# The Impact of Membership Overlap on the Survival of Online Communities

**Haiyi Zhu, Robert Kraut, Aniket Kittur**
Carnegie Mellon University, Pittsburgh, PA, USA
{haiyiz, robert.kraut, nkittur}@cs.cmu.edu

**ABSTRACT**
If people belong to multiple online communities, their joint membership can influence the survival of each of the communities to which they belong. Communities with many joint memberships may struggle to get enough of their members' time and attention, but find it easy to import best practices from other communities. In this paper, we study the effects of membership overlap on the survival of online communities. By analyzing the historical data of 5673 Wikia communities, we find that higher levels of membership overlap are positively associated with higher survival rates of online communities. Furthermore, we find that it is beneficial for young communities to have shared members who play a central role in other mature communities. Our contributions are two-fold. Theoretically, by examining the impact of membership overlap on the survival of online communities we identified an important mechanism underlying the success of online communities. Practically, our findings may guide community creators on how to effectively manage their members, and tool designers on how to support this task.

**Author Keywords**
Membership overlap, online communities, survival analysis

**ACM Classification Keywords**
H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction; K.4.3 [Computers and Society]: Organizational Impacts – Computer supported collaborative work.

## INTRODUCTION

Online communities are becoming an increasingly important way for work to be done in the Internet age. They aggregate the efforts of volunteers to produce complex artifacts such as the largest encyclopedia in human history (i.e., Wikipedia) or the software that powers the Internet (e.g., Apache). They are the basis of solving difficult software challenges (e.g., stackoverflow.com) or doing product design [28].

While the number of online communities to join is still growing exponentially, the growth of Internet users in developed countries has slowed over the last few years [11]. As a result, no single community is likely to monopolize its members' time. People are likely to belong to multiple communities that match their interests.

If people belong to multiple online communities, their joint membership can influence the survival of each of the communities to which they belong. On one hand, when people participate in many communities simultaneously, the time and effort they spend on one community will take time and effort from the others, reducing the resilience of them all. On the other hand, the knowledge, experience and social capital members obtain from one community can be transferred to other communities they concurrently participate in, and thus increasing the communities' ability to survive. For example, the spread of Wikipedia policy from the English Wikipedia to Wikipedia in other languages probably helped these smaller communities to thrive. Although the explosive growth of online communities and their impact on society have attracted hundreds of researchers to study the factors that lead to community success (e.g., [15], [21], [28]), very few of them have investigated how the relationship with other communities, including membership overlap, can influence their success. Wang et al. conducted a relevant study of Usenet groups, showing that sharing members with other groups reduced future growth rates, suggesting that membership overlap puts competitive pressure on online groups [26]. However, this research examined only the detrimental effects of membership overlap. We know of no research that has studied the potential benefits that membership overlap can bring to online communities.

This paper examines the effects of membership overlap on the survival of online communities. We use panel data from Wikia, a software platform that supports Wikipedia-like online communities. For example, there are Wikia communities organized around topics like movies (e.g., Star Wars), video games (e.g., World of Warcraft), and lifestyles (e.g., healthy recipes). Our analysis is based on archival data about 5673 communities from their inception to 2008. Our main finding is that higher levels of membership overlap was positively associated with greater survival of online communities. Furthermore, the beneficial effects of membership overlap on the survival of a particular, focal


*CHI 2014,* April 26–May 1, 2014, Toronto, Ontario, Canada.

ACM 978-1-4503-2473-1/14/04...$15.00.
http://dx.doi.org/10.1145/2556288.2557213

community were stronger when 1) the focal community is young; 2) the intersecting communities with which the focal community share members are mature; and 3) the shared members are core members in the intersecting communities. However, membership overlap is negatively associated with the survival when shared members are core in the focal community.

The contributions of this work are two-fold. First, we examine how membership overlap with other communities influences the survival of a focal community, providing new insight into mechanisms underlying successful online communities. Second, on the practical side, our findings may guide community leaders to better manage their members and build successful online communities

**Survival of Online Communities**

Research investigating the factors leading to continued functioning of online community falls into three categories: research on motivations of individual members in the community, research on dynamics of individual communities, and research on inter-community relationships. Research on the inter-community relationships is quite neglected.

The first type of research focuses on individuals in the community. The survival of online communities relies on the continuous participation of individual members. There is a large literature investigating the factors that motivate individuals to participate (e.g., [27] [19]). Weber [27] and Lerner and Tirole [17] use a cost-benefit framework for member motivation. The basic idea is that people act as if they are performing a calculation to assess the net benefit they will receive in return for their efforts in the community. The benefits include having enjoyment and fun [15][19], pursuing beliefs and values shared with other people [25], expressing humanitarian concerns for others [19], developing careers [15], and protecting oneself from negative emotions and enhancing positive attitudes [4]. One implication of this type of research is that online communities need to continuously provide benefits to members in order to keep active and healthy.

