

The Research on Knowledge Diffusion Based on Small World Network

Xinxin Feng¹, Baojiang Chen¹, and Huanzhi Zhu²

¹ Beijing University of Civil Engineering and Architecture, Beijing, China
{fengxinxin, chenbaojiang}@bucea.edu.cn

² Hebei United University, Tangshan, Hebei, China
693364322@qq.com.cn

Abstract. Small world is a sort of network between regular and random network that has the feature of both shorter characteristic path length and higher clustering coefficient, which is an appropriate model for the research of knowledge diffusion. There is simulation in this paper to certificate the validity of small world network in knowledge diffusion. In order to reflect the real world the model is completed by including expert as one influence factor and currency as the other to reflect knowledge transaction. Simulation results show that small world network do exist in knowledge diffusion.

Keywords: Small world, knowledge diffusion, expert.

1 Introduction

The varieties of elements in the nature construct numerous huge and complex systems. The diversity and uncertainty of these systems leads to the complexity of the research works on them. As the development of science and technology, people find that network is a very good expression to the complex systems. The theory of the complex networks was raised in the late twentieth century; it is the powerful tool for the research of complex system. The small world network is the latest hotspot in the research of complex networks. Lots of interesting experiments has been executed to uncover the phenomenon of “the world is small”. As Interdisciplinary, the small world theory has rapid growth and applied into many different fields.

At the same time, knowledge management is also the hottest topic in the twentieth century. The most important forms of knowledge management are diffusion. To take the small world theory into the study of knowledge management is a more and more extensively researched topic both in theory research and actual research.

In real world, the diffusion of knowledge has different forms inside different organized system. Such as the “team-work” or the “small organize” in some enterprise, or the “group teaching” in the school, are all the good research material for the study of small-world network.

2 Small World Theory

2.1 The Development of Small World Theory

Small world theory was originated in 1960's by the social psychologist Stanley Milgram in his research named "Tracking the shortest path to the rule of the social network in the United States". In this experiment, he asked the subjects to mail to a people who lived nearby Boston, and constrained that they can only mail to someone they know. In the end, he realized that to reach the target, there is average 6 times to forwarding the mail. This is the famous six degrees of separation, or the small world phenomenon[1]. This phenomenon reviews that anyone who doesn't know each other; they only need Six degrees of separation to get the link between them. Before this theory, most research of networks are based on its topological structure like regular or random. While in real world, it is neither regular nor random, but as a topological structure between them. In1998, Watts and Strogatz put forwards the small world network theory on the journal of science. They studied three real networks, Elegans worm network, US western power network and movie star corporation network, and realized that they all have the character of the small world networks[1].

2.2 The Structure and the Characteristic Parameters of the Small World Theory

Watts and Strogatz provided the construct algorithm of the small world theory: first begin with a regular annular network which has n vertexes and each vertex has k edges. The probability p represents the random degree of the network. Then breaks the edges and reconnects between the vertexes with limits of probability p with the restriction of no self-connection or duplicate connection, this will generate $pkn/2$ edges (it is called shortcuts)[2]. By adjust the value of p , the network can be changed to regular ($p=0$) and random ($p=1$), the so called small world is the between area of $0 < p < 1$, as shown in figure 1.

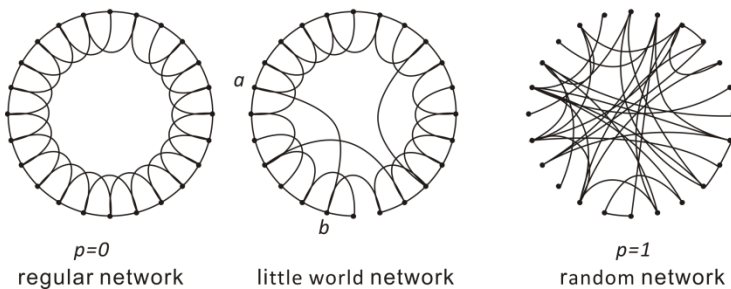


Fig. 1. The relationship of the regular, small world and random network

W-S' small world network model has two parameters to describe the features of the structure, one is the characteristic path length $L(p)$, the other is the clustering coefficient $C(p)$. $L(p)$ means from one vertex to the target vertex the minimum edges that must be passed, it reflect the difficulty of linking one vertex to the other. $C(p)$ shows

