiral J. Nebulax (Imperial Holovision) 17:16, 16 June 2006 (UTC)
- The Devaronian females from *Jedi: Aayla Secura* all looked more along the lines of Maladi than the Alien Anthology version. Kuralyov 17:19, 16 June 2006 (UTC)
  - I believe that the female Devaronians we see now are the more correct version, then. Fleet Admiral J. Nebulax (Imperial Holovision) 17:21, 16 June 2006 (UTC)

Figure 1. Example of a thread from the talk page for "Darth Maladi" on the Wookieepedia wiki, dedicated to Star Wars information. Users discuss the article's topic and how to improve the article.

Measures of integrative networks

We borrow from the communication and management literature on teams and networks to identify four overlapping, related measures of integrative communication structures: *centralization*, *core membership size*, *hierarchy*, *and density*.

Centralization refers to variance in how many others each person in a network is connected to. When a few people are much more connected than everyone else centralization is high. Integrative groups, therefore, have low centralization. While higher variance in *contributions* correlates with a peer production community's productivity (Kittur & Kraut, 2010; Qin et al., 2015), theories of group formation suggest that highly centralized communication networks make it difficult to form shared norms and coordinate who knows what. In addition to difficulty in coordinating, by definition a highly centralized community has many members who are not well connected to others and who therefore have less opportunity to become socially integrated.

We measure *centralization* in two ways: *indegree centralization* and *betweenness centralization*. To calculate indegree centralization, we first calculate the indegree for each person in a network. This is simply the number of incoming ties to a node (in our case, the number of different users who replied to or sent a message to a user two or more times). After calculating the indegree for each user, we follow Qin et al. (2015) and measure the Gini coefficient of the distribution of indegrees.

We follow a similar procedure to calculate *betweenness centralization*. Betweenness measures how often a person is on the shortest path between other people (Freeman et al., 1980). This measure is intended to capture the importance of a node, as those that sit in the middle of otherwise unconnected parts of a network can have increased influence (Burt, 2000). In order to control for the role of networks of different sizes, we normalize the betweenness score using the normalization found in Csardi and Nepusz (2006).[5] As with indegree, we measure the inequality in betweenness using the Gini coefficient.

Core membership size refers to how much of the group is well-connected to others and how many members are on the periphery. Integrative groups have a low proportion of their members on the periphery, allowing them to integrate the information from all members and to integrate the members into the group.

While there are a number of measures that could be used to identify peripheral members (such as the clustering coefficient), we believe the measure that best captures the theoretical construct is based on "coreness." The coreness of a node is defined as the largest k for which the node is in a subgraph where all members of the subgraph have at least k ties with the other nodes (Seidman, 1983).[6] We identify contributors as being central (i.e., nonperipheral) if their coreness is at least 3; we calculate the *core membership size* as the proportion of people in the "core" of each wiki's network. When this measure is high, it means that many of the users are part of a cohesive group, where all of the members of the group have interactions with others in the group.

Hierarchy in networks refers to how circular communication flows are. The intuition is that a network is hierarchical if commands and communication move in one direction but not the other. While hierarchy may reflect organization, it can also indicate problematic coordination and socialization processes. Integrative networks have low hierarchy and hierarchy has been found to have a negative association with work group outcomes (Balkundi & Harrison, 2006; Cummings & Cross, 2003).

Hierarchy is calculated following Cummings and Cross (2003), who use the hierarchy measure defined by Krackhardt (1994). Specifically, *hierarchy* is the proportion of paths in the graph which are not cyclical. For a given path from person v_i to person v_j, the path is cyclical if there also exists a path from v_j to v_i.

Density is the proportion of possible ties that actually exist in a network. In other words, it is a measure of how many connections to others people have on average, and more integrative groups have higher density. It seems like density should always contribute to coordination and social integration. Indeed, in work groups, density is generally, albeit weakly, associated with better performance (Balkundi & Harrison, 2006; Reagans & Zuckerman, 2001). However, in both early-stage open-source software (Hinds & Lee, 2009) and Wikipedia "project" groups (Qin et al., 2015), density has a negative relationship with productivity. One explanation is that maintaining a strong, dense social network may take time away from actually contributing to the artifact (Qin et al., 2015). On the other hand, if high density is an indicator of "too much" socializing, then we would expect that members with high density to be more socially integrated.

Dependent variables: productivity and longevity

We use these network measures to predict our two outcome measures. To measure productivity, we want to capture the amount of effort that community members are willing to contribute. One common measure of effort is the number of words added to an article by a contributor (Kittur et al., 2007). To calculate the productivity of the community, we count the total number of words added by contributors in the first 700 nonreverted edits. Because we look at the same number of edits for each wiki (rather than, e.g., the number of edits per month), this is also a rough measure of the per-edit effort that participants put into the project.[7]

To measure longevity, we count how long a wiki remains active following the 700th edit. We consider a wiki inactive at the start of the first 30-day period in which the wiki was edited by fewer than two people.

Controls

Many network measures have a relationship with the size of the network. For example, large networks typically have lower density, since the number of possible ties scales exponentially. Therefore, we include a control measure for *network size*. We also include *talk edits* (the number of total edits to talk pages) and *mean edge weight*, two controls for the overall amount of activity between members, in order to distinguish the structure of communication from the amount of communication. In order to help control for time-varying factors that may have played a role in shaping the growth patterns of particular wikis, we include *months since founding*.

The popularity of wiki topics is another potential confound for the relationships we seek to identify. For example, popular topics may attract more peripheral contributors than niche topics. Without access to a direct measure of popularity, we rely on several proxies. First, we include measures of how quickly the community is producing content. Specifically, we measure *days to 350 edits* and *days to 700 edits*. We also control for the number of *total editors* and *active editors* (editors with at least 10 edits). Finally, we include *contribution inequality*, which is calculated as the Gini coefficient of the number of nonreverted words added per editor. Similar measures have been found to be important predictors of performance in previous work on peer production communities (Kittur & Kraut, 2010). All measures were created using custom python scripts and the igraph python library (Csardi & Nepusz, 2006).

