I. INTRODUCTION

The research on the social networks of real world has a long history\cite{1}. As early as in 1920s and 30s, J. L. Moreno began to study the acquaintance groups\cite{2}. Studies focusing on various kinds of social networks have come out later, among which the most famous one is probably the “Six Degree of Separation” experiment by Milgram in 1965\cite{3}. In this experiment, participates were asked to pass a letter to an assigned target in Boston by delivering it to one of their acquaintances. For the letters that were successfully delivered, they passed through five to six persons on average. Most of the early studies were based on surveys, so
the scope was limited and the results were not precise enough. For example, when studying the friendship networks, the definition of “good friend” may be quite different from person to person. Recently, by using the databases on the internet, many large social networks have been studied without doing surveys, such as the science collaboration networks [4] and the collaboration network of film actors [5]. Besides the study of social networks, the information and biological networks have also been investigated, e.g., the internet [6], World Wide Web [7] and protein network [8]. Some tools and algorithms were developed to reveal the characteristic properties of networks belong to a specific type. For instance, several algorithms [9, 10] have been developed to find the communities or sub structures of social and biological networks.

In this paper, we will study the social network in the Complex System Summer School held in Beijing, China 2008 by surveys. The evolution of the networks is presented, as well as the correlation between individuals and the community structures.

The Santa Fe Complex System Summer School 2008 - Beijing (CSSS2008 hereafter) was held from June 30 to July 25, which enrolled 49 students in total, including 26 Chinese students (14 Males, 12 Females) and 23 foreign students (11 Males, 12 Females). Before the commencement of the summer school, most Chinese students received an English training course for one week. Among the foreign students, about half of them were from USA; others were from Argentina, South Korea, India, Singapore and European Union. During the summer school, every two students shared a standard room, one from Mainland China and the other from overseas. Besides, there were four student helpers from Institute of Theoretical Physics, CAS, who lived and studied together with the participants. The participants and the student helpers attended lectures for about 5 hours per day and ate together for meals. All participants were required to do projects in groups of 3 to 7 students, which should contain students from Mainland China and overseas. Members of the same group had dinner together outside the campus twice a week. Participants could enjoy various social activities in their spare time, including dancing, traveling etc. Most participants didn’t know each other at all before the summer school. Their personal relations were developed during these four weeks. Therefore, this summer school was a well defined system to investigate the emergence and evolution of friendship.

Our data were collected from surveys. Three same surveys were carried out from the beginning of the second week to the last day of the summer school. The first survey was
conducted from July 6 to July 11, the second one from July 13 to July 18 and the third one was conducted on July 25. The names of 48 students (one student was sick during the first week, thus was excluded) and 4 student helpers were listed in the questionnaire (see Appendix for detail). In each survey, the informants were asked to evaluate the familiarity between them and all other participants in five levels from “Very Unfamiliar” to “Very Familiar”. There’s no definition or description for these five levels, so the informants can have their own judgments. We got around 40-50 responses for each survey. For those who didn’t respond, the evaluations on him/her by others will be used as his/her response. After questionnaires were collected, we used score points 1 to 5 to represent the five levels from “Very Unfamiliar” to “Very Familiar”, thus we can generate a directed weighted social network for each survey. For simplicity, we will convert them into undirected unweighted networks in the following discussion.

II. BASIC PROPERTIES OF THE NETWORKS

To construct the undirected unweighted networks of CSSS2008, we use vertices and the links connecting them to represent the informants and the relations between them separately. The requirement for a link to established between node $i$ and $j$ is that $i$ chose “Very Familiar” (5 points) with $j$ in his questionnaires, and $j$ chose “Familiar” (4 points) or “Very Familiar” with $i$, or vice versa. The networks generated by the three surveys are named as Network 1, 2 and 3, according to the time order. Their topological structures are shown in Fig. 1. Table 1 lists the structural properties of these three networks.

