

Text Mining Model for Virtual Community User Portrait Based on Social Network Analysis

Deming Li

Abstract: With the rapid development of virtual communities, more and more customers participate in product innovation and knowledge sharing through virtual communities. Research on virtual community members, especially for community members, will help the community to manage the members and further promote community development and knowledge innovation. At present, the main difficulty in the study of community member user portraits lies in the user's grasp of user behavior data in the community. There is a large amount of structured data and semi-structured data in the community, which is crucial for the portrayal dimension of user portraits. This paper uses the association rule crawler algorithm to conduct community user behavior data association search, and uses text mining, social network analysis (SNA) and clustering technology to image users in the knowledge innovation community from the perspectives of professionalism, participation enthusiasm and network capability. The main result divides users into fancier, participator, and tourist.

Keywords: co-creation; reptile algorithm; SNA; text mining; user portrait; virtual community

1 INTRODUCTION

With the advent of the experience economy (Pin, Glimore 1998) [1], more and more users participate in the process of production, innovation, sales and other aspects. Due to the rapid development of the Internet and the accelerating diffusion of the information, the users could get more information and details about the products and features, which stimulates the enterprise to continually update the products [2]. However, in the face of such changes in the market, the enterprise is difficult to grasp the potential needs of users, so that users directly involved in the product innovation process have become the best solution. According to the existing research, users participating in product innovation will reduce the risk of enterprise product innovation, reduce price sensitivity of users and enhance the degree of tolerance to product defects and further improve customer satisfaction of products.

Compared to the traditional user participating in product innovation, NPD based on the virtual community has many unique traits as follows: 1) the number of users is numerous; 2) there is no limit to the time and place of participation; 3) the methods of participation are rich; 4) user has a great attitude [3]. How to identify the roles of these users and dig out the most useful users who can give the most contribution to the product innovation has become the hot point to the research field and enterprise. On the one hand, the existing theoretical basis for customer classification is derived from customer entities involved in innovation and the number of entities involved in the environment is limited and the means of participation are less; on the other hand, the form of the questionnaire survey is generally used and the data obtained in this form is subjective and the number of the collection is limited.

Through the web crawler technology, a large amount of data about the behavior of users participating in the product innovation like online hours, the number of posts, the number of responses and browses etc., is collected. Text mining could be used to extract the professional words and emotional words from message text to identify the profession and emotion of the users. The cluster theory and social network analysis are applied to identify the roles of these users and make the user portrait. This research

method is to supplement the existing theory of customer classification, and to provide theoretical basis and operation method for the maintenance of customer relationship.

2 LITERATURE REVIEW

As more and more users participate in the process of product innovation, the customer willingness, innovation ability, effort and individual situation are different, so that some of the users will fit into the process of product innovation and some users will not [4]. How to classify these users becomes the main concern of enterprises and scholars. The foreign scholars have carried out the research in this area earlier: Von Hippel (1986) distinguished the leading user from common users. Two specific characteristics were used to describe the leading users: 1) leading users meet new product or new demand a few months or years earlier than the general users; 2) the leading users can reap the benefits by solving their own needs [5]. In the study of the open source software community, according to the contribution degree of user participation in the new product development process, Hemetsberger and Pieters (2001) divided the users in the community into three kinds: the major contributors, contributors and the ordinary users [6]. In the virtual environment, the users who are involved in innovation are classified as: users with the high innovative ability to provide the most valuable innovations; early users, to provide valuable test input; heavy users will provide existing products defect information to help improve the next generation of products [7]. Through the statistics and observation of the posts' number, content, quality and the content of users' communication in online basketball shoe community, Füller, Jaweck (2007) found that active users only accounted for 3% of the total users, 20% of which is called the driven users and the other 80% users is the type of stimulus-driven [8]. Perks, Gemser (2015) identified the users as the mainstream users and leading user, which has a wealth of domain knowledge and advanced market demand [9]. Gustafsson, Kristensson (2015) only divided users into leading users and ordinary users [10]. Domestic scholars have done related research as follows: According to the type of participation, users can be divided to the

guiding type, partner type, the proposed type and the report type [11].

