ISSN:2231-0711
Jie Cheng et al | IJCSET |March 2013 | Vol 3, Issue 3, 71-75

Small World Phenomena

Jie Cheng[#], Mehdi R. Zargham[*], Shuguo Rang[&]

[#]*Department of Computer Science and Engineering University of Hawaii at Hilo*
200 W. Kawili Street Hilo, HI USA 96720

**Department of Computer Science,*
Southern Illinois University Carbondale, Carbondale, IL USA 62901

[&]*Department of Computer Science & Technology*
Heze Vocational College, Heze Shandong, P.R. China 274000

Abstract— **This paper offers an overview study of small world problem. We emphasize on introducing the basic concepts, and the empirical analysis and results of related literatures. We surveyed from Milgram's original experiments on small world problem to the current empirical study conducted by Leskovec on the large data sets on Microsoft messenger to uncover the structure and model of small world network.**

Keywords— **small world, six degree separation, random graph, power-law distribution.**

I. Introduction

The "small world problem" can be defined as: "What are the chances that two people chosen at random from the population will have a friend in common [1]?" Small world problem has been fascinating people from many disciplines, such as Sociologist, Mathematician, computer Scientist, Psychologist, historians, political scientist, and communication specialist [2]. Mathematician make mathematical model to reveal logical work of small world problem. Computer Scientists do experiments on large scale data to provide empirical evidence [3].

Small world networks are everywhere, such as email, online social network, World Wide Web, scientific and actors' collaboration network, and power transmission grid. Study of small world problems can help us understand why social network are organized the way it is and how information flows, such as how contagious disease spreads in the network.

In 1967, Milgram announced counter-intuitive statement that "The world is small indeed, separating average by six steps away" [2]. In other words, there are no more than six intermediate acquaintances between any two arbitrary people. "Six degree of separation" is another term associated with Milgram's discovery even though he did not use this term formally. There are many following work after Milgram's groundbreaking discovery. Killworth and Bernardn in 1978 conducted an experiment, which is reversal of Milgram's experiment to discover "how many of his or her acquaintances could be used as first steps in a small-world procedure, and for what reasons" [3].

In 1998, Walts and Strogats proposed a mathematical model of small world problem. Kleinfield tried to uncover the myth of small world problem by exploring the Milgram's archives in Yale library in 2002 [1].

As the technology advanced, large scale social data become available to scientists who could use the data set to study structure and developing models on small world network. Watts and Strogatz in 1998 carried on computer simulations on small world phenomenon to develop a model for it [4]. In 2008, Horvitz and Leskovec studied 30 billion conversations among 240 million Microsoft messenger users to verify Milgram's discovery and model the structure of small world network [5].

The paper is organized as follows: section 2 describes the origin of small world problem/experiment conducted by Milgram; Section 3 will explore the further study of small world problem**;** Section 4 presents the recent analysis and experiment of small world problem on large scale data set.

II. Origin Of Small World Problem/Experiment

Milgram conducted two experiments with $680 funding. The first experiment is to deliver the document starting from Kansas and trying to reach the target who is a divinity school student's wife lived in Cambridge. The second experiment starts from Nebraska and tries to reach to a target who is the stockbroker worked in Boston and lived in Sharon, Massachusetts. Both experiments will send a folder containing some documents to targets. Documents consist of information of targets, a list of rules for directing the participants to reach the target person and a roster which "includes the people's name on the chain in order to prevent endless loop" [2]. In addition, fifteen business reply cards are included in the document for helping to keep track of the progress. Everyone in the chain will fill out the card and returned to Milgram group. Participants will forward the document to the next person only based on first name [2].

As shown in figure 1, there are 44 complete chains out of starter of 160 in Nebraska study. The average chain length is about 6.6, varying from 2 to 10 intermediaries with median 5. In addition, by analyzing the features of the chain, Milgram also discovered "some pattern of contact of American Soceity" [2]. In terms of sex role, participants were three times less likely to send the documents to people of the opposite sex as to one with the same sex. Study also shows that participants tend to forward the folder to a friend and acquaintance rather than a relative.

