SMALL WORLD PHENOMENON IN INFORMATION NETWORKS: A REVIEW

Mr. Pranjal Kumar Bora[1], Sanjay Sharma[2]

[1]Asst. Professor, [2]Research Scholar

[1]Centre for Computer Science and Applications, Dibrugarh University

[2]Department of Mathematics, Dibrugarh University.

Abstract: The discovery of Small world phenomenon helps us to study real world complex networks and analyze the real world system in a different perspective which has received intensive interdisciplinary attention during the past several years. "Small world phenomenon", which is come to existence in 1960 by pioneering work of Stanely Milgram, which proves that we are connected by small chain acquaintances. It is found by the researchers that a small world network is often characterized by high connectivity and clustering. Here in this paper, it is try to introduced the influence of small world networks in different networks and tried to show its different attributes values on that networks which are calculated by the different researcher on different networks. Also try to emphasize on the designing model and new observation made by researchers and analyze it.

Keywords: Small world network, Complex networks

I. Introduction: Study of graph is one of the earliest discoveries stating by Euler's in 1736, popularly known as Seven Bridges Konigsberg problem. Complex Network Analysis (CNA) is a young field of research. A complex system is becomes complex networks where the vertices are elements of that system and the edges represent the relationship among them. Complex Network Analysis has been use in different areas from social networks to knowledge networks, from brain networks to banking networks and analyzes it time to time and applies it in real world applications. This real worlds networks behave like a small world networks which is neither regular nor random that demonstrate two basics properties i.e., average path length and high clustering coefficients (Watts and Storagtz, 1998)[3].

The "small world problem" is commonly explained as: "What are the chances that two people chosen at random from the population will have a friend in common [2]. Study of small world help us to know how the different networks are organized ,the efficiency and how the information like virus in computer networks flows, how HIV diseases spreads in a society, networks etc.

In 1967, Milgram [1] proves that "The world is small indeed, separating average by six steps away" which is also called six degree separation problems. In other words, there are no more than six intermediate acquaintances between any two arbitrary people. That is, in a real world graph one vertex is separated by any other vertex by less than six separations.

II. Basic properties of Small world model

A graph G consists of a set of nodes (or vertices) V and a set of edges E. If U and V are nodes and E (u, v) then they are connected by an edge. For an example of a graph, the Facebook network may be drawn as a graph wherever individuals are the nodes and two people are connected by an edge if they are friends. We will some time introduce weights on the graph, either on edges or nodes.

Clustering co-efficient: Mathematically, *C* is the proportion of edges e_i that exist between the neighbors of a particular node (*i*) relative to the total number of possible edges between neighbors (Bullmore and Sporns, 2009). The equation for *C* at an individual l node of degree k_i is:

$$C_i = \frac{2e_i}{k_i(k_i - 1)} \qquad (1)$$

Another definition of clustering coefficient is based on transitivity [5] which is defined as-

$$C^{\Delta} = \frac{3 \times \text{number of triangles}}{\text{number of paths of length 2}}, \qquad (2)$$

where a 'triangle' is a set of three nodes in which each contacts the other two. Both capture intuitive notions of clustering but, though often in good agreement, values for C^{ws} and C^{Δ} can differ by an order of magnitude for some networks.

Path length: Path length (L) is a measure of the distance between nodes in the network, calculated as the mean of the shortest geodesic distances between all possible node pairs:

$$L = \frac{1}{N(N-1)} \sum_{\text{ifn}, i \neq j} d_{ij} \qquad (3)$$

Where *dij* is the shortest geodesic distance between nodes ***i*** and ***j***.

A graph exhibits small world behavior if $L \geq L_r$ and $C >> C_r$ where L_r and C_r are the characteristic path length and clustering co-efficient.

