

TOWARD A SUSTAINABLE ONLINE Q&A COMMUNITY VIA DESIGN DECISIONS
BASED ON INDIVIDUALS' EXPERTISE: EVIDENCE FROM SIMULATIONS

By

Yuyang Liang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Media and Information Studies—Doctor of Philosophy

2020

ABSTRACT

TOWARD A SUSTAINABLE ONLINE Q&A COMMUNITY VIA DESIGN DECISIONS BASED ON INDIVIDUALS' EXPERTISE: EVIDENCE FROM SIMULATIONS

By

Yuyang Liang

Online Q&A communities have become an important channel for internet users to seek information and share knowledge. Existing research extensively focuses on the individual components of Q&A communities, such as content quality and user characteristics, but fails to provide a comprehensive understanding of the communities as complex social systems, whose behavior depends on the interactions of a large number of social agents. In this dissertation, I integrated the key components in online Q&A communities via agent-based modeling to provide a systematic examination of Q&A communities and help inform better community design to manage users' expertise. I conducted computer simulations and virtual experiments based on existing findings and theories as well as data from a large online Q&A community to understand how two design decisions, including expertise indication and question routing, influence the sustainability of a Q&A community as well as result in possible trade-offs involved in implementing these design decisions. Results indicate that these design decisions are likely to lead to a larger membership size and a higher rate of solved questions. In addition, implementing design decisions will also influence the member structure of a community. Question routing tends to prioritize experts' needs and benefits while expertise indication is more likely to attract beginners. These findings suggest that these design decisions should be leveraged according to the development stage a community is in. This research also demonstrates the value of agent-based modeling in terms of generating insights for Q&A community design by showing the underlying structural outcomes of the design decisions.

Copyright by
YUYANG LIANG
2020

*This dissertation is dedicated to Mom (黄小姬) and Dad (梁福璇).
Thank you for always believing in me.*

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
Chapter 1 INTRODUCTION.....	1
Chapter 2 LITERATURE REVIEW	6
Social Q&A.....	6
Knowledge Sharing and Online Community Design.....	9
Sustainability of Online Communities.....	11
Research Framework	21
Chapter 3 METHOD.....	24
Agent-based Modeling.....	24
Model Development.....	28
Model Calibration	35
Model Validation	43
Chapter 4 RESULTS.....	45
Success Rate.....	45
Membership Size	48
Average Membership Attrition	49
Percentages of Members by Expertise Level.....	49
Percentage of Solved Questions by Expertise Level	52
Chapter 5 DISCUSSION	57
Summary of Findings.....	57
Theoretical Implications	59
Managerial Implications	62
Limitations and Future Research	64
Chapter 6 CONCLUSION	68
APPENDICES	70
APPENDIX A Goodness-of-fit Plots for Fits of Attraction Rates and Contribution Likelihood	71
APPENDIX B Model Parameters and Values	75
APPENDIX C Computer Programs for Building the Agent-based Model	76
APPENDIX D A Model with Variable Member Expertise	78
REFERENCES	82

LIST OF TABLES

Table 1 Examples of Online Q&A Communities	7
Table 2 Platform Capabilities	29
Table 3 Individual Parameters	32
Table 4 The Percentiles of the Number of Solved Questions by an Individual (from 90 th to 100 th)	37
Table 5 Interaction Between Expertise Indication and Question Routing.....	46
Table 6 Interaction Between Expertise Level and Design Decisions	55
Table 7 Model Parameters, Explanation, and Values	75

LIST OF FIGURES

Figure 1 Key Elements of Online Community Sustainability.	13
Figure 2 Conceptual Framework of the Study	22
Figure 3 Flowchart of an Agent's Decisions in a Simulation Step	42
Figure 4 Interaction Effect on Success Rate Between Expertise Indication and Question Routing	47
Figure 5 Interaction Effect on Membership Size Between Expertise Indication and Question Routing.....	47
Figure 6 Interaction Effect on Membership Attrition Between Expertise Indication and Question Routing.....	48
Figure 7 Proportion of Low Expertise Members over the Simulation Period Under Each Condition.....	50
Figure 8 Proportion of Medium Expertise Members over the Simulation Period Under Each Condition.....	51
Figure 9 Proportion of High Expertise Members over the Simulation Period Under Each Condition.....	51
Figure 10 Percentage of Solved Questions Submitted by Low Expertise Members Under Each Condition.....	53
Figure 11 Percentage of Solved Questions Submitted by Medium Expertise Members Under Each Condition.....	53
Figure 12 Percentage of Solved Questions Submitted by High Expertise Members Under Each Condition.....	54
Figure 13 Goodness-of-fit Plots for Calibrating Attraction Rate by Fitting a Normal Distribution.	71
Figure 14 Goodness-of-fit Plots for Calibrating the Contribution Likelihood of Low Expertise Members by Fitting a Log-normal Distribution.....	72
Figure 15 Goodness-of-fit Plots for Calibrating the Contribution Likelihood of Medium Expertise Members by Fitting a Log-normal Distribution.	73

Figure 16 Goodness-of-fit Plots for Calibrating the Contribution Likelihood of High Expertise Members by Fitting a Log-normal Distribution.....	74
Figure 17 Expertise Growth Curve	78
Figure 18 Proportion of Low Expertise Members over the Simulation Period with Variable Expertise	79
Figure 19 Proportion of Medium Expertise Members over the Simulation Period with Variable Expertise	80
Figure 20 Proportion of High Expertise Members over the Simulation Period with Variable Expertise	80

Chapter 1

INTRODUCTION

The Internet has become one of the most efficient and essential channels for seeking and accessing knowledge and information. Besides search engines, online question & answer (Q&A) platforms have also emerged and thrived as a source for knowledge discovery. On these community-based platforms, people can post questions to seek information and help, or participate in discussions and share their knowledge and expertise by providing answers; hence, a social aspect is introduced to the information seeking and knowledge sharing process (Shah et al., 2009). Some of the sites are generic in terms of the topics of the questions (e.g., Yahoo! Answers & Quora) while others are domain-specific (e.g., Stack Overflow).

Srba and Bielikova (2016b) summarized three major concentrations of research on Q&A communities, including exploratory Q&A system and process studies, content and user characteristics examination and modeling, and algorithms supporting knowledge recommendation and retrieval. Among all the concentrations, much emphasis is placed on predicting, recommending, and retrieving good quality questions and answers (e.g., Agichtein et al., 2008; Li et al., 2012; Shah et al., 2014; Shah & Pomerantz, 2010; Toba et al., 2014). The quality of user-generated content is fundamental to the sustainability of Q&A communities, which allows a community to consistently provide resources and benefits to its members (Butler, 2001). Good quality information not only satisfies users' immediate needs but also contributes to the overall knowledge base that can be retrieved in the future (Anderson et al., 2012; Bian et al., 2009). Meanwhile, information quality is closely related to member retention and expansion, as existing users are likely to abandon the site if it fails to offer useful information and the

community becomes less attractive to new users as well (Liang & Introne, 2019; Srba & Bielikova, 2016a).

In comparison, little has been done to systematically examine Q&A communities from a broader socio-technical viewpoint. As a complex social system, whose performance and behavior are primarily the result of a large number of social agents' (e.g., human beings) interactions and decisions and intrinsically difficult to model, a Q&A community consists of several key components: users, moderators, information (i.e., questions and answers), and technical infrastructure. The connections and interactions between these components have a significant impact on the development of the system. Yet, despite the importance of studying this socio-technical system in its entirety, the majority of previous studies have extensively examined its individual components. Indeed, the integration of the key components can lead to improvements in the design of online Q&A communities because it can help to reveal the underlying structural outcomes of design decisions. Therefore, to extend these studies from the socio-technical perspective, this study aims to integrate its various components, including users, information, and technology, in order to offer a comprehensive and systemic understanding of how social Q&A communities sustain their development via the examination of their social structures and the dynamic of interactions between their members. Building on the integration, the study will further offer insights for community designers and moderators to better evaluate the impact and the trade-offs of various design decisions that can influence a community's sustainability.

From the perspective of information system and knowledge management, social Q&A represents an expertise sharing model, which puts more emphasis on interpersonal communications of knowledgeable actors (Ackerman et al., 2013). In the previous generation of

research, which can be described as the repository model, information and knowledge exist in physical records and documents, and is considered an externalized artifact or object (Ackerman et al., 2013). This body of work focuses on the technical aspects of knowledge repositories and the social context and usage of information artifacts. Since individuals retrieve knowledge mainly by searching the repository, interpersonal interactions are minimal in the process. As online communities and social media have become popular as platforms for knowledge sharing, the research focus has thus shifted away from the repository model to an expertise sharing model, where individuals' engagement and interactions are crucial in the process. Rather than finding information by searching knowledge repositories, in this stage, individuals obtain answers to their questions by finding a person with the right expertise. However, individuals' expertise is often in the form of tacit knowledge, and thus needs to be made explicit (Ackerman et al., 2013; Nonaka & Takeuchi, 1995). Therefore, it becomes critical for moderators and the designers of Q&A communities to properly design features and manage community members' expertise in order to facilitate the knowledge sharing process. As interpersonal interactions play a key role in the expertise sharing model, an integrative approach that directly connects design decisions with individuals' characteristics and behaviors becomes necessary.

As Ren and Kraut (2014a) point out, conducting research to inform the design of online communities relies on synthesizing multiple relevant propositions and theories and clearly identifying how particular design choices influence the community outcomes designers intend to achieve. Although existing research has specifically examined expertise ranking, assessment, and prediction in Q&A communities, integration of community characteristics and user motivations is still necessary to understand the effects of the design decisions based on user expertise. Hence, the key research question in this dissertation is: *how might the implementation of different design*

decisions shape the sustainability of Q&A communities? To address this question, agent-based modeling will be applied to provide a systematic examination. In addition to the findings of the existing social Q&A literature, the study will also integrate a wide variety of theories from other areas in social science, including the Resource-based Theory (Butler, 2001), the Expectancy Theory (Vroom et al., 2005), and the Intrinsic/Extrinsic Motivation Theory (Ryan & Deci, 2000). These theories have been individually applied in online community research and explained why and how individuals participate in online communities and what sustains the system. The integration will increase the number of variables being examined simultaneously, thus offering a more comprehensive and realistic depiction of this complex social system compared to testing each individual theory in isolation. As a result, these social science theories will become more useful in guiding Q&A community design (Ren & Kraut, 2014a). Also, virtual experiments are conducted based on the simulation model to provide a systematic examination of how community design influence users' interactions to shape the community. Hence, the understanding of social Q&A communities goes beyond the utility of information. Meanwhile, the study also has practical implications for community designers and moderators. Guided by a flexible computer model, community moderators and designers are able to foresee the potential outcome of their decisions rather than solely rely on intuitions or trial and error, thus proactively engineering information systems that are more efficient and attractive to users with various expertise and needs.

Simulation results of the agent-based model, which is built upon empirical data from a large online Q&A community, reveal the relationship between two measures community designers and moderators can apply to manage individuals' expertise and indicators of community sustainability. Specifically, when the following two design decisions (1) displaying

individuals' expertise level in a community (expertise indication); and (2) recommending a question to answerers whose expertise matches the difficulty of the question, are implemented, both the membership size of a community and the proportion of solved questions are likely to be higher (question routing). In addition, implementing design decisions will also influence the member structure of a community. Question routing tends to prioritize experts' questions so that they are more likely to stay in a community. On the other hand, expertise indication tends to attract more beginners as their questions are more likely to be solved. These findings thus shed light on the ways in which these design decisions should be leveraged according to the development stage a community is in, which will be discussed in detail in the discussion section.

The rest of the dissertation is organized as follows. In Chapter 2, I will conduct a brief review of the existing literature on social Q&A communities and examine the key concepts and factors involved in this study. Next, I will describe the conceptual model of this study and raise research questions.

In Chapter 3, I will introduce the agent-based model that is used to simulate online Q&A communities, including how the method is applied in other online community research and how the model is calibrated and validated in this study by the data from an existing Q&A community.

In Chapter 4, I will present the simulation results to show how the design decisions influence a series of community outcomes, including the percentage of solved questions (i.e., success rate), membership size, and member attrition.

In Chapter 5, I will summarize the results and discuss the theoretical and managerial implications of the study.

In Chapter 6, I will provide some concluding remarks of the study.

Chapter 2

LITERATURE REVIEW

Social Q&A

Definitions. Broadly speaking, social Q&A can refer to various types of platforms and services. Shah et al. (2009) summarized three major types of social Q&A services, including digital reference services, expert services, and community Q&A. Digital reference services involve reference librarians searching for information and providing answers back to library users, who are the questioners. Similarly, expert services are offered by various subject experts, who belong to commercial or non-commercial organizations. Both digital reference and expert services usually take place in the form of one-to-one interactions between a questioner (a service user) and an answerer (a librarian or an expert). In terms of social Q&A, one of the key features is that it allows individuals to ask and respond to questions in the form of social interactions involving multiple participants, rather than using keywords to obtain a list of documents in search engines. As the social component is fundamental in Q&A communities, in this study, I will follow Shah et al.'s (2009) definition, who defines social Q&A as an online service allowing a user to express their information need in natural language and other users to respond to questions; meanwhile, a community is built upon such social interactions between questioners and answerers.

Essentially, social Q&A sites are public collaboration systems on the internet where information is shared and distributed among users. Because content is generated by users' voluntary participation, there is no guarantee of answer quality; instead, questioners rely on "the wisdom of crowds—ask a hundred people to answer a question or solve a problem, and the average answer will often be at least as good as the answer of the smartest member"

(Surowiecki, 2005, p.11). The social interactions go beyond question asking and answering: users can also comment on the questions and answers, evaluate the quality of the information by voting, and earn rewards and recognition through their contributions. These social features facilitate the problem-solving and collaboration processes on these platforms and shape them into reservoirs of collective knowledge and wisdom. A good social Q&A platform can not only satisfy users' information needs but also serve as a public knowledge base with lasting value.

Table 1

Examples of Online Q&A Communities

Website	Founded	Features
Knowledge iN	2002	The first Q&A website.
Google Answers	2002	Fee-based. Discontinued in 2006.
Yahoo! Answers	2005	One of the leading Q&A sites on the web.
Baidu Knows	2005	The largest Q&A community in China.
Reddit	2005	Users can create subreddits to ask questions on various topics.
Stack Overflow	2008	Focuses on programming and software development.
Quora	2009	Users can suggest edits to answers that have been submitted by others.
Zhihu	2011	A big Chinese Q&A community with more than 100 million users.

The first online Q&A website, Knowledge iN, was set up in 2002 by a Korean corporation, Naver. After that, many online Q&A communities have begun to emerge on the internet around the world, which greatly expands the amount of information available on the web. Table 1 shows several examples of popular online Q&A communities.

Current research. Recent literature has focused on different, discrete aspects of Q&A sites. Key among these aspects is the content and the quality of the information provided (Srba & Bielikova, 2016b), which includes question topics (Nie et al., 2014), question quality (Z. Liu & Jansen, 2013; Yao et al., 2015) and answer quality (Gkotsis et al., 2014; Harper et al., 2008; Toba et al., 2014). Another focal area in extant research is the classification and modeling of users' behaviors and expertise (Furtado et al., 2013; Pal, Chang, et al., 2012; Zhang et al., 2007), which sheds light on the structures and the dynamics of various types of users in Q&A communities.

On the other hand, the amount of research regarding the underlying knowledge sharing process and the longitudinal evolution of the social system that supports it is relatively small (Srba & Bielikova, 2016b). Some studies have applied social network analysis to understand the global communication patterns in Q&A communities and their growth (Adamic et al., 2008; Rechavi & Rafaeli, 2012; G. Wang et al., 2013), and a few studies have examined the knowledge sharing process at the thread level (Liang, 2017; G. A. Wang et al., 2014).

As discussed, one of the directions to extend the existing research is to conduct a systematic investigation of how information quality, individuals' characteristics and activities, and community structure are connected and how the interplay between these elements inform the system design that promotes its sustainability. Based on the extant findings, the current study will take a comprehensive approach to understand how some of the design decisions

implemented by community moderators will have an impact on individuals' interactions, and how higher-level community outcomes will emerge from these interactions. Theoretically, this study can deepen the understanding of dynamics in Q&A communities by offering a systematic examination and an integration of theories and existing findings, which serves as the base for better Q&A community design. Meanwhile, from a practical point of view, a more granular approach to modeling community sustainability would be of great value for designers and moderators, as it helps to identify pathways through which particular design choices influence community performance, hence providing straightforward guidance on allocating resources and improving both efficiency and effectiveness in a community.

Knowledge Sharing and Online Community Design

Knowledge sharing in organizations. Supporting knowledge sharing and collaboration is fundamental to sustaining online Q&A communities as successful knowledge transfer and accumulation not only benefits community members themselves but also contributes to the community's greater worth (Faraj et al., 2011). Wang and Noe (2010) reviewed studies of individual-level knowledge sharing from the knowledge management perspective and presented a framework with five focus areas: organizational context, interpersonal and team characteristics, cultural characteristics, individual characteristics, and motivation factors. These studies being reviewed have shown how different levels of environmental and individual elements influence knowledge sharing intentions and behaviors via motivational factors. Similarly, Ardichvili (2008) also discussed the motivators, barriers, and enablers of knowledge sharing in online communities, and argued that community designers and members are co-creators of a vibrant and productive online community, and thus designers should encourage participation and remove barriers based on individuals' characteristics and community context.

Unlike previous research, which is situated in organizational settings, Faraj et al. (2011) argued that one of the fundamental characteristics of general online communities is fluidity, given that participants are not known to each other and join in with diverse interests and background. As a result, knowledge collaboration is established upon the absence of existing social ties in these communities. In fact, previous studies mainly focused on individuals' or communities' motivations in terms of knowledge contribution. Nonetheless, the fluidity perspective calls for more emphasis on the flow and connection of ideas, and technology platforms play a critical role in supporting these flows of ideas, activities, and interactions happening in online communities.

Still, factors pertinent to the design of the technology platforms have not been adequately examined. Successful knowledge sharing depends heavily on the proper design of the platforms as it directly influences how individuals interact with each other and the results of their interactions. Due to the fluid nature of online communities, without a well-designed platform that can effectively manage the resources, information seekers may find it difficult to connect with persons with the right expertise and knowledge contributors may become less motivated if their contributions do not benefit others.

Online Community Design. Some scholars and practitioners have already provided insights and suggestions in terms of online community design (e.g., Kraut et al., 2011; Preece, 2000). Community design is implemented via numerous large and small decisions, which involve community structure and architecture, site navigation, information distribution, and interactions (Ren et al., 2007). Hence, in order to better inform community design via research, it often requires expanding and revising existing theories and combining multiple theories and findings. Iriberri and Leroy (2009) reviewed online communities research on design and success

factors and proposed a lifecycle framework to understand the evolution of online communities. Specifically, the life cycle of an online community has five stages, which are labeled as inception, creation, growth, maturity, and death. In each stage, they also identified different technology features that will be suitable for the stage to achieve success. They argued that an integrated and organized view of success factors can better facilitate community development.

