AGENT-BASED MODELING TO INFORM ECOLOGICAL COEXISTENCE OF ONLINE COMMUNITIES: IMPACT OF COMMUNITY SIZE, PARTICIPATION COST, TOPIC BREADTH AND POPULATION INTEREST DISTRIBUTION

HAN LUYANG

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF INFORMATION SYSTEMS

NATIONAL UNIVERSITY OF SINGAPORE

2017

Supervisor:

Associate Professor Hahn Jungpil

Examiners:

Associate Professor Heng Cheng Suang

Dr Phan Tuan Quang

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Han Luyang

04 January 2017

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who helped me during the writing of this thesis.

My deepest gratitude goes first and foremost to my supervisor, Professor Hahn Jungpil, a respectable scholar who has provided me with much helpful guidance for this thesis. Without his enlightening inspiration, patient suggestion and consecutive discussion every week, I could not have finished my thesis. His impressive kindness and insight encourage me not only in this thesis but also my future life.

Second, I shall extend my thanks to all the professors who have taught me over the past two and half years of study. Their devoted teaching and interesting lectures have benefited me a lot when I prepared for my thesis.

Then, I would like to acknowledge my classmates and seniors who gave me their help and time in listening to me and help me work out my problems during the difficult time in my study as well as the thesis.

Finally and in particular, I would like to express my gratitude to my family and my girlfriend for their understanding and confidence in me through these years. Without their support, I could not have finished the thesis attentively.

CONTENTS

SUMMARY

Online communities have taken up a great part of people's daily lives, with the development of the Internet. Although huge in numbers, there exists a long tail phenomenon where only a few communities succeed and attract the majority of Internet users while the vast majority struggle for their survival until they fail. When various communities can (and do) coexist, it is important to understand which factors are important for them to maintain attraction and achieve success. The coexistence problem has been well explored in the organizational ecology literature. However, since there are both similarities and differences between online community and traditional organizations, whether organizational theories can be directly applied to the online context still needs to be cautiously explored. In this paper, we follow the roadmap provided by Davis et al. (2007) to conduct an agent-based modeling (ABM) simulation work to develop novel theory based on the previous literature. We find that in the scenario of two coexisting communities, both community size and participation costs can significantly affect the development of a community. A larger community can attract more active members who frequently login to it. Meanwhile, lower participation costs can encourage members' reading and posting behaviors. Moreover, we observe the important influence of the interest distribution of the user population on communities' topic trends. For a population that focuses on only one topic, a community can converge to the topic quickly regardless of

whether its initial topic is broad or focused. This simulation model provides not only theoretical implications to the literature but practical guidance to operators of online communities.

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Online communities have become one of the most popular Internet applications impacting peoples' daily lives. With the adoption of Web2.0 and other browser-based communication tools, it has gone through rapid development and is attracting numerous Internet users every day. It is reported that, among one billion Internet users, 84% have taken part in an online community (Horrigan 2001). In China, according to a report of the China Internet Network Information Center, half of the Internet users have participated in an online community by 2010. Many online communities have achieved great success. Facebook, for example, has around 1.7 billion users by 2016. In China, Sina Weibo has 160 million active users and Baidu Tieba has 200 million active users per month. However, many more online communities are small and struggle for their survival.

Such a long tail phenomenon[1] in online communities can be caused by several reasons. Membership size can benefit the sustainability of online communities. Research shows that membership size as a component of available resources can positively affect what members gain from a telecommunications network (Priem and Butler 2001). Reading and posting costs reduce members' intention

[1] Long tail: Statistically, it means the portion of the distribution having a large number of occurrences far from the "head" or central part of the distribution. In our case, a small number of online communities succeed at the "head" but the rest fail or struggle for survival at the long tail.

to participate in a community. There are many ways in which participation costs can increase. For example, a discussion community that adopts a rigorous and stringent identification process or that has a bad HCI design can increase both reading and posting costs. Researchers point out that discussion moderation at the community and individual levels can help members ease the information overload problem and hence reduce reading cost (Ren and Kraut 2014). The breadth of community-supported topics is another vital factor in the development of online community. Prior research suggests that a broader topic can result in a higher member commitment, thus enhance the participation intentions of community members(Ren and Kraut 2014).

Online communities usually have a voluntary structure which allows a member to change between each other with low switching cost(Bateman et al. 2011). At the same time, members' time and energy are limited. They need to decide how to effectively allocate these resources. The possibility of membership overlap when resources are constrained creates the competition phenomenon among communities. Although it is well explored in the organizational ecology literature, this ecological competition has not yet been fully understood in the context of online communities. One pioneering work discusses the relationship between membership overlap and member growth (Wang et al. 2013). Their results show that larger and older groups experience greater difficulty in growing their membership. Also, large groups are more vulnerable to the competitive pressure from the perspective of membership overlap.

In this study, we aim to investigate the coexistence of communities that share the same potential user population with one another. We will ask what factors can help communities surpass their opponents when they encounter competition. To be specific, for example, in a two-community scenario, will initial size bring insurmountable advantages to the larger community when they are faced with a smaller challenger? Will a user-friendly HCI design that reduces reading and posting costs help a community win over its rival? Will narrowing down topical breadth benefit a community's survival? Different from previous works, which usually adopt the case study approach or econometric analyses, this paper carefully follows the roadmap provided by Davis et al (2007) and conducts theoretical analyses using agent-based modeling and simulations.

The simulation method is an increasingly important methodological approach for developing theory. Especially, it is "very useful in the sweet spot between theory-creating research using such methods as inductive multiple case studies and formal modeling, and theory-testing research using multivariate, statistical analysis", as argued by Davis et al. (2007, p.481). In this paper, we first introduce theories related to the coexistence of communities. Building on these, we develop a computational model of online community participation and competition and conduct simulation experiments to identify the factors that affect the nature of ecological competition of in the two-communities scenario. Finally, we develop a simple regression model to validate our simulation results. The structure of the remaining report is as follows: In section 2, we introduce the organizational ecology literature and its relationship with online community coexistence. In section 3, we integrate several social science theories, termed as 'simple theory' in the Davis et al paper (Davis et al. 2007), that are related to this work. In section 4, we introduce the conceptual framework and theoretical constructs used in the simulation model. In section 5, we run experiments to build novel theory. In section 6, we use simple regression model to validate the results. Last but not the least, in section 7, we draw conclusion and discuss the implications and limitation.

2 Ecological Coexistence of Online Communities

2.1 Literature of Organizational Ecology

Organizations need resources to survive and operate. The resources can be tangible assets, such as land, funds and labors or intangible assets, such as brand and customer loyalty. Resources are constrained by the environment, such as, by the target market, the nature of opponents and cooperators (Astley and Fombrun 1983). To coexist in the same ecological system, organizational relationships are twofold. On the one hand, organizations can find legitimacy and opportunities to learn from each other, adopt strategic policies to adapt the environment and evolve a symbiotic relation (Aldrich and Ruef 2006). On the other hand, they need to compete for limited resources and follow the selection process of 'survival of the fittest' (McKenzie and Hawley 1968, Barnett and Carroll 1987).

Different factors are explored in the organizational ecology literature. Firstly, size is one of the most important variables that affect organizational performance. Empirical research indicates that larger organizations are more likely to survive longer than smaller ones (Aldrich and Auster 1986, Baum and Oliver 1996). This is not only because large organizations can access more resources and build useful connections with others, but also because large business can acquire more resources as needed by taking advantage of its strengths. Scholars term this phenomenon as the ‘liability of smallness’.

Secondly, barriers to entry can also affect organizational development. Defined as the cost that must be incurred by a new entrant to enter into a market that incumbents do not have or have not had to incur, entry barriers can help old organizations to keep their advantages and make it difficult for young organizations to enter the market. This partially causes the ‘liability of newness’ phenomenon in the ecology literature. For example, customers have loyalty to established products. This strong brand awareness increases switching cost and creates a barrier for new entrants.

Finally, organizations are also shaped by the environment, such as cultures, employees, and the target market. A study shows that organizational norms and environmental uncertainty have effects on entrepreneurial strategy and hence its coexistence in the organizational ecology (Russell and Russell 1992). Scholars points out that an organization’s strategies should be made according to its internal resources and skills as well as external environment (Grant 1991).