The second type of research investigates how the community-level characteristics influence the success of online communities. Research has explored two main types of community-level characteristics: composition (i.e., the makeup of the community, such as its size or age and gender composition) and structure (i.e., the patterns of the relationship among the members such as social network structure, leadership structure and governance structure). Examples of research investigating composition characteristics include Chen et al's work about diversity [7] and Butler's work on membership size and communication activity [6]. Examples of research examining structural characteristics include Kairam et al's work on members' social ties [15], Zhu et al's work on shared leadership [28], Choi et al's work on socialization [8], and O'Mahony and Ferraro's work on governance [21]. An implication of the research on community-level characteristics is that communities can become successful by adjusting their input (e.g., diversity of members, group size) and optimizing their internal structures (e.g., governance structure).

The third type of research investigating the survival of online communities adopts an ecological view. All online communities exist within a larger population of communities, with which they cooperate and compete. The relationship among these communities can affect the survival of all communities within a niche. Although there is a long tradition of ecological research about offline organizations [1], ecological research about online communities has been quite neglected. The only relevant research we know of is Wang et al's work about membership overlap on the growth of Usenet groups [26]. Wang et al took a competition view of membership overlap. They argue that an individual's time is scarce. When multiple online communities rely on the participation of the same members, the time members spent on one community takes time away from another community, thus reducing the chance of survival for both communities.

However, Wang et al. [26] did not completely characterize the effects of membership overlap on the survival of online communities. Research in organizational ecology has demonstrated that organizations that exist in a common population do not merely compete with each other, but can also learn strategies, practices and technologies from their "competitors" (e.g., [1]). For example, Ingram and Baum [13] found that the survival a hotel chain is positively related to the total operating experience other hotel chains had accumulated. Moreover, organizational behavior researchers (e.g., [20]) argue that shared team membership (i.e., membership overlap in work teams) can have positive effects on team productivity and team learning. Specifically, more shared membership and shared membership with more teams can improve a focal team's efficiency and diversity. Although these finding are based on research in offline organizations and groups, the mechanisms involved are likely to be applicable to online communities. Additional evidence is directly relevant to online communities. Hill and Shaw [9] have challenged the assumption that competition between projects is an important dynamic driving contribution to online communities. Hill and Shaw argue that the volunteer resources are not fixed and participation in one community does not necessarily take detract from participation in similar communities. Their analyses showed that the volume of contribution to pages within Wikipedia is positively related to the volume of contribution on related topics in other encyclopedia wikis run by Wikia. In sum, there are several reasons to believe that membership overlap might have positive as well as negative effects on the survival of online communities.

In the following section we will predict the effects of membership overlap on the survival of online communities. Particularly, we are interested in the conditions under which the beneficial effects of membership overlap are stronger. We use the following vocabulary in describing the hypotheses. A *focal community* is the community of interest (especially we are interested in its likelihood to survive). *Intersecting communities* are the communities with which a focal community shares members. *Shared members* are the participants who participated in both the focal community and the intersecting communities.

### Effects of Membership Overlap

We hypothesize that membership overlap can benefit online communities for three reasons. First, overlapping members may bring skills, knowledge, and experience they gain from their participation in one community to the others. According to theories of bridging social capital theory (e.g., [5]), people who participate in multiple communities connect relatively disconnected groups of people. These overlapping members can bring in valuable resources and novel information to the communities they belong to. For example, through participation, members learn basic technical skills (e.g., using editing tools in Wiki-like websites), implicit social skills (e.g., communicating and collaborating with other members) and community building skills (e.g., organizing activities, socializing new members, and resolving conflicts) [2]. The skills and knowledge may be transferred across communities when people participate in multiple online communities. Second, communities may gain diverse perspectives when their members participate in a variety of communities [20]. Research shows that a moderate level of diversity can increase productivity and decrease member turnover in online communities [7]. Therefore, a moderate level of membership overlap may positively affect the survival of online communities through increased diversity. Third, according to network diffusion theories ([15]), people are more likely to join a community if people in their social networks are already participating. Therefore, members participating in multiple communities might increase the probability that friends in one community will join in the other community, thus benefiting both communities.

At the same time, there are three reasons why high levels of membership overlap will harm online communities. By high levels, we refer to a large proportion of members belonging to many other communities. First, although Hill and Shaw showed that participating in two communities did not decrease contributions to either, there is still likely to be limits on members' time and effort. When individuals participating in too many communities exceed their limits, communities will start to compete with each other for their mutual members' time, thus reducing the chance of survival. Second, high levels of overlap might harm the survival of online communities by lowering members' identification with the communities. Common identity is a powerful way to keep members around in the community [22]. The basic cause of common identity is social categorization, in which people perceive themselves as members of a social category and contrast themselves with people outside the category [10]. However, as membership overlap becomes high, the boundaries between communities become ambiguous, which lowers people's identification with a certain community. With lowered group identification, people are less likely to participate, leading to decreased community survival. Third, high levels of membership overlap lead to high levels of diverse experiences which might harm the community by increasing the chances of conflicts. Chen et al [7] found out that diversity in experience in Wikipedia keeps members in the community only up to a point. Beyond that point (i.e., when the diversity is high), members are more likely to withdraw. In sum, high levels of membership overlap may decrease the chance of survival for online communities.

