Is Core-Periphery Network Good for Knowledge Sharing?

A Structural Model of Endogenous Network Formation on a Crowdsourced Customer Support Forum

Yingda Lu
Rensselaer Polytechnic Institute

Param Vir Singh,
Carnegie Mellon University

Baohong Sun[1]
Cheung Kong Graduate School of Business

Updated in March 2016

[1] Yingda Lu is Assistant Professor of Rensselaer Polytechnic Institute and Param Vir Singh is Associate Professor of Information Systems of Carnegie Mellon University. Baohong Sun is Distinguished Chair Professor of Marketing of Cheung Kong Graduate School of Business (New York). The authors thank ILAB at the Heinz College, Carnegie Mellon University for their generous support and data that made this study possible. We thank seminar participants at Emory University, Rice University, Texas A&M (economics department), the University of Southern California, the University of Santa Clara, University of Iowa, University of Maryland, and University of Washington at St Louis for helpful comments.

Is Core-Periphery Network Good for Knowledge Sharing?

A Structural Model of Endogenous Network Formation on a Crowdsourced Customer Support Forum

Abstract

Many companies have adopted technology driven social learning platforms such as social CRM (crowdsourcing customer support from customers) to support knowledge sharing among customers. A number of these self-evolving online customer support communities have reported the emergence of a core-periphery knowledge sharing network structure. In this study, we investigate why such a structure emerges and its implications for knowledge sharing within the community. We propose a dynamic structural model with endogenized knowledge-sharing and network formation. Our model recognizes the dynamic and interdependent nature of knowledge-seeking-and-sharing decisions and allows them to be driven by knowledge increments and social status building in anticipation of future reciprocal rewards from peers.

Applying this model to a fine grained panel data set from a social customer support forum for a telecom firm, we illustrate that a user in this community values from being linked to other individuals with higher social status. As a result, a user is more inclined to answer questions of those who are in the core (well connected) than the ones who are in the periphery (not well connected). We find that users are taking into account the expected likelihood of their questions receiving a solution before asking a question. With the emergence of core-periphery network structure, the peripheral individuals are discouraged from asking questions as their expectation of receiving a solution to their question is very low. Thus, the core-periphery structure has created a barrier to knowledge flow to new customers who need the knowledge the most. Our counterfactuals show that hiding the identity of the knowledge seeker or making the individual contributions obsolete faster helps break the core-periphery structure and improves knowledge sharing in the community.

Keywords: Structural Modeling, Social Networks, Web 2.0, learning by sharing, social media, discussion forums, social CRM

1. Introduction

The vast and expanding reach of Web 2.0 technologies has convinced companies of the potential of crowdsourcing platforms for knowledge sharing among customers of a company. These online communities harness the power of intelligence from customers, improve customer services and lower company operational costs. On a crowdsourced customer support platform of a firm, customers can post questions whenever they face problems using the product and learn by either finding existing answers to their question on the platform, or having others answer their new question within minutes. Table 1 shows an example of one question followed by answers to it on a typical customer support platform.

[INSERT TABLE 1 ABOUT HERE]

Knowledge sharing activities on the crowdsourced customer support platform help the company achieve diversified functions such as customer relationship management (social CRM), brand community building (e.g. Sephora's beauty advisor), product innovation (ideation), and etc. It significantly reduces the cost of customer service. Furthermore, a vigorous customer support platform could also help customers get solutions in a shorter time period. As customers become more knowledgeable about the products, they will have fewer problems and will be satisfied with the product. This not only improves customer satisfaction with the service, but also helps attract more customers to switch to online channels for customer support from the more costly Call-center Solutions for customer services.

A successful customer support forum can help company reduce operational costs as well. An anecdote suggests that one "diamond member" of Dell's community support forum helped Dell save up to an astonishing $1 million by handling other customers' questions (Li and Bernoff 2009). [2] In another example, HP consumer support forum is estimated to save $50 million dollars a year and help more than

[2] It is reported that companies will save at least $10 on a support call when a customer obtains solution from her peer customers on the customer support forum (Li and Bernoff 2009). However, the company not only saves cost because of the customer who asks this question, they also save cost because a lot more customers can solve their own problems by viewing this answer, instead of calling customer representatives for help. For example, if 100 customers solve their problems after viewing the solution, this single post can save the company $1000.

40 million HP customers solve their technology issues (Hewlett Packard, 2012). Given the increasing adoption of crowdsourced customer support platforms by firms, it is important to understand the factors that drive the efficiency of knowledge sharing on the social platform and examine whether the current platform design is aligned with the user decision process.

Customer support platforms typically reveal a knowledge sharing network structure that represents a core-periphery structure (Zhang et al. 2007, Adamic et al. 2008, Nam et al. 2009, Singh and Tan 2010). Central actors from the core group are active contributors to the community and are connected to both central and peripheral actors. Peripheral actors, by contrast, are connected to the central actors but not to each other. We use Figure 1 to document knowledge sharing relationships among online customer support forum adopters in our research setting. One of the properties of this knowledge sharing network is that there are a few users dominating the community. This type of network structure is widely observed in various online social communities, and is denoted as a core-periphery structure. Core-periphery structures have been documented in the sociology literature (McPherson et al 2001, Borgatti and Everett 2000) and in a number of Web 2.0 settings: open source software forums (Singh et al 2011; Singh and Tan 2010), blogs (Banos et al. 2013, Obradovic and Baumann 2009), and micro-blogging (Huang et al. 2013).

[INSERT FIGURE 1 ABOUT HERE]

Given the prevalence of core-periphery structure in customer support communities, we need to understand what drives the formation of core-periphery structure and how the core-periphery structure affects knowledge sharing within the community? The understanding of the network formation and evolvement will help us provide insights into the platform design for improved knowledge sharing. In this study, we are motivated to address the following three important research questions: 1) what drives the formation of core-periphery structure in customer support platforms? 2) how the core-periphery structure influences user participation and knowledge sharing on a customer support platform? 3) how should we improve the design of the platform to improve knowledge sharing?

We build a dynamic structural model that endogenizes knowledge sharing and network formation. In this model, users decide whether to ask a question, and whose question to answer to maximize a long-term utility that depends on knowledge, social status, altruism, reciprocal rewards and the cost of actions. The proposed model recognizes "learning from peers" by allowing decisions of asking questions to depend on how all of their peers will respond and knowledge from asking a question to get updated only when a solution is provided by the peers. By allowing users to decide whose questions to answer, our model also treats the formation of the network as an endogenous decision that is driven by knowledge accumulation and social status building within the community. This model is in the same spirit as the multi-agent dynamic game with imperfect information described by Ericson and Pakes (1995) and Benkard (2004). To overcome the estimation challenge related to the large number of agents in the social platform, we use Oblivious Equilibrium (hereafter OE) to approximate individual optimal decisions in Markov Perfect Equilibrium based on the algorithm proposed by Weintraub et al (2008). The OE provides an appealing behavioral model in our context, because the equilibrium concept allows individuals to only track peers' state at aggregate level rather than at individual level. We further incorporate the EM algorithm by Arcidiacono and Miller (2011) to control for unobserved heterogeneity.

We apply our model to a panel data set with history of user participation decisions collected on a crowdsourced customer support forum of a telecom company, and compare the estimation results with several benchmark models. We find that individuals derive greater social status based utility when they answer questions by those who are in the core (i.e. users who are active and already have a lot of knowledge sharing relationships) than otherwise. Thus, when choosing questions to answer, individuals pick questions by those who are already well connected and thus have a high social status. As result, a core-periphery structure emerges where these more connected individuals ask most questions and the others in the core answer their questions. Peripheral users, on the other hand, are discouraged from participating in the community as they are less likely to receive answers to their questions. This finding indicates that the existence of the core creates a barrier to knowledge sharing for the community. Through exploratory sensitivity analyses, we find that hiding the names of the knowledge seekers (but not the sharers) breaks the core-periphery structure of the online community. In addition, if we make individual contribution levels obsolete faster, individuals are more likely to contribute. Awarding knowledge seeking also significantly improves the knowledge sharing on the forum. These sensitivity analyses shed light on how crowdsourced customer support forums could be improved to increase the efficiency of knowledge sharing for both practitioners and researchers.

Our research makes several contributions. First, we present a dynamic structural model that captures individual interactive decision making on a social customer support platform. As a result, we are able to endogenize network formation and shed light on why core periphery structure emerges on such platforms. In contrast to most of the existing literature which treats the social network as an antecedent to knowledge sharing, we are among the first that endogenize network formation. Second, we are among the first few papers to show that core-periphery structures have a negative influence on knowledge sharing on social customer support forums. While the core-periphery structure encourages the core to be more active, it discourages the periphery to participate. As a result individuals who need the help the most do not receive any help. Finally, through counterfactuals, we suggest several design interventions that could break the core-periphery structure and increase the efficiency of knowledge sharing in the community. All the suggested design interventions can be easily applied in practice.

2. Literature Review

Our work is related to the literature on consumer learning, which focuses on understanding how individuals learn about the quality of a product through consumption (Erdem and Keane 1996), information gathering (Erdem et al 2005), exposure to quality signals contained in the price, advertising, branding (Erdem et al. 2008; Narayanan and Manchanda 2009), and peer choices (Zhang 2010, Iyengar et al. 2011, Van den Bulte and Lilien 2001). Most of the traditional learning models take an atomistic view of individual decision making and assume that customers will not strategically share information about the product except a few recent papers (for example, Chan, Li and Pierce 2013). We add to the existing literature on learning in the following ways. First, we investigate peer learning on a public forum, which is characterized by sharing and the externality of learning. This highly distinct learning mechanism inherently implies that any user's decisions cannot be made independently of others' decisions (that is, it implies interdependence) and individuals may strategically make decisions on seeking and sharing information. Second, the externality of learning on customer support platform also implies that there is a long-run spillover of knowledge throughout the community (that is, it implies dynamic and independent decision making process). This learning mechanism is quite different from observational learning (Zhang 2010) and social contagion (Iyengar et al. 2011, Van den Bulte and Lilien 2001) documented in the traditional literature that focuses on individual decision after they receive information inferred from others' decisions. Third, existing paper on peer learning, either observational learning or social contagion, assumes that knowledge exchange among friends (Aral and Walker 2011), or family (Beatty and Talpade, 1994) is exogenous and relationships among users are static. By comparison, we treat the knowledge seeking and sharing decisions as endogenous and dynamic. This permits us to investigate formation of network and examine how position in individual online social network affects knowledge sharing decisions.

Second, our paper is related to the literature on customer behavior in online communities. Researchers have only recently started to investigate the dynamics of social communities (Katona and Sarvary 2008). Mayzlin and Yoganarasimhan (2012) propose an analytical model analyzing how individual heterogeneity affects the ability to post breaking news and how the ability to find news in the blogs of others influences the bloggers' strategic link-formation decisions. Stephen and Toubia (2010) find that sellers in an online social-commerce marketplace derive significant benefit from connection with peers, and this benefit primarily comes from the accessibility enhancement of the network. Lu, Jerath and Singh (2013) model the decision of one individual trusting another whose reviews are found to be consistently helpful in an online review community. The question of why people contribute to online social media has received increasing attention in the marketing literature. Lurie et al. (2009) suggest that user identities, such as expertise, social connections and symbolic incentives (virtual points, in this case), can affect individual contributions to the community. Welser et al. (2007) find that in Q&A forums most of the activity is driven by a few highly active users.