the local characteristics of the network. That decides whether there are relatively stable sub systems in the whole network, in another word the degree of association of the neighbor vertexes. In the research of Watts and Strogatz, when $p \rightarrow 1$, $L \sim n/2k \gg 1$, $C \sim 3/4$; when $p \rightarrow 0$, $L \approx L_{random} \sim \ln(n)/\ln(k)$, $C \approx C_{random} \sim k/n \ll 1$ [3]. Conclusion: When network tends to regular, it has higher clustering coefficient, but longer characteristic path length; on the contrary, when network tends to random, it has lower clustering coefficient, but shorter characteristic path length, so any two vertexes in the network are easily to be linked in random network. When increase the numerical of p , network moves from the regular to the random, there is an area has the feature of both shorter characteristic path length and higher clustering coefficient, this area is called small world network.

3 Theory of Knowledge Diffusion

The knowledge flowerage can be divided into two style, transaction type and broadcast type. The difference between these two types is when during flowerage, there is condition or not. The former has no condition, and the rare has [4]. To compare the effect of the small world network and the scale-free network, the broadcast type is taken in this paper, because when there are conditions, the separation degree of the result will be poor.

Cowan and Jonard use network model to simulate knowledge diffusion in human society. They discover that the most efficient knowledge diffusion is happening in small world network, but also has the most unfair knowledge spread (which means that the knowledge spread is unfair for everyone)[3]. They put a limitation to the model, that the diffusion only occurs when the both side of the diffusion have the same degree of knowledge. This makes the diffusion process actually a swap process. But swap is not the only way of knowledge exchange. First, knowledge exchange is not barter. Second, sometimes knowledge exchange is one-way transfer when communicate with a friend. Pier Giuseppe and Richard (2001) have improved the model of Cowan and Jonard. But both of them didn't solve the problem mentioned above. And there is another situation that they didn't notice, knowledge retention. It's a very similar situation in company, because some knowledge is the core competitiveness of an enterprise, they will not share it. Even in personal contact, sometimes there will be retention too.

To isolate the effect away from the structure of network, which can make lots difference to knowledge diffusion; most researchers use the way of changing the edges connected to the vertex in the network. This method is concentrated on the density of the network but not structure. While in this research, structure is the most important factor but not density. So in the model of this paper, the number of edges is fixed, only the link of vertexes in space distribution is changed.

4 Model of Small World Network

Small world model is a sort of one-dimensional regular lattice with some randomness by breakpoint reconnects. The construction algorithm is: first begin with a regular

network ring, which has N vertexes, every vertex has n edges that connect with n neighbor, and meet the constraint $N \gg n \gg \ln(N) \gg 1$; then, for every edges, relink it to another vertex with probability p with no repeated edges. By this way, there will be $(p \cdot n \cdot N)/2$ longer edges to connect one vertex to another faraway vertex.

For example, to vertex A_0 , the vertexes connect with it are A_1, A_2, \dots, A_n , then the maximum connect number of vertexes will be $b = n \cdot (n + 1)/2$, the actual number of edges connected is m , then the gathering coefficient will be m/b . The gathering coefficient of the whole network is the average of all the gathering coefficient of every vertex calculated with this formula[5]:

$$C(p) = \frac{1}{N} \sum_{i \in I} \sum_{j, l \in \Gamma(i)} \frac{X(j, l)}{|\Gamma(i)|(|\Gamma(i)| - 1)/2}$$