Therefore, we hypothesize that membership overlap has a curvilinear effect on the survival of online community:

**Hypothesis 1.** *Moderate levels of membership overlap enhance community survival, but very low or very high levels of membership overlap diminish community survival.*

The beneficial effects of membership overlap on the survival of focal community might be moderated by the maturity of both the focal community and intersecting communities (i.e., ones with which the focal community shares members). Also the roles of shared members in both focal communities and intersecting communities may influence the effects of membership overlap.

Specifically, we hypothesize that the beneficial effects of membership overlap are stronger when the communities with which focal community shares members are more mature. First, mature communities are likely to have developed skills, knowledge, and ways of operating compared to young communities, and shared members provide the conduit to transfer these resources. Second, mature communities have longer operating history, which may enrich members' experience and enhance diversity. Third, more mature communities are often larger, providing more opportunities for the focal community to recruit. In sum, members who participate in more rather than less mature communities are likely to acquire useful knowledge and experiences, diverse perspectives, and contact with potential recruits, which in turn are more likely to benefit the other communities they simultaneously participate in.

| Effects of membership overlap on community survival | | | |
|---|---|---|---|
| | **Pros**<br>• Transfer knowledge<br>• Gain diverse perspective<br>• Recruit new members | **Cons**<br>• Compete for shared members' time and efforts | **Overall effects (H1)**<br>Moderate levels of membership overlap enhance community survival. Low or high levels of membership overlap diminish community survival. |
| **Maturity of the intersecting communities** | Pros dominate when the intersecting communities are mature | | **Moderating effects (H2a)**<br>Membership overlap is more likely to enhance community survival when the intersecting communities are mature. |
| **Maturity of the focal community** | Pros dominate when the focal community is young | | **Moderating effects (H2b)**<br>Membership overlap is more likely to enhance community survival when the focal community is young. |
| **Role of shared members in intersecting communities** | Pros are stronger when the shared members are core members in intersecting communities | | **Moderating effects (H3a)**<br>Membership overlap is more likely to enhance community survival when shared members are core in the intersecting communities. |
| **Role of shared members in the focal community** | | Cons are stronger when the shared members are core members in the focal community | **Moderating effects (H3b)**<br>Membership overlap is less likely to enhance community survival when shared members are core in the focal community. |

**Table 1. Summary of the hypotheses**

Moreover, we hypothesize that the beneficial effects of membership overlap are stronger when the focal communities are young. Online communities are fragile when they are young, and the majority never get off the ground. For example, SourceForge hosts over 300,000 software development projects, but 90% have fewer than four members ([23] p. 231). When they are young, communities have greater uncertainty about what their goals are, how to manage their members, and how to attract new members. Shared members who had experiences in other communities can benefit younger communities most since they can import technical skills, community building experience and human resources which are crucial to the survival of young online communities.

**Hypothesis 2a.** *Membership overlap is more likely to enhance community survival when the intersecting communities are mature*.

**Hypothesis 2b.** *Membership overlap is more likely to enhance community survival when the focal community is young.*

Furthermore, we hypothesize that the beneficial effects of membership overlap should be stronger when the shared members are core members in other communities. Most online communities have a core-peripheral structure [2]. Take Wikipedia as an example: peripheral members tend to participate in tasks that are useful but not crucial, such as correcting spelling and grammar errors. In contrast, core members tend to take on tasks central to the functioning of the communities, such as discussing policies, voting for or running for administrators, and socializing and educating newcomers [2]. Shared members who are core in other communities are more likely to have knowledge, experiences and social capital the focal community needs than are those who are peripheral in the other communities.

However, the beneficial effects of shared membership might be weakened when the shared members are core members in the focal communities. Core members carry on tasks central to the communities, which take much more time and efforts than peripheral members. In Wikipedia, administrators made 5010 revisions (a measure of contributions) on average [3], while the median number of revisions from non-administrators is 1. Therefore, when core members are participating in multiple communities simultaneously, they may reach limits of their energy, which decreases their participation in the focal community and decreases the likelihood of survival of the focal community.

**Hypothesis 3a.** *Membership overlap is more likely to enhance community survival when shared members are core in the intersecting communities.*

**Hypothesis 3b.** *Membership overlap is less likely to enhance community survival when shared members are core in the focal community.*

## METHOD

### Data collection

Wikia, a free web hosting service for wikis, provides the data for this research. A wiki is a type of website which

allows its users to add, modify, or delete its content via a web browser. Wikia is based on the same technology that powers Wikipedia. Wikis in Wikia cover a broad range of topics, including education, entertainment, finance, food and drink, gaming, politics, technology, sports and others.

Each wiki has project pages on which members can coordinate and organize the writing and the editing of articles. Once they have joined a wiki, members can create a personal profile to share information about themselves and interact with others. Since each wiki has a unique topic, dedicated pages to coordinate activities, and distinct places for users to interact with each other, we consider each wiki as an independent community.

Once a user creates an account in one wiki, this account can be used to participate in any other wiki in Wikia. The universal Wikia account allows us to track shared members among wikis. The dataset includes 5673 wikis from their inception to 2008. The oldest wiki has 7 years' history and the median age is 10 months.

