Selecting an Effective Niche: An Ecological View of the Success of Online Communities

Haiyi Zhu[1], Jilin Chen[2], Tara Matthews[2], Aditya Pal[2], Hernan Badenes[3], Robert Kraut[1]
[1]Carnegie Mellon University, Pittsburgh, PA, USA, haiyiz@cs.cmu.edu, robert.kraut@cmu.edu
[2]IBM Research -Almaden, San Jose, CA, USA, {jilinc, tlmatthe, apal}@us.ibm.com
[3]IBM, Buenos Aires, Argentina, hbadenes@ar.ibm.com

ABSTRACT
Online communities serve various important functions, but many fail to thrive. Research on community success has traditionally focused on internal factors. In contrast, we take an *ecological view* to understand how the success of a community is influenced by other communities. We measured a community's relationship with other communities—its "niche"—through four dimensions: topic overlap, shared members, content linking, and shared offline organizational affiliation. We used a mixed-method approach, combining the quantitative analysis of 9495 online enterprise communities and interviews with community members. Our results show that too little or too much overlap in topic with other communities causes a community's activity to suffer. We also show that this main result is moderated in predictable ways by whether the community shares members with, links to content in, or shares an organizational affiliation with other communities. These findings provide new insight on community success, guiding online community designers on how to effectively position their community in relation to others.

Author Keywords
Topic overlap; online communities; success; workplace.

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 have become increasingly important for both work and social relationships. Peer production communities [1], such as Wikipedia and open source software projects, aggregate the efforts of volunteers to produce complex artifacts, such as the largest encyclopedia in human history (Wikipedia) and software that powers the Internet (e.g., Apache). Businesses 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., [15]).

Development of Internet technologies has significantly reduced the cost of creating online communities, resulting in large populations of communities. For example, Usenet (now accessible on the web via Google Groups) had over 189,000 active newsgroups as of 2005 [21]. Wikia is a well-known platform hosting more than 350,000 Wikipedia-like communities. Facebook provides infrastructure to host over a quarter of a billion Facebook groups [13]. However, communities in these platforms are not equally successful or active. On Wikia, 22% of the communities received no contributions one month after being created. On Facebook, where members create well over 100,000 new groups a day, 20% have no content production after the first day they were created and 53% have stopped all activity within three months of creation [13].

An important success factor for any community is its relationship with other communities, such as how the topics it covers and members it attracts relate to those of other communities. For example, if employees in a company have already set up many communities on the topic of Java programming, a newly created community on Java may be doomed to failure, because it directly *competes* with many established communities on the same topic for a shared pool of members. On the other hand, a new community on the Eclipse programming environment—an overlapping but still distinguished topic—might flourish, because many of the existing Java communities have members who use Eclipse and have the knowledge to contribute, a relevant but not redundant content base, and thus *complement* the new community. Due to these interactions, anyone starting a new community will have to carefully define its niche by examining other related communities, and may even decide a new community is not needed.

In this work we studied community success from an *ecological view* by examining how a community's activity level is impacted by its niche, i.e., its relationship with other communities in an ecosystem. We use the word *ecosystem* to mean the collection of all communities in a given environment, such as a shared technology platform or organization. Of the various dimensions defining a niche, we focus particularly on topic, because a community's topic strongly influences its scope, its audience, and the type of

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

ACM 978-1-4503-2473-1/14/04...$15.00.
http://dx.doi.org/10.1145/2556288.2557348

content that is relevant. We measure a community's topic niche through its topic overlap with other communities, and propose a series of hypotheses describing how a community's topic overlap affects its activity level. Beyond topic overlap, we also hypothesize how other dimensions of niche, such as shared members, content linking, and offline organizational affiliation can interact with topic overlap to impact activity level. We test our hypotheses on the internal use of online communities within a large global company. We used a mixed-method approach, combining quantitative analysis of 9,495 communities and qualitative interviews of community users.

The contributions of this work are two-fold. Theoretically, we show how a community's relationship with other communities in a larger ecosystem influences its activity levels, and gain new insights on important mechanisms that affect community success in large ecosystems. Practically, our findings may guide community creators on how to effectively position new communities within an ecosystem, and tool designers on how to support creators with this task.

RELATED WORK

Prior research on factors leading to the continued success of online communities fall into two main categories: individual community dynamics and inter-community relationships.

Individual community dynamics: A large body of literature investigates how community-level characteristics influence the success of online communities. This research has focused on two kinds of community-level characteristics: composition (i.e., the makeup of the community, such as its size or composition [3] [4]) and structure (i.e., the patterns of relationships among members, such as social network, leadership and governance structures [10] [16] [23]). The assumption of this group of research is that communities can achieve continued success by adjusting their composition (e.g., diversity of members, group size) and optimizing their internal structures (e.g., governance).