Analysis

We estimate the relationship between project structure and project outcomes by regressing our project-level measures of productivity and community survival on our project-level measures of network structure after the first 700 edits. Because our measures of productivity are all over-dispersed counts, we use negative binomial linear regression to model productivity, modeled using the brm function from the brms package in R (Bürkner, 2017), using the default priors. To estimate community survival, we fit Cox proportional-hazards models using the coxph function from the survival package in R (Therneau, 2022). In each model, we added polynomial terms for significant controls until model fits failed to improve based on likelihood-ratio tests.

Our Supplementary Appendix shows summary statistics for each of our variables. As is typical in peer production, many of our measures are highly skewed. To address this, we use the natural log of *days to 350 edits*, *days to 700 edits*, *total editors*, *active editors*, *network size*, *mean edge weight*, and *density* in all models where they appear.

Results

Figure 2 shows the results of the full model used to predict productivity, with 95% credible intervals.[8] Surprisingly, we find that the only network measure whose 95% credible interval does not cross zero is density. Even here, the effect is so uncertain that when taken together, including the network measures does not appear to improve the fit of the model over a controls-only model, based on a leave-one-out (LOO) comparison (expected log pointwise predictive density (ELPD) difference $= -0.12$, $SE = 5.43$). It is worth noting that density is consistently positive across nearly all of our robustness check models. However, even if a real relationship exists, the LOO comparison suggests that it is very weak. In short, we did not find evidence for H1a.

Our second set of results, predicting the longevity of a community, is presented in Figure 3. Again, the model including network measures surprisingly does not provide a better fit than the controls-only model ($\chi^2 = 4.6$, $df = 5$, $p = .47$) and none of the individual network measures have a significant relationship with the likelihood of survival.[9] As with H1a, we do not find evidence for H1b.

Discussion

We hypothesized that nascent peer production projects would benefit from integrated, nonhierarchical communication in order to coordinate their work and create a long-lasting community. However, our results provide evidence that early communication structure does little to predict a group's productivity or longevity. Overall, these findings suggest two surprising features of early-stage projects: First, they do not require structured communication in order to coordinate their work and second, social integration does not increase project-level productivity or longevity. This is different from finding that nascent peer production projects are unstructured: There was a large amount of variation along each of our measures of network structure. Rather, we found no strong relationship between structural measures and either of our outcomes. We develop theoretical explanations below based on the unique features of peer production.

Stigmergic coordination

We turn first to the role of communication in coordinating work. As we have argued, in situations that require

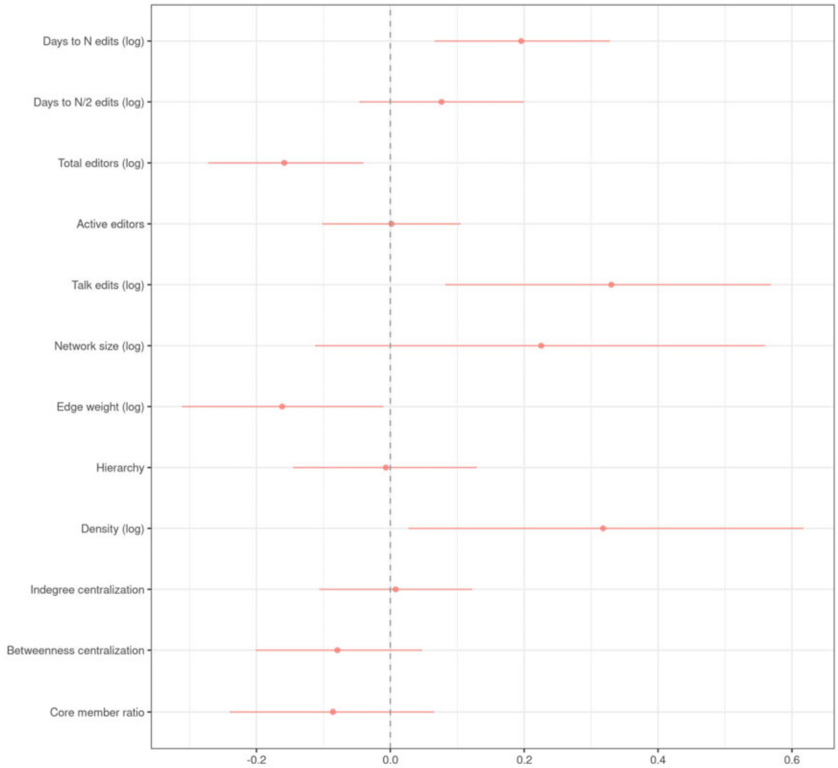

Figure 2. Scaled regression coefficients predicting the number of nonreverted words added in the first 700 edits. Polynomial control terms are excluded for clarity.