![FIG. 1: The topological structures of Network 1, 2 and 3 (From left to right).](image)

The degree of a node is the number of edges of that node. In the friendship social network, degree means the number of friends of that node, thus $\langle k \rangle$ is the average number
TABLE I: Basic Properties of CSSS 2008 Social Networks

<table>
<thead>
<tr>
<th></th>
<th>Network 1</th>
<th>Network 2</th>
<th>Network 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average degree $\langle k \rangle$</td>
<td>9.23</td>
<td>14.58</td>
<td>18.31</td>
</tr>
<tr>
<td>Clustering coefficient $C$</td>
<td>0.33</td>
<td>0.45</td>
<td>0.51</td>
</tr>
<tr>
<td>Average shortest path length $L$</td>
<td>2.08</td>
<td>1.76</td>
<td>1.65</td>
</tr>
<tr>
<td>Diameter $D$</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Modularity $Q$</td>
<td>0.34</td>
<td>0.27</td>
<td>0.25</td>
</tr>
</tbody>
</table>

of friends of all vertices of that network. Furthermore, the spread of the node degrees can be characterized by the degree distribution $P(k)$, which gives the fraction of nodes in the network with degree $k$. Figure 2 shows the degree distributions of the three networks. As time passed by, the participants became more and more familiar with each other, which can be reflected from the evolution of both $\langle k \rangle$ and $P(k)$: The average number of friends of the participants $\langle k \rangle$ goes from 9 in Network 1 to 18 in Network 3; The degree distribution $P(k)$ also moves to the large $k$ side, which indicates the portion of people has large number of friends increases.

In social networks, a friend of your friend has a big chance to be your friend, the clustering coefficient $C$ is often used to quantify this probability. The local clustering coefficient of node $i$ is defined as $C_i = \frac{2N^i_\Delta}{k_i(k_i - 1)}$, where $N^i_\Delta$ is the number of triangles connected to node
\(i\) and \(k_i\) is \(i\)'s degree. \(C_i\) measures the closeness of \(i\)'s friends. The clustering coefficient of the network is the average of \(C_i\) over all nodes, \(C = \frac{\sum_i C_i}{N}\). From Table I it can be observed that clustering coefficient \(C\) increased in the process of time, meaning that on average, the relations between the friends of a participant became closer and closer. The network’s diameter and the average length of the shortest path decreased with time. The evolution of all these quantities shows that with the passing of time, the participants became more and more familiar.

III. CORRELATION BETWEEN RECIPROCAL EVALUATIONS

The friendship network should be directed, e.g., you are not necessary the best friend of your best friend. For the facility of data processing and discussions, we converted the directed networks to undirected ones. But is it reasonable to do so? To answer this question, the correlations between reciprocal evaluations are examined. Here correlation means the linear relation between one’s evaluations on others and their feedbacks. If participant \(i\)'s evaluation on another participant \(j\) is exactly the same as \(j\) evaluation on \(i\), then the correlation between the reciprocal evaluations of \(i\) and \(j\) is perfect. But if \(i\) chose “Very Familiar” with \(j\), while \(j\) chose “Very Unfamiliar” with \(j\), then the correlation of \(i\) and \(j\) is bad. In this section, the Pearson’s correlation coefficient (PCC hereafter) is used to measure the correlation between reciprocal evaluations, which is defined as:

\[
r_k(a, b) = \frac{\langle ab \rangle - \langle a \rangle \langle b \rangle}{\sqrt{(\langle a^2 \rangle - \langle a \rangle^2)(\langle b^2 \rangle - \langle b \rangle^2)}}
\]

where \(a\) and \(b\) are the evaluation of \(i\) on \(j\) and \(j\)'s evaluation on \(i\) respectively. \(\langle a \rangle(\langle b \rangle)\) is the average evaluation of \(i(j)\) on all other participants. \(-1 \leq r \leq 1\). \(r_k(a, b) = 1\) stands for perfect correlation between the reciprocal evaluations of \(i\) and \(j\), while \(r_k(a, b) = 0\) indicates that there’s no correlation between \(i\) and \(j\).

For the convenience of our further discussion, participants are divided into four groups, namely Non-Chinese Females (NCF), Non-Chinese Males (NCM), Chinese Females (CF) and Chinese Males (CM). PCC of every participant is calculated for the three surveys, the mean values are all around 0.6, with a slight increase in the last survey. So the correlations between reciprocal evaluations are quiet good, thus our method to convert the system into undirected networks is reasonable and reliable. Fig. 3 shows all participants’ PCC in the
second survey.