Inter-community relationships: Though most online communities cooperate and compete within a larger population of communities, only a few researchers have investigated community activity from an ecological perspective. The recent book by Kraut and Resnick [12] surveys hundreds of research papers and proposes design claims about building successful online communities. Among the 176 claims, 171 are about internal dynamics. We know relatively little about how success is influenced by external factors, such as other related communities. The closest prior work are Wang et al. [21] and Zhu et al. [24] which examine the impact of membership overlap on community activity. Wang et al. argued that membership overlap caused competition among communities for member time and attention that reduced the chance their opportunities for growth [21]. Zhu et al. [24] built on Wang et al.'s work, finding that moderate levels of membership overlap between communities may bring benefits that out-weigh the negatives, such as knowledge transfer and new member recruitment. However, research on inter-community relationships is in its infancy and many open questions remain. We contribute to this emerging area of study by examining the impact of shared topics, members, content, and offline organizational affiliation.

ECOLOGICAL VIEW OF COMMUNITY SUCCESS

To further explore inter-community relationships, we examine the online community success from an ecological perspective. This perspective is based on *organization ecological research*, which examines traditional organizations such as hotel chains and newspaper publishers ([1] [7]). Organization ecology research suggests that two ecosystem mechanisms—competition and complementarity—influence the success of organizations [1]. However, prior work has not studied how these mechanisms manifest in online communities, something we contribute in this paper.

Online Communities' Competition and Complementarity

Competition is a core concept in organization ecology. Organizations compete with others in the same ecosystem for common resources [7]. Furthermore, the *intensity* of competition between organizations is largely a function of how similar their resource requirements are: the more similar their resource requirements, the greater the potential for intense competition [7].

Applying this finding to online community ecosystems, we would expect communities to compete with each other for common resources such as members' attention and efforts. Members have a certain amount of time in the day, some of which they may allocate to community participation, but it is not possible for them to keep track of what is going on in all the communities in a large ecosystem. Competition might result in decreased activity in the communities vying for member attention, which is a common resource.

Complementarity in organization ecology describes benefits organizations may get from the existence of "competitors". Researchers in offline organizations found evidence that knowledge and operating experience can be transferred among similar organizations, thus increasing the survival rate of the organizations. For example, Ingram and Baum [8] found that a hotel chain's survival rate was positively related to the total operating experience accumulated by other hotel chains in the same country.

Similarly for online communities, members who join more than one community in an ecosystem may share their knowledge across communities. Community leaders can benefit by learning from the success and failure experiences of other similar communities. Complementarity might result in increased activity in the communities that share knowledge and experience.

Effects of Topic Overlap on Community Success

In this paper, we apply the mechanisms of competition and complementarity to explain different success levels across

an ecosystem of related online communities. We center our exploration on understanding the effects of *topic overlap* on community success, because a community's topic defines its content scope and member audience, thus centrally defining its relationship to other communities in the ecosystem. We also study the moderating effects of other dimensions that help define a community's niche, including shared members, content linking, and shared offline organizational affiliation with other communities.

To estimate community success in this paper, we use the overall activity (i.e., number of posts created, commented on, and viewed) in the community. Multiple researchers argue that these are reasonable approximations of community success, since ongoing activity and interactions among members are necessary for a healthy community and volume indicates levels of engagement and value [17].

When communities have *higher topic overlap* (i.e., more communities in the ecosystem with similar topics), communities have *more intense competition* for members' time, and hence *lower activity levels*. Thus, we hypothesize that competition between communities leads to a negative relationship between topic overlap and activity level. See row 1(a) of Table 1 for an illustration of this prediction.

On the positive side, when communities have *higher topic overlap*, they are more likely to complement each other by increased *learning and content sharing*. However, we predict that this benefit will slow down (or even plateau) as the topic overlap becomes higher. We base this prediction on the mechanism behind previously studied "learning curve" plateaus [22]: as topic overlap and sharing increase, there is less new information and experience available for a community to learn from. Furthermore, low topic overlap will hurt communities, because there will be less able to learn or borrow content from other complementary communities. Thus, we hypothesize that complementarity, as manifested through learning and content sharing between communities, leads to a positive relationship between topic overlap and activity level with diminishing returns. See row 1(b) of Table 1 for an illustration of this prediction.

When we put these predictions for competition and complementarity together, we expect that the effects of topic overlap should have a curvilinear shape (see the right-most column of Table 1, row 1). Too little or too much topic overlap will negatively impact a community's activity, for the arguments outline above: either complementarity will be too low or competition too high. Only when the topic overlap is moderate will the activity level be highest.

Hypothesis 1. There is a curvilinear relationship between the topic overlap of a given community with other communities and the activity level of this community. *Low* topic overlap and *high* topic overlap results in *low* activity level, while *moderate* topic overlap results in *highest* activity level.

Hypothesis 1 is about the general effects of how other communities with similar topics in the ecosystem can influence the activity of a given community. However, each of the other communities in the ecosystem does not equally influence a given community. For example, communities that share both members and topics should impact each other even more than communities that only share topics. Therefore, in the following section we propose hypotheses about the moderating effects of other ecosystem relationships, including shared members, shared content (approximated by measuring content linking), and shared offline organizational affiliation. Understanding the moderating effects of these other relationship aspects can provide a more complete view of the ecosystem's impact on community activity, as well as further our understanding of the underlying mechanisms of competition and complementarity.