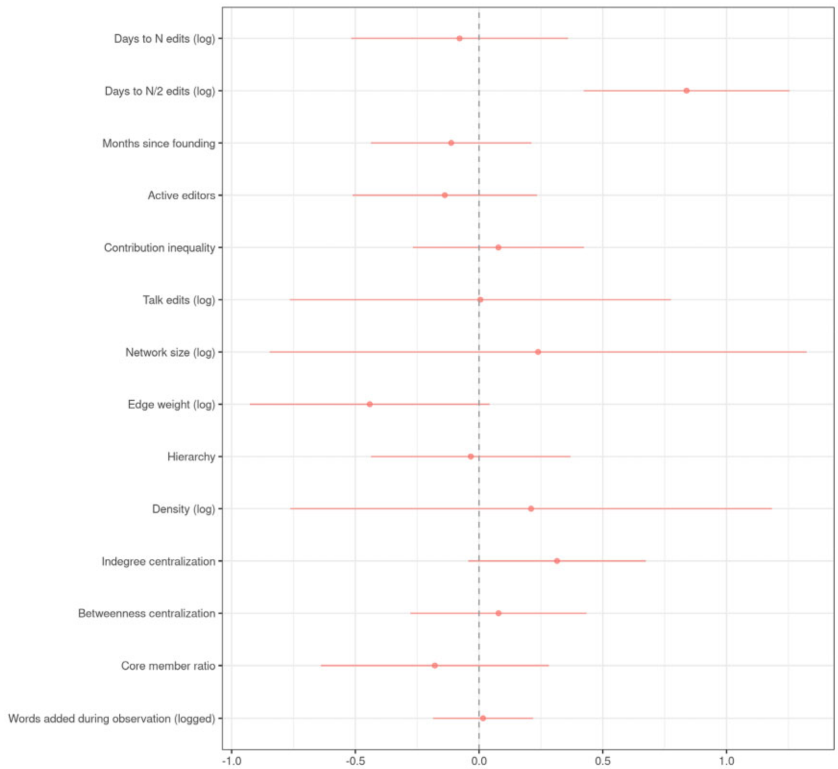

Figure 3. Scaled regression coefficients predicting the hazard of a wiki becoming inactive. Polynomial control terms are excluded for clarity.

collaboration and the integration of different perspectives, decentralized, integrative networks are ideal. On the other hand, straightforward tasks benefit from a centralized network, where one or a few people can organize who does what (Leavitt, 1951). The inequality in levels of participation in peer production suggests that they fall into the latter category (Matei & Britt, 2017). However, if this were the case then we should have found hierarchical (rather than integrative) structures to be associated with productive, long-lived communities.

We do not see that, suggesting that there must be another explanation. One key feature of wikis is that work is organized around a shared, continuously updated artifact. The pages on a wiki are updated as soon as an editor makes an edit. This affordance allows contributors to use the artifact itself to perform "stigmergic" coordination. Stigmergy is a concept that comes from research on social insects like ants and refers to coordination that happens through modifying the environment (e.g., an ant's pheromone trail) rather than through direct communication. Bolici et al. (2016) identify ways that stigmergy applies to coordination in online communities.

For example, users can leave explicit or implicit signals of needs in the artifact (Hinds & Lee, 2009). Changes to wikis can coordinate work, negotiate norms, and integrate newcomers. Some edits are explicitly stigmergic in that they provide trails for others to follow. For example, by creating a link to a nonexisting page, a contributor can signal the need for someone to create that page. Figure 4 shows a more subtle example of how edits can coordinate and provide norm instruction. Fandom provides a page history interface, which shows how a page has changed over time. In this example, the user *DolAmroth123* has taken the "(aged 41)" content added by *Roberto4554* and moved it into its own spot in a predefined template. Without explicit communication, this edit by *DolAmroth123* both enforces a norm about how information should be organized and teaches *Roberto4554* (and others who see the page) about how to use the template feature of the website.

Our findings about the relative unimportance of communication structure, combined with theories of stigmergic communication and coordination, suggest a possible tradeoff between social structure and project structure. When the structure of a project is explicit and tasks are straightforward, as in many early-stage peer production projects, there are few social interdependencies. Many simple coordination tasks can be performed through the wiki itself and thus do not require complex social structures. This theory suggests an explanation for findings in the peer production literature that projects tend to become more structured and hierarchical over time (Halfaker et al., 2013; Shaw & Hill, 2014; TeBlunthuis et al., 2018). In contrast with work groups, the work of a typical peer production project may be simpler in early stages. As projects grow and become more complex, it becomes more difficult to signal needs through the artifact and structured coordination is needed.

Limits of social integration

Even more surprising than our finding that highly integrative networks are not more productive, we find no evidence that communication structure has any relationship to a community's longevity. This suggests that either social integration does not matter very much in peer production or that wiki talk page networks are not capturing the socialization and socializing that is happening. For example, users might be socializing face-to-face or using other online platforms. While we cannot dismiss the latter out of hand, it is evident that talk pages are being used by communities for communication and coordination. Looking at our data, we find that communities make a median of 89 edits to talk pages. This is a large enough number of contributions to suggest that a substantial portion of communication happens through these channels.

How can we reconcile findings about the importance of social integration in work groups (Cummings & Cross, 2003) and large peer production projects (Bryant et al., 2005; Halfaker et al., 2013) with our finding that communication structure does not help to predict nascent peer production community survival?

One possible explanation is that the affordances of peer production make signals of socialization much weaker. In a face-to-face work group, it is often quite obvious who is a

