

The Impact of Membership Overlap on the Survival of Online Communities

Completed Research Paper

Haiyi Zhu

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15217
haiyiz@cs.cmu.edu

Robert E. Kraut

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15217
robert.kraut@cmu.edu

Aniket Kittur

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15217
nkittur@cs.cmu.edu

Abstract

Online communities play an important role in society. 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 greater survival rate of online communities relative to lower levels of membership overlap. Furthermore, we find that it is beneficial for new communities to have shared members who play a central role in other mature communities. These findings provide new insight into an important mechanism underlying successful online communities, contribute to theories of organization science, and provide several actionable steps for the hosts and creators of online communities.

Keywords: Membership overlap, online communities, survival analysis

Introduction

Online communities, which provide virtual spaces where people around the world can interact around a shared purpose, have become increasingly important for both work and social relationship. Peer production communities (Benkler 2006) such as Wikipedia or open source software projects aggregate the efforts of volunteers to produce complex artifacts such as the largest encyclopedia in human history or the software that powers the Internet. Communities such as health support groups enhance opportunities for interpersonal relationships that span geographic boundaries. Business use online communities to fulfill different organizational goals, such as facilitating team work, providing technical support, brainstorming innovative ideas, connecting employees, and facilitating peer-to-peer customer support (e.g., Muller et al. 2012; Jeppesen and Frederiksen 2006). Development of Internet technologies has significantly reduced the costs of creating online communities on various platforms. For example, the venerable Usenet (now accessible on the web via Google Groups) had over 189,000 active newsgroups as of 2005 (Wang et al. 2012a). Facebook provides infrastructure to host online groups, and there are more than 100,000 new Facebook groups created a day (Kraut et al. Forthcoming). There is also tremendous amount of enterprise communities for the internal collaboration and learning among employees inside companies. For example, Matthews et al. reports that there are more than 111,557 IBM connection communities supporting the usage of 487,941 IBM employees (Matthews et al. 2013).

Because so many online communities exist, no single one is likely to monopolize its members' time. The growth of Internet users in developed countries has slowed over the last few years, while the number of Internet offerings – sites to visit and groups to join – is still growing exponentially (International Telecommunications Union 2012). If the 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 reliance 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, 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., Kraut & Resnick 2012, Lakhani and Wolf 2003, O'Mahony& Ferraro 2007), very few of them have investigated how the relationship with other communities, such as membership overlap, can influence the success of online communities. Wang et al. has conducted a relevant and interesting study about membership overlap on member growth in Usenet groups (Wang et al. 2012a). However, they only focus on the detrimental effects of membership overlap. To the best of our knowledge, there is no research studying 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 the panel data from Wikia, a technical platform that supports Wikipedia-like online communities. For example, there are Wikia communities organized around topics like movies such as Star Wars, video games such as the World of Warcraft, and lifestyles such as healthy recipes. Our data includes archival data of 5673 communities from their inception to 2008. Our main finding is that higher levels of membership overlap is positively associated with the survival of online communities compared to lower levels of membership overlap. Furthermore, the beneficial effects of membership overlap on the survival of focal community are stronger when 1) the focal community is a new community; 2) the intersecting community with which the focal community share members is mature; and 3) the shared members are core members in the intersecting communities. However, membership overlap is negatively associated with the survival rate when shared members are core members in the focal community.

The contributions of this work are two-fold. First, we examine how the membership overlap with other communities influences the survival of the focal community, which provides new insight into one of the main mechanisms underlying successful online communities. Second, our findings provide practical guidance for the hosts and creators of online communities. The proliferation of communities that already exist on the Internet brings in opportunities and challenges to businesses and individuals who are interested in launching new communities. Our findings suggest that community creators cannot only spend time on the internal design of the community, but need to consider the impact of the larger ecosystem of the 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. The 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., Hertel et al. 2003, Weber 2004, Nov 2007). Weber (2004) and Lerner and Tirole (2005) 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 he or she will receive in return for his or her efforts on the community. The benefits include having enjoyment and fun (Lakhani and Wolf 2003, Nov 2007), pursuing beliefs and values shared with other people (Steward& Gosain 2006), expressing humanitarian concerns for others (Nov 2007), learning (Palloff and Pratt 1999), developing careers (Lakhani and Wolf 2003), and protecting oneself from negative emotions and enhancing positive attitudes (Burke et al. 2010, Wang et al. 2012b). One implication of this group 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. There are two main categories 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 (Chen et al 2010) and Butler's work on membership size and communication activity (Butler 2001). Examples of research examining structural characteristics include Kairam et al's work on members' social ties (Kairam et al. 2012), Zhu et al's work on shared leadership (Zhu et al. 2012), Choi et al's work on socialization (Choi et al. 2010), and O'Mahony and Ferraro's work on governance (O'Mahony and Ferraro 2007). The assumption of this group of research is that communities can achieve continued success 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 uses an ecological view. Most 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. Unlike the first two types of research, research on inter-community level has been quite neglected. The only relevant research is Wang et al's work about membership overlap on the growth of online groups (Wang et al. 2012a). Wang et al took a competition view of membership overlap. They argue that an individual's time is scarce. When multiple online communities are relying on the participation of the same members, the members' time spent on one community will take time away from another community, thus reducing the chance of survival for both communities.

