Summer 6-30-2018

A Study on Online-Shopping Opinion Leaders Based on Massive Data

Chong Zhang
College of Information Sciences, Beijing Language and Culture University, China

A Study on Online-Shopping Opinion Leaders Based on Massive Data

Chong Zhang[*]

College of Information Sciences, Beijing Language and Culture University, China

Abstract: The online communities are important channels for consumers to distribute, transmit and access word-of-mouth information. However, the influence power of the users has a significant difference. This article selects about 200 Online Communities (taobao wangwang group), based on more than one million taobao real transaction data provided by nearly fifty thousand members of the communities, from the perspective of social network to build weighted and directed graph of online shopping interpersonal influence relations and users' influences model, mining online shopping opinion leaders and exploring the community online communication, spread and effect regularities to provide theoretical basis and practice guidance for word-of-mouth marketing.

Keywords: virtual community, word-of-mouth marketing, online shopping, opinion leaders

1. INTRODUCTION

With the rapid development of e-commerce, the phenomenon of online information asymmetry is growing, Numerous studies have demonstrated that word-of-mouth can reduce the degree of information asymmetry and influence consumers decision-making. However, a lot of word-of-mouth information release in online community such as BBS, QQ group. The online community is an important channel for word-of-mouth information to distribute, transmit, and access.

In the process of interpersonal interaction and word-of-mouth behavior, the existence of opinion leaders has become a consensus. Compared with normal members of groups, the opinion leaders are considered to have greater influence on consumers. In the past study of opinion leaders' recognition and feature research, the vast majority of research has focused on online community such as BBS or SNS, using the post or replies to map the relationship between users, but mining opinion leaders or highly influential users in online-shopping field are very rare, especially the research of online shopping opinion leaders based on actual purchase behavior and transaction data is blank. This article selects about 200 TaoBao wangwang groups as the research object, based on more than one million TaoBao real transaction data provided by nearly fifty thousand members of the communities, from the perspective of social network to build weighted and directed graph of online shopping interpersonal influence relations and users' influences model, mining online shopping opinion leaders and exploring the community online communication, spread and effect regularities to provide theoretical basis and practice guidance for word-of-mouth marketing.

2. CONCEPTUAL MODEL AND RESEARCH DESIGN

2.1 Conceptual model

Model's basic idea: there are a few members in the community, at first ,member 1 purchase the goods; the entire process(before-purchase, purchase and post-purchase) can produce various kinds of information related to goods and services; there are two kinds of forms of information, the first is the cognition, attitude, evaluation about goods and services, that is word-of-mouth information; The second is behavior information, the purchase behavior is also a kind of information, which can be others perceived and impact on others.

---

[*]　Chong Zhang. Email: zhangchong@blcu.edu.cn

These information are created in these communities, and spread them to other members ,such as member 2 member 3 and member 4,　through community platforms. the information receiver will be affected, showing the change of attitude and purchase behavior about goods. The most direct effect shows as following purchase, such as members 2 also purchased the same goods after member 1 purchasing them, therefore, we can say there's obvious interpersonal influence relation between member 1 and 2.Similarly, the interpersonal influence relation also can exist between others members, such as between member 3 and member 4. The interpersonal influence relation presents as net-like structure.

2.2　the Identify of the Interpersonal Influence Relationship

In behavior influence, the most direct expression is the relationship between the purchase and following purchase as member1 firstly buy a commodity, then member 2 in the same group was affected and followed to buy the same commodity over a period of time. From this view to measure consumer influence relationship will be highly accurate and objective. Therefore, this article's empirical research will take "purchase - following purchase" as important factors affecting the relationship between group members to construct the shopping influence relationship network.

3.　BUILD THE INFLUENCE RELATIONSHIP NETWORK

3.1 Introduction of the Data

This research uses the real transaction data of taobao users from taobao website (www.taobao.com),which is the largest c2c e-commerce website of China. Each data recorded a transaction. The contents of each record is as follows:

- Trading　time:　xx month xx day xx hour xx minute
- Buyer ID：the buyer's unique identifier in a transaction, it can be a string variable,
- Buyers' wangwang group: the number of wangwang group that buyers joined
- Seller ID：the seller 's unique identifier in a transaction,
- Trading goods ID: the goods' unique identifier in a transaction.
- Trading goods'name: goods display name, such as "super fantasty small-cake sleepshirts"

These variables above describes the buyer , the seller (trade subjects) and trading goods' information (the object) in any transaction in detail. "Buyers' wangwang group" is to determine whether a buyer belongs to the same online community.

