

DIFFUSION OF INNOVATION AND COLLECTIVE
ACTION IN COMPLEX NETWORKS

MURAT MUSTAFA TUNÇ

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of the requirements for the degree of
Master of Science

Sabancı University

August, 2015

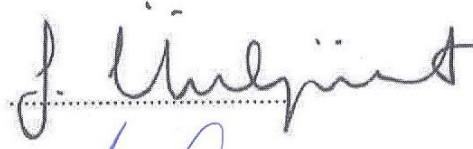
DIFFUSION OF INNOVATION AND COLLECTIVE ACTION IN
COMPLEX NETWORKS

APPROVED BY:

Prof. Dr. Ali Rana Atılgan
(Thesis Supervisor)



Assoc. Prof. Dr. Tongu Ünlüyurt



Asst. Prof. Dr. Ahmet Onur Durahim



DATE OF APPROVAL: ...04.08.2015

© Murat Mustafa Tunç 2015

All Rights Reserved

DIFFUSION OF INNOVATION AND COLLECTIVE ACTION IN COMPLEX NETWORKS

Murat Mustafa Tunç

Master of Science in Industrial Engineering

Thesis Supervisor: Ali Rana Atılgan

Abstract

In this thesis, we implement the existing models for the diffusion of innovation and we present a novel model for the collective action in complex networks. We also create a new network called the Robin Hood network using poor get richer approach. The diffusion models are divided into three components, namely node, link and neighbor diffusion. Each diffusion model is simulated in ring lattice, small worlds, random networks and scale free networks in order to analyze the effect of network topology. The simulation results suggest that low level of clustering coefficient and average path length, as well as the presence of hubs are efficient for the innovation to become widespread. Comparative results for the collective action show that the presence of hubs is insignificant for determining the size of the collective subset. On the other hand, the tightly-knit communities are effective for the collective decision making. We also provide sufficient parameters for the modelling of networks where both the diffusion and collective action take place. According to our results, the clustering plays an important role in both diffusion and collective action. However, the presence of hubs is not a must for collective action.

Keywords: Collective Action, Diffusion, Clustering, Small Worlds, Robin Hood Network

KARMAŐIK AĐLARDA YENİ FİKİR YAYILIMI VE ORTAKLAŐA DAVRANIŐ

Murat Mustafa Tunç

Endüstri MühendisliĐi Yüksek Lisansı

Tez Danıőmanı: Ali Rana Atılgan

Özet

Bu tezde karmaőik aĐlarda hali hazırda var olan yeni fikir yayılım modellerini kullanıyor ve yepyeni bir ortaklaőa davranıő modeli sunuyoruz. Ek olarak fakirden zengin yapan bir yaklaőım ile Robin Hood aĐı adında yeni bir aĐ sunuyoruz. Yayılım modellerini düĐüm, baĐ ve komőu yayılımları olarak üç bileőene ayırıyoruz. Her bir yayılım modeli kullanılarak aĐ yapısının etkilerini incelemek için halka örü, küçük dünya, rastlantısal ve orantısız aĐlarda benzetim yapıyoruz. Benzetim sonuçları gösteriyor ki hem kümelenme katsayısının ve ortalama yol uzunluĐunun düşük seviyelerde olması hem de merkez düĐümlerin varlıĐı yeni fikirlerin geniő alana yayılması için etkilidir. Ortaklaőa davranıő modelinden elde edilen karőılaőtırmalı sonuçlar ise merkez düĐümlerin varlıĐının ortaklaőa davranan düĐümler alt kümesinin boyutunu belirlemede yetersiz kaldıĐını göstermektedir. Öte yandan, sıkı sıkıya kenetlenmiő topluluklar ortaklaőa karar verme konusunda daha etkilidirler. Ayrıca, hem yayılım hem de ortaklaőa davranıőların olduĐu aĐların modellenmesi için elveriőli deĐiőkenler de sunuyoruz. Sonuçlarımıza göre kümelenme hem yayılımda hem de ortaklaőa davranıőta önemli rol oynamaktadır. Fakat merkez düĐümlerin varlıĐı ortaklaőa davranıő için olmazsa olmaz deĐildir.