From the experiment, Milgram concludes that "social communication is sometimes restricted less by physical distance than by social distance" [2]. In the experiment, some of the folders were moved from Nebraska to neighborhood of Boston, which is 1000 miles away. However, the folder went around and around in the neighborhood of target. The chain never got complete; the folder eventually never reached the target.

Milgram's experiment inferred two important points as below. 1) There are plenty of short paths exit in the social network. Apparently, it is quite difficult for each participant with limited local information about the social structure to choose correct shortest social path to reach to the target. This kind of search is called social search. 2) However, Milgram points out that: with collective effort people indeed can find these shortest paths. This search follows geometric progression which shows exponential growth or exponential decay. "With a few moves, the search extends to an enormous number of persons"[2]. Therefore, we can understand that in average, "five people apart between any two arbitrary persons" as "five circles of acquaintance apart" [2].

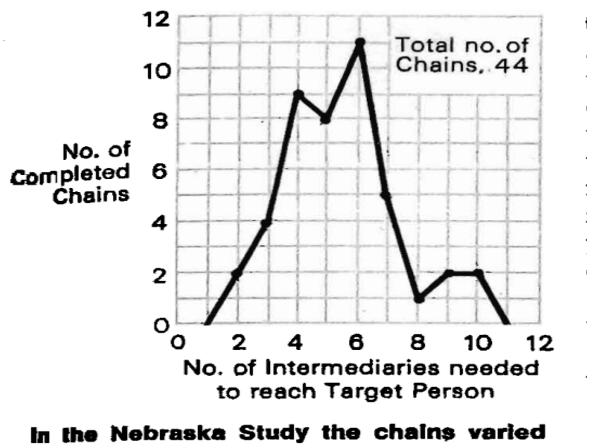

Figure 1: Length of Complete Chain; taken from [2].

In 1969, Travers and Milgram conducted a variation of original Milgram's experiment [3]. This experiment follows the same procedure of Milgram's experiment but "starting population" varies. There are 296 initial starters who would forward the document to reach the target. These starters were divided into three groups: "Nebraska stockholder" consists of blue-chip stock owner in Nebraska; "Nebraska random" group were chosen from Nebraska large population. "Boston random" group were from Boston population.

There is only one target person who was a stockholder who works in Boston Proper and lives in Sharon, Massachusetts, suburb of Boston. His detailed information was included in the forwarding document in order for the authors to discover which type of information would help participants to reach the target.

Result of this experiment shows that 29% of starting people (64 out of 296) successfully reached the target. Figure 2 shows the distribution of completed chain length. There are two possible reasons for incomplete chains: 1) participants are not highly motivated to send the document to the next people. 2) Participants didn't know to whom s/he should send the document.

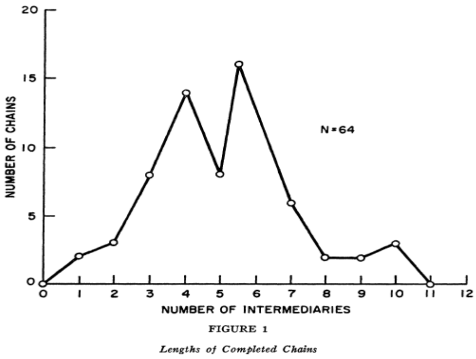

Figure 2: Length of Complete Chain; taken from [3].

The mean of chain length turned out to be sensitive to the place of residence of starters and target. Boston random group (4.4 intermediaries) has less chain length than Nebraska random group (5.7 intermediaries). In addition, Nebraska stockholder group can reach contact through business contacts; therefore, Nebraska stockholder group has 5.4 intermediaries, as opposed to a mean length of 5.7 for Nebraska random group. Also, this experiment shows the common channel, which is the intermediary who shows in more than one chain.

Even though "The world is small indeed, separating average by six steps away" announced by Milgram is counter-intuitive, people accepted this theory in a very short period of time. Kleinfield explored Milgram's papers in Yale archive and other related literature review.