2.2 Ecology and Coexistence of Online Communities

Although it is well explored in the organizational ecology literature, few researches have been done under the online context. Whether conclusions from the organizational ecology literature can be directly transferred to this new scenario is unclear since both similarity and difference exist between traditional organizations and online communities.

Like traditional organizations, online communities have boundary-maintaining mechanisms to distinguish members from nonmembers and sometimes limit group content to only its members (Wang et al. 2013). Online communities can also be affected by its members, environments, and targets. For example, the nature of what members discuss determines the type of the community. At the same time, differences are twofold. On the one hand, smallness and newness don’t necessarily imply liability and disadvantages. A study indicates larger and older online groups experience greater difficulty in further growing their membership (Wang et al. 2013). On the other hand, the entry cost is low due to global reach, flux membership and low switching costs of online communities.

3 Theory Integration in Agent-Based Models

Many social science theories explain what factors make online communities successful. According to the resource-based theory, scholars find large groups provide more resources and informational benefits to their members (Priem and Butler 2001). According to information overload theory, people benefit less when they deal with too many messages because of limited capacity to process information (Jones et al. 2004). According to theories of altruistic behaviors, members can gain positive self-evaluation from actions such as answering questions (Wasko and Faraj 2005). According to the collective effort model, people are less willing to contribute when the group is large as they believe others will do so (Karau and Williams 1993). According to group identity theory, similarity among group members can lead to stronger attachment to a group (Hogg and Terry 2000). In small group studies, scholars find that people can benefit from interpersonal bonds as they interact more (Ren et al. 2007). Also, intrinsic and extrinsic motivation, in other words, enjoyment and reputation, can contribute to communities' survival (Ridings and Gefen 2004, Wasko and Faraj 2005). In summary, there are six sources of benefits a member can derive from an online community: 1) accessing information; 2) sharing information; 3) identity-based attachment; 4) bond-based attachment; 5) enjoyment; and 6) reputation (Ren and Kraut 2014).

Although quite a lot of social science studies explain communities' survival and sustainability, most studies suffer from two limitations. First, conclusions drawn from different theories may contradict with each other when they are from different perspectives. For example, resource-based theory points out that large groups can benefit members more and encourage their participation, while the collective effort model indicates that large groups may discourage members' contribution. Second, social factors found in these studies cannot guide community designers to improve community performance. For example, we know that benefit from sharing information can positively affect community's activity. However, what we can do at the community design level to increase it is still not clear.

The simulation method is a good way to overcome these two limitations. In this paper, we integrate those simple theories into a conceptual framework and follow the roadmap provided by Davis and colleagues (Davis et al. 2007) to simulate dynamics of an online community. Also, we add various treatments, which are directly related to community design, into the model.

4 Conceptual Framework of the Agent-Based Model

In our framework, we depict the model dynamics from three different levels: population, community and member. Briefly speaking, a population refers to the set of people (agents) who may be interested in the target online communities. A community means an online application that attracts people together and has a clear boundary of membership. And an agent is an entity who can enter a community, by which he gains the membership and participates in the community such as reading and posting, or can exit a community, by which he loses the membership and interest from the community. In the following sections, we show the attributes of these three levels in details.

4.1 Population

A population consists of agents who may be interested in the target online communities. Different online communities are designed to attract different people. One of vital parameters to distinguish this difference is the interest of an agent. In this paper, we categorize the population into four types according to their distinction in ***interest generation process***: (1) a focused interest population; (2) a broad interest population; (3) a hybrid population consisting of focused and broad interest agents (half-half interest) and; (4) a general interest population.

The interest generating process is used to generate agents' interests according to certain design principles. We define three different ways to simulate an agent's interests. Firstly, Interest type-1 process. A member can be interested in all topics and spread his attention uniformly across different interests. Secondly, Interest type-2 process. A member can be interested in a specified topic and focus on it. Finally, Interest type-3 process. An agent can be interested in not all but some of the interests and devote his energy to them. In the broad interest population, every agent generates his interest by the type-1 process. In the focus interest population, every agent generates his interest by the type-2 process. In half-half interest population, half of the agents generate their interest by the type-1 process and the rest adopts the type-2 process. In the general interest population, we use the mixture of these three processes. More specifically, in this thesis, the proportion for type-1, type-2 and type-3 generation processes is 0.3, 0.3 and 0.4 respectively. We summarize the details in the Table 4.1.1.

Table 4.1.1 Interest Generation Process and Population Types

ppl\ratio\interest	Interest 1[2]	Interest 2	Interest 3
Broad Interest PPL	1	0	0
Focus Interest PPL	0	1	0
Half-Half PPL	0.5	0.5	0
General Case PPL	0.3	0.3	0.4

[2] **Interest Generation Process** Interest 1: members' interests are uniformly distributed among all the topics; Interest2: members only focus on the first one among all topics; Interest 3: members can have at most 3 interested topics and uniformly distribute their interests among these topics.

4.2 Community

A community is an online application that has a clear boundary of membership. An agent needs to sign up to become a member and login to read and post messages. Members may post new messages or reply to existing messages to create message threads. Different communities have different mechanism designs for the information sharing process on it. This difference can cause varying levels of effort a member need to pay to derive benefits. For example, a design with complex authentication increases reading costs and hinders members' reading behaviors. Also, an unfriendly text editor can result in the cost of participation such as posting and replying. Hence, one important attribute in the community level is the cost of participation, which includes two aspects: ***reading costs and posting costs***.

4.3 Member

The micro level entity in this conceptual framework is the member. Members are a kind of agents who have memberships. An agent chooses to enter or exit a community to gain or lose his membership. Once in a community, an agent becomes a member and can conduct reading or posting behaviors. The behavior is driven by the benefits and costs embedded in the design of the community. In other words, an agent has two important attributes, which are member benefits and behaviors.

Firstly, we model the six types of benefits related to members reading and posting behaviors. In Table 4.3.1, we provide an overview of these benefit implemented in the model. Pseudo codes for them are added in the Appendix.

Table 4.3.1 Six Sources of Member Benefits

Benefits	Manipulation	Theory
Accessing Information	A member can gain benefits when the topic matches his interests; The benefit is marginally decreasing as the number of messages increases.	Resource-based Theory, Information overload theory
Sharing Information	A member gains benefits when he believes the group is small or others don't contribute too much	Collective effort model
Identity-Based Attachment	A member gains benefit when his interest matches with the community's target topic.	Group identity theory
Bond-Based Attachment	A member gains benefit when he has reciprocal interaction with others.	Reciprocity
Enjoyment	A member gains benefit when he feels enjoyed, different among individuals.	Intrinsic values
Reputation	A member gains benefit when his contribution is recognized by others in the community	Extrinsic values

