Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

A customer network value model based on complex network theory

Rong Liu, Jun Ma, Jiayin Qi
School of Economics and Management
Beijing University of Posts and Telecommunications
Beijing, P.R. China
liurongbsr@gmail.com, qijiayin@139.com

Bin Wu
School of Computer Science and Technology
Beijing University of Posts and Telecommunications
Beijing, P.R. China
wubin@bupt.edu.cn

Chen Wang
IBM China Research Laboratory
IBM
Beijing, P.R. China
wangcwc@cn.ibm.com

***Abstract*—Network value has attracted more and more attentions from scholars and enterprisers. Information through network can essentially impact everyone, while everyone in the network has the structure function to the network and may also influence others. But the contribution to the network is not the same for every one. Based on the complex network theory and customer activity, a customer network value model is proposed, which is from the influence value and existence value. Then a case study was done with the data of a city branch of China Telecom. Finally, the management application based on customer network value is proposed.**

***Keywords*—customer network value, complex network, customer activity**

I. INTRODUCTION

Since 1990's, with the rapid development of information technology, especially for the Internet, we have entered a network era [1]. The network impact and value is so huge that it attracts researchers and enterprisers more and more attentions. But how to measure the network value is still a tough task.

Sarnoff's law, Metcalfe's law and Reed's law are three principles to evaluation network value, and they are about the technological leverage effect [2]. Dr. Dunbar, an anthropologist who studied the scales of group behavior of primates and its relation to the size of the relevant part of the brain, pointed out that regardless of the function and the purpose of the group, the natural groups of around 150 persons are likely to be formed in human society [3], so people's social capability is limited. Considering social capability, Noriaki Yoshikai proposed a modified network value model [2]. But all these models are about the whole network value, not the individual member contribution value to the network, in other words, not the customer network value.

The other relevant research to customer network value is Work-of-Mouth (WOM) value, which is an important part of customer value. WOM is a network phenomenon [4], and has received extensive attention from both academics and practitioners for decades [5]. John E. Hogan (2003) first introduced the WOM value in Customer Lifetime Value Model [6], and then Hyunseok Hwanag (2004), Hua Zhao (2005), Jonathan Lee (2006) modified the WOM value, and improved the CLV model [7][8][9]. Yingying Zhang (2007) intruduced the personalized product recommendation influences to customer value hierarchy research [10]. Form the WOM demonstration, the individual social ability is considered and WOM model can calculate individual customer's network value. However, the model is very complex and too many parameters need to be set, what's more, it ignores the integrity of the network and the importance of each customer to the whole network.

Complex network can deal with these problems. Euler's celebrated 1735 solution of the Königsberg bridge problem is often cited as the first true proof in the theory of networks [11]. Complex network research can be thought as lying on the intersection between graph theory and statistical mechanics, which confer a truly multidisciplinary nature to this area [12]. With the statistical mechanics, the nodes and edges are recoded, and so are interact and interdependent relationships. The graph theory can clearly describe the network and exhibit it. There are six important characteristics variables: degree, clustering coefficient, betweenness centrality, diameter, average path length and connected component, which can exhibit the network clearly.

In this paper, we combine the complex network theory which customer individual attribute to build customer network value model. With the degree, clustering coefficient, betweenness centrality, and customer activity, a customer network value is proposed, and then with the data of a city of China telecom, a case is done.

II. VARIABLES INVOLVED TO CUSTOMER NETWORK VALUE

Networks are all around us, and we are ourselves, as individuals, the units of a network of social relationships of different kind and, as biological systems, the delicate result of a network of biochemical reactions [13]. Customer network

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

value is the expected profit from sales to other customers whom she/he may influence to buy, the customers those may influence, and so on recursively [14]. So customer network value is similar with WOM value. While in this paper the customer network value is defined as the contribution value of a customer to the whole network, and WOM value is one of the important.

Since now, there are two ways to measure the network value. One is from network effect and the other is from word-of-value.