Third, our paper is also related to research on user behavior and knowledge sharing in online discussion forums. Adamic et al (2008) analyze Yahoo Answers and find that more knowledge seekers receive higher quality answers when the participant interests in the community are focused. Analyzing Naver's question and answer forums, Nam et al (2009) find that user participation is driven by their desire to gain higher social status in the community. Lakhani and Von Hippel(2003) report that individuals in Apache online Q&A community answer questions by others in the anticipation that this may help them in getting answers for their own questions in future. Zhang et al. (2007) document the presence of a small group of users who answer each other's questions (core) while remaining users participating little on Java forums. Most of the literature on online forums employs reduced form approach, and treats the social network as an antecedent to an outcome of economic interest. In contrast, we treat the social network as a consequence of the strategic utility-maximizing actions of individuals, and endogenously form directed links among users through knowledge seeking and sharing. This allows us to conduct several sensitivity analyses to explore different ways to improve the performance of the platforms. By incorporating individual decision of answering questions into our framework, we also evaluate the impact of learning by sharing and spill-over effect in knowledge sharing.

Our modeling approach is inspired by the recent development in social network formation (Katona and Sarvary 2008) where researchers employ dynamic structural models to better illustrate the dynamics of individual decision making within social networks (Hartmann 2010). Methodologically, our research is related to the emerging literature on dynamic-competition games. Many studies have developed models to incorporate strategic interactions among forward-looking actors in various contexts: firm entry/exit (Weintraub et al. 2008, Aguirregabiria and Mira 2007), technology adoption (Ryan and Tucker 2012), product adoption (Kumar et al. 2010), blogging behavior (Huang et al. 2013) and etc. In this paper, we apply this framework to the context of an online crowdsourced customer support platform and illustrate how individuals take into account their peers' decisions on this public learning platform. In terms of estimation strategy, we use Oblivious Equilibrium (Weintraub et al. 2008) to approximate the Markov Perfect Equilibrium strategies.

3. Generalized Model Specification for Knowledge Sharing Platform

3.1 Industry Background

Online discussion forums have been widely adopted to engage customers, support peer learning and reduce operational cost of the company. A firm sponsored crowdsourced customer support platform works as follows. Typically, once registered, users are able to ask and answer questions freely on the platform. Whenever they encounter problems with the product or services that the firm offers, they can freely post their questions on the platform. After the question is posted, any member of the community can propose a solution to address this problem by providing an answering post following the corresponding question post. Because threads containing questions and answers are on the forum 24/7, any user of the platform can also obtain knowledge whenever she visits the threads on the platform. When a user asks or answers a question, her social status, number of questions asked, number of solutions provided and number of kudos received are displayed along with his/her user name. By clicking on the user name, other users can further find her personal information and all the history of asking and answering questions.

Many customer support platforms have introduced features to encourage desired behavior of participants. For example, to help users locate correct answers, some platforms allow users who ask questions to label correct answers as *solution*. Thus all other users can quickly locate the answers that actually solve corresponding problems without going through other answers. Furthermore, some platforms also display the number of views for each post along with the content of the post. This number can be used to measure how many users benefit from each post. An illustrative schema of the forum is shown in Figure 2.

[INSERT FIGURE 2 ABOUT HERE]

3.2 Decision of Asking and Answering Questions

Consider a context where there are *N* individuals on the forum. During each of the time periods $t \in \{1,2,..,T\}$, every individual $i \in \{1,2,\dots,N\}$ makes two decisions: whether to ask a question and whether to answer a question. More specifically, we use a_{it} to denote the decision that individual i asks a question in period t:

$$(1) \qquad a_{it} = \begin{cases} 1, & if\ i\ asks\ question\ at\ time\ t \\ 0, & otherwise \end{cases}$$

Here $a_{it} = 0$ also incorporates the possibility of other knowledge seeking alternatives. We use the dummy variable s_{ijt} to denote the binary decision of an individual i deciding to answer a question posted by individual j at time t.

$$(2) \qquad s_{ijt} = \begin{cases} 1, & if\ i\ answers\ a\ question\ asked\ by\ j\ at\ time\ t \\ 0, & otherwise \end{cases}$$

Users may answer multiple questions asked by different knowledge seekers during the same period. We use the vector $\boldsymbol{s_{it}}$ to represent the set of answering decisions for individual i at time *t*. Note that when an individual chooses to stay as an observer on the forum, our model treats it as a choice not to ask or answer, and a_{it} as well as s_{ijt} remain zero. s_{ijt} is both *i* and *j* specific in this model. This specificity means that we consider the source of the question and allow the user to decide whose question to answer. This dyadic nature of answering decisions permits us to model the possible strategic interactions among individual users and to investigate the fundamental drivers for the formation of the interactive (question-answer) network. This is among the first papers that introduce dyadic individual interactions to choice models (Ma et al. 2010).

3.3 Per-Period Utility Function

The per-period utility function of user participation can be written as a function of knowledge, social status, altruism/self satisfaction and costs of asking and answering questions:

$$(3) \qquad U_{it}\,(\boldsymbol{K}, \boldsymbol{R}, \boldsymbol{Z_i}, a_{it}, \boldsymbol{s_{it}}, \boldsymbol{\varepsilon_{it}}) = \alpha_1\ K_{it} + \alpha_2\ R_{it} + \alpha_3\ \textstyle\sum_j s_{ijt} - C\ \ (a_{it}, \boldsymbol{s_{it}}, \boldsymbol{Z_i}) + \varepsilon_{it}(a_{it}, \boldsymbol{s_{it}}),$$

where K_{it} is the knowledge level accumulated by individual i at time t.[3] R_{it} is the social status level for individual i at time t. $\boldsymbol{Z_i}$ denote i's individual characteristics, such as tenure on the forum, that may affect the cost of asking and answering questions. These factors account for potential observed heterogeneity in costs across individuals. We let ε_{it} denote the action specific private shock that is only observable by the focal individual. We assume that ε_{it} has a type-I extreme-value distribution and that private shocks are *iid* across participants and periods. $\sum_j s_{ijt}$ is the number of questions answered by individual *i* in period *t*. C is the cost of asking and answering questions. Hence, $\alpha_3\ \sum_j s_{ijt}$ represents the altruistic utility derived from helping others. Note that α_3 cannot be separately estimated from cost of answering as both are linearly related to number of answers provided by a user in a period. Hence, we normalize α_3 to zero. As a result, the estimated cost of answering is the net of cost of answering and the altruistic utility gained from answering.

3.3.1 Knowledge Updates

[3] We also estimate a model with squared term of knowledge level in utility function to capture the potential nonlinear impact of individual knowledge on utility function. However, because the estimated coefficient for square term is insignificant, we do not include this term in our model for simplicity.

We use the term *knowledge* to represent individual knowledge accumulated on the customer support platform, rather than a measure of the overall knowledge level of the individual. The accumulation of knowledge helps the customer solve problems related to the product or services offered by the firm. Knowledge gain can provide utility as the individual is better able to use the product and gain the maximum out of it.

An individual can update her knowledge by two processes. First, an individual's knowledge is updated when someone answers her question. Second, answers to all questions are stored on the online platforms and users have 24/7 access to them. This means that all users of the platform can benefit and increase their knowledge by viewing these answers. In other words, a user can increase her knowledge even without asking a question.

However, not all answers are beneficial for members of the community. In particular, some answers may be incorrect and cannot help users who post the question. Thus we only consider the answers that are identified as "solutions" in the knowledge updating process. As discussed earlier a large number of platforms employ mechanisms to identify these solutions among answers. We will elaborate these designs and how we identify solutions in this research context in Section 4.

We use $\boldsymbol{SN_{it}}$ to denote the set of solutions posted for individual i's question at period t. We model K_{it} , the knowledge level of individual *i* at time t, as follows:

$$(4)\ K_{it} = \beta_k K_{it-1} + k_1 I\left(\sum_{j\in N, j\neq i} s_{jit-1} \in \{\boldsymbol{SN_{it-1}}\}\right)$$

$$+k_2\ \sum_{o\in N, o\neq i} View_{i,a_{ot-1}} I\left(\sum_{j\in N, j\neq i} s_{jot-1} \in \{\boldsymbol{SN_{it-1}}\}\right)$$

β_k is a discount parameter that accounts for possible decay of knowledge. This is because individuals may forget knowledge over time, and knowledge itself may also get outdated and become less useful (Ryu et al 2005, Boulding 1996). This discount parameter also helps to ensure individual knowledge level is bounded.

The knowledge updating equation captures both learning by asking a question and passive learning, two salient features introduced by customer support platforms. The second term in equation 4 recognizes learning from asking a question in which a user asks questions and learns from reading the solutions provided to her specific question. By abusing the notation, $I\left(\sum_{j\in N,j\neq i} s_{jit-1} \in \{\boldsymbol{SN_{it-1}}\}\right) = 1$ if user i received a solution to her question at time $t-1$ and zero otherwise.

The third term captures knowledge gain through passive learning where a user gains knowledge by browsing through solutions provided to questions asked by others. $I\left(\sum_{j\in N,j\neq i} s_{jot-1} \in \{\boldsymbol{SN_{it-1}}\}\right) = 1$ if there is a solution provided by another user to someone else's question other than i or j. Thus $\sum_{o\in N,o\neq i} I\left(\sum_{j\in N,j\neq i} s_{jot-1} \in \{\boldsymbol{SN_{it-1}}\}\right)$ represents the total number of questions which are posted by individuals other than i that received a solution. Notice that individual will gain knowledge from a solution contributed by others only if she views the answer. We define $View_{i,a_{it-1}} = 1$ if individual i views the solution posted for o's question, and zero otherwise.

If k_1>0 and k_2>0 then the two components in knowledge updating rule, $I\left(\sum_{j\in N,j\neq i} s_{jit-1} \in \{\boldsymbol{SN_{it-1}}\}\right)$ and $\sum_{o\in N,o\neq i} View_{i,a_{ot-1}} I\left(\sum_{j\in N,j\neq i} s_{jot-1} \in \{\boldsymbol{SN_{it-1}}\}\right)$, imply that user *i*'s decisions are not independent of the decisions of her peers. For example, individual i's decision about whether to ask a question at each period *t* would depend on her expectation on whether her question may receive a solution. Furthermore, the knowledge updating rule also allows a single user's decision to increase the future knowledge levels of all her peers when the passive learning is in place: solutions posted to the community increase the knowledge level of the whole community.