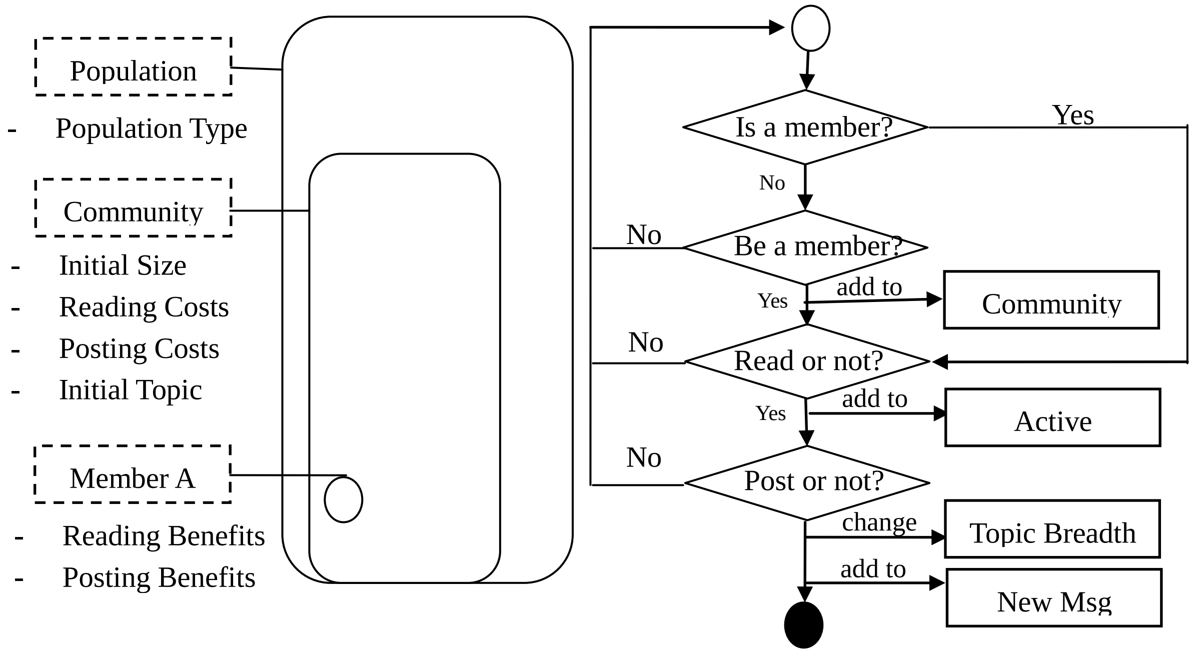
Secondly, we model four different behaviors of an agent. (1) Entry behavior: In each period, an agent who is not a member of the focal community will consider or not to sign up to become a member. The probability is proportional to community size as well as community activity (Priem and Butler 2001). (2) Reading behavior: for an agent who is already a member of the focal community, he can decide whether or not to read message threads based on the comparison between personal reading benefits and the community's reading costs. When the agent starts to read, he will look through messages in a reverse chronological order (i.e., most recently posted messages first) and stops when he runs out of time, energy or when there are no additional messages to read. (3) Posting behavior: for an agent who has read threads during a session, he can decide to post a new thread or reply to an existing thread base on his posting benefits and the community's posting costs. For posting a new thread, the thread topic is related to the poster's personal interests as well as threads he reads during the session; for replying, the topic is related to personal interests, threads read and topic of the replied thread. (4) Exit behavior: A member who doesn't post for a week consecutively will exit the community and never come back.

Thirdly, we make connections between member benefits and behaviors based on the literature. To be specific, reading behavior is driven by reading benefits which include information access benefits and enjoyment benefits. And posting behavior is motivated by posting benefits which includes sharing information benefits and reputation benefits. Moreover, both behaviors are affected by identity-based attachment and bond-based attachment benefits.

4.4 Treatments of Online Community

In this thesis, we seek to explore the effects of four constructs on communities' performance. (1) Population: similar to traditional organizations, an online community can be largely shaped by the environment, where the nature of the population of target members is the most important element. (2) Initial size: the initial members' contributions can make a large difference in the communities' development and evolution. They can help to conduct propagation through word of mouth as well as through threads. (3) Reading and posting costs: Different interface design and membership policy can induce various reading and posting costs. For example, a public discussion forum has lower reading costs compared with a private one that asks for strict authentication. Both design principles have their own strengths. On the one hand, low costs can lower the entry barrier, which encourage more participation. On the other hand, high costs can filter loyal and core members who help the community's development. When two communities coexist, which principle can bring advantages is of our interest. (4) Target topic: a community can be a general discussion forum covering a variety of topics or a focused forum dealing with only one topic. The effect of this mechanism design can be twofold. For one thing, a broad topic community can attract more users by its ample and various contents. For another, users can be tired of searching useful information when they have unique topic concern. To summarize, the conceptual framework is shown as followed in Figure 4.4.1.

Figure 4.4.1 Conceptual Framework of Agent Based Model



5 Simulation Experiments and Results

5.1 Simulation Process

In this section, we describe the simulation experiment by varying the five parameters under two different scenarios: single community and two communities. The five parameters are those a community designer can control as introduced in Section 4.4. We first experiment under a single community scenario to observe the dynamics of performance measurement and to provide a comparable baseline for the two-community scenario. After that, we run for two-communities to obtain the results relevant to our research objectives.

The details of experimental setup are as follows. (1) The total number of potential users in the population is fixed at 5000 agents; (2) There are four different kinds of populations based on users' interest profiles; (3) Initially there are two types of community sizes; (4) There are three levels of reading costs as well as posting costs. We assume that posting costs are always higher than reading costs because members consume more time when posting compared to when just reading. Therefore, there are three pairs of read-post costs in total; (5) There are two types of community topical breadth – broad vs. focused. These are summarized in Table 5.1.1. We note that although this thesis uses specific values as an example, we conduct sensitivity analysis of the parameters to check the robustness of the results.

Table 5.1.1 Initial size, Participation Cost and Topical Breadth

Size	Value	Read,Post	Value	Topic	Value
Large	500	Low,Medium	0.2,2	Broad	$[0.1,\dots,0.1]_{10}$
Small	100	Low,High	0.2,3.8	Focus	$[1,0,\dots,0]_{10}$
		Medium,High	2,3.8		

To explore the dynamics of the communities, we monitor 4 performance measurements (Table 5.1.2).

Table 5.1.2 Measurement Performance

Measurement	Details
Group Size Over Time	The number of memberships.
Active Group Size Over Time	The number of members who exhibit reading behaviors
Topic Breadth Over Time	Topic breadth of a group, inversely measured by HHI[3].
New Message Post Over Time	The number of messages posted every day.

The simulation procedure is as follows:

(1) In period $t = 0$, initialize the population of potential members according to the population interest generation process for the treatment .

(2) In period $t = 0$, initialize the community(communities) with members according to its(their) initial community size(s) by randomly drawing

[3] HHI: Herfindahl-Hirschman Index. It is the sum of the squares of the topic shares in the community. High HHI (near 1) means the community is totally concentrated on one topic and low HHI(near 0) means the community's topics are uniformly distributed.

agents from the population.

(3) In each period $t = 1, \ldots, 150$, agents surf on the Internet in a random sequence. For non-member agents, they will first consider whether to sign up for in a community based on the community size and community activity. If so, they login the community and participate activities starting from the next period; for members, they will consider whether to login to the community and start reading posts/threads.

(4) In each period $t = 1, \ldots, 150$, a member decides to login according to her expected reading benefit and the community's reading costs. A member reads threads until he runs out of messages or energy, which is determined by the benefits to her. After reading, a member needs to decide whether to post a new thread or reply a comment to an existing thread.

(5) In each period $t = 1, \ldots, 150$, a member decides to post according to her expected posting benefit and the community's posting costs. A member will post a new thread with probability of 0.5 or reply a comment. If a member doesn't post for seven days consecutively, she will leave the community and lose her membership.

(6) Repeat step 2-5 for 200 rounds by using Monte Carlo methods.

(7) Repeat step 1-6 across four population types.

5.2 Single Community

Experimental Design

In this section we choose a community with large initial size, medium reading costs, high posting costs and broad topic as a baseline to investigate the single community dynamic. We intend to validate that the community model works as expected in this context. Furthermore, the single community's performance can be used as a control group for comparing the case when two communities' coexist.

Simulation Results

In Figure 5.2.1, we show that community size and active member size change over time. We highlight that we choose general interest population to illustrate the findings because results are similar across population types. Firstly, the stable state of community size is around 5000, which means that most of members post at least once within a week and few people leave the community. Secondly, half of members are active and have reading behaviors at stable state. Thirdly, the gap between community size and active member size is enlarged over time, which suggests that members are reducing their frequency of visiting this community, the negative side of large size. Finally, 95% confidential intervals of community size and active member size are narrow. This shows the robustness of our results.

