User Behavior and IM Topology Analysis

Q. Yan, X. Huang
School of Economics and Management, Beijing University of Posts and Telecommunications,
Beijing, China, 100876
Email: yq_10@sohu.com

***Abstract*—The use of Instant Messaging, or IM, has become widely adopted in private and corporate communication. They can provide instant, multi-directed and multi-types of communications which make the message spreading in IM different from those in WWW, Blog and email systems. Groups have great impacts on message spreading in IM. The research demonstrates the power law distribution of groups in MSN with parameter γ ranging from 0.76 to 1.22. Based on an online survey, IM user behavior is analyzed from the aspects of message sending/receiving and contacts maintaining. According to the results, degree distribution of users has a peak value and doesn't present power law character. This may indicate that social networks should be a prospective direction for the research on IM topology.**

***Index Terms*—instant messaging, topology, user behavior**

I. INTRODUCTION

The use of Instant Messaging, or IM, has become widely adopted in private and corporate communication. A recent report from IDC estimated that around 12 billion instant messages are sent each day [1]. Users may use IM to find their friends and partners on networks and exchange information and play games. So far there are many popular IM such as MSN, Tencent QQ, Fetion of China Mobile and so on. IM based communications usually have the following characters:

- Instantaneity: messages are exchanged in real time;
- Multi-direction: there are point to point and point to multi-points communications;
- Diversity: the messages exchanged can be in the forms of text, voice or video and etc.

According to the survey of iResearch Consulting Group, IM in China have experienced the following four phases, as illustrated in fig. 1. Nowadays, IM have had enormous numbers of users. For example, the number of users of Tencent QQ has reached a hundred millions till year 2007. IM users constitute a giant virtual community. In such community, users present new behavior characters powered by the IM technologies. Groups and contacts present the relation between IM users in the virtual community.

This work was supported by the National Natural Science Foundation of China (Grant No. 70671010) and the Key Project of Chinese Ministry of Education (Grant No. 108011).

Groups are virtual organizations composed of IM users, which are classified according to specialties, domains or interests and so on. Contacts are partners of IM users in digital space. The way of joining or quitting a group, adding or deleting a contact directly affects the topology of IM as well as the spreading of messages in digital space.

In this paper, we firstly study the group topology in IM through analyzing the actual data of MSN, and then analyze the user behavior and related topology based on an online survey. The results point out the power law distribution of the group's degree while the user topology presents different characters.

The remainder of this paper is organized as follow. Section 2 introduces the related researches. Section 3 analyzes the topology of groups in IM. The online survey and related analysis about IM user behavior are described in section 4. Finally, our conclusions are outlined in section 5.

II. RELATED RESEARCHES

In the past few years, complex networks theory has been widely used to study the relationship between the topology and evolvement of the Internet. Some researches pointed out the power law distribution of WWW and Internet [2][3][4]. These achievements have had great impacts on the study of the evolvement and security of the Internet. But so far, there are only a few researches about the IM topology based on complex networks.

Regarding IM topology, there are mainly two kinds of researches. One is case study of IM topology aimed to analyze the statistical characters of IM networks. The other is the study of the evolving models of IM.

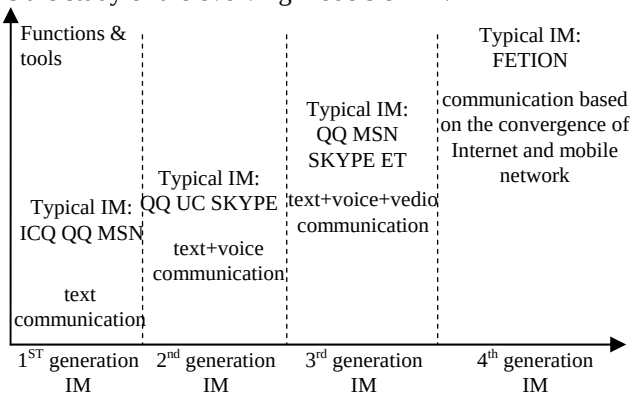

Figure 1. Development of IM in China

Current case studies mainly focus on whether IM networks have the characters of small world or scale free networks. Smith studied the IM database of nioki.com and constructed a social network [5]. He depicted the IM network as a directed graph with 50158 nodes and about 0.5 million edges. Smith pointed out that nioki.com IM network presented small world characters with cluster coefficient 0.33. Nioki.com IM network is also a scale free network with in-degree 9.1 and out-degree 8.2. In 2005, Cheryl D. Morse and Haining Wang analyzed another IM network [6]. They also pointed out that IM network has the characters of small world and scale free networks.