The network effect means that the value of a network depends on the number of users of the network. The more users, the greater the value of a network and the network effect increases. Noriaki Yoshikai summarized the Sarnoff's law, Metcalfe's law and Reed's law, and proposed a new model based on technological and social network [2]. No matter what form of the model, the network value is impacted by:

1. The number of a network member;

2. The number of friends she/he has.

TABLE I. NETWORK VALUE MODEL BASED ON NETWORK EFFECT DEMONSTRATION

	Samoff	Metcalfe	Reed	Noriaki Yoshikai
Possible Transaction	Receive Broadcast	Connect to Peers	Jion/Creats Groups	Reed's law , network constricts factors and human ability
Value of N member networks	N	N^2	2^N	
Combined Value of N and M member networks	$N+M$	$(N+M)^2$	2^{N+M}	$\sum_{k=1}^{m} \frac{N(N-1)\cdots(N-K+1)}{K(K-1)\cdots 1}$
Core Technology	Transponder	Channel Switch, Router	Technology for forming groups	Technology and social network

The other demonstration is WOM. Lots of scholars have done research in this area, and proposed models to measure WOM. And also, it is one part of Customer Lifetime Value. Jung and Euiho Suh proposed a CLV model considering the WOM [15]. And the equation is as the following:

$$LTV_i = \sum_{t_i=0}^{N_i} \pi_p (1+d)^{N_i - t_i} + k_i \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{\pi_f(t_i)+B(t_i)}{(1+d)^{t_i - N_i}} \quad (1)$$

k_i is the parameter of WOM, while the WOM is positive,

$$k_i = 1 + \frac{\sum_{j=1}^{n-n_0} (R_j \times P_j + C_0)}{R_i} \quad (2)$$

while the WOM is negative,

$$k_i = 1 - \frac{\sum_{j=1}^{n} (R_j \times P_j + C_0)}{R_i} \quad (3)$$

The WOM is positive or negative depends on the customer's satisfaction. n is the number of customers influenced by customer i; n_0 is the number of old customer who won't be impacted by customer i; R_j is the expected profit of j who is affected by customer i, P_j is the possibility of customer j buying behavior; C_0 is the average cost; and R_i is the expected profit of customer i in the future.

Other WOM models may be different, but while building the WOM model, the researchers generally consider the following impact factors:

1. The number of customers that customer i influences;
2. Customer satisfaction;
3. The expected profit;
4. The customer cost.

The last two variables are always evaluated on the assumption of experience, so the accuracy is suspected, while the first two can be evaluated via data mining and questionnaire.

So there are three calculable variables related to customer network value calculable, which are the number of network member, the influence capability of customer i, and the satisfaction of customer i.

However, both the network effect and WOM ignore the different importance of customer to the whole network. So in this paper, we introduce two complex network variables, clustering coefficient and betweenness centrality, to build a new customer network value model.

The customer network value (CNV) is the contribution of a customer to the whole network. In this paper, we analyze two contributions, one is the existence value, and the other is the influence value. The existence value is the contribution to the network structure, to the integrity and stability, and once the customer churns, the structure of the network will be destroyed, and the existence value is the basis of the network, lots of other values are based on the platform. While the influence value is one kind of values based on the network, which is the customer create based on the network, and the most common form is the WOM.

So the customer network value is impacted by the following five variables:

N: the number of network members/the scale of customers;

k_i: node degree/ the friend number of customer i;

S_i: satisfaction of customer i;

c_i: clustering coefficient of customer i;

and b_i: betweenness centrality of customer i.

III. CUSTOMER NETWORK VALUE MODEL

A network is a set of items, which we call vertices or sometimes nodes, with connections between them, called edges [11]. A complex network can be represented as a graph, and consists of two sets, nodes set V, and edges set E. And a network is indicated as G (V, E). There are undirected and directed, unweighted and weighted networks as figure 1. The network built in this paper is an undirected and unweighted one as the segment I in figure 1 shows.

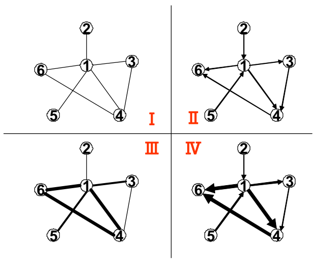

Figure 1. Four kinds of networks

A. node degree k_i

The degree k_i of a node i is the number of edges incident with the node. For Figure1 I, k_1 is 5, k_2 is 1, k_3 is 2, k_4 is 3, k_5 is 1 and k_6 is 2.

The most basic topological characterization of a graph can be obtained in terms of the degree distribution $P(k)$, and it is defined as the probability that a node chosen uniformly at random has degree k or, equivalently, as the fraction of nodes in the graph having degree k.