Kleinfield argued that even though mathematician developed computer model to study how small world operate in a logical way, empirical evidences are lacked to support this idea. Indeed, he believes that world we lived in is actually separated by social barriers, such as race and gender. Kleinfield also suggested that empirical studies are needed for questions like "Is six degrees of separation a large or a small number?"[1]. "We are used to thinking of "six" as a small number, but in terms of practical connections, "six" may be a huge number indeed." [1].

III. MODEL OF SMALL WORLD NETWORK

Watts and Strogatz in 1998 carried on computer simulations on small world phenomenon [4]. The simulation not only captured the Milgram's discovery that the random individual in a social network can be reached on the average by a small number of intermediaries, but also measure the local clustering as local information for any individual on the large social network to search for targets.

Watts and Strogatz claimed that each of us exits in a small community, in which an individual's friends are each other's friends. This small community thus has high clustering, which means highly connected. With small probability, someone in this community may have contact with someone in a different community. Even though this "long range contact" [4] may be rare, it makes possible for our large social network to become well connected with shortest distance with average about 6 intermediaries.

Watts and Strogatz performed simulation on the ring lattice as shown in figure 3. Initially, a ring structure was created which is not completely connected but displays a high degree clustering (left most on figure 3). Then the network is rewired by redirecting the outgoing edges of each node to some other nodes with controlling probability p. Return value of a random generator which distributes uniformly between 0 and 1 can be used to decide whether to rewire an edge or not. If return value is less than p, rewiring needs to be done; otherwise, keep edge unchanged and continue to consider the next node.

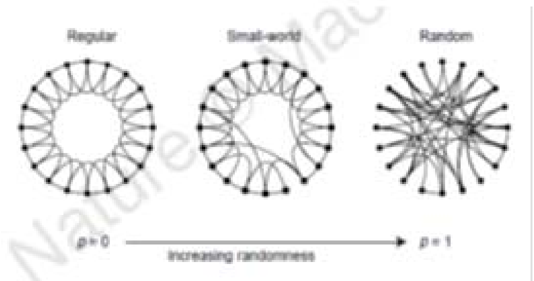

Figure 3: Lattice ring. The first ring represents the initial ring structure; the second ring represents the structure after random rewiring for certain number p; the last ring represents the random rewiring structure for p=1; taken from [3].

The graph constructed based on p = 1.0 is close to Erdos Enyi random graph [5]. “Strictly speaking, an Erdos-Renyi graph is obtained by starting with n isolated nodes, visiting each of the n(n − 1) possible node pairs, and selecting with probability p a node pair for a direct connection with an edge. [6]” In the Erdos-Renyi graph, the probability of the node degree d follows the binomial distribution

$$prob(d) = \binom{n-1}{d} p^{d}(1-p)^{n-1-d}$$

Let z = (n-1) p describes the average degree of a node

$$prob(d) = \binom{n-1}{d}\left[\frac{z}{n-1}\right]^{d}\left(1-\frac{z}{n-1}\right)^{n-1-d} \approx \frac{z^{d}}{d!}e^{-z}$$

When number of nodes in the network n approaches to infinity, $\frac{z^d}{d!}e^{-z}$ approaches to a exact number. In other words, when the network is large enough, the distribution of the node degree will follow Poisson distribution. This will cause the social network graph to be associated with a scale according to the Poisson law [14].

Unfortunately, the average degree distribution in our real life, such as World Wide Web does not follow Poisson distribution. In 2008, Horvitz and Leskovec studied 30 billion conversations among 240 million Microsoft messenger users [5]. A graph with 180 million nodes and 1.3 billion undirected edges, largest social network built up to date, was constructed and analyzed. The result of this study shows nothing like the Binomial random model. As shown in figure 4, the degree distribution of MSN messenger communication follows the power law distribution, that is $prob(d) = c/d^{\alpha}$, [5] which is totally different from random graph following binomial distribution as shown in figure 5.