Anahtar Kelimeler: Ortaklaőa Davranıő, Yayılım, Kümelenme, Küçük Dünya, Robin Hood aĐları

*This work is dedi-
cated to my beloved parents & my sis-
ter, Ebru & my precious, Gizem & IE grads,
especially Özgün, Ece, Can, Faran, Yağmur,
Bahar, Başak, İhsan, Amir, Burcu, Merve,
Berk, Sonia & my blues brothers, Ali
& Çağdaş & my band, Moobs &
countless friends who be-
lieved in me*



Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor Ali Rana Atılgan for his help and encouragement during the course of my master's thesis.

I want to thank Tongu Ünlüyurt and Ahmet Onur Durahim for accepting to be part of thesis jury.

I gratefully acknowledge the funding received from TÜBİTAK BİDEB to complete my master degree.

I also would like to express many thanks to Sabancı University for the scholarships received and for becoming my home for the last seven years.

Contents

1	Introduction	1
2	Diffusion of Innovation with Infinite Speed in One-Dimensional Networks	5
2.1	Active Link Diffusion with Infinite Speed	6
2.2	Active Node Diffusion with Infinite Speed	9
2.3	A Different Application: Infection and Vaccination	11
2.4	Future Models	14
3	Diffusion of Innovation with Infinite Speed in Two-Dimensional Networks	15
3.1	Active Link Diffusion with Limited Information	17
4	The Diffusion of Innovation in Complex Networks	20
4.1	Diffusion Models	23
4.1.1	Node Diffusion	23
4.1.2	Link Diffusion	26
4.1.3	Neighbor Diffusion	27
4.2	The Effect of Clustering Coefficient and Average Path Length in Diffusion . .	29
4.3	The Effect of Hubs in Diffusion	35
4.4	Time to Equilibrium in Diffusion	38
5	Collective Action	45
5.1	Definitions	45
5.2	Collective Action Algorithm	46
5.2.1	Finding the Giant Cluster of Potentials	46
5.2.2	Finding the Collective Subset	46
5.3	The Effect of Clustering Coefficient and Average Path Length in Collective Action	48
5.4	The Effect of Hubs in Collective Action	49
5.5	Conclusions	50
5.5.1	Motifs for Collective Subset in 2D Regular Lattice	53

List of Figures

2.1	A one-dimensional network with N nodes and $N - 1$ links	5
2.2	An example of active link diffusion with infinite speed in one-dimensional networks	7
2.3	An example for the active node diffusion with infinite speed in one-dimensional networks	10
2.4	An example for the active link diffusion in one-dimensional networks	13
3.1	An illustrative example of the diffusion model with perfect information in two-dimensional networks	16
3.2	An illustrative example of the diffusion model with limited information in two-dimensional networks	17
3.3	An example of the active link diffusion model with limited information in two-dimensional networks	18
4.1	An example for the node diffusion in two-dimensional networks	25
4.2	2D regular lattice and ring lattice with N nodes and $2N$ links	26
4.3	The fraction of adopters with respect to the fraction of potential nodes in node diffusion in 2D regular lattice and ring lattice	27
4.4	The fraction of adopters with respect to the fraction of active links in link diffusion in 2D regular lattice and ring lattice	28
4.5	The fraction of adopters with respect to the fraction of potential nodes in neighbor diffusion in 2D regular lattice and ring lattice	29
4.6	The clustering coefficient and average path length with respect to the rewiring probabilities	31
4.7	Ring lattice, small world and random networks	31
4.8	The effect of rewiring probabilities in node diffusion	32
4.9	The effect of clustering coefficient and the average path length in node diffusion	32
4.10	The effect of rewiring probabilities in link diffusion	33
4.11	The effect of clustering coefficient and the average path length in link diffusion	33
4.12	The effect of rewiring probabilities in neighbor diffusion	34