### Analysis strategy: survival analysis

The purpose of the analysis is to estimate how membership overlap influences the survival for online communities. Because Wikia communities are organized to produce content, we consider a community "alive" (i.e., active) if it is producing content and "dead" or at least dormant when it stops. We conduct a survival analysis, a statistical technique for modeling time to an event [24]. While survival analysis can be used to analyze death in biological organisms, it is appropriate for modeling many other types of event histories, like an appliance's time to failure, the time until an ex-smoker resumes smoking and or the time until a restaurant goes out of business. Unlike conventional regression techniques, it is robust to censored data, in which the event of interest does not occur during the period of observation. Because membership overlap for a given community varies over time, we used discrete time proportional hazard models [14]. The unit of analysis is the community-month. We used *ln(t),* where *t* denotes the month, as the baseline hazard function.

### Measurement

*Dependent variable*

- **Community dormancy.** We define a community to be dormant (the inverse of active) in a given month if the community did not have any activity (including discussion pages and community pages) in the given month and the preceding two months. Community dormancy is a binary variable. This variable is assigned to 1 if the community was dormant during month *t*; it is assigned 0 if the community was still active in month *t*. A dormant community can subsequently become active again. Dormancy is ambiguous and thus the data are right censored when the month *t* is within three months of the end of the data collection period. [14].

*Independent variables*

- **Membership overlap.** We consider two communities as sharing a member if the member made revisions to both communities in a given month. Members who made revisions to more than 10 communities simultaneously (in any given month) are excluded because they are often either Wikia administrators or non-human software agents (i.e., "bots"). The percentage of these users is 0.2%. We used the same membership overlap measurement as Wang et al. [26]. They first counted the number of members that the focal community shared with another community (i.e., the amount of overlap between two communities). Then, they calculated the sum of the overlap between the focal community and all the other intersecting communities. Finally, they calculated membership overlap by dividing this sum by the focal community size (see formula (1)). This is equivalent to calculating the mean shared membership per focal community member (see formula (2)). This measure considers both the proportion of members who participate in multiple communities and the number of other communities they participate in.
- **Mature intersecting communities overlap.** This variable is used to measure the degree of overlap with mature intersecting communities, based on a median split of community age. That is, it is the average number of mature communities a member belongs to per focal community member. Specifically, formula (1) is adjusted so that *number of shared members* is added only when *community j* is mature. A mature community is one that has existed for at least 10 months, which is the median community age.
- **Young intersecting communities overlap.** This variable is used to measure the degree of overlap with young intersecting communities (communities younger than 10 months). To calculate this variable, formula (1) is adjusted so that *number of shared members* is added only when *community j* is less than 10 months old.
- **Mature focal community overlap.** We differentiate whether the focal community is mature or not. When the focal community is younger than 10 months, this measure

$$Membership\ overlap_{community\ i} = \frac{\sum_{Any\ community\ j\ (j \neq i)} Number\ of\ shared\ members\ between\ i, j}{Number\ of\ members\ in\ community\ \ i} \quad (1)$$

$$= \frac{\sum_{Any\ member\ M\ in\ community\ i} Number\ of\ other\ communities\ \ M\ participating\ in}{Number\ of\ members\ in\ community\ \ i} \quad (2)$$

is zero. When the focal community is at least 10 months old, this variable is equal to membership overlap.

- **Young focal community overlap.** We differentiate whether the focal community is young or not. When the focal community is 10 months or older, this measure is zero. When the focal community is less than 10 months old, this variable is equal to membership overlap.
- **Mature intersecting x mature focal, mature intersecting x young focal, young intersecting x mature focal, and young intersecting x young focal.** These four variables are intended to investigate interaction between the maturity of the focal community and its intersecting communities.
- **Core in intersecting communities overlap.** We calculate this measure by focusing on shared members who are core members in the intersecting communities. We define core members as those in the top 25% of degree centrality in the co-authorship network. We define co-author relationship as editing the same community page in the same period of time (a month) at least once prior to the given month. Note that this definition of core members does not make much sense if the community size is too small. Therefore, we only define people who are top 25% degree centrality in communities with at least eight numbers as core members. Otherwise, they are peripheral members. To calculate this measure, formula (1) was adjusted so it included only the number of shared members who were core in the intersecting communities.
- **Peripheral in intersecting communities overlap.** Similarly, we calculate this measure by focusing on the shared members who are peripheral members in the intersecting communities (i.e., in the bottom 75% of the degree centrality distribution or in communities smaller than eight).
- **Core in focal community overlap.** Similarly, we calculated this measure by focusing on the shared members who were core in the focal community (i.e., in the top 25% of the degree centrality distribution in focal communities with at least eight membership).
- **Peripheral in focal community overlap.** Similarly, we calculated this measure by focusing on the shared members who are peripheral members in the focal community (i.e., in the bottom 75% of the degree centrality distribution in the focal community or in focal communities with at least eight membership).
- **Core in intersecting x core in focal, core in intersecting x peripheral in focal, peripheral in intersecting x core in focal, and peripheral in intersecting x peripheral in focal.** These four variables are designed to test the interaction effects of members' roles in intersecting communities and focal community.

