

# Research on the Relationship Network in Customer Innovation Community based on Text Mining and Social Network Analysis

Aiqin ZHOU, Yusen ZHOU

**Abstract:** Relationship is the focus of the current study in the social phenomenon with social network theory, which is mainly about its meaning and strength. However, a different object, different relationship. Social network theory insists that the actor's behavior is the result of the limitations and opportunities of many relationships that occur simultaneously and interaction. The behavior and characteristics of the whole group are also dependent on the integration of multi-dimensional relationships. There are multi-dimensional relationships among customers participated product innovation in the customer innovation community. Since the huge number of customers in customer innovation community and the complex relationships among the customers, the method is different in traditional ways. Therefore, this paper combines associated crawler algorithm, text mining, and social network analysis to study network relationship types, network structure and the relevance of the customer innovation community. Firstly, this paper analyzes the relationship type and the relationship network according to previous studies. Secondly, reptile technology is used to obtain structured data in the customer community. After cleaning and pre-processing, the data is transformed into relational data from the original structure, with format  $1069 \times 1069$  size matrix. Analyzing the structure of relationship network using social network analysis methods and tools, the results show that interactive network, social network, and knowledge-sharing networks are all sparse network. Thirdly, the correlation among the relationship networks is studied. The results demonstrate that it is higher than the correlation between the interactive network and the knowledge-sharing network and lower than the social network correlated with the other two networks.

**Keywords:** associated crawler algorithm; customer innovation community; relationship network; social network analysis; text mining

## 1 INTRODUCTION

With the rapid development of the Internet, customers continue to improve their ability to obtain information, their rights and status are gradually increased. This phenomenon leads to the rapid development of product innovation community with customers as the main participants. On the one hand, a user is not satisfied with the unilateral acceptance of the products provided by the producer. They want to put forward their own unique needs of the product, participate in product design and improvement process. On the other hand, as producers, they welcome customers to participate in the creative design and continuous improvement process to obtain the accurate needs of customers and improve customers' loyalty to the product. The development of the Internet makes this possible. Product innovation community has gradually become the main platform for customers to participate in product innovation online. The way in which customers participate changes from physical participation to online participation. This new way receives a wide range of customer support and develops fast. It arouses great concern among the scholars and the producers.

The existing study on customer innovation community mostly focuses on customer behavioral characteristics at the micro-level but lacks holistic research at the macro-level. Now, the social network analysis method put all customers into the network, doing the study at the macro-level. Social networks focus on the interaction and relationships between people. Because interaction affects behavior. A pair of actors may keep one relationship, such as employer and employee relationship. They may keep many kinds of relationships, such as emotional relationship, knowledge-sharing relationship, and so on. The network is a collection of relationships. It is used to describe the relationship or the connection pattern. The social network refers to social members and their collections. In the customer innovation community, the most basic relationship between members is interaction relationship. Through interaction, members

participate in product innovation, then establish other relationships. As the frequency and intensity of interaction increases, members become familiar. Members with the same knowledge background and interests establish interpersonal relationships, that is, social relations. If the interactive topic is knowledge-related, the relationship that members establish is a knowledge-sharing relationship. This paper will study the interaction, social, and knowledge-sharing relationships that customers generate in the innovation network in the process of product innovation.

This paper is divided into seven chapters. Chapter 1 analyzes the background of the customer innovation community and its importance for relationship and relationship study. Chapter 2 studies and summarizes the existing literature, and provides a theoretical basis for the follow-up study. Chapter 3 analyzes the relationship type and the relationship network in the customer innovation community and puts forward three assumptions. Chapter 4 uses the crawler algorithm to capture customer interaction, social and knowledge sharing data, and builds a matrix by applying text mining and co-occurrence algorithms. Chapter 5 analyzes the network structure and characteristics of different relationship types. Chapter 6 analyzes the relevance of the three relational networks. Chapter 7 summarizes and discusses the previous studies.

## 2 LITERATURE REVIEW

Freeman [1] proposed the concept of innovation networks in 1991; he believes that the network of innovators is equivalent to innovation networks. Rheingold [2] believes that the virtual community is a new type of social organization, with four characteristics: (1) Express freedom of the organization members; (2) Lack of centralized control to the members; (3) Many-to-many transmission; (4) Autonomous interaction between members. Customer innovation community is an information technology-based innovation support service