Figure 5.2.1 Single Community-Community Size and Active Member Size

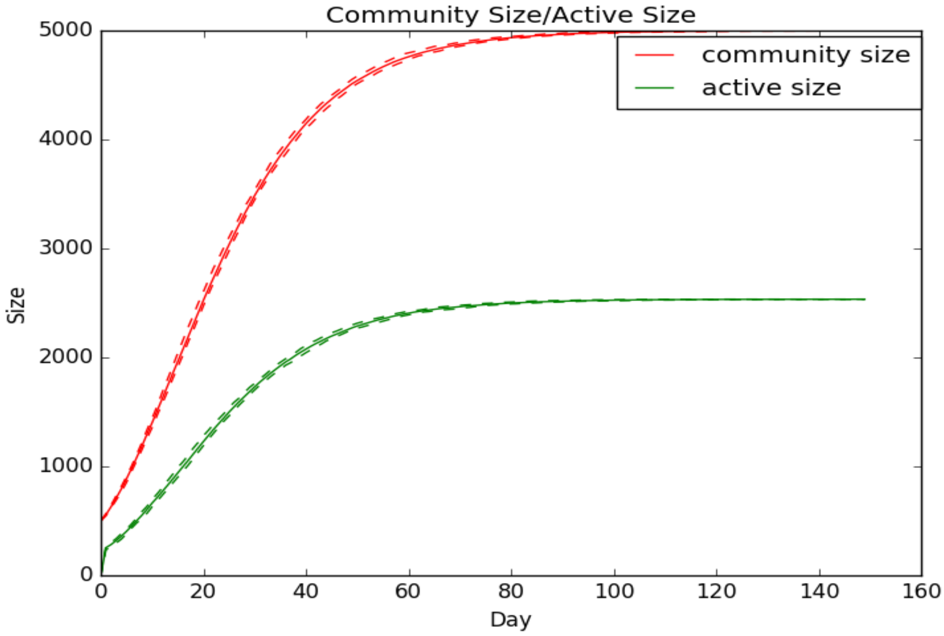
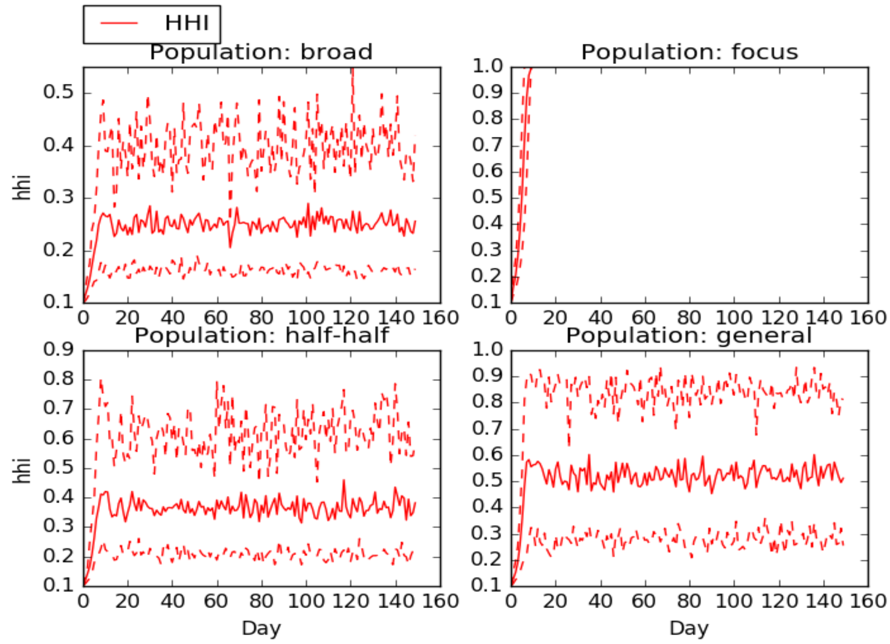


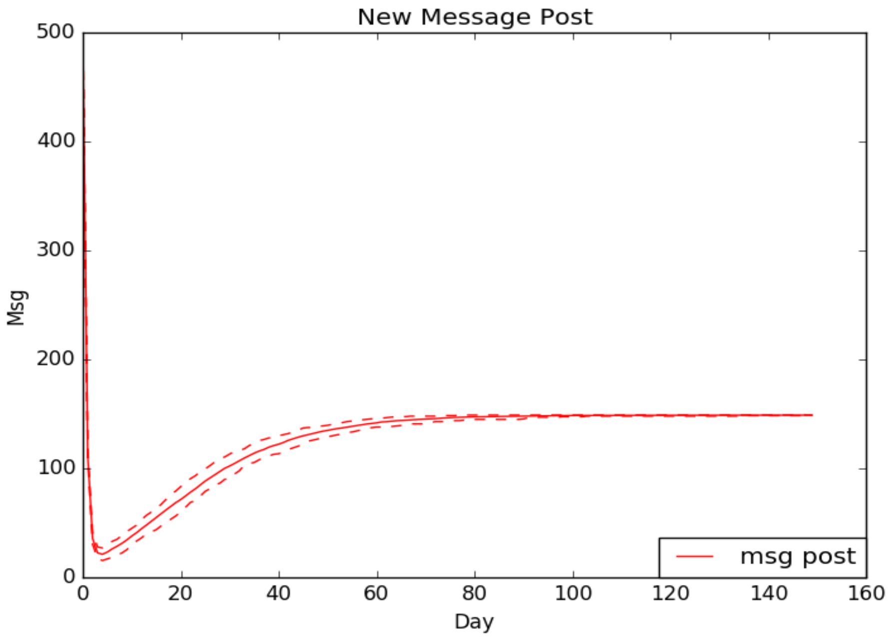
Figure 5.2.2 shows us the topic breadth of the community over time. First of all, an obvious conclusion is that population type can largely affect the topic breadth of a community. When members have flattened interests, the topic breadth of the community seems diverse. However, when members are especially interested in one topic, this community will turn into a focused one even though it starts off as a community supporting a broad set of topics. Moreover, compared to focused interest population, other three types have more flexible topic breadths, where their 95% confidential intervals are broader.

Figure 5.2.2 Single Community-Topic HHI



In Figure 5.2.3, we show new messages posting of the focal community. Consistent with Figure 5.2.1, we choose the case of general interest population to analyze the results. Firstly, we highlight the sharp drop at the beginning is induced by program design, where we assume each member will post a message initially (the high level of new message posting at time 0) and only 20% messages can be saved from one day to next (drop quickly in first few days). Therefore, this phenomenon does not have much theoretical implications. Secondly, comparing the number of active members in Figure 5.2.1, that of members who post and contribute remains low which is around 150 at the stable state.

Figure 5.2.3 Single Community-New Messages Post



5.3 Two Communities

Experimental Design

In this section, we use the previous single community as a baseline and compare it with an opponent community with only one parameter changed to test the main effects. Then, we change two parameters each time to test the interaction effects. For each scenario, we used Monte Carlo method over 200 rounds. The results are robust when we choose a different community setting as the baseline. Table 5.3.1 is used to summarize all scenarios.

Table 5.3.1 Two Communities' Scenarios

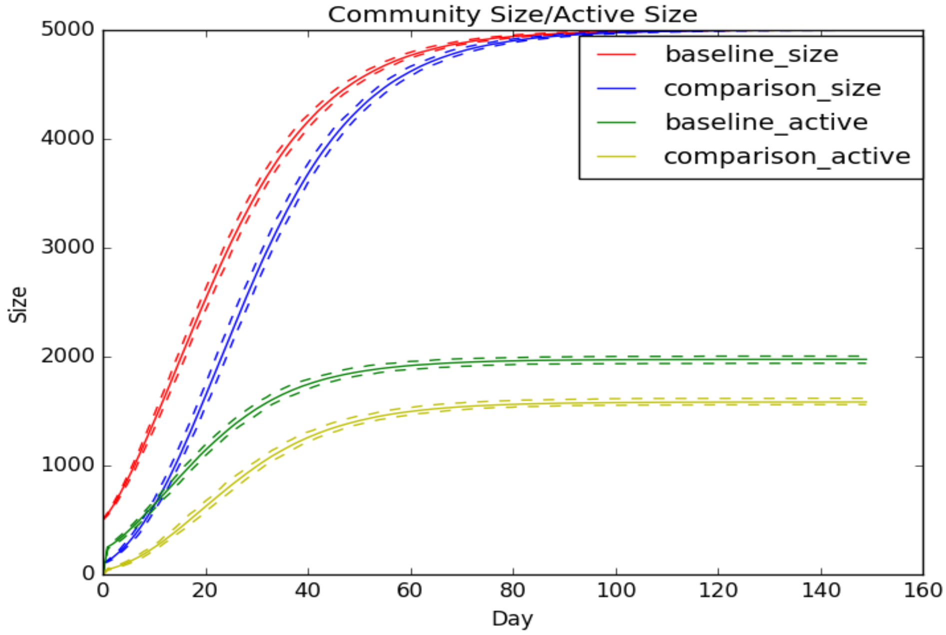
Main Effect	
Initial Size	(Large,Medium,High,Broad) v.s. (Small,Medium,High,Broad)
Cost	(Large,Medium,High,Broad) v.s. (Large,Low,MediumBroad)