Moderating effects of shared members

Shared members are the medium by which knowledge and experiences are transferred between communities, as well as the resources that communities compete for. Having shared members might intensify both complementarity and competition processes. Therefore, topic overlap with communities that share members should have a stronger curvilinear effect on the activity level (i.e., steeper increase and then steeper decrease), compared to the same amount of topic overlap but no shared members. See Table 1, row 2.

Hypothesis 2. The effects of topic overlap are stronger for communities that share members than for communities that do not share members.

Moderating effects of content linking

In an online setting, it is common that communities link to content in other relevant communities. These linking relationships on one hand encourage knowledge sharing and enhance complementarity (see Table 1, row 3(b)). But on the other hand, linking may intensify competition because the existence of "potential competitor communities" is more visible to members. Members may find the linked-to community more useful and spend more time there instead (see Table 1, row 3(a)).

Hypothesis 3. The effects of topic overlap are stronger with linked communities than with non-linked communities.

Moderating effects of offline organizational affiliation

For many communities, members share not only their online affiliation, but also their offline affiliation. In some communities, people get to know each other or are affiliated in an offline setting, and then maintain social or work contact in online communities. Examples include enterprise communities where employees, who already have their affiliations in a hierarchical company, participate in online communities to fulfill business-centric goals such as learning, collaboration and professional networking [15]. In other cases, communities in which people mainly interact online also have offline structures determined by members'

geographic or demographic distribution. For example, Wikipedia has language-based sub-communities and geographic-based local chapters. The offline organizational affiliation might influence the strength of competition and complementarity among the online communities.

We propose that the *competition* is *stronger* among communities that *share* offline organizational affiliation compared to communities that do not. High turn-over is an issue for most communities, and so their continued activity depends on the supply of new members [12]. According to prior research, network diffusion is one of the major mechanisms of community growth [10], i.e., new individuals participate because of their offline ties to current community members. Therefore, offline organizational affiliations often define a pool of people who can become new members in the online communities. Communities within the same offline organizational group recruit from a new member pool that overlaps, intensifying competition. See row 4(a) of Table 1 for an illustration of this prediction.

Separately, we propose that the *complementarity* (e.g., experience learning and content sharing) might be *stronger* if communities are from *different* offline organizational affiliation. According to weak tie theories [6], communities with members from different networks might provide more novel information and experiences than communities with

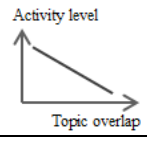
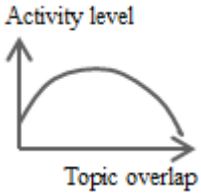
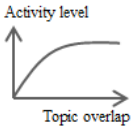
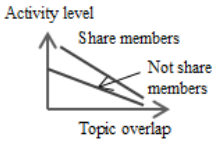
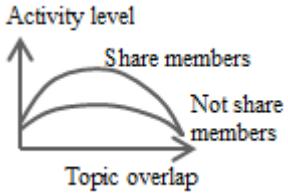
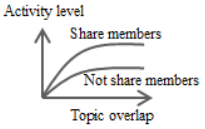
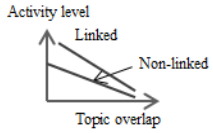
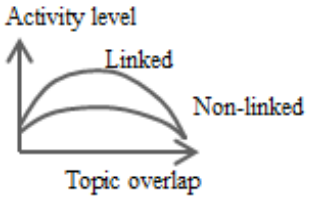
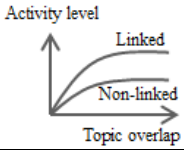
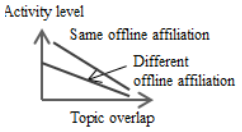
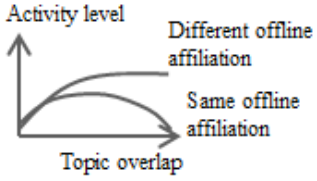
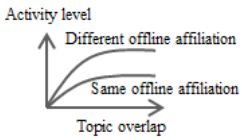
	Mechanisms		Overall effects
(1) How does topic overlap influence activity level?	**(a) Competition:** Dilute members' time and attention.		**Hypothesis 1:** 
	(b) Complementarity: Share information on common topic and learn success and failure experience from each other.		
(2) How do shared members moderate the effects of topic overlap?	**(a) Competition:** Competition is stronger for communities that share members.		**Hypothesis 2** 
	(b) Complementarity: Complementing is stronger for communities that share members because shared members can transfer knowledge and experience.		
(3) How does content linking moderate the effects of topic overlap?	**(a) Competition:** Competition is stronger if the communities are linked with each other because it is easier for members to go from one to the other.		**Hypothesis 3** 
	(b) Complementarity: Complementing is stronger if the communities are linked because information is easier to access and transfer.		
(4) How does shared offline organizational affiliation moderate the effects of topic overlap?	**(a) Competition:** Competition is stronger for communities that share the same offline organizational affiliation because they share the same new member pool and their growth space overlaps.		**Hypothesis 4** 
	(b) Complementarity: Complementarity is stronger for communities that do not share the same offline organizational affiliation because communities from a different network are more likely to bring in valuable information and experiences.		