platform. In the process of innovation, company forms information sharing and R & D cooperation relationship, trust and mutual benefit relationship. This connects the members participating in the open innovation into a complex network system. This network system contains a variety of knowledge elements and information transmission circuit, forming a large number of linear and non-linear action processes (Xia Enjun [3]). The customer innovation community this paper studies is product innovation that is based on customer innovation as the innovation subject. Customers participate in product innovation activities in innovative communities. In this process, customers generate interaction and communication and establish a variety of relationships. Many scholars have conducted a profound study of these relationships. Johnson K. et al. [4] believes that knowledge-sharing is a matter of respect and understanding among community members. Many researchers define knowledge-sharing as knowledge-sharing behavior. Davenport [5] defines knowledge-sharing as conscious exchange behavior. Bartol & Srivastava [6] define knowledge-sharing as sharing of organizational information, ideas, advice and expertise among individuals, the sharing knowledge including explicit knowledge and tacit knowledge. Koh & Kim [7] quantify the number of knowledge-sharing by analyzing the content exchanged by community members. They quantify the knowledge-sharing through the log records on the server. Chun and Kwak [8] compare the similarities and differences between the activity network and friends, network in the online community. They study whether the interaction relationship between members is similar to their friend's relationship and whether members' communication is a continuation of their friend's relationship. The study found the members' interactive networks are entitled to value and direction. The network topology is similar to friends network structure. Lampel & Bhalla [9] argue that altruism and reciprocity are the causes of knowledge-sharing behavior and that sharing behavior is affected by individuals' identity sense to the group. Wesley Shu et al. [10] believe that payback expectation, self-esteem will affect individual's identity sense to the social network, and further affect individual's knowledge-sharing willingness and knowledge-sharing behavior in the social network. Yang Chen [11] argues that, in the scientific research field, to find the right scholars to cooperate is conducive to knowledge discovery and exchange. In addition, scholars recommendation is mainly based on the similarity of professional knowledge and social network proximity.

From the literature, we can find that in the innovation community, there are many kinds of relationships between customers. However, previous scholars' study on customer relationships in innovative communities is limited to single kind of relationship. Therefore, this article will study a variety of customer relationships, and analyze the relevance of different relationships.

### 3 RELATIONSHIP TYPE AND CORRELATION ANALYSIS

#### 3.1 Interactive Relationship and Interactive Network

Interaction is the basis of customer innovation community. Interactions produce knowledge exchanges, which form the knowledge resources that all customers can share. At the same time, interaction makes unfamiliar

members familiar, and then establishes stable interpersonal relationships. Therefore, the study of interaction relationship between the members is one of the most important contents to study members' participation in product innovation activities in the online community.

In the customer innovation community, users can post a post to explain any issues that arise during the use of the product and suggest new features. Other members reply to the post, thus establishing a relationship between the members, that is, interaction relationship. All members of the community form the network of interpersonal relationships in the online community. In the customer innovation community, the interaction frequency is often used to measure the relationship between members. The more exchanges, the greater the willingness of members to interact.

The nodes in the network represent the posters and replies in the network. The edge represents the relationship between the two nodes. Arrows start at the replies, point to posters. It is directional. Often there are multiple interactions between two members of the innovative social network. Members will also reply to their posts.

#### 3.2 Social Relationship and Social Network

Members of the online community usually do not know each other in the real world. Based on similar interests, hobbies, knowledge backgrounds, sharing knowledge and experience, and so on, they establish interpersonal relationships in the online community, that is, social network. In the online community, the anonymity of membership makes interpersonal relationships virtual, random and free. The interaction between members is almost free from the real world rules.

Members of the customer innovation community are all customers of the company product. They have the same or similar product experience. When the members who post a topic, discuss or suggest, have the same awareness, interest or knowledge background, they will be attracted to each other to establish friendly relationship. In the community, this kind of friendly relationship has no direction. As the number of people increases and the interaction between members increases, the friend's relationship is increased. This gradually forms a social network.

#### 3.3 Knowledge-sharing Relationship and Knowledge-sharing Network

The strength of the knowledge-sharing relationship is measured by the similarity of the members' knowledge structure. The more similar knowledge structure is, the closer the knowledge is. In the process of knowledge transfer and absorption, the content of knowledge transfer depends on the members' common knowledge system. Knowledge-sharing is easier if both parties pass on shared knowledge. In a virtual community, a post on a related topic will attract other members to reply and participate in the discussion. By clearing up and processing, the results of the discussion become common knowledge in the community. These members are all co-contributors to the community's knowledge. Therefore, the more members discuss, the more knowledge they contribute, the stronger

the knowledge-sharing relationship is. In general, in the early days of the online community, the knowledge storage is small. As members gradually increase, knowledge interaction increases, knowledge gradually increases. It is changed from disorderly state to orderly and sorted state, forming an orderly knowledge system. After that, members can easily and purposefully share knowledge or demand knowledge.