In this formula, if $j \in \Gamma(i)$, then $X(j, l) = 1$; or else $X(j, l) = 0$. L as the average distance is to reflect the overall features of the network structure. If two vertexes in the network can be connect with several edges that connect with each other sequential, then we called these two vertexes is reachable. The minimum number of edges to connect two vertexes is called the distance. If the two vertexes is not reachable, then we define their distance is infinity. The network's average distance is calculated by average all the distances of very pair of vertexes in the network. If we take $d(i, j)$ as the distance of vertex i to vertex j , then the average distance of the network will be[5]:

$$L(p) = \frac{1}{N} \sum_{i \in I} \sum_{j \neq i} \frac{d(i, j)}{N - 1}$$

When $p=0$ and N is big enough, $C(0) = 3/4 \cdot (n - 2)/(n - 1)$, $L(0) \sim N/(2n)$; when $p=1$ and N is big enough, $C(1) \sim n/N$, $L(1) \sim \ln N / \ln n$. The result is higher degree of aggregation always leads to longer distance, and lower degree of aggregation leads to shorter distance. While in small world networks model, $L(p)$ is very close to $L(1)$, and $C(p)$ is much bigger than $C(1)$. To fulfill this, shortcut is added into the network to make the average distance shorter, this method is called reconnect. By this way, L is approaching the level of random network, while C is approaching the level of regular network.

Corresponding to real world, people or enterprise (called agent in this paper) is the vertex of the network, and the link of them is the edges. The clustering coefficient C expresses the gathering level of the agent in the real world. When C is high, there are a lot of clusters in this network, like enterprise alliance for the corporation or the coterie for personal. Distance L is representing the tightness of the agent. If L is short, means that two agents has very close relationships, like the parent and subsidiary companies, the link between them is called strong ties. On the contrary, that is the weak tie [6]. Convenient for research, take the weights of the agent to 1, and suppose the edges are undirected. With these assumptions the model mentioned above then can be put into research. For every agent in the network, a vector is defined to

indicate their knowledge level, the dimension of the vector indicate the types of the knowledge they have. The knowledge level of the agent will change by knowledge diffusion as time goes by.

Expert is another issue that should be mentioned. In real world, there are always some people they grasp much more knowledge than others, but they usually will have some reservations in diffusion. For example when an expert wants to maintain his core competencies, he will stop to diffuse his knowledge to those non-experts when they reached some height [7]. So in this model, we assume that only part of knowledge will absorb by the audience during knowledge diffusion. A coefficient α is introduced in, and $0 < \alpha < 1$, it means the ratio of diffusion result with the initial difference of the knowledge level from the both side of the diffusion. By define $V_{i,k}(t+1) = V_{i,k}(t) + \alpha \cdot (V_{j,k}(t) - V_{i,k}(t))$, the reservation of knowledge diffusion is reflected by this formula. The more difference of the knowledge level, the more the receiver can acquire knowledge, and the incremental of each knowledge diffusion will decrease. It is according to the real situation.

5 Simulation and Analysis

5.1 Simulation

In this paper, Matlab was used as the analysis tool. The network is designed as: 500 vertexes and every vertex have 20 edges. So in the model we define $N=500$, $n=20$. In this analysis, the only variable is the probability p . The average distance L and average clustering coefficient C are all calculated with this parameter p .

To characterize small world networks, a notion of far-edge is introduced in to identify the “faraway” connection in the network. In figure 1, the edge from vertex a to vertex b, are the so-called far edge.

The algorithm design is as follow: From the beginning, the network is a regular network, as the reconnection with probability p from 0 to 1, the average distance L and average clustering coefficient C changed [8]. To reflect the changing trend, $L(p)/L(0)$ and $C(p)/C(0)$ should be taken to standardize the L and C . The quantity $L(p)$ denotes the average length of the shortest path between any two vertexes, and the $C(p)$ denotes the average of C_v over all the vertexes, while C_v is the number of edges connecting the neighbors of vertices v , normalized with respect to the maximum number of possible edges between these neighbors[2]. With this definition, we introduced the variable f as the ratio of far edges with the total number of edges. When p is very small as $p \leq 0.1$, f is approximately equal with p . When p get bigger, f grow slowly. When $p=1$, f saturates to a value of about 0.2. So in this paper, the simulation is based on the variable of f , but not p . Thus we plot $C(f)/C(0)$ and $L(f)/L(0)$ as functions of f . Actually this figure has the same expression of $C(p)/C(0)$ and $L(p)/L(0)$ [2], as shown in figure 2.