4.13	The effect of clustering coefficient and the average path length in neighbor diffusion	34
4.14	An illustration of the rich get richer model	35
4.15	An illustration of the poor get richer model	36
4.16	An illustration of the random get richer model	37
4.17	The degree distribution of rich get richer model in logarithmic scale.	37
4.18	The degree distribution of poor get richer model in logarithmic scale	38
4.19	The degree distribution of random get richer model in logarithmic scale	39
4.20	The effect of hubs in node diffusion	39
4.21	The effect of hubs in link diffusion	40
4.22	The effect of hubs in neighbor diffusion	41
4.23	The time to equilibrium in node diffusion for ring lattice, small world and random networks	42
4.24	The time to equilibrium in node diffusion for hub networks	42
4.25	The time to equilibrium in link diffusion for ring lattice, small world and random networks	43
4.26	The time to equilibrium in link diffusion for hub networks	43
4.27	The time to equilibrium in neighbor diffusion for ring lattice, small world and random networks	44
4.28	The time to equilibrium in neighbor diffusion for hub networks	44
5.1	Finding the giant cluster of potential nodes in 2D regular lattice	47
5.2	Finding the collective subset in 2D regular lattice	48
5.3	The effect of clustering coefficient and average path length in collective action	50
5.4	The effect of hubs in collective action	51
5.5	An illustrative example that the collective subset is independent from the size of the giant cluster of potential nodes.	52
5.6	An illustrative example that the collective subset is subject to the position of the potential nodes in 2D regular lattice	53
5.7	Collective subset motifs in 2D regular lattice.	54

List of Tables

2.1	The computational results for the time to cover all the nodes in one-dimensional networks and the analytical solution	9
2.2	The computational results for the time that vaccination catches the epidemic .	14
3.1	The computational results for the time to reach to the target node in two-dimensional regular lattice	19

1 Introduction

The structure of relationships shapes the social interactions and decisions. For instance, the transmission of virus depends on the nearest contacts of the transmitter. Likewise, the job offerings are usually accessed via friends. Therefore, the contacts or friends play a significant role for decision making. Understanding the local structure of relations can help better prediction for the outcomes. Thus, the study of networks is important to anticipate the outcomes of the economic, political and social decisions. In recent years, the study of networks has attracted many researchers from various disciplines. Sociologists, economics, computer scientists, mathematicians and statistical physics have studied networks from different perspectives. Easley and Kleinberg's book is very influential for the collection of the networks literature [1].

There are so many examples in the literature that prove network influences behavior. In particular, how people find about their jobs has been studied extensively. Myers and Shultz [2] surveyed textile workers in the late 1940s. When Myers and Shultz asked the textile workers about how they found their first job, 62% of the workers replied that they found their first jobs through a contact who had already been working in the textile industry. In contrast, only 23% of the textile workers found their first jobs by direct applications. Another similar study by Rees and Shultz [3] suggests that 37% of typists, 65% of janitors, 23% of accountants, 57% of electricians and 73% of material handlers found their jobs through contact. There is also a famous study by Granoveter [4] that suggests the weak ties (i.e. casual acquaintances rather than close friends) play an important role for finding jobs. The literature for the network influence over labor markets can be found in a recent study by Ioannides and Loury [5].

The diffusion in networks has been also studied extensively. The summary of the diffusion literature can be found in Jackson and Jariv [6]. In particular, Coleman et al. [7] studied how the new drugs are prescribed by the doctors. They showed that the adoption of new drugs by doctors differs depending on the degree of the doctors' contact. Another study for the diffusion is the adoption of hybrid corn among the farmers [8] [9]. The hybrid corn was developed by mixing the genetics of different corn species and marketed in the 1930s. The adoption of hybrid corn took different years depending on the states; however the adoptions over the years have S-shapes for every state. The adoption starts slowly, but after a threshold it becomes widespread. These S-shapes can help better predictions for the time that an innovation becomes widespread throughout the network.

The role of individuals for diffusion also shows the influence of local structures in the networks. The role of opinion leaders in the societies has been studied by several researchers [10] [11]. Katz and Lazarsfeld [10] interviewed women in Illinois for a variety of topics like fashion, movies, etc. They traced the changes in the opinion which eventually led to the presence of the opinion leaders. Lazarsfeld et al. [11] studied the voting decisions in Ohio for the presidential campaign. Their results suggested that the opinion leaders plays significant role in voting decisions.

The diffusion of new products and the success of new products are also shown to be influenced by certain individuals in the networks. Fick and Price [12] identified the market mavens and showed that whether the product reaches to the market mavens affects the success of the new products. They showed that consumers recognize the market mavens in their social networks and make use of them. Tucker also [13] analyzed the effect of influence for the adoption of video messaging technology in the employees of investment banks.