3.3.2 Online Social Status Updates

There are various reasons why people benefit from contributing to customer support communities even in the absence of direct monetary benefits. First, individuals obtain social recognition from community by contributing to the community (Nam et al. 2009). The more they contribute to the community, the more likely that they will be recognized as ones with competence and commitment. As a result, the community will ascribe a higher social status to them. Second, it may be attributed to reciprocal rewards from helping others. For example, individuals may expect that in the future others may respond back as a result of their present contributions (Lakhani and Von Hippel 2003). Or, an individual may be motivated to give back to the community to reciprocate support that he/she may have received in the past (Lampel and Bhalla 2007). When individuals actively contribute to the community, their peers on the community all benefit from their voluntary input because of the passive learning effect. Hence, members of the community are also more likely to award individuals who actively contribute to the platform.

We use the frequency of user contributed solutions to approximate a user's contribution to the community. This is consistent with the common practice that a large number of social customer support platforms where they rate or rank users based on their cumulative frequency of posting solutions.[4] To be more specific, we use A_t to denote an $N \times N$ adjacency matrix that measures the intensity of interaction between i and j (in our context, it is the number of solutions provided) up to time t. The $(i,j)th$ element of A_t is the number of solutions provided by i for individual j's question up to time t. Defining $\boldsymbol{\chi_t}$ as an *N*x1 vector containing perceived contribution level for users at the beginning of period t, then we can calculate individual contribution level in the system according to $\boldsymbol{\chi_t} = \beta_r \boldsymbol{\chi_{t-1}} + (\boldsymbol{A_{t-1}} - \boldsymbol{A_{t-2}}) \cdot \mathbf{1}_{N\times 1}$. β_r is a discount factor capturing depreciation of contribution -- that newly created contributions should have more weight compared to the older ones. $\mathbf{1}_{N\times 1}$ is a vector with ones. $(\boldsymbol{A_{t-1}} - \boldsymbol{A_{t-2}}) \cdot \mathbf{1}_{N\times 1}$ is an *N*x1 vector representing additional number of answers for individuals at period $t-1$.

[4] We do not incorporate the answers that are not solutions in the social status updating rule here because answers that are incorrect or not helpful will not be recognized by other members of the community. Nonetheless, in one of our robustness check, we extend our model to incorporate the impact of providing a non-solution answer on online social status. We find that our estimation results remain the same. In particular, the impact of non-solution answers on online social status is statistically insignificant. As a result, we ignore the non-solution answers in our main framework for simplicity.

Extensive sociology and psychology literature has shown that building social status may go far beyond a simple measure of individual contribution level. For example, Bonacich (1987) shows that entities wield power or benefit from being central in a community and that entities closely connected with other central participants are considered to have high status. In our context, this finding means that those who post solutions for questions posted by high-status participants may obtain a higher perceived social status themselves. Solving questions by individuals whom the community considers knowledgeable and higher status indicate that the solution provider is more knowledgeable, active, and resourceful. In other words, individual network position could also contribute to their social status, denoted as *network position effect* in this paper. This may provide an additional incentive for individuals to answer questions from high social status individuals. This consideration requires us to explicitly incorporate the dyadic relationship between individuals, which is defined for any pair of individuals by how many answers they provide to each of other users.

To take into account the possibility that those who solve questions posted by high-status participants should obtain a higher perceived social status themselves, we write the *social status* of individual i by summing contribution level and network position effect:

$$(5) \qquad \boldsymbol{x_t} = \boldsymbol{\chi_t} + \gamma_1 \boldsymbol{A_{t-1}} \boldsymbol{\chi_t}$$

$\boldsymbol{x_t}$ is an Nx1 vector denoting the *absolute social status* of individual i at the beginning of period t. The second component $\boldsymbol{A_{t-1}} \boldsymbol{\chi_t}$ is an Nx1 vector capturing the network position effect. Each element in this vector is calculated as the contribution provided by i weighted by the contribution level of the one who posted the question. Here, γ_1 measures the impact of contribution level of the knowledge seeker on the knowledge sharer's network position level. We do not impose any constraints on the sign of γ_1, thus allow it to be empirically estimated. If $\gamma_1 > 0$, users benefit more from solving questions of a high social status individual.

The measure in equation (5) is an absolute measure of social status. In reality, individuals care about their relative rank in the community, rather than the absolute rank (Frank 1985, Kumar et al. 2010). To adapt to this observation, we model user psychological perception of social status (French and Raven 1959), and define the *social status score* in our context by

$$(6) \qquad R_{it} = \frac{x_{it} - \min(\boldsymbol{x_t})}{\max(\boldsymbol{x_t}) - \min(\boldsymbol{x_t})}$$

R_{it} can be viewed as a relative social status score calculated based on summary statistics of contribution frequency and association with others. This is consistent with the practice of many forums that rank users based on their contribution level. This also guarantees that social status is bounded.

Note that the social status updating process is different from the knowledge updating process for the following two reasons. First, users improve their social status mainly by answering questions. However, they improve their knowledge mainly by asking questions and reading the answers posted by others. Second, the knowledge updating rule depends on the quality of answers but not on who answered them. However, the social status updating rule accounts for whose question is answered.

3.4 Costs of Asking and Answering Questions

While knowledge and social status are influenced by the decisions of whether to ask and answer questions, there are also costs associated with each one of these decisions. When asking a question, a user needs to invest time in posting the question on the forum and in carefully phrasing it so that people in the community can correctly understand it. When answering a question, she needs to first think about the answer and then express it on the forum in an organized and clear manner. Both of these two processes are time consuming. To be more specific, we write the costs for each time period as

$$(7) \qquad C\ (a_{it}, \boldsymbol{s_{it}}, \boldsymbol{Z_i}) = C_a\ (a_{it}, \boldsymbol{Z_i}) + C_s\ (\boldsymbol{s_{it}}, \boldsymbol{Z_i}),$$

where $C_a\ (a_{it}, \boldsymbol{Z_i})$ is the cost function of asking question and $C_s\ (\boldsymbol{s_{it}}, \boldsymbol{Z_i})$ is the cost function of answering questions. We further model the cost of asking/answering questions to be a linear function of user characteristics such as tenure on the platform.

3.5 Question Heterogeneity

The questions asked on the forum could vary in terms of quality. We explain how we identify the quality of a question in Section 4. The status gained and cost incurred from answering questions could vary with quality. As a result we model the cost function for answering a question as:

$$(8) \qquad C_s\ (\boldsymbol{s_{it}}, \boldsymbol{Z_i}) = \sum_j (c_{s,0}(1 + c_s^H I(a_{jt} = H)) + \boldsymbol{c_{s,1}} \boldsymbol{Z_i}) s_{ijt}.$$

This cost function specification allows us to capture heterogeneity in question quality. $I\left(a_{jt} = H\right) = 1$ if j asks a high quality question and 0 otherwise. c_s^H represents the additional cost of answering a high quality question. While our model only allows the quality of question to affect the cost of answering it, answering a high quality question may also provide utility in terms of status gain. Because, both cost and status gain from answering high quality question would be linear function of question quality they cannot be estimated together. As a result, c_s^H represents the net of status gain and cost of answering a high quality question. $c_{s,0}$ represents the baseline cost of answering a question. And $\boldsymbol{c_{s,1}} \boldsymbol{Z_i}$ captures the impact of $\boldsymbol{Z_i}$ on cost of answering a question.

3.6 Unobserved Heterogeneity

We take into account observable heterogeneity by incorporating available individual characteristics in the cost function. However, there could be potential unobserved heterogeneity. For example, users may have different intrinsic cost of asking and answering questions on the platform; some users may value the social status more than others; and some users may feel happier when they become the top contributor of the community compared to others. Thus our model also needs to capture the unobserved heterogeneity among users. Controlling for unobserved heterogeneity is particularly important as core users could be inherently different from peripheral users which could provide an explanation of different behavior of the two groups on the forum.

We assume there are p types of individuals. We use q_{ip} to denote the probability that individual i is of type $p \in \{1,2,\dots,P\}$. $\boldsymbol{\pi}$ is the distribution of unobserved individual heterogeneity. Individuals of different types have different specifications of utility function:

$$U_{it}^{p}(\boldsymbol{K},\boldsymbol{R},\boldsymbol{X_i},a_{it},\boldsymbol{s_{it}},\boldsymbol{\varepsilon_{it}}) = \alpha_1^p K_{it} + \alpha_2^p R_{it} - (C_a^p(a_{it},\boldsymbol{Z_i}) + C_s^p(\boldsymbol{s_{it}},\boldsymbol{Z_i})) + \varepsilon_{it}(a_{it},\boldsymbol{s_{it}}), \quad (9)$$

We also allow the effect of network position effect on social status (γ_1) to be different across segments. This is because users in different segment may have different perception of the additional social status they can derive from answering questions from high social status individuals.

We also model the cost function to account for potential unobserved heterogeneity among users. Users in different segments may have different costs of answering questions. This could be one reason why different segments may end up with different propensity to answer. Hence, the answering cost function is modeled as: $C_s^p(\boldsymbol{s_{it}},\boldsymbol{Z_i}) = \sum_j (c_{s,0}^P(1 + c_s^H I(a_{jt} = H)) + \boldsymbol{c_{s,1}}\boldsymbol{Z_i}) s_{ijt}$. And the asking cost function of modeled as $C_a^p(\boldsymbol{a_{it}},\boldsymbol{Z_i}) = c_{a,0}^P + \boldsymbol{c_{a,1}}\boldsymbol{Z_i}$.

3.7 Probability of Answer Getting Selected as Solution

While individuals incur a cost for answering a question, not all answers are identified as solutions. Because answering a question incurs cost and an individual who provides an answer gains more utility when her answer is selected as a solution, she would consider the expected probability that her answer may get selected as a solution when she decides to answer. We model the propensity of an answer being identified as solution as a function of individual state levels. Hence, the probability that an answer by individual j for question by i at period t is identified as solution is modeled using a logistic regression as:

$$Pr\big(s_{jit} \in \{SN_{it}\} | s_{jit} = 1\big) = \frac{\exp(\beta_{j1}+\beta_{j2}K_{jt}+\beta_{j3}R_{jt})}{1+\exp(\beta_{j1}+\beta_{j2}K_{jt}+\beta_{j3}R_{jt})} \quad (10)$$

3.8 User Dynamic Problem, Oblivious Equilibrium, and Estimation

Knowledge and social status are state variables, and their impacts on utility carry on over time. Accordingly, this dynamic and interactive decision making can be best approximated by assuming each individual maximizes her long term utility that depends on decisions of others:

$$(11) \qquad E\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t}\, U_i\ \ (\tau) \middle| \boldsymbol{h_t}\right],$$

where β is the discount factor indicating how much the individual values future utility. In this model setup, the state at time period t, denoted as $\mathbf{h_t} \in \mathcal{H}$, $\mathbf{h_t} = (\mathrm{h}_{1,\mathrm{t}}, \dots, \mathrm{h}_{\mathrm{N,t}})$, where $\mathbf{h_{i,t}} = \{\mathrm{K}_{\mathrm{i,t}}, \mathrm{R}_{\mathrm{i,t}}\}$. Individuals make decisions to maximize their discounted long-term utility based on the information available to them at time t. Realizing that their states and decisions are interdependent, all users will incorporate the expected responses from their peers when making decisions about asking and answering questions that maximize their own long-term utility.