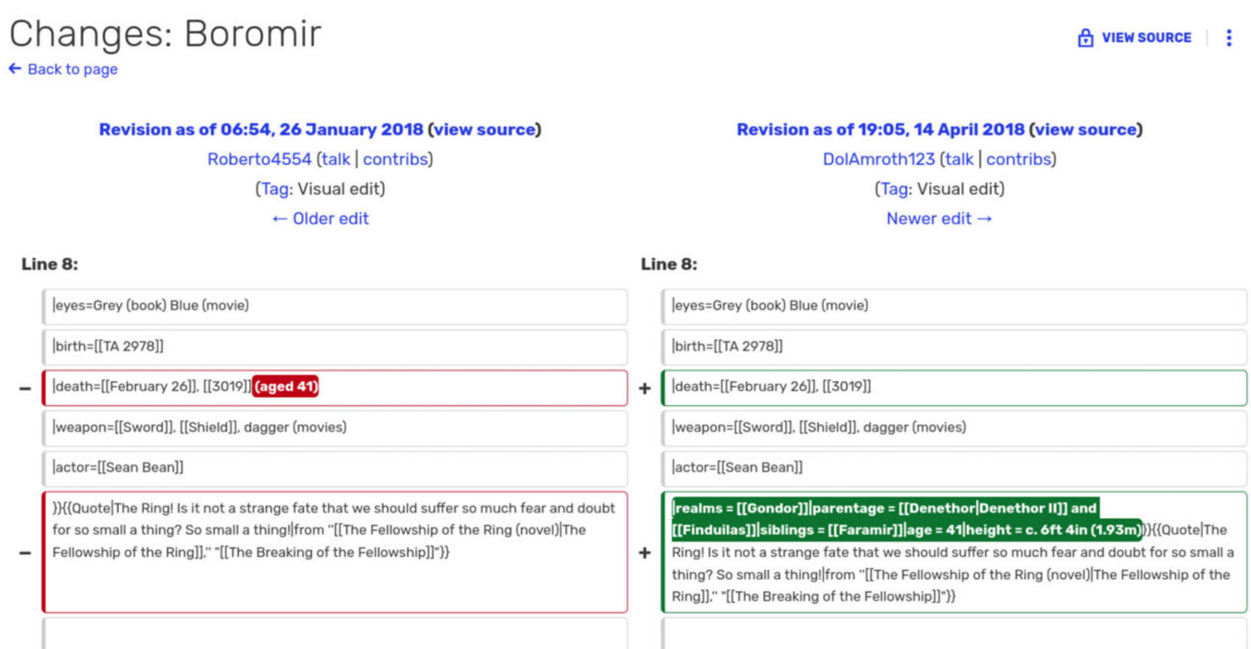

Figure 4. User DolAmroth123 moves information into a predefined template. This shows other users where and how information should be organized on this wiki without any explicit coordination or communication.

friend with whom via many verbal and nonverbal communications (e.g., facial expressions or invitations to go to lunch). While the communications on Fandom are technically more open—in the sense that all talk pages are public—conversations often happen in obscure corners of the wiki, where others are unlikely to ever visit. These affordances likely decrease the salience of group-level measures of social integration. Alternatively, one explanation may be that the visibility of all communications helps community members to develop a group-level identity without strong dyad-level relationships (Hwang & Foote, 2021).

And what about the theory that early groups would need to communicate more in order to create norms and goals? Our results suggest that in peer production this sort of negotiation may be less necessary. On Fandom, as in many other peer production contexts, new communities are either explicitly or implicitly part of a larger ecosystem. It is likely that new communities start with a set of default norms, based on acceptable behavior in related projects (Fiesler et al., 2018). In addition, many wikis have a fairly explicit, visible, and unchangeable goal from the moment they are recorded as a URL. For example, the wiki at lotr.fandom.com will be focused around Lord of the Rings content; there is very little room for negotiation of this high-level goal.

Finally, unlike in a work team where participants have contractual obligations to the organization, the cost of exit from peer production projects is incredibly low. Instead of "storming" and "norming" (Tuckman, 1965), those who disagree with the goals of a wiki may be more likely to create a new project or simply not contribute at all. By changing the relative cost of leaving, peer production communities may encourage people to choose "exit" rather than "voice" (Hirschman, 1970). A complementary implication is that the affordances of peer production platforms may make it difficult for projects to succeed when norms can't be imported or goals are not clear and must be negotiated. Indeed, Hill (2011) argues that a well-defined goal helped Wikipedia succeed where other attempts to build an online encyclopedia fell short.

Taken together, these findings suggest that peer production may work because the affordances allow for a rotating group of participants to contribute efficiently by eliminating many of the costs of coordination and relationship management that are needed for comparable offline projects.

Limitations

One possible explanation for our null results is that our sample size was not large enough to detect the relationships hypothesized. We gain confidence that this is not the case for several reasons. First, studies of work groups that have presented statistically significant estimates of the relationship between communication structure and performance have typically used datasets of fewer than 100 teams, a full order of magnitude smaller than ours (e.g., Balkundi & Harrison, 2006). Even with a larger dataset and more statistical power, one explanation for a null finding might be that the relationship in peer production is simply noisier and more difficult to detect. This seems unlikely because our measures of network structure have comparable variance to similar measures from the work group literature. For example, our measure of hierarchy ($\mu = 0.47$; $SD = 0.24$) has a similar distribution to hierarchy measured by Cummings and Cross (2003) in work groups ($\mu = 0.57$; $SD = 0.28$) and our measure of density ($\mu = 0.24$; $SD = 0.14$) is slightly smaller but has similar variance to the density measured by Wise (2014) in work groups ($\mu = 0.34$; $SD = 0.23$). As a result, we believe that our sample of nearly 1,000 projects should have been more than adequate to identify relationships between network structures and group outcomes similar in magnitude to those found in previous studies of work groups and that any relationships that exist in our data are likely much weaker.

While our study benefits from using a large population of wikis across a number of different domains, it is still possible that our findings may reflect idiosyncrasies of the software platforms and user interface configurations of Fandom wikis as they existed in 2010. Other peer production communities have different goals, norms, and affordances, and the landscape of technologies has changed since these data were collected. We do not believe there are obvious reasons that the key findings of this article would not apply to wikis today or to other peer production contexts, but the extent to which they are generalizable is an addressable empirical question.

Additionally, the analysis presented here is based on digital trace data taken from a single platform and is subject to the sorts of caveats that apply to other similar studies. For example, it does not capture interactions that occur "off-wiki" either in face-to-face meetings or through other technological systems. Although it was much less common in 2010, many peer production communities today use chat tools like Discord as additional communication channels. Researchers using newer data could be less confident that talk pages capture the communication occurring and may need to incorporate multiple channels in order to build believable communication networks. A related problem is that Fandom is a pseudonymous system. The same individual might edit from multiple accounts or sometimes edit anonymously but we treat each of these separate accounts as a different individual—potentially undermining the precision of several of our measures.