III. Watts- Strogatz Model

Graphs found in any networks where there is a relation or link between two objects and this small world phenomenon exists in that graph where it is used to study the feasibility, how good that networks is in terms of communications, networks flows. In the Watts-Strogatz model we take the ring lattice and rewire every edge to a random node in the graph with some probability "p". Watts and Strogatz showed that there is a region of with some "p" state that the model has both short distances and a high clustering coefficient. The Watts and Strogatz, 1998[1] model begins with a ring of n nodes, each node connected to its nearest neighbors out to some range K. Each edge in turn is rewired to a new target node with probability "p". The WS model shows that p=0 gives a regular network ,with high clustering but high path length; p=1 gives a pseudo-random network, with low clustering and path length ; and intermediate p values give small-world networks with high clustering and low path length.

Watts and Strogatz, 1998[1]; Collins and Chow 1998 and Watts 1999, design small world that lie somewhere in between regular lattices and random graphs, in which the following properties are shown:

(i) Local neighborhood is preserved - as for regular lattices, and

(ii) Diameter of the network, quantified by average shortest distance between two vertices, increases logarithmically with the number of vertices n - as for random graphs.

Two parameters were formally introduced earlier on by Watts and Strogatz to quantify these two properties: clustering coefficient C (a local property) and characteristic path length L (a global property). The clustering coefficient measures the average probability that two nodes with a mutual "friend"(node) will be connected. It is the average number of edges existing in the clique divided by the maximum possible number of edges in the clique. Meanwhile, characteristic path length is defined as the length of the shortest path (i.e. smallest number of edges) required to connect one node to another, averaged overall all pairs of nodes.

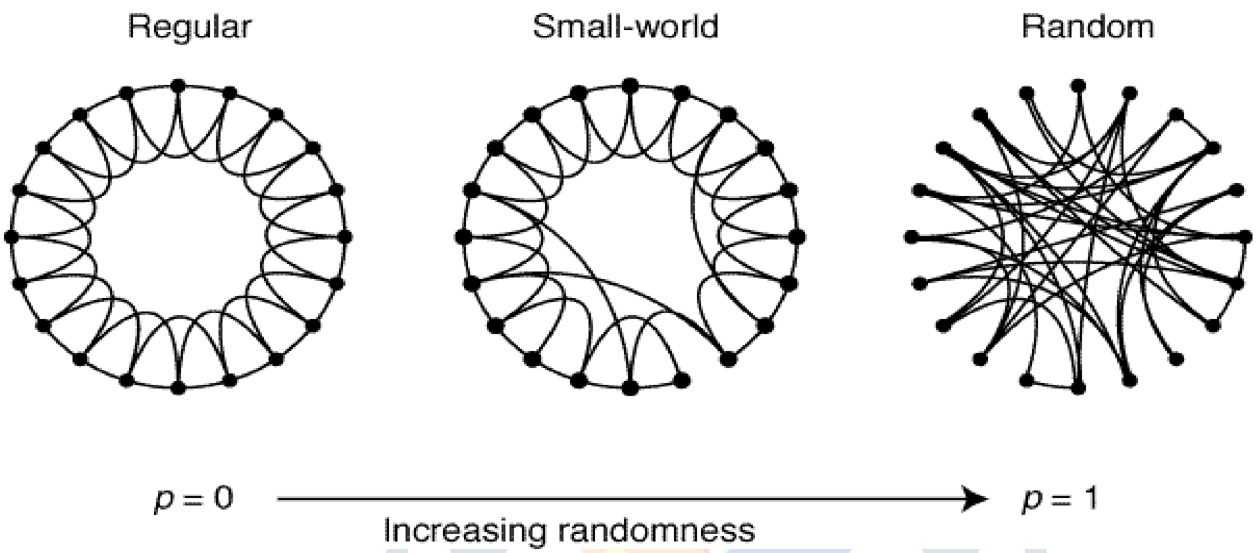

Fig1: Regular, Random vs Random Networks (Watts and Strogatz 1998)

Here in Fig1 shows that regular networks have high clustering and high distance. Again, in case of random networks it exibits low distance and low clustering. Small world networks appear in between of this two networks with high clustering and low distance.This rewiring of the nodes happend with increase of the probability range from 0 to 1.

Also Newman and Watts (1999) proposed a model where new links are added between pairs of sites which is choosen at random which does not change the size of the networks.