Yao Yuanyuan studied the evolving characters of IM networks based on Clique-VGBA model. Yao analyzed the in-degree and out-degree of nodes in IM network and pointed out that the degree presented power law characters [7].

Jure Leskovec and Eric Horvitz analyzed the static and dynamic communication and network characteristics of all 240 million users of Microsoft Instant Messenger who were active during the month of observation, (June 2006) [8]. They found that about half of all conversations occurred between 2 people. The complementary half of conversations involved 3 or more people. They found that longer paths exist in the graph, with lengths up to 29. They also found that the network is well clustered, with a clustering coefficient that decays with exponent -0.37. They analyzed the characters of the IM conversations including the number of users per session, number of buddies per user, Conversation duration and so on. Several power law properties were pointed out in their research, such as the power law distribution of conversation duration with exponent 3.6.

Aforementioned researches all focus on the degree distribution of nodes (in other words, IM users). However, groups are very important communication communities for users in IM. Groups have great impacts on the message spreading in IM. So we analyzed not only the degree distribution of users but also that of groups in this paper. This case study is based on the data collected from MSN website and an online survey.

III. GROUP TOPOLOGY IN IM

Groups are virtual organizations composed of IM users, which are classified according to specialties, domains or interests and so on. The number of users varies in different groups. Each user can join one or more groups or join none of the existing groups. We firstly analyze the group topology since groups have played important roles in the message spreading in IM. The following study is based on MSN which is one of the most popular IM at present.

In MSN, groups are organized as a tree structure. For example, the structure of group "automotive" is depicted as fig. 2.

By the end of Jan 2006, there are 17 types of groups and 113 sub-types in MSN as described in table 1.

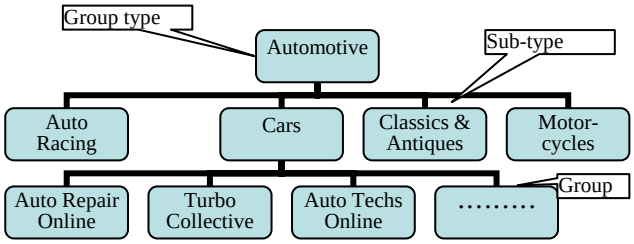

Figure 2. Group organization in MSN

TABLE 1. GROUPS IN MSN

Group Type	Number of Sub-types
Automotive	4
Computers & Internet	7
Games	11
Home & Families	5
Money & Investing	4
Organizations	8
Places & Travel	11
Schools & Education	6
Sports & Recreation	4
Business	7
Entertainment	8
Health & Wellness	3
Lifestyles	5
News & Politics	3
People	9
Religion & Beliefs	9
Science & History	9

During the experiment, we collected the data about groups in MSN. The data is up to the end of Jan 2006. In MSN, groups can be in more than 20 kinds of languages. We only collected the data of groups in English. The primary analysis of the collected data is described in table 2.

Fig. 3 shows the user proportion of different kinds of groups. For example, the total user proportions of "Religion & Beliefs", "people" and "entertainment" reach 70%. In MSN, 17.6% of groups take up nearly about 70% of the total users.

TABLE 2. STATISTIC OF GROUPS AND USERS IN MSN

Group Catalog	Number of Groups	Number of Users	Users per Group	Percent to the Total Groups	Percent to the Total Users
religion&beliefs	19889	11010949	553.62	0.0374	0.3466
people	206206	7055121	34.21	0.3873	0.2221
entertainment	78062	3510938	44.98	0.1466	0.1105
place&travel	22396	1454487	64.94	0.0421	0.0458
games	25214	1290913	51.20	0.0474	0.0406
money&investing	2684	997653	371.70	0.0050	0.0314
lifestyle	24067	976226	40.56	0.0452	0.0307
computer&internet	10910	939257	86.09	0.0205	0.0296
home&families	54825	877503	16.01	0.1030	0.0276
sports&recreation	21247	752157	35.40	0.0399	0.0237
health&wellness	9596	747701	77.92	0.0180	0.0235
school&education	20902	634916	30.38	0.0393	0.0200
organization	12466	357656	28.69	0.0234	0.0113
automotive	6131	350626	57.19	0.0115	0.0110
business	9807	336069	34.27	0.0184	0.0106
news&politics	4684	294484	62.87	0.0088	0.0093
science&history	3272	183688	56.14	0.0061	0.0058
Totoal	532358	31770344	59.68		

According to fig. 4, type "people" and "entertainment" take up more than 50% of the total groups.