Table 1. The effects of topic overlap on community activity.

members from the same network. See row 4(b) of Table 1 for an illustration of this prediction.

When we put these two predictions together, the resulting hypothesis is illustrated in the right-most column of Table 1, row 4, and described here:

Hypothesis 4. Topic overlap with communities that do *not* share offline organizational affiliation has a *greater increase* and a *smaller decrease* on activity level. Topic overlap with communities that *do* share offline organizational affiliation, has a *smaller increase* and a *greater decrease* on activity level.

METHOD

We test our hypotheses in the context of an enterprise online community platform. Here we describe the platform and our quantitative analysis and interview methods.

Study platform

This research was conducted in a global enterprise offering technology products and services to businesses. The company widely encouraged employee leadership of, and participation in, internal online communities and made commercial technology, Connections Communities ("Communities"), available to all employees. All communities we studied used this tool, which enabled leaders to easily create a community space with various social tools like forums, blogs, wikis, files, and bookmarks. As a result, there was a proliferation of communities and widespread membership, with over 166 thousand communities and over 580 thousand distinct members over five years. Communities ranged in size from a couple to tens of thousands. Many employees were members of multiple communities.

Connection Communities within the company studied provides a good platform to test the impact of ecosystems on community success for three reasons: First, the Connections platform supports the fundamental features that define online communities: (1) members have a shared goal/activity that provides the primary reason for belonging to the community, (2) members engage in repeated active participation, (3) members have access to shared resources, (4) there is reciprocity of information and services between members, and (5) there is a shared context of social conventions, language and protocols [17]. Second, the vast number of communities in the company studied has resulted in a community ecosystem crowded with similar communities, enabling the study of topic overlap. Third, members of Connections communities are authenticated, enabling us to collect data on their offline organizational affiliation. This provides a unique opportunity to examine hypothesis 4.

Due to the enterprise nature, Connection communities still differ from public online communities in several ways. For example, the Connection communities share organizational context, have business-centric goals, and members are authenticated. However, there is no strong reason to believe that these differences will confound the impact of ecosystem factors on community success. Therefore, we believe our results can generalize across most online communities.

Analysis strategy

We used a mixed methods approach to characterize our findings from both qualitative and quantitative perspectives. We chose 9495 active communities and ran our quantitative analyses on historical data to test the relationship between topic overlap and activity level. Meanwhile, we also conducted interviews with active community members to provide rich descriptions and concrete examples of the phenomena studied.

Quantitative analysis method

Data collection

We selected the 10K communities that had most recently been updated over a 14-day period prior to March 28th 2013. 9495 communities remained in the dataset after excluding those using non-English languages. We collected data at two time points: March 28th 2013 and June 9th 2013. In the analysis, the outcome variable is the activity of the community between March 28th 2013 and June 9th 2013; the independent variables and control variables (including topic overlap, number of members, and age) were collected in March 28th 2013.

Dependent variable

- **Activity level.** To measure the activity level, we calculated the sum of the counts of new content produced (the number of new wiki edits, wiki comments, forum topics, forum replies, blog entries, blogs comments, idea entries, idea comments, file entries, file comments, bookmarks, and activity entries) and the counts of content consumed (number of blog views, idea views, and file downloads), in the three-month period note above. We calculated the sum of *production* and *consumption* counts because (1) both are widely used measures of community activity [5, 9], (2) production and consumption highly correlate with each other, and (3) we found that the result is the same if we considered these two measures separately. We log transformed this variable in the analysis.

Control variables

- **Number of members.** We included the number of community members as control variable. We define members as those who have edited any page of a community at least once, not just those people whose names appear on the member list. The reason is that, by definition [17], the members in communities should engage in repeated active participation.
- **Age of the community**. We included the age of the community as control variable. We measured age in number of months

Independent variable

- **Topic overlap.** We operationalized the topic overlap of one community as the sum of content similarity between the focal community and all the other communities in our dataset. We represented the content of each community through a vector of TF-IDF (term frequency-inverse document frequency) scores, where each score represented how important a word was to the content of a given community [19]. The TF-IDF score increases proportionally to the frequency of the word in the given community, but is offset by the frequency of the word in all the communities. Then for the focal community, we calculated the cosine similarity between its TF-IDF vector and the TF-IDF vectors of all other communities, and summed these similarity scores together. Take the Java Developer community in Table 2 as example, the topic overlap of the Java Developer community is the sum of its cosine similarity with all the other communities: 0.9+0.4+0.1=1.4.
- **Topic overlap with shared members**. This variable measures the topic overlap with communities that share members. We calculated this measure by only summing the similarity of communities that shared members with the focal community. For example, since the Java Developer community only share members with the Software Engineer community (see Table 2), its *topic overlap with shared members* is 0.9. This measurement is operationalized as the sum (not mean) of the similarity score because the underlying competition and complementarity effects are stronger both when many communities overlap a little and when few communities overlap a lot. For example, a community will likely learn comparably from 30 communities that share members and some topic relevance, OR from 3 communities that share members and nearly identical topic focuses.
- **Topic overlap without shared members**. This variable measures the topic overlap with communities that do not share members. We calculated this measure by only including communities that do not share members. Therefore, the value of the Java Developer community is 0.5 for this measure (see Table 2).
- **Topic overlap with linked communities.** We defined two communities as linked if one had hyperlinks that directed to pages of the other community. We calculated this measure by only including linked communities. The value of the Java Developer community is 1.3 for this measure (see Table 2).
- **Topic overlap with non-linked communities.** We calculated this measure by only including communities not linked with focal community. The value of the Java Developer community is 0.1 for this measure (Table 2).
- **Topic overlap in the same offline organizational affiliation.** The variable measures the topic overlap with communities that share offline organizational affiliation. We define communities as sharing offline organizational affiliation when they are from the same business division. Connections communities are inside a large global company, which includes divisions such as Marketing, Software Development, Hardware Development, Business Services, and Research. We operationalized the division of each community as the division of the majority of community owners. We calculated this measure by including communities from the same division as the focal community. The value of the Java Developer community is 0.9 for this measure (Table 2).
- **Topic overlap in different offline organizational affiliation.** The variable measures the topic overlap with communities that belong to different offline organizational affiliation. We calculated this measure by including communities that belong to different divisions. The value of the Java Developer community is 0.5 for this measure (see Table 2).