3.2　Online interpersonal influence' measuring and judging

Based on the transaction data above, if the trading behavior between two users satisfy the following three conditions, then we can determine that there exist influence relationship. Judging conditions are as follows:

　　• Firstly, the users must have influence/transmission approach, which means that the users belong to the same wangwang group The community's members possess the online transmission and communication approaches.

　　•Secondly, the users can affect others' behaviors in the same community, that is "purchase － following purchase" behavior, if the users in the same community buy a goods of the same stores, so there's may exist influence relationship.

　　•Finally, the two successive purchasing behavior must satisfy a certain time range, and the interval is

generally less than 30 days.

Having the three conditions, for example: User A and B belong to a same online community. A bought of goods 1 in a store .A week after, B also bought the same goods, or buy other goods of the same stores. Then we say B followed A to purchase, so this article judged that B's purchasing behavior is influenced by A.

We calculate the interaction relationship between any two members within community, and part of the calculation results shown in table 3.1 . The first column in the table is the name of the goods; the second column is influencers' ID; the third column is the affected' ID; the fourth column is the number of influence.The first record in the table shown that in "Tea life" want want group, users "97649038" influence user "70568760" on shopping behavior by 11 times.

Table 3.1   The sample of online shopping interpersonal influence determination results

|			
Tea life	97649038	70568760	11
	70022244	408408681	9
	70022244	97649038	9
	70022244	70568760	8

3.3 Build the directed and weighted network

After measuring the relationship between interpersonal influence of online shopping, we will draw on the relevant methods of graph theory , use the directed and weighted network to describe the relationship in the community, every community will has its own network. This method will help to describe the influential relationship more intuitively, and more convenient for further study the characterizations of the relationship. In this paper, the data model is defined as G=(V,E,W), which describes the relationship between interpersonal influence of online shopping.

The nodes in the node set V represents the users from the same community.

"E" represents the edges of the network. " Edge<Vi,Vj>∈E " represents the relationship between Vi and Vj, which means if Vi has an impact on Vj, there will be an edge or connection pointed from Vi to Vj .

" W " represent the weight of the edge. " Wij" discribes the intensity of the effective relationship. In the network, the value of W represents the frequency of the influence. For example, if Vj has 5 shopping actions which are influenced by Vi, the weight Wij between Vi and Vj will be 5. The directed and weighted network not only describes the direction of influence between users, but also describes the intensity of the influential relationship.
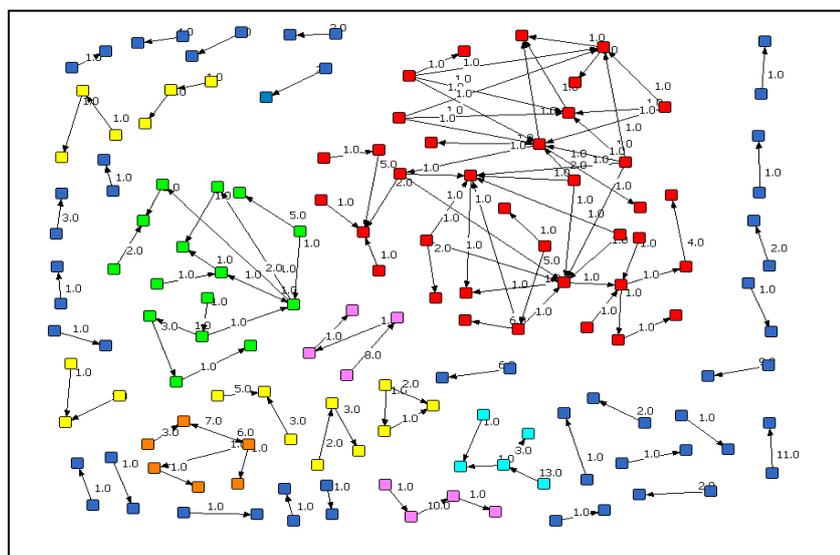
Picture 3.1 the network of the relationship

4. THE RECOGNITION OF ONLINE SHOPPING OPINION LEADERS

4.1 The Construction of the Community Members online shopping impact index

In this study, the evaluation of the influence from the community members will be described through two aspects, the breadth and the depth of the influence. In the directed network, the out-degree of nodes means the numbers of users which directly influenced by it, it measures the breadth of the direct influence. The in-degree means the number of the users who has influenced this node, measuring the acceptance of this node. High out-degree means wider breadth of the node. High in-degree means this node is much easier to be influenced, which shows that the trend following is more obvious.