Additionally, Kraut et al. (2011) further incorporated findings from social psychology, organizational behavior, economics, and other social science research to inform online community design. Social science theories can not only provide ideas to solve design problems, but also predict the consequences of various design decisions. However, they also pointed out that one challenge in this line of research is how to appropriately apply the findings and implications from empirical studies and social science theories to online communities with various contexts and characteristics. Particularly, studies examining online community design from the knowledge sharing perspective is relatively limited, with motivations to share and community engagement being the focus (e.g., Chung et al., 2016; Hall & Graham, 2004). As knowledge sharing and collaboration are some of the most common activities in many online communities, in addition to motivating individuals to participate, it is also important to understand how to properly manage the knowledge as resources to benefit the participants via implementing different design decisions so that successful community building can be achieved.

Sustainability of Online Communities

Community sustainability. Social structures are sustainable when the provided benefits outweigh the cost of participation (Levine & Moreland, 1994). Users will continue to engage in the social interactions with others when they think what they obtain from the interactions outweighs the time and effort they spend (Vroom et al., 2005). Hence, the sustainability of online

communities relies on users' continuous participation and consistent contributions (Butler, 2001; Wasko & Faraj, 2005). At the same time, users are able to continuously derive benefits from a community when the social system is sustainable over the long term (Butler, 2001). Particularly in social Q&A communities, members usually come with different interests, expertise, and activity levels, which bring different resources to the community (Furtado et al., 2013). In the meantime, members' behaviors are interdependent: an individual's decision to contribute or stay in the community can be influenced by others' behaviors and changing over time. Through the interactions between the community members, knowledge and information are created and exchanged; and these social interactions play a vital role in determining the quantity and quality of the created content. Thus, a sustainable system helps to support more knowledge sharing initiatives and develop stronger relationships and coordination among its members (Wasko & Faraj, 2005).

To understand how online groups and communities are sustained, Butler et al. (2007) discussed several key factors that are necessary to enable online group communication and sustain social interactions, as shown in Figure 1. First, it is critical for a community to have substantial financial support and investment, which cover a variety of expenditures, like software, hardware, and personnel. Online groups and communities may have various sources of funding, such as membership fees, advertising revenue, and internal funds. Only when operating costs are covered can online groups and communities maintain their infrastructure and support their development.

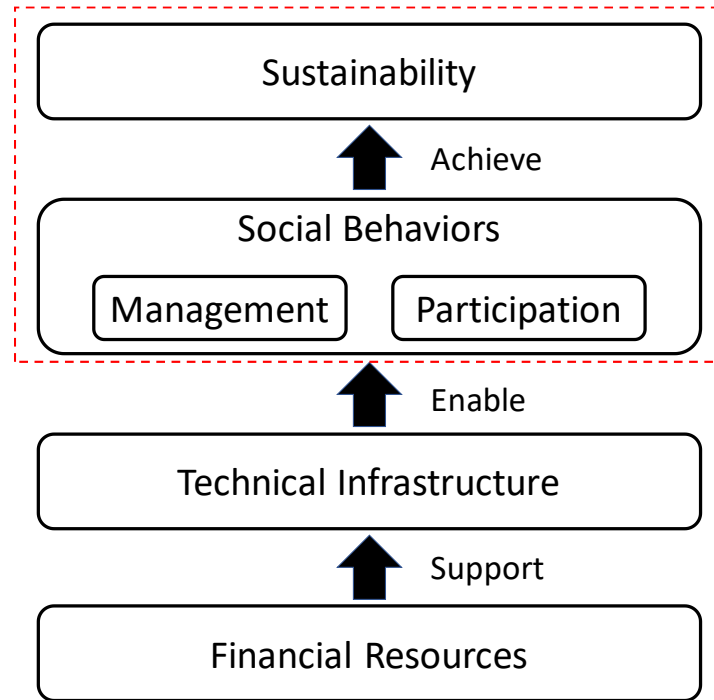


Figure 1 Key Elements of Online Community Sustainability.

Technical infrastructures make it possible for users to utilize various tools and mechanisms to fulfill their needs via social interactions. In terms of knowledge sharing and management, modern internet technologies offer efficient and effective tools for knowledge creation, inquiry, storage, and distribution. Furthermore, users can evaluate and improve existing knowledge by interacting with one another via these technologies. Without proper technical infrastructures, social interactions on the internet will become difficult or even nonexistent.

To ensure technical infrastructures function properly, infrastructure administration provided by technical specialists is necessary and critical. In addition to maintaining technical components and fixing issues, it is also important to improve existing infrastructure and develop new functionalities to satisfy the growing needs of users, groups, and organizations. An up-to-date infrastructure plays a vital role in attracting new users and retaining existing ones.

Technical infrastructures provide space and tools to make group and community communication possible, and these tools and infrastructures need to be constantly utilized and maintained by members and moderators to sustain viable online groups. Therefore, social behaviors are necessary to sustain online groups and communities over time. These social behaviors include two important dimensions: social management and active participation. Internally, rules and regulations—i.e., *social management*—must be implemented to control improper use of the infrastructures, such as letting newcomers understand community norms, discouraging the misuse of community resources, interpreting community rules and resolving disputes, and punish those who engage in abusing the system or other inappropriate behaviors. Furthermore, desirable behaviors should be publicly recognized and rewarded to encourage more appropriate and constructive use of the community. Oftentimes, these management measures, both technical and social, are implemented through community design, which includes the navigation architecture, interaction features, organization structures, regulation policies, and so on (Ren et al., 2007).

External promotion is another essential part of social management. Online communities are likely to collapse if there are no incoming new members while membership size is shrinking. It is thus critical to recruiting new members via promoting the community to the public so that more people can participate and bring in additional resources to the community. This can be done by interpersonal communication, like word-of-mouth, or through explicit promotion in other online spaces. Meanwhile, attracting new members also helps to encourage existing ones to interact with newcomers and contribute to the community.

Another fundamental component of social behaviors is *active participation* from community members. The communication process between members makes benefit provision

and acquisition possible. Participation in online communities is often in the form of creating and consuming content and relies on members' time, effort, and attention, which can be the most basic and important type of investment. Therefore, the goal of building solid technical infrastructures and conducting effective social management is to create a pleasant and reliable environment to engage more members in communication activities and ultimately sustain an online community.

Among all the key factors mentioned above, I will primarily focus on the *social management and behavior elements* and associated *technical features*, including community design and management, and users' participation under different design decisions. More specifically, in the following, I will discuss how the sustainability of Q&A communities is indicated by membership size and the outcome of communication activities, and how it is associated with user expertise and community structure.

Community membership size. In order to attract users to participate and get engaged in online communities, these communities need to maintain a sufficiently large size (Butler, 2001; Markus, 1987). In fact, community size is often considered as one of the key indicators of community sustainability (Arguello et al., 2006; Butler et al., 2014; Ma & Agarwal, 2007). When a social structure maintains a sufficiently large and stable size, it shows its members that social interactions are active and vibrant on the platform (Markus, 1987), and a large amount of resources is available (Butler, 2001); thus, the members are more likely to derive benefits from the system, which is critical for the system to demonstrate its value and remain attractive (Butler et al., 2007; Ren et al., 2012). Nevertheless, changes in community size are the result of many users' decisions based on their beliefs, preferences, and interactions with other members, which cannot be directly controlled (Butler et al., 2007). Therefore, community size is the outcome of

the interplay between individual choice and characteristics, communication activity, and technological features (Butler et al., 2014).

Particularly in online Q&A communities, Dev, Geigle, Hu, Zheng, and Sundaram (2018) examined the relationship between community size and long-term health and sustainability of the community. The study demonstrates the dependency between community sustainability and membership size: content generation depends on user generation and content types; however, measures of community health (e.g., the percentage of questions being answered) can decrease as the size grows. Such observation is in accordance with evidence found in other types of communities as individuals' willingness to contribute can be negatively impacted by an increase in community size (Jones et al., 2004; Kraut et al., 2012). One reason that Q&A communities can fail at scale is the increasing number of incoming negligent and undesired users generating low-quality content, which leads to a higher percentage of unsolved questions. Therefore, in order to fully understand community sustainability, in addition to membership size, we also need to take into account the user composition and the interaction network structure of a community.

User composition and network structure. In online communities, users can often be categorized in various roles. Specifically, with respect to Q&A communities, a large body of research has sought to identify common user roles based upon social networks and structural signatures. The most salient roles in these online knowledge sharing spaces include question people, answer people, and discussion people (Adamic et al., 2008; Fisher et al., 2006; Nam et al., 2009). Question people start new threads and ask questions, answer people respond to existing posts, and discussion people post questions and also actively participate in other members' discussions (i.e., they do both). These studies have painted a general picture of the

composition of users in Q&A communities and offered the basis for further understanding the social structure and dynamics of the system.

Even though individuals may occupy the same social role, they can still be different with respect to many attributes, such as demographics, socioeconomic status, values, beliefs, expertise, and experience (Jehn et al., 1999). In fact, members of Q&A communities come to participate and make contributions for various motivations, such as recognition and reputation (Tausczik & Pennebaker, 2012; Wei et al., 2015), altruism (Wasko & Faraj, 2005), self-identity and group bonding (Bateman et al., 2011; Ma & Agarwal, 2007), and financial incentives (Hsieh et al., 2010). Therefore, based on their motivations, members will bring different resources to the community, which are also often associated with their roles (Gleave et al., 2009).

Furthermore, such heterogeneity of members plays an important role in community sustainability (Markus, 1987; Oliver et al., 1985). The variance of resources and interests creates possibilities for members to obtain resources they are looking for from others; while at the same time giving them purposes by being able to benefit other communication partners through providing (distinct) resources they possess. Meanwhile, members can obtain access to a broader range of expertise and knowledge in a diverse community (J. Chen et al., 2010). Hence, social groups are more likely to sustain their long-term development when their members are characterized by heterogeneity (Markus, 1987; Oliver et al., 1985), and the subsequent dynamic balance of resources is key to the sustainability of the system (Welser et al., 2007). However, at the same time, it is important to note that too much diversity may increase disagreement and conflict within groups, thus reducing group cohesion and causing members to leave (J. Chen et al., 2010).

An additional model for understanding online communities is the core-periphery model, which exhibits a power-law pattern with respect to members' participation (Raban & Harper, 2007; Raban & Rafaeli, 2007). Active and knowledgeable members are often considered the core of an online community, who are a small portion of the total membership. Outside of the core are the peripheral members, who are usually less active and have lower expertise in the subject matter. Core members are connected to each other and peripheral members while peripheral members are only connected to the core. This type of network structure has been shown to be beneficial for the sustainability of online Q&A communities (Lu et al., 2014; Singh et al., 2011; Zhang et al., 2007). This structure can maintain a dynamic balance in terms of benefit seeking and provision, facilitate knowledge sharing, and is beneficial to the management of online communities (Bulgurcu et al., 2018). Yet despite these benefits, a core-periphery structure may create barriers for peripheral members to contribute (Lu et al., 2014).

Based on the discussion above, the model developed in this study will incorporate individuals' characteristics as well as the community structural patterns found in previous research, with an emphasis on information seeking and sharing behaviors.

Communication activities. In online communities, communication activities are the basis of the process by which resources are transformed into benefits, and they are the key element connecting members with community sustainability (Butler, 2001). Particularly in Q&A communities, question posting and answering are the core communication activities. Both questioners and answerers are obtaining benefits from the Q&A process, in the sense that questioners post questions so that answerers are able to post replies and gain recognition by showing their knowledge and expertise while answerers can provide solutions to satisfy questioners' information needs (Gleave et al., 2009). If no question is posted by information

seekers or questions fail to elicit replies from answerers, members are unlikely to receive any benefits from the community. Therefore, effective communication activities facilitate the knowledge sharing and problem-solving process in Q&A communities, as well as the development of stronger relationships and coordination among members (Wasko & Faraj, 2005). Consequently, the percentage of solved questions is often used as an indicator of the sustainability of Q&A communities (Dev et al., 2018; Srba & Bielikova, 2016a). When a community sees a relatively high proportion of unsolved questions, members are less likely to join and stay in the community as they fail to get their needs satisfied (Liang & Introne, 2019).

The outcome of the Q&A process has been one of the primary focus areas of existing studies. A large body of literature has sought to directly evaluate the quality of the answers (Harper et al., 2008; Tausczik & Pennebaker, 2011), or predict which answer is likely to be selected as the best answer (Adamic et al., 2008; Tian et al., 2013). Others have examined questioners' satisfaction with the answers received (Y. Liu et al., 2008). Moreover, some studies have sought to connect the quality of information with user types. Good quality answers are often provided by a small number of active users (Mamykina et al., 2011; Nam et al., 2009). Further, Furtado et al. (2013) provided a detailed investigation of the productivity of different types of users on Stack Exchange communities and showed how these users' behaviors change over time. Srba and Bielikova (2016b) also studied Stack Overflow on a longitudinal basis and associated the increasing rate of unanswered questions of the site with the increasing number of novice and churning users.

Still, a systematic investigation of the outcome of Q&A activities is lacking (e.g., Dev et al., 2018), which can limit the understanding of how to prevent failures and preserve the long-term sustainability of Q&A communities. From the system design perspective, in addition to the

features associated with Q&A discussions and participants, it is also important to consider what interventions can be employed to moderate the activities so that beneficial behaviors can be promoted and questions are more likely to be solved.

User expertise. In online Q&A communities, knowledge and information are no longer external artifacts; instead, they are situated in interpersonal communications. In other words, social interactions are an essential part of the system (Ackerman et al., 2003, 2013). Successful knowledge seeking and sharing relies on finding people with the right expertise and motivating them to contribute. Thus, user expertise plays a key role in community sustainability and it is thus crucial for community designers and moderators to better manage individuals' expertise to facilitate knowledge exchange by actively implementing different design decisions.

A large body of literature has examined user expertise in online Q&A communities, which predominantly focuses on expertise identification and assessment. User expertise can be measured on a global community level or on a specific topic level, which can also be referred to as user reputation and topic authority, respectively (Srba & Bielikova, 2016b). One common group of approaches to measure expertise is the graph-based method: individuals' interactions are transformed into a social network and various ranking algorithms (e.g., PageRank) are employed to assess each individual's expertise based on centrality measures (Aslay et al., 2013; Zhang et al., 2007). Another group of approaches is to predict expertise based on the track record of each individual, such as content, the number of posts, etc. (Movshovitz-Attias et al., 2013; Pal, Harper, et al., 2012).

Nonetheless, some important issues with respect to user expertise have not been well addressed by the extant literature. Srba and Bielikova (2016b) pointed out that current approaches fail to manage members' expertise properly, as they place too much emphasis on

questioners by directing most of the questions to the experts but tend to neglect the expectations of answerers, which can potentially result in an overuse of the capacity of highly knowledgeable members. Without proper moderation, community contributors are likely to leave, and the community will be inundated with low-quality content and become unsustainable. Therefore, a deeper exploration and examination of system design interventions to understand how to utilize members' expertise more efficiently is necessary.

Research Framework

Figure 2 presents the conceptual framework of this study. Individuals come to online Q&A communities with various interests, preferences, and expertise. From the resource-based perspective, community members can be seen as providers of different kinds of resources, including attention, social engagement, and information. Community members also have needs, which may be met via the resources that others provide (Butler, 2001; Butler et al., 2007). Meanwhile, individual's resources and needs are likely to be heterogeneous in online communities and such heterogeneity is a key factor in shaping collective goods as well as sustainability in social systems (Markus, 1987; Oliver et al., 1985). Furthermore, members interact with each other via posting questions and replies in Q&A communities. These communication activities enable these members to exchange resources; without effective interactions, a community may fail to sustain its development as members are not able to obtain benefits (Butler, 2001).

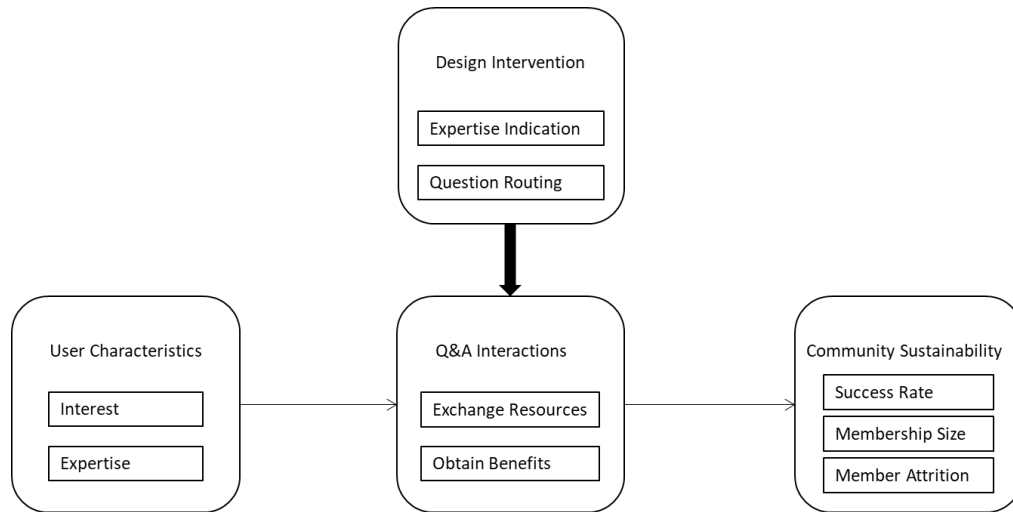


Figure 2 Conceptual Framework of the Study

One challenge for community designers and moderators is the management of the resources, especially with respect to an individual's expertise, as Q&A communities are a place where individuals seek and share knowledge and expertise. On the one hand, identifying and indicating an individual's expertise can benefit problem-solving processes as it is easier for questioners to find out information and solutions they need based on answerers' expertise. On the other hand, however, showing an individual's expertise requires proper procedures and mechanisms to assess an individual's knowledge and experience on the subject matter. Furthermore, it may discourage individuals with low expertise from contributing to the community since experts are more recognizable and tend to attract more attention.

Question routing is another design intervention community moderators can employ to potentially achieve more efficient problem-solving. This approach recommends questions to possible answerers who have interests in and are suitable for solving them (Guo et al., 2008; Srba & Bielikova, 2016a). Question routing consists of three basic components: (a) a question profile representing its topics/difficulty; (b) a user profile representing their expertise/interest; (c) a

mechanism matching the question profile with all relevant user profiles (Guo et al., 2008). Therefore, question routing is also closely associated with individuals' expertise. As an individual-level content moderation mechanism, question routing can be more effective in retaining members, especially when the message volume is large. Nonetheless, the individual convenience offered by this approach may come at the expense of overall community health due to a narrower scope of the community messages (Ren & Kraut, 2014b).