	(Large,Medium,High,Broad) v.s. (Large,Low,High,Broad)
Topic Breadth	(Large,Medium,High,Broad) v.s. (Large,Medium,High,Focus)
Interaction Effect	
Size x Cost	(Large,Medium,High,Broad) v.s. (Small,Low,Medium,Broad)
	(Large,Medium,High,Broad) v.s. (Small,Low,High,Broad)
Size x Topic	(Large,Medium,High,Broad) v.s. (Small,Medium,High,Focus)
Cost x Topic	(Large,Medium,High,Broad) v.s. (Large,Low,Medium,Focus)
	(Large,Medium,High,Broad) v.s. (Large,Low,High,Focus)

Simulation Results

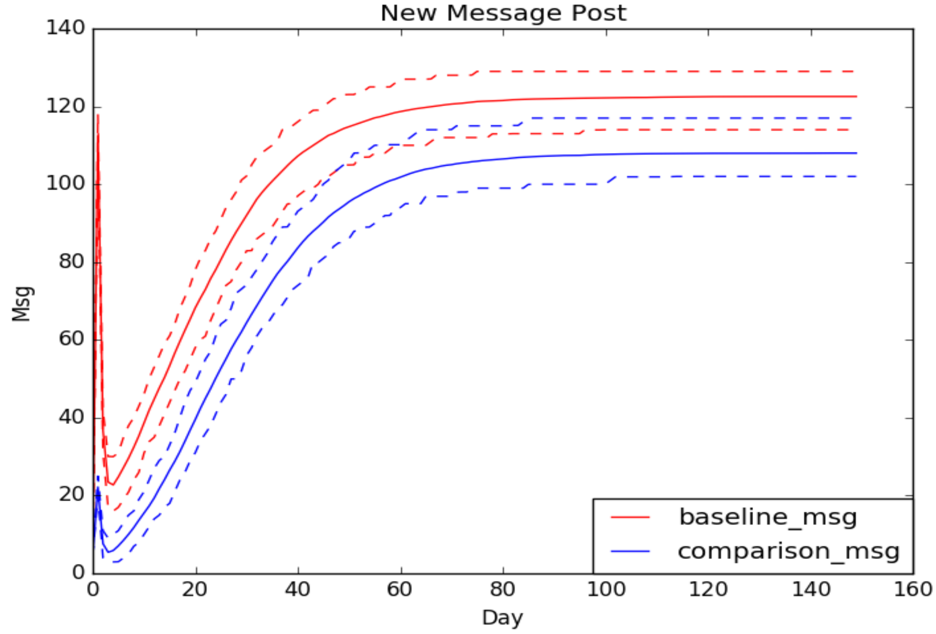
Main effect – Initial size. We check the main effect of initial size. Because we don't observe much difference across populations, we choose the scenario of the general population to highlight the results. In this scenario, the baseline is a large community with 500 members initially while the comparison has just 100 members. As we can see in Figure 5.3.1, in terms of member size, an initially large community doesn't necessarily maintain its initial advantage over the small one at the stable state. However, that the smaller community has a sharper gradient implies that smaller groups can attract members more rapidly than its larger opponents. As for active member size, initially large community does have advantages over small one at the stable state with a gap of more than 500 members. The results are robust in 200 rounds since 95% confidence intervals are very narrow.

Figure 5.3.1 Two Communities-Effect of Size on Community Size and Active Member Size



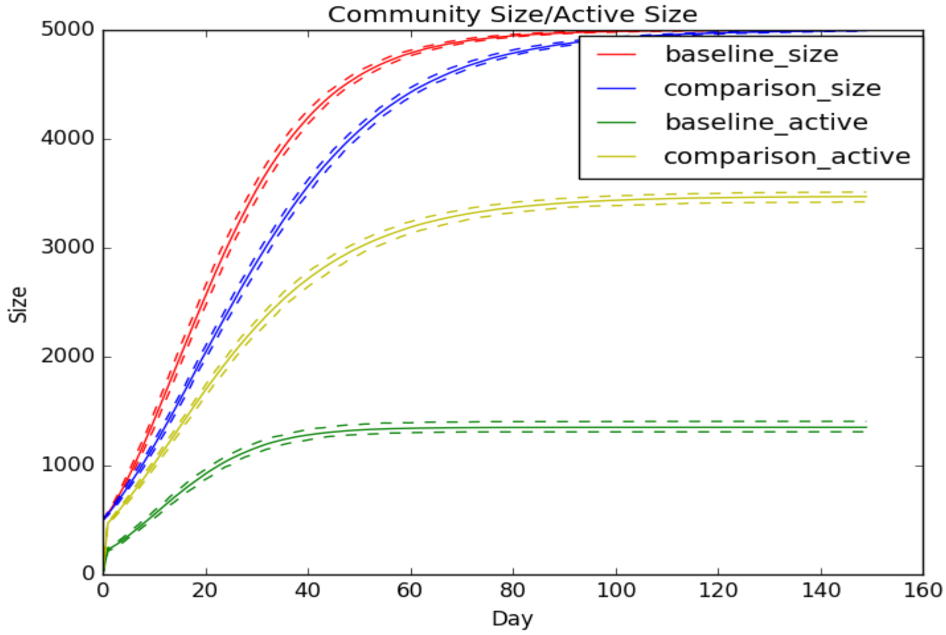
Initial size not only has effects on community size and active member size, but also on the community activity such as posting of new messages. As shown in Figure 5.3.2, the difference between large and small communities in new messages posting is similar to that of active member size, where the larger community has advantages.

Figure 5.3.2 Two Communties-Effect of Size on New Message Posting



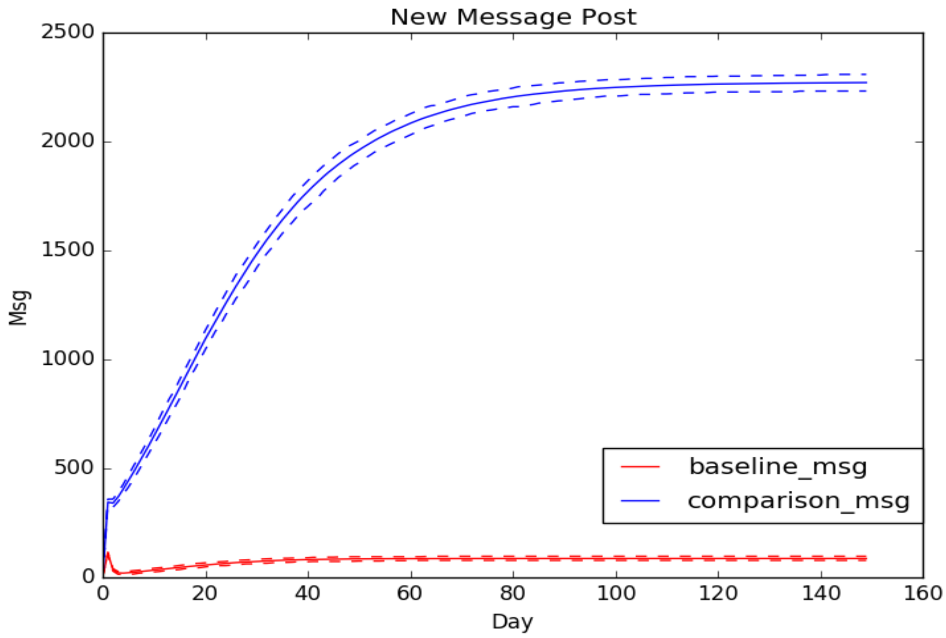
Main effect – Participation Cost We check the effects of participation cost on community size and active member size by altering reading and posting costs at the same time. The baseline community is of medium reading costs and high posting costs while the comparison community is of low reading costs and medium posting costs. As shown in Figure 5.3.3, in terms of community size, participation cost makes no difference between the baseline and comparison communities at the stable state where both communities reach around 5000 members. As for active member size, participation cost is a significantly influential factor. With lower participation costs, the comparison community encougages members to read more on it and hence attract more active members compared to the baseline community.