In the network we have built before, the relationship of influence between users not only has direction, but also has weight "W", the weight "W" represents the intensity of the relationship. We defined the number of impact as the weight of the relationship. If user i influenced user j, there will be a connection pointed from node i to node j, $E_{ij}=1$. If the number of impact from node i to node j is n, the weight of influence relationship between i and j will be n. The weight reflects the depth of the impact between member i and j, depicts the loyalty or stickiness.

Therefore, we construct an evaluation index of the influence on online shopping of the members from the community. This index shows the members' influence in the whole network.

4.2 The reorganization result of high influential members

In this study, we measured the influence index of 50 thousand members from 2 hundred communities. We have found that the exponents also distribute as long-tailed. The minority members as head has stronger influence, and the tail represents most of ordinary customers.

Table4. 1 the influence index of community members

Out-degree (Influence Breadth ）		
	max	79
	min	0
	mean	4.32

From the table, we can see that the max out-degree of the virtual community reached at 79, which means that some member in this virtual community can influence 79 other members, the breadth of influence is wide. Fwhich means there might be some members who have much stronger influence.

4.3 High impact users identification test based on user interview

In order to test the validity and accuracy of the identification methods, we use a method that make user survey in the small scale to verify the high impact users. This method is used because the test work has some difficulties itself. First, the high impact users belong to minority groups. For the identification rate is 3 percent, which means there may be 3 users in 100 of Taobao users. It is very difficult to find a large number of effective samples in reality. If we choose to deliver the questionnaires to random users, most of them will be meaningless for the test. This method is inefficient. Second, the test work shown in this paper need to identify the users' identities, which means we have to get the users' Taobao ID. Then we can compare the survey result with the identification result, which also has difficulties in operation. Therefore, the author finally choose to use the first

method , interviewing in the small scale, to simply test the identification result with the help from Taobao's user survey department.

4.3.1 The interviewee

The interviewees are the users whose node degrees are more than 4. These users influenced others or have been influenced in the measuring result. They belong to the " charactistic users". We finally got 19 users to interview.

4.3.2　Interview outline

The interviews are conducted around the following questions, and the interview time is about 10 to 20 minutes.

1) How many times per month do you shop on Taobao?

2) Do you ask for advice from colleagues or friends while shopping on Taobao?

3) How do you get advice from others?

4) Do you shop because of colleagues or friends' recommendation in this two month?

5) Do you think it is more satisfied to shop following the colleagues' recommendation?

6) Do you still pay attention to colleagues or friends' advice and experience sharing after shopping?

7) Do you share you shopping experiences during or after shopping?

8) How do you share and communicate your shopping experience with others?

9) Did someone shop because of your recommendation in this two months?

we compared the users' answers with the measuring result. If the users answer " don't know " or " don't remember ", the interview will be invalid. If the answers are " no " or something like this, we determine that the degree of the node is 0 in this paper. If the answers are "occasionally", " sometimes", " no many" or something, the degrees of nodes will be 1 to 3. If the answers are " absolutely", " always" and so on, the degree of the nodes will be 4 and above. Finally the investigation result will be compared with the measuring result to test.

4.3.3　Analysis of the Test Result

In these 19 users' interviews, one is invalid, and the following is the other 18 analysis results. the test of accuracy, 15 users' interview results are the same as the measuring result, the other 3 are different. The accuracy rate is 83.3 percent, which means that the high impact users' identification in this paper has high accuracy.

For the ways to spread the word-of-mouth information, the most frequency method is "oral communication among colleagues and friends around". The interviewees said, " they asked for advice from the colleagues and friends " and " they always recommend the satisfactory goods to colleagues and friends" . Another common way is spreading among friends on some instant messengers like QQ, MSN, Wangwang and so on. The interviewees said , " everybody has different hobbys and interests, colleagues and myself maybe not interested in the same things, so they like to share the shopping information with friends who interested in the same things on QQ and MSN. "

4.4　Category attribute characters of members with high influence

Many researches about opinion leaders indicate that generally they have some professional knowledge and high influence in some professional field. So, when shopping online in the virtual community, are those members with high influence possess the same characters, which means that dose their influence only work in some specific field? In chapter one and two of this research, we only consider the whole influence of members, but don't consider the professional field their influence work in. In this character, we will focus on the specialty of high influential members which we call the category attribute character. It subdivides influence of members

into attribute dimension, which shows the characters of members with high influence accurately.