The knowledge of the customer innovation community comes from its members. Knowledge-sharing is the foundation of knowledge that members have. As the source of knowledge, the knowledge-sharing relationship of members forms a knowledge-sharing network. The strength of knowledge-sharing relationships is measured by the similarity of members' knowledge structure. The higher similarity of the members' knowledge, the closer their knowledge relationship is. On the contrary, the greater difference in the members' knowledge structure, the weaker their knowledge relationship.

### 3.4 Relationship between the Three Networks

#### (1) The Relationship between Interactive Network and Knowledge-Sharing Network

In the customer innovation community, the interaction of members is the basis of the knowledge-sharing relationship. Interaction behavior is the external manifestation of knowledge-sharing activities. On the one hand, members absorb knowledge in the interaction process and change their knowledge structure. Members' knowledge-sharing relationships also change. On the other hand, as the knowledge structure changes, members' relationship becomes closer, members' willingness and inclination to interact gets stronger. Common experience will strengthen the interaction of members.

In the virtual community, the initial interaction of members happens often not because of acquaintance but based on their interests and knowledge. At first, only a few members have knowledge relationships. They interact with each other to create a few interactive relationships. At this time, the knowledge relationship of the members has an important impact on their interaction relationship. With the establishment and development of interaction, members become more familiar and more of them have the same knowledge. Then, the knowledge relationship of members changes. Therefore, the knowledge and the interaction of the virtual community members is not static. It continues to develop in coordination with the knowledge-sharing activities. There is a close relationship between them. So we make the first hypothesis:

Hypothesis 1: There is a positive correlation between the interactive network and the knowledge-sharing network in the customer innovation community.

#### (2) The Relationship between Interactive Network and Social Network

The interpersonal relationships in the virtual community are divided into two kinds. The first kind of interpersonal relationship is friend relationship, such as relationship on Facebook. This kind of network relationship is a strong relationship. The second kind is concern relationship, such as relationship on Twitter. The relationship in the customer innovation community is similar to the second kind. Members do not establish

relationships based on social purposes but based on hobbies, common knowledge background and experience sharing. The strength of virtual network interpersonal relationship is smaller than that in the actual world; the impact is also smaller. But it has a beneficial effect on the interactive network members. A member will be concerned quicker when a friend releases a theme. The discussion between friends is also closer than between other members. So we make the second hypothesis:

Hypothesis 2: There is a positive correlation between the interactive network and the social network in the customer innovation community.

#### (3) Social Network and Knowledge-sharing Network

An individual's attitude to a certain behavior depends on his beliefs, such as values and identity. In the process of knowledge-sharing in the community, members' knowledge structure, interest and values influence knowledge-sharing behavior. Members' value recognition helps to promote the occurrence and continuation of knowledge-sharing. The opposite will hinder knowledge-sharing. So we make the third hypothesis:

Hypothesis 3: There is a positive correlation between social network and knowledge-sharing network in the customer innovation community.

## 4 DATA COLLECTION AND PROCESSING

Xiaomi community is a communication platform established by Xiaomi Company for Xiaomi mobile phone users and enthusiasts. They communicate, discuss, share using experience, ask for help, make advice, and complaint in the community. It is currently one of the best online communities for customers to participate in product innovation.

Xiaomi community customers' product innovation participation data can be acquired through web crawler technology. This paper chooses "Xiaomi bus" theme data since it was established half a year ago as a sample. It has 529 topics, 1069 participants.

### 4.1 Data Processing Procedure

Through the associated crawler algorithm, the user's interaction data, social data and user's published topic data are collected in the Xiaomi community, and the three data are transformed into interaction matrix, social matrix and user's knowledge sharing matrix through different data processing methods. The data processing procedures are shown in Fig. 1.

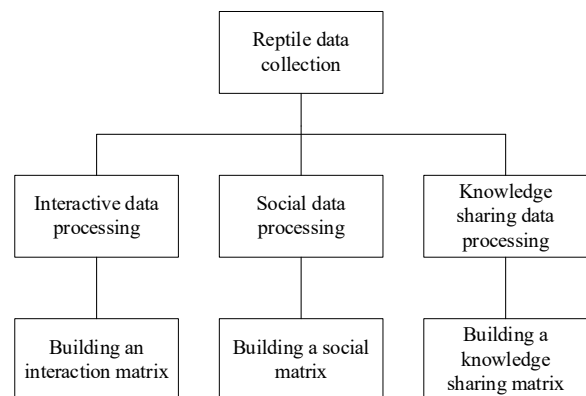


Figure 1 Data processing procedures

### 4.2 Associated Crawler Algorithm

There are a large number of structured and semi-structured data in the customer innovation community and forum page style is shown in Fig. 2. The traditional crawling algorithm mainly uses the fuzzy C-means method or the PSO web crawling method is hard to get the linked data from forum weds. In this paper, community associated crawler algorithm is adopted, and adaptive learning method is used for data relevance search and clustering. Combined with hierarchical segmentation method, community customer interactive data, social data and knowledge-sharing are extracted.