While we tried to make principled decisions about how to operationalize the constructs that we measured, other approaches are reasonable. In particular, our approach takes a liberal view of what it means to be in communication with others online by creating edges to up to five previous editors. We undertook a number of analyses in order to gauge how much this decision influences our key findings. First, we looked at the size of talk pages at the point of data collection. If they are quite small, then it seems reasonable that an editor could read an entire page and be in conversation with previous editors of the page. In addition, if pages are very small, this limits the impact of our approach to connect editors liberally. A dot plot of the distribution of edits per page is shown in Supplementary Figure S6, which shows that these pages are indeed very small. On average, talk pages have only 2.91 edits and 1.78 editors (with a median of 1 edit and 1 editor).

In addition, we performed a set of four robustness checks intended to test the impact of more conservative specifications of communicative acts. In two of the models, we build edges based on temporal proximity, creating edges only with those who edited the page in the previous 7 days and 1 day, respectively. In a third model, we simply mirror earlier approaches (e.g., Keegan et al., 2012), only creating an edge to one previous editor. The fourth model looks only at communications which occur on User Talk pages. The results of these robustness checks are shown in Figures 5 and 6 and full regression

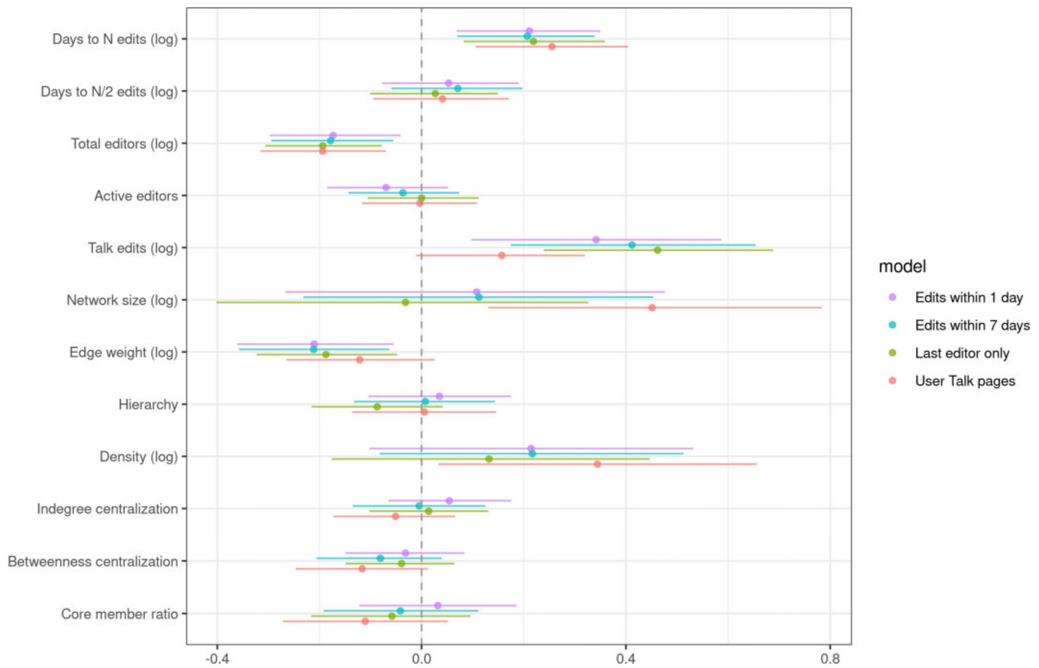

Figure 5. Estimated coefficients when predicting productivity for more restrictive models.

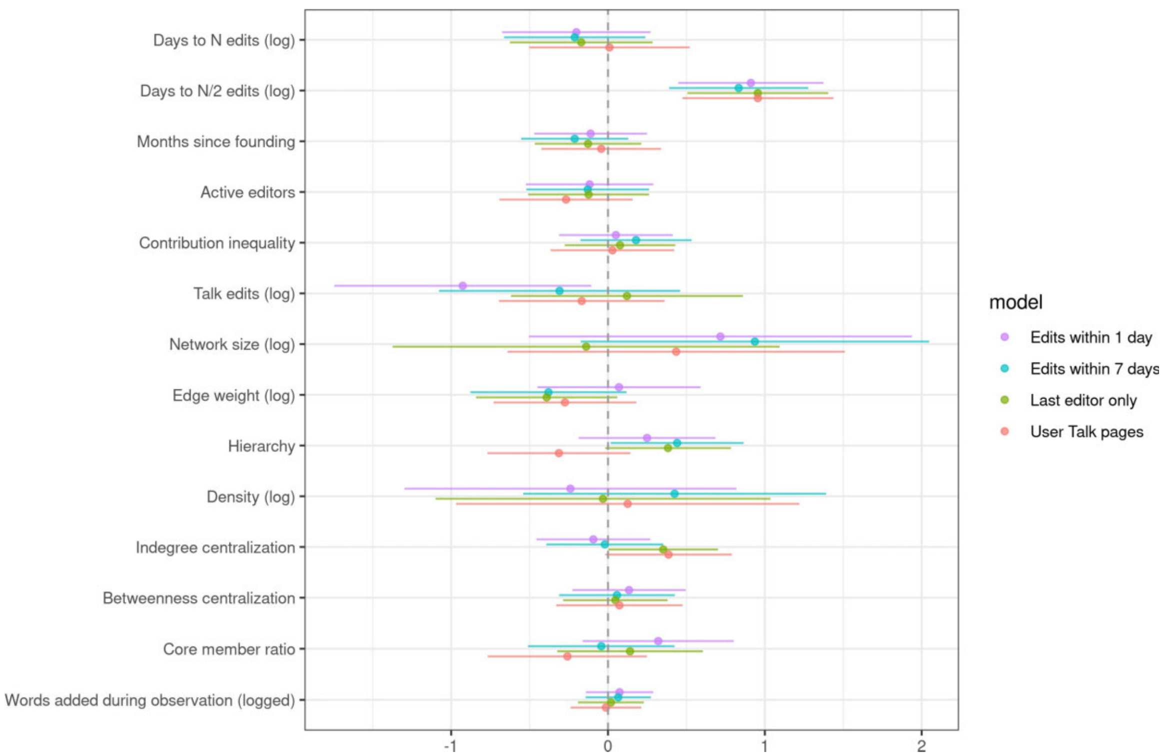

Figure 6. Estimated coefficients for hazard model predicting community inactivity for more restrictive models.

tables are included in the Supplementary Appendix. While there are some interesting differences between these models, the overall story is the same: there is no clear relationship between any network measures and either productivity or survival.