Nodes with the maximum score have more friends and more influence. So the higher score k_i has, the higher the CNV_i will be, and there is a positive relationship between k_i and CNV_i.

B. Clustering coefficient c_i

Clustering coefficient C, a measure introduced by Watts and Strogatz in Ref., defined as follows [16]. A quantity c_i (the local clustering coefficient of node i) is expressing how likely for two neighbors j and m of the node i. The local clustering coefficient of defined as the ratio between e_i (the actual number of edges among node i neighbors) and $k_i(k_i-1)/2$ (the maximum possible number of edges among node i neighbors).

$$c_i = \frac{2e_i}{k_i(k_i-1)} \tag{4}$$

And the clustering coefficient of the graph is then given by the average of over all the nodes in the graph.

$$C = \frac{1}{N}\sum_{i=1}^{N} c_i \tag{5}$$

Clustering coefficient is supposed to the density of network, and the C is one mark of the connections between customers. The bigger c_i is, the more connections among customer i friends have, and the less importance of customer i will be, for the network won't be destroyed if the customer churns.

C. Betweenness centrality b_i

Betweenness centrality has a strong relationship with the shortest paths. The shortest paths play an important role in the transport and communication within a network. Suppose one needs to send a data package from one computer to another one through the Internet: the geodesic provides an optimal path way, since one would achieve a fast transfer and save system resources [17]. It is useful to represent all the shortest path lengths of a graph G as a matrix D in which the entry d_{ij} is the length of the geodesic from node i to node j. The maximum value of d_{ij} is called the diameter of the graph, and will be indicated in the following as Diam(G). A measure of the typical separation between two nodes in the graph is given by the average shortest path length, also known as characteristic path length, defined as the mean of geodesic lengths over all couples of nodes [16].A problem with this definition is that L diverges if there are disconnected components in the graph. One possibility to avoid the divergence is to limit the summation in formula only to couples of nodes belonging to the largest connected component [16]. An alternative approach, that is useful in many cases, is to consider the harmonic mean of geodesic lengths and to define the so-called efficiency of G as [18]:

$$E = \frac{1}{N(N-1)}\sum_{i,j\in N, i\neq j} d_{ij} \tag{6}$$

Such a quantity is an indicator of the traffic capacity of a network, and avoids the divergence of formula, since any couple of nodes belonging to disconnected components of the graph yields a contribution equal to zero to the summation in formula.

The communication of two non-adjacent nodes, say j and k, depends on the nodes belonging to the paths connecting j and k. Consequently, a measure of the relevance of a given node can be obtained by counting the number of geodesics going through it, and defining the so-called node betweenness. Together with the degree and the closeness of a node (defined as the inverse of the average distance from all other nodes). The betweenness is one of the standard measures of node centrality, originally introduced to quantify the importance of an individual in a social network [19]. More precisely, the betweenness b_i of a node i sometimes referred to also as load, is defined as:

$$b_i = \sum_{j,k\in N, j\neq k} \frac{n_{jk}(i)}{n_{jk}} \tag{7}$$

n_{jk} is the number of shortest paths connecting j and k, while $n_{ij}(i)$ is the number of shortest paths connecting j and k and passing through i. The concept of betweenness can be extended also to the edges. The edge betweenness is defined as the number of shortest paths between pairs of nodes that run through that edge [20].

Nodes with the maximum score are assumed to be important for the graph to stay interconnected. These high scoring nodes are the 'weak nodes' that interconnect clusters of nodes, and removing them leads to unconnected clusters of nodes.

So the higher score of betweenness, the more important, and the customer network value will be higher.

D. *Customer satisfaction* S_i

Customer satisfaction has significant impact to WOM [6][7][8][9]. And a satisfied customer will be more active, and buy more products and have positive WOM, while if a customer unsatisfied, she/he will be less active and may lead to negative WOM, and P_i is a variable to measure customer's activity [21]. So we use P_i to measure the customer satisfaction, and the higher P_i is, the higher CNV_i will be. That means there is a positive relationship between P_i and CNV_i. SMC model can used to predict the customer activity P_i [22][23].

E. *Customer network value model*

As the above, the measurement of the four variables is settled, but the contributions and dimensions are different.

k_i and P_i measure one customer's influence to other customers, while c_i and b_i measure one customer's importance to the network structure. So the value of k_i and P_i is for influence, and the value of c_i and b_i is for existence.