Table 1 The classification of users

| Guiding type | Partner type |
|--|--|
| technology forecasting, identifying problems and opportunities, generating and defining concepts | cooperative development, resources sharing, discussion and application |
| Proposed type | Report type |
| test, customer research, market experience | product discussion, focus group, trial version test |

Based on the participation degree, customer value and customer behavior in product innovation, users are divided into leading users, growing customer, avoidant customer, lag type of customer [12]. Sukhodolov et al. (2013) identified the leading customer by collecting auto forum data and combining with the leading users' features such as customer demand leading degree, professional knowledge level, and label praise degree and network connection strength through the cluster analysis [13]. However, the defect of this method is division of the category of users by the advantages and disadvantages of the content of a topic; the amount of data is too small, and the recognition of the trust is low. Existing research also pointed out that users are often difficult to directly participate in the core technology research and development of products, and mainly reflected in demand for expression, creative generation, design, information feedback and sharing and other aspects [14].

3 THE BEHAVIOR OF USERS IN VIRTUAL COMMUNITY

In the virtual community, the behaviors of customer participation in product innovation include browsing, posting themes, response, reply to other users, participating in the official activities such as voting, testing, cooperative development organization [15]. Several forms of customer participation in product innovation in the network community can be divided into the following situations:

(1) In the community, users post themes to propose product demand, suggestions or complaints. They also discuss with others and hope to capture the attention of enterprises.

(2) Users rely on other's product innovation related topics and discuss with them.

(3) Users participate in the activities of relevant product innovation, such as questionnaire survey, product testing, collaborative development, etc.

The customer's feature classification is based on the quality of the content, the release of frequency, professional level, and also includes the member's personal information. Fuller, Jaweck (2007) [16] collected the data from basketball shoe forum and observed the behavior of members of the community to analyze the communication content between the members qualitatively and discuss with the industry experts.

By combining the above theoretical analysis and taking XIAOMI community as an example, the original index and derivative index for the customer participation the product innovation is proposed. The data of the original index could be collected directly in the community; the data of derivative index could be obtained through the

calculation of the original index. The indicators are shown in the following Tab. 2 and Tab. 3.

Table 2 The original index

| The Original Index | | Content |
|--------------------------------------|-----|--|
| Total number of the customer's theme | TT | The total number of the customer's themes |
| Bug feedback | BF | Bug feedback from users in the process of testing or using the product |
| Idea | IA | Feedback from users on the product function demand |
| Response | RE | Total number of responses to other' themes |
| Online hours | OH | Total number of online hours |
| Words of themes | WT | Total number of words of the customer's themes |
| The number of browses | TB | Total number of browses from other users |
| The number of responses | TR | Total number of responses from other users |
| Registration time | RT | Customer registration time |
| The number of friends | TF | The total number of the customer' friends in the community |
| Sum of professional words | SPW | The sum of professional words from message text |
| Sum of emotional words | SEW | The sum of emotional words from message text |

Table 3 The derived index

| Derived index | Calculation formula | Content |
|-------------------------------|----------------------|---|
| Innovation theme | $IT = BF + IA$ | Total number of innovation themes |
| The ratio of innovation theme | $RIT = (BF + IA)/TT$ | The ratio of innovation themes accounted for the proportion of the total number of themes |
| Innovation frequency | $IF = (BF + IA)/OH$ | Frequency of customer's innovation theme |
| Average browse | $AB = TB/OH$ | Average number of browse per innovation theme |
| Average response | $AR = TR/OH$ | Average number of responses per innovation theme |
| Average words | $AW = WT/TT$ | Average number of words per innovation theme |

To reduce the impact of the unnecessary index for classification results, the index for cluster analysis includes IT, BF, IA, RIT, IF.

4 DATA ACQUISITION BASED ON ASSOCIATION RULE CRAWLER ALGORITHM

There are a large number of structured and semi-structured data in the customer innovation community. The traditional crawling algorithm mainly uses the fuzzy C-means method, the PSO web crawling method, the hierarchical segmentation clustering crawling method, etc., and the collection effect of structured data is better [17]. However, acquisition and matching with semi-structured data is less effective. In this paper, community network member attribute association mining algorithm is adopted, and adaptive learning method is used for data relevance search and clustering. Combined with hierarchical segmentation method, community user behavior, community attributes and scene information 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 initial network topology is as shown in the Fig. 1.