However, Wang et al. (2012a) 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 a multiplicity of strategies, practices and technologies employed by their "competitors" (Miner and Haunschild 1995, Ingram and Baum 1997, Baum and Shipilov 2006). For example, Ingram and Baum (1997) found that a hotel chain's survival rates is positively related to the total operating experience other US hotel chains had accumulated. Moreover, organizational behavior researchers (e.g., O'Leary et al. 2011) argue that multiple team membership (i.e., membership overlap in work teams) can have positive effects on team productivity and team learning. Specifically, increased number and variety of multiple team membership can improve efficiency and increase diversity, thus benefiting team productivity and learning. The first two pieces of evidence are based on research in offline organizations and groups, but the mechanisms are likely to be applicable to online communities. The third piece of evidence is directly relevant to online communities. Hill and Shaw (Forthcoming) have challenged the assumption that competition between projects is an important dynamic driving contribution to online communities. Their analysis showed that the amount of contributions to pages in topic areas within Wikipedia is positively correlated to the amount of contributions on other encyclopedia wikis on overlapping topics run by Wikia. 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. 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 understanding the conditions under which the beneficial effects of membership overlap is stronger.

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 the classic social network and bridging social capital theory (Granovetter 1973, Bourdieu 1985, and Burt 1992), people who participated 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,

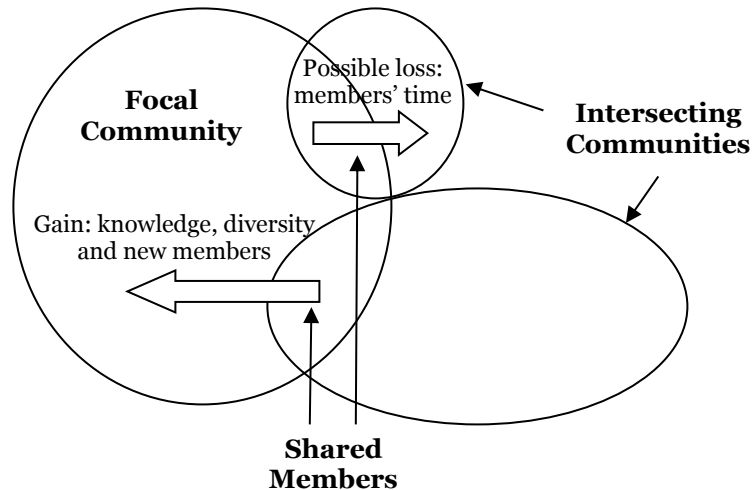


Figure 1. Illustrate the relationships between communities which share members

members learn basic technical skills (e.g., using editing tools in Wiki-like websites and using version control tools in open sourcing projects), 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) (Bryant et al. 2005). 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 (O’Leary et al. 2011). Research shows that a moderate level of diversity can increase productivity and decrease member withdrawal in online communities (Chen et al. 2010). Therefore, a moderate level of membership overlap may positively affect the survival of online communities through increased diversity. Third, according to social network theories (Kairam, et al. 2012), people are more likely to participate in a community if people in their social networks are already in the community. Therefore, members participating in multiple communities might make it more likely that their friends in one community will also become members in the other community, thus benefiting the growth of both communities.

At the same time, high levels of membership overlap are likely to impede online communities’ active and healthy functioning for three reasons. By high levels of membership overlap, 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 a limit on individual’s 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 (Ren et al. 2007). 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 (Tajfel et al.1971; Hogg and Turner 1985). However, as membership overlap becomes high, the boundaries between communities become ambiguous and less important, which lowers people’s identification with a certain community. With lowered group identification, people are less likely to participate, leading to decreased survival rate of online communities. 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 (2010) 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 the survival of online communities, but very low or very high levels of membership overlap impede the survival of online communities.*

The beneficial effects of membership overlap on the survival of focal community might be moderated by the maturity of both the focal community and the 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 enhances diversity. Third, mature communities often have more members for the focal community to recruit. In sum, member who participate in mature communities are more likely to acquire necessary knowledge and experiences, gain diverse perspective, and have the opportunities to recruit more people, which in turn is more likely to benefit the survival of other communities the member simultaneously participates in.

Moreover, we hypothesize that the beneficial effects of membership overlap are stronger when the focal communities are new. New online communities are fragile and many of them never really get off the ground. For example, SourceForge hosts over 300,000 software development projects, but 90% have fewer than four members (Resnick et. al. 2012, p. 231). IBM reports that well over 50% of its in-house, online communities intended to connect employees or to connect employees to clients fail; they have no activity within three months of formation. New 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 new communities most since they can bring in technical skills, community building experience and human resources which are crucial to the survival of new online communities.

Hypothesis 2a. The beneficial effects of membership overlap are stronger when the intersecting communities (i.e., the ones which the focal community overlaps with) are mature.

Hypothesis 2b. The beneficial effects of membership overlap are stronger when the focal community is new.