Research method: further analyze those members with high influence whose out-degree is not null, refine the number of total influence according to category, calculate the number of total influence of this member and the number of influence under all categories, and find the category with the maximum number of influence. If that maximum number owns equal or greater than a half proportion of the number of total influence, this member has higher influence in that category and is the highest influential member during this study. Or, the influence of this member is a little dispersive or average, and category attribute is not obvious.

On the whole, in 11583 members whose out-degree is bigger than zero, there are 10988 members whose category with the maximum number of influence is equal or greater than a half proportion, which means about 95 percentage of members have category attribute on influence. To exclude the disturbance of those members whose out-degree is 1 or 2, we only consider those members whose out-degree is equal or greater than 3. The outcome shows that in 5034 members, there are 4439 members whose category with the maximum number of influence is equal or greater than a half proportion, which means about 88 percentage of members have category attribute on influence.

In a word, statistics shows that most (about 88%) high influential members have obvious category attribute , which means the specialty always reflects on one kind of category, and they are more familiar or professional with one kind of category, instead of with many kinds of categories. We classify and collect those high influential members. From the diagram, we can see that high influential members from these fields such as clothes for female, cosmetic, food and maternal and child supplies have almost a half proportion of all fields. Those categories are goods for use and need higher degree of trust.

Table 4.2 top10 categories forming high influential members

Category ID	Category name	Quantity of high influential members	Percentage
16	clothes for female/ladies boutique	657	0.1487
1801	beauty and skin care/bodybuilding/essential oil	446	0.1005
50002766	snacks/nut/tea/specialty	430	0.0968
50014812	diaper/care/feeding/cart bed	394	0.0887
50008165	clothes for child/shoes for child/parent-child	323	0.0727
29	pet/food and supplies for pets	261	0.0588
50010788	cosmetic/perfume/tools for make-up	157	0.0353
25	toy/model/comic/early education/puzzle	156	0.0351
50008090	3C digital accessory market	152	0.03424
50016422	grain and oil/vegetable and fruit/aquatic/fast food	132	0.02974
50022517	clothes for pregnant/supplies for pregnant/nutrition	118	0.02658
35	milk powder/solid food/nutrition	113	0.02546

5. CONCLUSION

Though the view of social network, this study is based on the real transaction data of Taobao user. It judges the relationship between people in the community by the behavior of buy and follow-buy. It structures the directional weighted network which displays the influential relationship between people in the community and also structures the judgment model of influence of members in the community, which identifies high influential

members in the community. By verifying those high influential members in the community, the outcome shows that the identification of high influential members has high accuracy which is more than 83%. The further study of high influential members shows that those members' influence works on one specific classify of goods, which means that they have the ability to influence others in some professional fields.

ACKNOWLEDGEMENT

REFERENCES

[1]   Momtaz N J,et al. Identifying opinion leaders for marketing by analyzing online   social networks[J].International Journal of Virtual Communities and Social Networking,2011,3(1):43-59.

[2]   Hazeldine M F.An exploratory role analysis of opinion leaders,adopters,and communicative adopters with a dynamically continuous innovation[J].Journal of Applied Business Research,2010,26(4):117-130.

[3]   Forbes L P. Does Social media influence comsumer buying behavior? An investigation of recommendations and purchases[J]. Journal of Business&Economics Research,2013,11(2):107-112.

[4]   Rogers E M.Diffusion of innobation[M].New York:Free Press,2003.

[5]   Shoham A and Ruvio A. Opinion leaders and followers:A replication and extension[J]. Psychology & Marketing,2008,25(3):280-297.

[6]   Huffaker D. Dimensions of leadership and social influence in online communities[J].Human Communication Research,2010,36(4):593-617.

[7]   Doumit G,et al. Opinion leaders and changes over time: A survey[J]. Implementation Science,2011,6(1):117-123.

[8]   Ghosh R,Leman K. Predicting influential users in online social networks//Proceedings of the 4th KDD workshop on Social Network Analysis. Washington.DC,USA,2010

[9]   Malmgren R D,Stouffer D B,Motter A E,Amaral L A N. A Poissonian explanation for heavy tails in e-mail communication. Proceedings of the National Academy of Sciences.2008,105(47):18153-18158.