Name	Division	Cosine similarity between TF-IDF (Share members or not) (Linked or not)			
		1. Java Developer	**2.**Software Engineer	**3.**Data Analytics	**4.**Human Resource
1. Java Developer	Engineer		0.9 (1) (1)	0.4 (0) (1)	0.1 (0) (0)
2. Software Engineer	Engineer	0.9 (1) (1)		0.6 (1) (1)	0.1 (0) (0)
3. Data Analytics	Market	0.4 (0) (1)	0.6 (1) (1)		0.2 (1) (0)
4. Human Resource	HR	0.1 (0) (0)	0.1 (0) (0)	0.2 (1) (0)	

Table 2. Hypothetical names and values for four communities to illustrate how the measures are calculated

All the independent variables are normalized to [0,1]. Also, note that the four niche dimensions (i.e., topic overlap, shared members, content linking and shared offline organizational affiliation) are independent and only minimally correlated. Take topic overlap and shared members as example: because each member has multiple interests and needs they tend to join many communities with very different topics, and thus topic irrelevant communities may also share members. The statistics confirm the above observations: the correlation between topic overlap and shared members is 0.16 in our dataset.

Qualitative analysis method

To supplement our quantitative analysis, talked with members of the communities we were analyzing to understand if our conclusions were accurate and to contribute detailed descriptions of the mechanisms studied. We interviewed 10 members about their experience participating in communities with high and low topic overlap, managing their time between multiple communities, and their practices around sharing information between communities. We referred to a list of 5 communities they had contributed to when we asked questions about these topics, in order to keep the discussion grounded in actual communities and experiences.

We randomly sampled from a pool of members who had contributed to at least 5 of the communities in our dataset of 9495 communities (described above), where at least one of those communities had low topic overlap (bottom 20% of our dataset) and one had high topic overlap with other communities (top 20%). These criteria selected members who were at least moderately active in communities with a variety of job roles, an average of 19 years of experience (ranging from 1 to 33 years) from across the organization. We followed a grounded theory approach of adding participants and analyzing data as we went, stopping when we reached a point of information saturation [20]. Three researchers attended each interview, one to ask questions and the others to take detailed notes. Interviews were semi-structured, lasted 30-45 minutes, were conducted via phone and audio recorded. We analyzed the detailed notes using open coding, and then analyzed the concepts and categories from our initial coding for themes. Below we include those themes that are relevant to our quantitative findings.

RESULTS

The effects of topic overlap (Hypothesis 1)

We hypothesized that for a given community, there is a curvilinear relationship between its topic overlap with other communities and its activity level (Table 1, row 1). As shown in Figure 1, low levels of topic and high topic overlap led to low activity levels. Moderate levels of topic overlap led to the highest activity levels. Model 1 in Table 4 shows that the curvilinear relationship is statistically significant. The linear term of topic overlap is significantly positive (*coef.*=3.30, $p<0.01$), while the quadratic term is significantly negative (*coef.*= -14.2, $p<0.01$). These results *confirm hypothesis 1*.

The qualitative interviews confirmed these quantitative results, suggesting that *competition* and *complementarity* were key mechanisms behind them. Five out of 10 participants discussed themes related to the importance of complementarity regarding topic overlap. Specifically, participants discussed how topically related communities in the ecosystem shared the same content to mutual benefit, as described by H1:

> *"I'm in [my division's sales community] and [the sub-division's sales community]. I know they have a lot of the same information... for example, if [my division's sales community] post [sales] about [our sub-division's product], it'll probably show up in [our sub-division's sales community]. But something like [my division's sales community] is much more broad, so it's going to have a lot more information."*

Competition was particularly salient for participants, as 7 out of 10 discussed its importance. About communities that shared topics, participants discussed the importance of finding information, and how fewer communities on a topic made this easier and, conversely, too many competed for their attention and made it difficult. W1 describes:

> *"I find it very difficult to find the information I need in communities… There's a [Product] Program Team Community, there's a [Product] Development Community, and I think there's at least a couple of others… The fact that there are a lot of different [Product] communities… I don't know which one to look at."*

Variable Name	N	Mean	S.D.
Age of the community	9495	15.5	14.5
Number of members	9495	14.9	54.0
Activity level (logged)	9495	3.23	2.51
Topic overlap	9495	0.31	0.18
Topic overlap with shared members	9495	0.02	0.04
Topic overlap with linked community	9495	0.01	0.04
Topic overlap in the same offline org. affiliation	9495	0.10	0.19

Table 3. Descriptive statistics

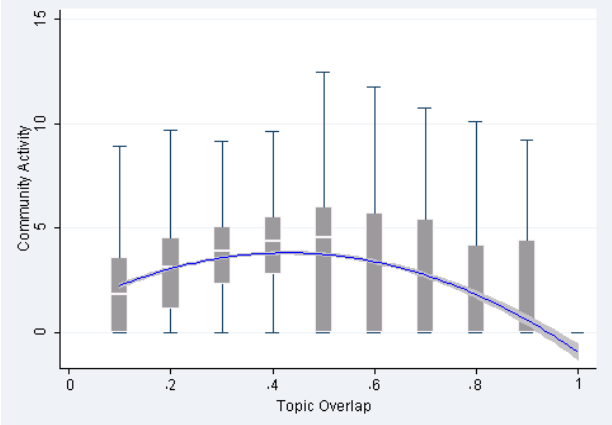


Figure 1. Relationship between topic overlap and activity. We showed the quadratic prediction plots with 95% confident interval as well as the box plots.

Participants also described how they determined which community to join or visit when such competition occurred. Commonly cited factors were a large community size (4 participants), frequent updates (3), high quality content (3), and that key people for the community's topic were members (e.g., known subject matter experts) (2), e.g.:

> *"If you looking for an industry one, you'll come up with about a hundred different ones. Some created by three people in Finland. So for me, the criteria is, which is the biggest, which has the people that I recognize as being the subject matter experts in that area… finding the ones that looked like they covered the most ground and probably were the most active and had the most information." (A1)*

Moderating effects of shared members (Hypothesis 2)

Shared members are medium to transfer the knowledge and, as well as a valuable resource communities compete for. We therefore hypothesized that topic overlap should have a *stronger* curvilinear effect in communities that shared members than in communities that do not share members. As shown in Figure 2 and Model 2 of Table 4, for communities with shared members, topic overlap's effects are of higher magnitude (*linear coef.* = 26.6, *quadratic coef.* = -36.9), while for communities without shared members

Explanatory variables	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.	Model 4 Coef.
Topic Overlap (v1)	**3.30****			
Quadratic term of v1	**-14.2****			
Topic overlap with shared members (v2)		**26.6****		
Quadratic term of v2		**-36.9****		
Topic overlap without shared members (v3)		**2.25****		
Quadratic term of v3		**-13.4****		
Topic overlap with linked communities (v4)			**24.8****	
Quadratic term of v4			**-28.4****	
Topic overlap with non-linked communities (v5)			**2.36****	
Quadratic term of v5			**-16.4****	
Topic overlap in the same offline org affiliation (v6)				**-2.73****
Quadratic term of v6				**-2.13****
Topic overlap in different offline org affiliation (v7)				**8.42****
Quadratic term of v7				**-8.42****
Number of members	**8.95e-4****	**3.33e-3****	**5.02e-3****	**6.30e-3****
Community age	**9.29e-4****	**-3.48e-3**	**3.90e-3***	**-4.17e-3***
R-square	**0.13**	**0.21**	**0.21**	**0.31**

** p<0.01, *p<0.05

Table 4. The effects of topic overlap (model 1) and the moderating effects of shared members (model 2), content linking (model 3), and offline organization affiliation (model 4) on the community activity.

the effects are of much lower magnitude (*linear coef.* = 2.25, *quadratic coef.* = -13.4). These results indicate that there are stronger competition and stronger complementarity effects between communities that share both topics and members, *confirming hypothesis 2*.

The qualitative interviews provided further insights on the role of shared members. For complementarity, 8 members described specific instances when they shared content between two topically-similar communities, e.g., H1:

> *"The Consulting by Degrees Community is actually the… parent community of the U.S. Philadelphia [Community]… So sometimes if we see something in the Consulting by Degrees Community that we want to specifically share with our group of Philadelphia folks we might post it again in our group, just to bring more attention."*

Several of the 7 out of 10 participants noted above who discussed competition, emphasized that competing for a shared member base between topically-similar communities harmed those communities, e.g.:

> *"Your user base is spread or is divided into these various communities… People just go and create communities without paying attention if there is something already out there… They keep creating communities with content that is already out there. And then those communities start dying out and their activity is pretty low." (S1)*

Moderating effects of content linking (Hypothesis 3)

We predicted that content linking makes knowledge sharing easier, while also intensifying competition by making members more aware of related communities. As shown in Figure 3 and in Model 3 of Table 4, for linked communities, topic overlap's effects are of higher magnitude (*linear coef.* = 24.8, *quadratic coef.* = -28.4), while for unlinked communities the effects are of much lower magnitude (*linear coef.* = 2.36, *quadratic coef.* = -16.4). These results *confirm hypothesis 3*.