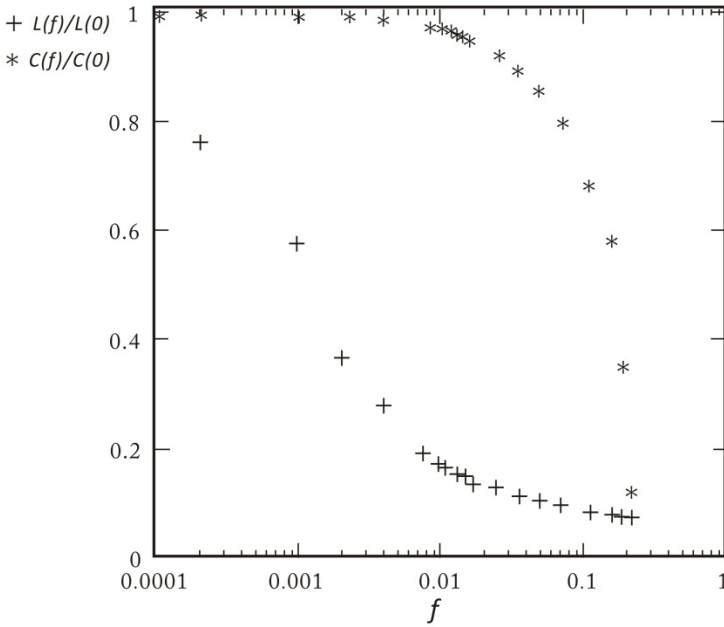


Fig. 2. $C(f)/C(0)$ and $L(f)/L(0)$ as functions of f

5.2 Analysis

From figure 2, we could clearly find that with the change of p , the network experience 3 stage: regular, small world and random. In first stage, f is close to 0 correspond to the area 0.001 to 0.01, reconnection is seldom happened, the network reflect the characteristic of the regular style. In this region L and C both have bigger numeric. As f getting bigger in region 0.01 to 0.1, L declined rapidly, this reviewed that the reconnection of several edges could make much reflection on the construct of the network, especially on distance. But to the average clustering coefficient C there's not much influences, which means the changing of f didn't affect the local relationship of the network, this phenomenon is called small world. When f increases more than 0.1, L is stabilized, while C declined sharply, the network shows the characteristic of random network.

The small world network can be identified within the region of $C(p)/C(0) \approx 1$ and $L(p)/L(0) \approx L(1)/L(0)$, this region is correspond to $f \approx 0.01$. So f can be used to characteristic network between regular, small-world and random. It is more Intuition than p in identifies the style of the network.

In the simulation, the knowledge of each agent is expressed with a five dimension vector. The initial knowledge level of every vertex is given randomly from 0 to 9. The ratio of expert in this model is set to 1%[9]. That is, pick 1 agent as expert and assigns its level of knowledge to 9 from every 100 agent. To simulate the situation of

retention, when one agent's knowledge is grown up to 7 by diffusion, 90% of the experts will stop diffusion to him to maintain their own superiority. That mentioned before to identify the effect of diffusion is set to 50%, which means that only half of the knowledge gap can be transferred in every diffusion [7], [9]. Another premise is that diffusion is happened in all the edges that connect to the vertex. To separate the small world region, the reconnection ratio is grow with the step of 0.001, with every value of p , there are 8 times of diffusion [7]. One time of diffusion means diffuse from one vertex by all the edges connected with it.

In the network of the real social environment with strong ties, people or enterprise is very familiar with each other. Since people have the same knowledge background, there's actually not too much knowledge to diffuse. While in social environment with weak ties, the different clusters can be connected, that will generate new channel for knowledge and conception diffusion [4]. As in real world, good friends usually exchange their knowledge and conceptions, but this knowledge is much the same. While people with ordinary relationships, they won't exchange their knowledge unreserved, retention will happen more or less. But when these clusters are connected, new knowledge and conception with much effect can be diffused. This phenomenon will also happen between the enterprises. This is the advantages of the weak ties. The reconnection of the small-world is the same effect as the weak ties. Also reconnection has little effect on the local environment of the network, but has big effect on the whole structure. That reflect the meaning of $L(p)$ and $C(p)$ in the small-world. When $L(p)$ is smaller, the net connection is shorter, which means the knowledge diffusion is faster. When $C(p)$ is higher, the degree of aggregation is higher, it is good for the knowledge diffusion in local circumstances but not good for other clusters and the whole network.