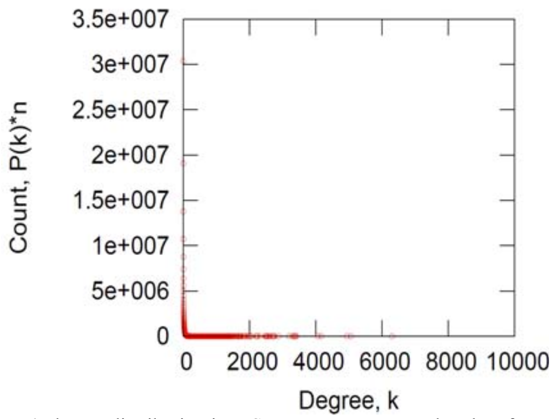

Figure 4: degree distribution in MSN Messenger network; taken from [5].

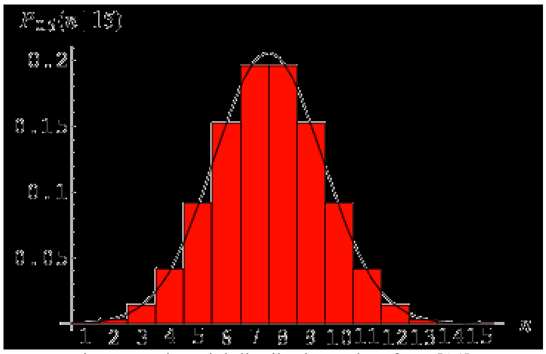

Figure 5: Binomial distribution; taken from [14].

In addition, 1000 nodes were randomly sampled in order to calculate the shortest path between each of these nodes. As a result, “the average path length is 6.6” with median at 7 [5]. In other words, any random pair of nodes can be reached in less than 8 hops [5].

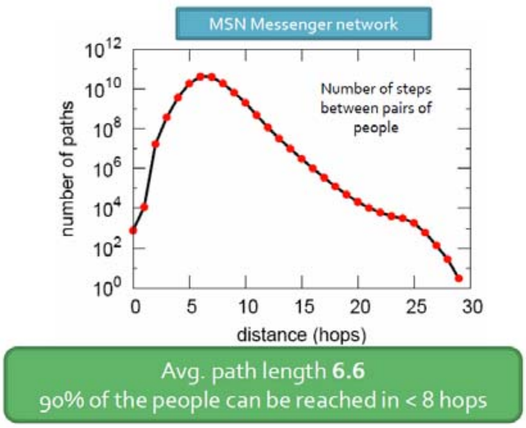

Figure 6: Shortest distance in MSN Messenger network; taken from [14].

Figure 7 shows that the diffence between the random network and network with power law distribution. As we can see, by following the power law distribution, there exist highly connected nodes in the network. In contrast, the random network shows no prominent highly connected nodes.

Therefore, networks following the power law distributions are also called scale free network.

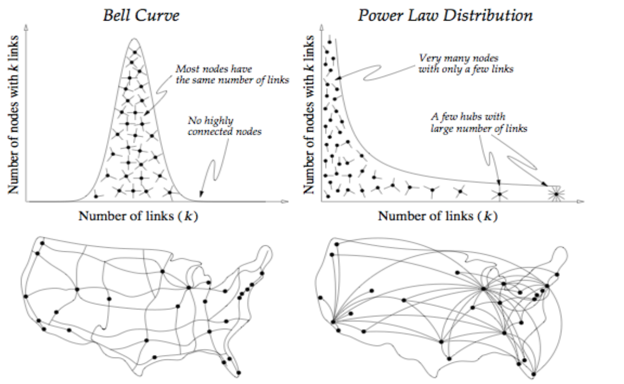

Figure 6: comparison of random network and scale free network ; taken from [14].

The power law distribution is shown as below: $prob(d) = c/d^{\alpha}$
Here, c and α are both constants; α is typically in the range [2, 3]. Table 1 displays α value in different networks. For example, Web graph has in degree power law distribution with $\alpha_{in} = 2.1$ and out degree power law distribution with $\alpha_{in} = 2.4$.