Suppose k is the degree of one group which means the number of users included in the group. Denote by $P(k)$ the probability that the users number of a group is k. In MSN, $P(k)$ has the following characters:

For all the groups in MSN, $P(k) \sim k^{-\gamma}$, $\gamma = 1.22$ as illustrated in fig. 5.

In each kind of groups, $P(k)$ presents power law distribution. For example, in "Automotive" $\gamma = 0.76$ as illustrated in fig. 6. In "business", $\gamma = 0.87$ as illustrated in fig. 7. In "entertainment", $\gamma = 1.18$ as illustrated in fig. 8.

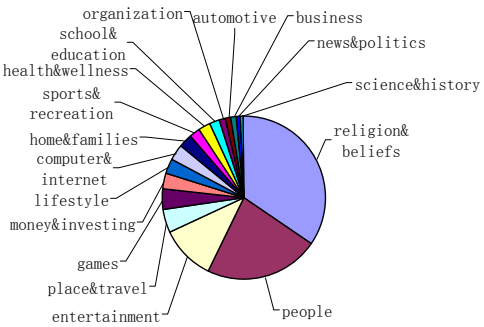

Figure 3. User proportions of different group types

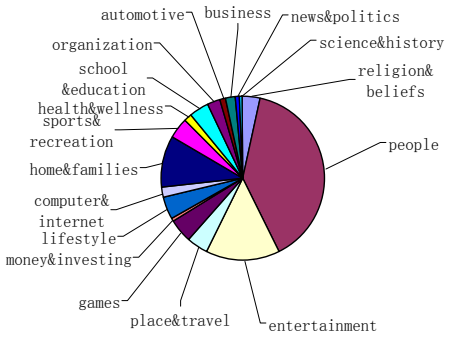

Figure 4. Group proportions of different group types

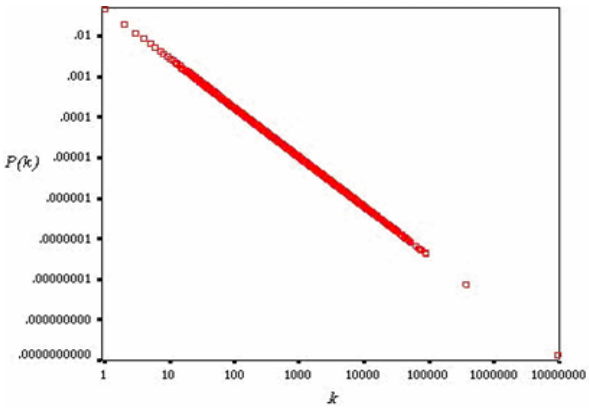

Figure 5. $P(k)$ Distribution of all the groups

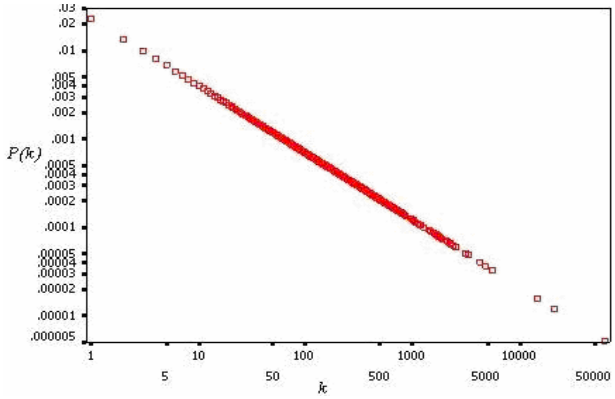

Figure 6. $P(k)$ Distribution of groups in "Automotive"

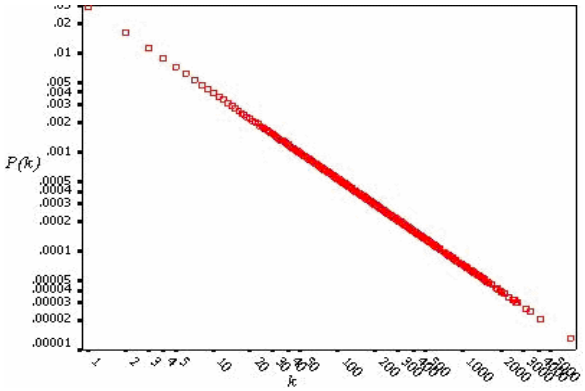

Figure 7. $P(k)$ Distribution of groups in "business"

