Six Degrees of Separation in Online Society

Lei Zhang*
Tsinghua-Southampton Joint Lab on Web Science
Graduate School in Shenzhen, Tsinghua University
Shenzhen, Guangdong Province, P.R.China
zhanglei@sz.tsinghua.edu.cn

Wanqing Tu
Department of Computer Science
University College Cork
Cork, Ireland
wanqing.tu@cs.ucc.ie

Abstract— Six degrees of separation is a well-known idea that any two people on this planet can be connected via an average number of six steps. Having succeeded in the real world, the theory even directly or indirectly motivated the invention of online societies. However, no much effort has been paid on checking if the theory really holds for online societies whose connection pattern may not be identical to the real world. This paper tries to give an answer to the question by both mathematical modeling and online measurements. The mathematical approach formulates the problem as a Minimum Diameter Problem in graph theory and evaluates the maximum and average values of the number of connections between any two random-selected community members. Measurements are conducted in three different kinds of online societies, namely ArnetMiner for academic researchers, Facebook for students, and Tencent QQ for teenagers in China. Analysis of these measurements verifies our theoretical findings.

Keywords- *Six degrees of separation, social network, minimum diameter problem, small world*

I. INTRODUCTION

The theory of six degrees of separation states that any two random-selected people on this world can get to know each other by no more than six steps of intermediate friend chains (Figure 1). Although it is very hard to figure out the actual origin of the theory, there seem to be a common agreement that the term of six degrees of separation was popularized by John Guare in 1990’s by his famous intricately plotted comedy.

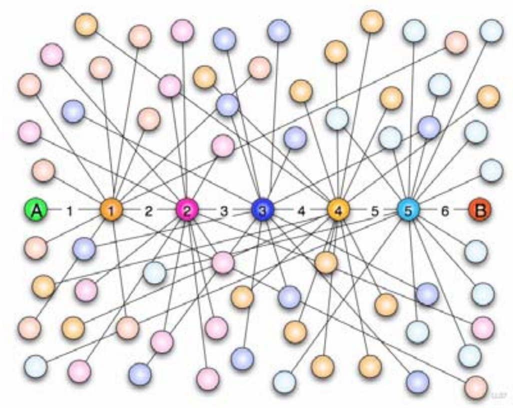

Figure 1 Illustrative example of the six degrees of separation theory. (Wikipedia).

As the theory is widely accepted by more and more people, it is also considered as the motivation of online Social Network Services (SNS). Many Web 2.0 websites are based on the idea that the users would greatly increase their social capital simply because they would be able to know almost everyone on this planet within six steps of hops.

However, the connection pattern among online society members does not need to be exactly the same as the real world. This lead to a number of unsolved questions: Will the theory of six degrees of separation hold for the virtual world? Is the value of six too large or too small to connect two random members?

*Correspondence Author

Some experiments have been conducted on the popular Facebook platform [1] by collecting the profile information of volunteer members provided they are willing to download and install an application. The results seem to indicate the correctness of the theory with an average of 5.73 hops. But the homogeneous characteristics of all participants weaken the result, making us less confident to guarantee it is still true for a diverse range of users with different interests and backgrounds. Furthermore, except the few experiment studies, there has never been a mathematical proof or analysis on the correctness of the six degrees of separation theory in the virtual online society world.

Therefore, in this paper, we would like to investigate the problem by both mathematical modeling and measurement analysis on different online societies. The conclusion is that the degree roughly holds for all kinds of online societies we studied, but the accurate degree is dependent on the size, structure, connectivity, and other metrics of the society.

The paper is structured as follows. First we review the related work on the theory of six degrees separation in Section II. Special attention is paid to previous experiments on online social networks. Then we present the problem discussed in this paper and formulate it as the Minimum Diameter Problem using a graph theory approach in Section III. Measurement results and analysis conducted on three kinds of online societies are presented in SectionIV. Finally we summarize our conclusions in Section V.

II. RELATED WORK

The theory of six degrees of separation seems to be a common sense whose origin is very difficult to trace. But we can still identify Milgram's "Small World Problem" in 1967 as one of the earliest scientific studies on verifying the theory [8]. In the experiment, some random-chosen people succeeded in passing information packages to another random-chosen people via a chained path with length 5.5 or six.

In 2001, Duncan Watts, a professor at Columbia University, repeated Milgram's experiment on the Internet, using an e-mail message with 48,000 senders and 19 targets. It was found that the average number of intermediaries was around six [10].

A newer result of 6.6 average degrees was obtained by Microsoft researchers recently after examining 30 billion electronic messages [5].