The current study will examine the trade-offs of the design decisions managing online Q&A discussions based on individuals' expertise, including expertise indication and questions routing. I aim to investigate the following questions by simulating virtual online communities using agent-based modeling: (1) *How do expertise indication and question routing influence the sustainability of Q&A communities;* (2) *What are the trade-offs between the design decisions regarding various community outcomes, namely success rate, membership size, and member attrition?*

Chapter 3

METHOD

In this study, I built an agent-based model drawing insights from existing findings as well as empirical data from an online Q&A community to examine the research questions. In the following sections, I will first introduce agent-based modeling and its application in online community research. I will also discuss the data source and how the model will be developed, calibrated, and validated based on the relevant findings and the data.

Agent-based Modeling

Computational simulations have become one of the most effective approaches to understanding social systems due to their capability in handling longitudinal and nonlinear social processes (Davis et al., 2007). Simulations are models that represent some of the characteristics of real-world processes, systems, events, and interactions via parameters calibrated upon some observations in real life (Lave & March, 1993; Law & Kelton, 2013). One of the advantages of simulations is allowing the manipulation of parameters to represent possible conditions, and thus virtual experiments can be conducted to systematically examine proposed research questions (K. M. Carley, 2001). Furthermore, simulations also help to yield propositions that can be used to inform both theory development and system design practices (Butler et al., 2014).

As a form of computational simulation, agent-based modeling “enables a researcher to create, analyze, and experiment with models composed of agents that interact within an environment” (Gilbert, 2008). Researchers can often use the agents to represent various physical and social entities, such as human beings, organizations, animals, and particles. In a simulated environment, agents will follow certain stipulated rules to perform a series of actions and interact

with each other, so as to imitate and examine a wide variety of physical and social phenomena such as human communication and particle movement.

Compared with other types of mathematical modeling, agent-based modeling can be employed in situations where system-level characteristics and structures are the results of individual-level agents' actions and interactions (Ren & Kraut, 2014a). Hence, it is suitable for examining relationships in complex social systems, such as online communities, which are often nonlinear, non-deterministic, and evolutionary. It not only helps to understand how system-level patterns emerge from individuals' interactions over time but can also demonstrate how the variations in a set of factors affect the development of the system in a rigorous manner. From a socio-technical perspective, the simulated model offers insights for the mechanisms behind individuals' social behaviors and thereby contributes to the development of theories. At the same time, it also informs the design and management of the system by focusing on the variables of interest and the potential outcomes these variables will lead to (Ren & Kraut, 2014b). Additionally, in terms of theory development, agent-based modeling is especially suitable for bottom-up theorizing and for exploring, elaborating and extending underdeveloped theories with modest empirical or analytical grounding (Davis et al., 2007; Klein & Kozlowski, 2000).

Several studies have been conducted to examine membership size and user commitment in online communities via agent-based modeling (e.g., Butler et al., 2014; Ren & Kraut, 2014b; Schweitzer & Garcia, 2010). Butler et al. (2014) extended the attraction-selection-attrition theory developed in traditional organizational settings (Schneider, 1987; Schneider et al., 1995) to online communities through introducing new technological features in their agent-based model. Specifically, the study theorizes how the time and effort required to engage with content and how

consistent the content topics are in an online community affect its sustainability, represented by its size and resilience.

Furthermore, Ren and Kraut (2014b) demonstrated the value of agent-based modeling in examining dynamics in complex social systems and generating insights for online community design. The authors examined the effects of different design choices (topical breadth, message volume, and discussion moderation) on member commitment and contribution. They argued that the design of complex social systems requires to consider a larger set of parameters than social science research does, and agent-based modeling can synthesize findings from multiple social theories so that they can be applied to inform the design of online communities.

Although scarce in existing literature, agent-based modeling is also applied in the setting of online Q&A communities. For example, Aumayr and Hayes (2014) show the dynamics in a Q&A community can be effectively modeled by a small set of agent attributes, including expertise, activity, and question-to-answer ratio. The study also finds out that the recency of content seems to be more important than the actual content itself in terms of accurately capturing the interactions and dynamics.

Additionally, scholars have been leveraging the method to examine knowledge sharing behaviors (Hall & Graham, 2004; Jiang et al., 2014; Jolly & Wakeland, 2009; Kane & Alavi, 2007; Nissen & Levitt, 2004; Xia et al., 2013). These studies demonstrate the benefits and advantages of agent-based simulations, particularly in the context of organizations. The focus of these studies has mainly been placed on simulating the flow of knowledge and employee behaviors and interactions within organizations. For example, Wang et al. (2009) used agent-based models to help decision-makers to understand how knowledge sharing results from the interaction between employee behaviors and organizational interventions so that managers can

better devise and review policies and interventions to support more effective knowledge sharing. As the extant research has been centered around the management of employees, it is equally important to apply this method to examine the design of information systems and technologies that can facilitate (the management of) the knowledge sharing process, given the fact that enterprise social media have been gaining popularity within organizations as a platform for enabling serendipitous, informal, and collaborative knowledge-sharing (Kane, 2017; Leonardi, 2014; Leonardi et al., 2013; Osch et al., 2015).

To summarize, these studies offered a useful basis for the examination of social Q&A communities via agent-based modeling while also underscoring the applicability of this method. Unlike the existing studies, which put more emphasis on the high-level community outcomes, I will focus specifically on problem-solving interactions between users and evaluate the direct outcome of these interactions, i.e., whether a problem is solved or not. In Q&A communities, users interact with each other through question posting and answering, and when the questioner thinks the question has been sufficiently answered, the status of the question, which is visible to all users, will be changed from unsolved to solved. These features can provide a more concrete context to the model development process compared to the previous studies building on generic discussion-based communities (Aumayr & Hayes, 2014).

Ren and Kraut (2014a) prescribe a seven-step roadmap to building agent-based models, which are followed in the current study, namely: (1) evaluate the appropriateness of agent-based modeling for the proposed research questions; (2) define boundary conditions and build a conceptual model; (3) translate the conceptual model into computational representations; (4) implement the model; (5) demonstrate the internal and external validity of the model; (6) experiment with the model; and (7) publish the model and results. Particularly, (1) and (2) have

been discussed in Chapter 2. (3), (4), and (5) correspond to the following Model Development, Model Calibration, and Model Validation sections, respectively. (6) and (7) will be conducted in Chapter 4 and Chapter 5.

Model Development

The agent-based model built in this study is essentially a discrete event simulation model, which advances in distinct “steps”. In each step, all agents are randomly activated to perform actions and system-level outcomes will emerge from these simultaneous actions of the autonomous agents. More specifically, the model consists of two parts: a platform and its users. A platform is a passive agent where users interact with each other. It accepts, disseminates, and moderates messages (including questions and answers) created by users. All community members interact with one another on a single platform. Users are modeled as active agents who can join a community, read and post messages, and leave a community. Over the process of simulation, the interactions between users and the platform and among users themselves may lead to different results of communication and community sustainability. The development of the model and the agents’ actions is similar to other studies employing agent-based modeling to investigate online community dynamics and interactions (Aumayr & Hayes, 2014; Butler et al., 2014; Ren & Kraut, 2014b).

Modeling a platform. In the model, a platform serves three main functions, including the design decisions: *member tracking*, *question routing*, and *expertise indication* (Table 2). An individual choosing to join a community will go through the member registration process, informing the platform to add the individual to the community member list, and thus the individual will be granted permission to submit messages to the platform. On the other hand, when an individual decides to leave the community, they will be marked as inactive on the

platform and removed from the model. Hence, the platform maintains a list of community membership based on individual member's decision to join or leave, i.e., *member tracking*.

Table 2

Platform Capabilities

Platform Capabilities	Explanation
Member Tracking	Maintaining a list of members based on individuals' decisions to join or leave a community.
Question Routing	When implemented, a question will be more likely to rank higher on a member's reading list if the question's difficulty is closer to the member's expertise; otherwise, all questions will be listed in reverse chronological order.
Expertise Indication	When implemented, each member will have a direct indicator of their expertise; otherwise, all members' expertise is implicit.

Question routing refers to a platform's capability to display the questions in a certain order to the members. When entering the platform, a member will be shown a list of questions submitted by the others. All questions concern a single subject matter (e.g., math, computer programming, etc.), while each question is modeled to have different levels of difficulty based on the questioner's expertise (represented by a value between 0 and 1 exclusively). If no routing is applied, the questions will be listed in reverse chronological order, with the newest post shown first. If question routing is implemented, the question list is tailored to each member, and

questions matching a member's expertise will be likely to have a higher priority on the list. In other words, if the absolute difference between a question's difficulty and a member's expertise is smaller, the question will be assigned a larger weight thus being more likely to rank higher on the question list.

Expertise indication is a platform's ability to identify and exhibit a member's level of expertise regarding their knowledge on the subject matter. When an individual joins a community, its platform will prompt the individual to report their experience with the subject matter. With expertise indication implemented, each member will have an indicator of their expertise (such as a badge next to their username) so that a questioner can directly tell how experienced an answerer is, which has an impact on how likely the question will be solved. Otherwise, all members will appear to be more homogeneous and their expertise will become implicit. When the expertise levels are visible, a questioner will decide whether the question is solved based on the number of answers from each level of members; otherwise, the likelihood is predicted by the total number of answers received in a question.

Modeling individuals. In this model, individuals are modeled as autonomous agents who possess different expertise, contribution likelihood, and preferences and their participation actions include joining and leaving the platform, as well as reading and posting messages (including questions and answers). Additionally, a member who posts a question can decide whether the question is solved or not based on the answers they receive. Following the expectancy theory (Vroom et al., 2005) and the resource-based theory (Butler, 2001), members take time and effort to participate in online communities and they also derive benefits from their participation; hence, their actions are motivated by evaluating the participation costs and benefits, and when benefits exceed costs, members will continue their engagement in a

community; otherwise, they will stop participating and leave the community. Community-level outcomes will then emerge from individual differences and their interactions over time.

Individuals are modeled by three innate characteristics, which remain stable over the course of a simulation, including *expertise*, *contribution likelihood*, and *contribution predilection* (Table 3). *Expertise* (E) is represented by a value between 0 and 1, which indicates a member's knowledge and experience with respect to the subject matter. An individual will further be categorized as one of the three levels of expertise based on the value: low, medium, and high. In this model, expertise is considered as the core characteristic of an individual. A high value of expertise suggests that an individual has extensive knowledge and experience of the subject matter so they are more likely to offer viable solutions to the questions, while low values indicate that an individual's knowledge and experience are limited. Additionally, expertise not only determines the difficulty of the question posted but also influences the individual's contribution likelihood and predilection.

Table 3

Individual Parameters

Individual Parameters	Explanation
Expertise (E)	The indicator of an individual's knowledge and experience of the subject matter, which ranges from 0 to 1 and has three levels (low/mid/high).
Contribution Likelihood (CL)	The likelihood of an individual submitting a message to the community platform, which ranges from 0 to 1.
Contribution Predilection (CP)	The probability of an individual's decision to submit a question to the community platform, which ranges from 0 to 1.
Participation Cost (C)	The time and effort needed to compose a message and submit it to the community platform.
Participation Benefit (B)	The value derived from a solved question an individual is involved via asking the question or offering an answer.

An Individual's decision to post messages in a community is modeled probabilistically as the *contribution likelihood* (CL). The likelihood is a value between 0 and 1, indicating how likely an individual will compose a message and submit it to the community platform. An individual with a low value is less likely to contribute messages and often considered as a "lurker"; while

an individual with a high value is more active on the platform to make contributions. Meanwhile, the likelihood also depends on an individual's expertise: experts are more likely to be active contributors than individuals with low expertise. Furthermore, during each simulation step, an individual may have multiple opportunities to submit messages and the contribution likelihood will decay as the number of submitted messages increases so that an individual is unlikely to post new messages if they have already contributed many during the same time period.

In addition, there are two types of contributions in a Q&A community: questions and answers. Therefore, the *contribution predilection (CP)* is used to model the probability of an individual's decision to submit a question or an answer to the community platform, which is represented by a value between 0 and 1. Specifically, when the value is close to 1, it indicates that an individual will be more likely to submit a question and a value close to 0 suggests that an individual will be more likely to post an answer. Again, the contribution predilection depends on an individual's expertise, and experts will post answers more often while individuals with low expertise will post more questions. Together, individuals' contribution behaviors are probabilistically determined by expertise, contribution likelihood, and predilection.

When participating in a community to seek information and share knowledge, members spend time and effort, i.e., the cost of participation, to obtain certain benefits (S. Wang & Noe, 2010; Wasko & Faraj, 2005) (Table 3). Specifically, the benefits include accessing information to satisfy one's needs and sharing information as a way to demonstrate competence and provide positive self-evaluation (Ren & Kraut, 2014b; Wasko & Faraj, 2005). A member's ongoing engagement with a community is affected by their evaluation of the costs and the benefits of participation. The previous attributes (i.e., expertise, contribution likelihood, and predilection) determine how individuals contribute to the platform, whereas participation costs and benefits

are associated with the outcome of participation. Following Butler et al. (2014), for each submitted message, the *participation cost* (C) is a fixed value, which is between 0 and 1. On the other hand, an individual is more likely to derive benefits from their participation when the question they involve in (either by asking the question or offering an answer) gets solved. Hence, for each solved question, the *participation benefit* (B) an individual will receive is calculated as the cost plus a function of the contribution likelihood ($CL - CL^2$). In other words, when an individual has a relatively high or low interest to contribute, the net benefit derived from a solved question ($B - C$) is smaller. In contrast, an individual with a moderate level of interest tends to receive larger net benefits (Butler et al., 2014). During each simulation step, each individual agent will update the net benefit they derive from their participation ($B \times N_1 - C \times N_2$, where N_1 is the number of solved questions and N_2 is the number of questions the individual is involved in). Once an individual's assessment of the net benefit falls below zero, they will stop participating in the community and thus be removed from the platform.

Community Outcomes. This study will focus on several community-level outcomes that emerge from the community design decisions and the interactions of community members. The first outcome is *success rate*, or the percentage of solved questions in a certain time period. The second one is *community size*, which is the number of members who remain and continue to participate in the community. In addition, *membership attrition* measures the number of individuals who leave the community at the end of a particular time period. All these outcomes indicate how sustainable a Q&A community is regarding helping its members to find solutions to their questions as well as maintaining and expanding its size.

Model Calibration

The next step is to connect the model parameters with the features and characteristics observed in real data, which is called model calibration (Bratley et al., 1987; A. Chen & Edgington, 2005). The calibration step ensures that a model can produce results that match real-world phenomena within reasonable accuracy by tuning a model's rules and parameters (K. Carley, 1996; Ren & Kraut, 2014a).

In this study, I obtained data from a large online Q&A forum `/r/excel`, which is a sub-community (subreddit) launched in 2009 on Reddit.com, and features questions and answers concerning Microsoft Excel and VBA (Visual Basic for Applications) programming. On this platform, users can ask questions by starting new posts and later replies are organized as grouped messages, known as discussion threads. Once the problems are solved, they should change the status of the questions to 'solved', as requested by the community rules. In addition, users whose answers are accepted as solutions will be awarded virtual points, and all members can upvote questions and answers if they think the content is useful.

The `/r/excel` forum provides an excellent opportunity for the research questions because the Reddit Q&A forum is a stable and successful community with more than 110,000 subscribers. Given the large and diversified user base and active interactions among users, this community is thus suitable for understanding how the heterogeneity and the balance of resource exchanging between users are related to the development of the community. Meanwhile, this community is actively managed by a group of moderators, with both community and user-level routing, which facilitates the investigation of the effects of routing. This community represents a specific type of social Q&A platform where users mainly look for instrumental and factual

information as opposed to emotional support; hence, the analysis will specifically focus on knowledge seeking and offering behaviors.

The dataset contains a trace of 29-month of activities in the community, starting from January 1, 2015, resulting in a dataset containing 32,733 questions and 193,769 replies in total. The data collection period is chosen to be long enough to observe adequate and meaningful changes. To perform model calibration, 80% of the questions (26,186) are randomly selected.

When an individual becomes a new member of a community, their expertise needs to be specified. As individuals with extensive knowledge are more likely to provide viable solutions to questions they see in the community, to calibrate expertise, I examined the number of questions solved by each individual. The percentiles are shown in Table 4, starting from the 90th to the 100th, suggesting that the majority of the members did not solve any questions during the data collection period (about 90%) while less than two percent of the members solved more than five questions. Based on the results, an individual's expertise will fall into one of the three levels, including low, medium, and high, with different probabilities. More specifically, the expertise is modeled as a random number with a value ranging from 0 to 1. The value will fall between 0 and 0.33 with a probability of 90%, indicating a low level of expertise (corresponding to individuals who never solved a question). It will fall between 0.33 and 0.66 with a probability of 8%, indicating a medium level of expertise (corresponding to individuals who solved one to five questions). The expertise will take a value between 0.66 and 1 with a probability of 2%, representing a high level of expertise (corresponding to individuals who solved more than five questions).

Table 4

The Percentiles of the Number of Solved Questions by an Individual (from 90th to 100th)

Percentile	Number of Solved Questions by an Individual
90%	0
91% - 95%	1
96%	2
97%	3
98%	5
99%	13
100%	1290

In addition, the number of individuals becoming new members at the beginning of each simulation step is calibrated by the number of new members per day in the calibration dataset. I used a series of goodness-of-fit plots (density plot, cumulative distribution function plot, probability-probability plot, and quantile-quantile plot) to determine which probability distribution best fits the empirical that. Results suggest that a normal distribution with $\mu = 24.38$ and $\sigma = 10.35$ best fits the empirical data. Thus, the number of new individual agents entering the

model at the beginning of each simulation step will be randomly drawn from this normal distribution.

For each individual, the number of messages contributed per day over the observation period is used to determine the contribution likelihood, which represents the probability of an individual submitting messages to the community platform. The contribution likelihood is examined separately for each expertise level, and probability plots are applied to decide which distribution fits the data best. For the low expertise level, the empirical data follows a log-normal distribution with $\mu = -4.59$ and $\sigma = 1.1$; for the medium level, the data follows a log-normal distribution with $\mu = -4.17$ and $\sigma = 1.17$; for the high level, the data follows a log-normal distribution with $\mu = -1.98$ and $\sigma = 1.12$.

The contribution predilection is calibrated by the proportion of questions posted by each expertise level of individuals. Of all questions posted during the data collection period, 90% come from low expertise individuals, 7% come from the medium level and 3% come from individuals with high expertise. Therefore, if a low expertise individual decides to submit a new message (either a question or an answer), the message will be a question with a probability of 90%. An individual with medium expertise will post a question with a probability of 7% and a high expertise individual will post a question with a probability of 3%.