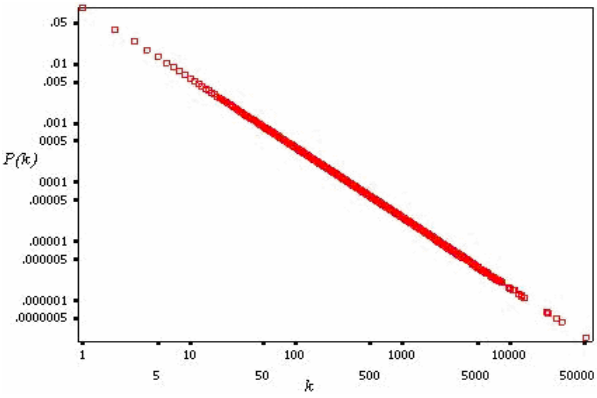

Figure 8. $P(k)$ Distribution of groups in "entertainment"

Ⅳ. USER BEHAVIOR AND TOPOLOGY ANALYSIS

Besides group topology, user behavior characters and related topology are also very important for the message spreading in IM. In our previous work, we have introduced the user topology briefly [9]. In this paper, we provide more details about the user behavior analysis. This is conducted from three aspects: 1) Basic properties of IM users, 2) behavior regarding message sending/receiving, 3) behavior regarding contacts maintaining. These aspects will be discussed in the following subsections.

A. Basic Properties of IM Users

The survey collected 221 valid questionnaires. Among all the survey objects, 63.3 percent are male and 36.7

percent are female. 68.3 percent of these users are with educational degree of bachelor. The main majority are young people with age from 19 to 25, as illustrated in fig. 9. Fig. 10 illustrates the distribution of average online duration every day. Table 3 describes how long the survey objects have used IM software.

B. Behavior Regarding Message Sending/Receiving

The survey shows that IM users mainly use point-to-point mode to communicate with others. Fig. 11 depicts the frequency that a user sponsors a talk with IM contacts. Fig. 12 depicts the frequency a user posts messages actively in a group. In both modes, males are more active than females. Fig. 13 illustrates the talking frequencies of users with different genders.

According to the survey, users mainly use IM in point-to-point mode merely to keep in touch with others. Work and study are the second purposes for using IM software. But in group mode, users mainly use IM of work and study purposes, as illustrated in fig. 14 and 15.

Users talk more frequently in point-to-point and group mode when the online times get longer. The talking frequency is correlated to online duration, which is significant at 0.01 significance level by chi square method. In group mode, users browse the messages from others only occasionally if they don't sponsor a talk actively. The browsing frequency is correlated to online duration, which is significant at 0.05 significance level by chi square method.

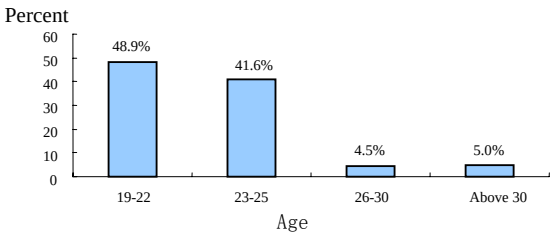

Figure 9. Age distribution of the survey objects

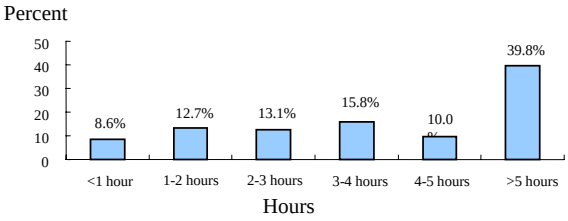

Figure 10. Distribution of average online duration per day

TABLE 3. IM USING DURATION

duration	percentage(%)	cumulative percentage(%)
more than 5 years	38.5	38.5
4-5 years	19.5	58.0
3-4 years	27.1	85.1
2-3 years	8.6	93.7
1-2 years	5.4	99.1
less one year	0.9	100.0