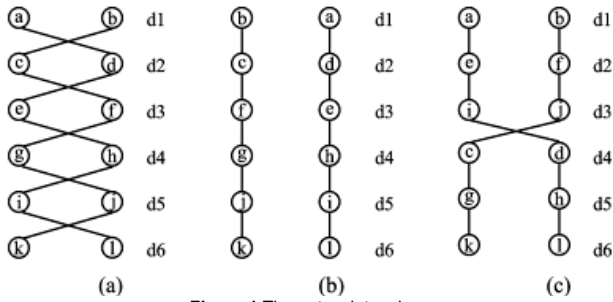


Figure 1 The network topology

Table 4 The main function of association rules

```
#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
def apriori(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
```

Table 4 The main function of association rules

```
#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
def apriori(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
```

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}^*

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

where $d_{out}(u)$ 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:

$$CW_{\min}^{l_M} = CW_{\min}^{l_{M-1}} \times (1 + \bar{D}_{l_{M-1}})^x = CW_{\min}^0 \times (1 + \bar{D}_{l_0})^x \times (1 + \bar{D}_{l_1})^x \times \dots \times (1 + \bar{D}_{l_{M-1}})^x \quad (2)$$

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:

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

In the formula, $adj(a, c)$ indicates the number of $a \rightarrow c$ paths, and $\beta_d \in (0, 1]$, the fuzzy directivity constraint control method is used to cluster the community user behavior attributes to improve the community discovery ability. The main functions are implemented in Tab. 4.

The customer innovation community behavior attribute data and text data obtained by the association rule crawler algorithm are shown in Tab. 5 and Tab. 6.

5 USER MESSAGE CONTENT TEXT ANALYSIS

This chapter uses text mining technology to perform text mining on member messages, and analyzes the professionalism, enthusiasm, and emotion of members' speeches. By using the Jieba Chinese word segmentation tool, the user's message is segmented, combined with the existing mobile phone professional vocabulary and emotional word vocabulary, the professional vocabulary and emotional vocabulary in the message are extracted, and the number of messages, professional vocabulary and emotional vocabulary is counted. Emotional vocabulary has positive vocabulary and negative vocabulary, and the enthusiasm of users is measured by the ratio of positive vocabulary to categorical vocabulary. The processing could be seen from Fig. 2.

Table 5 The behavior attribute data

| UID | Friends | Reply | Themes | Time Online (Hours) | Scores | Experience | Prestige | Contribution |
|------------|---------|--------|--------|---------------------|--------|------------|----------|--------------|
| 46832945 | 0 | 427 | 81 | 132 | 6033 | 5713 | 0 | 3 |
| 1148539658 | 0 | 967 | 100 | 84 | 3750 | 3177 | 0 | 0 |
| 482377212 | 0 | 908 | 88 | 53 | 4597 | 4073 | 0 | 0 |
| 449135173 | 0 | 159 | 19 | 12 | 1086 | 991 | 0 | 0 |
| 1545952186 | 0 | 118 | 33 | 207 | 4119 | 3940 | 0 | 3 |
| 44432451 | 0 | 125 | 40 | 146 | 1262 | 1106 | 0 | 0 |
| 135743799 | 167 | 116219 | 1081 | 2006 | 236838 | 137286 | 2124 | 200 |
| 794209690 | 1 | 360 | 1 | 40 | 716 | 517 | 0 | 0 |
| 159116352 | 0 | 87 | 2 | 20 | 1604 | 1549 | 0 | 0 |
| 159681699 | 0 | 6 | 1 | 0:00 | 97 | 136 | 0 | 0 |
| 156611235 | 0 | 1647 | 30 | 162 | 5945 | 5025 | 0 | 0 |

Table 6 The text data

| UID | Message Titles | Message Contents |
|------------|--|--|
| 46832945 | Two functional recommendations | Just arrived today, the first feeling first to talk about! Is dialing. I like it. Let me talk about two points first, the first hope to add a dial button switch like the 2.3 system! Just press the dial button and do not appear the number button. |
| 1148539658 | I hope to join the smart stroke dialing | Recently updated 2.10.12 joined the new feature phonetic dialing. Plus a friend in the forum said this smart stroke dialing function! Reminds me that smart stroke dialing is actually a feature I have been using before, in the past with Nokia In the era of N72, there is an input method called the national pen! The biggest reason for like the national pen is that you can dial the smart pen directly on the numeric keypad. |
| 482377212 | About the ringtone settings, read about a minute each time! This experience is very bad! | About the ringtone settings, read about a minute each time! This experience is very bad! I have feedback this before, but still have not solved it. I believe this is very simple to do. The previous 2.3 setup experience is much better! Although my ringtone library is relatively large, I still read it in seconds~~~~ I hope the development team pays attention! |