*Control variables*

- **Number of members.** This variable is the number of members who made revisions to any page (including discussion pages) in the community in the given month.
- **Amount of activity.** This variable is the number of total revisions that members made to the articles in the community in the given month.
- **Wikia staff.** This variable indicates the number of Wikia administrators who made revisions to the articles in the community in the given month.
- **ln(t).** This variable represents the baseline hazard function, where t denotes the month.

Note that all the independent variables and *number of members* and *amount of activity* were log transformed in the analysis to reduce non-normality in the data. Because the *number of articles* was highly correlated with *number of members* and *amount of activity*, we did not include it in the analysis.

## RESULTS

Table 2 shows the descriptive statistics. The mean of *community dormancy* is 0.13, which means that on average in any given month 13% communities have been inactive for at least three months. The mean of *membership overlap* in all the communities is 1.13, indicating that, on average in any given month members in a community tend to participate in one other community.

### Interpreting the Results

Tables 3-5 show the results of survival analysis, reporting hazard ratios and their 95% confidence intervals. A hazard ratio is the ratio of the risk of a community becoming dormant in a given month-long period associated with a one unit change in the explanatory variables. A hazard ratio smaller than 1 indicates the decreased rate of dormancy (i.e., increased survival rate), while a hazard ratio larger than 1 indicates the increased rate of dormancy (i.e., decreased survival rate).

| | Mean | S.D. |
|---|---|---|
| **Variables internal to the community** | | |
| Community dormancy | **0.13** | **0.34** |
| Number of members | **17.69** | **141.56** |
| Amount of activity | **508.91** | **2983.4** |
| Wikia staff | **0.83** | **1.90** |
| **Membership overlap variables** | | |
| Membership overlap | **1.13** | **1.51** |
| Mature intersecting communities overlap | **0.80** | **1.15** |
| Young intersecting communities overlap | **0.33** | **0.74** |
| Mature focal community overlap | **0.48** | **1.07** |
| Young focal community overlap | **0.65** | **1.33** |
| Core in intersecting communities overlap | **0.20** | **0.44** |
| Peripheral in intersecting comm. overlap | **0.93** | **1.31** |
| Core in focal community overlap | **0.03** | **0.10** |
| Peripheral in focal community overlap | **1.10** | **1.51** |
| N=5673 communities, 37665 community-month observations | | |

**Table 2. Descriptive Statistics**

### Testing Hypothesis 1: Effects of membership overlap

Table 3 test hypothesis 1, i.e., a curvilinear relationship between membership overlap and community survival. The analysis tested both linear and quadratic terms for membership overlap. We see that the hazard ratio of linear term of membership overlap is significantly smaller than 1 (H.R. = .922, 95% C.I. is [.869, .978], p<0.01), which shows that as membership overlap increases so does community survival. A community where members are on average also members of one other community is 7.8% more likely to be active in a typical month than a community where members do not belong to any other communities. Figure 1 show this result graphically. We divided the community-month observations into two equal-sized groups, those with high and membership overlap, and plotted community survival separately for each group. Communities with high levels of membership overlap are more likely to survive, compared with communities with low levels of membership overlap[1]. However, the hazard ratio for quadratic term is not significant (H.R. = 1.06, 95% C.I. is [.980, 1.14]), indicating that community survival is not highest at intermediate values of membership overlap. Therefore, the curvilinear effects are not confirmed.

### Testing Hypothesis 2: Moderating effects of the maturity of the communities

Table 4 shows analysis testing the moderating effects of the maturity of the communities. Model 1 in Table 4 examines two types of membership overlap: overlap with mature communities (i.e., *mature intersecting communities overlap*) and overlap with young communities (i.e., *young intersecting communities overlap*). We can see that the hazard ratio of *mature intersecting communities overlap* is significantly smaller than 1 (H.R. = .880, p<0.01) while the hazard ratio of *young intersecting communities overlap* is significantly larger than 1 (H.R. = 1.20, p<0.01). The results suggest that overlapping with mature communities is beneficial but overlapping with young communities is harmful. In Model 2, we examine the influence of membership overlap on two types of focal community: young and mature. Young communities tend to benefit from membership overlap (H.R. = .861, p<0.01) while mature communities do not (H.R. = 1.18, p<0.01). Model 3 shows the interaction between the types of focal communities and the types of intersecting communities. Membership overlap is most beneficial when young focal communities are overlapping with other mature intersecting communities (H.R. = .794, p<0.01), and membership overlap is least beneficial when mature focal communities are sharing

[1] Note that the survival curves in Figure 1 (in Figure 2-3 as well) are not monotonically decreasing. The reason is that, although the estimated survival rate of any single community is monotonically decreasing as ages, each community may have different levels of membership overlap at different months. The data in Figure 1 is on community-month level rather than on community level, so the curves are not strictly monotonic.

| Explanatory variables | Hazard Ratio (H.R.) | [95% Conf. Interval] |
|---|---|---|
| Membership Overlap | .922** | [.869 .978] |
| Quadratic Term for Membership Overlap | 1.06 | [.980, 1.14] |
| Number of members | .229** | [.202, .260] |
| Amount of activity | .704** | [.697, .722] |
| Wikia staff | .847** | [.816, .880] |
| Ln(t): baseline hazard function | .690** | [.673, .708] |