The effect of neighbors for the behaviors has also been studied. The social links and connections influence the behaviors in various contexts. For instance, pineapple farmers in Ghana are shown to have a tendency to follow other farmers [14]. Christakis and Fowler [15] studied a large social network over 32 years and showed that if a person had a friend who became obese,

then that person's chances of becoming obese increased by 57%. Likewise, if the person had a sibling who became obese, then the increase is by 40% and by 37% if the person had a spouse who became obese. Banerjee et al. [16] studied the peer effects of the diffusion of microfinance participation in 75 villages in South India. They showed that the information passing parameters are significant in the social networks. According to their results, if a person participates, then it is more likely that the person will transmit the information to its neighbors. In a different setting that the network influences behavior, Glaeser et al. [17] showed evidences that the social interactions are important for determining who becomes a criminal in youth.

Collective action is another research topic for the influence of networks. Granovetter [18] suggested a model in which people make choices depending on their neighbors in social networks. In the Granovetter model, each agent responds to the previous stage's distribution of adopters. The shape of the fraction of adopters over time determines the equilibrium, namely tipping points, such that a slight change in the fraction of adopters can shift the population significantly. Chwe modeled a coordination game in which people tells their willingness to participate to their neighbors [19]. Chwe also modeled collective action model in which people reason about their neighbors before they get into action [20]. In the reasoning model, what the person knows about what their neighbors' know play significant role for the decision making. Acemoğlu et al. studied the opinion fluctuations and disagreement in the social networks [21]. In their study, the results suggest that when the agents decides collectively and collaboratively, there will be a persistent disagreement and that the network will never reach to a consensus for the opinions due to the presence of stubborn agents.

In this thesis, we slightly modify the diffusion of innovation models from the literature, and implement the diffusion models in complex network to analyze the effect of clustering, average path length and the hubs. To see the hub effect, we introduce a brand new network called the Robin Hood network using the poor get richer approach. In the Robin Hood network there are no hubs. We compare the poor get richer approach with the rich get richer approach (i.e. Scale Free network) in which the hubs are present. Moreover, we present a novel collective action model. The collective action model is also simulated in complex networks. The results

indicate that the presence of hubs is not a must for collective action, whereas the tightly-knit communities positively effect the size of collective subset. On the other hand, the better diffusion comes with lower clustering and smaller paths, in which the presence of hubs is significant.

In the following chapters, we will focus on diffusion and collective action in complex networks. In the second chapter, the diffusion with infinite speed will be analytically solved and simulated in one-dimensional networks. In the third chapter, the focus will be on the two dimensional networks and diffusion models. In the fourth chapter, more complex networks namely ring lattice, small world and random networks will be simulated for the diffusion models. Furthermore, the effect of clustering coefficient, average path length and the hubs will be analyzed. In the fifth chapter, we will define a model for collective action. The effects of the structure of networks in the collective action model will also be presented. In the sixth chapter, we will summarize the findings and remark the future research areas.

2 Diffusion of Innovation with Infinite Speed in One-Dimensional Networks

Imagine a one-dimensional (1D) network consisting of N nodes and $N - 1$ links between the nodes.

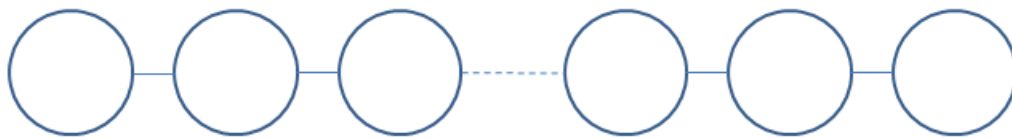


Figure 2.1: A one-dimensional network with N nodes and $N - 1$ links.

Let us define time as discrete time clicks: $t = 1$, $t = 2$, $t = 3$ and so on. At each time click an innovation diffuses throughout the network. The innovation can be anything, for example; adaptation of a new technology, an epidemic, or a new behavior. Let's assume that the innovation diffuses only one way, from the left node to the right node. Let us also assume that the innovation initiates from the left-most node. We can limit the number of nodes that the innovation diffuses at each time click. For instance, we can assume that the innovation diffuses to only one node at each time click. Then it will take approximately N time clicks for the diffusion to cover all nodes in the network. Likewise, if there are only two nodes that the innovation diffuses at each time click, then in $N/2$ time clicks all the nodes will be covered.