TABLE 1
A VALUES IN DIFFERENT NETWORKS [14]

	α	References
Web Graph	$\alpha_{in} = 2.1$, $\alpha_{out} = 2.4$	[7]
Autonomous systems:	= 2.4	[8]
Actor Collaborations	= 2.3	[9]
Citations to papers	≈ 3	[10]
Online social network	≈ 2	[11]

Power law distribution can be used not only to explain why the popular nodes gets more popular, which is rich get richer phenomena, but also to build effective crawlers in the modern search engine. For example, in the World Wide Web, generative model are used to model the growing of the web sites, such that the behaviors of linking new web page to existing ones. Preferential attachments are foundation for most generative models. [12] "one node is added to the network with m new edges to existing nodes selected according to some probability distribution (a function of the some characteristic of existing nodes). [12]" BA model is one of the well-known preferential attachment models, which states that the probability of node j, a newly created web page, connect to node i, an existing webpage depends on the degree of node i. The problem associated with BA model show as the following: 1) It is based on impractical assumption that web authors have global information of web degree. 2) The model allows the popularity of the page alone be the factor to affect how the web pages receive the links. The consequence is that the oldest nodes will have the highest degree [12].

In order to resolve the second problem of BA model, one extension of BA model states that an existing node i, the probability of new node j connect to node i depends on the fitness of page i. The pages with highest degree will be those with highest fitness when the time is long enough. This variation of BA extension still yields power-law degree distribution of the entire Web.

Another extension of BA model, Mixture model includes two parts: first part is the same with BA model; second part states that a new web page j uniformly connected to an existing node. The positive point of this model is that it can "fit in both power distribution of the entire web and unimodal degree distribution of subset of webpages such as university, company, or newspaper homepages" [12]. However, it still relies on complete knowledge of degree. Also, it cannot apprehension the reasoning that leads authors to choose pages to link [12].

Copying Model does not need complete knowledge of node degrees. For each new node, a prototype of existing node i is chosen randomly. This new node either connects its k^{th} link to this node i with probability α or connect to k^{th} link of node i with probability 1- α. This model also results in power-law distribution. This model will also result in popular nodes will get more popular.

IV. CONCLUSIONS

The world is small indeed; any two random people are separated by a few steps away. These small world networks have certain characteristics such as the small network diameters and large clustering coefficients. The degree distribution follows the power law distribution. People with local information can indeed collectively find such short path to reach the target. Empirical studies, including the original Milgram's experiments and recent large msn data set analysis have verified the structure and model of small world problem.

REFERENCES

[1] J. Kleinfeld. *Could it be a Big World After All? The `Six Degrees of Separation' Myth.* Society, 2002.
[2] S. Milgram. *The small world problem*. Psychology Today 1(1967).
[3] P. Killworth and H. Bernard, *Reverse small world experiment.* Social Networks 1, 1978.
[4] J. Travers and S. Milgram. *An experimental study of the small world problem*. Sociometry 32, 1969.
[5] J. Leskovec, E. Horvitz. *Worldwide Buzz: Planetary-Scale Views on an Instant-Messaging Network*. Proc. International WWW Conference, 2008.
[6] P. Erdos, A. Renyi. *On the evolution of random graphs*. Magyar Tud. Akad. Mat. Kutato Int. Koezl., 1960
[7] F. Menczer. *Growing and Navigating the Small World Web by Local Content.* Proc. Natl. Acad. Sci., 99(22): 14014-14019, 2002.
[8] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener. *Graph structure in the Web*. Computer Networks, 33, 2000
[9] M. Faloutsos, P. Faloutsos, C. Faloutsos. *On Power-Law Relationships of the Internet Topology*. In Proc. SIGCOMM, 1999.

[10] L. Barabasi, R. Albert. *Emergence of scaling in random networks.* Science, 286, 1999.

[11] S. Redner. *Citation statistics from 110 years of Physical Review.* Physics Today 58, 49-54, 2005.

[12] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, M. Hurst. *Cascading Behavior in Large Blog Graphs*. In Proc. SIAM International Conference on Data Mining, 2007.

[13] F. Menczer. *Growing and Navigating the Small World Web by Local Content.* Proc. Natl. Acad. Sci., 99(22): 14014-14019, 2002.

[14] J. Leskovec. *Power Laws and Rich-Get-Richer Phenomena* http://snap.stanford.edu/class/cs224w-2011/slides/11-powerlaws.pdf