The four variables range is:

$$k_i \in [1, N];$$
$$c_i \in [0,1];$$
$$b_i \in [1, C_N^2];$$
$$P_i \in (0,1);$$

To be comparable, k_i and b_i need to be dimensionless, since the magnitude of k_i is N, and b_i is N^2, so the formulas we use are:

$$k_i^{'} = \frac{\sqrt{k_i} - \sqrt{k_{\min}}}{\sqrt{k_{\max}} - \sqrt{k_{\min}}}; \qquad b_i^{'} = \frac{\ln b_i - \ln b_{\min}}{\ln b_{\max} - \ln b_{\min}}. \tag{8}$$

The specific function depends on the magnitude, and the basic principle is that the maximum of $k_i^{'}$ and $b_i^{'}$ should be no more than 100.

So the customer network value model is:

$$CNV_i = w_1 k_i^{'} P_i + w_2 b_i^{'} (1 - c_i) \tag{9}$$

w_1 is the influence value weight coefficient, w_2 is the existence value weight coefficient, and $w_1 + w_2 = 1$. The value of w_1 and w_2 depend on the decision-makers. At the beginning of the network, the existence may be more important, while at the stable period, the influence may be more important.

IV. A CASE STUDY

We use the *Network Workbench Tool* to analyze the network characteristics, and compute k_i, c_i and b_i. As for the customer activity P_i, we use *Matlab7.3.*

Two months call detail information of SCDMA customer, from September 1 to October 31, 2007, in a city branch of China Telecom was collected. SCDMA customers are one type of mobile customer serviced by a fix-line telecom company in China before Feb.2009.

The call detail information contains the following variables:

TABLE II. THE VARIABLE INS CALL RECORDS INFORMATION

SOURCEID
CALLINGAREACODE
CALLINGNUM
CALLEDAREACODE
CALLEDNUM
STARTDATETIME
ENDDATETIME
DURATION

A. *Three variables based on complex network* k_i, c_i *and* b_i

To build the call network, data cleaning rules are:

1. CALLINGAREACODE= CALLEDAREACODE;
2. CALLEDNUM belongs to the set of CALLINGNUM;
3. For the network is undirected, A calls B is the same as B call A, and ones A calls B, there is an edge between AB, no matter the times and duration of the call;
4. If CALLINGNUM=CALLEDNUM, delete the record.

After data cleaning, there are 10988 call records remained, which contains 6571 numbers. So the SCDMA call network has 6571 nodes and 10988 edges, the ratio of edges to nodes is 1.67.

The diameter is 19, and the average shortest path is 5.89, which matches the “six degrees of separation”. The network includes 271 connected components, and the biggest contains

5895 nodes, which accounts for 89.71%. The network statistical characteristics are:

The average degree is 3.34. The maximum score of the node degree is 342, which means this customer has called other 342 SCDMA customers, or she/he has 342 friends using SCDMA. And he can influence 342 customers in the network.

There are 5044 (76.76%) customers whose c_i is 0, which means the 5044 customers' friends don't know each other, once they churn, the whole network will be influenced, and the connectivity will be destroyed. Only 261 (4%) customers c_i is 1, which means the 261 customers' friends know each other, so their churn will be less important.

The highest score of b_i is 7022350, which means more than 7 million shortest path pass though this customer, so if the customer churn, the diameter and the average shortest path will be lengthen and the connection of the network may be impacted. There are 3666 customers whose b_i bigger than 6571, and the average of the b_i is 35454.81.

B. *Customer activity* P_i

To predict the P_i, two months, from September 1 to October 31, 61 days call records was collected, so T=61. After the "MADS Positive basis 2N" was analyzed, each customer's P_i is calculated. Using SMC model [23], the results of the four parameters are: α=-0.0676; β=0.2195; r=-1.0445; s=2.8710. The mean of P_i is 0.4641, and table III is some examples.