In this research we discovered that the average knowledge level is not the monotonic function with p , it reach maximum at the region of small-world area ($0.005 < p < 1$), then as p grow up, the average knowledge descend slowly and finally remained at a certain area. This come to the conclusion that the higher gathering degree and the shorter average distance make the most efficient knowledge diffusion. $L(p)$ and $C(p)$ are not irrelevant.

The gap between the different agents is another focus in research to show the fair degree of knowledge diffusion. After several diffusions, the difference of knowledge level is quite similar; this phenomenon is different with the research of Cowan and Jornard's. In their model, the difference of the knowledge level is the maximum in the small world network, because if the reconnection ratio p is equal to 0 or very small, the gathering of groups is high. It leads to the diffusion of knowledge in the group is very fast, while the diffusion between the groups are very slow. This will lead to such result: group without profession will remain a low level of knowledge after diffusion; while the group with profession will get a higher knowledge level after diffusion. This will lead to knowledge gap. When the parameter p getting higher, the degree of aggregation will get lower, this means the small group gradual disintegration. In the other word, the difference of the network becomes smaller. But in our model, two important influencing factors are introduced, one is the currency which indicates the knowledge transaction between the corporations in real world, the other is the

unilateral knowledge diffusion which indicates the situation between personal. These two situations of knowledge diffusion reduce transaction costs and improve the efficiency of the transaction. Offset the adversely affected of the high degree of aggregation to some extent, which leads to reduce the difference of knowledge level between the agents in the network and makes the diffusion more equitable in the network.

6 Conclusion

By including expert as one influence factor and the currency in knowledge transaction, the small world network model is more completed to reflect the real world. Simulation results show that small world network do exist in knowledge diffusion. In small world network, the whole knowledge level is the highest among the regular ($p=0$) and random ($p=1$) network after knowledge diffusion. And it has fairer diffusion opportunity.

But in this paper it is a relatively ideal state. In real world, people wish that the group has a faster knowledge diffusion speed and still remain the knowledge distribution inhomogeneity. This can both integrate the group's overall capacity, and remains the competitive pressures within the group which can lead to the long-term development of the group. The result in this paper not touches this area yet. The other factor that not discussed in this paper is the innovation of knowledge which still needs further research.

References

1. Cowan, R., Jonard, N.: Network Structure and the Diffusion of Knowledge. In: MERIT Working Papers, 99–128 (1999)
2. Pandit, S.A., Amritkar, R.E.: Characterization and control of small-world networks. arXiv: chaos-dyn/9901017v1 (January 16, 1999)
3. Cowan, R., Jonard, N.: Statistical mechanics of complex network. *Economic Dynamics & Control* 28, 1557–1575 (2004)
4. Plouraboue, F., Strayer, A., Zimmermann, J.-B.: Learning Induced Criticality in Consumers' Adoption Pattern: A Neural Network Approach. *Economics of Innovation and New Technology* 6, 73–90 (1998)
5. Latora, V., Marchiori, M.: Economic Small-world Behavior in Weighted Networks. *The European Physical Journal* 32(2), 249–263 (2003)
6. Kleinberg, J.: The Small World Phenomenon and Decentralized Search. *SIAM New.* 37(3), 1–2 (2004)
7. Feng, H., Li, Z.: Network Model of Knowledge Diffusion and Its Implications. *Journal of the China Society for Scientific and Technical Information* 25(1) (February 2006)
8. Watts, D.J., Strogatz, S.H.: Collective dynamics of "small-world" network. *Nature* 393(6), 440–442 (1998)
9. Computer Programs for Social Network Analysis [EB/OL], <http://www.insna.org/INSNA/softinf.html>, 2007-03-27