Log likelihood = -11571.206 ** p<0.01, *p<0.05

**Table 3: Predicting the effects of membership overlap on survival (Hypothesis 1)**

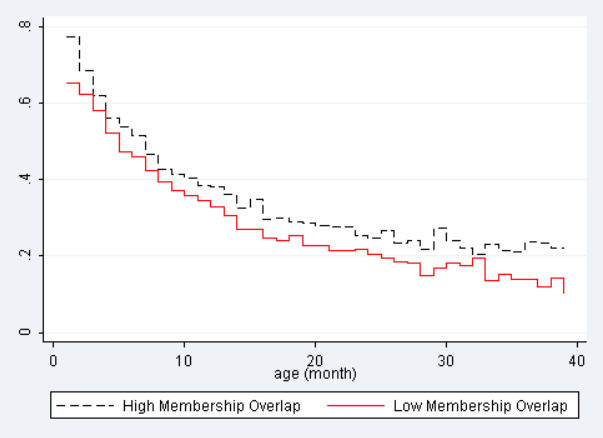


**Figure 1. Average survival rate for communities with different levels of membership overlap.** *This visualization corresponds to the results in Table 3.*

members with young intersecting communities (H.R. = 1.45, p<0.01). In sum, we found broad support for hypothesis 2.

We show the effects of different types of intersecting communities visually in Figure 2. We divide the observations into two buckets: high and low *mature intersecting communities overlap*. In the visualization, we

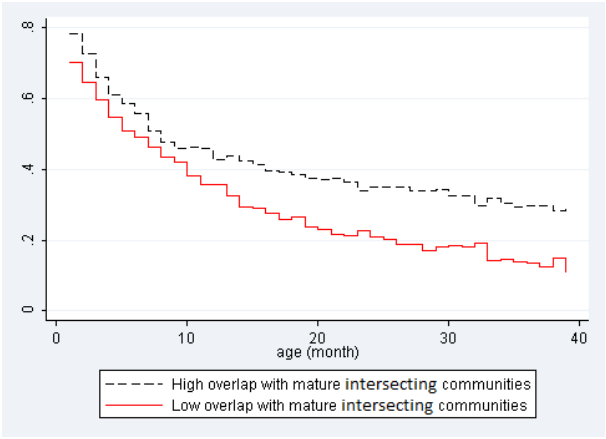


**Figure 2. Average survival rate for communities with different levels of overlap with mature intersecting communities.** *This visualization corresponds to Model 1 in Table 4*

| Explanatory variables | Model 1 H.R. | Model 1 [95% CI] | Model 2 H.R. | Model 2 [95% CI] | Model 3 H.R. | Model 3 [95% CI] |
|---|---|---|---|---|---|---|
| Mature intersecting communities overlap | .880** | [.831, .931] | | | | |
| Young intersecting communities overlap | 1.20** | [1.03, 1.18] | | | | |
| Mature focal community overlap | | | 1.18** | [1.09, 1.26] | | |
| Young focal community overlap | | | .861** | [.816, .908] | | |
| Mature intersecting x mature focal | | | | | 1.02 | [.923, 1.12] |
| Mature intersecting x young focal | | | | | .794** | [.740, .851] |
| Young intersecting x mature focal | | | | | 1.45** | [1.26, 1.69] |
| Young intersecting x young focal | | | | | 1.03 | [.955, 1.11] |
| Number of members | .224** | [.198, .254] | .228** | [.201, .258] | .225** | [.198, .255] |
| Amount of activity | .705** | [.688, .723] | .704** | [.687, .722] | .706** | [.689, .724] |
| Wikia staff | .853** | [.822, . 886] | .845** | [.814, .877] | .854** | [.823,. 888] |
| Ln(t): baseline hazard function | .697** | [.679, .715] | .653** | [.634, .673] | .658** | [.639, .678] |

Log likelihood = -11533.524 ** p<0.01, *p<0.05

**Table 4: The moderating effects of tenure of communities (Hypothesis 2)**

| Explanatory variables | Model 1 H.R. | Model 1 [95% CI] | Model 2 H.R. | Model 2 [95% CI] | Model 3 H.R. | Model 3 [95% CI] |
|---|---|---|---|---|---|---|
| Core in intersecting communities | .755** | [.678, .840] | | | | |
| Peripheral in intersecting communities | 1.03 | [.977, 1.08] | | | | |
| Core in focal community | | | 2.14 | [.203, 22.5] | | |
| Peripheral in focal community | | | .949* | [.907, .992] | | |
| Core in intersecting x core in focal | | | | | 804** | [14.7, 44000] |
| Core in intersecting x peripheral in focal | | | | | .754** | [.677, .839] |
| Peripheral in intersecting x core in focal | | | | | .017 | [.000, 3.51] |
| Peripheral in intersecting x peripheral in focal | | | | | 1.03 | [.978, 1.08] |
| Number of members | .229** | [.202, .259] | .226** | [.199, .256] | .227** | [.200, .258] |
| Amount of activity | .703** | [.686, .721] | .704** | [.687, .721] | .703** | [.686, .721] |
| Wikia staff | .845** | [.813, . 877] | .845** | [.814, .877] | .844** | [.813,. 877] |
| Ln(t): baseline hazard function | .696** | [.678, .714] | .691** | [.674, .709] | .696** | [.678, .714] |

Log likelihood = -11557.379 ** p<0.01, *p<0.05

**Table 5: The moderating effects of roles of shared members (Hypothesis 3)**

can see that communities with high overlap with mature communities are more likely to survive. We do not include a graph comparing mature and young focal communities because it is difficult to visualize the influence of membership overlap on different age periods using survival curves.