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 (Bryant et al. 2005, Preece and Shneiderman 2009). 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 (Bryant et al. 2005). 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 tasks. In Wikipedia, administrators made 5010 revisions (a measure of contributions) on average (Burke and Kraut 2008), 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 communities.

Hypothesis 3a. The beneficial effects of membership overlap are stronger when the shared members are core members in the intersecting communities.

Hypothesis 3b. The beneficial effects of membership overlap are weakened when the shared members are core members in the focal communities.

Methods

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, science, sports, technology, toys, travels and others. Any user can create articles in wikis. In addition to articles pages, each wiki has dedicated project pages on which users can coordinate and organize the writing and the editing of articles. Wikis also have user personal pages for their users to share information about themselves and interact with each other. Since each wiki has a unique topic, dedicated pages to coordinate activities, and unique places for users to interact with each other, we consider each wiki as an independent online community.

One Wikia account can be used across all its Wikis. 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. Our 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

The purpose of the analysis is to estimate how membership overlap influences the chance of survival for online communities. Because Wikia communities are organized to produce content, we consider a community alive if it is producing content and dead when it stops. We conduct a survival analysis, a statistical technique for modeling time to an event (Singer and Willett 2003). While survival analysis can be used to analyze death in biological organisms (Collett 2003), 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 the discrete time proportional hazard models (Jenkins 2005). The unit of analysis is the community-month. We used $\ln(t)$, where t denotes the month, as the baseline hazard function. We use the following vocabulary in describing the analysis. A *focal community* is the community of interest. *Intersecting communities* are the communities with which the focal community shares members. *Shared members* are the participants who participated in both the focal community and the intersecting communities.

Measurement

Dependent variable

- **Community death.** We define a community as dead in a given month if the community does not have any activity in the subsequent three months. Community death is a binary variable. This variable is assigned to 1 if the community is dead during this month; it is assigned to 0 if the community is still alive in this month. A community can become alive later on after it is pronounced dead. If we cannot determine whether the community is dead or not (especially when the observation is close to the end of the data collection date), community death is assigned to 0, which is a case of right-censoring in the discrete time survival analysis (Jenkins 2005).

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 in any given month are excluded because they are often either Wikia administrators or non-human software systems ("bots"). We used the same membership overlap measurement as Wang et al. (2012a). They first counted the number of members that the focal community shared with another community (i.e., the degree of overlap between two communities). Then, they calculated the sum of the degree of overlap between the focal community and each of the other intersecting communities. Finally, they calculate membership overlap by dividing the sum by the focal community size (see formula (1)). This is equivalent to 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.

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

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

- **Mature intersecting communities overlap.** This variable is used to measure the degree of overlap with the mature intersecting communities. We calculate this variable in a similar way to calculating membership overlap except that we only count the mature communities which the focal community has shared members with. Specifically, formula (1) is adjusted so that *number of shared members* is added only when *community j* is mature. Mature communities are defined as communities no younger than 10 months (≥ 10 months). Ten months is the median of community age in the dataset.
- **New intersecting communities overlap.** This variable is used to measure the degree of overlap with new 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 new.
- **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 is assigned zero. When the focal community is 10 months or older, this variable is equal to membership overlap.
- **New focal community overlap.** We differentiate whether the focal community is new or not. When the focal community is 10 months or older, this measure is assigned zero. When the focal community is younger than 10 months, this variable is equal to membership overlap.
- **Mature intersecting x mature focal, mature intersecting x new focal, new intersecting x mature focal, and new intersecting x new focal.** These four variables are intended to investigate the interaction effects of maturity of focal community and the intersecting communities. *Mature intersecting x mature focal* indicates the degree of overlap with mature intersecting communities when the focal community is also mature. Therefore, *mature intersecting x mature focal* is equal to *mature intersecting community overlap* when the focal community is mature; it is equal to zero when the focal community is new. We calculated the remaining three variables (i.e., *mature intersecting x new focal*, *new intersecting x mature focal*, and *new intersecting x new focal*) similarly.
- **Core in intersecting communities overlap.** We calculate this measure by focusing on the shared members who are core members in the intersecting communities. We define core members as those top 25% in terms of degree centrality in the co-authorship network. By co-authorship, we mean the people who edit the same community page in a period of time (a month) are co-authors. Note that “core members” do 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 no less than eight numbers as core members. Otherwise, they are peripheral members. To calculate this measure, formula (1) should be adjusted so that only the number of shared members who are core members in the intersecting communities is added.
- **Peripheral in intersecting communities overlap.** Similarly, we calculate this measure by focusing on the shared members who are peripheral members in the intersecting communities.
- **Core in focal community overlap.** Similarly, we calculate this measure by focusing on the shared members who are core members in the focal community.
- **Peripheral in focal community overlap.** Similarly, we calculate this measure by focusing on the shared members who are peripheral members in the focal community.

- **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 intended to investigate the interaction effects of members' roles in intersecting communities and focal community. To calculate *core in intersecting x core in focal*, the formula (1) is adjusted to only account shared members who are core members in intersecting communities and at the same time core members in the focal community. We calculated the other three variables similarly.