The MPE would be an appropriate solution concept for this dynamic interactive game. However, the computational burden of explicitly solving for a MPE (e.g., Pakes and Mcguire 1994) is prohibitively huge due to the high dimensionality of our state space. To circumvent the computational burden of iteratively solving the dynamic game model with large number of agents, we approximate the MPE by OE specified in Weintraub et al (2008), and adapt the EM algorithm (Arcidiacono and Miller 2011, Chung et al. 2014) to control for unobserved heterogeneity.

There are two main advantages of using OE in our context. First, it has been shown that OE works very well in approximating MPE when there are a large number (>10) of agents in the model (Weintraub et al. 2008). Given the large number of agents (>1500) in our context, the performance of OE is particularly good for modeling individual decisions in our context. Second, OE also provides an appealing behavioral model that is highly consistent with individual decision patterns in our setting. Because there are many users, it is unrealistic for individuals to keep track of the states of all other users on the forum. Given an individual's limited capability of processing information, it is more reasonable to assume that users merely trace statistics at aggregate level, such as distribution of individuals at different status levels. This is consistent with the assumption of OE where in a market with many agents, each individual makes nearly optimal decision based on her own state as well as the long-run average market state, rather than keeping track of everyone's states. Here the market state means the frequency distribution of individuals across the states. We use $\mathbb{Z}$ to denote the expected distribution of states of the market in the long run. Hereafter, we use $\boldsymbol{\sigma_i}$ to denote OE strategy to ask and answer question as a function of her own state variables and the private shock for individual i: $\boldsymbol{\sigma_i} : \boldsymbol{h_i} \times \mathbb{Z} \times \boldsymbol{\varepsilon_i} \mapsto \boldsymbol{A_i}$, where $\boldsymbol{A}_i$ is the set of all actions individual i can take. Individual also expects her peers to use the publicly known OE strategy to make their own decisions. We use $\boldsymbol{\sigma_i'}$ to denote the OE strategy for individuals other than i. Because individual decisions depend only on their state levels in OE, the oblivious value function can be defined below:

$$(12) \qquad \bar{V}_i(\mathbf{h_{i,t}}|\boldsymbol{\sigma_i'}, \boldsymbol{\sigma_i}) = E\left[\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} U_i(\tau) \middle| \boldsymbol{h_{i,t}}\right]\right]$$

Here, the oblivious value function $\bar{V}_i(\mathbf{h_{i,t}}|\boldsymbol{\sigma_i'}, \boldsymbol{\sigma_i})$ is the expected net present value of an individual with state $\mathbf{h_{i,t}}$ and follows oblivious strategy $\boldsymbol{\sigma_i'}$, when her peers with long-term state $\mathbb{Z}$ follow strategy $\boldsymbol{\sigma_i}$ Then a strategy profile $\boldsymbol{\sigma}^* = \{\boldsymbol{\sigma_1^*}, \dots, \boldsymbol{\sigma_N^*}\}$ is an OE strategy for a market with a large amount of agents, in which this strategy optimizes an oblivious value function as below.

$$(13) \qquad sup_{\sigma_i^*} \bar{V}_i\left(\mathbf{h_{i,t}}\middle|\boldsymbol{\sigma_i^*}, \boldsymbol{\sigma_i}\right) = \bar{V}_i\left(\mathbf{h_{i,t}}\middle|\boldsymbol{\sigma_i}, \boldsymbol{\sigma_i}\right), \forall i, \mathcal{H} .$$

We use $\boldsymbol{\Theta}_p = \{\alpha_1^p, \alpha_2^p, c_{a,0}^p, \boldsymbol{c_{a,1}}, c_{s,0}^p, \boldsymbol{c_{s,1}}, k_s, k_x, \gamma_1, \gamma_2\}$ to represent the structural parameters for user with type p. We can write the likelihood of observing asking and answering decision history conditional on observed individual states, long-run average market state and unobserved individual type:

$$(14) \qquad L_p\left(\boldsymbol{d_i} \middle| \boldsymbol{H}_i, \boldsymbol{Z}, \boldsymbol{\sigma_i}; \boldsymbol{\Theta}, \pi_p\right) = \pi_{ip} \prod_t l(\boldsymbol{d_{it}}|\boldsymbol{H}_{it}, \boldsymbol{Z}, \boldsymbol{\sigma_i}; \boldsymbol{\Theta})$$

where $\boldsymbol{d_{it}}$ is the observed individual i's asking and answering decision at period t, $l(\boldsymbol{d_{it}}|\boldsymbol{H}_{it}, \boldsymbol{Z}; \boldsymbol{\Theta})$ is the likelihood of observing decision $\boldsymbol{d_{it}}$ at period t given their own status and long-run average market states.

Integrating over all unobserved individual types, the likelihood of observing individual i's activity history is:

$$(15) \qquad L(\boldsymbol{d_i}\ |\boldsymbol{H}_i, \boldsymbol{Z}, \boldsymbol{\sigma_i}\ ; \boldsymbol{\Theta}, \pi) = \sum_{p=1}^{P} L_p(\boldsymbol{d_i}\ |\boldsymbol{H}_i, \boldsymbol{Z}, \boldsymbol{\sigma_i}\ ; \boldsymbol{\Theta}, \pi_p)$$

Then the log-likelihood function for all users on the forum is:

$$(16) \qquad \mathcal{L} = \prod_{i=1}^{N} L(\boldsymbol{d_i}\ |\boldsymbol{H}_i, \boldsymbol{Z}, \boldsymbol{\sigma_i}\ ; \boldsymbol{\Theta}, \pi)$$

We can estimate this likelihood function by adapting the EM algorithm by Arcidiacono and Miller (2011). Intuitively, this algorithm allows us to infer user types based on their observed behaviors. For estimation purposes, we discretize the knowledge and social status. We fix the discount factors $\beta = \beta_{r,1} = \beta_{r,2} = 0.95$, and $\beta_k = 0.95$ in the estimation. Appendix A presents more detailed information on identification and estimation.

4. Empirical Result

4.1 Data Description

Crowdsourced customer support platforms have been widely adopted by industry. In this case, we apply our proposed framework on a crowdsourced customer support forum of a telecom company which sells 2G and 3G services to their customers. Instead of hiring hundreds of customer assistants to answer customers' questions, this company asks customers to post their questions on the platform, and encourage peer customers to answer these questions. As a matter of fact, the dominating channel in which customers can ask question is through this crowdsourced customer support platform, except the questions which are very specific to customer account. Customers can ask questions ranging from "why the 3G internet speed is so slow" to "why carphone warehouse won't give me a PAC code". This support forum is completely customer driven. Roughly 70% of questions received their first answer within five minutes from their peer customers on this platform.

Our estimation dataset contains all user activities on the customer support forum between Sep 2009 and May 2011. We split the sample into warm-up dataset (Sep 2009-Sep 2010) and estimation dataset (Oct 2010-May 2011). We select individuals who have answered at least one question on the forum in the estimation dataset.[5] As a result, our estimation dataset consists of 1558 individuals' activities over 30 weeks. We define one week as one time period. The warm-up dataset is used for constructing the initial values for state variables to be used at the beginning of the estimation dataset. Further, users need to learn about the platform and get comfortable with using it. The warm-up period allows for this learning to happen. Our estimation dataset covers the period when the platform was relatively mature and the data indicated the system was in equilibrium.

We employ two mechanisms to identify solutions. Individuals who ask the question can label one answer as *verified answer* of the question to indicate that this answer help solve their problems. In the meanwhile, after an answer is posted on this customer support forum, other customers on the forum can give it a "*kudo*" as a way to acknowledge the usefulness of this answer. Whether an answer received a kudo will be displayed as part of that post, and everyone can see this. Both of these two mechanisms are widely used among other major customer support forums. The main difference between these two mechanisms is that verified answers can only be identified by individuals who ask the question, while virtual award can be given by any member of the community. Both of these two designs help identify solutions to answers on the forum, and questions that are not labeled as solutions or receive no kudos are mostly the ones that will not help customers address their problems. In this paper, we use the term "*solution*" to denote the answers that were labeled verified answer or were given a kudo.

Because this platform also allows users to give kudos to the question as a way to indicate that they feel this question is helpful to the community. In other words, the number of kudos a question receives

[5] We also exclude company employees from our estimation dataset because employees may have different incentives of knowledge contribution compared with users who are customers of the company. Given that only around 1% of answers are posted by company employees, and that none of them were in the top contributor list, excluding knowledge contributions from company employees will only have minimum impact on our estimated user knowledge contribution decision.

can be regarded as a signal for the quality of the question.[6] We define a question to be of high quality if this question receives at least one kudo from peers. We find that the correlation between a question's quality and the seeker's status is only 0.1542. While this correlation is significant it also reveals that a high status in the community is not acting as a proxy for high quality questions.

In our estimation dataset, a total of 3748 questions were asked, and 2684 solutions were provided to these questions. The majority of the answers (87.89%) were posted the same day as the corresponding question was asked. Table 2 provides some sample statistics from the estimation sample. As highlighted by the core-periphery structure shown in Figure 1, the participation on the forum is heavily skewed. While, approx. 82% users provide less than 6 solutions each, approx. 2% provide more than 50 solutions each. The question asking rate is even more skewed with approx. 89% asking less than 6 questions each and approximately 1% asking more than 20 questions each. The answering rate is though less skewed with approx. 55% providing less than 6 answers each and approx. 10% providing more than 50 answers each.

[INSERT TABLE 2 ABOUT HERE]

This customer support forum also displays the total number of views for each post. To construct individual level viewing history, we simulate individuals views of each post based on their individual characteristics. The details of this simulation are specified in Appendix A.

4.3 Estimation Results

[INSERT TABLE 3 ABOUT HERE]

The number of unobserved segments is identified by comparing Akaike Information Criterion and Bayesian Information Criterion. A model with two unobserved states outperforms all other models (with no unobserved heterogeneity or more than two unobserved segments) based on these statistics. Table 3 presents the estimation results of our proposed model with two segments. We find 77.66% of users are of Type 1 on this customer support forum, while the remaining users are of Type 2. Core individuals are more likely to be Type 2 individuals while peripheral individuals are more likely to be Type 1 individuals.

[6] We thank the associate editor for pointing us to this way of measuring question quality.

Customers in different segments demonstrate different preferences on knowledge and social status. Type 1 customers derive 0.042 and 2.927 units of utility per unit of knowledge and social status respectively. Type 2 customers derive 0.070 and 4.040 units of utility per unit of knowledge and social status respectively.[7] The social status of the person whose question a user solves has a significant positive effect (0.069 for Type 1 customers and 0.097 for Type 2 customers) on updating the social status of the user. This indicates that a user would derive higher social status if she solves questions of other high status individuals than otherwise.