As mentioned above, an individual may submit more than one message during one simulation step. The model assumes that as the number of contributed messages increase, the probability of submitting additional messages will decline. However, this cannot be directly observed from the calibration data, therefore the values will be selected based on prior research or inferred by other observable data (Law & Kelton, 2013). For each additional answer contributed by an individual, the probability will reduce by five percent. On the other hand,

given that the majority (70%) of the members submit less than two questions over the observation period, the decay of the likelihood to submit additional questions is larger than the one for answers. Specifically, for each additional question submitted, the probability will reduce by approximately 63% (i.e., $1 - 1/e$, where e is the base of the natural logarithm).

Since the participation cost is not directly observable from the calibration dataset and is not the primary focus of the study, following Butler et al. (2014), the cost for submitting a question is set as 0.3 while the cost for an answer is 0.075 (given that on average the length of a question is four times longer than the one of an answer). Similarly, the benefit of posting an answer will be a quarter of the one for submitting a question.

There are two design decisions examined in this study: question routing and expertise indication. If question routing is implemented, the community platform will customize a list of questions for each individual to read based on their expertise. First, the platform will calculate the absolute difference between the difficulty of the submitted questions and each individual's expertise, which is a value between 0 and 1. Based on the difference, the platform will assign a weight to each question, which is one minus the difference, and thus the larger the difference, the lower the weight. Next, the platform will order the questions using the weights as probabilities, so the larger the weight, the more likely the question will rank higher on the list. If question routing is not implemented, the questions will be ordered in reverse chronological order so newer questions will rank higher on the list and all individuals will see the same list.

Additionally, when reading through the list of questions, each individual will also evaluate the difficulty of the questions by themselves and if the individual thinks the difference between the difficulty of a question and their expertise is relatively large (greater than 0.3), the likelihood to answer the question is reduced by 20%.

When expertise indication is indicated, each individual's expertise level will be displayed on the community platform. As questioners are able to see the expertise level of the answerers, their decisions on whether the questions are solved are based on the number of answers from members with each expertise level as well as the expertise level of the questioners themselves. Using the calibration data, the process is modeled by logistic regression, shown as follows. Specifically, the log-odds of a question being solved are estimated by plugging in the numbers of answerers to the equations. Then, the probability can be calculated based on the log odds.

$$\text{logit}(P(\text{a question from a low expertise member being solved}))$$

$$\begin{aligned} &= .72 + .11 \times \# \text{ of answers from high expertise members} \\ &\quad - .06 \times \# \text{ of answers from medium expertise members} \\ &\quad - .14 \times \# \text{ of answers from low expertise members} \end{aligned}$$

$$\text{logit}(P(\text{a question from a medium expertise member being solved}))$$

$$\begin{aligned} &= .64 + .13 \times \# \text{ of answers from high expertise members} \\ &\quad - .13 \times \# \text{ of answers from medium expertise members} \\ &\quad - .14 \times \# \text{ of answers from low expertise members} \end{aligned}$$

$$\text{logit}(P(\text{a question from a high expertise member being solved}))$$

$$\begin{aligned} &= .84 + .13 \times \# \text{ of answers from high expertise members} \\ &\quad - .01 \times \# \text{ of answers from medium expertise members} \\ &\quad - .42 \times \# \text{ of answers from low expertise members} \end{aligned}$$

If expertise indication is not implemented, questioners will decide whether their questions are solved based on the total number of answers received, which is also modeled by logistic regression.

$$\text{logit}(P(a \text{ question being solved})) = .16 + .07 \times \# \text{ of answers}$$

To summarize, the agent-based model developed in this study consists of a community platform and a population of individuals. The community platform tracks individuals' member status and activities, and two design decisions can be implemented: question routing and expertise indication. These decisions will determine how the individuals read and answer questions and affect how the questions will be solved. The population of members is modeled by specifying the number of new members entering the community platform as well as distribution parameters for expertise, contribution likelihood, and predilection. The distribution of expertise indicates how much knowledge and experience the individuals have on the subject matter. Based on the expertise, the distributions of contribution likelihood and predilection describe how individuals vary in terms of their message contribution tendency and type. Individuals can perform a set of actions on the platform based on some specified rules, including reading and posting messages, as well as changing the status of their questions. Meanwhile, there are costs and benefits associated with the actions, and individuals will calculate the net benefit to decide if they will continue to participate in the community.

Figure 3 depicts a flowchart of an agent's decisions in a simulation step. At the beginning of each step, the number of new members is drawn from the calibrated normal distribution. Each of the new and existing members will perform their actions following the flowchart and the outcome of their actions is the result of the calibrated parameters and specified rules, which are generally represented as probabilities. Each of the design decisions and agents' actions will be implemented via individual functions in the computer program.

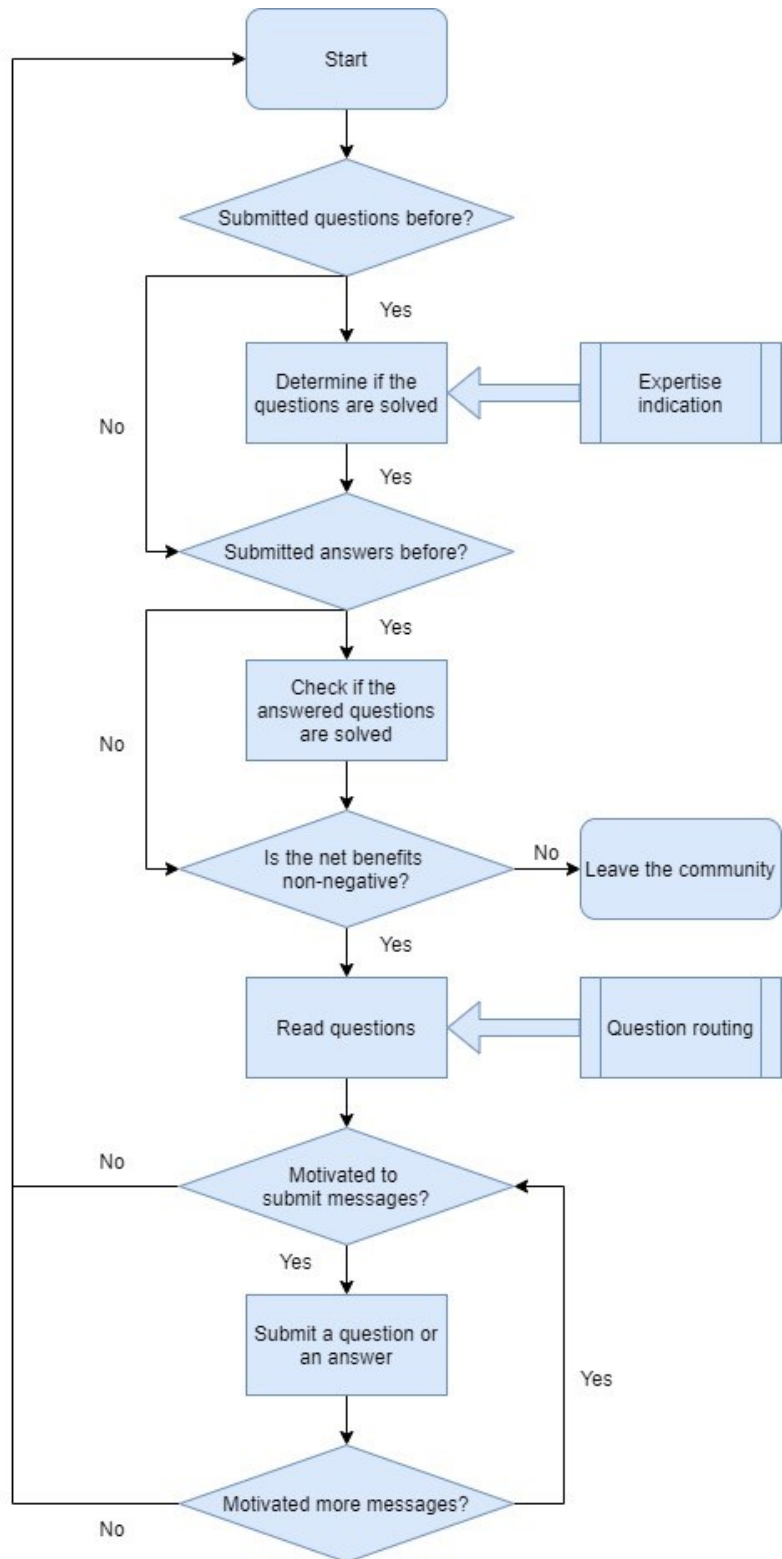


Figure 3 Flowchart of an Agent's Decisions in a Simulation Step

Model Validation

Based on the calibrated parameters, model validation is conducted to compare model predictions with the corresponding data in a holdout sample so as to assess how well the two match (Ren & Kraut, 2014a). Both calibration and validation steps are necessary in building an agent-based model to ensure external validity so that the outcomes generated by the model reflect the phenomena in the real world (Taber & Timpone, 1996). Hence, the model can become a reasonable basis for the development of insights and propositions regarding the design of online Q&A communities.

The agent-based model is developed using Mesa (Masad & Kazil, 2015), which is an agent-based modeling framework in Python, an interpreted, high-level, general-purpose programming language. Mesa is an open-source, Apache 2.0 licensed Python package that allows the easy implementation of the agent-based model with built-in components (e.g., spatial grids and agent schedulers). In fact, it is a Python counterpart to other popular multi-agent programmable modeling environments, such as NetLogo, Repast, or MASON. Additionally, Mesa is customizable based on the needs of users via programming. It also provides a browser-based interface to visualize the simulation process and the results. Further, another advantage of using Mesa is that users can conduct virtual experiments and other statistical analyses by combining the results produced by Mesa with other powerful data analysis tools in Python. Given these features, Mesa is an ideal tool for the study to examine the implications of the design decisions of Q&A communities. The computer program implementing the model in this study is attached in the Appendix.

The validation sample in this study contains 20% of the data which is not used for the calibration process. To perform the validation, the calibrated parameters were implemented in

the model, which is built in Mesa, to simulate 20 online communities. The percentage of solved questions (i.e., success rate) and the average number of answers in each question were recorded in each simulation. One-sample t-tests were then conducted to determine if the simulated values are statistically different from the empirical data in the validation dataset (percentage of solved questions = 0.64, average number of answers = 1.9). Results suggest that both the percentage of solved questions ($M = 0.66$, $SD = 0.05$, $t(19) = 1.79$, $p = 0.09$) and the average number of answers ($M = 1.78$, $SD = 0.27$, $t(19) = -1.99$, $p = 0.06$) in the simulation data are not statistically different from the empirical ones. Therefore, the validation shows that the model can reasonably approximate online Q&A communities found in the real world regarding the size and the outcome of the Q&A interactions.

Chapter 4

RESULTS

A 2 (with/without expertise indication implementation) \times 2 (with/without question routing implementation) factorial design virtual experiment is performed to examine the impact of the design decisions on three visible community outcomes: success rate, membership size, and membership attrition. In each combination, I ran 30 iterations with 30 seed members and 400 steps. The seed members are essentially the same as the other members later entering the community and can be seen as the early members of the community. The steps are operationalized as days in real life and one step represents one day; moreover, the first 35 steps are used for model initialization and hence are not included in the data analysis. In total, there are 120 simulated communities ($N = 30 \times 2 \times 2$), with 365 simulated steps each.

Success Rate

A two-way analysis of variance (ANOVA) was conducted to examine the effects of expertise indication and question routing on the success rate, which is the proportion of solved questions. Results show that the main effects are both significant: for expertise indication, $F(2, 117) = 285.38, p < 0.001$; and for question routing, $F(2, 117) = 101.28, p < 0.001$. Further, the interaction between the two main effects is examined via linear regression. The overall model is significant ($F(3, 116) = 140.56, p < 0.001$), and 78% of the variance can be accounted for by the model. As presented in Table 5, on average, the success rate is 11% higher when question routing is implemented and the other conditions remain the same; similarly, on average, the success rate is 17% higher when expertise indication is implemented with all other conditions being equal. Additionally, the interaction effect is negative and significant. As Figure 4 indicates, when question routing is not applied, i.e., questions are displayed in reverse chronological order,

the implementation of expertise indication results in a greater increase in the predicted success rate.

Table 5

Interaction Between Expertise Indication and Question Routing

	<i>Dependent variable:</i>		
	Success rate	Membership size	Membership attrition
	(1)	(2)	(3)
Question routing implemented	0.11*** (0.01)	698.03*** (79.82)	-2.05*** (0.21)
Expertise indication implemented	0.17*** (0.01)	1,125.33*** (79.82)	-3.22*** (0.21)
Interaction	-0.05*** (0.02)	-211.67 (112.88)	0.95*** (0.30)
Intercept	0.51*** (0.01)	6,160.27*** (56.44)	9.68*** (0.15)
Observations	120	120	120
Adjusted R ²	0.78	0.79	0.79
F Statistic (df = 3; 116)	140.56***	146.64***	149.20***

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

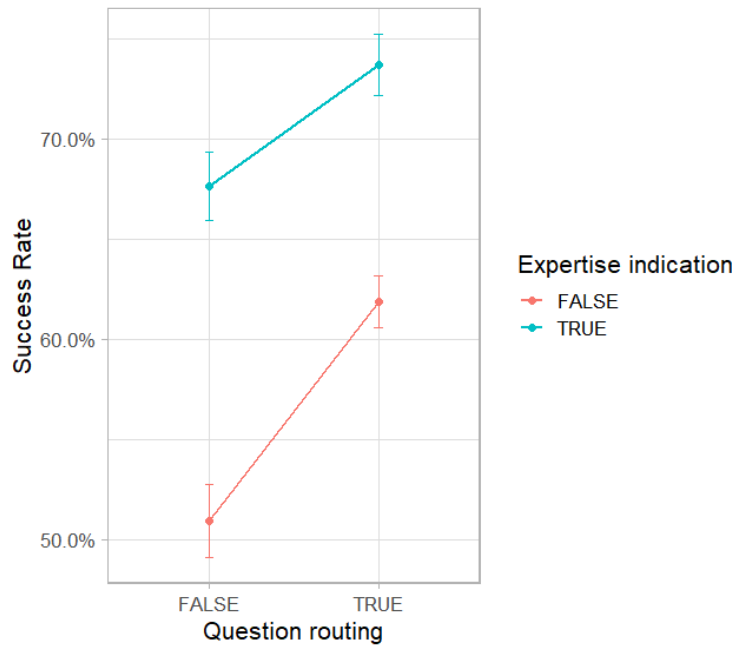


Figure 4 Interaction Effect on Success Rate Between Expertise Indication and Question Routing

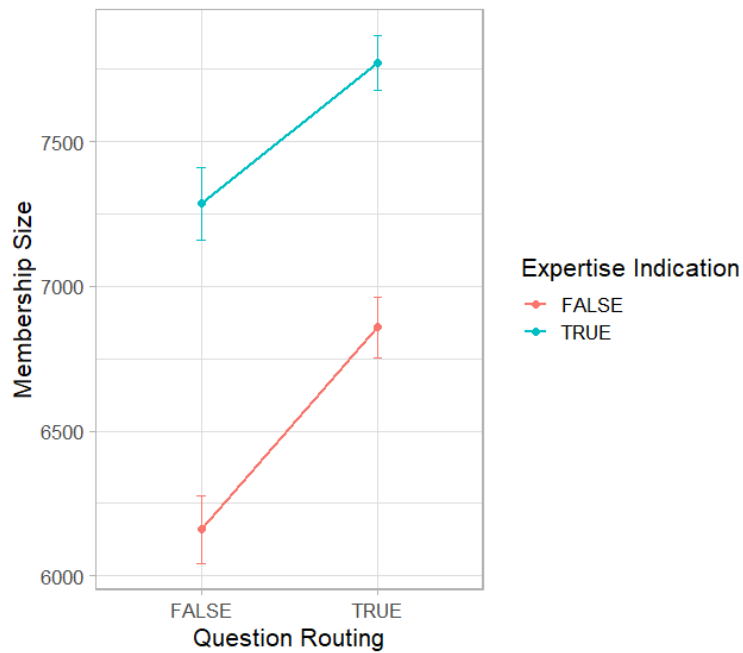


Figure 5 Interaction Effect on Membership Size Between Expertise Indication and Question Routing

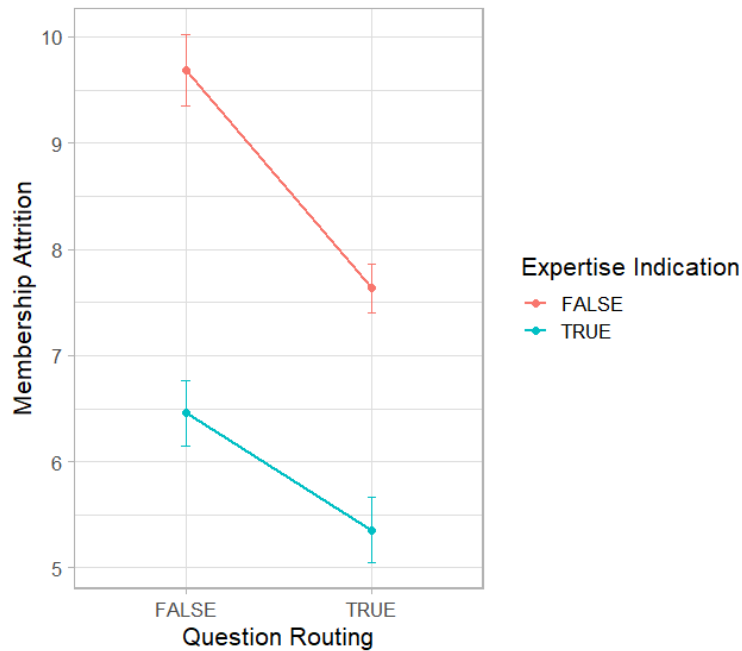


Figure 6 Interaction Effect on Membership Attrition Between Expertise Indication and Question Routing

Membership Size

Results from the two-way ANOVA show that both expertise indication and question routing have significant main effects on the membership size. Specifically, for expertise indication, $F(2, 117) = 319.43, p < 0.001$; and for question routing, $F(2, 117) = 107.78, p < 0.001$. Linear regression was also conducted, which is significant ($F(3, 116) = 146.64, p < 0.001$), and 79% of the variance can be explained by the model. As shown in Table 5, a community by to have 698 more members when question routing is implemented while the other conditions remain the same; similarly, a community will have 1,125 more members on average when expertise indication is applied with the other conditions being equal. However, no significant interaction effect has been found (Figure 5).