Control variables

- **Number of members.** This variable is the number of members who made revisions to the articles 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 represents the baseline hazard function, where t denotes the month.

Note that all the independent variables and *number of members* and *amount of activity* are log transformed in the analysis.

Results

Descriptive statistics are shown in Table 1. The mean of *community death* is 0.13, which means that 13% communities die per month. The mean of *membership overlap* in all the communities is 1.13, indicating that, on average members in any community tend to at least participate in one other community simultaneously in any given month. Then, to examine the moderating effects of the maturity of intersecting communities, we split the *membership overlap* into two variables: *mature intersecting communities overlap* (i.e., the degree of membership overlap with mature communities) and *new intersecting communities overlap* (i.e., the degree of membership overlap with new communities), with mean values as 0.80 and 0.33 respectively (the sum is equal to 1.13). Similarly, we split the *membership overlap* into the remaining pairs to examine the effects of maturity of focal community, roles of shared members in focal community and in intersecting communities.

Table 1: Descriptive Statistics of Main Variables

Variable Name	N	Mean	S.D.
Community death	37665	0.13	0.34
Number of members	37665	17.69	141.56
Amount of activity	37665	508.91	2983.38
Wikia staff	37665	0.83	1.90
Membership overlap	37665	1.13	1.51
Mature intersecting communities overlap	37665	0.80	1.15
New intersecting communities overlap	37665	0.33	0.74
Mature focal community overlap	37665	0.48	1.07
New focal community overlap	37665	0.65	1.33
Core in intersecting communities overlap	37665	0.20	0.44
Peripheral in intersecting communities overlap	37665	0.93	1.31
Core in focal community overlap	37665	0.03	0.10
Peripheral in focal community overlap	37665	1.10	1.51

Table 2-4 shows the results of survival analysis. We report the hazard ratios and their 95% confidence interval. Hazard ratio is the ratio of the risk of a community dying 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 death (i.e., increased survival rate), while a hazard ratio larger than 1 indicates the increased rate of death (i.e., decreased survival rate).

Testing Hypothesis 1: Effects of membership overlap

Table 2 shows the results of analysis examining hypothesis 1. In the model we tested both linear and quadratic terms of 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], $p < 0.01$) and the hazard ratio of quadratic term is larger than 1 but not significant (H.R. = 1.06, 95% C.I. is [.980, 1.14]). The results confirm the linear effects of membership overlap on decreasing the death rate. There is some evidence of the curvilinear effects but they are not statistically significant. Therefore, hypothesis 1 is partially supported. Note that the control variables (i.e., *number of members*, *amount of activities*, *Wikia staff* and the baseline hazard function) are all associated with decreased rate of death, which is reasonable.

The results are visualized in Figure 2. We first estimate the survival rate (i.e., the probability to survive) for each community in any given month. Then we divide the community-month observations into two buckets with equal number: high membership overlap and low membership overlap. Figure 2 shows the average survivor rate for different levels of membership overlap. From the figure, we can see that communities with high levels of membership overlap are more likely to survive, compared with communities with low levels of membership overlap. Note that the survival curves in Figure 2 (in Figure 3-5 as well) are not monotonically decreasing as normal Kaplan-Meier survival curves (Kaplan and Meier 1958). 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. For example, average survival rate of high membership overlap observations in month 1 might contain a different set of communities compared to the average survival rate with high membership overlap in month 2. Therefore, curves in Figure 2 are not strictly monotonic.

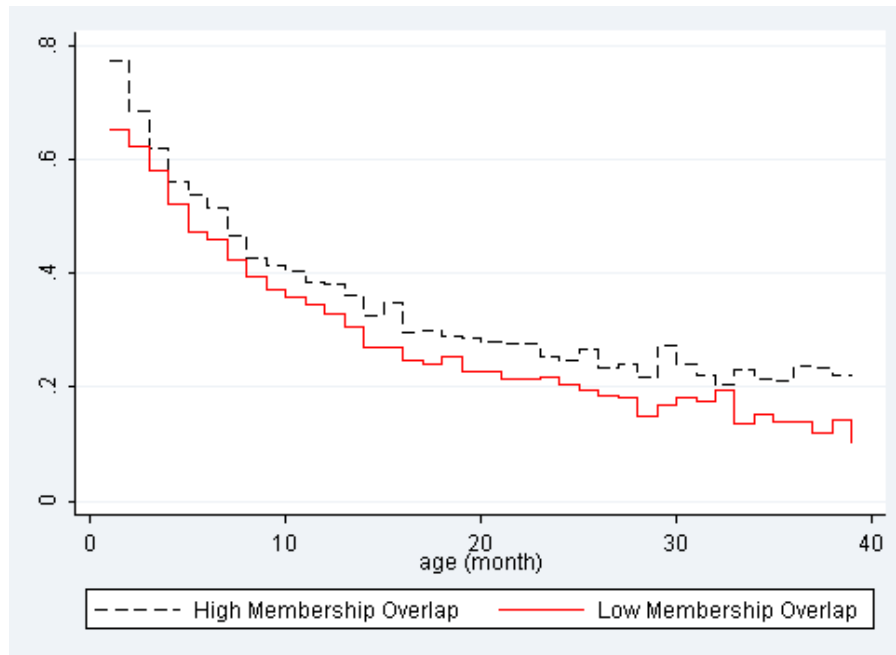


Figure 2. Average survival rate for communities with different levels of membership overlap