Based on the information transfer model of the community network and the mining of the characteristics of the community network user behavior information, the data crawling is performed.

The fuzzy attribute clustering method is used to cluster the user behavior attribute of the community network, and the preference attribute value of the community network user behavior feature reptile is recorded as  $C_{uv}^*$ .

$$C_{uv}^* = \sqrt{\frac{d_{in}(v)}{d_{out}(u) + d_{in}(v)}} \times C_{uv} \tag{1}$$

where  $d_{out}(v)$  is the starting point crawler link trajectory;  $u$  is the community attribute mixing recommendation association information set, that is, the  $u$  degree set, using the fuzzy decision making method, calculating the time trajectory set from the time  $T_0$  to the crawler end user position, and obtaining the user behavior The eigenvalues of the association rules mining of attributes are:

$$CW_{min}^{lM} = CW_{min}^{lM-1} \times (1 + \overline{D}_{lM-1})^x = CW_{min}^0 \times (1 + \overline{D}_{l0})^x \times (1 + \overline{D}_{l1})^x \times \dots \times (1 + \overline{D}_{lM-1})^x \tag{2}$$

The feature vector of the cluster output is self-correlated feature template matching method to achieve information fusion. The fitness function of information fusion is:

$$\beta_d = (MPDist - d + 1) / MPDist, d \in [2, MPDist] \tag{3}$$



Figure 2 The forum page

Based on the information transfer model of the community network and the mining of the characteristics of the community network customer behavior information, the data crawling is performed. The associated crawler algorithm is shown in Tab. 1.

Table 1 Associated crawler algorithm

```
#Build a collection of all item candidate sets
def createC1(dataSet):
    C1 = []
    for transaction in dataSet:
        for item in transaction:
            if [item] not in C1:
                C1.append([item])
    return list(map(frozenset, C1))
#A set of items L that meets the minimum support is generated from the candidate set.
def scanD(D, Ck, minSupport):
    ssCnt = {}
    for tid in D:
        for can in Ck:
            if can.issubset(tid):
                if not ssCnt.__contains__(can): # ssCnt[can] = ssCnt.get(can,0)+1
                    ssCnt[can] = 1
                else:
                    ssCnt[can] += 1
    numItems = float(len(D))
    retList = []
    supportData = {}
    for key in ssCnt:
        support = ssCnt[key] / numItems
        if support >= minSupport:
            retList.insert(0, key)
    supportData[key] = support
    return retList, supportData
#Generate K item candidate set C from Lk
def aprioriGen(Lk, k):
    retList = []
    lenLk = len(Lk)
    for i in range(lenLk):
        for j in range(i + 1, lenLk):
```

```

if len(Lk[i] | Lk[j])==k:
retList.append(Lk[i] | Lk[j])
return list(set(retList))
#Generate all frequent subsets
defapriori(dataSet, minSupport=0.5):
C1 = createC1(dataSet)
D = list(map(set, dataSet))
L1, supportData = scanD(D, C1, minSupport)
L = [L1]
k = 2
while (len(L[k - 2]) > 0): # k=2
Ck = aprioriGen(L[k - 2], k)
Lk, supK = scanD(D, Ck, minSupport)
supportData.update(supK) # supportData
L.append(Lk)
k += 1
return L, supportData
    
```

Three types of data are collected through the crawler algorithm:

**(1) Interactive data.** The type of interaction data between the collected users includes the customer ID and the ID of all the customers' replying to the published topic. The data types are shown in the following Tab. 2:

Table 2 Data type of interactive data

Customer ID	Respondent's ID
A1	A2
	A3
	A5

**(2) Social data type.** The type of social data collected between the users includes the user ID and the ID of all the friends of the user. The data types are shown in the following Tab. 3:

Table 3 Data type of social data

Customer ID	Friends'ID
A1	A3
	A6
	A7

**(3) Knowledge sharing data type.** The collected knowledge-sharing data mainly includes the user ID and all

the topic content published under the ID. The data format is shown in the following Tab. 4.

Table 4 Data type of knowledge-sharing data

ID	Content
A1	Can music be controlled in the lock screen state with a wire headset?
A2	Can yellow page content support offline caching?
A3	Can a common contact for a new SMS message be deleted from a private SMS contact?
A4	It is strongly recommended that the backup function back up the HD avatar of the phone contact.
A5	File cleaning, new features, very good look.