### Testing Hypothesis 3: Moderating effects of the roles of the shared members

Table 5 shows the results of the moderating effects of roles of shared members in focal communities and intersecting communities. Model 1 shows that a community where members are on average also core members of one other community is 24.5% more likely to be active in a typical month than a community where shared members are not core in any other communities. (H.R. = .755, p<0.01). In contrast, they gain no benefit from sharing members who are peripheral members in intersecting communities (H.R. = 1.03, 95% C.I. is [.977, 1.08]). Model 2 suggests communities are more likely to be active if they share their peripheral members with other communities are beneficial for the focal communities (H.R. = .949, p<0.01). However, they get no benefit from sharing their core members (H.R. = 2.14, 95% C.I. is [.203, 22.5]). In contrast, Model 3 shows that shared members who are both core members in focal community and intersecting communities are associated with significant decrease in the likelihood of survival of focal community (H.R. = 804, p<0.01). Note that the hazard ratio and its value for *core in intersecting x core in focal* is large, probably because it is rare in the dataset for shared members to be core in both the focal and intersecting communities. Communities are most likely to be active when they have shared members who are peripheral members in focal community and core members in intersecting communities (H.R. = .754, p<0.01).

We draw survival curves to show the results graphically. Figure 3 shows that communities with their core members participating in other communities are less likely to survive,

compared to those communities with fewer core members participating in other communities. In sum, we found support for hypothesis 3.

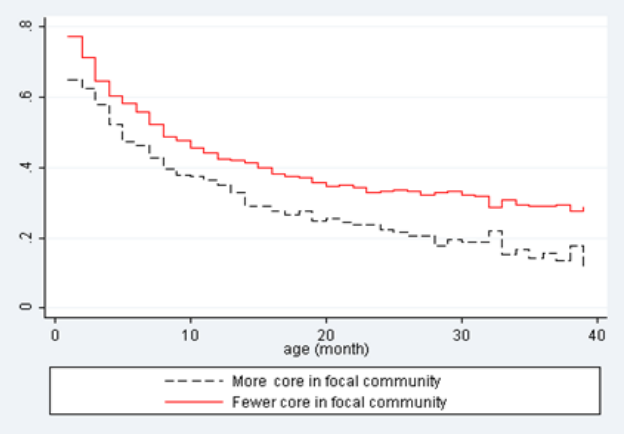


**Figure 3. Average survival rate for communities varying *core in focal community* (i.e., shared members who are core members in focal community).** *This visualization corresponds to Model 2 in Table 5*

## DISCUSSION

This paper examined the effects of membership overlap on the survival of online communities. With archival data from 5673 Wikia communities, we found that 1) higher levels of membership overlap are associated with increased community activity; 2) the beneficial effects of membership overlap are especially strong when the focal community was young and the intersecting communities were mature; 3) membership overlap increases the chances of survival more when the shared members are core members in the intersecting communities but reduces the chance of survival when the shared members are core members in the focal community.

Although we predicted that membership overlap should have a curvilinear effect on community survival, our results only confirmed the linearly positive relationship (see Table 3). Our results contrast with those of Wang et al [26], who found a negative relationship between membership overlap and community growth for Usenet groups. The reason of these different findings might be that membership overlap was much higher in the Usenet groups that Wang et al. studied, with Usenet group members participating in 7.56 additional groups, compared to the Wikia communities we studied, where members participated in 1.13 additional communities on average. It is possible that that the overall effects of membership overlap on the survival rate are indeed curvilinear as hypothesized, but the current study and Wang et al.'s study of Usenet groups were studying different locations in the membership overlaps distribution.

Our results have guidance for community practitioners. The proliferation of communities that exist on the Internet brings in uncertainty to community managers and creators. Our results show that communities can potentially benefit from other communities in the environment. Specifically, in the communities we studied, the beneficial effects of membership overlap (i.e., learning, knowledge sharing, diverse perspectives and new member recruiting) outweigh the negative effects (i.e., competition for the members' efforts), resulting in increased capability to survive. To exploit the beneficial effects of membership overlap, community practitioners can design recruiting strategies to specifically target members who have experience in other mature communities, especially those core members in other communities.

This study is also subject to limitations. First, our data analysis provides limited support for understanding *why* the membership overlap is associated with community survival. It would be more convincing if mediating variables which directly relate to membership overlap and the survival rate of community could be included in the analysis. Example mediating variables might include organization or content similarity between communities (which are indicators of learning and knowledge sharing) and diluted members' attention and efforts (which is an indicator of competition). We will investigate these in future research.