One final additional analysis of our network measures tests whether we are simply measuring noise or whether these measures represent something persistent and unique about a community. To do this, we created networks based on the 700th to 1400th main namespace edit, and compared these with our original networks, created from the 1st to the 700th edits. As reported in Supplementary Table S10, we find that all of the network measures are strongly correlated, with correlations from 0.16 to 0.38 and p-values all $<.001$.

The Supplementary Appendix also includes robustness checks that test an alternate measure of productivity. In these models, we measure the number of nonreverted words added to wikis in the 30 days *following* data collection. As with our alternate network measures, there are some interesting differences, but the overall story remains the same.

Conclusion

In a large-scale study of 999 peer production wikis, we find that communication network structures are poor predictors of both community productivity and longevity. These findings contradict results from earlier studies of offline work groups and teams as well as our theory-derived hypotheses that integrative communication structures would support effective nascent peer production projects. A number of theories suggested that these sorts of structures would support effective groups in any context. Indeed, theories about the formation of norms and social identity suggest that integrative structures should be especially important in nascent online groups, which organize the collaborative, interdependent work of small groups of people with limited formal hierarchies, roles, or assignments. Our findings do not support our hypotheses and undermine these earlier theoretical claims.

Our findings provide a number of directions for future work. First, while we looked at group-level measures of social integration, future work could look at the individual-level mechanisms behind our findings. In particular, we suggest the importance of research on whether an individual's location in a communication network is related to their retention or motivation. In another direction, future work should investigate the ways that communication networks evolve as communities grow and age. For example, it may be that *changes* to network structures are more predictive of group outcomes than overall measures of network structures. Our understanding of these processes would also benefit from qualitative research. For example, we suggested the importance of stigmergy; qualitative research could examine how contributors conceive of and use artifacts as a medium of communication. Qualitative research could also help us to understand better how group identity forms in nascent online groups.

Our findings also suggest opportunities for designers and design researchers. First, we suggest that a focus on tools that promote interpersonal relationships may not be necessary. Group identity and willingness to contribute appear to depend less on dyadic relationships than research in related contexts would suggest, and spending effort in that area may be less effective than designers would hope. On the other hand, it may be worth exploring interfaces that make the stigmergic aspects of projects more salient, such as by highlighting missing links or unanswered questions.

Why do communication networks—important predictors of group performance outcomes across diverse domains—not predict productivity or survival in peer production? Our findings suggest that the relationship of communication structure to effective collaboration and organization is not universal but contingent. While all groups require coordination and undergo social influence, groups composed of different types of people or working in different technological contexts may have different communicative needs. Wikis provide a context where coordination via stigmergy may suffice and where the role of cheap exit as well as the difficulty of group-level conversation may lead to consensus-by-attrition. Future work should evaluate these propositions directly.

Supplementary material

Supplementary material is available online at *Journal of Computer-Mediated Communication*.

Data availability

All of the code and data needed to replicate our study are provided in a Harvard Dataverse at https://doi.org/10.7910/DVN/48QC7B.

Conflicts of interest: None declared.

Open science framework badges

Open Materials

The components of the research methodology needed to reproduce the reported procedure and analysis are publicly available for this article.

Open Data

Digitally shareable data necessary to reproduce the reported results are publicly available for this article.

Preregistered

Research design was preregistered.

Notes

1. See Jacobs and Watts (2021) for a recent, interesting exception to this general finding.
2. Specifically, we removed edits made by users with "bot" at the end of their name and repeated edits where the same author made at least 20 identical edits.
3. Wikis have a number of "namespaces," such as talk pages, user pages, and administrative pages. The primary content of the wiki, such as articles, is typically located in the main namespace.
4. In separate robustness checks reported in the online supplement, we ran models using snapshots at 500 and 900 edits and found similar results.
5. Specifically, we multiply each betweenness score by $\frac{2}{(n-3n+2)}$, where n is the number of nodes.
6. More detail about this measure is provided in the Supplementary Appendix.
7. An alternative measure of productivity based on words added in the month following data collection is discussed in the *Limitations* section and reported in the Supplementary Appendix.
8. Full regression results and robustness checks are available in the Supplementary Appendix.
9. A model allowing for time-varying measures produced similar results.

Funding

This research was partially funded by the National Science Foundation (IIS-1617468 and IIS-1617129).

Acknowledgments

This work relied on use of the University of Washington's Hyak computing cluster. Bingjing Tang and Purdue Univerity's Statistical Consulting Services provided advice on our statistical analyses.Earlier versions of this article received very helpful feedback from participants at the INSNA Sunbelt, American Sociological Association, and the International Communication Association conferences, as well as members of the Community Data Science Collective and Seungyoon Lee.

References

Aldrich, H., & Ruef, M. (2006). *Organizations evolving* (2nd ed.). SAGE Publications.

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal*, *49*(1), 49–68. https://doi.org/10.5465/amj.2006.20785500

Benkler, Y. (2002). Coase's penguin, or, Linux and the nature of the firm. *Yale Law Journal*, *112*(3), 369–446. https://doi.org/10.2307/1562247

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Benkler, Y. (2016). Peer production and cooperation. In J. M. Bauer & M. Latzer (Eds.), *Handbook on the economics of the internet* (pp. 91–119). Edward Elgar.