Table 2: Predicting the effects of membership overlap on survival (Hypothesis 1)

Explanatory variables	Hazard Ratio (H.R.)	[95% Conf. Interval]
Membership Overlap	.922**	[.869, .978]
Quadratic Term of 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

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

Table 3 shows the results of our analysis testing for the moderating effects of the maturity of the communities. Model 1 in Table 3 examines two types of membership overlap: overlap with mature communities (i.e., *mature intersecting communities overlap*) and overlap with new communities (i.e., *new 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 *new 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 new communities is not. In Model 2, we examine the influence of membership overlap on two types of focal community: new and mature. New 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 new 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 members with new 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 3. As we do in Figure 2, we divide our observations into two buckets: high overlap with mature communities and low overlap with mature communities. In the visualization, we can see that communities with high overlap with mature communities are more likely to survive. We do not include a graph comparing mature focal communities and new focal communities because it is difficult to visualize how the influence of membership overlap changes at different time periods using survival curves.

Table 3: The moderating effects of tenure of communities (Hypothesis 2)

Explanatory variables	Model 1		Model 2		Model 3	
	H.R.	[95% CI]	H.R.	[95% CI]	H.R.	[95% CI]
Mature intersecting communities overlap	.880**	[.831, .931]				
New intersecting communities overlap	1.20**	[1.03, 1.18]				
Mature focal community overlap			1.18**	[1.09, 1.26]		
New focal community overlap			.861**	[.816, .908]		
Mature intersecting x mature focal					1.02	[.923, 1.12]
Mature intersecting x new focal					.794**	[.740, .851]
New intersecting x mature focal					1.45**	[1.26, 1.69]
New intersecting x new 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	

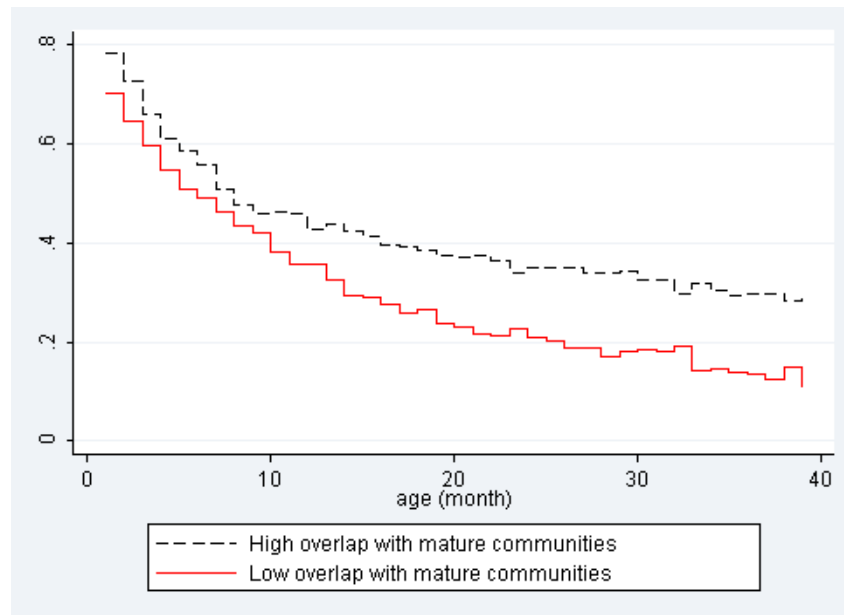


Figure 3. Average survival rate for communities with different levels of membership overlap with mature communities

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

Table 4 shows the results of the moderating effects of roles of shared members in focal communities and intersecting communities. Model 1 shows that it is more beneficial for the focal community if the shared members are core members in the intersecting communities (H.R. = .755, $p < 0.01$), compared to having shared members who are peripheral members in intersecting communities (H.R. = 1.03, 95% C.I. is [.977, 1.08]). Model 2 suggests that more core members in the focal community participating in other communities might harm the survival of focal community but it is not statistically significant (H.R. = 2.14, 95% C.I. is [.203, 22.5]). In contrast, peripheral members participating in other communities are beneficial for the focal communities (H.R. = .949, $p < 0.01$). 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, $p < 0.01$), while shared members who are peripheral members in focal community and core members in intersecting communities are most beneficial (H.R. = .754, $p < 0.01$). Note that the mean value and the variance of hazard ratio of *core in intersecting x core in focal* is large. The reason might be that there is quite low chance of having members who are core members in both communities in the dataset.

We also draw survival curves to show the results graphically. Figures 4 and 5 visualize these results using approach similar to Figure 2 and 3. Figure 4 shows that communities with more shared members who are core members in other communities are more likely to survive than communities with fewer shared members who are core members in other communities. Figure 5 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.