Second, our study used community activity and dormancy as a proxy for community success, while in reality success can be measured in many aspects such as quality of deliverables in Wikipedia-like communities and progress towards particular business-oriented goals in enterprise communities. Nonetheless, as activity level is indeed a widely-used measure of community success, we believe our results are still valuable. Future research could extend this work, by incorporating more nuanced success measures as appropriate.

Lastly, we used a homogeneous platform in Wikia. Doing so was important for our research for two reasons: 1) we were able to compare across communities since they shared the same UI and backend; and 2) we were able to track member migration across communities since member identifiers were Wikia-wide. However, one caution in generalizing from this homogeneous system is that knowledge, experience, and human capital may be easier to transfer among similar types of organizations or projects than they would be in more heterogeneous environments of communities. We would like to examine communities with different UIs and affordances in future research in order to understand how these findings are similar or different in heterogeneous communities.

## CONCLUSION

Online communities play an important role in society. In this study, we study the effects of membership overlap on the survival of online communities. These findings provide new insight into an important mechanism underlying successful online communities and practical implication for the hosts and creators of online communities.

**ACKNOWLEDGMENTS**

This work was supported by NSF grants IIS-0808711, OCI-0943148, IIS-0968484, IIS-1111124, and Google.

**REFERENCES**

1. Baum, J. A., & Shipilov, A. V. (2006). *1.2 Ecological Approaches to Organizations.* The SAGE Handbook of Organization Studies, 55.
2. Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proc. SIGGROUP'2005*. ACM Press.
3. Burke, M., & Kraut, R. (2008). Mopping up: modeling wikipedia promotion decisions. In *Proc. CSCW'2008*. ACM Press.
4. Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. In *Proc. CHI'2010.* ACM Press.
5. Burt, R.S. (1987) Social contagion and innovation: Cohesion versus structural equivalence, *American Journal of Sociology*, 92, 1287-1335.
6. Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research*, 12(4).
7. Chen, J., Ren, Y., & Riedl, J. (2010). The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proc. CHI'2010*.ACM Press.
8. Choi, B., Alexander, K., Kraut, R. E., & Levine, J. M. (2010, February). *Socialization tactics in Wikipedia and their effects*. In *Proc.* CSCW'2010. ACM Press.
9. Hill, B. M. and Shaw, A. (Forthcoming) "Is volunteer labor a "fixed and finite" resource? Evidence from peer production." Working Paper.
10. Hogg, M. A., and J. C. Turner 1985 'Interpersonal attraction, social identification and psychological group formation'. *European Jour of Social Psychology* 15: 51–66.
11. International Telecommunications Union (2013) Key ICT indicators for developed and developing countries and the world, 04 Sept 2013.
12. Iriberri, A. and Leroy, G. A life-cycle perspective on online community success. *ACM Comput. Surv*. 41, 2 (2009), 1–29.
13. Ingram, P. and Baum, J.A.C. (1997) 'Opportunity and constraint: Organizations' learning from the operating and competitive experience of industries', *Strategic Management Journa*l, 18: 75–98.
14. Jenkins, S. P. (2005). Survival analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.
15. Kairam, S. R., Wang, D. J., & Leskovec, J. (2012). The life and death of online groups: predicting group growth and longevity. In *Proc. WSDM'2012*. ACM Press.
16. Lakhani, K., and Wolf, R. "Why hackers do what they do: Understanding motivation and effort in free/open source software projects." (2003).
17. Lerner, Josh, and Jean Tirole. (2005). The economics of technology sharing: Open source and beyond. *Journal of Economic Perspectives.* 19 (2): 99-120.
18. Muller, M., Ehrlich, K., Matthews, et al. (2012) Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. In *Proc. CHI '12*, 2815-24. ACM Press.
19. Nov, O. (2007). What motivates wikipedians?. *Communications of the ACM*, 50(11), 60-64.
20. O'leary, M. B., Mortensen, M., & Woolley, A. W. (2011). Multiple team membership: A theoretical model of its effects on productivity and learning for individuals and teams. *Academy of Management Review*, 36(3), 461-478.
21. O'Mahony, & Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal,* 50(5), 1079-1106.
22. Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, 28(3), 377-408.
23. Resnick, P., Konstan, J., Chen, Y., & Kraut, R. E. (2012). 6 Starting New Online Communities. *Building successful online communities: Evidence-based social design*, Cambridge, MA: MIT Press.
24. Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis.* New York: Oxford University Press.
25. Stewart, K. J., & Gosain, S. (2006). The impact of ideology on effectiveness in open source software development teams. *MIS Quarterly*, 291-314.
26. Wang, X., B. Butler, Y. Ren. (2013). The impact of membership overlap on the growth: An ecological competition view of online groups. *Organization Science*. 24(2), 414-431.
27. Weber, Steven. 2004. *The success of open source.* Cambridge, MA: Harvard University Press.
28. Yu, L., and Nickerson, J. V. (2011) Cooks or cobblers?: crowd creativity through combination. In *Proc. CHI'2011*.ACM Press.
29. Zhu, H., Kraut, R., & Kittur, A. (2012). Effectiveness of shared leadership in online communities. In *Proc. CSCW'2012*. ACM Press.