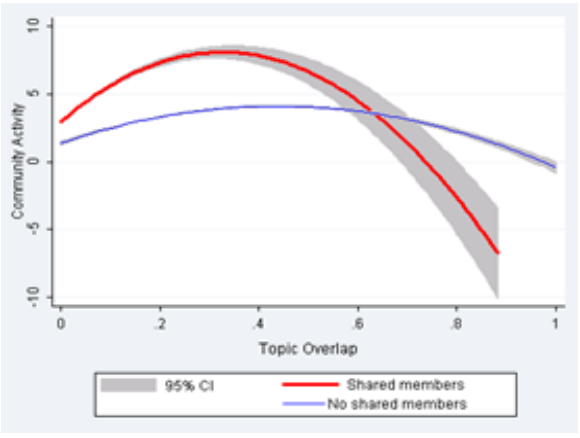


Figure 2. Moderating effects of shared members.

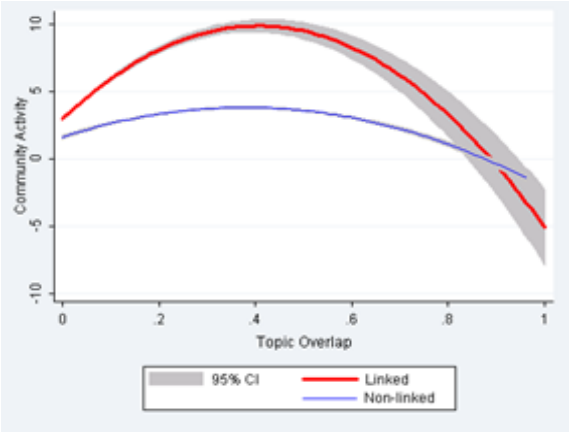


Figure 3. Moderating effects of content linking.

Moderating effects of shared offline affiliation (Hypothesis 4)

We predicted that sharing offline organizational affiliation intensifies competition and reduces complementarity. Indeed, while in all other conditions topic overlap's linear effect is positive, when communities share the same offline affiliation, the linear effect turned negative (*coef* = -2.73), as shown in Figure 4 and Model 4 in Table 4. This result indicates that the shared affiliation has indeed intensified the detrimental effects of topic overlap and reduced its benefits, *confirming hypothesis 4*.

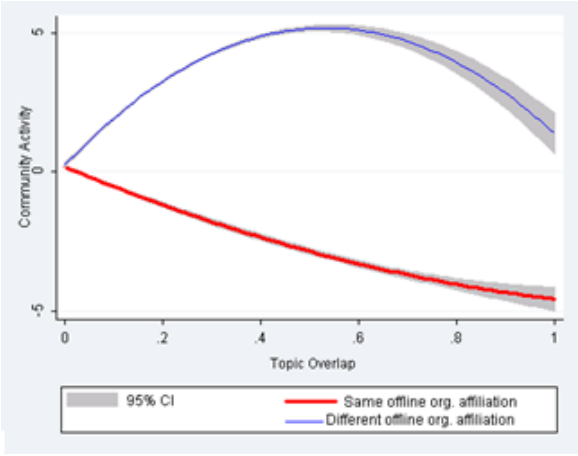


Figure 4. Moderating effects of offline organizational affiliation.

DISCUSSION

Theoretical contributions

Our paper investigated organizational ecology theories in an online enterprise setting, a condition that prior work has not studied. Our results largely confirmed prior theory in this new condition: Communities that overlap in niche within the same ecosystem both complement and compete with each other. The benefits of complementarity dominate when overlap is low, while the drawbacks of competition dominate when the overlap is high. These effects lead to a sweet-spot, where communities with a moderate overlap achieve the highest activity levels.

By studying niche through four different dimensions—topic, members, content, and offline affiliation—we also uncovered new nuanced insights. For instance, we have found that sharing members and linking content intensifies the effect of topic overlap, making complementarity and competition stronger, and making the sweet-spot sweeter. We also found that sharing offline organization affiliation makes topic overlap more harmful, making more specialized communities more desirable. On the other hand, not sharing offline affiliations makes topically-similar communities more likely to flourish. This latter insight might explain the huge success of Facebook copies in other countries, such as Chinese RenRen. (Chinese RenRen is a clone of Facebook launched in 2005), despite their similarity to Facebook in almost all other aspects.

Practical implication

We believe the theoretical findings of this work have direct value to leaders and managers of online communities. When creating a new community, leaders often have a topic in mind but are concerned if the new community will gain support from similar communities in the ecosystem, or if it will die from fierce competition.