Average Membership Attrition

Again, the main effects of expertise indication and question routing on the average membership attrition are significant based on the two-way ANOVA. Regarding expertise indication, $F(2, 117) = 306.25, p < 0.001$; and for question routing, $F(2, 117) = 100.76, p < 0.001$. The results of the linear regression are presented in Table 5. The overall model is significant ($F(3, 116) = 149.20, p < 0.001$) and 79% of the variance can be accounted for by the model. The results also indicate that a community is likely to lose about two more members per step (day) if question routing is not implemented, or about three more members per step if expertise indication is not implemented, with the other conditions remaining the same. Meanwhile, the interaction effect is positive and significant suggesting that when question routing is not implemented, the implementation of expertise indication results in a larger decrease in the average membership attrition (Figure 6).

Percentages of Members by Expertise Level

Overall, the adjusted R^2 values of the ANOVA analyses suggest high effect sizes of the models. Besides significance testing, I also examined the changes in the percentages of members by expertise level to show how the simulated communities evolve under different design decisions. Based on the calibration results, the probability of a new member who has a low, medium, or high level of expertise is 90%, 8%, and 2%, respectively. When only expertise indication is implemented, the proportion of low expertise members keeps increasing over time (Figure 7). On the other hand, for the other three conditions, the proportion decreases at the beginning and starts to increase after a certain point. Moreover, the proportion of low expertise members at the end of the simulation is higher than the calibrated percentage (90%) when neither of the design decisions is implemented or only expertise indication is implemented. When only

question routing is employed, the proportion is about the same as the calibrated percentage. If both decisions are employed, the proportion drops below the calibrated percentage.

In terms of the members with a medium level of expertise, for all four conditions, the proportion increases at the beginning and starts to decline after a certain point, as shown in Figure 8. Particularly, when only expertise indication is implemented, the proportion at the end of the simulation is lower than the calibrated value (8%) while the other three remain similar to the calibration results.

When neither decision or only expertise indication is implemented, the proportion of high expertise members decreases over time and becomes lower than the calibrated value (2%) at the end of the simulation. The proportion remains relatively stable for the other two conditions and the proportion is highest when both decisions are implemented, which is depicted in Figure 9.

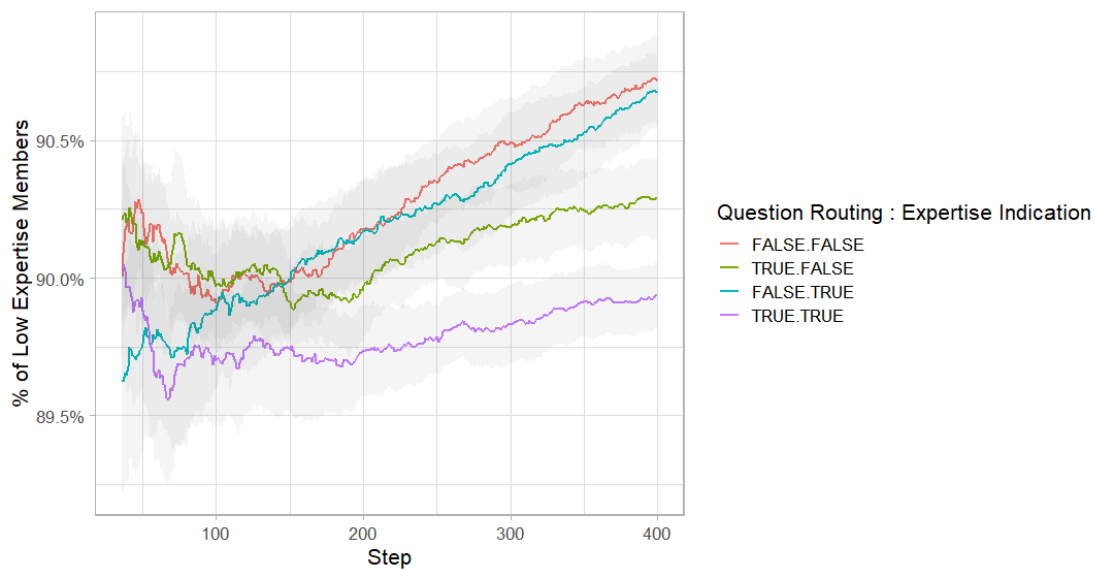


Figure 7 Proportion of Low Expertise Members over the Simulation Period Under Each Condition

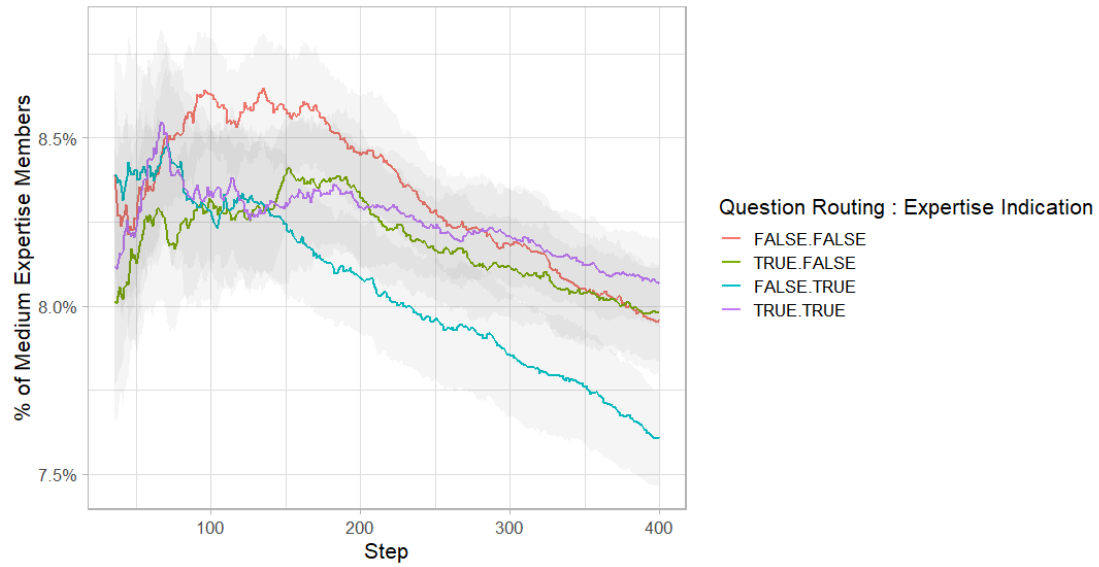


Figure 8 Proportion of Medium Expertise Members over the Simulation Period Under Each Condition

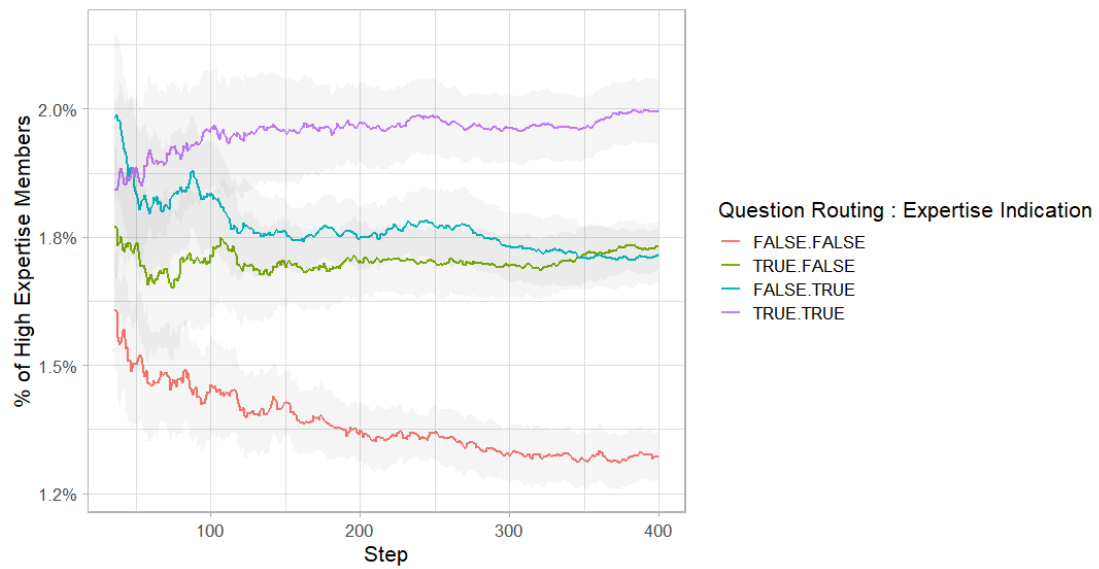


Figure 9 Proportion of High Expertise Members over the Simulation Period Under Each Condition

Percentage of Solved Questions by Expertise Level

Across all three expertise levels, the percentage of solved questions, i.e., the success rate, is the lowest when neither design decision is implemented and highest when both are applied on the platform. Among questions submitted by low expertise members (Figure 10), compared to the condition where only question routing is employed, the success rate tends to be higher when only expertise indication is implemented. On the contrary, among medium (Figure 11) and high expertise members' questions (Figure 12), the implementation of question routing is more likely to produce a higher success rate than applying expertise indication.

Additionally, the interactions between the expertise level and the design decisions are examined via linear regression. As Table 6 indicates, the interactions between the expertise level and question routing are significant while the interactions between the expertise level and expertise indication are not. Specifically, when question routing is implemented, on average, the success rate of low expertise members' questions is 15% lower than the ones from high expertise members; similarly, the success rate of mid expertise members' questions is 4% lower, with the other conditions being identical. Therefore, high expertise members' questions are more likely to be solved when question routing is employed.

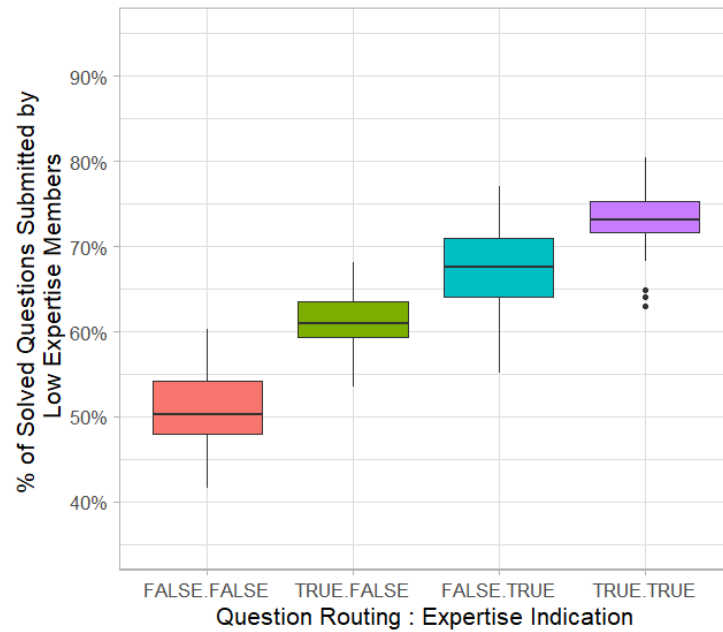


Figure 10 Percentage of Solved Questions Submitted by Low Expertise Members Under Each Condition

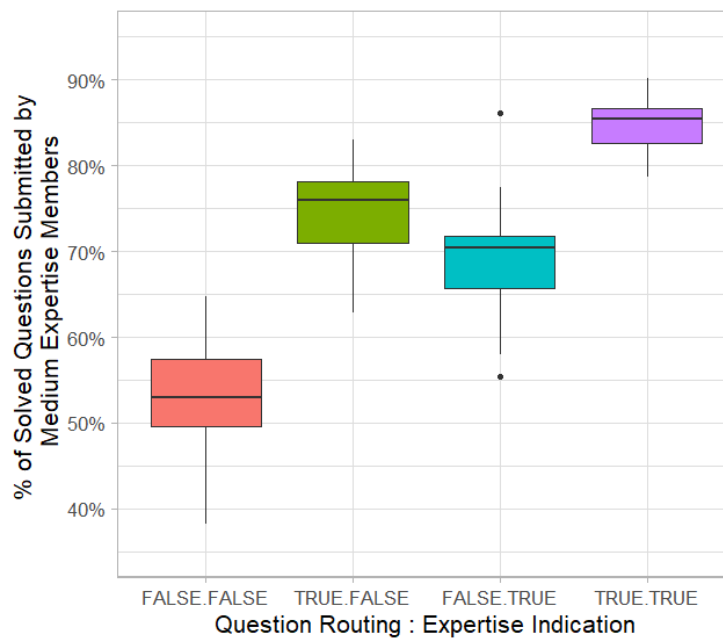


Figure 11 Percentage of Solved Questions Submitted by Medium Expertise Members Under Each Condition

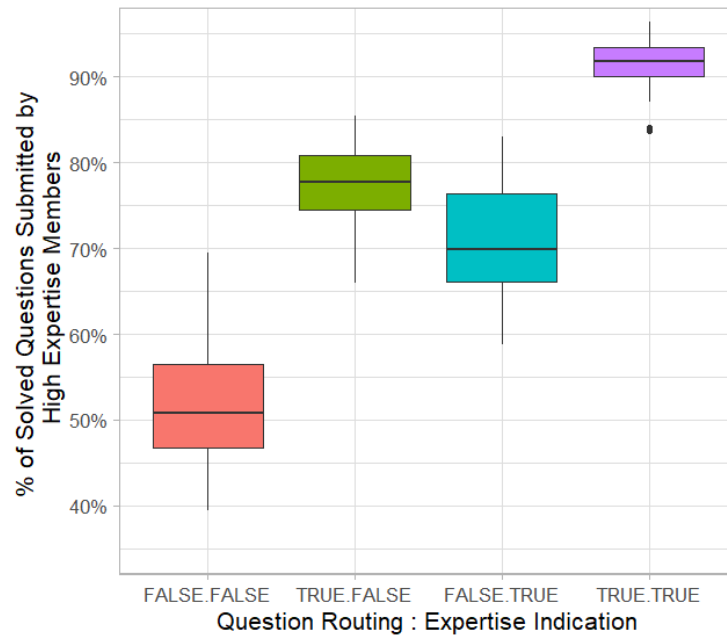


Figure 12 Percentage of Solved Questions Submitted by High Expertise Members Under Each Condition

Table 6

Interaction Between Expertise Level and Design Decisions

	<i>Dependent variable:</i>
	Success Rate
Low expertise	-0.01 (0.01)
Mid expertise	0.01 (0.01)
Question routing implemented	0.19*** (0.01)
Expertise indication implemented	0.25*** (0.01)
Low expertise : Expertise indication implemented	-0.03 (0.02)
Mid expertise : Expertise indication implemented	-0.03 (0.02)
Low expertise : Question routing implemented	-0.15*** (0.02)
Mid expertise : Question routing implemented	-0.04** (0.02)

Table 6 (cont'd)

	<i>Dependent variable:</i>
	Success Rate
Expertise indication implemented : Question routing implemented	-0.05*** (0.02)
Low expertise: Expertise indication implemented : Question routing implemented	0.002 (0.03)
Mid expertise : Expertise indication implemented : Question routing implemented	-0.01 (0.03)
Intercept	0.52*** (0.01)
Observations	360
Adjusted R ²	0.85
F Statistic (df = 11; 348)	186.01***

*Note:** $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Chapter 5

DISCUSSION

Summary of Findings

The current study investigated how two design decisions, question routing and expertise indication, influence the sustainability of social Q&A communities through agent-based modeling, a computational simulation method to examine how system-level outcomes emerge from individual-level interactions. The results indicate that implementing the two design decisions helps a Q&A community to increase its membership size as well as success rate, and reduce membership attrition. Question routing enables the community to recommend questions to suitable answerers in terms of expertise, which can reduce the number of unsolved questions. In contrast, if question routing is applied, other high expertise members are more likely to come across and answer difficult questions, especially when the message volume is large in the community, which helps the community attract and retain more users. In the meantime, question routing also encourages members to answer questions, as the model assumes that individuals are prone to offer answers when the difficulty of a question is within a certain range of their expertise.

Expertise indication makes members' expertise level visible in the community so it is easier for questioners to identify information that is more useful and reliable based on the answerers' expertise, thus facilitating the problem-solving process. In other words, members are more likely to gain benefits from the questions they asked, and thus continue to participate in the community.

Furthermore, these design decisions have different effects on members with different expertise levels. When question routing is implemented, the questions submitted by low

expertise members are less likely to be solved than the ones from medium and high expertise members. This is due to the fact that with question routing, low expertise members' questions tend to rank lower on medium and high expertise members' reading lists. The difference between the success rates is even larger when both decisions are implemented, where the success rate of high expertise members' questions is the highest, followed by medium and low expertise members' questions. Therefore, members with higher expertise tend to benefit more with the implementation of question routing. Without question routing, these members, who tend to ask difficult questions, may find it hard to obtain satisfactory answers within a relatively short amount of time, thus becoming less engaged in or even leaving the community.

Additionally, if both decisions are implemented, the distribution of members with different levels of expertise in the simulated data is close to the one in the calibration data. Nonetheless, the proportion of high expertise members is lower than the calibrated value if only one or neither of the decisions is implemented, suggesting that these members are more likely to leave the platform as the participation costs exceed the benefits they obtain. On the other hand, under the same conditions, the percentage of low expertise members is higher than the calibrated value, especially when question routing is not applied. In fact, since low expertise members are the majority of the platform, they contribute most of the questions, which are more likely to be read and answered by the others. However, when both decisions are implemented, the platform tends to prioritize questions submitted by higher expertise members, thus reducing the benefits low expertise members can obtain from the platform. As a result, these members are less likely to stay involved in the community.

Theoretical Implications

Through building an agent-based model, the current study incorporates some of the findings from existing literature on online Q&A community research and provides a detailed examination of how the sustainability of the system is shaped by individual activities and interactions, depending on the design of the system. More specifically, the study contributes to the online Q&A community literature by showing how individuals' behaviors and community outcomes are influenced by different design decisions over time and the possible trade-offs of these decisions.

The study focuses on two design decisions, question routing and expertise indication, both of which are based on individuals' knowledge and experience of a certain subject matter. As one of the most important resources in Q&A communities, individuals' expertise is critical in facilitating the problem-solving process and sustaining the development of a community. However, for information seekers, finding suitable answerers who are likely to provide satisfactory solutions in a reasonable time can often be challenging, especially in a large community. These measures pertain to the recognition and the allocation of individuals' expertise and can be leveraged to improve the efficiency of the system and contribute to better management of individuals' expertise in the community. The results suggest that a Q&A community can achieve better overall outcomes (i.e., a higher success rate of solving questions, a larger membership size, and lower membership attrition) by implementing the design decisions. In previous studies, psychology and human-computer interaction scholars have already demonstrated that the outcome of online Q&A interactions is affected by the expertise difference between questioners and answerers (Isaacs & Clark, 1987; Pollack, 1985). Although questioners are more likely to benefit from answers coming from individuals with more expert on the topic, a

larger expertise difference does not increase the benefits the questioners will receive (White & Richardson, 2012). Therefore, question routing helps to match questioners with answerers who have an appropriate expertise level so that the most expertise answerers, who usually account for a very small proportion of the community members, can be reserved for the most knowledgeable questioners. Moreover, compared to the questions submitted by novices, some of the experts may be more interested in handling difficult questions asked by other experienced members. Question routing thus benefits these members by recommending questions matching their interests, which encourages them to continuously participate in the community.