TABLE III. THE RESULT OF THE P_i

number	X=(x,t$_1$,t$_2$)	P
85****00	(1,22,61)	0.056
85****05	(2,54,61)	0.693
85****12	(1,4,61)	0.001
85****19	(2,59,61)	0.908
85****35	(2,24,61)	0.038
85****78	(1,53,61)	0.670
85****82	(1,60,61)	0.954
85****83	(65,61,61)	1.000

C. *Customer network value* CNV_i

The managers of the SCDMA customer think the influence value is the same important as the existence value. So $w_1 = w_2 = 0.5$, and the CNV model is:

$$CNV_i = 0.5k_i^{'}P_i + 0.5b_i^{'}(1-c_i) \quad (10)$$

And the customer network value distribution as the Figure 2:

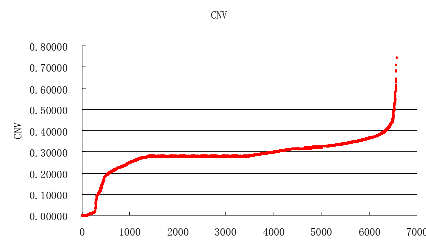

Figure 2. The distribution of CNV

Most of the CNV is from 0.2 to 0.4, and it accounts for 87.95%, and 8.54% are lower than 0.2, 3.51% are high than 0.4.

The maximum CNV score is 0.7453, the minimum is 0.007, and the mean is 0.2844. Figure 3 is the distribution from the influence and existence dimensionalities. Most of the customers have high existence value ($b^{'}(1-c)$) and low influence value ($k^{'}P$), and it is consistent with the fact.

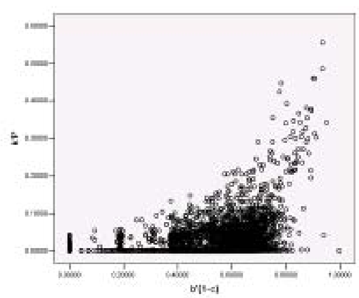

Figure 3. Influence and existence value

V. MANAGEMENT APPPLICATION

Customer value has different forms, and this paper analyzes the essential value from influence and existence value. Nowadays, lots of uncountable researches have been done for customer influence, while the existence value is still ignored. And it is very dangerous, especially for the company at the beginning period. The customer churn management has referred to the customer existence value, but it only considers the profit value, and ignores the demolishment to the network.

Based on our customer network value, from the two dimensions, we segment customers as Figure 4:

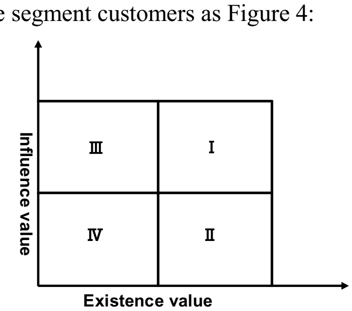

Figure 4. Customer segment based on CNV

Segment Ⅰ is the most important part. On one hand, the customer in this segment are important to the whole network, so customer retention management should pay more attention to them, on the other hand, their influence value are high too, so their influence capability should be used efficiently, and strengthen their WOM.

Segment Ⅱ is the second important part. In this paper, the existence value is the basic value, so it has precedence over influence value. For these customers in segment Ⅱ, churn prediction and retention management are the two emphases.

Segment Ⅲ is the third important part. To the customers in segment Ⅲ, their influence capability is high, while their contribution to the whole network is less. The marketing for these customers is the core important. To make satisfaction and loyalty campaigns plan for them to improve their loyalty and stickiness.

Segment Ⅳ is the less important and less attention should be paid for this part. But if there are too many customer, we can deeper classify them via customer profitability.

In the CNV model we proposed, there are 2 parameters need to be confirmed. And it depends on the enterprise life time stage and the management need.

VI. CONCLUSION

In this paper, based on the complex network and SMC, we integrate the network effect and WOM, analyze four variables, and build a customer network value model, from influence value and existence value. Then a case study in telecommunication industry is done and the application based on customer network value is proposed.

The customer network value measurement in this paper is an exploratory study, and the future work may be:

1. The improvement of CNV model. The model in this paper is the weighted summation of influence and existence value. With the deeper research on the relationship between influence value and existence value, the model can be refined.

2. Customer relationship management based on CNV. In this paper, we only analyze the segment management based on CNV, and how to use reward program or others strategies to raise customer network value is another direction of future research.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Project No.:70701005), the Specialized Research Fund for the Doctoral Program of Higher Education of the People's Republic of China (Project No.: 20070013014) and the Open Research Fund between Beijing University of Posts and Telecommunications and IBM.