Table 4: The moderating effects of roles of shared members (Hypothesis 3)

Explanatory variables	Model 1		Model 2		Model 3	
	H.R.	[95% CI]	H.R.	[95% CI]	H.R.	[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

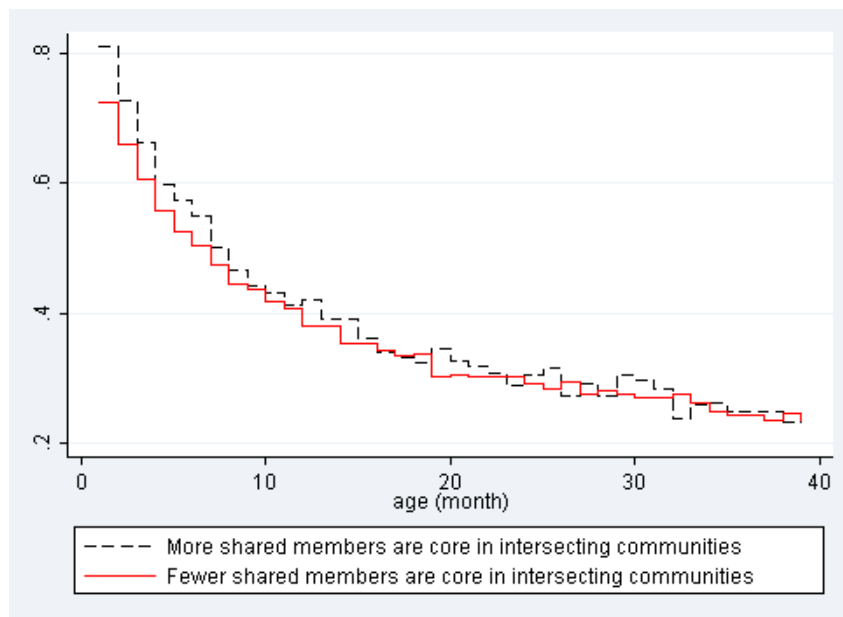


Figure 4. Average survival rate for communities varying the numbers of shared members who are core members in intersecting communities

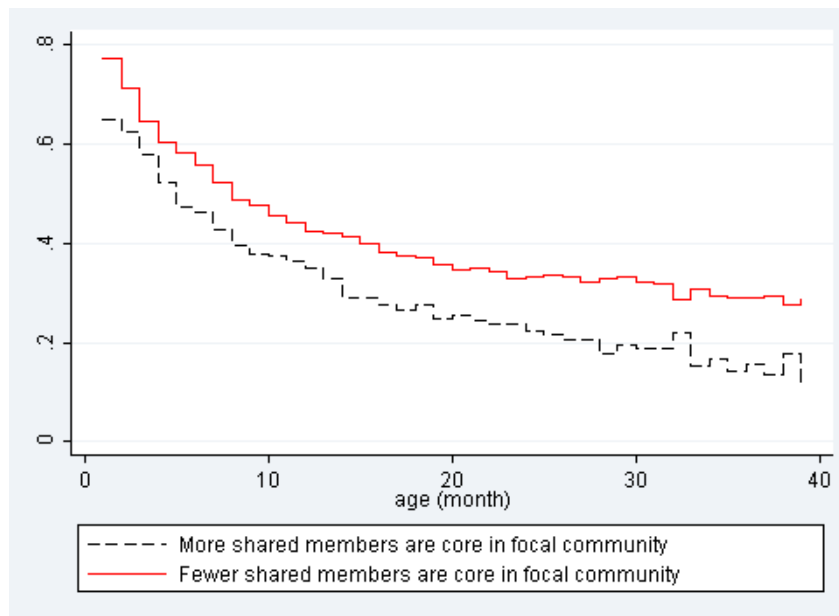


Figure 5. Average survival rate for communities varying the numbers of shared members who are core members in the focal community

Discussion

In this paper, we seek to examine the effects of membership overlap on the survival of online communities. With archival data of 5673 Wikia communities, we found that 1) higher levels of membership overlap are associated with increased chance of survival compared to lower levels of membership overlap; 2) the beneficial effects of membership overlap are especially strong when the focal community is new and the intersecting community is mature; 3) membership overlap increases the chances of survival when the shared members happen to be core members in the intersecting communities but decreases the chance of survival when the shared members are core people in the focal community.

Our study shows that membership overlap has a linearly positive relationship with community survival for the average Wikia community in our dataset (see Table 2), while Wang et al (2012a) showed a negative relationship between membership overlap and community growth for average Usenet groups. The reason might be that Wang et al. was investigating a more “crowded” population which membership overlap falls in the higher end of the spectrum (the mean of membership overlap is 7.56), while Wikia communities we studied were less crowded and in the lower end of the spectrum (the mean of membership overlap is 1.13). Therefore, we believe that the overall effects of membership overlap on the survival rate should still be curvilinear as we hypothesized. Furthermore, the Usenet population studied by Wang et al. was a mature population and those groups had 20 years of history prior to the study. The Wikia communities we investigated were relatively new communities. The oldest Wikia community in our dataset was seven years old by the end of data collection and the median age was 10 months old. Mature communities often already have rich operating experiences and large participant pools, so the potential learning benefits from other community are limited while the competition with other communities may still be prominent. Model 2 in Table 3 suggests that the impact of membership overlap for an average mature community’ survival is negative while the impact of membership overlap for an average young community is positive, which might also contribute to the difference between the findings of our paper and Wang et al’s study (2012a).