The costs of answering and asking questions differ based on the user type. Type 1 users incur higher costs of asking and answer questions compared to Type 2 users' costs. Type 2 users may be better at framing and articulating a question leading to lower cost of asking a question compared to Type 2 users. The cost of answering a question captures both the baseline cost of answering and the altruistic utility from helping others. Type 2 users could be deriving higher altruistic utility or could be better at framing and articulating their answers leading to lower baseline costs. The cost of answering questions also depends on the quality of the question. The coefficient before the additional cost of answering a high quality question (c_s^H) is negative and significant. This indicates that the net utility of answering a high quality question is higher compared with that from a non-high quality question. As we explained in previous section, the estimated c_s^H represents the net of additional cost minus utility of answering a high quality question. Thus our results suggest that users derive positive net additional utility from answering a high quality question. Tenure of the user on the forum significantly affects one's cost of asking a question (0.045) but not answering a question. A Type 1 customer with Tenure value of 10 would incur a cost of 2.863-0.0.045*10 for asking a question. In comparison, a Type 2 customer with Tenure value of 10 would incur a cost of 2.106-0.0.045*10 for asking a question.

[7] The magnitude of these two coefficients doesn't mean customers value social status a lot more than knowledge. One of the reasons that contributes to this difference is that the scale of knowledge level is much larger compared with that of social status. Even for one of the later robustness check where we use absolute social status, the scale of knowledge level is still much larger compared with that of absolute social status.

There is significant though very small amount of passive learning. The passive learning coefficient k_x is 0.005. This indicates that individuals primarily learn when peers answer their own questions, while the effect of passive learning on user knowledge update is minimal.[8]

Compared to Type 1, Type 2 customers derive significantly greater utility per unit of social status and knowledge. Further, Type 2 customers have lower cost of asking and answering questions than Type 1. Thus Type 2 customers are more incentivized to be engaged in customer support forum because they derive higher utility from their seeking and sharing activities on the platform.

It is important to notice that whenever an individual asks or answers a question, the additional utility from the knowledge and/or social status increments in the current period generally cannot compensate for the incurred cost. This sacrificial behavior can be justified when the individuals make participation decisions in anticipation of future reciprocal rewards, as we will discuss in the next subsection.

4.5 Dynamic and Interdependent Decision Making

We now report the equilibrium decision rules resulting from the users' dynamic interactions. To understand how customers share knowledge on this crowdsourced customer support platform, we focus on describing customer decisions of whose question gets solutions (Figure 3A) and whether to ask question given own social status level (Figure 3B).

Whose Question Gets Solved

[INSERT FIGURE 3A ABOUT HERE]

We estimate policy functions (probability of asking and answering) that vary with individual type and question quality. In Figure 3A we plot the weighted average (for individual type and question quality) probability of a user receiving a solution as a function of her social status. As the figure shows, the higher

[8] Because, the passive learning coefficient is sensitive to the passive learning attributed to an individual in our viewing simulation, lack of detailed viewing data may lead to potential bias. While we observe the number of views a solution received, we do not know who viewed it. If we have over attributed the viewing of answers to users in our sample then the estimated passive learning coefficient would be negatively biased. In contrast, if we have under attributed the viewing of answers to users in our sample then the passive learning coefficient would be positively biased.

the social status of the knowledge seeker, the more likely her question will receive a solution. By contrast, users with lower social status are less likely to get help from the community. One of the potential reasons for this result is our estimated positive and significant network position effect, which means that answering a question posted by a higher social status peer will increase an individual's *social status* more than answering a question asked by a lower social status peer.

Whether to Participate

[INSERT FIGURE 3B ABOUT HERE]

Figure 3B demonstrates how the probability of posting a question (weighted average over individual type and question quality) is driven by an individual's social status. It is interesting to observe that a user is more likely to ask a question when she has high social status. This finding can be explained by her anticipation of future reciprocal rewards from her peers. In particular, high status individuals are more likely to receive solutions to their questions. Hence, users are more likely to seek knowledge from the public forum when they have high social status in the community. This could be one of the most important drivers of the emergence of core-periphery knowledge sharing structure on the customer support platform. This result implies that customers with low social status will be discouraged from participating in online activities and become less active in online communities.

These results shed some light on the incentives for individual contribution to the community from a dynamic perspective. While previous literature on incentive of individual contribution focuses more on static reasons such as altruism, we show that there is another layer of incentives involving the dynamic interaction among all the users and the future payoffs reciprocated by the community. This observation is consistent with the concept of "reciprocal altruism" that is established in the social psychology literature. As Trivers (1971) states: "<u>Altruism</u>, defined as an act of helping someone else although incurring some cost for this act, could have evolved since it might be beneficial to incur this cost if there is a chance of being in a reverse situation where the person whom I helped before may perform an altruistic act towards me."

4.6 The Formation of the Core-periphery Structure and the Efficacy of Knowledge Sharing

We now explore the impact of core-periphery structure on the knowledge sharing of the forum, and how to better encourage users to share their knowledge on such platforms, one of the research questions listed in the introduction. The dynamics shown in Figure 3A and 3B suggest that individuals with high social status are more likely to ask questions because they are more likely to receive solutions to their questions from their peers on the platform. As a result, we observe that a few individuals in the community ask and answer a large number of questions on the platform, and generate most of the activities on the platform. Meanwhile, the sizeable portion of the community becomes *peripheral* members who are relatively inactive. This is consistent with Figure 1 we show in *Introduction*. The core-periphery structure leaves knowledge seekers with low social status positions in disadvantageous situations. Once a core appears, it reinforces itself through the pattern of future interactions among its participants. Over time, this decision process will result in a small inner core within which the users have the privilege of answering each other's questions, while the questions posted by users outside the circle are likely to receive much less attention.

To examine whether the formation of core-periphery affects knowledge accumulation, we compare the growth of knowledge for users who are within the core to that of those outside it in three different network structures with different degree of core-periphery structure. We start with the individual activity history by the end of our observation period. To be consistent with the estimation results in which 348 individuals are classified as Type 2 users, we denote the 348 individuals with highest social status as core members, and the remaining 1210 individuals as peripheral members. In the first network, we randomly add 500 directed ties in the community. In the second network, we randomly assign 500 directed ties only among core members. In the third network, we randomly assign 500 directed ties only among peripheral members. These three networks have different core-periphery degree (see Appendix B for details on the measurement). We simulate individual decision in the subsequent 20 periods based on our estimated OE strategy using these three networks respectively. We conduct this simulation 100 times, and calculate the expected knowledge levels both for individuals in core group as well as for those in peripheral group at the end of the 20 periods respectively.

[INSERT TABLE 4 ABOUT HERE]

As we can see from Table 4, the core-periphery structure is most salient for the second network, and it is least salient for the third network. Meanwhile, core members obtain the highest knowledge level in the second network in which the core-periphery structure is most salient, followed by the first and second network. On the other hand, peripheral members obtain the highest knowledge level in the third network where the core-periphery structure is least salient. That is to say, when the degree of core-periphery structure in the network becomes salient, the rate of knowledge increments is faster among the users who are within the core and is much slower among the rest. These results imply that the endogenously formed core impedes effective knowledge sharing within community by discouraging the peripheral members to participate. In practice, individuals who are of low social status are likely to be the newcomers who need more help. However, they are less likely to receive help from the community.

4.7 Alternate Policy Analyses

In the previous sections, we illustrate the dynamics of knowledge seeking and sharing decisions, and the formation of the network on the crowdsourced customer support platform. In this subsection, we conduct a series of alternate policy analyses which focus on improving the design of the platform. These analyses either take effect through changing exogenous variables, or by improving the mechanism of the knowledge sharing on the platform. To conduct these policy analyses we first change the appropriate parameter values and then solve for the equilibrium. In the first analysis, we investigate the how user's would respond if recent contributions are given even more weight when attributing social status. We perform this analysis by altering the discount factor of individual contribution level. In the second analysis, we examine how individuals would respond if we were to anonymize the identity of knowledge seeker. In the third analysis, we quantify the impact of monetary incentive on knowledge seeking. These sensitivity analyses further our understanding of the impact of different components on the knowledge sharing on crowdsourced customer support platform. They also shed some light on how the incentive for participants and design of platforms should be improved in the future. Table 5 shows the percentage change in the number of questions asked and answered, the percentage change in the community knowledge increments under alternative designs, and compares these statistics with those under the original design. These numbers are also reported separately for users in core and peripheral positions.

[INSERT TABLE 5 ABOUT HERE]

Change the decay of contribution level

In this sensitivity analysis, we change the discount factor of contribution level β_r in the social status updating rule. In reality, management of the platform can change the discount factor of contribution level by revising the design of the platform. For example, instead of showing the total number of answers provided by a user on her profile page, the platform can show the number of her answers in the past six month. This could decrease the discount factor, because contributions six month ago will not be counted toward individual social status level. In other words, when we decrease discount factor, individual contributions to the community will become obsolete faster. The impact of decreasing discount factor is of two folds. First, a smaller discount factor may discourage individuals from contributing to the forum. This is because their contribution will soon become outdated, and the incentive of contributing to the community decreases. Second, a smaller discount factor could also encourage individuals from contributing to the forum. This is because the gap of contribution levels between core and periphery members become smaller, thus peripheral users obtain more benefit from participating in the forum. As a result, we may see an increase in contribution from peripheral users. To evaluate the resulting impact of reduced discount factor on knowledge sharing (β_r), we conduct a sensitivity analysis where the discount factor of contribution level is set to 0.90. In other words, posted answers lose half of their influence on social status within around six weeks in this sensitivity analysis.

We can infer the impact of reducing the discount factor on knowledge sharing from comparing the statistics in Table 5. After we reduce the discount factor of social status from 0.95 to 0.9, users tend to contribute more to the community. The expected number of solutions provided increases from 0.058 to 0.063. As a result, this design change improves the knowledge sharing on this customer support platform, and the average community knowledge level increases from 8.132 to 8.940. Results of this policy analysis suggest that when contribution level is discounted fast, individuals are more likely to contribute. The degree of core-periphery structure also slightly decreases under the new policy.

Hide identity of knowledge seeker

We next conduct a policy analysis that requires the knowledge seekers to hide their identities while allowing the knowledge sharers to still build their status. More specifically, we assume that the change occurs at the end of our observation window. The existing social status and knowledge levels are preserved. After this time, however, whenever an individual asks a question, their identities are anonymized. In other words, we do not allow the social status of the knowledge seeker to influence users' decision on whose question to answer. In this setting, all of the users are still motivated to contribute and build their social status purely from number of solutions provided. Without knowing the source of the questions, however, the knowledge sharers cannot selectively answer the questions asked by high social status users. We simulate individual behavior for the subsequent 20 periods, starting from the last period in our dataset.

In Table 5, we observe that with this minor change to the design of the forum, users are more likely both to ask questions (from 0.080 to 0.092) and to answer questions (from 0.058 to 0.064) on average. These increases occur because without knowing the source of the question, users treat all peers and their questions equally. As a result, otherwise low social status users, who consist of the majority of the users on the forum, are more likely to obtain help from the public forum, thus they are more likely to seek knowledge as well as contribute to the community.