Our results suggest that these concerns are not misplaced, and our models suggest that the responses of the ecosystem can be partially predicted beforehand. For instance, if a proposed community has a high topic overlap with many existing communities, and many of these communities share the same offline affiliation with the proposed community, it may be better to not start the new community but instead join an existing one. On the other hand, if a proposed community is only moderately overlapped with other communities' topic, has already gained support from these communities' members, and does not share offline affiliation with these communities, the community should be created as-is because it will likely succeed. For situations in between these two examples, various strategies might be taken, such as specializing the niche to avoid competition, changing the niche so as to leverage members and contents in related communities, or making the community independent of existing offline organizations.

Because many ecosystems are very large, with 100K communities or more, it may be impossible for community leaders to understand them. Our work informs tools, such as visualization or analytic systems, aimed to solve this problem. These tools should enable leaders to get an overview of a community ecosystem to understand its topic distributions, how many members gravitate toward different topics, and how communities relate to offline affiliations. The tools could assess a proposed niche and suggest modifications to improve the chance of success. These tools could also point designers to relevant content to bootstrap their community. Our interviews suggest that members also suffered when too many communities covered a topic. Other tools could help them identify the right set of communities to join to best fit their topic interests.

Limitation and future research

First, while we proposed the underlying mechanisms that drive the observable variables, our quantitative data analysis by itself cannot directly prove the existence of these mechanisms. Nonetheless, our use of both quantitative and qualitative methods, results which strongly agreed with each other, helps alleviate this concern.

Secondly, our study used whether the community has activity or not as a proxy for community success, while in reality success can be measured in many aspects such as quality of deliverables (in Wikipedia) and progress towards particular goals (in enterprise). Nonetheless, as activity

level is indeed a widely-used measure of community success [9, 17], we believe our results are still valuable. Future research could extend this work, by incorporating more nuanced success measures as appropriate.

Lastly, the *importance* of a community's topic might be a confounding factor, because it could be argued that a more important topic may result in more members and more activities. We however believe such proposed effects do not necessarily happen, because while more people will be interested in important topics, they will also likely have more communities to choose from, in the end balancing out the success of each individual community. As a result, we believe competition and complementarity are indeed the mechanisms driving our findings, and suggest future work to measure topic importance separately and study its effects.

CONCLUSION

We take an ecological view to understand the impact of a given community's position in a larger population of communities on its activity level. These findings provide new insight into an important mechanism underlying successful online communities, and may provide valuable guidance for the hosts and creators of online communities.

ACKNOWLEDGMENTS

This work was supported by IBM and NSF grant IIS-0808711.

REFERENCES

1. Baum, J. A., & Shipilov, A. V. (2006). 1.2 Ecological Approaches to Organizations. The SAGE Handbook of Organization Studies, 55.
2. Benkler, Y. (2006). The wealth of networks: How social production transforms markets and freedom. Yale University Press.
3. 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.
4. Chen, J., Ren, Y., & Riedl, J. (2010). The effects of diversity on group productivity and member withdrawal in online volunteer groups. In Proc. of CHI (821-830).
5. Cothrel, J.P. Measuring the success of an online community. Strategy & Leadership 28, 2 (2000), 17–21.
6. Granovetter, M. S. (1973). The strength of weak ties. American journal of sociology, 78(6), 1360.
7. Hannan, M. T., & Freeman, J. (1977). The population ecology of organizations. American journal of sociology, 929-964.
8. 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.
9. Iriberri, A. and Leroy, G. A life-cycle perspective on online community success. ACM Comput. Surv. 41, 2 (2009), 1–29.
10. Kairam, S. R., Wang, D. J., & Leskovec, J. (2012). The life and death of online groups: predicting group growth and longevity. In Proceedings of WSDM (673-682).
11. Kim, A.J. (2000) Community building on the web : Secret strategies for successful online communities. Peachpit Press.
12. Kraut, R. E., & Resnick, P. (2012) Building successful online communities: Evidence-based social design. MIT Press.
13. Kraut R.E. & Fiore A. T.(2014) The role of founders in building online groups. In Proc. of CSCW.
14. Lazar, J. and Preece, J. (1998) Classification schema for online communities. Proc. of AISAC, 84–86.
15. Muller, M., Ehrlich, K., Matthews, T., Perer, A., Ronen, I., and Guy, I. (2012) Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. Proc. of CHI, 2815–2824.
16. O'Mahony, & Ferraro, F. (2007). The Emergence of Governance in an Open Source Community. Academy of Management Journal, 50(5), 1079-1106.
17. Preece, J., & Maloney-Krichmar, D. (2003). Online communities. In J. Jacko & A. A. Sears (Eds.), Handbook of human-computer interaction (596-620).
18. 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.
19. Salton G, Buckley C (1988). Term-weighting approaches in automatic text retrieval. Information Processing and Management 24 (5): 513–523.
20. Seidman, I. (1998) Interviewing as Qualitative Research: A Guide for Researchers in Education And the Social Sciences. Teachers College Press, New York.
21. 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.
22. Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. Dec. Sci., 10(2), 302.
23. Zhu, H., Kraut, R., & Kittur, A. (2012). Effectiveness of shared leadership in online communities. In Proc. of CSCW (407-416).
24. Zhu, H., Kraut, R., & Kittur, A. (in press). The Impact of Membership Overlap on the Survival of Online Communities. In Proc. of CHI.