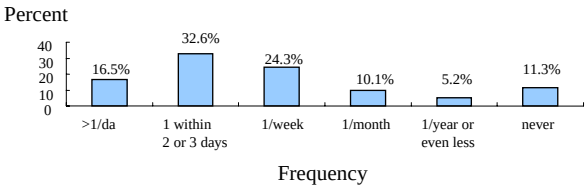

Figure 11. Frequency a user sponsors a talk

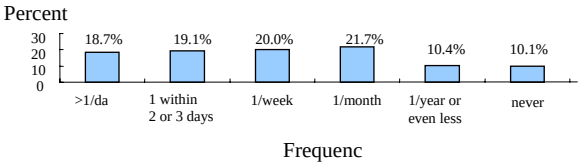

Figure 12. Frequency a user posts messages actively in a group

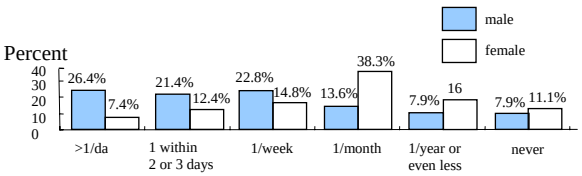

Figure 13. Talking frequencies of users with different genders

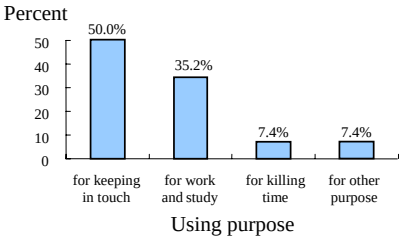

Figure 14. Using purposes in point-to-point mode

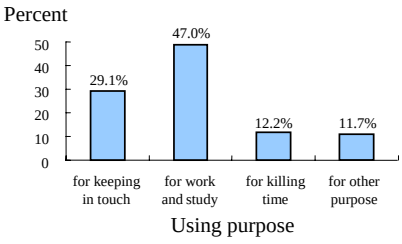

Figure 15. Using purposes in group mode

So far, some IM tools provide the function of "stealth". If one user uses stealth mode when he/she login into IM, his/her online status is not available for others. This function is useful if users don't want to be interrupted by strangers in cyberspace when they are using IM tools. Users in stealth mode mainly use IM tools to communicate with relatives. Their talking frequency is obviously lower than that of others who don't use stealth mode. In our survey, 79.2 percent of users usually select stealth mode when they use IM tools.

Garbage information (such as advertisements and pornography linkages) has been one of the main negative factors hindering the development of IM market. In our survey, most people received garbage information very month or even more frequently, as illustrated in fig. 16.

C. Behavior Regarding Contacts Maintaining

Among all contacts of the IM users, 53.5% are relatives and friends, 28.3% are work partners and 6.1% are friends in cyberspace. Fig. 17 illustrates the type distribution of IM users' contacts. Fig. 18 illustrates the distribution of contacts amount of IM users. Similarly, 59.1% of the talking peers are relatives and friends, 21.3% are work partners and 9.6% are friends in cyberspace as illustrated in fig. 19.

Regarding the attitude to strangers' invitation, 88.3% of IM users require inviters pass the identification before chatting online. Even though, 30% IM users in our survey select to refuse all invitations from strangers on the networks. And 63.9% of IM users say it depends whether accepting or refusing strangers' invitation, as illustrated in fig. 20. In addition, female users are more apt to refuse invitations from strangers than male users. And the frequencies of inviting strangers actively by male users are higher than that by female users. The attitude to strangers is correlated to user's gender, which is significant at 0.01 significance level by chi square method.