Now we define a new function $r_d$ as

$$r_d(i, \text{group } x) = r_k(i, \text{students in group } x) - r_k(i, \text{all participants except } i)$$

(2)

here $i$ can be any participant of CSSS2008 and $x$ is a group out of NCF, NCM, CF and CM. If $i$ belongs to group $x$, then $r_d(i, \text{students in group } x)$ should be replaced by $r_d(i, \text{students in group } x \text{ except } i)$. A positive(negative) value of $r_d(i, \text{group } x)$ indicates the correlation between $i$ and group members of $x$ is better(worse) than the correlation between $x$ and everyone else in the summer school. $r_d$ as a function of every participant and the four groups are calculated for all the three surveys. It is surprising to find that $r_d(i, \text{the group } i \text{ belongs to})$ is always positive in all the three surveys, meaning that the correlation between within-group reciprocal evaluations is always better than the correlation between global reciprocal evaluations. $r_d$ of the second survey is shown in Fig. 4 as an example. From our result, it is found that people belong to the same sex and ethnic group can better evaluate the friendship between them, compared with evaluating the friendship with people of different sex/ethnic groups.
FIG. 4: \( r_d \) as a function of each individual and each group of the second survey. The difference between the within-group correlation and global correlation is shown in red color, which is positive for all informants.

IV. COMMUNITY-LIKE STRUCTURES IN THE NETWORKS

A general property of social networks is that they contain compact sub-graphs\[1\]. Within these sub-graphs connection are dense, while between them connection are less dense. In this part, we will find the \( k \)-clique components\[11\] and the community structures of social networks of CSSS2008.

A \( k \)-clique is a complete sub-graph of \( k \) vertices, and two cliques are adjacent if they share \( (k-1) \) vertices. \( k \)-clique percolation component denotes the maximal sub-graph made of adjacent \( k \)-cliques. By finding the clique percolation components (CPC hereafter) in a network, the compact groups and the overlapping vertices can be revealed. Since the links
in the social networks we discussed before are too dense to find the \( k \)-clique components, here we increase the threshold of the establishment of links. Two nodes were connected only if they chose “Very Familiar” with each other. Fig. 5(a) shows the 3-clique and 4-clique components in the social network based on the first survey (VF Network 1). While Fig. 5(b) contains the 4-clique (thin lines) and 5-clique components in the social network based on the second survey (VF Network 2).

![Network 1](image1)

(a)The 3-clique (thin lines) and 4-clique (thick lines) percolation components in VF Network 1

![Network 2](image2)

(b)The 4-clique (thin lines) and 5-clique (thick lines) percolation components in VF Network 2

FIG. 5: The \( k \)-clique percolation components in VF Network 1 and 2. Vertices connected by links with the same color belong to the same clique percolation component. Red vertices are shared by two CPCs or above.

The largest cluster in Fig. 5(a) (colored in purple) is a compact group of Chinese students who received English training together; they lived together for one more week than others, which lead to a closer relation between them. The other clusters are the project groups that formed very early, so the group members became very familiar with each other before the first survey. In Fig. 5(b), the cluster of Chinese students is still the largest, while there are more clusters of project groups. Besides these two types, a Spanish-speaking cluster (colored in green) was also existed.
To sum up the clique percolation analysis, $k$-clique percolation components driven by friendship, projects and language are found in the social networks based on the surveys, which consists with the real scenario.

Participants of CSSS2008 contacted with their roommates and members of the same project frequently, so it is reasonable to assume that the relations between them are close. In fact, this assumption is verified by the survey results. In Network 2, about 80% participants have connections with all other people in their project group, over 90% participants have links with their roommates. Beside the contacting with roommates and project members, people also develop their personal relations through various in-campus and out-campus activities. By removing the links between roommates and members of same projects in Network 1, 2 and 3, social networks driven by these activities are obtained, which are named as Activity Network 1, 2 and 3. The hierarchical agglomerative algorithm\cite{10} is used to detect community structures in these activity networks of CSSS2008, the modularity $Q$ are also calculated. $Q$ is often used to quantify the compactness of the community. $Q = \sum_i (e_{ii} - a_i^2)$, where $e_{ii}$ is the fraction of edges connecting vertices in community $i$, and $a_i$ is the fraction of ends of edges that are attached to vertices in community $i$. High $Q$ value indicates for strong community structure in the network. Fig. 6 shows the community structures of Activity Network 2 and 3 respectively. The modularity $Q$ of Activity Network 1, 2 and 3 are listed in Table I.