### 4.3 Data Matrix Conversion

Interactive data and social data are structured data after being acquired by the crawler algorithm. The co-occurrence algorithm can directly convert interactive data and social data into an interaction matrix and a social matrix, while the knowledge sharing data is unstructured data and needs to pass natural language. After the processing method extracts the keywords and forms the structured data, the co-occurrence algorithm is used to generate the knowledge sharing matrix. The process of generating the knowledge sharing matrix is shown in Fig. 3.

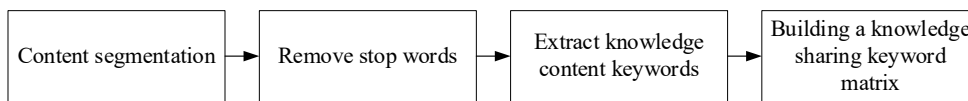


Figure 3 Knowledge sharing matrix transformation procedure

Table 5 Mobile phone concept dictionary

Dial-up	Call	SMS	SIM	Contact	Life yellow page	Lock screen
Status bar	Theme	Camera	Album	Music	Video	Browser
Download management	Email	File management	Notes	Compass	Recorder	Radio
Clock	Calculator	Local Backup	Dnd mode	Application Lock	Security center	App store
System update	Account	Voice assistant	Coin	Wallet	Input method	Life
Cloud	Sweep	VIP	Finance	Quick pass	Card	Smart family
Mobile avatar	Application dual open	Screenshot	Global search	Feedback	Multitasking	Remote assistance
Payment	Game acceleration	Smart assistant	Network disk	Forum	Red packet	Screen recording

Table 6 Keywords from text extraction

ID	Keywords
A1	Music/Lock screen/Status/Wire headset/Control/Music
A2	Yellow Page/Content/Support/Offline Cache
A3	SMS/Common Contacts/Private/SMS Contact
A4	Suggestion/Backup function/Phone Contact/HD avatar
A5	File cleanup/New feature

This paper adopts the Jieba segmentation algorithm, and uses the Chinese standard to disable the dictionary to remove the stop words, and extracts the keywords of the topic content according to the mobile phone related concept dictionary (some concept words are shown in Tab. 5), as shown in Tab. 6. Natural language processing and co-occurrence matrix algorithm are shown in Tab. 7.

**Table 7** Natural language processing and co-occurrence matrix algorithm

```

Input path: the path of text
Outputresult{<filename,category>:
Read text path
Read the stop words list
#Remove stop words
    Stpwrldst = readfile(path)
Load the dictionary
    Jieba.load(path)
#Segment
Segment
for text in file:
    segs = Jieba.cut(text)
Remove stop words
    Forseg in segs:
#Extract keywords
ifseg not in stpwrldst and seg.isdigit() == False:
sentence += " " + seg
Extract the feature words

#Building a knowledge sharing matrix
data = pd.read_excel(r'E:\Python\data.xlsx',header=None)
keyword = (set(i.split("/")) for i in data.loc[:,2])
keyword = set.union(*keyword)
togo = pd.DataFrame(np.zeros([len(keyword),len(keyword)]),columns=keyword,index=keyword)
for i in data.iloc[:,2]:
    line = i.split("/")
    togo.loc[line,line] = togo.loc[line,line] + 1
for i in range(len(togo)):
    togo.iloc[i,i]=0
    
```

**4.4 Data Processing Results**

Its customer interaction data, social data and knowledge-sharing data form 1069×1069 scale interaction matrix, social relations matrix and knowledge-sharing matrix. These three types of matrices are shown in the following Tab. 8, Tab. 9 and Tab. 10.

In the interactive matrix, ID of the first line represents the post, ID of the first column represents the replies. Except for the diagonal cells, the values of other cells represent the time the reply is posted to the poster.

**Table 8** Interactive relationship data matrix

ID	A1	A2	A3	A4	A5	A6
A1	-	0	0	1	2	1
A2	0	-	0	1	1	1
A3	0	0	-	0	0	1
A4	0	0	0	-	0	0
A5	0	0	0	0	-	0
A6	0	0	0	0	0	-

**Table 9** Social relationship data matrix

ID	A1	A2	A3	A4	A6	A7
A1	-	1	0	1	0	1
A2	1	-	0	0	0	0
A3	0	0	-	0	0	1
A4	1	0	0	-	0	0
A6	0	0	0	0	-	0
A7	1	0	1	0	0	-

The social relationship matrix is a symmetric matrix. The values of the matrix are only 1 and 0. When the value is 1, the two members have friend relationship. When the value is 0, the two members do not have friend relationship.