REFERENCES

[1] Xiaofan Wang, Xiang Li and Guanrong Chen. Complex network theory and applications[M]. Beijing: Tsinghua University Press, 2006.

[2] Noriaki Yoshikai. A Network Value Model and its Application to Technology Evaluation for the On-line Community [C]. Information and Telecommunication Technologies, 2005. APSITT 2005 Proceedings. 6th Asia-Pacific Symposium on, Yangon, 2005, 399-404.

[3] Robin Dunbar. Grooming, Gossip and the Evolution of Language [M]. London: Faber and Faber, 1996.

[4] Paul Dwyer. Measuring the value of electronic word of mouth and its impact in consumer communities[J]. Journal of Interactive Marketing, 2007, 21(4): 63-79.

[5] [5] Arnaud De Bruyn and Gary L. Lilien. A multi-stage model of word-of-mouth influence through viral marketing[J]. Intern. J. of Research in Marketing, 2008, 25: 151-163.

[6] John E. Hogan, Katherine N. Lemon, Barak Libai. What is the true value of a lost customer? [J]. Journal of Service Research, February 2003, 5(3): 196-208.

[7] Hyunseok Hwanag,Taesoo Jung,Euiho Suh. An LTV Model and Customer Segmentation Based on Customer Value:A Case Study on the Wireless Telecommunication Industry [J]. Expert Systems with Applications, 2004, 26: 181-188.

[8] Hua Zhao and Xiulan Xia. Improvement of CLV Model Based on the Effect of Public Praise[J]. China Business and Market, 2005, 19(12): 42-45.

[9] Jonathan Lee, Janghyuk Lee & Lawrence Feick. Incorporating Word-of-mouth effects in estimating customer lifetome value. Data Marketing & Customer Strategy Management, 2006, 14(1): 29-39.

[10] Yangming Zhang, Jiayin Qi, Huaying Shu & Jiantong Cao. Personalized Product Recommendation based on Customer Value Hierarchy[C]. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, Montreal, QC, Canada, 2007, pp. 3250-3254.

[11] M. E. J. Newman. The Structure and Function of Complex Networks[J]. SIAM Review, 2003, 45(2): 167-256.

[12] Costa, Luciano da F., Rodrigues, Francisco A., Travieso, Gonzalo & Boas, P. R. Villas. Characterization of Complex Networks: A Survey of measurements. Advances in Physics, 2007, 56(1): 167-242.

[13] S. Boccaletti, V. Latorab, Y. Morenod, M. Chavezf, D.-U. Hwanga. Complex networks: Structure and dynamics[J]. Physics Reports, 2006, 424: 175 – 308.

[14] Pedro Domingos, Matt Richardson. Mining the Network Value of Customers[C]. International Conference on Knowledge Discovery and Data Mining, Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, California, 2001, pp. 57-66.

[15] Hyunseok Hwanag, Taesoo Jung & Euiho Suh. An LTV model and customer segmentation basedon customer value: a case study on the wireless telecommunication industry[J]. Expert Systems with Applications, 2004, 26: 181-188.

[16] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393: 440-442.

[17] R. Pastor-Satorras, A. Vespignani. Evolution and Structure of the Internet: A Statistical Physics Approach [M], Cambridge: Cambridge University Press, 2004.

[18] R. Criado , A. García del Amo , B. Hernández-Bermejo and M. Romance. New results on computable efficiency and its stability for complex networks[J]. Journal of Computational and Applied Mathematics, 2006, 192(1): 59-74.

[19] S. Wasserman & K. Faust. Social Networks Analysis [M]. Cambridge: Cambridge University Press, 1994.

[20] M.E.J. Newman. Scientific collaboration networks: Network construction and fundamental results. Phys. Rev. E64(2001), 016131.

[21] Werner Reinartz, V Kumar. On the Profitability of Long-life Customer in a Noncontractual Setting: An Empirical Investigation and Implication for Marketing Strategy Development and Implementation [J]. Industrial marketing Management, 2001, 30: 315-319.

[22] Jiayin Qi, Huaying Shu. Customer Value's evaluation, modeling and decision making [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2005 .

[23] Jiayin Qi, Huaizu Li, Huaying Shu and Liangjun Qin. Positive Research of SMC Models in IT Distribution Market Industry [J]. Chinese Journal of Management Science, Vol. 11, No. 6, 2003, pp.71-76.