2.1 Active Link Diffusion with Infinite Speed

In the active link diffusion model, the nodes are static, however the links are dynamic. In other words, at each time click the links are active or not, depending on a random probability distribution. Let's assume that the probability of any link to be active is $p \sim (0, 1)$ and p is fixed for all the time clicks. Although p is constant at all times, the active links are not dependent on the previous time's active link distribution. That is to say, the links are active or not independent from their previous active status. There is no memory between the active link distributions at any time clicks. Note that at any time, there will be $(N - 1) * p$ active links on average.

We consider a diffusion process as follows: At any time, the innovation can diffuse via the active links with an infinite pace. In other words, the innovation diffuses to all the nodes that are attached actively (i.e. with active links) to a node that has already been diffused. If a node is diffused, we say that the node is adopted. Once a node is adopted, then it will stay adopted for the rest of the time. Initially all the nodes are not adopted. At time $t = 1$, we start the diffusion from the leftmost node in the network (i.e. the initial adopter). At time $t = 1$, all the nodes that are connected with the initial adopter via active links will be adopted. At time $t = 2$, the links active status will change, and the innovation will continue to diffuse through active links. The simulation stops when all the nodes are adopted. An example is presented in Figure 2.2.

Before moving to more complicated networks, let us answer a couple of questions here. The first question that comes to mind is that how many time clicks does it take for the diffusion to cover all the nodes. In order to answer this, we need to find the number of nodes that become adopted at any time click. The answer partially boils down to the average node cluster size connected with active links, since the diffusion has infinite speed over the network. In other words, if you put your finger to a node at any time click, what is the average size of the node cluster that is connected via active links?

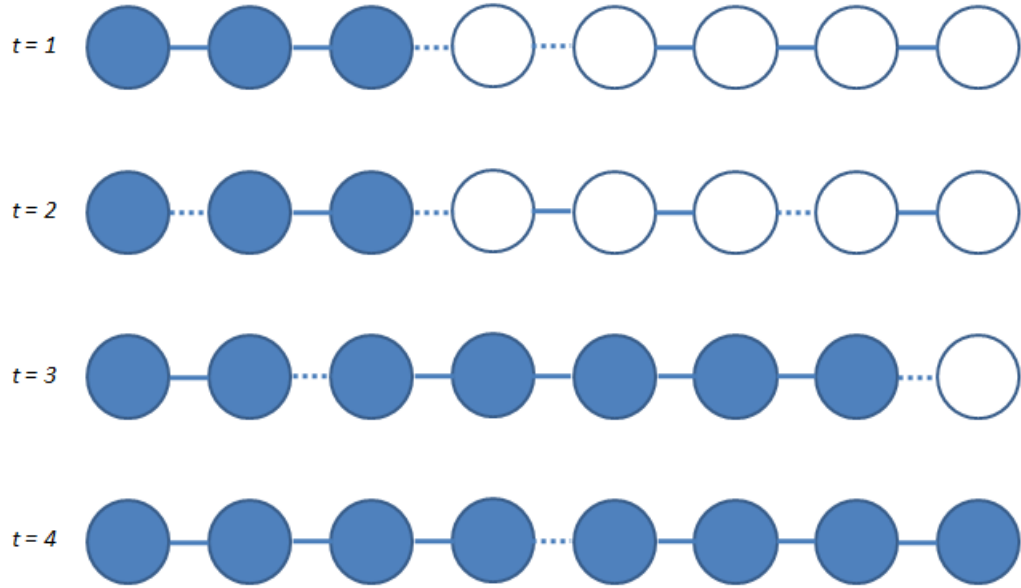


Figure 2.2: An example of link diffusion with infinite speed in a one-dimensional network with 8 nodes and 7 links. The active links are represented as solid lines and the adopted nodes are represented as filled circles. The active link probability p is 0.8. At time $t = 1$, the initial adopter (the leftmost node) diffuses to its neighbor node. Likewise, the neighbor node diffuses to its neighbor on the right. The diffusion continues until there is no active link. At time $t = 2$, there is no diffusion, since the link between 3rd and 4th node is not active. It takes 4 time clicks for the diffusion to cover all the nodes in the network.