Figure 5.3.3 Two Communities-Effect of Cost on Community Size and Active Member Size



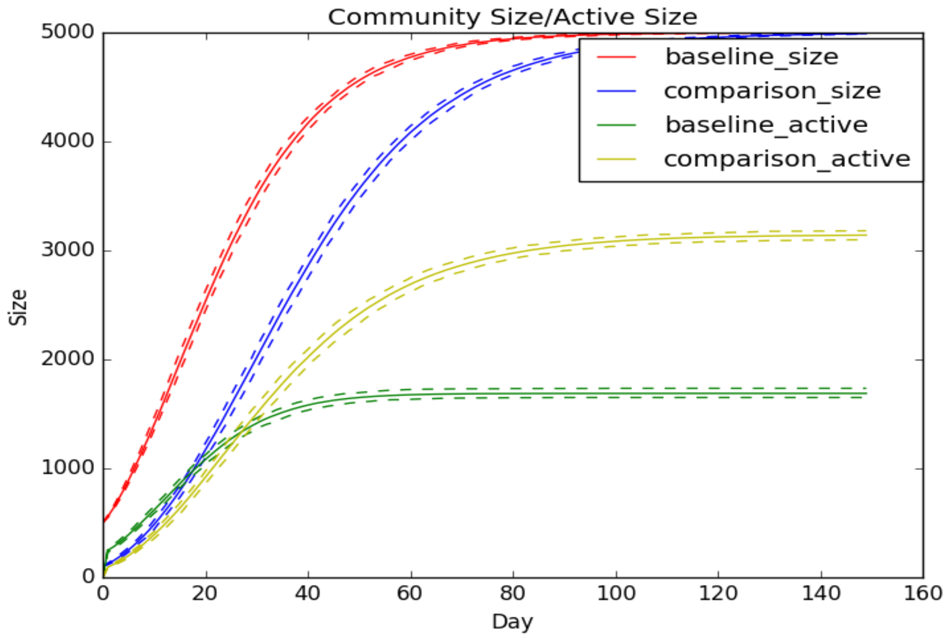
At the same time, community activities such as new messages posting are also largely affected by participation costs. This is shown in Figure 5.3.4. The results are as expected because with low posting costs, the community design for participating is friendlier and hence members are more willing to post new threads and reply with comments.

Figure 5.3.4 Two Communties-Effect of Cost on New Messages Post



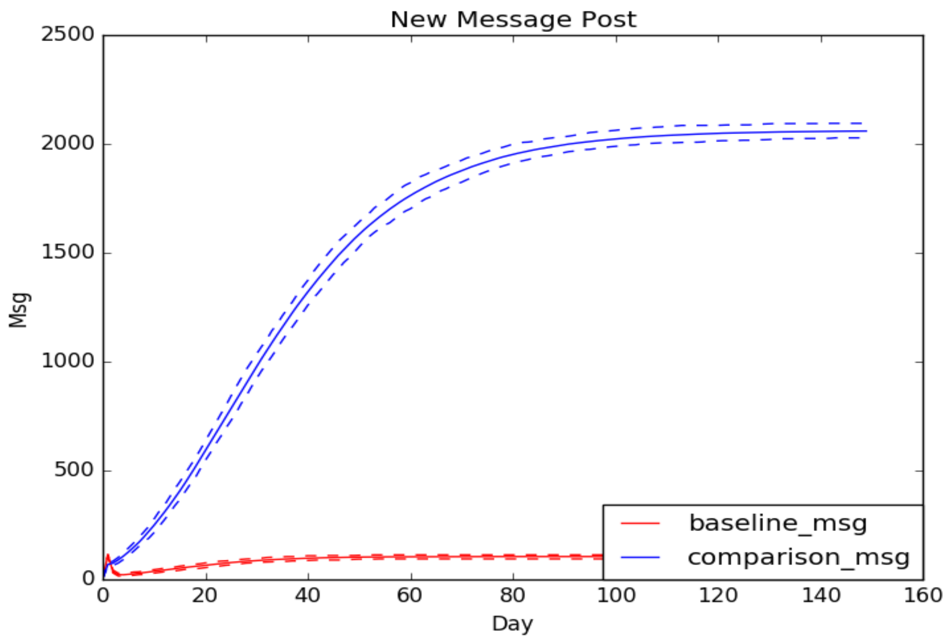
Interaction effect – Size and Cost. In Figure 5.3.5, we show the interaction effect of initial size and participation costs on community size and active member size. Because there are not many differences across populations, we choose the general generation to highlight the results. Firstly, the results are robust with 95% confidence intervals being very narrow. Secondly, at stable state, both baseline and comparison community reach same level of community size. The difference relies in the slope of two communities, which implies with limited potential member population, small and more friendly designed (low participation cost) community can attract new members more rapidly compared to larger but less friendly ones. Finally, for active member size, it is shown that although initially large community can help to gain more users at the beginning, the gap is narrowed because participation costs of the small community are low, which allows the smaller community to finally overtake its opponent.

Figure 5.3.5 Two Communties-Interaction Effect of Size and Cost on Community Size and Active Member Size



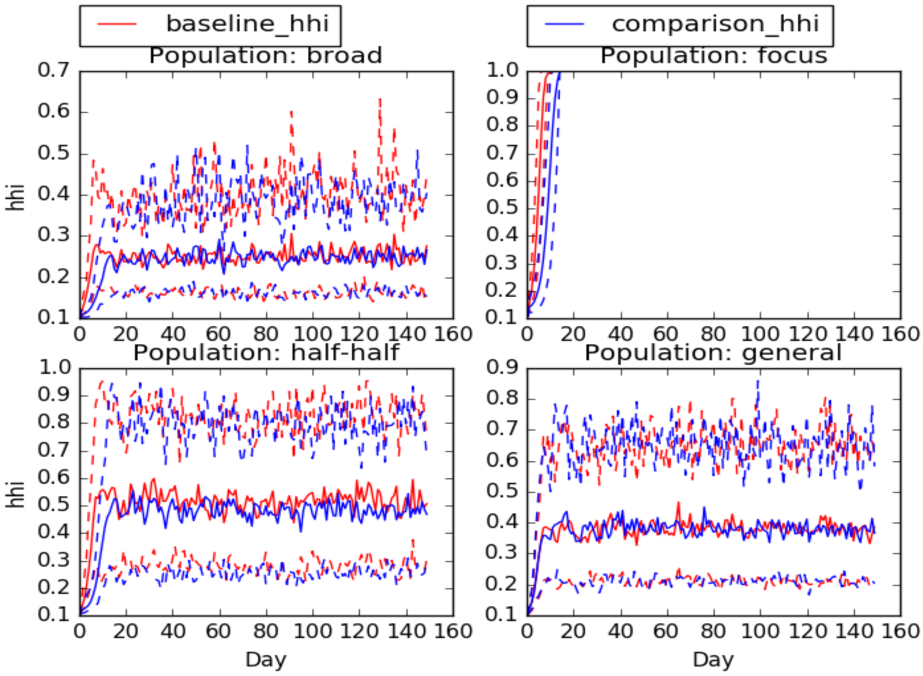
In Figure 5.3.6, we check the influence of initial size and participation costs on community activity. This result again shows the important interaction effects of both factors. At the beginning, large community generates more messages because of its size. However, over time, the small community catches up and overtakes it very quickly.

Figure 5.3.6 Two Communities-Interaction Effect of Size and Cost on New Message Post



Main effect – Population. In Figure 5.3.7, we show the effect of population type on community topic breadth. Although the scenario we choose is that the baseline and comparison community are different from each other only in their initial size, the results are very similar across different situations. We check the influence of every factor such as size and costs on topic breadth, none of them but population type makes significant difference. For example, in a focused population, both communities whose initial topics are focused converge to 1 in topic HHI, which implies both communities focus on one topic too. Another point we need to note is that the topic breadth has a large variation in all the population types except the focused one. This can be explained by the flexibility of broad interest members.

Figure 5.3.7 Two Communities-Effect of Population on Community Topic HHI



Propositions. We summarize the results we derive visually in Table 5.2.2.

Table 5.3.2 Propositions

Proposition	Description	Effect	Figure
1	Initial size positively affects active member size.	Main	5.3.1
2	Initial size positively affects main new messages posted.	Main	5.3.2
3	Participation cost negatively affects active member size.	Main	5.3.3
4	Participation cost negatively affects new messages posted.	Main	5.3.4
5	Population types have influence on topic breadth.	Main	5.3.7
6	Initial size and participation cost have interaction effect on active member size.	Interaction	5.3.5
7	Initial size and participation cost have	Interaction	5.3.6

interaction effect on new messages posted.

6 Regression Model and Results

In this section, we use a simple regression model to validate the proposition we derive in previous section.