In knowledge-sharing matrix, the rows and columns of the matrix are IDs of the members participating in the topics. The values of the cells on the diagonal represent the topic number of the member participating in. For example, in the table "huazz145" participated in 64 topics. Except the diagonal, the value of the cell represents the topic

number that both members are participating. For example, "Huazz145" and "OnlySayBye" participate in five topics together.

**Table 10** Knowledge-sharing relationship data matrix

ID	A1	A2	A3	A4	A5	A5	A6
A1	2	1	1	1	2	2	1
A2	1	5	1	1	1	5	5
A3	1	1	2	1	1	2	2
A4	1	1	1	4	1	4	4
A5	2	1	1	1	2	2	1
A6	2	5	2	4	2	64	15

**5 RELATIONAL NETWORK VISUALIZATION**

Put the relationship matrix data into the UCINET software to calculate the basic attributes of the three networks, such as the number of nodes, the number of edges, the isolated point, the network density, and the clustering coefficient.

(1) Cluster coefficient

It is used to measure the probability of links between two neighboring nodes in a social network, indicating the closeness of the relationship between actors in the network. Given the degree of node  $i$  is  $k_i$ , there are up to  $C_{k_i}^2 = k_i(k_i - 1) / 2$  edges in  $k_i$  neighbors. In the actual network, the clustering coefficient of node  $k$  is:

$$C_i = \frac{E_i}{(k_i(k_i - 1)) / 2} = \frac{2E_i}{k_i(k_i - 1)} \tag{4}$$

(2) Network density

The density of the network map is the ratio between the number of connections actually present in the graph and the maximum number of possible connections. In addition, the density of directed networks and undirected networks is also different. Directed networks calculate density by point-in and dot-out, respectively, while undirected

networks use only degrees of degrees as the calculation standard.

Directed graph:  $Den = 2M / N(N - 1)$  (5)

Undirected graph:  $Den = M / N(N - 1)$  (6)

(3) Average path length

Consider an unweighted graph  $G$  with the set of vertices  $V$  let  $d(v_1, v_2)$ , where  $v_1, v_2 \in V$  denote the shortest

distance between  $v_1$  and  $v_2$ . Assume that  $d(v_1, v_2) = 0$  if  $v_1 = v_2$  or  $v_2$  cannot be reached from  $v_1$ . Then, the average path length  $l_G$  is:

$$l_G = \frac{1}{n(n-1)} \sum_{i,j} d(v_i, v_j) \tag{7}$$

The results are shown in the Tab. 11:

Table 11 Properties of the three relationship networks

Relationship networks	Nodes	Edges	Isolated point	Network density	Cluster coefficient	Average path length
Interactive network	1069	1840	52	0,0017	0,084	3,022
Social network	1069	11	1055	0,00001	0,7	1,474
Knowledge-sharing network	1069	9767	2	0,0093	0,376	2,997

Put the three relationship matrix data into NET REVIEW, the community map of the relationship networks is as follows:

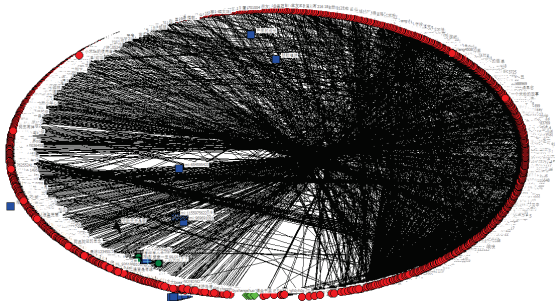


Figure 4 Interactive relationship network community map

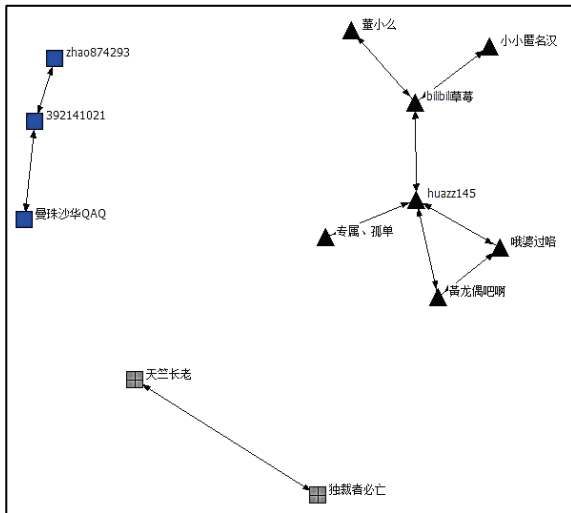


Figure 5 Social relationship network community map

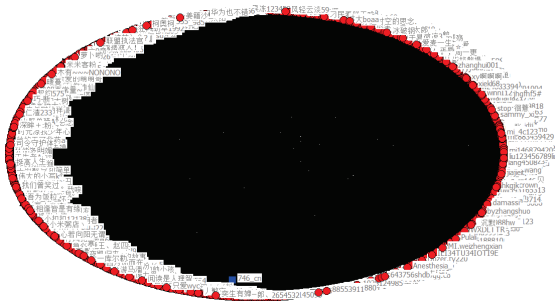


Figure 6 Knowledge-sharing relationship network community map

6 NETWORKS CORRELATION ANALYSIS

Through the previous theoretical analysis, we can conclude that there is an association between interactive network, social network, and knowledge-sharing network in the innovative community. They jointly affect customers' participation in product innovation activities. From the empirical analysis of the data, we can see that interactive network, social network and knowledge-sharing network are all sparse, especially social network. So we need to quantify the three relationships.

Many a software provides correlation analysis. But traditional correlation analysis requires variables to be independent from each other, while relational data is just the opposite. QAP (Quadratic Assignment Procedure) is a method to compare the similarity of two square matrices according to the values of the elements in the square matrix. The correlation coefficients between the matrices are obtained by comparing every element value. It performs a nonparametric test on the coefficients at the same time.

The calculation steps are as follows:

(1) First calculate the initial correlation coefficient between the matrices.

(2) Randomly substitute the rows and columns of one of the matrices. Calculate the correlation coefficient between the replaced matrix and the other matrix. Repeat this process several times until you get a correlation coefficient distribution.

(3) Finally, compare the initial correlation coefficient and the correlation coefficient distribution after replacement. Determine the relevance of the two matrices by examining whether the correlation coefficient is in the rejection region or the acceptable region. If the significance level is below 0,05 there is a statistically significant relationship between the two matrices.

The interactive network is asymmetric network, which is a multi-valued directed network, while social network and knowledge-sharing network are symmetric networks. So the interactive network needs to be symmetric before the correlation analysis. Then use the QIN analysis of UCINET software to calculate the correlation of two matrices.

(1) Interactive Network and Social Network

Write the interaction matrix and the social matrix into UCINET to obtain the correlation coefficient and the  $p$  value matrix. The result is as follows:

1) correlation coefficient

Table 12 Correlation coefficient matrix

	Interactive network	Social network
Interactive network	1,000	0,018
Social network	0,018	1,000

2) *p* value

Table 13 *p* value calculation matrix

	Interactive network	Social network
Interactive network	0,000	0,061
Social network	0,061	1,000

The results show that the correlation coefficient between the two matrices is only 0,013, the *p* value of QAP is 0,061 not less than 0,05. Statistically, there is no correlation between the interactive network and the social network.

(2) Social network and knowledge-sharing network

Write the social matrix and the knowledge-sharing matrix into UCINET to obtain the correlation coefficient and the *p* value matrix. The result is as follows:

1) correlation coefficient

Table 14 Correlation coefficient matrix

	Social network	Knowledge-sharing network
Social network	1,000	0,036
Knowledge-sharing network	0,008	1,000

2) *p* value

Table 15 *p* value calculation matrix

	Social network	Knowledge-sharing network
Social network	0,000	0,173
Knowledge-sharing network	0,173	0,000

The results show that the correlation coefficient between the two matrices is only 0,036; the *p* value of QAP is 0,173 not less than 0,05. Statistically, there is no correlation between the social network and the knowledge-sharing network.

(3) Interactive network and knowledge-sharing network

Write the interactive matrix and the knowledge-sharing matrix into UCINET to obtain the correlation coefficient and the *P* value matrix. The result is as follows:

1) correlation coefficient

Table 16 Correlation coefficient matrix

	Interactive network	Knowledge-sharing network
Interactive network	1,000	0,095
Knowledge-sharing network	0,095	1,000

2) *p* value

Table 17 *p* value calculation matrix

	Interactive network	Knowledge-sharing network
Interactive network	0,000	0,000
Knowledge-sharing network	0,000	0,000

The results show that the correlation coefficient between the two matrices is only 0,095 and the *p* value of QAP is less than 0,05. Statistically, there is a positive correlation between the interactive network and the knowledge-sharing network.

The results of network correlation analysis show that both the interactive network and the knowledge-sharing network are not positively correlated with social network. But the interactive network and the knowledge-sharing network are positively correlated.

Customer innovation community belongs to functional community. When comparing with Twitter, Facebook and other social networks, it has lower social property. Except product enthusiasts and the company's employees, most community members only log in and use the community when they need it. Normally, they do not interact through the community. Although members in the community will become familiar in the interaction process, and develop friends relationships, the friends relationships do not always mean that the interaction between them is very active. There are a lot of silence and non-active relationships. The users' invisible behavior (such as browsing the web, or only looking for topics of interest) reduces the clustering coefficient. This makes the network structure looser, the average path longer, and the network connection performance worse.