A Facebook platform application named "Six Degrees" has been developed to calculate the degrees of separation between different people [6]. With 4.5 million users the average separation is 5.73 degrees, whereas the maximum value is 12. There was another experiment on the same platform but failed to yield any valuable results [7].

All of the above research works are just measurement attempts. The lacking of mathematical participation has lasted until the mathematicians of the American Mathematician Society defined the notion of collaboration distance, an analogy to the degrees of separation, to represent the co-author connections among themselves [11]. But that is still a statistical approach without any mathematical analysis.

III. PROBLEM FORMULATION

By abstracting an online society in to a graph, whose nodes are society members and links are friend connections, we can formulate the theory of six degrees of separation as the Minimum Diameter Problem.

Definition 1. *Given any two nodes A and B in graph G, the distance d_{AB} between the two nodes is defined as the number of hops of the shortest path connecting them.*

Definition 2. *The diameter D_G of graph G is defined as the maximum distance value for any pair of nodes (A,B) in the graph.*

$$D_G = max_{A,B \in G} d_{AB}$$

Following the previous definitions, the theory of six degree of separation can be formulated as follows:

$$D_G <= 6$$

Therefore, in order to test the theory, for a given online society graph, we need to calculate mathematically the diameter of it and then compare with six.

Previous study reveals that the diameter of a sparse random graph can be derived in the following form:

$$D_G = c\ln n + o(\ln n).$$

For WWW and online societies, they demonstrate a power-law distribution, where the number of degree-d nodes is proportional to d^{β}. Then the diameter can be further expressed as a function of β.

Therefore, the crucial problem becomes determining the value of *beta* for a given kind of online society. That would be presented in the next Section.

IV. Measurement Studies

In this section we present our measurement results of degrees of separation in three different online societies.

- **ArnetMiner**

ArnetMiner is an online database built by Tsinghua University using semantic web technologies. Currently it contains a society of 0.5M academic researchers [1].

We randomly select 100 pairs of the researchers and search for all feasible paths connecting them. An example is shown in Figure 2, where the randomly selected pair of researchers are connected via a number of other researchers. Each connect is based on co-authorship of papers or projects.

Figure 2 Illustrative example of collaboration distance in ArnetMiner.

The feasible paths are sorted using different metrics. We first rank all of them based on ascending order of path length (defined as number of hops). The length of the shortest path is taken as the degree of separation among the two researchers. We then repeat this procedure for all the 100 pairs and plot the distribution curve for them. The result is shown in Figure 3. The result is quite surprising since most researchers (over 80%) can be connected in less than 3 hops, which is much smaller than what the theory of "Six Degrees of Separation" predicted.

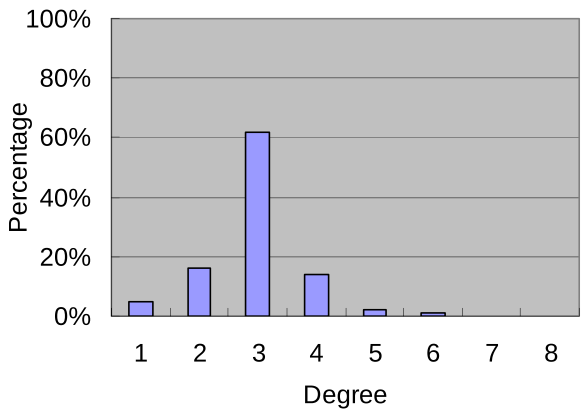

Figure 3 Degree Distribution of shortest path in ArnetMiner.

In the first experiment, we only select the shortest path and don't care about whether the links in the path are strong or weak. In the second set of experiments, we associated weights to indicate the strength of the links. If two researchers have more co-authored papers and projects, the link connecting them will be stronger. We then sort all feasible paths connecting the two in descending order of path strength. The length of the strongest path is taken as the degree of separation among the two researchers. We then repeat this procedure for all the 100 pairs and plot the distribution curve for them. The result is shown in Figure 4. The degree is much larger than the first experiment since we have to use longer paths in order to make it strong. But the theory of "Six Degrees of Separation" is still observed here.

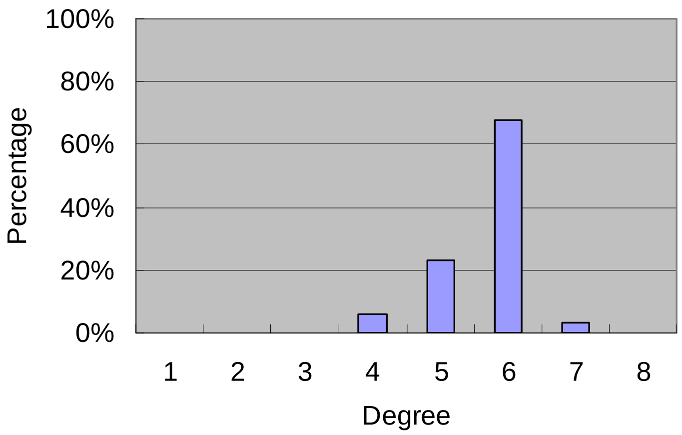

Figure 4 Degree distribution of strongest path in ArnetMiner.