6.1 Simple Regression Model

The following module is used:

$$Y_{base-comp} = \alpha_i + \beta_{i1} ppl_{broad} + \beta_{i2} ppl_{focus} + \beta_{i3} ppl_{halfhalf} + \beta_{i4} size_{small} + \beta_{i5} read_{low} + \beta_{i6} post_{med} + \beta_{i7} topic_{focus} + \gamma INTERACTION + \varepsilon_i$$

Where $Y = \{Size, Active, TopicHHI, Msg\}$ for community size, active member size, (inverse) topic breadth and new messages. γINTERACTION in above equations refers to: $\gamma_{i1} size_{small} * read_{low} + \gamma_{i2} size_{small} * post_{med}$, i=1, 2, 3, 4 for $Y = \{Size, Active, TopicHHI, Msg\}$ respectively.

We summarize independent variables in Table 6.1.1.

Table 6.1.1 Independent Variables

	Type	Description
ppl_{broad}	Binary	=1 if the population is broad
ppl_{focus}	Binary	=1 if the population is focused
$ppl_{halfhalf}$	Binary	=1 if the population is hybrid(half-half)
$size_{small}$	Binary	=1 if two communities' initial sizes are different, i.e., compared community is initially small

$read_{low}$	Binary	=1 if two communities' reading costs are different, i.e., compared community has low reading costs
$post_{med}$	Binary	=1 if two communities' posting costs are different, i.e., compared community has medium posting costs
$topic_{focus}$	Binary	=1 if two communities' initial topics are different, i.e., compared community focuses on only one topic

6.2 Data Description

In this part, we collect our data by running each scenario 50 rounds. In total, we get 3600 observations to run the regression.

Table 6.2.1 Data Description

	Mean	Std	Min	Max
ΔSize	5.87	7.191	-4	37
ΔActive	-1206.99	1061.17	-2390	508
ΔTopic	0.02	0.05	-0.15	0.23
ΔMsg	-726.27	-151.00	-2256	32
ppl_{broad}	0.25	0.433	0	1
ppl_{focus}	0.25	0.433	0	1
$ppl_{halfhalf}$	0.25	0.433	0	1
$size_{small}$	0.44	0.496	0	1
$read_{low}$	0.67	0.470	0	1

$\boldsymbol{post_{med}}$	0.33	0.470	0	1
$\boldsymbol{topic_{focus}}$	0.44	0.496	0	1

6.3 Regression Results

For each performance measure, we first run the regression for the main effect. Then we add interaction effect into the model. The results are consistent with our proposition.

In Table 6.3.1, we show the regression results of community size. Although we don't observe much difference in community size from previous part, in both models, the main effects of initial size and participation cost are well recognized (p<0.05). This indicates that initially large community has its advantage over the small one with 3.5 and 0.9 larger in main effect and interaction effect models respectively. Also, different from what we can expect, a community with high participation costs can lead to higher community size. In model 2, the results imply that there exists an interaction effect between size and participation costs. For example, a small community whose posting costs are high can attract 6.8 more members on average. Even though the coefficients are significant, they are much smaller when compared with the total number for community size.

Table 6.3.1 DV: Community Size

	Model1-Main Effect			**Model2-Interaction Effect**		
	B	Std.Error	sig	B	Std. Error	Sig
(Constant)	-1.485	.191	.000	.270	.186	.147

ppl_broad	.035	.144	.811	.035	.124	.780
ppl_focus	-.164	.144	.255	-.164	.124	.185
ppl_halfhalf	-.010	.144	.946	-.010	.124	.937
size_small	3.533	.125	.000	.902	.170	.000
read_low	2.507	.144	.000	1.451	.170	.000
post_med	12.590	.125	.000	10.316	.131	.000
topic_broad	-.110	.125	.379	-.110	.107	.305
size_small*read_low				.535	.227	.019
size_small*post_med				6.824	.227	.000

In Table 6.3.2, we show the regression results with active member size as the dependent variable. Most main effects are significant with $p<0.05$. The most influential main effect is that of reading cost. A reduction in reading cost can result in around 2100 more active members according to Model 2. The initial size still acts as a positive factor impacting active member size. The interaction between with size and reading costs shows that community size can ease the problems brought by opponents' reduced cost advantage to certain degree. An interesting result is that population type actually affect the active member size. Compared to the general population type, the baseline community suffers active member loss with coefficient of -70.9 in the population with broad interests and gains members with coefficient of 18.9 in that with focused interests.

Table 6.3.2 DV: Active Size

	Model1-Main Effect			**Model2-Interaction Effect**		
	B	Std.Error	sig	B	Std.Error	sig
(Constant)	-90.271	5.449	.000	39.379	4.110	.000

ppl_broad	-70.859	4.119	.000	-70.859	2.740	.000
ppl_focus	18.923	4.119	.000	18.923	2.740	.000
ppl_halfhalf	-62.099	4.119	.000	-62.099	2.740	.000
size_small	543.105	3.567	.000	348.630	3.752	.000
read_low	-2015.710	4.119	.000	-2171.925	3.752	.000
post_med	44.031	3.567	.000	32.337	2.906	.000
topic_focus	-1.043	3.567	.770	-1.043	2.373	.660
size_small*read_low				274.171	5.034	.000
size_small*post_med				35.082	5.034	.000

In Table 6.3.3, we summarize the results of the regression analysis with topic breadth(HHI) as the dependent variable. Most of the main effects are significant except community size. In the general population, with the constant equals to 0.025, the baseline community has a relatively focused topic than the compared community. However, in both broad and focused populations, the results reverse with the differences of (0.025-0.034=) 0.009 and (0.025-0.035=) 0.01. Although significant, these differences are really negligible compared with the range of HHI, which is [0,1]. Combined with what we observe in Section 5.3, we can conclude that the population type affects both communities' topic breadths simultaneously but it does not induce many differences between them. Moreover, participation costs can affect the difference of topic breadth. Higher participation costs will increase the topic breadth by 0.01 (i.e., make the topics slightly more focused). This may implies that high costs filter out core members and hence generate more focused topics.

Table 6.3.3 DV: Topic Breadth (*HHI*)

	Model1-Main Effect			Model2-Interaction Effect		
	B	Std.Error	sig	B	Std. Error	Sig
(Constant)	.025	.003	.000	.026	.003	.000
ppl_broad	-.034	.002	.000	-.034	.002	.000
ppl_focus	-.035	.002	.000	-.035	.002	.000
ppl_halfhalf	.014	.002	.000	.014	.002	.000
size_small	-.001	.002	.653	-.003	.003	.331
read_low	.009	.002	.000	.008	.003	.002
post_med	.010	.002	.000	.008	.002	.000
topic_focus	.001	.002	.449	.001	.002	.449
size_small*read_low				.001	.003	.827
size_small*post_med				.004	.003	.271

In Table 6.3.4 summarizes the regression results with new message posts as the dependent variable. As expected, initial size and participation costs do affect new messages posting significantly (p<0.05). The positive coefficient of initial size indicates that the large size community can induce more posting of messages. Both reading and posting costs negatively influence the gap between the baseline and the comparison. This implies that lower participation costs lead to higher message volume. In terms of population, compared with all other groups, the baseline generates more messages on average in the focused population and less messages in the three other population types. In model 2, the interaction effects are significantly positive. This means that although high participation costs can negatively affect message posting but the initially larger community size can ease this problem.

Table 6.3.4 DV: New Messages Post

	Model1-Main Effect			Model2-Interaction Effect		
	B	Std.Error	sig	B	Std.Error	Sig
(Constant)	-82.766	3.661	.000	-32.881	2.717	.000
ppl_broad	6.157	2.767	.026	6.157	1.811	.001
ppl_focus	54.480	2.767	.000	54.480	1.811	.000
ppl_halfhalf	24.451	2.767	.000	24.451	1.811	.000
size_small	93.579	2.397	.000	18.752	2.480	.000
read_low	-115.237	2.767	.000	-145.838	2.480	.000
post_med	-1888.175	2.397	.000	-1951.686	1.921	.000
topic_focus	-.330	2.397	.891	-.330	1.569	.833
size_small*read_low				16.974	3.327	.000
size_small*post_med				190.533	3.327	.000

7 Discussion and Limitation

In this section, we firstly summarize our main finding. Then we discuss their theoretical and practical contributions. Finally, we point out the limitations and potential future research to extend this study.