More importantly, we can see from Table 5 that the degree of core-periphery structure decreases from 0.0541 to 0.0490. The average knowledge accumulated for each individual at the end of the observation period increases from 8.132 to 8.670 on average. This suggests that this slight modification of the existing design encourages knowledge sharing within the community. We can further differentiate between the core and peripheral individuals in terms of the effect of the design change on their knowledge improvement from the last column. The peripheral members receive a 7.94% knowledge-increment improvement (from 6.815 to 7.356) while the privileged core individuals also receive improvement from 12.712 to 13.240, a 4.15% increase. In other words, anonymizing knowledge seeker's identity help knowledge gain for both core and periphery individuals and makes the core-periphery structure less salient.

Reward knowledge seeking behavior

The design for most of the popular online customer support forums emphasize on highlighting the ones who share most knowledge in the community. However, the impact of rewarding asking of questions is not clear both for researchers and practitioners. In this sensitivity analysis, we reward knowledge seeking behavior by giving reward that equals 10% of the baseline cost of asking questions ($c_{a,0}$) when individuals ask a question instead. By giving individuals extra incentive for asking questions, more questions will be posted on the platform and the community will also see more answers. The higher number of answers further leads to more questions in the next period. As a result of this "ripple effect", more knowledge can be shared among members of the community, and the knowledge level of the community will increase. Intuitively, because a higher knowledge level could increase the probability of answering questions, a higher knowledge level will in turn reinforce the knowledge sharing within the community.[9] Our sensitivity analysis indicates that by giving reward that equals 10% of the baseline cost of asking questions ($c_{a,0}$), the probability of asking question increases from 0.080 to 0.098, the expected number of solutions provided

[9] Here, we assume that the quality level of answers and the probability of viewing answers will be exogenous and will not change because of the new policy. To consider the impact of the new policy on quality and viewing of answers will require us to explicitly model individual decision of providing answers with different quality, as well as decision of viewing each answer. We leave this extension for future research.

for each user increases from 0.058 to 0.071, and the platform knowledge level increases by 23.36%, a much larger improvement compared with other two policy changes.

4.7 Robustness Checks

We conducted a number of robustness checks to see how our results vary when we relax some of our assumptions and to check for alternate explanations. First, it has been shown that discount factor estimation is quite problematic and requires unique (rare, present in very few contexts) exclusion restrictions where some state variable that affect state transitions do not enter flow utility (Magnac and Thesmar 2002). Our data does not allow any such exclusion restriction. Hence, we cannot estimate the discount factor. Thus we follow the traditional literature by assuming the value of the discount factors. We report the results in Table 6. The results remain qualitatively the same when we vary the value of the discount factors. Second, we have fixed the values for discount factor for status depreciation and knowledge depreciation at 0.95. To test the robustness of our results to these discount factors we vary the values of these discount factors and report the results of the estimated model in Table 6. As the results indicate our results remain qualitatively the same. Third, we used a relative measure for social status. It is possible that individuals may only care about their own social status value rather than in relation to others. We conducted an analysis where we replaced the relative social status with absolute social status. While the parameters for social status in the utility function change, all the key findings remain consistent. Fourth, our model accounts for reciprocity at community level. That is, when a user contributes more to the community, she is able to receive more help from the community. Another type of reciprocity, i.e. dyad level reciprocity, cannot be directly accounted for in our model. This dyad level reciprocity means that user A responds to user B's question if B has answered A's question in the past. Our data suggests that among all answers posted on the platform only 2.3% of them are such reciprocal answers. We estimate a model where we remove all such reciprocal answers from the data. The results are provided in Table 6 and show that our main results stay qualitatively similar. Fifth, our dataset suggests that among all the observation in our dataset, less than 0.3% of the time users ask more than one question in one period. Thus it is reasonable to simplify our model to a binary discrete choice model. Nonetheless, we conduct a robustness check where we employ an ordered logit model to capture the instances where multiple questions were asked. Our main results remain the same.

[INSERT TABLE 6 ABOUT HERE]

One potential explanation in any network formation is homophily. We argue that the core-periphery structure cannot be explained by pure homophily because while it can provide a reason why core connects with core, it cannot explain why core also connects with periphery and why periphery does not connect with other periphery. Further, even in core-core relationships it cannot explain why even core individuals would prefer to connect with high status individuals even in the core. If homophily were the primary driver for core-core relationship, our results about network position effects should not hold when we consider relationship formation among only core individuals. To check this, we conduct a robustness test. Specifically, in this robustness check, we model only the asking and answering decisions of the top 348 individuals based on their social status. We select 348 individuals because this is consistent with the 348 individuals who are classified as Type 2 users. We apply the estimation model without unobserved heterogeneity on this sub dataset, because almost all of them are Type 2 individuals. We report the results of this estimation in Table 6. The results from this estimation are quite close to results for Type 2 users in Table 3. The network position effect is positive and significant in this analysis indicating that even when individuals decide which core individual's question to answer they are more likely to pick the one with higher social status.

[INSERT TABLE 7 ABOUT HERE]

Finally, ideally one would want to model individual knowledge sharing as a sequential process where users visit the forum at different times and can answer only those questions which haven't received solutions yet. However, there are two difficulties. First, it would significantly increase the computational burden for model estimation. In particular, in our dynamic structural model setting, we would need to model individual's timing of answering questions, as well as how they predict others' timing of answering questions. More importantly, we would need to know whether and when individuals browse each one of the threads to know how many answers are already provided for the question. In the absence of such information, we performed one analysis which may shed light on this issue. We know when a question is posted and when a solution is posted. We use this information to construct a distribution of whether a question would be unanswered when a user with certain state variable values visits the forum. We incorporate this in our model. We assume that the visit timing is exogenous. As a result as time goes on a user may have fewer unanswered questions available which he/she could answer. We estimate this model. The only key difference from our main results is that under this model, the cost of answering a question goes down significantly. None of the other parameter estimates change significantly. Hence, we can interpret the simultaneous decision model in our framework as the aggregation of a sequential decision model for each thread.

5. Conclusions

Our model provides a framework for managers to analyze the factors that influence user contribution level and to evaluate their designs for customer support forums. As many firms are adopting crowdsourced customer support platforms for knowledge sharing, idea generation, project management, customer service, and identifying sales and marketing opportunities, it is important to understand the fundamental drivers of user behavior to increase the return on investment (ROI). Understanding the dynamics behind the individual participation decisions becomes even more critical with the fast adoption of social CRMs. Gartner predicted that spending on social CRM is expected to exceed $1 billion in 2013 which is approximately 8% of all the CRM spending in that year. However, companies also face significant challenge in developing online communities for social CRM. In another report from Gartner in 2008, more than half of the companies which had established an online community for customer support fail in two years. Thus it is very important for both practitioners and researcher to understand how to improve the design of these platforms.

We recognized the dynamic and interdependent decision-making process and built a dynamic structural model to investigate the users' knowledge-seeking and knowledge-sharing decisions. Applying the model to a customer support forum, we found the following results. (1) Knowledge seeking and sharing on the platform is driven by the knowledge and social status of both the users themselves and their peers in the community. We showed that sharing knowledge with peers can be better explained by dynamic, interactive decision making in anticipation of future reciprocal rewards from the community. This result was supported by our findings that users are more likely to share to help peers higher social status, (2) The formation of the core-periphery structure results from the strategic interactions described above. The users strategically choose to answer the questions asked by the users in the core to improve their social status and hence to obtain greater future reciprocal rewards. (3) Users located within the core have the advantage of obtaining help from each other and meanwhile exclude other users from participating. Thus, the "free-riding" behavior of the peripheral contributors may be an equilibrium result because the existence of a core discourages peripheral users from actively asking questions on the platform. (4) Exploratory sensitivity analysis found that breaking the core-periphery structure by hiding the knowledge seeker's identity can improve knowledge sharing; offering slightly more weight to recent contribution encourages contribution; and awarding knowledge seeking is very effective in accumulating knowledge on the platform.

Our manuscript provides a framework that examines the fundamental mechanism of individual behaviors in the context of knowledge sharing forum. This framework focuses on the strategic interactions among customers on the forum, because almost all of the contents are generated through interactions among users. This characteristic is also shared among other social media platforms. Thus this framework on the dynamic individual decision process can be extended to other social media contexts, such as blogs, social CRM, targeting online influential customers, crowdsourcing initiatives and etc.

Our research has some limitations, which open exciting avenues for future research. First, we did not incorporate effort as part of individual decision process. We did not consider the possibility that individual may invest more effort so that his answer is more likely to be selected as a solution. Second, many new features such as subscription of discussion threads and sharing posts with friends, have been gradually introduced to customer support platforms. Future research can incorporate these features into the model and examine how they modify the main findings. Third, platforms have different designed in terms of social status. While social status features have been widely adopted among social media platforms, there are platforms without this feature nonetheless. Our framework can be extended to this scenario by setting parameters for social status to zero. In addition, there are also platforms in which social status measures are discrete. We can extend our model by changing social status updating rule accordingly. Fourth, there are data limitations in our work. While we have the accurate information about views that a post received, we do not have information on who viewed the post. As a result, the passive learning estimated by us could be biased. In reality, individuals would not make simultaneous answering decisions. Individuals may arrive sequentially and attempt to answer questions which are open at that time. Further, individuals may also consider other answers which have already been provided to a question before attempting to answer a question. We do not have access to data that can allow us to model this process. Future research should consider modeling user decision making in a sequential manner. Last, we treat the quality of the question to be exogenous and do not treat quality of the question as a decision variable for a user when asking a question. We believe a whenever a user is facing a problem with the product or service he/she would consider posting the question irrespective of how others may find it useful.

In this study, we showed how efficiency of knowledge sharing will change given different policy designs. The corresponding managerial implications can be widely applied to a wide spectrum of customer support platforms. However, there are also a few cases where the results could have different managerial implications for firms with different objectives for their platforms. For instance, for a firm who employs discussion forum with the purpose of identifying experts from customers to improve their product, having more core members is more important compared with knowledge level of the community. While our framework can be applied in these scenarios, we should interpret the results with caution.

References

Adamic, Lada A., Jun Zhang, Eytan Bakshy and Mark Ackerman. 2008. "Knowledge Sharing and Yahoo Answers: Everyone Knows Something". *Proceedings of the 17th International Conference on World Wide Web*, pp. 665-674, Bejing, China.

Aguirregabiria, Victor and Pedro Mira. 2007. "Sequential Estimation of Dynamic Discrete Games". *Econometrica* Vol. 75, No. 1, pp 1-53.

Aral, Sinan and Dylan Walker. 2011. "Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks". Management Science. Vol. 57, No. 9, pp. 1623-1639.

Arcidiacono, Peter and Robert Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity". *Econometrica*. Vol. 79, No. 6. Pp. 1823-1867.

Argote, Linda, Bill McEvily and Ray Reagans. 2003. "Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes". *Management Science*. Vol. 49, No. 4, pp. 571-582.