Contribution

As organizations and individuals increasingly leverage online communities for work support and relationship maintenance, the online community space becomes more and more crowded. Understanding

the implication of such crowding is critical to building vibrant communities (Wang et al 2012a). Researchers and practitioners cannot only spend time on the internal design of the community, but need to consider the potential benefits and challenges caused by the crowded environment. Our paper contributes new insights to understand the influence of other communities in environment on the survival of the focal community. Through shared membership, an online community can learn from other similar communities within the same environment and improve their ability to survive. Moreover, our study examines the conditions under which the gain from other communities is strengthened. For example, the gain of membership overlap is more pronounced for new communities compared to mature communities. The beneficial effects of membership overlap are stronger when the shared members are core members from other mature communities.

Our results have design implications for community practitioners. For example, new community founders can design recruiting strategies to specifically target members who have experience in other mature communities, which might increase the new community's ability to survive. Furthermore, community leaders should monitor the level of membership overlap as an index of potential learning opportunities (especially for younger communities) and an index of competition (especially for older communities).

Besides contributing to the management of online communities, our work also contributes to organization science by extending the multi-team membership research. Simultaneous membership on more than one group is not just common in online setting but also prevalent in offline organizations. According to surveys conducted by the organization scholars (O'Leary, Mortensen and Wooley 2011), simultaneous membership in more than one team appears to be the norm for at least 65 percent of knowledge workers across a wide range of industries in United States. Despite the prevalence of membership overlap, there is only a handful of organization science research investigate this problem (Zika-Viktorsson, Sundstrom, & Engwall, 2006, O'Leary, Mortensen and Wooley 2011). We extend the multi-team membership research by identifying and empirically testing the benefit and withdraw of multi-team membership in the online setting.

Limitation and Future Research

This study is also subject to important limitations. First, our data analysis provides limited support for "why" the membership overlap can impact 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 improved community organization (which is an indicator of learning gain) and diluted members' attention and efforts (which is an indicator of competition). We are going to investigate these in the future study.

Secondly, our study only uses whether the community has activity or not as a proxy for community performance. Although activity level is a general performance metric and is widely used in many prior research (e.g., Iriberry, and Leroy 2009, Preece 2000), online communities have diverse goals, and more nuanced performance measures should be applied. For example, online health support communities should promote a specific type of activities -- discussions containing useful information and providing emotional support -- in order to be effective (Wang et al. 2012b). For peer production communities such as Wikipedia and open source software, the quality of the deliverables (i.e., encyclopedia articles and software) is another important performance measurement. In future study, in addition to the general metric (i.e., activity level), we will identify performance metrics specific to the studied communities and test the effects of membership overlap on those metrics.

Third, we only examined the overlap between communities inside Wikia. However, it is very likely that Wikia users also participate in communities outside Wikia like Wikipedia or even more different communities such as an open source software project or a Facebook group. This issue raises an interesting question of how the similarity between focal community and intersecting communities affects the effects of membership overlap. One hypothesis is that communities are more likely to gain benefits from similar communities because the knowledge, experiences and human capital are easier to transfer among similar types of organizations or projects. Heterogeneous communities, on the hand, might be more likely to compete with each other for members' time. We will investigate this in future study.

Finally, we conducted community-level analysis, but the communities might not be sufficiently independent. For example, through sharing members, the increased survival rate of new communities

might relate to the decreased survival rate of old communities. The benefits of having members who are core members in other communities might be associated with the loss of those communities whose core members are participating actively elsewhere. In future study, we would like to resolve the interdependence by conducting analysis on the community pair level (i.e., examining the relative change for pairs of communities which share members) or on the contributor level (i.e., tracking the contribution change of members for each community they participate in).

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. By analyzing the historical data of 5673 Wikia communities, we find that higher levels of membership overlap are associated with higher survival of online communities relative to lower levels of membership overlap. Furthermore, we find that it is beneficial for new communities to have shared members who play a central role in other mature communities. These findings provide new insight into an important mechanism underlying successful online communities, contribute to theories of organization science, and provide several actionable steps for the hosts and creators of online communities.

Acknowledgments

We thank all the ICIS reviewers for their thoughtful comments. We also thank Benjamin Mako Hill from University of Washington and Lisa Yu from Carnegie Mellon University for their insightful feedback and editing help.