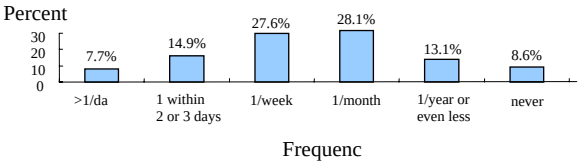

Figure 16. Frequency of receiving garbage information

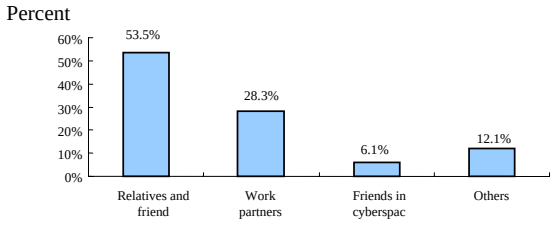

Figure 17. Distribution of IM users' contacts

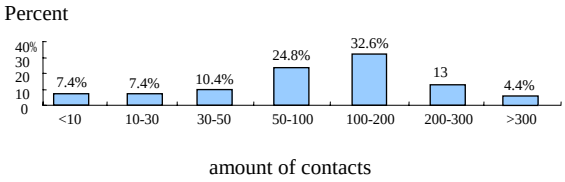

Figure 18. Distribution of contacts amount

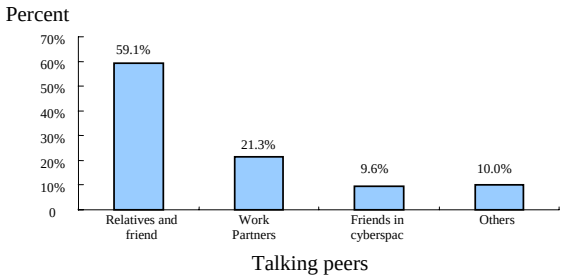

Figure 19. Talking peers distribution of IM users

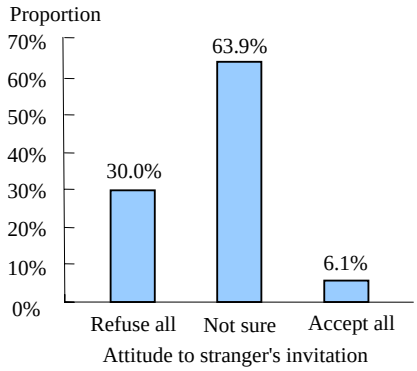

Figure 20. Distribution of the attitude to stranger's invitation

The cliques based on IM tools also present to be stable according the survey. The amount of contacts may maintain unchanged in a quite long time. As fig. 21 illustrated, 31.7% of IM users only add contacts for one time every year on average. And 53.5% of IM users only remove contacts for one time very year. The stabilization of the cliques in cyberspace is also quite similar with that in real life.

Fig. 22 is the statistical result of the numbers of groups every user connects to. The survey also shows that users with age from 19 to 25 usually connect to more groups than others, male users connect to more groups than female users, and more contacts one user has, more groups he/she connects to.

Suppose k' is the degree of a user which means the amount of contacts of the user. Denote by $P(k')$ the probability that the contacts number of a user is k'. Based on the online survey, $P(k')$ presents different distribution character compared with $P(k)$. $P(k')$ has a peak value and doesn't present Power Law character, as illustrated in fig. 23. In addition, according to the survey, more contacts one user has, more groups he/she creates. The amount of contacts is correlated to the amount of groups one user creates, which is significant at 0.01 significance level by chi square method. This implies that users with higher degrees are more apt to be administrators of groups.