- **Facebook**

Facebook [3] is a social utility that connects people with friends and others who work, study and live around them. Unlike ArnetMiner, most users of Facebook are university students, which represent a different group of interests and life style. It is reported that Facebook has replaced MySpace to be the No. 1 social network in US [2].

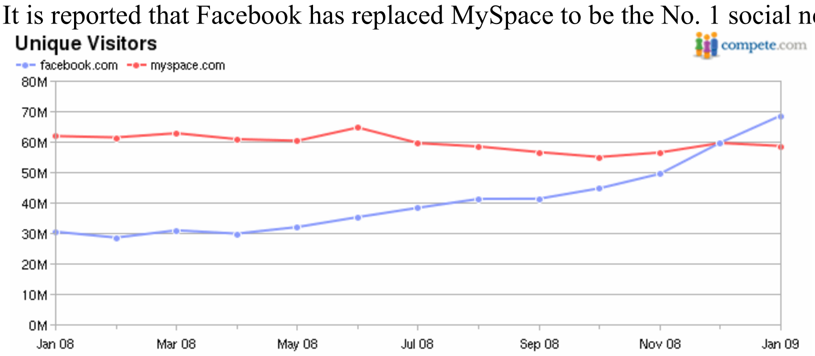

Figure 5 Unique visitor trends of Facebook and MySpace.

Based on the previous mathematical analysis in section III, we need to determine parameters for $D_G = a + c\log n$, where n is the number of accounts in Facebook. According to Facebook statistics, the current number of active accounts is 175 million and the average number of friends is 120 [4].

Note that Facebook user graph obeys the power-law distribution with parameter β, the distribution of number of friends is given as follows:

$$\lambda_\beta = i^{-\beta}/\zeta(\beta), \text{ where } \zeta(\beta) = \Sigma\, n^{-\beta}.$$

Computation result shows a β value of 2.97, which gives a graph diameter as a function of β whose value is

2.94, independent of the number of nodes in the power-law graph. That also justifies that future growth of Facebook won't increase the diameter significantly.

- **Tencent QQ**

Tencent QQ is an IM tool with 783.4M users in China [9] and QQZone is one of its applications which allows the users to create their own mini flashy homepage with idiot-proof templates. QQ and QQZone represents an online society full of teenagers.

By crawling QQZone, we use a robot to collect information of all visible friends as well as all message reply posts. They will all be viewed as friends of the page owner with strong links. Statistics show that the average number of friends is 94, which yields a rich connected graph topology.

Using exactly the same mathematical model above, we find an average degree of separation at the value of 3.12 with the power-law parameter of 2.81.

V. Conclusions

We re-visited the well-known theory of six degrees of separation in the context of online society. Both mathematical analysis and measurements are used to suggest that the theory still holds. We also find that the maximum and average value of the degree depend on the characteristics of the society.

Further investigation is needed to study the effect of having groups in online societies. For example, almost 50% of QQ users join groups whose size is often larger than 100 and even goes to 500. This would significantly increase the average number of friends for all group members. Considering the fact the groups are more and more popular, this would very likely bring the degree of separation to the floor.

Another interesting thing is to focus only on strong connections like classmates, colleagues, neighborhoods and filter out those weak ones.

References

[1] ArnetMiner, http://www.arnetMiner.com.

[2] Compete, http://www.compete.com.

[3] Facebook, http://www.facebook.com.

[4] Facebook statistics, http://www.facebook.com/press/info.php?statistics.

[5] Microsoft, Proof! Just six degrees of separation between us, http://www.guardian.co.uk/technology/2008/aug/03/internet.email.

[6] Six Degrees Application, Facebook, http://apps.facebook.com/six_degrees_app/.

[7] Six Degrees of Separation, Facebook, http://www.facebook.com/apps/application.php?id=4616854023.

[8] The Small World Problem, http://en.wikipedia.org/wiki/Small_world_experiment .

[9] Tencent QQ, http://www.qq.com.

[10] Duncan J. Watts, Peter Sheridan Dodds,M. E. J. Newman, Identity and Search in Social Networks, Science 17 May 2002:
Vol. 296. no. 5571, pp. 1302 – 1305.

[11] http://www.ams.org/mathscinet/collaborationDistance.html.