References

- Baum, J. A., & Shipilov, A. V. (2006). 1.2 Ecological Approaches to Organizations. *The SAGE Handbook of Organization Studies*, 55.
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.
- Bryant, S. L., Forte, A., & Bruckman, A. (2005, November). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work* (pp. 1-10). ACM.
- Bourdieu, P. (1985). The forms of capital. In JC Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241-258). New York: Greenwood.
- Burke, M., & Kraut, R. (2008, November). Mopping up: modeling wikipedia promotion decisions. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 27-36). ACM.
- Burke, M., Marlow, C., & Lento, T. (2010, April). Social network activity and social well-being. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 1909-1912). ACM.
- Burt, R.S. (1987) "Social contagion and innovation: Cohesion versus structural equivalence," *American Journal of Sociology*, 92, 1287-1335.
- Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research*, 12(4), 346-362.
- Chen, J., Ren, Y., & Riedl, J. (2010, April). The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 821-830). ACM.
- Choi, B., Alexander, K., Kraut, R. E., & Levine, J. M. (2010, February). Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 107-116). ACM.
- Collett, D. (2003). *Modelling survival data in medical research* (Vol. 57). CRC press.
- Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 78(6), 1360.
- Hertel, G., Niedner, S., & Herrmann, S. (2003). Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. *Research policy*, 32(7), 1159-1177.

- Hill, B. M. and Shaw, A. (Forthcoming) "Is volunteer labor a "fixed and finite" resource? Evidence from peer production." Working Paper.
- Hogg, M. A., and J. C. Turner 1985 'Interpersonal attraction, social identification and psychological group formation'. *European Jour of Social Psychology* 15: 51–66.
- International Telecommunications Union (2013) Key ICT indicators for developed and developing countries and the world, <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>, Geneva, accessed 4 Sept 2013
- Iriberry, A. and Leroy, G. A life-cycle perspective on online community success. *ACM Comput. Surv.* 41, 2 (2009), 1–29
- Ingram, P. and Baum, J.A.C. (1997) 'Opportunity and constraint: Organizations' learning from the operating and competitive experience of industries', *Strategic Management Journal*, 18: 75–98 (Summer Special Issue).
- Jenkins, S. P. (2005). Survival analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.
- Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science*, 17(1), 45–63.
- Kairam, S. R., Wang, D. J., & Leskovec, J. (2012, February). The life and death of online groups: predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 673–682). ACM.
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), 457–481.
- Kraut, R. E. & Resnick, P. (2012). *Building successful online communities: Evidence-based social design*. Cambridge, MA: MIT Press.
- Kraut et al. forthcoming. The role of the founder in building online groups.
- Lakhani, K., and Wolf, R. "Why hackers do what they do: Understanding motivation and effort in free/open source software projects." (2003).
- Lerner, Josh, and Jean Tirole. 2005. The economics of technology sharing: Open source and beyond. *Journal of Economic Perspectives*. 19 (2): 99–120.
- Matthews, T., Whittaker, S., Badenes, H., et al. (2013) Community insights: helping community leaders enhance the value of enterprise online communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 513–522.
- Miner, A.S. and Haunschild, P.R. (1995) 'Population level learning', in L.L. Cummings and B.M. Staw (eds), *Research in Organizational Behavior*, 17. Greenwich CT: JAI Press. pp. 115–66.
- Muller, M., Ehrlich, K., Matthews, T., Perer, A., Ronen, I., & Guy, I. (2012, May). Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 2815–2824). ACM.
- Nov, O. (2007). What motivates wikipedians?. *Communications of the ACM*, 50(11), 60–64.
- 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.
- O'Mahony, & Ferraro, F. (2007). THE EMERGENCE OF GOVERNANCE IN AN OPEN SOURCE COMMUNITY. *Academy of Management Journal*, 50(5), 1079–1106.
- Palloff, R. M., and Pratt, K. 1999. *Building learning communities in cyberspace*. San Francisco: Jossey-Bass Publishers.
- Preece, J. (2000) *Online communities: Designing usability and supporting sociability*. Wiley, 2000.
- Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1), 13–32.
- Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, 28(3), 377–408.
- Resnick, P., Konstan, J., Chen, Y., & Kraut, R. E. (2012). 6 Starting New Online Communities. *Building successful online communities: Evidence-based social design*, 231.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.
- Stewart, K. J., & Gosain, S. (2006). The impact of ideology on effectiveness in open source software development teams. *Mis Quarterly*, 291–314.

- Tajfel, H., M. G. Billig, R. P. Bundy, and C. Flament 1971 'Social categorization and intergroup behaviour'. *European Journal of Social Psychology* 1/2: 149–178
- Wang, X., B. Butler, Y. Ren. (2012a). The impact of membership overlap on the growth: An ecological competition view of online groups. *Organization Science*
- Wang, Y., Kraut, R. E., & Levine, J. M. (2012b). To Stay or leave? The Relationship of emotional and informational support to commitment in online health support groups CSCW'2012: Proceedings of the ACM conference on computer-supported cooperative work. New York: ACM Press.
- Weber, Steven. 2004. *The success of open source*. Cambridge, MA: Harvard University Press.
- Zhu, H., Kraut, R., & Kittur, A. (2012, February). Effectiveness of shared leadership in online communities. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 407-416). ACM.
- Zika-Viktorsson, A., Sundstrom, P., & Engwall, M. 2006. Project overload: An exploratory study of work and management in multi-project settings. *International Journal of Project Management*, 24(5): 385-394.