Table 7 The text mining algorithm

```

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

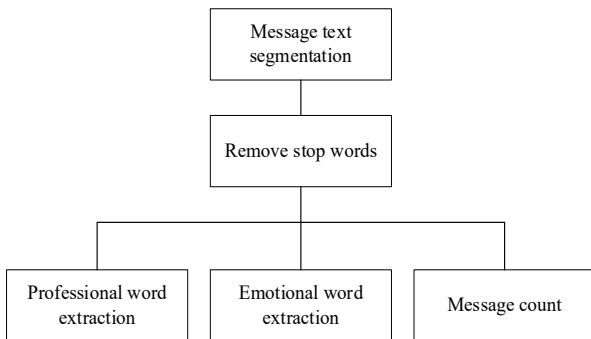


Figure 1 The text mining processing

The core text-mining algorithm is listed as Tab. 7.

The results of the professional words extraction and the emotional words extraction could be seen as Tab. 8 and Tab. 9.

Table 8 The extraction of professional words

| UID | Professional Words | Sum |
|------------|---|-------|
| 46832945 | Camera, photo album, browser, compass, porting, application dual open, input method, card, notification bar | 10 |
| 1148539658 | Status bar, system update, voice assistant | 3 |
| 482377212 | Multitasking, feedback, remote assistance, screen recording, click, global search | 6 |
| 449135173 | Compass, recorder, radio, player, flow control | 5 |
| | | |
| 1545952186 | System settings, weather, hovering ball, app store, security center, screenshots, account, gallery | 9 |

Table 9 The extraction of emotional words

| UID | Positive emotional words | Negative emotional words | Emotional value |
|------------|---|--|-----------------|
| 46832945 | Very good, great, praise, perfect, happy, proud | None | 5 |
| 1148539658 | Satisfied, happy | Angry | 2 |
| 482377212 | Happy, ecstasy | Garbage, disappointment | 0.5 |
| 449135173 | None | Tired, stunned, unhappy, junk, depressed, suck | -5 |
| | | | |
| 1545952186 | None | None | 0 |

The sum of users' message words are counted and listed in Tab. 10.

Table 10 The sum of users' message words

| UID | Message words quantity statistics |
|------------|-----------------------------------|
| 46832945 | 101794 |
| 1148539658 | 52826 |
| 482377212 | 57437 |
| 449135173 | 74654 |
| | |
| 1545952186 | 7427 |

6 DATA PREPROCESSING

The establishment of the marine knowledge interest community is for fans who are interested in marine knowledge, communicating with each other. They could discuss with each other, share the knowledge or publish questions. It is one of the best virtual communities to involve users to participate in innovation.

In the community, every member has a unique record with a nickname and a UID. It records users' registration time, online hours, behaviors data like the quantity of theme, response, etc.

The data used in this paper is collected from an interesting group, about 150 members.

Table 11 The data of the original index

| ID | Theme | Innovation Theme | Browse | Response | Online Time |
|----|-------|------------------|---------|----------|-------------|
| 1 | 224 | 9 | 49950 | 457 | 4695 |
| 2 | 343 | 48 | 421156 | 4152 | 6072 |
| 3 | 1179 | 931 | 1681162 | 5663 | 6092 |

Table 12 The data of the derived index

| ID | Innovation Theme | Innovation Proportion | Innovation Frequency |
|----|------------------|-----------------------|----------------------|
| 1 | 9 | 0.04 | 0.0019 |
| 2 | 48 | 0.14 | 0.0079 |
| 3 | 931 | 0.79 | 0.1528 |

In-Degree centrality defines the number of comments received by each member of the network and indicates a user's popularity and prestige within the network. Out-Degree Centrality is defined as the amount of a node's outgoing relationships. The out-degree measures the number of comments each user has written on the ideas of other community members.

In the following matrix, the line represents the users, who post themes; the list represents the users, who reply to others themes. X_{ij} is the number of columns J users replying to line I customer's themes. The relationship of the mutual responses between users is shown in Tab. 13:

Table 13 The data of the relationship of users

| ID | 1 | 2 | 3 | 4 |
|----|---|---|---|---|
| 1 | — | 1 | 0 | 2 |
| 2 | 4 | — | 3 | 2 |