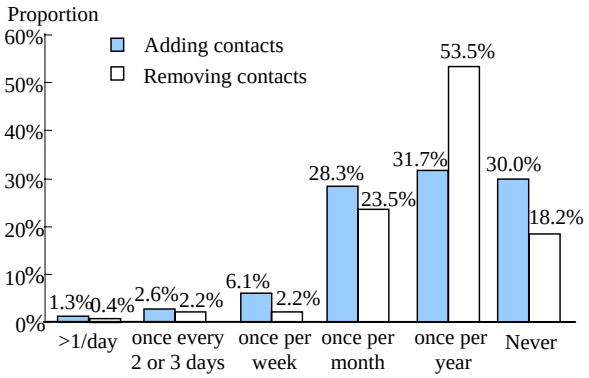

Figure 21. Proportion distribution of adding or removing contacts

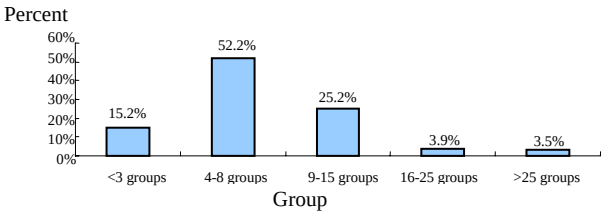

Figure 22. Statistical result of the numbers of groups every user connects to

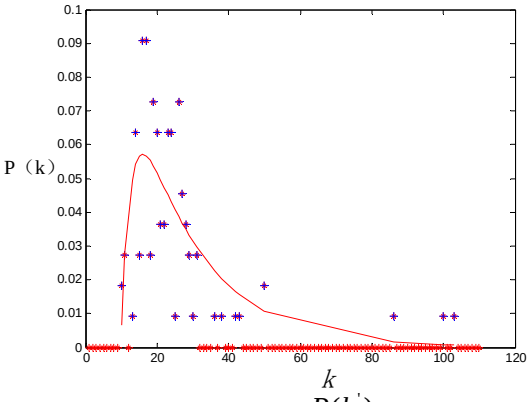

Figure 23. Curvilinear fit of $P(k^{'})$ distribution

Because in our research most contacts of IM users are relatives and friends, we suppose the topology of IM users has the character of social networks. Since the number of samples in our survey is limited, this proposition needs further verification. However, social networks should be a prospective direction for the research on IM topology.

Ⅴ. CONCLUSIONS

IM have become very popular communication tools nowadays. Numerous users and groups of instant communication systems constitute a complex network of information exchange.

In IM, there are point to point and point to multi-points communications, which makes the message spreading in IM different from those in WWW [10] and Blog space [11]. In addition, the character of instantaneity of the communications in IM decides that message spreading rules for email systems [12] can not be applied directly to IM. These differences may be derived from the different topology characters. So far most researches focus on the degree distribution of nodes. However, groups also have important roles in message spreading in IM. So in this paper we analyzed not only the degree distribution of users but also that of groups.

Our research demonstrates the power law distribution of groups in MSN with parameter γ ranging from 0.76 to 1.22. Based on an online survey, IM user behavior is analyzed from the aspects of message sending/receiving and contacts maintaining. According to the results, degree distribution of users has a peak value and doesn't present power law character. This result may caused by the limited numbers of samples in our survey. But this also indicates that social networks should be a prospective direction for the research on IM topology.

REFERENCES

[1] IDC Market Analysis. Worldwide Enterprise Instant Messaging Applications 2005-2009 Forecast and 2004 Vendor Shares: Clearing the Decks for Substantial Growth, 2005.

[2] Lada A. Adamic, Bernardo A. Huberman and etc. Power-Law Distribution of the World Wide Web. *Science*, 287, 2000: 2115a.

[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. Computer. *Communication Review*, 1999,29 (4):251-262.

[4] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett*. 87, 258701 (2001).

[5] Smith R D. Instant messaging as a scale-free network. *e-print cond-mat*/0206378. 2002.

[6] Cheryl D. Morse and Haining Wang. The Structure of an Instant Messenger Network and its Vulnerability to Malicious Codes. *SIGCOMM Poster* .2005.

[7] Yao Yuanyuan. Internet topology study and its application in IM network modeling. *Dissertation*. China: Zhengzhou University.2006.

[8] Jure Leskovec, Eric Horvitz. Worldwide Buzz: Planetary-Scale Views on an Instant-Messaging Network. *Microsoft Research Technical Report. MSR-TR-2006-186*. 2007,6.

[9] Yan Qiang, Shu Huaying, Huang Xiaoyan and etc. Complex network demonstration: topology of instant communication systems. *In Proc. of International Conference on Systems, Man and Cybernetics, Quebec, Canada*. Oct. 2007, 3245-3249

[10] Yan Qiang, Shu Huaying. The way to control the harmful information on the Internet. *High Technology Letters*. 2006,5: 60-63.

[11] D. Gruhl, R. Guha, David LibenNowell and A. Tomkins. Information Diffusion through Blogspace. *WWW2004*, May 17–22, 2004, New York, New York, USA. ACM 158113844X/04/0005.

[12] M. E. J. Newman, Stephanie Forrest, and Justin Balthrop. Email networks and the spread of computer viruses. *Phys. Rev. E*, 66(035101), 2002.

Qiang Yan has ever studied economics, medicine and got bachelor and master degree respectively in these fields. Later, he entered Beijing University in China to research on information and network security and got PhD degree in this subject. He has more than ten years' experiences in the field of computer science with the working experience in Microsoft and ZTE corp. During 2004 to 2006, Dr. Yan pursued the postdoctoral research in Beijing University of Posts and Telecommunications. He is now associate professor of the School of Economics and Management. His current researches focus on Internet security governance, complexity system theory and IT service management.

Xiaoyan Huang is a graduate student of School of Economics and Management, Beijing University of Posts and Telecommunications. Her research focuses on management science and engineering, telecommunication operation management.