The major findings are twofold. On the one hand, we find several useful propositions from the perspective of two communities' coexistence ecology. Firstly, we find a strong main effect of initial community size on both active member size and new messages posting. An initially larger community can attract more active members who frequently login and read threads. Also, with more reading behaviors, members are more likely to post new threads or make comments. Second, we find the main effect of communities' participation cost on both active member size and new message posting. Lower participation costs can encourage members to post more threads and enhance the performance of community in terms of activity. Third, there exists an interaction effect of initial community size and participation costs. For an initially small community, lower participation costs can compensate for this shortcoming and lead to a larger of active member size and greater community activity. Finally, an interesting finding is that the community's topic is significantly affected by the population types regardless of what the community's initial topic is set to be.

On the other hand, we can derive more insights by comparing the results of single community and that of two communities. First of all, communities can make the cake bigger in the market through cooperation. However, the coexistence can also bring competitions between communities that do not promise the win-win situation even if the market is larger. For example, in one community scenario, active member size is around 2500 at the stable state. When two communities coexist, this number of both communities in total is enlarged to 3500~4500. Nevertheless, the number of the baseline alone decreases to 1500~2000. Furthermore, core members, who frequently post in the community, are not affected by its challengers but by its easiness of community designs (i.e., participation costs). For example, in one community scenario, the number of new messages posting and hence core members at the stable state is around 150. It remains quite constant in coexistence scenario whatever its opponent is. Last but not the least, ecological coexistence can affect the convergence of topic breadth of focal community in different ways under various population types. For example, in purely broad and focused interest population, the baseline converges to 0.25 in both scenarios. However, in hybrid interest population (half-half), the baseline's topic breadth is more focused in coexistence scenario than that of single community (0.5 v.s. 0.35). while the situation is reversed in general interest population(0.4 v.s. 0.5).

Our findings have both theoretical and practical contributions. In terms of the theoretical literature, there are two streams of related research. On the one hand, the organizational ecology literature points out that limited resources force organizations to compete with each other and the most suitable one survives. On the other hand, the online community literature explores intrinsic and extrinsic factors that drive people to participate in online communities. This paper fills the gap between these two literatures by introducing influential factors from the online community literature to the context of ecological coexistence of two communities. Firstly, we show that some conclusions in organizational ecology literature can be used in the online community context although there exist differences between online community and traditional organizations as we state previously. For example, the coexistence can enhance the development of online communities as a whole when they cooperate. At the same time, it can also bring competitions that do not promise the win-win situation. Secondly, different from traditional organizations, when faced with challenges, online communities do not lose their core members easily, the size of which is largely decided by participation costs under community operators' control. Finally, we find that topic breadth of online communities is largely shaped by user population, and this is not mentioned by previous literatures as far as we know. Also, the ecological coexistence can change the convergence of topic breadth.

In terms of practical contributions, this work implies that a community designer can think about popularizing their service initially to achieve a higher level of initial size. Also, we highlight the importance of mechanism design to promote the participation (i.e., different ways to reduce reading and posting cost). Moreover, it is suggested that a community is shaped in large part by its members therefore the nature of the target market (i.e., the discussion of interests) also need careful consideration.

There are also limitations in this study. Firstly, we assume two communities coexist under a fully overlapping population. This is true given the ‘global reach’ characteristic of the Internet. However, more and more communities start to consider refining the markets thus the target users may not be as homogeneous as our model sets them up to be. This implies that member overlap can be investigated further. Secondly, although we allow members to choose threads to which they are most likely to reply, we do not directly consider user similarity in this context directly, which however is frequently observed in reality. Therefore, our model does not reflect the homophily phenomenon. Finally, we do not have empirical data to test what we find in the simulation part. Future works may manipulate the model parameters and modify them to adapt different types of communities for more nuanced insights.

8 Appendix: Member Benefits

Pseudo-code for calculating benefit of accessing information

Initial information accessing benefit to zero

FOR <message> IN <messages read today>:

Calculate marginally diminishing factor f(n)

Calculate match between <member's ***interest***> and <message's ***topic***> SUM(interest*topic)

Increase information accessing benefit by f(n)*SUM(interest*topic)

Pseudo-code for calculating benefit of sharing information

Initial information sharing benefit to zero

IF <intrinsic benefit> or <identity benefit> or <bond benefit> GREATER THAN 0.3:

IF <***numbers*** of messages post today> LESS THAN <community ***size***>:

    Increase information sharing benefit by 0.5*(size- numbers)/ size

IF <member's ***contrib***ution> GREATER THAN 0.5*<***maximum*** of contributions>:

    Increase information sharing benefit by 0.5*(contrib-0.5*maximum)/contrib

IF <community ***size***> GREATER THAN 1000:

    Mutiply information sharing benefit by 1-(size-1000)/(size+1000)

Pseudo-code for calculating benefit of identity

Initial identity benefit to zero

Calculate the match between <member's ***interest***> and <community's current

topic ***trend***>

Increase identity benefit by SUM(interest*trend)

Pseudo-code for calculating benefit of bond

Initial bond benefit to zero

FOR <member***_i***> and <member***_j***> in <community’s member list>:

IF member i and j interact with each other ever:

    Calculate the marginal diminishing factor f(m)

    Increase benefit of bond by f(m)*1

Pseudo-code for calculating benefit of enjoyment

Initial enjoyment benefit to zero

Draw enjoyment benefit from uniform distribution

Pseudo-code for calculating benefit of reputation

Initial reputation benefit to zero

IF <member’s ***contrib***ution> GREATER THAN 0.1*<***maximum*** of contributions>

    Increase reputation benefit by 1-0.1*maximum/contrib

9 References

Aldrich, H. and E. R. Auster (1986). "Even dwarfs started small: Liabilities ofage and size and their strategic implications." Research in organizational behavior **8**(1986): 165-186.

Astley, W. G. and C. J. Fombrun (1983). "Collective Strategy: Social Ecology of Organizational Environments." Academy of management review **8**(4): 576-587.

Barnett, W. P. and G. R. Carroll (1987). "Competition and mutualism among early telephone companies." Administrative Science Quarterly: 400-421.

Bateman, P. J., et al. (2011). "Research note-the impact of community commitment on participation in online communities." Information Systems Research **22**(4): 841-854.

Baum, J. A. and C. Oliver (1996). "Toward an institutional ecology of organizational founding." Academy of Management Journal **39**(5): 1378-1427.

Davis, J. P., et al. (2007). "Developing theory through simulation methods." Academy of management review **32**(2): 480-499.

Grant, R. M. (1991). "The resource-based theory of competitive advantage: implications for strategy formulation." California management review **33**(3): 114-135.

Hogg, M. A. and D. I. Terry (2000). "Social identity and self-categorization processes in organizational contexts." Academy of management review **25**(1): 121-140.

Horrigan, J. (2001). Online communities: Networks that nurture long-distance relationships and local ties, Pew Internet & American Life Project.

Jones, Q., et al. (2004). "Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration." Information Systems Research **15**(2): 194-210.

Karau, S. J. and K. D. Williams (1993). "Social loafing: A meta-analytic review and theoretical integration." Journal of personality and social psychology **65**(4): 681.

McKenzie, R. D. and A. H. Hawley (1968). On human ecology: Selected writings, University of Chicago Press.

Priem, R. L. and J. E. Butler (2001). "Is the resource-based “view” a useful perspective for strategic management research?" Academy of management review **26**(1): 22-40.

Ren, Y., et al. (2007). "Applying common identity and bond theory to design of online communities." Organization studies **28**(3): 377-408.

Ren, Y. and R. E. Kraut (2014). "Agent-based modeling to inform online community design: Impact of topical breadth, message volume, and discussion moderation on member commitment and contribution." Human–Computer Interaction **29**(4): 351-389.

Ridings, C. M. and D. Gefen (2004). "Virtual community attraction: Why people hang out online." Journal of Computer-Mediated Communication **10**(1): 00-00.

Russell, R. D. and C. J. Russell (1992). "An examination of the effects of organizational norms, organizational structure, and environmental uncertainty on entrepreneurial strategy." Journal of management **18**(4): 639-656.

Wang, X., et al. (2013). "The impact of membership overlap on growth: an ecological competition view of online groups." Organization Science **24**(2): 414-431.

Wasko, M. M. and S. Faraj (2005). "Why should I share? Examining social capital and knowledge contribution in electronic networks of practice." MIS quarterly: 35-57.