| 3 | 2 | 1 | — | 0 |
| 4 | 0 | 0 | 0 | — |

7 VIRTUAL COMMUNITY USER PORTRAIT

7.1 Cluster Analysis

The customer sample is divided into k class, that is G_1, G_2, \dots, G_k . $X_i^{(t)}$ represents the customer I . The number of the G_t -type customer is the center of gravity of G_t . Sum of squares of deviations of G_t is

$$S_t = \sum_{i=1}^{n_t} \left(X_i^{(t)} - \bar{X}^{(t)} \right)' \left(X_i^{(t)} - \bar{X}^{(t)} \right) \tag{4}$$

K class customer's internal sum of deviation square is

$$S_t = \sum_{i=1}^k S_t = \sum_{n=1}^k \sum_{i=1}^{n_t} \left(X_i^{(t)} - \bar{X}^{(t)} \right)' \left(X_i^{(t)} - \bar{X}^{(t)} \right) \tag{5}$$

The specific approach is to take n samples into a class, and then each time to narrow a class, each of the differences between the sum of the square and to increase, select the S to increase the smallest class merger, knowing that all the samples belong to a class so far.

Table 14 The result of Cluster

| Cluster classification | Sum | Theme | Enthusiasm | Innovation | Profession | Emotion |
|------------------------|-----|-------|------------|------------|------------|---------|
| | | AVG | AVG | AVG | AVG | AVG |
| 1 | 42 | 1.29 | 1.14 | 0.14 | 0.04 | 0.002 |
| 2 | 50 | 6.16 | 2.84 | 3.32 | 0.18 | 0.023 |
| 3 | 58 | 85.4 | 67.8 | 17.6 | 0.48 | 0.126 |

From the results of cluster, the user's portrait could be divided to three categories as in Fig. 3.

(1) **Fancier.** For this type of users, the sum of theme, innovation proportion, and innovation frequency are obviously more than the other two types of users. It shows that the users have a strong willingness to participate in the innovation of the products and putting into action. The high efficient innovation frequency means that the main time can be used for product innovation.

The high number of friends and responses shows that such users' social relationships in the community are rich,

and could be able to actively participate in the discussion of the topic.

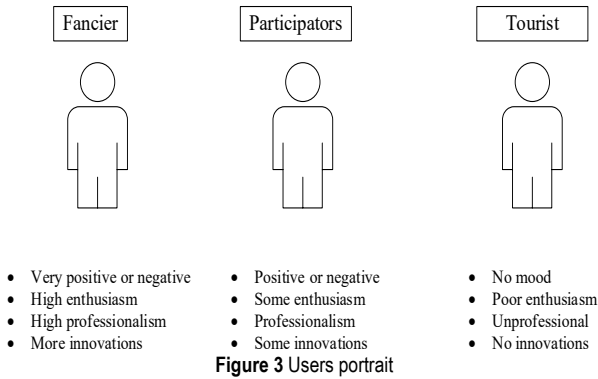


Figure 3 Users portrait

Users' average online time is much higher than the other two types of users, and the registration of the time is earlier. They are the earlier users of this product and continue to pay attention to follow-up products, of which 95% of the users continued activity to 2016.

From the average browse, reply and an average number of words of the themes, we can find that the index of this type of users is higher than of the other two types of users, especially the average number of the words of the themes. Through the detailed analysis of the samples, we also found that this type of users could describe the problems they found or new demands they require more clearly. For the questions, they could give their opinions or propose solutions. From the above data analysis, we could define this type of users as the positive users.

(2) **Participator.** Most of indexes are less than the third type of users, but more than the first type of users. This kind of users have the willingness to participate in the product innovation, and to a certain extent, to participate in the product innovation. They have weak social relationship with other members and are less involved in the topic. They are easily affected by internal and external factors and brand loyalty is low.

(3) **Tourist.** This kind of users have low willing to participate in the product innovation, they have less themes about innovation and rarely participate in the topic discussion. But they may join the community earlier and be positive in the community. Most of the themes posted in the community are not related to product innovation. Some of them are very concerned about social relations and very enthusiastic to participate in the discussion, which is nothing about product innovation.