IV. Small World Theory results

Small world concept apply and study by number of researcher like Korte & Milgram 1970,Stevenson et al. 1997 ,Newman 2000b in Social Network Study. also by Albert, Jeong & Barabási 1999, Giustiniano & Carignani 1999, Allen 2000 in Internet & Telecommunication Technology, in life science by Wagner & Fell 2000 ,Gleiss et al. 2000 ,Bagnoli & Bezzi 2000 etc. Also recently this complex network analysis done on different transportation network like urban road network, worldwide airport network, airport network in china.

Researchers use different small world properties in real world system using different tools and methods. Such an example like affect of spread of diseases or epidemics in social networks have studied and then find out the affected one from the community and isolated them, by Moore and Newman,1999.

In the case of the film industry where the number of vertices (n) was 225,226 and average number of edges per vertex (k) was 61, its characteristic path length (L_{flim}) was found to be 3.65 and clustering coefficient (C) was film film 0.79. In comparison, a random graph with the same "n" and "k" was found to have an L of 2.99, a rather close number to L , and C of 0.00027 that was much smaller than C .

Mark D Humphries,, Kevin Gurney defined a precise measure of 'small-world-ness' *S* based on the tradeoff between high local clustering and short path length. They try to examining the behavior of S on large data sets and found that all these systems were linked by a linear relationship between their *S* values and the network size *n*. In this below table shows the small world properties in different class of networks and calculate its attributes values. Here, n indicates nodes, m indicates edges, edge density ξ, ⟨*k*⟩ is the expected value of the degree across the network, C^{WS} indicates clustering coefficients in WS model, C^{Δ} indicates clustering coefficients based on transitivity, small-worldness' S^{Δ} , L is for minimum path length, Entries '-' indicate missing data; n/a indicates values that could not be computed.

Table 1: Table of Small-Worldness values and other topological properties of information networks [5]

class	#	network	*n*	*m*	<k>	ξ	L	C^{Δ}	C^{WS}	S^{Δ}	S^{WS}	*P(WS)*
Social	1	Dolphins[f]	62	159	5.13	0.084	3.36	0.31	0.26	2.8	2.35	0.64
	2	film actors	449913	25516482	113.43	2.561024	3.48	0.2	0.78	627	2446	0.95
	3	company	7673	55392	14.44	0.002	4.6	0.59	0.88	228	341	0.77
	4	math	253339	496489	3.92	1.661025	7.57	0.15	0.34	11666	26443	0.7
	5	physics	52909	245300	9.27	1.861024	6.19	0.45	0.56	2026	2521	0.73
	6	biology	1520251	11803064	15.53	161025	4.92	0.088	0.6	9089	61967	0.88
	7	email	59912	86300	1.44	4.861025	4.95	-	0.16	-	40524	n/a
	8	email address	16881	57029	3.38	461024	5.22	0.17	0.13	1301	995	0.64
	9	student	573	477	1.67	0.0029	16.01	0.005	0.001	1.34	0.27	n/a
	10	newspaper	459	1422	6.2	0.0135	2.98	-	0.02	-	1.67	n/a
	11	US directors	11057	74414	13.46	0.0012	5.19	0.56	0.87	315	494	0.77
	12	UK directors	8850	39741	8.98	0.001	6.46	0.61	0.89	386	561	0.71
	13	German	4185	30438	14.55	0.0035	6.4	0.72	0.93	100.71	129.7	0.79
Information	14	WWW nd.edu	269504	1497135	5.56	461025	11.27	0.11	0.29	3453	9104	0.81
	15	Roget's	1022	5103	4.99	0.0098	4.87	0.13	0.15	23.54	27.17	0.76
	16	word	112	425	7.59	0.0684	2.54	0.16	0.17	2.13	2.34	0.74
	17	book	105	441	8.4	0.081	3.08	0.35	0.49	3.09	4.33	0.71
Technological	18	Internet	10697	31992	5.98	5.661024	3.31	0.035	0.39	98.09	1093	0.83
	19	power grid	4941	6594	2.67	5.461024	18.99	0.1	0.08	84.45	67.56	0.8
	20	train routes	587	19603	66.79	0.114	2.16	-	0.69	-	4.26	n/a
	21	software	1439	1723	1.2	0.0017	2.42	0.07	0.082	1403	1644	n/a
	22	software	1377	2213	1.61	0.0023	1.51	0.033	0.012	285.26	103.73	n/a
	23	electronic	24097	53248	4.42	1.861024	11.05	0.01	0.03	33.5	100.5	0.91
	24	peer-to-peer	880	1296	2.95	0.0034	4.28	0.012	0.011	5.26	4.82	0.85