Activity Network 2 contains three communities, one is Chinese dominate, one is Non-Chinese dominate, with a small community as a bridge between them. Three Chinese students are in this bridge community, one is a student helper, another two are participants who are very willing to help both Chinese and Non-Chinese. There are also three Non-Chinese students in it, one is a American born Chinese, another two are from outside US and Europe; so they serve as a cultural bridge. As time went by, the bridge community split and merged into the two large communities in Activity Network 3, with inter-community connections became denser. The modularity $Q$ is small for all three activity networks and keeps on decreasing as time passed by, indicates the strength of community structure in CSSS2008 was weak and decayed with time.
FIG. 6: The community structures in Activity Network 2 (Upper Panel) and 3 (Lower Panel). Symbols of different shapes and colors are nodes belong to different groups respectively. • for Non-Chinese Females, ■ for Non-Chinese Males, ♦ for Chinese Females and ▲ for Chinese Males.

V. CONCLUSION

To sum up our study, we have carried out surveys to study the time evolution of the friendship network of CSSS 2008. The topological structures and basic properties are then discussed. Correlation between reciprocal evaluations is quantified by Person’s correlation coefficient, which shown that people can well evaluate the friendship between them and people belonging to the same sex and ethnic group. At last, we obtained the clique percolation components and community structures to reveal the compact sub-groups of the networks.

VI. ACKNOWLEDGEMENT

This collaboration was initiated at the 2008 China Complex Systems Summer School, co-sponsored by the Santa Fe Institute and the Institute of Theoretical Physics of the Chinese Academy of Sciences. The research and education programs of SFI are supported by core funding from the U.S. National Science Foundation and by gifts and grants from individuals, corporations, other foundations, and members of the Institute’s Business Network for Complex Systems Research. SFI received direct support for the 2008 China CSSS from the
US NSF under award OSIE-0623953.


Appendix: The Sample Questionnaire

This is a sample of the paper-base questionnaire, to protect the personal privacy of participants, their names will be appeared as "student name 1". The contents of the online survey platform are the same as this one.

**Final Survey On CSSS 2008 Social Network**

Please kindly take about 5 minutes to answer the following questions. Thank you for your kindly help!

**Section 1 : Personal Particulars**

Your personal information and answers are confidential to our project group only. Your personal information is only used to construct and analyze the network and will not appear in the project report.

1. Please write down your name here (distinguishable nickname is acceptable, e.g. Alex)

**Section 2 : The Relationship Between You and Other Students.**

Please choose one option from "Very Familiar" to "Very Unfamiliar" to describe the familiarity of you and the students listed below. You can also choose not to respond. The sequence of the names follows the facebook.

<table>
<thead>
<tr>
<th>Family Name A - G</th>
<th>Very Familiar</th>
<th>Familiar</th>
<th>Neutral</th>
<th>Unfamiliar</th>
<th>Very Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student Name 1</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>..............</td>
<td>...........</td>
<td>..........</td>
<td>............</td>
<td>..............</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family Name H - N</th>
<th>Very Familiar</th>
<th>Familiar</th>
<th>Neutral</th>
<th>Unfamiliar</th>
<th>Very Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student Name 12</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>..............</td>
<td>...........</td>
<td>..........</td>
<td>............</td>
<td>..............</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family Name O - T</th>
<th>Very Familiar</th>
<th>Familiar</th>
<th>Neutral</th>
<th>Unfamiliar</th>
<th>Very Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student Name 26</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>..............</td>
<td>...........</td>
<td>..........</td>
<td>............</td>
<td>..............</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family Name U - Z</th>
<th>Very Familiar</th>
<th>Familiar</th>
<th>Neutral</th>
<th>Unfamiliar</th>
<th>Very Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student Name 35</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>..............</td>
<td>...........</td>
<td>..........</td>
<td>............</td>
<td>..............</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Staff Members</th>
<th>Very Familiar</th>
<th>Familiar</th>
<th>Neutral</th>
<th>Unfamiliar</th>
<th>Very Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student Helper 1</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>..............</td>
<td>...........</td>
<td>..........</td>
<td>............</td>
<td>..............</td>
</tr>
</tbody>
</table>

*Thank You Very Much for Helping Us ^_^, Good Luck Everyone!!!*

*Also available online at*  http://mmlab.itsc.cuhk.edu.hk/SurveyASP/Survey.aspx?sid=1337