Meanwhile, social Q&A sites rely on user-generated content to answer users' questions and establishing trust in other users is thus important for the development of the communities, especially when the volume of information is large and users need cues of trustworthiness to sift through the available information in order to find out viable solutions (Golbeck & Fleischmann, 2010). Therefore, the indication of an individual's expertise can serve as a strong cue of the reliability of the information the individual provided, and it becomes easier for the questioners to obtain information that satisfies their needs. Also, expertise indication can help answerers to build reputations and establish trust with other users so that they will be more visible in the community and their contributions will be more noticeable as well.

Nonetheless, implementing these design decisions comes with trade-offs. Individuals who are active and have higher expertise levels are more likely to obtain benefits from the community when question routing and expertise indication are applied; on the other hand, the community may become less attractive to novices and light users. As mentioned above, Q&A communities often exhibit a core-periphery structure, where a small group of core members supports a much larger group of peripheral members via active participation. These design

decisions help to maintain a relatively stable core in the community by facilitating the connections between the central users; however, they also tend to reduce the possibility for the peripheral users to establish connections to the core. Thus, the core-periphery structure will be strengthened. Based on the data from a social customer support forum, Lu, Singh, and Sun (2014) found that individual members are more inclined to answer questions submitted by core members and therefore overtime the core-periphery structure creates barriers for peripheral users to seek knowledge as their expectations of receiving a solution is low. Nevertheless, it is worth noting that in Q&A communities, core members are more likely to be answerers while peripheral individuals tend to be questioners. It is hence critical to maintaining a stable group of core members so that they can consistently make contributions to support the peripheral members as the community expands (Liang & Introne, 2019; Welser et al., 2007). Therefore, the implementation of the design decisions should take into account maintaining an appropriate balance between members with different levels of expertise and activity. Since an online Q&A community will experience different stages during its development, the design and moderation of the community should change accordingly. There is unlikely to be a universally optimal design for all circumstances.

Furthermore, the agent-based model built in this study provides a more comprehensive and extensible examination of how Q&A communities operate based on individuals' interactions and decisions. The simulations can not only predict the community outcomes of the design decisions but also offer a more granular view of how the community evolves over time. Additionally, the study connects many findings and propositions in existing social Q&A and online community literature, thus creating a more comprehensive picture to inform social Q&A community theory and design.

From a broader viewpoint, this study contributes to the larger online community and knowledge management literature in several ways. First, given that the current knowledge research management research predominantly concentrates on individuals' motivations and characteristics, the current study extends the research by examining individuals' knowledge sharing behaviors and the design factors that support these interactions. Second, the study also provides more specific design insights regarding information seeking and expertise management that will not only apply to Q&A communities but also contribute to the success of other types of online communities by helping to facilitate the flow of information and promote effective interactions between individuals. Moreover, compared to the extent online community and knowledge management literature employing agent-based modeling to examine the design of information systems and communities (e.g., Nissen & Levitt, 2004; Ren & Kraut, 2014b), the current study provides greater fidelity and insights into understanding how individuals achieve different communication outcomes under different design decisions, and how community sustainability is further driven by the outcomes.

Managerial Implications

Based on the findings of the study, to better manage online Q&A communities via implementing the design decisions, community moderators and managers may want to consider several key issues pertaining to community structure and characteristics. For many new communities, the primary goal is usually to attract more new users and reach critical mass. At this stage, the amount of content is relatively low so community moderators should focus more on helping the newcomers to find solutions and facilitate the problem-solving process via expertise indication, which will expand the membership size and a community can maintain a certain level of traffic to sustain its development. As a community grows, question routing may

become important, especially when the community is susceptible to the increasing amount of low-quality content created by undesired users (Srba & Bielikova, 2016a). Also, with question routing, it is easier for core members to connect with each other, thus contributing to a stronger core to support more community activities.

Another key issue to take into account is the structure of a community. Typically, an online Q&A community consists of a large number of infrequent users as questioners and a small number of active and knowledgeable users as answerers (Mamykina et al., 2011). However, when high expertise members in the community also actively ask questions, community moderators should be cautious about employing the design decisions, as these questions are likely to draw more attention from the community members thus hindering newcomers from seeking help (Lu et al., 2014). In this case, community moderators may want to adjust the underlying algorithm to promote the priority and visibility of the questions from newcomers and beginners.

Moreover, identifying and assessing members' expertise is the first step towards implementing these design decisions. In fact, use expertise is one of the focuses of social Q&A research (e.g., Movshovitz-Attias et al., 2013; Pal et al., 2011; Pal, Harper, et al., 2012). Perhaps the most straightforward way to evaluate individuals' expertise is through self-reporting, i.e., ask the individuals to assess their expertise themselves. Another common approach is to let other users evaluate the quality of the answers and give points to the answerer for good quality contributions. Additionally, other features, including both textual and non-textual ones, have also been employed to better assess and predict individuals' knowledge and experience. Community moderators can thus combine available features and information to properly evaluate users' expertise level in order to implement the design decisions.

The model can further be extended and incorporated into the system so that community designers and moderators can utilize the model to predict potential outcomes of different design decisions based on the specific context of the community and the characteristics of its members. As Ren and Kraut (2014a) pointed out, beyond theory development, agent-based models can serve as a dynamic decision-making tool for designers to simulate experiments by varying multiple parameters at the same time. Hence, it is easier for community designers to foresee potential problems and take more proactive measures to facilitate members' interactions.

In addition, as more and more people are transitioning to an online collaborative working environment, especially in times of public health crisis, it becomes more imperative for organizations and enterprises to develop and improve the design of online knowledge sharing and collaboration platforms. Given individuals' heavy reliance on these technologies, the current study sheds light on understanding how to achieve better management of members' expertise and facilitate the information seeking process by leveraging the design of technology. Further, the model can be modified to incorporate the influx of a large number of new members to the platform when a crisis happens to examine how implementing different design decisions helps the system to adapt to the change.

Limitations and Future Research

The current study has several limitations in terms of its scope and the research method. The sustainability of online Q&A communities is a relatively broad issue and the study only focuses on a subset of the critical factors and the relevant design decisions. Meanwhile, the simulation model is established on assumptions that simplify the reality in order to capture the gist of individuals' motivations and interactions and make the model interpretable. Thus, in this section, I will discuss the limitations of the study and how it can be extended for future research.

First, it is important to specify the boundary conditions of the study and the model. As mentioned, technological infrastructures and financial resources are closely related to the sustainability of Q&A communities, which include topics about data storage and processing, user interface design, business operations, and so on. Nonetheless, these topics require other domains of knowledge, such as computer science and business administration, and thus are beyond the scope of the study. Future research can extend the study by combining the technological and/or financial elements of Q&A communities and explore how these elements interact with the social ones to provide a more comprehensive view of the sustainability of the system.

The simulation is based on one type of Q&A community where its members are primarily seeking and offering information support. In some other types of Q&A communities, such as health, interpersonal relationship, or hobby-related ones, individuals may also look for emotional support and networking. Individuals' motivations and the way they interact differ in part with respect to the context of a community. For example, an individual in a health support Q&A community may benefit not only from their question being solved but also from being emotionally supported by their peers. Therefore, to generalize the findings of the study to a broader set of Q&A communities, one may want to integrate additional factors into the model. One potential direction for future research is to consider relational benefits and examine the outcomes of design decisions associated with members' social relations in a community.

For simplicity's sake, the model assumes that individuals' expertise, interests, and preferences are formed before they join the simulated platform and remain unchanged over time. Nevertheless, this assumption may not hold in reality as it is likely that individuals can gain more experience and knowledge through their participation and interactions with others, and their behaviors will also change accordingly. Therefore, future studies can further model individuals'

characteristics from a dynamic perspective. For instance, one can first empirically investigate how a novice becomes a more knowledgeable member and shifts their behavior from submitting questions to posting answers; and then turn the empirical findings into parameters and functions in an agent-based model for more realistic simulations. In Appendix D, the original model is extended by allowing members' expertise to be variable. The results are mainly the same except for the changes in member composition over time, which will be discussed in detail in Appendix D. Additionally, future research can incorporate more variations of the model parameters to further examine the impact of agents' interactions. For instance, studies in the future can apply mathematical and probabilistic models to simulate participation benefits and costs. Future studies can also model the initial stage of a community separately to examine how it will impact the development of the community.

Another direction for future research is to explore the user composition of online Q&A communities. Unlike the one I simulated in this study, some communities may consist of a smaller number of beginners and novices and a bigger portion of knowledgeable members and experts. It is interesting to examine if the effects of the design decisions still exist as the difference in members' expertise is smaller in this case. In addition, future studies can also include undesired groups of users in the model, such as churners and vandalizers. In fact, the emergence and prevalence of these users have become one of the problems impacting the sustainability of the system (Srba & Bielikova, 2016a). It is hence critical to understand what design decisions can be utilized to reduce and eliminate the negative impacts and promote community health.

Moreover, similar research can also be conducted in organizational and enterprise settings where organizational knowledge and expertise sharing have been benefited by the

adoption of enterprise social media (ESM). Despite sharing important similarities with general online communities in terms of the fluidity of activities and collaboration among individuals who are usually unknown to each other (Faraj et al., 2011), ESM exhibit several distinctive characteristics as individuals' participation can often lead to certain offline benefits including positive performance reviews and promotions (Bulgurcu et al., 2018). As a relatively new domain of inquiry, the study of knowledge sharing and collaboration via ESM relies on the existing theories and findings of online communities outside of enterprise settings. Meanwhile, extending the current research to model knowledge sharing interactions in organizations can yield novel insights about system dynamics and the impact of design decisions regarding problem-solving and expertise sharing that may not be observed outside enterprises.

Chapter 6

CONCLUSION

A successful online Q&A community enables its members to find and share their knowledge and expertise and is an indispensable source of information on the World Wide Web. Its sustainability not only relies on the members' contributions but also on community designers' and moderators' careful management and interventions. In this study, I provided a socio-technological perspective to investigate the design of Q&A communities as a social system, particularly focusing on user expertise. This investigation thus emphasizes interpersonal interactions in the knowledge sharing process, i.e., finding people with the right expertise. One of the strengths of this study is that it integrates the key components in the system, including individuals and information, into a dynamic process. By doing so, it allows researchers and practitioners to take a closer look at how system-level outcomes are influenced by different design decisions over time. Meanwhile, such integration makes it easier and more straightforward for designers and moderators to actively make decisions to promote the health of a community based on a systematic view of members' characteristics, motivations, and interactions.

Being an emergent social system that involves multiple types of actors interacting with each other by means of technology infrastructures and functions, online Q&A communities pose a challenge for researchers who are interested in understanding the underlying dynamics and processes due to their inherent complexity. The current study incorporates various theoretical perspectives, and combines empirical data with computational simulations, to develop a dynamic model that is relatively parsimonious yet able to present fundamental understandings of how a Q&A community sustains its development. Moreover, the flexibility of the model allows for

further integration of additional insights and features so that the model can be expanded and augmented according to researchers' and designers' specific goals and needs. Building upon this study, more experiments and tests can be conducted to deepen the theoretical understanding of online Q&A communities and inform better system design.

APPENDICES

APPENDIX A

Goodness-of-fit Plots for Fits of Attraction Rates and Contribution Likelihood

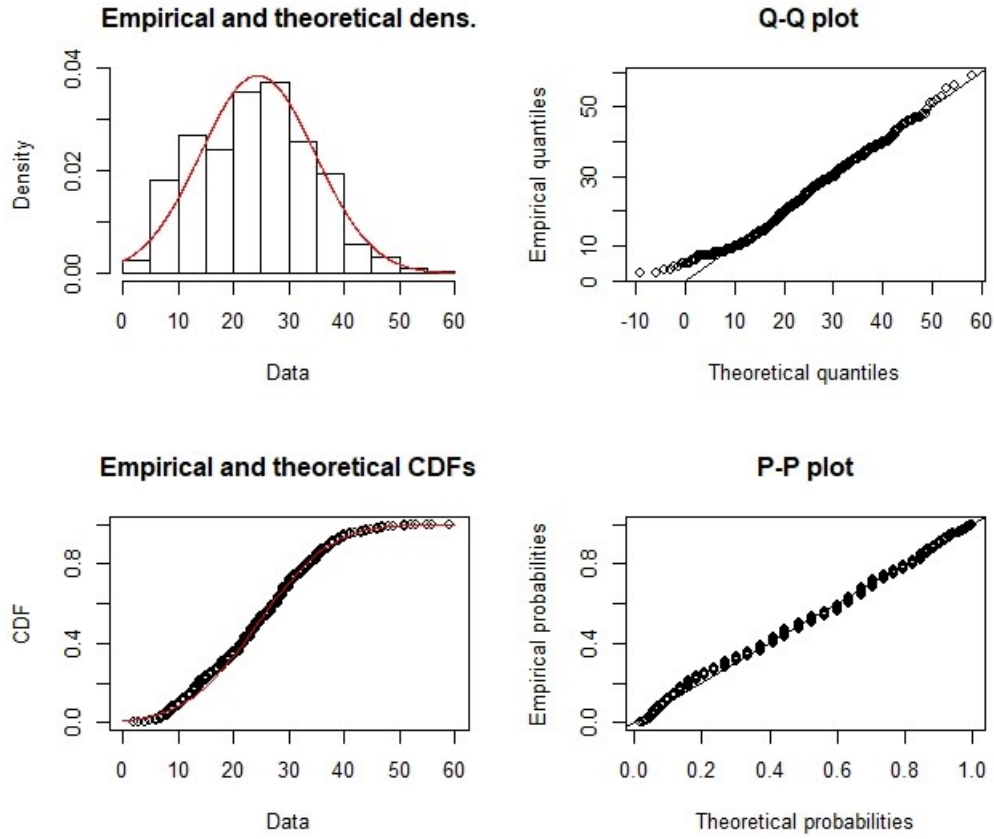


Figure 13 Goodness-of-fit Plots for Calibrating Attraction Rate by Fitting a Normal Distribution.

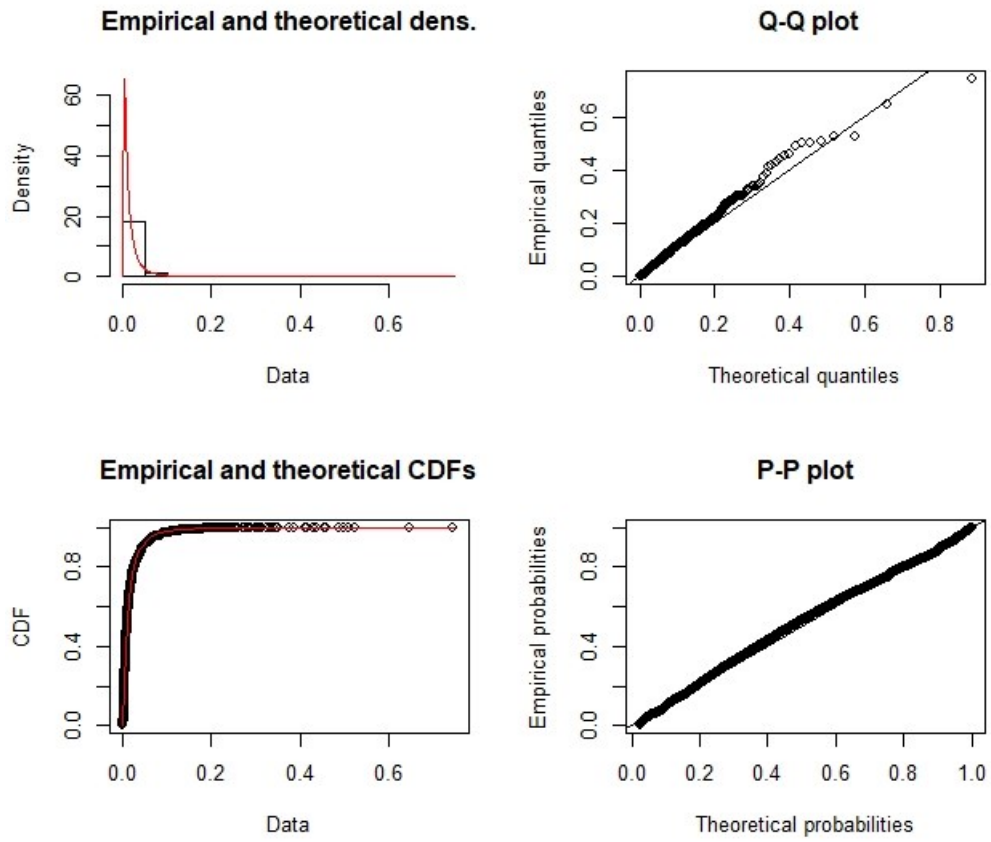


Figure 14 Goodness-of-fit Plots for Calibrating the Contribution Likelihood of Low Expertise Members by Fitting a Log-normal Distribution.

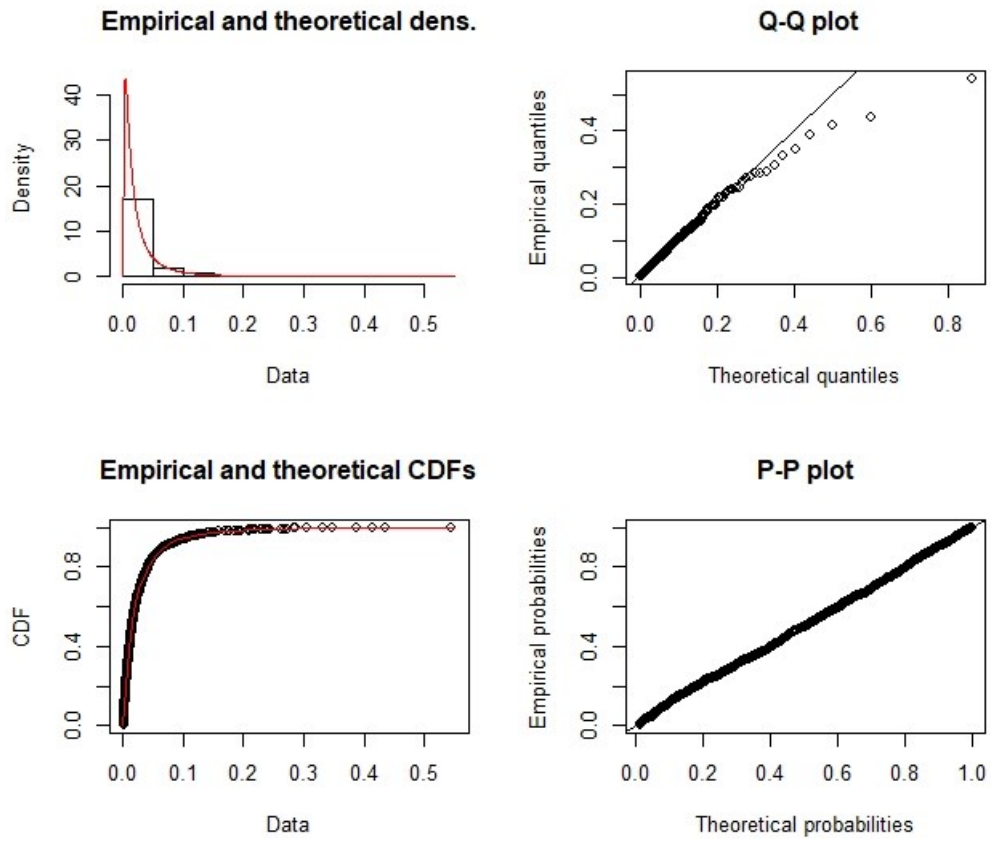


Figure 15 Goodness-of-fit Plots for Calibrating the Contribution Likelihood of Medium Expertise Members by Fitting a Log-normal Distribution.