7.2 Social Network Analysis

Social network is a collection of social actors and their relationships. Social network analysis is a quantitative study of the relationship between these social actors. Relationship graph and relational matrix are two common ways to describe social network relations. Based on the customer behavior of posting and replying in the community, relation matrix is established and the degree centrality is used to analyse the network construction. The centrality of a point is the number of nodes directly connected to a node, which is a pointer to measure the position of the actor in the network. It is used to measure the interaction ability of the moving person in the network.

In general, if the node's in-degree is large, it means that the node in the network has great influence and a more important position in the network; if the node's out-degree is large, it means that the node likes to communicate with other nodes in the network and is more active, belonging to the active actors in the network.

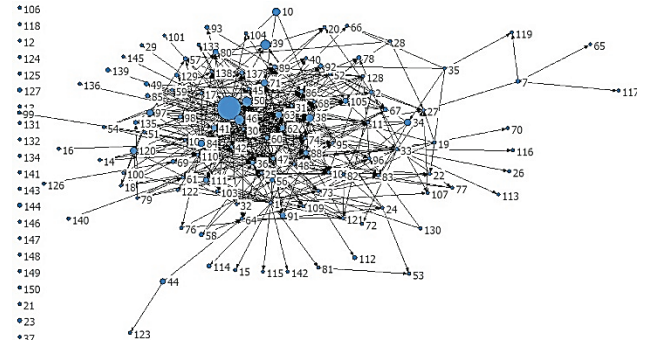


Figure 4 The relationship of users in community

Table 15 In-degree and out-degree of part of users in the social network

| ID | Out-Degree | In-Degree | ID | Out-Degree | In-Degree |
|-----|------------|-----------|-----|------------|-----------|
| 62 | 83 | 28 | 3 | 4 | 153 |
| 86 | 58 | 5 | 46 | 19 | 105 |
| 84 | 56 | 10 | 68 | 19 | 2 |
| 110 | 40 | 0 | 63 | 19 | 42 |
| 74 | 34 | 2 | 80 | 18 | 4 |
| 38 | 33 | 9 | 45 | 17 | 10 |
| 41 | 32 | 15 | 60 | 0 | 108 |
| 111 | 30 | 3 | 88 | 16 | 14 |
| 25 | 27 | 10 | 131 | 0 | 0 |
| 89 | 26 | 0 | 132 | 0 | 0 |

Table 16 The proportion of non-isolated points and isolated points

| Category | Number | Proportion |
|---------------------|--------|------------|
| Non-isolated points | 118 | 78.6% |
| Isolated points | 32 | 21.4% |

Network density

Density (matrix average) = 0.0053

Standard deviation = 0.0837

Table 17 The classification of users in product innovation community

| Category | Non isolated point | Isolated point | Proportion of Isolated point (%) |
|-----------|--------------------|----------------|----------------------------------|
| Cluster 1 | 42 | 11 | 26.2 |
| Cluster 2 | 50 | 15 | 30.0 |
| Cluster 3 | 58 | 6 | 10.3 |
| Sum | 118 | 32 | 21.4 |

8 CONCLUSION

Through the data collecting from a marine knowledge interest community, cluster theory and SNA are used to identify the roles of users in product innovation in the community. Five kinds of users in the virtual community are classified, they include creative users, innovation users, discussion users, common users, guests. Different kinds of users have different behavior, influence on others and impact on the product innovation. They are in the different position in the network.

The deficiency of this paper is that the research object is a partial network, and does not cover the whole network in the community. Future work will be about the impact these users have on the product innovation, the relationship between the roles and influence.

Acknowledgements

This work was supported by the National Grand Fundamental Research 863 Program of China (No. 2012AA011201), the Major Program of the National Natural Science Foundation of China (No. 91318301), the National Nonprofit Industry Research (No. 201313009-08), the Foundation for Outstanding Young Scientist in Shandong Province (No. BS2015DX017), and the Fundamental Research Funds for the Central Universities (No. 15CX02050A).

9 REFERENCES