Benkler, Y., Shaw, A., & Hill, B. M. (2015). Peer production: A form of collective intelligence. In T. W. Malone & M. S. Bernstein (Eds.), *Handbook of collective intelligence* (pp. 175–204). MIT Press.

Bolici, F., Howison, J., & Crowston, K. (2016). Stigmergic coordination in FLOSS development teams: Integrating explicit and implicit mechanisms. *Cognitive Systems Research*, *38*, 14–22. https://doi.org/10.1016/j.cogsys.2015.12.003

Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/1099203.1099205

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Burt R. S. (2000). The Network Structure Of Social Capital. *Research in Organizational Behavior*, . 22, 345–423. https://doi.org/10.1016/S0191-3085(00)22009-1.

Cheshire, C., & Antin, J. (2008). The social psychological effects of feedback on the production of Internet information pools. *Journal of Computer-Mediated Communication*, *13*(3), 705–727. https://doi.org/10.1111/j.1083-6101.2008.00416.x

Crowston, K., & Howison, J. (2005). The social structure of free and open source software development. *First Monday*, *10*(2). https://doi.org/10.5210/fm.v10i2.1207

Crowston, K., & Howison, J. (2006). Hierarchy and centralization in free and open source software team communications. *Knowledge, Technology & Policy*, *18*(4), 65–85. https://doi.org/10.1007/s12130-006-1004-8

Crowston, K., Wei, K., Li, Q., & Howison, J. (2006). Core and periphery in free/libre and open source software team communications. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences, 2006, HICSS '06*, 6 (p. 118a). https://doi.org/10.1109/HICSS.2006.101

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695, 1–9.

Cummings, J. N., & Cross, R. (2003). Structural properties of work groups and their consequences for performance. *Social Networks*, *25*(3), 197–210. https://doi.org/10.1016/S0378-8733(02)00049-7

DeSanctis, G., & Monge, P. R. (1998). Communication processes for virtual organizations. *Journal of Computer-Mediated Communication*, *3*(4). https://doi.org/10.1111/j.1083-6101.1998.tb00083.x

Ellemers, N., De Gilder, D., & Haslam, S. A. (2004). Motivating individuals and groups at work: A social identity perspective on leadership and group performance. *Academy of Management Review*, *29*(3), 459–478. https://doi.org/10.5465/amr.2004.13670967

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, *22*(5), 1224–1239. https://doi.org/10.1287/orsc.1100.0614

Faraj, S., von Krogh, G., Monteiro, E., & Lakhani, K. R. (2016). Online community as space for knowledge flows. *Information Systems Research*, *27*(4), 668–684. https://doi.org/10.1287/isre.2016.0682

Fiesler, C., Jiang, J. A., McCann, J., Frye, K., & Brubaker, J. R. (2018). Reddit rules! Characterizing an ecosystem of governance. *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 72–81). Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/icwsm.v12i1.15033.

Foote, J., & Contractor, N. (2018). The behavior and network position of peer production founders. In G. Chowdhury, J. McLeod, V. Gillet, & P. Willett (Eds.), *iConference 2018: Transforming Digital Worlds* (pp. 99–106). Springer. https://doi.org/10.1007/978-3-319-78105-1_12

Foote, J., Gergle, D., & Shaw, A. (2017). Starting online communities: Motivations and goals of wiki founders. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)* (pp. 6376–6380). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025639

Freeman, L. C., Roeder, D., & Mulholland, R. R. (1980). Centrality in social networks: II. Experimental results. *Social Networks*, *2*(2), 119–141. https://doi.org/10.1016/0378-8733(79)90002-9.

Fulk, J., Flanagin, A. J., Kalman, M. E., Monge, P. R., & Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, *6*(1), 60–87. https://doi.org/10.1111/j.1468-2885.1996.tb00120.x

Gibbs, J. L., Kim, H., & Ki, S. (2016). Investigating the role of control and support mechanisms in members' sense of virtual community. *Communication Research, 46(1), 117–145.* https://doi.org/10.1177/0093650216644023

Gibson, C. B., & Gibbs, J. L. (2006). Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly*, *51*(3), 451–495. https://doi.org/10.2189/asqu.51.3.451

Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, *57*(5), 664–688. https://doi.org/10.1177/0002764212469365

Hill, B. M. (2011). *Almost Wikipedia: What eight collaborative encyclopedia projects reveal about mechanisms of collective action*. Berkman Center for Internet and Society Luncheon Series Presentation.

Hinds, D., & Lee, R. M. (2009). Communication network characteristics of open source communities. *International Journal of Open Source Software and Processes*, *1*(4), 26–48. https://doi.org/10.4018/jossp.2009100102

Hinds, P. J., & Bailey, D. E. (2003). Out of sight, out of sync: Understanding conflict in distributed teams. *Organization Science*, *14*(6), 615–632. https://doi.org/10.1287/orsc.14.6.615.24872

Hinds, P. J., & Kiesler, S. (Eds.). (2002). *Distributed work*. MIT Press.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard University Press.

Hogg, M. A., & Terry, D. I. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review*, *25*(1), 121–140. https://doi.org/10.5465/amr.2000.2791606

Howison, J., Inoue, K., & Crowston, K. (2006). Social dynamics of free and open source team communications. In Damiani, E., Fitzgerald, B., Scacchi, W., Scotto, M., & Succi, G. (Eds.), Open Source Systems (pp. 319–330). Springer US. https://doi.org/10.1007/0-387-34226-5_32

Hwang, S., & Foote, J. (2021). Why do people participate in small online communities? *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 462:1–462:25. https://doi.org/10.1145/3479606

Jacobs, A. Z., & Watts, D. J. (2021). A large-scale comparative study of informal social networks in firms. *Management Science*, mnsc.2021.3997. https://doi.org/10.1287/mnsc.2021.3997

Katz N., Lazer D., Arrow H., & Contractor N. Network Theory and Small Groups. *Small Group Research*. 2004. *35* (3) 307–332. https://doi.org/10.1177/1046496404264941.