Banos, R. A., Borge-Holthoefer, J., Wang, N., Moreno, Y. & González-Bailón, S. (2013) Diffusion dynamics with changing network composition. Preprint, http://arxiv.org/abs/1308.1257

Beatty, Sharon and Salil Talpade. 1994. "Adolescent influence in Family Decision Making: A Replication with Extension". *Journal of Consumer Research*. Vol. 21, No. 2, pp. 332-341.

Benkard, Lanier. 2004. "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft". *Review of Economic Studies*. Vol 71, No. 3, pp 581-611.

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures". *The American Journal of Sociology*. Vol. 92, No. 5, pp. 1170-1182.

Borgatti, Stephen and Martin Everett. 2000. "Models of Core-periphery Structures". *Social Networks*. Vol. 21, No. 4, pp. 375-395.

Boulding, Kenneth. 1966. "The Economics of Knowledge and the Knowledge of Economics". *The American Economic Review*. Vol. 56, No. ½, pp. 1-13.

Chan, Tat Y., Jia Li and Lamar Pierce. 2013. "Learning from Peers: Knowledge Transfer and Sales Force Productivity Growth". Working Paper. Washington University in St. Louis.

Chung, Doug, Thomas Steenburgh and K. Sudir. 2014. "Do Bonuses Enhance Sales Productivity? A Dynamic Structural Analysis of Bonus-Based Compensation Plans". *Marketing Science*, Vol. 33, No. 2, pp. 165-187.

Erdem, Tulin and Michael Keane. 1996. "Decision Making Under Uncertainty: Capturing Dynamic Brand Choice Process in Turbulent Consumer Good Market". *Marketing Science*. Vol. 15, No. 1, pp. 1-20.

Erdem, Tulin, Michael Keane, T. Sabri Öncü, and Judi Strebel. 2005. "Learning About Computers: An Analysis of Information Search and Technology Choice". *Quantitative Marketing and Economics*. Vol. 3, No. 3, pp. 207-247.

Erdem, Tulin, Michael Keane and Baohong Sun. 2008. "A Dynamic Model of Brand Choice when Price and Advertising Signal Product Quality". *Marketing Science*, 27(6), 1111-1129.

Ericson, Richard and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work". *The Review of Economic Studies*. Vol. 62, No. 1, pp. 53-82.

Frank, Robert. 1985. "The demand for unobservable and other nonpositional goods". The American Economic Review. .Vol. 75, No. 1, pp. 101-116.

French, John and Bertram Raven. 1959. "The Bases of Social Power". *Classics of Organization Theory,* by Jay Shafritz, J. Ott and Yong Jang, Cengage Learning.

Fisher, Danyel, Marc Smith and Howard T. Welser. 2006. "You Are Who You Talk To: Detecting Roles in Usenet Newsgroups". *Proceedings of the 39th Hawaii International Conference on System Sciences*. Vol. 3.

Gartner. 2011. "Gartner Says Spending on Social Software to Support Sales, Marketing and Customer Service Processes Will Exceed $1 Billion Worldwide By 2013". http:// http://www.gartner.com/-it/page.jsp?id=1541415

Hartmann, Wesley. 2010. "Demand Estimation with Social Interactions and the Implications for Targeted Marketing". *Marketing Science*. Vol. 29, No. 4, pp. 585-601.

Hewlett Packard. 2012. "Engaging and Energizing Experts for Superior Support". http://groundswell-discussion.com/groundswell/awards2012/detail.php?id=875.

Huang, Yan, Param Singh and Anindya Ghose. 2013. "A Structural Model of Employee Behavioral Dynamics in Enterprise Social Media". Forthcoming at *Management Science*.

Iyengar, Raghuram, Christopher Van den Bulte and Thomas Valente. 2011. "Opinion Leadership and Social Contagion in New Product Diffusion". *Marketing Science*. Vol. 30, No.2, 195-212.

Katona, Zsolt and Miklos Sarvary. 2008. "Network Formation and the Structure of the Commercial World Wide Web". *Marketing Science*. Vol. 27, No. 5, pp. 764-778.

Kumar, Vineet, Ramayya Krishnan, and Baohong Sun, 2010. "Measuring Dynamic Effect of Promotion through Social Network." Working paper, Carnegie Mellon University.

Lakhani, Karim and Eric von Hippel. 2003. "How Open Source Software Works: "Free" User-to-User Assistance". *Research Policy*. Vol. 32, No. 6, pp. 923-943.

Lampel, Joseph and Ajay Bhalla. 2007. "The Role of Status Seeking in Online Communities: Giving the Gift of Experience". *Journal of Computer Mediated Communication*. Vol. 12, No. 2, pp. 434-455.

Li, Charlene and Josh Bernoff. 2009. "Helping the Groundswell Support Itself: Customers Supporting Customers Through Social Technologies". Harvard Business School Press.

Lu, Yingda, Kinshuk Jerath and Param Singh. 2013. "The Emergence of Opinion Leaders in a Networked Online Community: A Dynamic Model with Time Dynamics and a Heuristic for Fast Estimation". *Management Science*. Vol. 59, No. 8, pp. 1783-1799.

Lurie, Nicholas, Hai Che and Allen Weiss. 2009. "Helping Strangers: Who Contributes to Online Communities, How Much Do They Contribute, and Why". Working Paper. Georgia Institute of Technology.

Ma, Liye, Kannan Srinivasan, Baohong Sun. 2010. "A Dynamic Competitive Analysis of Content Production and Link Formation of Internet Content Developers". Working Paper. Carnegie Mellon University.

Magnac, Thierry and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes". Econometrica. Vol. 70, No. 2, pp. 801-816.

Mayzlin, D. and H. Yoganarasimhan, (2012), "Link to Success: How Blogs Build an Audience by Promoting Rivals". *Management Science*. Vol. 58, No. 9.

McAfee, AP. 2006. "Enterprise 2.0: The Dawn of Emergent Collaboration". *MIT Sloan Management Review*.

McPherson, Miller, Lynn Smith-Lovin and James Cook. 2001. "Birds of a Feather: Homophily in Social Networks". *Annual Review of Sociology*. Vol. 27, pp. 415-444.

Nam, Kevin, Mark Ackerman and Lada Adamic. 2009. "Questions in, Knowledge iN? A Study of Naver's Question Answering Community". CHI 2009, Boston

Narayanan, Sridhar and Puneet Manchanda. 2009. "Heterogeneous Learning and the Targeting of Marketing Communication for New Products*". Marketing Science*. Vol. 28, No. 3, pp. 424-441.

Obradovic, Darko and Stephan Baumann. 2009. "A Journey to the Core of the Blogsphere". *Advances in Social Network and Mining*. July, pp. 1-6.

Pakes, Ariel and Paul McGuire. 1994. "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model". *The RAND Journal of Economics*. Vol. 25, No. 4, pp. 555-589.

Ryan, Stephen and Catherine Tucker. 2012. "Heterogeneity and Dynamics of Technology Adoption". *Quantitative Marketing and Economics*. Vol. 10, No. 1, pp. 63-109

Ryu, Chungsuk, Yong Jin Kim, Abhijit Chaudhury and H. Raghav Rao. 2005. "Knowledge Acquisition via Three Learning Process in Enterprise Information Portals: Learning-by-Investment, Learning-by-Doing, and Learning-from-Others". *MIS Quarterly*. Vol. 29, No. 2, pp. 245-278.

Singh, Param and Yong Tan. 2010. "Developer Heterogeneity and Formation of Communication Networks in Open Source Software Projects." *Journal of Management Information System*. Vol. 27, No. 3, pp. 179-210.

Singh, Param, Yong Tan and Vijay Mookerjee. 2011. "Network Effects: The Influence of Structural Social on Open Source Project Success". *Management of Information System Quarterly*. Vol. 35, No. 4, pp. 813-829

Stephen, Andrew and Olivier Toubia. 2010. "Deriving Value from Social Commerce Networks". *Journal of Marketing Research*. Vol. 47, No. 2, pp. 215-228.

Trivers, Robert. 1971. "The Evolution of Reciprocal Altruism". *The Quarterly Review of Biology*. Vol 46, No. 1,pp 35-57.

Van den Bulte, Christophe and Gary Lilien. 2001. "Medical Innovation Revisited: Social Contagion versus Marketing Effort". *American Journal of Sociology*. Vol. 106, No. 5, pp. 1409-1435.

Wasko, Molly and Faraj Samer. 2005. "Why Should I Share? Examine Social and Knowledge Contribution in Electronic Networks of Practice". *MIS Quarterly*. Vol. 29, No. 1, pp. 35-57.

Weintraub, Gabriel, Lanier Benkard and Benjamin Van Roy. 2008. "Markov Perfect Industry Dynamics with Many Firms". *Econometrica*. Vol. 76, No. 6, pp. 1375-1411.

H. T. Welser, E. Gleave, D. Fisher, and M. Smith. 2007. "Visualizing the Signatures of Social Roles in Online Discussion Groups". *Journal of Social Structure*, Vol. 8. No. 2, pp. 564-586.

Zhang, Juanjuan. 2010. "The Sound of Silence: Observational Learning in the U.S. Kidney Market". *Marketing Science*. Vol. 29, No. 2, pp. 315-335.

Zhang, Jun, Mark Ackerman and Lada Adamic. 2007. "Expertise Networks in Online Communities: Structure and Algorithms". *Proceedings of the 16th International Conference on World Wide Web*, pp. 221-230. New York, USA.

Table 1. An Example of Knowledge Seeking and Sharing in Customer Support Forum*
Topic: WIFI issues

Post Type	Author	Time	Content
Question	**A**	**13/04/2013**	I have got a new sony experia E. I have enter the APN settings and I have connected to 3G but I can't connect to wifi every time I try to it

		9:39 PM	says 'connecting' and then it just says saved, secured with WAP and it keeps doing that every time I try connecting! HEELPP! please!
Answers:	**B**	**13/04/2013 9:48 PM**	This is a quick suggestion based on my experience of the screens on budget Experias; enable the tickbox that shows the password as you type instead of stars. It may be that it's triggering some letters twice so the password is saved but incorrect. Once it's right it should join immediately.
	C	**13/04/2013 9:49 PM**	Are you entering your router or the wifi hotspot password in correctly? You should also note that if you are connected to 3G (mobile internet) and wifi, it will try to connect to 3G automatically unless you change this in options on android (had to do this for my Galaxy S3)
	D	**13/04/2013 10:01 PM**	Do you have any sort of security set up on your router such as MAC address filtering, or a device access list that might be refusing access to your phone? Can you also confirm that the password is correct by reconnecting another device?
	E	**13/04/2013 10:50 PM**	The wifi settings on your phone have nothing whatsoever to do with the SIM card...You are either trying to connect to your home broadband router, or a public wifi hotspot and the message you're seeing (which is actually "Secured with **WEP"**(*Wired Equivalent Privacy*)), and this is possibly why your device will not connect. WEP is not secure and some new devices may refuse to connect to it. Remove the network from your saved network list first, then re-add it but tick the box to show password as you type it, to ensure that you are entering it correctly and not confusing number 0 with a O or number 1 with letters I or lower case l. WEP passphrases use the numbers 0-9 and letters A-F only. You might also check the settings in the router, in case MAC filtering has been enabled. If set, only devices on an approved list can connect to the network.