Another popular research issue is the size exploration of World Wide Web and its small world behavior, particularly the unique hyperlink structure (Albert, Jeong and Barabási, 1999; Adamic, 1999; Watts, 2000; Allen, 2000). The total number of pages on the Web is estimated to be over 8 x 10^8 documents (Albert et al., 1999) and the total number of hyperlinks would never be a stable figure due to constant updates. When Albert et al. (1999) studied the World Wide Web in their research, they used robots to obtain certain statistics and found that the web page hyperlinks follow the power law. They concluded that two randomly chosen documents on the web are on average 19 clicks away from each other.

In an another study Adamic instead looked at the World Wide Web from the perspective of Watts-Strogatz model in which properties L and C were calculated based on all sites and a subset of ".edu" sites. Firstly, she considered "undirected shortest paths" which had an average path length of 3.1 "clicks" between any two connected sites. Then she analyzed directional paths between sites, used 64,826 sites, and found that the average path length was 4.228 (adamic 1999).

Parongama Sen, Subinay Dasgupta,Arnab Chatterjee,P.A.Sreeram,G.Mukherjee and S.S.Manna recently carried out a experiment on Indian railway networks[6].They take the stations as nodes of the network and a train which stops at any two stations as the link between the stations. This table contains a total of L=579 trains covering N=587 station. They found the path length $l_{(ij)}$ is 2.16 and the clustering coefficient C(N) found 0.69. They also observed that the mean distance of Indian railway network varies logarithmically with the number of nodes with a high value of the clustering coefficient, which shows that Indian railway network behaves like a small-world network. Besides this they also calculated degree distribution of numbers of trains which stop arbitrary stations also calculate correlation of IRN.

In an another study on "The anatomy of the facebook social graph"[7],they studied the numerous features of facebook like degree distribution, small world effect, mixing patterns ,degeneracy graph.They carried out this research on three observation, global structure of the network for active user,on average local clustering coefficient and assortativity of the graph .They analyzed this reaearch on two scale, facebook user as in globel scale and as in US based facebook user.The small world effect and six degree separation were confirmed in globel scale and found average distance between user on facebook in may 2011 was 4.7 while the average distance for U.S. based user is 4.3.Also found 92% of all pairs of facebook users were within five degree separation and 99.6% were six degree separation. For US based user 96% were in five degree and 99.7% were within six degree which proves the six degree separation concept in facebook.The degree correlations for facebook is r=.226 ,calculated using Pearson correlation coefficient. They characterized the structure of the facebook graph by the many metrics and tools which was one the largest experiment of this kind.

Fowler, James H [8], studied relationship between small worldness properties on voter's personal networks and voter turnout. Experiments carried out data from 2176 voters surveyed in Huckfeldt and Sprague's 1966, Indianapolis-St. Lousis election duty. They found the influence of committed voters on clustering increases as well in other voter's personal networks. They also proved the six degree separation concept on voters that a friend is a friend of friends is 0.61 and probability of two friends conversation is 0.47.

Besides these, researchers have focused and applied different aspects of different networks on small world. Vito Latora and Massimo Marchiori [10] introduced the efficiency of a network and try to calculate how the information exchanges over the networks and show efficiency in terms of globally and locally in networks like brains, communications and transportation networks. They found E_{glob} and E_{loc} for WWW is .28 and .36,similarly E_{glob} and E_{loc} internet .29 and .26.Also calculated efficiency for unweighted and weighted networks in Boston underground transportation networks(MBTA) E_{glob} and E_{loc} .10 and .006 (unweighted) and E_{glob} and E_{loc}

Another new method was presented by Faraz Zaidi [12] which creates a graph with small world properties. They randomly generate a graph and replace each node with cliques on different sizes and connectivity between cliques is choosing randomly. Also calculated relative density and modularity of this clustered networks had community structures. Firstly they generate random networks and replace with the triad which is set of three node connected by three edges then an edge is placed between two triads which exhibits the two structural properties of small world network.