Keegan, B., Gergle, D., & Contractor, N. (2012). Staying in the loop: Structure and dynamics of Wikipedia's breaking news collaborations. *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (pp. 1:1–1:10). Association for Computing Machinery. https://doi.org/10.1145/2462932.2462934

Kittur, A., Chi, E., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the Bourgeoisie. *Alt.CHI*. Association for Computing Machinery.

Kittur, A., & Kraut, R. E. (2010). Beyond Wikipedia: Coordination and conflict in online production groups. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)* (pp. 215–224). Association for Computing Machinery. https://doi.org/10.1145/1718918.1718959

Kotlarsky, J., van den Hooff, B., & Houtman, L. (2015). Are we on the same page? Knowledge boundaries and transactive memory system development in cross-functional teams. *Communication Research*, *42*(3), 319–344. https://doi.org/10.1177/0093650212469402

Krackhardt, D. (1994). Graph theoretical dimensions of informal organizations. In K. M. Carley & M. J. Prietula (Eds.), *Computational organization theory* (pp. 89–111). Lawrence Erlbaum Associates.

Lakhani, K. R., & von Hippel, E. (2003). How open source software works: "Free" user-to-user assistance. *Research Policy*, *32*(6), 923–943. https://doi.org/10.1016/S0048-7333(02)00095-1

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology*, *46*(1), 38–50. https://doi.org/10.1037/h0057189

Marwell, G., & Oliver, P. (1993). *The critical mass in collective action: A micro-social theory*. Cambridge University Press.

Matei, S. A., & Britt, B. C. (2017). *Structural differentiation in social media: Adhocracy, entropy, and the "1% effect"*. Springer.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273–283. https://doi.org/10.1037/0021-9010.85.2.273

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. Oxford University Press.

Narayan, S., TeBlunthuis, N., Hale, W. S., Hill, B. M., & Shaw, A. (2019). All talk: How increasing interpersonal communication on wikis may not enhance productivity. *Proceedings of the ACM: Human-Computer Interaction*, *3*(CSCW), 101:1–101:19. https://doi.org/10.1145/3359203

Nov, O. (2007). What motivates Wikipedians? *Communications of the ACM*, *50*(11), 60–64. https://doi.org/10.1145/1297797.1297798

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Harvard University Press.

Qin, X., Cunningham, P., & Salter-Townshend, M. (2015). The influence of network structures of Wikipedia discussion pages on the efficiency of WikiProjects. *Social Networks*, *43*, 1–15. https://doi.org/10.1016/j.socnet.2015.04.002

Reagans, R., Zuckerman, E., & McEvily, B. (2004). How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative Science Quarterly*, *49*(1), 101–133. https://doi.org/10.2307/4131457

Reagans, R., & Zuckerman, E. W. (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science*, *12*(4), 502–517. https://doi.org/10.1287/orsc.12.4.502.10637

Harper, R., Harper, F. M., Sara Drenner, S., & Terveen, L.G. (2012). Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. *MIS Quarterly*, *36* (3) 841. https://doi.org/10.2307/41703483.

Rogers, E. M. (1962). *Diffusion of innovations*. The Free Press of Glencoe.

Roth, C., Taraborelli, D., & Gilbert, N. (2008). Measuring wiki viability: An empirical assessment of the social dynamics of a large sample of wikis. *Proceedings of the 4th International Symposium on Wikis* (pp. 27:1–27:5). Association for Computing Machinery. https://doi.org/10.1145/1822258.1822294

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review*, *68*(2), 195–222. https://doi.org/10.2307/1519766

Schweik, C. M., & English, R. C. (2012). *Internet success: A study of open-source software commons*. MIT Press.

Scott, C. R. (2007). Communication and social identity theory: Existing and potential connections in organizational identification research. *Communication Studies*, *58*(2), 123–138. https://doi.org/10.1080/10510970701341063

Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, *5*(3), 269–287. https://doi.org/10.1016/0378-8733(83)90028-X

Shaw, A., & Hill, B. M. (2014). Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication*, *64*(2), 215–238. https://doi.org/10.1111/jcom.12082

TeBlunthuis, N., Shaw, A., & Hill, B. M. (2018). Revisiting "The rise and decline" in a population of peer production projects. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 355:1–355:7). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173929

Therneau, T. M. (2022). *A package for survival analysis in R*. https://cran.r-project.org/web/packages/survival/citation.html.

Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, *63*(6), 384–399. https://doi.org/10.1037/h0022100

Tushman, M. L. (1979). Work characteristics and subunit communication structure: A contingency analysis. *Administrative Science Quarterly*, *24*(1), 82–98. https://doi.org/10.2307/2989877

Van De Ven, A. H., Delbecq, A. L., & Koenig, R. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, *41*(2), 322–338. https://doi.org/10.2307/2094477

Van Knippenberg, D. (2000). Work motivation and performance: A social identity perspective. *Applied Psychology*, *49*(3), 357–371. https://doi.org/10.1111/1464-0597.00020

von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. (2012). Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quarterly*, *36*(2), 649–676. https://doi.org/10.2307/41703471

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). Springer. https://doi.org/10.1007/978-1-4612-4634-3_9

Wiesenfeld, B. M., Raghuram, S., & Garud, R. (1998). Communication patterns as determinants of organizational identification in a virtual organization. *Journal of Computer-Mediated Communication*, *3*(4), 777-790. https://doi.org/10.1111/j.1083-6101.1998.tb00081.x

Wise, S. (2014). Can a team have too much cohesion? The dark side to network density. *European Management Journal*, *32*(5), 703–711. https://doi.org/10.1016/j.emj.2013.12.005

Zhu, H., Kraut, R. E., & Kittur, A. (2012). Effectiveness of shared leadership in online communities. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 407–416). Association for Computing Machinery. https://doi.org/10.1145/2145204.2145269