*These asking and answering posts source from a thread in a real online discussion forum.

Table 2. Data Description

Variables	Statistics
Number of Users	1558
Number of Periods	30
Total Number of Questions	3748
Total Number of Solutions	2684
Average Number of Questions Asked per User	2.41
Average Number of Solutions per User	1.73
Mean of Tenure (Number of Weeks Registered Before Period 0)	25.08
Average Number of Views per Post	130.59

Table 3. Parameter Estimation

Variable	
Utility Function Parameters for Type 1 Customers	
Impact from Knowledge (α_1)	0.042***
Impact from Social Status(α_2)	2.927***
Network Position Effect on Social Status (γ_1)	0.069***
Constant for Cost of Asking a Question ($c_{a,0}$)	2.863***
Constant for Cost of Answering a Question ($c_{s,0}$)	18.975***

Percentage of Customer in this Type	77.66%
Utility Function Parameters for Type 2 Customers	
Impact from Knowledge (α_1)	0.070***
Impact from Social Status(α_2)	4.040***
Network Position effect on Social Status (γ_1)	0.097***
Constant for Cost of Asking a Question ($c_{a,0}$)	2.106***
Constant for Cost of Answering a Question ($c_{s,0}$)	11.852***
Percentage of Customer in this Type	22.34%
Cost Function Parameters	
Impact of Tenure on Cost of Asking Question ($c_{a,1}$)	0.045***
Impact of Tenure on Cost of Answering Question ($c_{s,1}$)	0.014
Other Parameters	
Effect of Knowledge Spill Over (k_2)	0.005 [a]
Additional Cost of Answering a High Quality Question (c_s^H)	-2.371***

*** The 99% confidence interval does not include zero.
** The 95% confidence interval does not include zero.
* The 90% confidence interval does not include zero.
[a] The 90% confidence interval does not include one. We normalize $k_1 = 1$ for identification purpose.

Table 4. The Formation of Core-periphery Structure and Knowledge Level

	Core-periphery Degree[b]	Average Knowledge Level for Core Group Members[a]	Average Knowledge Level for Peripheral Group Members[a]
Network 1	0.0509	13.119	6.655
Network 2	0.0641	13.418	6.413
Network 3	0.0484	12.482	7.818

a The knowledge level reported here is measured at the end of the simulation periods.
b.Here, core individuals are selected as a cohesive group of 348 individuals who closely communicate with each other, and periphery individuals are the remaining ones who may loosely connected with someone in the core group, but have rare connection with other periphery ones (Borgatti and Everett, 2000).

Table 5. Sensitivity Analysis

	Probability of Asking Questions			Expected Number of Solutions Provided			Degree of Core/ Periphery	Average Community Knowledge		
	Core	Periphery	Total	Core	Periphery	Total		Core	Periphery	Total
Benchmark	0.198	0.046	0.080	0.143	0.034	0.058	0.0541	12.712	6.815	8.132
Faster Decay	0.217	0.053	0.090	0.153	0.037	0.063	0.0521	15.084	7.173	8.940
Anonymity	0.192	0.063	0.092	0.139	0.039	0.061	0.0490	13.240	7.356	8.670
Reward Qs	0.234	0.059	0.098	0.161	0.044	0.071	0.0536	17.173	7.978	10.032

Table 6. Robustness Tests

Variable	Absolute Social Status	$\beta_k = 0.95$ $\beta_r = 0.85$	$\beta_k = 0.85$ $\beta_r = 0.95$	$\beta = 0.8$	Non-Reciprocal Answers
Utility Function Parameters for Type 1 Customers					
Impact from Knowledge (α_1)	0.045***	0.048***	0.101***	0.053***	0.039***
Impact from Social Status(α_2)	0.241***	2.840***	3.136***	3.457***	2.793***
Network Position Effect on Social Status (γ_1)	0.064**	0.064**	0.072**	0.075**	0.065**
Constant for Cost of Asking a Question ($c_{a,0}$)	3.091***	3.135***	3.167***	2.540***	2.931***
Constant for Cost of Answering a Question ($c_{s,0}$)	18.922***	19.395***	19.166***	18.053***	19.257***
Percentage of Customer in this Type	76.96%	78.18%	77.09%	78.95%	78.37%
Utility Function Parameters for Type 2 Customers					
Impact from Knowledge (α_1)	0.074***	0.069***	0.159***	0.078***	0.065***
Impact from Social Status(α_2)	0.337***	3.922***	4.197***	4.469***	3.884***
Network Position Effect on Social Status (γ_1)	0.093***	0.93***	0.110***	0.103***	0.094***
Constant for Cost of Asking a Question ($c_{a,0}$)	2.214***	2.382***	2.235***	1.931***	2.029***
Constant for Cost of Answering a Question ($c_{s,0}$)	11.768***	12.765***	12.238***	11.528***	12.135***
Percentage of Customer in this Type	23.04%	21.82%	22.91%	21.05%	21.63%
Cost Function Parameters					
Impact of Tenure on Cost of Asking Question ($c_{a,1}$)	0.046**	0.052***	0.037**	0.040***	0.051***
Impact of Tenure on Cost of Answering Question ($c_{s,1}$)	0.013	0.016	0.013	0.017	0.015
Other Parameters					
Effect of Knowledge Spill Over (k_2)	0.004 [a]	0.005 [a]	0.004 [a]	0.005 [a]	0.005 [a]
Addl. Cost of Answering a High Quality Question (c_s^H)	-2.360***	-2.898***	-2.650***	-2.063***	-2.516***

*** The 99% confidence interval does not include zero.
** The 95% confidence interval does not include zero.
* The 90% confidence interval does not include zero.
[a] The 90% confidence interval does not include one. We normalize $k_1 = 1$ for identification purpose.

Table 7. Model Estimation with Only Core Members.

Variable	
Utility Function Parameters for Core Customers	
Impact from Knowledge (α_1)	0.078***
Impact from Social Status(α_2)	3.952***
Network Position Effects on Social Status (γ_1)	0.095***
Constant for Cost of Asking a Question ($c_{a,0}$)	2.313***
Constant for Cost of Answering a Question ($c_{s,0}$)	11.631***
Cost Function Parameters	
Impact of Tenure on Cost of Asking Question ($c_{a,1}$)	0.049***
Impact of Tenure on Cost of Answering Question ($c_{s,1}$)	0.013
Other Parameters	

Effect of Knowledge Spill Over (k_2)	0.006 [a]
Additional Cost of Answering a High Quality Question (c_s^H)	-2.460***

*** The 99% confidence interval does not include zero.
** The 95% confidence interval does not include zero.
* The 90% confidence interval does not include zero.
[a] The 90% confidence interval does not include one. We normalize $k_1 = 1$ for identification purpose.

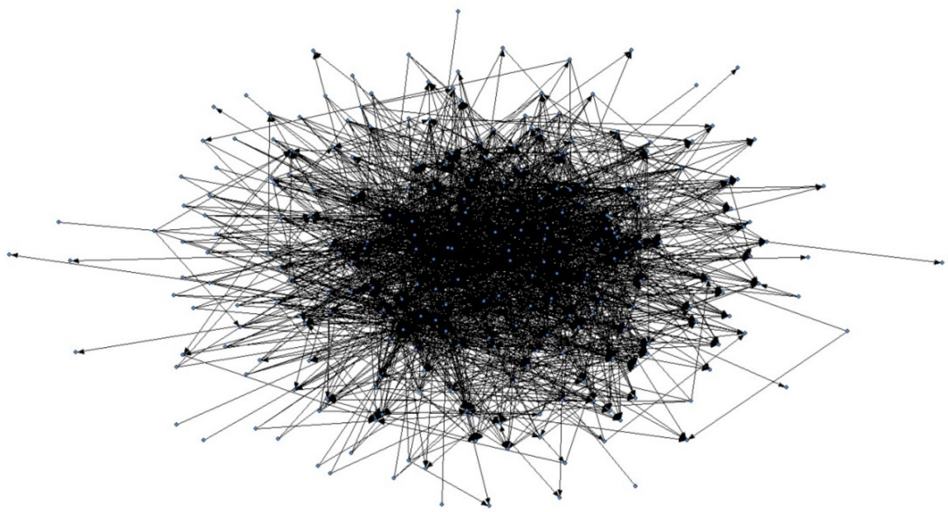

Figure 1. The Core/Peripheral Network Structure[a]

a. Individuals are represented by spheres. Lines connecting two individuals represent the presence of a knowledge-sharing relationship between them. The arrow heads point towards the individual who answers the question. More active participants are indicated by larger spheres.

Figure2. An Illustrative Schema of the Customer Support Forum

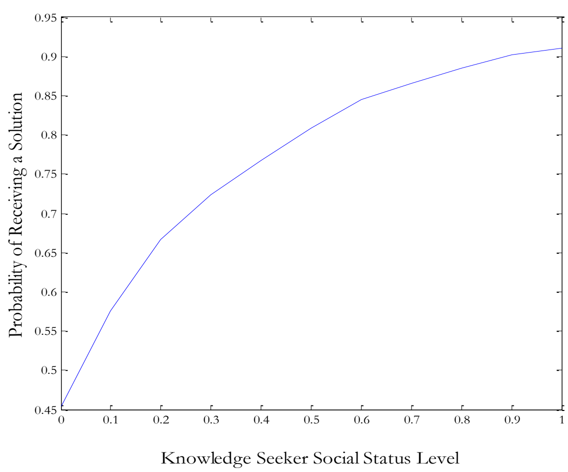

Figure 3A. Probability of Receiving a Solution to a Question Given Knowledge Seekers' Social Status Level

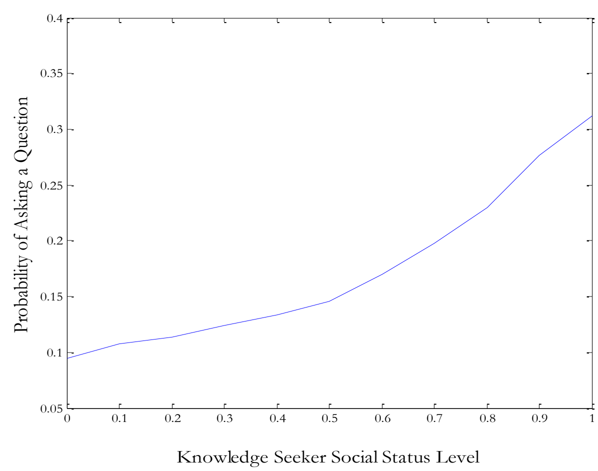

Figure 3B. Probability of Posting a Question given Own Social Status Level

Figure 3. Individual Optimal Decision Rules