By Zengwang and Daniel [13] study small world networks from the perspective of autocorrelation and investigates small world properties in spatial networks. Here they used Moran's *I* and Getis-Ord's G to study in rewiring process of the networks. They applied these techniques in three different geographical scale, national level for US interstate high way, metropolitan level for road networks in Huston-Galveston area and intra-city level for Boston subway networks. The auto-correlations results on this networks capable of detecting the critical threshold in the emergence of small world phenomenon.

V. Conclusion: The main aim of this paper is to explore the areas related to the study on small world phenomenon also try to cover different methods, techniques that are applied to generate small world effect in a network. These fields could be anything that might include from social, physical sciences and to even life sciences etc where properties of a complex networks exhibits. The world is now a day's small enough than ever before to communicate. Any random two people are easily connected due to social networks or use of other forms of networks. Small world have some properties which shows short cut way of information flowing which is try to show here by defining clustering co-efficient and shortest length path length of numbers of information networks also able to calculate the efficiency of that networks as mentioned above.

REFERENCES:

1] Milgram S .1967. The small world problem. Psychol Today 2:60–67.

2] Jie Cheng , Mehdi R. Zargham , Shuguo Rang. March 2013. "Small World Phenomena" IJCSET, Vol 3, Issue 3, 71-75.

3] M. E. J. Newman, D. J. Watts, S. H. Strogatz. Feb 2002. Proceedings of the National Academy of Sciences. 99 (suppl 1) 2566-2572; DOI: 10.1073/pnas.012582999.

4] Giuliano Andrea Pagani and Marco Aiello. 2012. "The Power Grid as a Complex Network : a Survey", March22.

5] Humphries, Mark D. and Kevin N. Gurney. 2008. "Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence." PLoS ONE 3: 167 - 256.

6] Parongama Sen, Subinay Dasgupta, Arnab Chatterjee, P.A.Sreeram, G.Mukherjee, and S.S.Manna. 2003. "Small-world properties of the Indian railway network".

7] Ugander, J., Karrer, B., Backstrom, L. & Marlow, C. 2011. "The Anatomy of the Facebook Social Graph". *CoRR*, abs/1111.4503.

8] Fowler, James H .2005. "Turn out in a Small World". Philadelphia,PA:Temple University Press.

9] Uzzi, Brian and Amaral, Luis A. N. and Reed-Tsochas, Felix. 2003. "Small-World Networks and Management Science Research: A Review". European Management Review, Vol. 4, Issue 2, pp. 77-91.

10] Latora V, Marchiori M. 2001. "Efficient behaviour of small-world networks". PhysRev Lett 87: 198701.
11] D.J.Watts and S.H. Strogatz.1998."Collective dynamics of 'small-world' networks".Nature(London) *Nature* volume 393, pages 440–442.

12] Zaidi, F. 2012, 21 february."Small world networks and clustered small world networks with random connectivity". Soc. Netw. Anal. Min 3: 51.

13] Zengwang Xu , Daniel Z. Sui. 3 April, 2007. "Small world charterristics on transportation networks: a perspective from network autocorrelation". Springer-Verleg.

14] Newman MEJ,Watts DJ. 1999. "Renormalization group analysis of the small-world network model". Phys Lett A 263:341–346.

15] Newman ME, Girvan M. 2001. “Finding and evaluating community structure in networks” . Phy Rev E Stat Nonlin Soft Matter Phy 64.

16] H. Herzel. 1998. How to quantify ‘Small World’ Networks?. Fractals 6(4) .

17] Albert R, Jeong H, Barabasi AL .1999. “Diameter of the World Wide Web”. Nature 401:130–131.

18] Gaure, J, “Six degree of Separation : A Play” ,Vintage ,New York, 1990.

19] Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. 2011. The ubiquity of small-world networks. *Brain connectivity*, *1*(5), 367–375. doi:10.1089/brain.2011.0038.