In the study of online social network, Xu Xiang & Zhang Sai [12] pointed out a similar phenomenon. The study shows that many studies only focus on the statement of friends relationship, ignoring the actual user interaction. It is possible that when A and B have established friends relationship, C and D have not, but the frequency of communication between A and B is much lower than between C and D. Therefore, interactive network or knowledge-sharing network does not have a positive correlation with social network.

Through the analysis above, it can be concluded that in the customer innovation community, social relationship cannot enhance the interaction or knowledge-sharing between members directly. Interaction relationship will affect knowledge-sharing relationship directly; the more interactive the more knowledge-sharing.

## 7 CONCLUSION

Through the previous study, we found that in the customer innovation community, customers involved in product innovation process form interaction relationship, social relationship and knowledge-sharing relationship. Then they further form interactive network, social network and knowledge-sharing network. By collecting the structured data from Xiaomi community, cleaning, preprocessing and converting the data, then further studying, we get the results. The results show that the three networks are all sparse, especially the social network.

Through QAP method, we analyze the correlation of the three relationship networks. The results show that the interactive network has a high correlation with the knowledge-sharing network, and the interaction of customers with similar knowledge background is strong. But both the interactive network and the knowledge-sharing network are not positively correlated with social network. This shows that social environment in the



community is weak, members communicate little. In the development of customer innovation community, we need to focus on creating good social environment in the community, organize some activities in which customers participate jointly, to establish relationship. This encourages customers to participate in product innovation in the community for long, and improve interaction and knowledge-sharing frequency.

## 8 REFERENCES

- [1] Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social networks*, 13(2), 141-154. [https://doi.org/10.1016/0378-8733\(91\)90017-N](https://doi.org/10.1016/0378-8733(91)90017-N)
- [2] Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*. MIT Press. <https://doi.org/10.7551/mitpress/7105.001.0001>
- [3] Enjun, X., Huaihai, Z., & Ming, Z. (2014). Study on Key Variable Measure of Open Innovation Community Network. *Journal of Beijing Institute of Technology (Social Science Edition)*, 16(3), 43-50.
- [4] Johnson, R. K., Boström, B., & Van de Bund, W. (1989). Interactions between *Chironomus plumosus* (L.) and the microbial community in surficial sediments of a shallow, eutrophic lake. *Limnology and Oceanography*, 34(6), 992-1003. <https://doi.org/10.4319/lo.1989.34.6.0992>
- [5] Davenport, T. H., De Long, D. W., & Beers, M. C. (1998). Successful knowledge management projects. *Sloan management review*, 39(2), 43.
- [6] Bartol, K. M. & Srivastava, A. (2002). Encouraging knowledge sharing: The role of organizational reward systems. *Journal of Leadership & Organizational Studies*, 2002, 9(1), 64-76. <https://doi.org/10.1177/107179190200900105>
- [7] Koh J, Kim Y G. (2004). Knowledge sharing in virtual communities: an e-business perspective. *Expert systems with applications*, 26(2), 155-166. [https://doi.org/10.1016/S0957-4174\(03\)00116-7](https://doi.org/10.1016/S0957-4174(03)00116-7)
- [8] Chun, H., Kwak, H., & Eom, Y. H. (2008). Comparison of online social relations in volume vs interaction: a case study of cyworld. *Proceedings of the 8<sup>th</sup> ACM SIGCOMM conference on Internet measurement*. ACM, 57-70. <https://doi.org/10.1145/1452520.1452528>
- [9] Lampel, J. & Bhalla, A. (2007). The role of status seeking in online communities: Giving the gift of experience. *Journal of Computer-Mediated Communication*, 12(2), 434-455. <https://doi.org/10.1111/j.1083-6101.2007.00332.x>
- [10] Shu, W., Chuang, Y. H., & Lin, C. S. (2009). Knowledge Sharing Investigation in Social Networks. *Proceedings of the eighth international conference on information and management sciences*, 8, 67-77.
- [11] Chen, Y. *Study partners in the social networking platform*. China University of Science and Technology, 2015.
- [12] Xu, K., Sai, Z., & Hao, C. (2014). Measurement and Analysis of Online Social Network. *Chinese Journal of Computers*, 37(1), 165-188.

### Contact information:

#### Aiqin, ZHOU

School of Economics and Management,  
Tsinghua University  
E-mail: zhousaq@sem.tsinghua.edu.cn

#### Yusen, ZHOU

(Corresponding author)  
Neoma Business School,  
E-mail: zhouysen@126.com