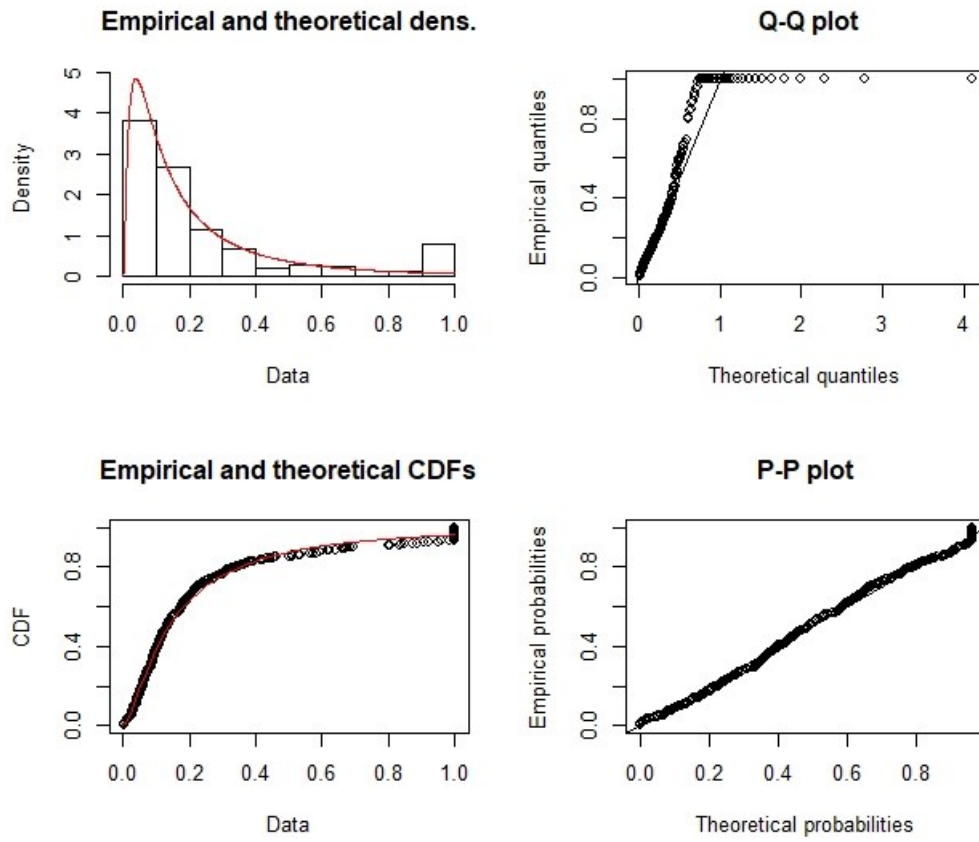


Figure 16 Goodness-of-fit Plots for Calibrating the Contribution Likelihood of High Expertise Members by Fitting a Log-normal Distribution.

APPENDIX B

Model Parameters and Values

Table 7

Model Parameters, Explanation, and Values

Parameters	Explanation	Values
Attraction Rate (A)	The number of new members joining the community in each time period.	Randomly draw from a normal distribution with $\mu = 24.38$ and $\sigma = 10.35$.
Expertise (E)	The indicator of an individual's knowledge and experience of the subject matter, which ranges from 0 to 1 and has three levels (low/mid/high).	Low: 90% of the chance fall between (0, 0.33); Mid: 8% of the chance fall between (0.33, 0.66); Hi: 2% of the chance fall between (0.66, 1).
Contribution Likelihood (CL)	The likelihood of an individual submitting a message to the community platform, which ranges from 0 to 1.	Low: log-normal with $\mu = -4.59$ and $\sigma = 1.1$; Mid: log-normal with $\mu = -4.17$ and $\sigma = 1.17$; Hi: log-normal with $\mu = -1.98$ and $\sigma = 1.12$.
Contribution Predilection (CP)	The probability of an individual's decision to submit a question to the community platform, which ranges from 0 to 1.	Low: 90% of the chance to post a question; Mid: 7% of the chance to post a question; Hi: 3% of the chance to post a question.
Participation Cost (C)	The time and effort needed to compose a message and submit it to the community platform.	0.3 to post a question; 0.075 to post an answer
Participation Benefit (B)	The value derived from a solved question an individual is involved via asking the question or offering an answer.	$C + CL - CL^2$

APPENDIX C

Computer Programs for Building the Agent-based Model

As mentioned in Chapter 3, the agent-based model is implemented via Mesa (Masad & Kazil, 2015), which is an agent-based modeling framework in Python 3+. In this section, a brief description of how the model is developed will be provided. The source code is available upon request.

The simulation process contains two core classes: one for the overall model and the other for the agents. The model class holds the model-level attributes, manages the agents, and handles the outputs of the simulations. Each instantiation of the model class will be a specific simulation run. Each model will contain multiple agents, all of which are instantiations of the agent class.

Specifically, there are two types of agents within the agent class: a community agent and community member agents. Each simulation run will have only one community agent, representing a Q&A community platform. The community agent can store a list of questions and the answers posted by the member agents, and it can also record the membership status of each member. Additionally, the community agent is initialized by specifying the design decisions being implemented (question routing and expertise indication) and execute corresponding functions.

On the other hand, there are multiple member agents in each simulation run, representing the members participating in the community. Basically, each member agent is initiated by determining its expertise based on the calibrated probability distributions and it can record a list of questions it submitted to the community and a separate list for the replies it posted. Furthermore, a member agent can execute several functions to represent different activities in the community, including reading a question, posting a question or an answer, change the status of a

question to solved, and calculate participation costs and benefits. The list of a question a member agent will read depends on whether question routing is implemented; the way a member agent determines if a question is solved is based on whether expertise indication is employed. Each member agent will execute the functions in the following order: (1) change the status of the posted questions; (2) calculate benefits and costs; (3) read posts; and (4) write posts.

Within the model class, each simulation run (i.e., running the model for multiple iterations) is initiated by creating a community agent, specifying the number of seed member agents, and create member agents according to the number. Moreover, at the beginning of each iteration, new member agents will be created, the number of which is drawn from the calibrated normal distribution. Then, member agents will perform a series of activities, which will further generate results for analysis.

APPENDIX D

A Model with Variable Member Expertise

In this model, each member's expertise is assumed to be variable based on their question answering behaviors. A member's expertise will increase when the question they answered gets solved, and the amount of increase is based on the cumulative distribution function of a beta distribution ($\text{Beta}(1.2, 1)$). As shown in Figure 17, the x-axis represents the original expertise and the y-axis indicates the new expertise, and the amount of increase is the highest for members with a medium level of expertise (around 0.5).

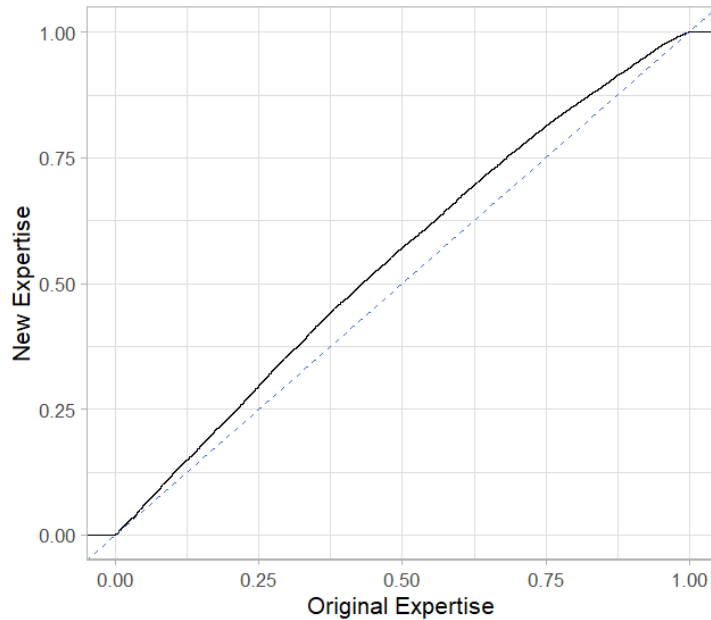


Figure 17 Expertise Growth Curve

The simulation results of this model are very similar to the one with fixed members' expertise, except the changes in member composition over time (shown in Figure 18, Figure 19, and Figure 20). As members' expertise is variable, the proportion of high expertise members increases over time, especially when both design decisions, question routing and expertise

indication, are implemented. In general, the proportion of low expertise members decreases over time, except for the case when neither decision is implemented. In terms of the members with a medium level of expertise, when question routing is implemented, the proportion increases over time; on the other hand, the proportion decreases in the scenario where neither decision is implemented. The proportion remains stable when only expertise indication is implemented.

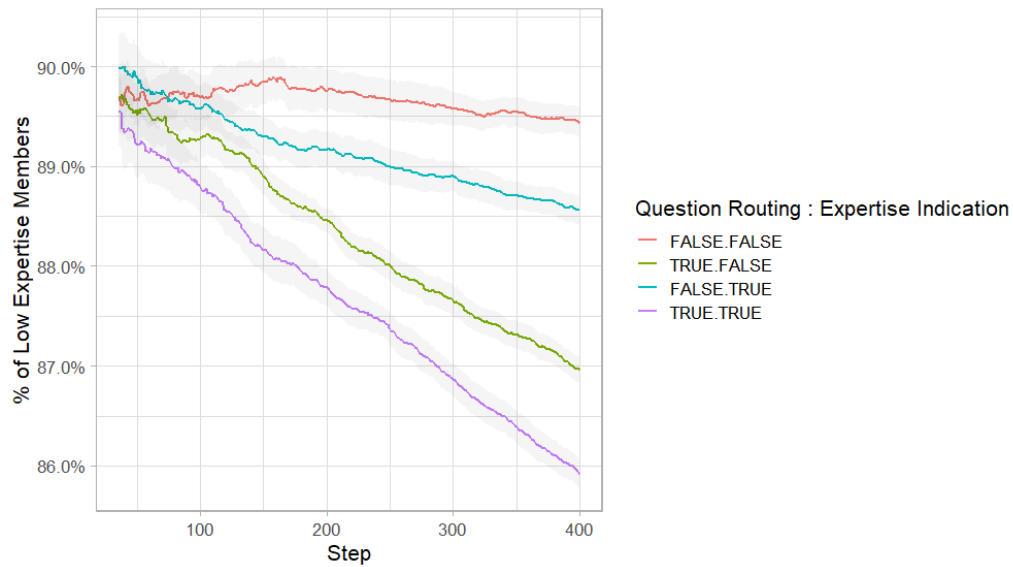


Figure 18 Proportion of Low Expertise Members over the Simulation Period with Variable Expertise

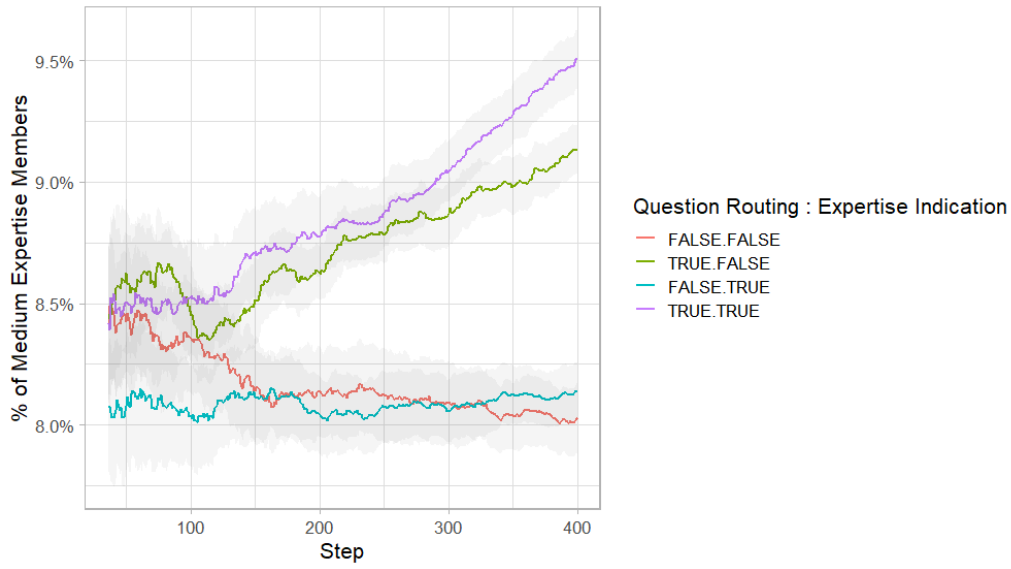


Figure 19 Proportion of Medium Expertise Members over the Simulation Period with Variable Expertise

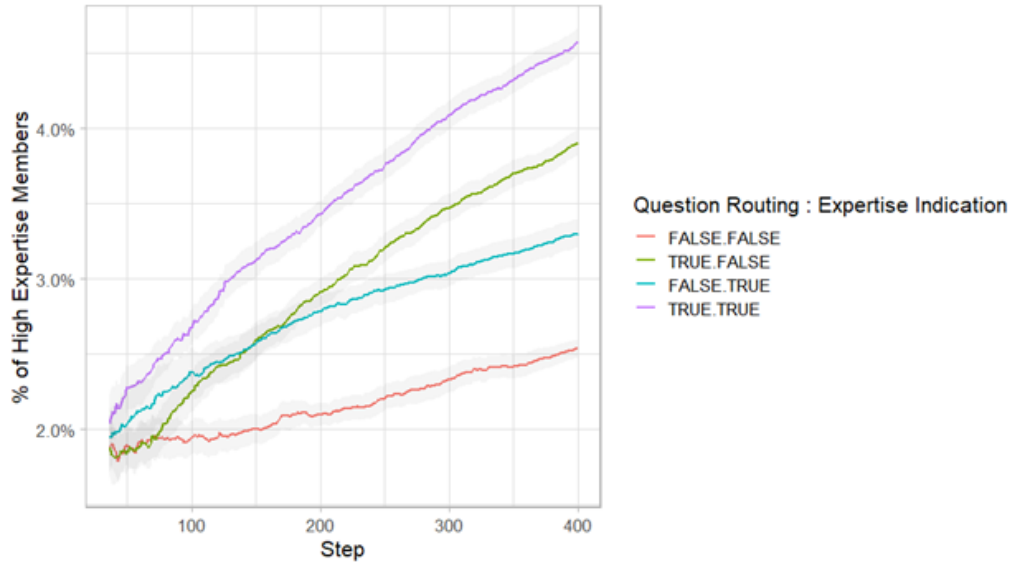


Figure 20 Proportion of High Expertise Members over the Simulation Period with Variable Expertise

Therefore, the results suggest that question routing particularly benefits low expertise members to become more experienced as these members are more likely to see questions that match their expertise level, making it easier for them to participate. In terms of expertise indication, it mainly helps questioners to identify the best answers, thus increasing the success rate. Although more questions being solved are also important for answerers to gain experience, the effect is relatively moderate compared to question routing.