- [1] Pine, B. J. & Gilmore, J. H. (1998). Welcome to the experience economy. *Harvard business review*, 76, 97-105.
- [2] Bettiga, D., Lamberti, L., & Noci, G. (2018). Investigating social motivations, opportunity and ability to participate in communities of virtual co-creation. *International journal of consumer studies*, 42(1), 155-163. <https://doi.org/10.1111/ijcs.12409>
- [3] Kim, S., Song, K., & Coppersmith, S. (2018). Creating an Interactive Virtual Community of Linguistically and Culturally Responsive Content Teacher-Learners to Serve English Learners. *Contemporary Issues in Technology and Teacher Education*, 18(2), 442-466.
- [4] Hsieh, S. H., Tseng, T. H., & Lee, C. T. (2018, January). Drivers of Online Brand Community Value Creation: The Role of Psychological Empowerment. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2018.265>
- [5] Von Hippel, E. (1986). Lead users: a source of novel product concepts. *Management science*, 32(7), 791-805. <https://doi.org/10.1287/mnsc.32.7.791>
- [6] Hemetsberger, A. & Pieters, R. (2001, May). When consumers produce on the Internet: An inquiry into motivational sources of contribution to joint-innovation. In *Proceedings of the Fourth International Research Seminar on Marketing Communications and Consumer Behavior*, La Londe, 274-291.
- [7] Füller, J. & Matzler, K. (2007). Virtual product experience and customer participation—A chance for customer-centred, really new products. *Technovation*, 27(6-7), 378-387. <https://doi.org/10.1016/j.technovation.2006.09.005>
- [8] Füller, J., Jawecki, G., & Mühlbacher, H. (2007). Innovation creation by online basketball communities. *Journal of Business Research*, 60(1), 60-71. <https://doi.org/10.1016/j.jbusres.2006.09.019>
- [9] Gemser, G. & Perks, H. (2015). Co-creation with customers: An evolving innovation research field. *Journal of Product Innovation Management*, 32(5), 660-665. <https://doi.org/10.1111/jpim.12279>
- [10] Bartikowski, B., Laroche, M., Jamal, A., & Yang, Z. (2018). The type-of-internet-access digital divide and the well-being of ethnic minority and majority consumers: A multi-country investigation. *Journal of Business Research*, 82, 373-380. <https://doi.org/10.1016/j.jbusres.2017.05.033>
- [11] Morgan, T., Obal, M., & Anokhin, S. (2018). Customer participation and new product performance: Towards the understanding of the mechanisms and key contingencies. *Research Policy*, 47(2), 498-510. <https://doi.org/10.1016/j.respol.2018.01.005>
- [12] Min Hyejong, Junghwan Yun, & Youngjung Geum. (2018). Analyzing dynamic change in customer requirements: An approach using review-based Kano analysis. *Sustainability* 10(3), 746. <https://doi.org/10.3390/su10030746>
- [13] Sukhodolov, A. P., Popkova, E. G., & Kuzlaeva, I. M. (2018). Perspectives of internet economy creation. In *Internet Economy vs Classic Economy: Struggle of Contradictions*, Springer, Cham, 23-41. https://doi.org/10.1007/978-3-319-60273-8_3
- [14] Von Hippel, E. & Katz, R. (2002). Shifting innovation to users via toolkits. *Management science*, 48(7), 821-833. <https://doi.org/10.1287/mnsc.48.7.821.2817>
- [15] Abhari, K., Davidson, E., & Xiao, B. (2018). Classifying motivations in social product development networks: a discriminant analysis of actor profiles. <https://doi.org/10.24251/HICSS.2018.067>
- [16] Elmezeny, A., Wimmer, J., dos Santos, M. O., Orlova, E., Tribusean, I., & Antonova, A. (2018). Same but Different: A Comparative Content Analysis of Trolling in Russian and Brazilian Gaming Imageboards. *Game Studies*, 18(2).

Contact information:

Deming LI
 School of Economics and Management,
 Beijing Jiaotong University,
 No. 3, Shangyuancun, Haidian District, Beijing, China
 E-mail: demingli_bjtu@163.com