REFERENCES

REFERENCES

- Ackerman, M. S., Dachtera, J., Pipek, V., & Wulf, V. (2013). Sharing Knowledge and Expertise: The CSCW View of Knowledge Management. *Computer Supported Cooperative Work (CSCW)*, 22(4–6), 531–573. <https://doi.org/10.1007/s10606-013-9192-8>
- Ackerman, M. S., Pipek, V., & Wulf, V. (Eds.). (2003). *Sharing expertise: Beyond knowledge management*. MIT Press.
- Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008). Knowledge sharing and yahoo answers: Everyone knows something. *Proceedings of the 17th International Conference on World Wide Web*. <https://doi.org/10.1145/1367497.1367587>
- Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. *Proceedings of the International Conference on Web Search and Web Data Mining*, 183–194. <https://doi.org/10.1145/1341531.1341557>
- Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2012). Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 850–858. <https://doi.org/10.1145/2339530.2339665>
- Ardichvili, A. (2008). Learning and Knowledge Sharing in Virtual Communities of Practice: Motivators, Barriers, and Enablers. *Advances in Developing Human Resources*, 10(4), 541–554. <https://doi.org/10.1177/1523422308319536>
- Arguello, J., Butler, B. S., Joyce, L., Kraut, R., Ling, K. S., & Wang, X. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 959. <https://doi.org/10.1145/1124772.1124916>
- Aslay, Ç., O'Hare, N., Aiello, L. M., & Jaimes, A. (2013). Competition-based networks for expert finding. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '13*, 1033. <https://doi.org/10.1145/2484028.2484183>
- Aumayr, E., & Hayes, C. (2014). Modelling User Behaviour in Online Q&A Communities for Customer Support. In M. Hepp & Y. Hoffner (Eds.), *E-Commerce and Web Technologies* (Vol. 188, pp. 179–191). Springer International Publishing. https://doi.org/10.1007/978-3-319-10491-1_19

- Bateman, P. J., Gray, P. H., & Butler, B. S. (2011). The Impact of Community Commitment on Participation in Online Communities. *Information Systems Research*, 22(4), 841–854. <https://doi.org/10.1287/isre.1090.0265>
- Bian, J., Liu, Y., Zhou, D., Agichtein, E., & Zha, H. (2009). Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement. *Proceedings of the 18th International Conference on World Wide Web*, 51–60. <https://doi.org/10.1145/1526709.1526717>
- Bratley, P., Fox, B. L., & Schrage, L. E. (1987). *A Guide to Simulation*. Springer New York. <https://doi.org/10.1007/978-1-4419-8724-2>
- Bulgurcu, B., Van Osch, W., & Kane, G. C. (Jerry). (2018). The Rise of the Promoters: User Classes and Contribution Patterns in Enterprise Social Media. *Journal of Management Information Systems*, 35(2), 610–646. <https://doi.org/10.1080/07421222.2018.1451960>
- 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. <https://doi.org/10.1287/isre.12.4.346.9703>
- Butler, B. S., Bateman, P. J., Gray, P. H., & Diamant, E. I. (2014). An Attraction-Selection-Attrition Theory of Online Community Size and Resilience. *MIS Quarterly*, 38(3), 699–728. <https://doi.org/10.25300/MISQ/2014/38.3.04>
- Butler, B. S., Sproull, L., Kiesler, S., & Kraut, R. E. (2007). Community Effort in Online Groups: Who Does the Work and Why? In S. P. Weisband (Ed.), *Leadership at a distance: Research in technologically-supported work*. Lawrence Erlbaum Associates.
- Carley, K. (1996). Validating Computational Models. *Working Paper*.
- Carley, K. M. (2001). Computational Approaches to Sociological Theorizing. In J. H. Turner (Ed.), *Handbook of Sociological Theory* (pp. 69–83). Springer US. https://doi.org/10.1007/0-387-36274-6_4
- Chen, A., & Edgington, T. (2005). Assessing Value in Organizational Knowledge Creation: Considerations for Knowledge Workers. *MIS Quarterly*, 29(2), 279. <https://doi.org/10.2307/25148680>
- Chen, J., Ren, Y., & Riedl, J. (2010). The effects of diversity on group productivity and member withdrawal in online volunteer groups. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 821. <https://doi.org/10.1145/1753326.1753447>
- Chung, N., Nam, K., & Koo, C. (2016). Examining information sharing in social networking communities: Applying theories of social capital and attachment. *Telematics and Informatics*, 33(1), 77–91. <https://doi.org/10.1016/j.tele.2015.05.005>

- Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007). Developing Theory Through Simulation Methods. *Academy of Management Review*, 32(2), 480–499. <https://doi.org/10.5465/amr.2007.24351453>
- Dev, H., Geigle, C., Hu, Q., Zheng, J., & Sundaram, H. (2018). The Size Conundrum: Why Online Knowledge Markets Can Fail at Scale. *Proceedings of the 2018 World Wide Web Conference*, 65–75. <https://doi.org/10.1145/3178876.3186037>
- Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge Collaboration in Online Communities. *Organization Science*, 22(5), 1224–1239. <https://doi.org/10.1287/orsc.1100.0614>
- Fisher, D., Smith, M., & Welser, H. (2006, January). You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. *HICSS 2006 - 39th Hawaii International International Conference on Systems Science*.
- Furtado, A., Andrade, N., Oliveira, N., & Brasileiro, F. (2013). Contributor profiles, their dynamics, and their importance in five q&a sites. *CSCW*, 1237. <https://doi.org/10.1145/2441776.2441916>
- Gilbert, G. N. (2008). *Agent-based models*. Sage Publications.
- Gkotsis, G., Stepanyan, K., Pedrinaci, C., Domingue, J., & Liakata, M. (2014). It's all in the content: State of the art best answer prediction based on discretisation of shallow linguistic features. *Proceedings of the 2014 ACM Conference on Web Science*, 202–210. <https://doi.org/10.1145/2615569.2615681>
- Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. A. (2009). A Conceptual and Operational Definition of “Social Role” in Online Community. *2009 42nd Hawaii International Conference on System Sciences*, 1–11. <https://doi.org/10.1109/HICSS.2009.6>
- Golbeck, J., & Fleischmann, K. R. (2010). Trust in Social Q&A: The Impact of Text and Photo Cues of Expertise. *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47*, 77:1–77:10. <http://dl.acm.org/citation.cfm?id=1920331.1920442>
- Guo, J., Xu, S., Bao, S., & Yu, Y. (2008). Tapping on the potential of q&a community by recommending answer providers. *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, 921. <https://doi.org/10.1145/1458082.1458204>
- Hall, H., & Graham, D. (2004). Creation and recreation: Motivating collaboration to generate knowledge capital in online communities. *International Journal of Information Management*, 24(3), 235–246. <https://doi.org/10.1016/j.ijinfomgt.2004.02.004>

- Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 865. <https://doi.org/10.1145/1357054.1357191>
- Hsieh, G., Kraut, R. E., & Hudson, S. E. (2010). Why pay?: Exploring how financial incentives are used for question & answer. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 305. <https://doi.org/10.1145/1753326.1753373>
- Iriberry, A., & Leroy, G. (2009). A life-cycle perspective on online community success. *ACM Computing Surveys*, 41(2), 1–29. <https://doi.org/10.1145/1459352.1459356>
- Isaacs, E., & Clark, H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1), 26–37.
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups. *Administrative Science Quarterly*, 44(4), 741. <https://doi.org/10.2307/2667054>
- Jiang, G., Ma, F., Shang, J., & Chau, P. Y. K. (2014). Evolution of knowledge sharing behavior in social commerce: An agent-based computational approach. *Information Sciences*, 278, 250–266. <https://doi.org/10.1016/j.ins.2014.03.051>
- Jolly, R., & Wakeland, W. (2009). Using Agent Based Simulation and Game Theory Analysis to Study Knowledge Flow in Organizations: The KMscape. *International Journal of Knowledge Management*, 5(1), 17–28. <https://doi.org/10.4018/jkm.2009010102>
- Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration. *Information Systems Research*, 15(2), 194–210. <https://doi.org/10.1287/isre.1040.0023>
- Kane, G. C. (2017). The evolutionary implications of social media for organizational knowledge management. *Information and Organization*, 27(1), 37–46. <https://doi.org/10.1016/j.infoandorg.2017.01.001>
- Kane, G. C., & Alavi, M. (2007). Information Technology and Organizational Learning: An Investigation of Exploration and Exploitation Processes. *Organization Science*, 18(5), 796–812. <https://doi.org/10.1287/orsc.1070.0286>
- Klein, K. J., & Kozlowski, S. W. J. (Eds.). (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. Jossey-Bass.
- Kraut, R. E., Resnick, P., & Kiesler, S. (2011). *Building successful online communities: Evidence-based social design*. MIT Press.

- Kraut, R. E., Resnick, P., Kiesler, S., Burke, M., Chen, Y., Kittur, N., Konstan, J., Ren, Y., & Riedl, J. (2012). *Building successful online communities: Evidence-based social design*. Mit Press.
- Lave, C. A., & March, J. G. (1993). *An introduction to models in the social sciences*. University Press of America.
- Law, A. M., & Kelton, D. W. (2013). *Simulation modeling and analysis* (Fifth edition). McGraw-Hill Education.
- Leonardi, P. M. (2014). Social Media, Knowledge Sharing, and Innovation: Toward a Theory of Communication Visibility. *Information Systems Research*, 25(4), 796–816. <https://doi.org/10.1287/isre.2014.0536>
- Leonardi, P. M., Huysman, M., & Steinfield, C. (2013). Enterprise Social Media: Definition, History, and Prospects for the Study of Social Technologies in Organizations. *Journal of Computer-Mediated Communication*, 19(1), 1–19. <https://doi.org/10.1111/jcc4.12029>
- Levine, J. M., & Moreland, R. L. (1994). Group Socialization: Theory and Research. *European Review of Social Psychology*, 5(1), 305–336. <https://doi.org/10.1080/14792779543000093>
- Li, B., Jin, T., Lyu, M. R., King, I., & Mak, B. (2012). Analyzing and predicting question quality in community question answering services. *Proceedings of the 21st International Conference Companion on World Wide Web - WWW '12 Companion*, 775. <https://doi.org/10.1145/2187980.2188200>
- Liang, Y. (2017). Knowledge Sharing in Online Discussion Threads: What Predicts the Ratings? *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. <https://doi.org/10.1145/2998181.2998217>
- Liang, Y., & Introne, J. (2019, January). Social Roles, Interactions and Community Sustainability in Social Q&A Sites: A Resource-based Perspective. *2019 52nd Hawaii International Conference on System Sciences*. <http://hdl.handle.net/10125/59717>
- Liu, Y., Bian, J., & Agichtein, E. (2008). Predicting information seeker satisfaction in community question answering. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 483. <https://doi.org/10.1145/1390334.1390417>
- Liu, Z., & Jansen, B. J. (2013). Factors influencing the response rate in social question and answering behavior. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 1263. <https://doi.org/10.1145/2441776.2441918>
- Lu, Y., Singh, P. V., & Sun, B. (2014). Is Core-Periphery Network Good for Knowledge Sharing? A Structural Model of Endogenous Network Formation on a Crowdsourced

Customer Support Forum. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.2486892>

- Ma, M., & Agarwal, R. (2007). Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. *Information Systems Research*, 18(1), 42–67. <https://doi.org/10.1287/isre.1070.0113>
- Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design lessons from the fastest q&a site in the west. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2857. <https://doi.org/10.1145/1978942.1979366>
- Markus, M. L. (1987). Toward a “critical mass” theory of interactive media universal access, interdependence and diffusion. *Communication Research*, 14(5), 491–511.
- Masad, D., & Kazil, J. (2015). Mesa: An Agent-Based Modeling Framework. *Proceedings of the the 14th Annual Scientific Computing with Python Conference*, 51–58.
- Movshovitz-Attias, D., Movshovitz-Attias, Y., Steenkiste, P., & Faloutsos, C. (2013). Analysis of the reputation system and user contributions on a question answering website: StackOverflow. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13*, 886–893. <https://doi.org/10.1145/2492517.2500242>
- Nam, K. K., Ackerman, M. S., & Adamic, L. A. (2009). Questions in, knowledge in?: A study of naver’s question answering community. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 779. <https://doi.org/10.1145/1518701.1518821>
- Nie, L., Zhao, Y.-L., Wang, X., Shen, J., & Chua, T.-S. (2014). Learning to Recommend Descriptive Tags for Questions in Social Forums. *ACM Transactions on Information Systems*, 32(1), 1–23. <https://doi.org/10.1145/2559157>
- Nissen, M. E., & Levitt, R. E. (2004). Agent-based modeling of knowledge dynamics. *Knowledge Management Research & Practice*, 2(3), 169–183. <https://doi.org/10.1057/palgrave.kmrp.8500039>
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.
- Oliver, P., Marwell, G., & Teixeira, R. (1985). A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action. *American Journal of Sociology*, 91(3), 522–556.
- Osch, W. van, Steinfield, C. W., & Balogh, B. A. (2015). Enterprise Social Media: Challenges and Opportunities for Organizational Communication and Collaboration. *2015 48th Hawaii International Conference on System Sciences*, 763–772. <https://doi.org/10.1109/HICSS.2015.97>

- Pal, A., Chang, S., & Konstan, J. (2012). Evolution of Experts in Question Answering Communities. *Sixth International AAAI Conference on Weblogs and Social Media*, 274–281. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4653>
- Pal, A., Farzan, R., Konstan, J. A., & Kraut, R. E. (2011). Early Detection of Potential Experts in Question Answering Communities. In J. A. Konstan, R. Conejo, J. L. Marzo, & N. Oliver (Eds.), *User Modeling, Adaption and Personalization* (pp. 231–242). Springer Berlin Heidelberg.
- Pal, A., Harper, F. M., & Konstan, J. A. (2012). Exploring Question Selection Bias to Identify Experts and Potential Experts in Community Question Answering. *ACM Transactions on Information Systems*, 30(2), 1–28. <https://doi.org/10.1145/2180868.2180872>
- Pollack, M. E. (1985). Information sought and information provided: An empirical study of user/expert dialogues. *ACM SIGCHI Bulletin*, 16(4), 155–159. <https://doi.org/10.1145/1165385.317486>
- Preece, J. (2000). *Online Communities: Designing Usability and Supporting Socialbilty* (1st ed.). John Wiley & Sons, Inc.
- Raban, D. R., & Harper, F. M. (2007). Motivations for Answering Questions Online. In T. Samuel-Azran & D. Caspi (Eds.), *New Media and Innovative Technologies*. Ben Gurion University of the Negev Press.
- Raban, D. R., & Rafaeli, S. (2007). Investigating ownership and the willingness to share information online. *Computers in Human Behavior*, 23(5), 2367–2382. <https://doi.org/10.1016/j.chb.2006.03.013>
- Rechavi, A., & Rafaeli, S. (2012). Knowledge and Social Networks in Yahoo! Answers. *2012 45th Hawaii International Conference on System Sciences*, 781–789. <https://doi.org/10.1109/HICSS.2012.398>
- Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. E. (2012). Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. *MIS Quarterly*, 36(3), 841–864. <https://doi.org/10.2307/41703483>
- Ren, Y., & Kraut, R. E. (2014a). Agent Based Modeling to Inform the Design of Multiuser Systems. In J. S. Olson & W. A. Kellogg (Eds.), *Ways of Knowing in HCI* (pp. 395–419). Springer New York. https://doi.org/10.1007/978-1-4939-0378-8_16
- Ren, Y., & Kraut, R. E. (2014b). 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. <https://doi.org/10.1080/07370024.2013.828565>

- Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies*, 28(3), 377–408.
<https://doi.org/10.1177/0170840607076007>
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67.
<https://doi.org/10.1006/ceps.1999.1020>
- Schneider, B. (1987). THE PEOPLE MAKE THE PLACE. *Personnel Psychology*, 40(3), 437–453. <https://doi.org/10.1111/j.1744-6570.1987.tb00609.x>
- Schneider, B., Goldstein, H. W., & Smith, D. B. (1995). THE ASA FRAMEWORK: AN UPDATE. *Personnel Psychology*, 48(4), 747–773. <https://doi.org/10.1111/j.1744-6570.1995.tb01780.x>
- Schweitzer, F., & Garcia, D. (2010). An agent-based model of collective emotions in online communities. *The European Physical Journal B*, 77(4), 533–545.
<https://doi.org/10.1140/epjb/e2010-00292-1>
- Shah, C., Kitzie, V., & Choi, E. (2014). Questioning the Question—Addressing the Answerability of Questions in Community Question-Answering. *2014 47th Hawaii International Conference on System Sciences*, 1386–1395.
<https://doi.org/10.1109/HICSS.2014.180>
- Shah, C., Oh, S., & Oh, J. S. (2009). Research agenda for social Q&A. *Library & Information Science Research*, 31(4), 205–209. <https://doi.org/10.1016/j.lisr.2009.07.006>
- Shah, C., & Pomerantz, J. (2010). Evaluating and predicting answer quality in community QA. *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 411. <https://doi.org/10.1145/1835449.1835518>
- Singh, P. V., Tan, Y., & Mookerjee, V. (2011). Network Effects: The Influence of Structural Capital on Open Source Project Success. *MIS Quarterly*, 35(4), 813–829.
<https://doi.org/10.2307/41409962>
- Srba, I., & Bielikova, M. (2016a). Why is Stack Overflow Failing? Preserving Sustainability in Community Question Answering. *IEEE Software*, 33(4), 80–89.
<https://doi.org/10.1109/MS.2016.34>
- Srba, I., & Bielikova, M. (2016b). A Comprehensive Survey and Classification of Approaches for Community Question Answering. *ACM Transactions on the Web*, 10(3), 1–63.
<https://doi.org/10.1145/2934687>
- Surowiecki, J. (2005). *The wisdom of crowds* (1. ed). Anchor Books.

- Taber, C., & Timpone, R. (1996). *Computational Modeling*. SAGE Publications, Inc.
<https://doi.org/10.4135/9781412983716>
- Tausczik, Y. R., & Pennebaker, J. W. (2012). Participation in an online mathematics community: Differentiating motivations to add. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 207. <https://doi.org/10.1145/2145204.2145237>
- Tausczik, Y. R., & Pennebaker, J. W. (2011). Predicting the perceived quality of online mathematics contributions from users' reputations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1885.
<https://doi.org/10.1145/1978942.1979215>
- Tian, Q., Zhang, P., & Li, B. (2013). Towards Predicting the Best Answers in Community-based Question-Answering Services. *Seventh International Conference on Weblogs and Social Media, ICWSM 2013*.
- Toba, H., Ming, Z.-Y., Adriani, M., & Chua, T.-S. (2014). Discovering high quality answers in community question answering archives using a hierarchy of classifiers. *Information Sciences*, 261, 101–115. <https://doi.org/10.1016/j.ins.2013.10.030>
- Vroom, V., Lyman, P., & Lawler, E. (2005). Expectancy theories. In J. B. Miner (Ed.), *Organizational behavior 1. Essential theories of motivation and leadership* (pp. 94–113). M.E. Sharpe.
- Wang, G. A., Wang, H. J., Li, J., Abrahams, A. S., & Fan, W. (2014). An Analytical Framework for Understanding Knowledge-Sharing Processes in Online Q&A Communities. *ACM Transactions on Management Information Systems*, 5(4), 1–31.
<https://doi.org/10.1145/2629445>
- Wang, G., Gill, K., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2013). Wisdom in the social crowd: An analysis of quora. *Proceedings of the 22nd International Conference on World Wide Web*, 1341–1352. <https://doi.org/10.1145/2488388.2488506>
- Wang, J., Gwebu, K., Shanker, M., & Troutt, M. D. (2009). An application of agent-based simulation to knowledge sharing. *Decision Support Systems*, 46(2), 532–541.
<https://doi.org/10.1016/j.dss.2008.09.006>
- Wang, S., & Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human Resource Management Review*, 20(2), 115–131.
<https://doi.org/10.1016/j.hrmr.2009.10.001>
- Wasko, M. M., & Faraj, S. (2005). Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly*, 29(1), 35–57.

- Wei, X., Chen, W., & Zhu, K. (2015). Motivating User Contributions in Online Knowledge Communities: Virtual Rewards and Reputation. *2015 48th Hawaii International Conference on System Sciences*, 3760–3769. <https://doi.org/10.1109/HICSS.2015.452>
- Welser, H. T., Gleave, E., Fisher, D., & Smith, M. (2007). Visualizing the Signatures of Social Roles in Online Discussion Groups. *The Journal of Social Structure*, 8(2).
- White, R. W., & Richardson, M. (2012). Effects of expertise differences in synchronous social Q&A. *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '12*, 1055. <https://doi.org/10.1145/2348283.2348466>
- Xia, H., Du, Y., & Xuan, Z. (2013). Structural Evolution in Knowledge Transfer Network: An Agent-Based Model. In R. Menezes, A. Evsukoff, & M. C. González (Eds.), *Complex Networks* (Vol. 424, pp. 31–38). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30287-9_4
- Yao, Y., Tong, H., Xie, T., Akoglu, L., Xu, F., & Lu, J. (2015). Detecting high-quality posts in community question answering sites. *Information Sciences*, 302, 70–82. <https://doi.org/10.1016/j.ins.2014.12.038>
- Zhang, J., Ackerman, M. S., & Adamic, L. (2007). Expertise networks in online communities: Structure and algorithms. *Proceedings of the 16th International Conference on World Wide Web*, 221. <https://doi.org/10.1145/1242572.1242603>