The probability of a fixed cluster size, say a , is the following:

$$P(\text{Cluster Size} = a) = p^a * (1 - p)^2 \quad (1)$$

There are a active links and two inactive links (one in both sides) in a cluster of size a . However, there are a couple of points to clear. Firstly, in fact $a - 1$ active links can connect a cluster of size a . However, this difference becomes very small for the sake of the average cluster size calculations. Average cluster size is basically the sum over all possible a 's. Therefore, the difference between $a - 1$ and a can be negligible for determining the average cluster size. Secondly, on the edge of the network there can be only one inactive link (since the other side will be empty). But, if the network size is large enough, the probability of a cluster to be in the edge of a network is also negligible.

There are several intermediate steps before finding the average cluster size which can be found in Stauffer and Aharony [22]. After following these steps, the average cluster size can be found as:

$$\text{Average Cluster Size } (S) = \frac{1 + p}{1 - p} \quad (2)$$

Why is it important to know the average cluster size? Since, at any time click the innovation diffuses throughout the whole cluster. However, the average cluster size is not the answer that we pursue: How many nodes does the innovation diffuse at each time click? Some of the nodes in the cluster might have already been adopted. So, how can we know the number of nodes that are not adopted in the cluster?

Let's consider the example above in Figure 2.2. In the example, the diffusion continues from the rightmost node of the adopted nodes. The rightmost adopted node is the 3rd node in $t = 1$ and $t = 2$, 7th node in $t = 3$ and 8th node in $t = 4$. If we figure out the position of the rightmost adopted node in any cluster, then we can understand the average number of nodes that the innovation diffuses.

In fact, the rightmost adopted node can be in any position with the same probability. It can be at the very beginning, with probability $1/S$, where S is the cluster size. In that case, the innovation will diffuse to $S - 1$ nodes, excluding the already adopted. Using the same logic, the following formula can be obtained.

$$\text{Diffusion Rate} = \frac{1}{S} * (S - 1) + \frac{1}{S} * (S - 2) + \frac{1}{S} * (S - 3) + \dots + \frac{1}{S} * 0 \quad (3)$$

After a few simplifications, the number of adopted nodes becomes:

$$\text{Diffusion Rate} = \frac{P}{1 - p} \quad (4)$$

So, how long does it take for the diffusion to cover all the nodes in the network? After finding the average number of nodes that become adopted at any time, the answer is easy. The number

of nodes in total divided by the diffusion rate will be the time to cover all the network.

$$\text{Time to Cover All the Network} = \frac{N * (1 - p)}{p} \quad (5)$$

Above we provided exact analytical solution for the time to cover all the network. Now, we compare our analytical solution with the computational simulation results. The main purpose for the comparison is twofold: Firstly, we can be confident about the implementation of the active link diffusion algorithm to the computational environment. Secondly, we can be sure that our analytical solution is correct. In Table 2.1, the computational results and the analytical results are presented. Since the numbers are very close, we can conclude that both computational algorithm and our analytical solution are consistent. As N goes to infinity, both computational results and analytical solution will converge.

p	N	<i>Computational Time to Cover All</i>	<i>Analytical Time to Cover All</i>
0.8	100	25.17	25
0.8	200	50.83	50
0.5	100	99.47	100

Table 2.1: The computational results for the time to cover all the nodes in network and analytical solutions are compared. The computational results are averaged across 1000 simulation runs. Since the numbers are close enough, we can be confident about our analytical solution and the computational implementation of the proposed algorithm. As N goes to infinity, both analytical and computational results will converge.

2.2 Active Node Diffusion with Infinite Speed

It is interesting to think what happens if the nodes would be dynamically changing instead of the links. Therefore, the innovation diffuses through the active nodes, not the active links. Does the time to cover all change with these settings? The answer is no, there won't be any change as long as the time to cover all the network is concerned.

Let's consider the same network and the same diffusion process with only one change: Instead of the links, the nodes are dynamically active or not depending on a fixed probability. An

example is illustrated in the figure below.

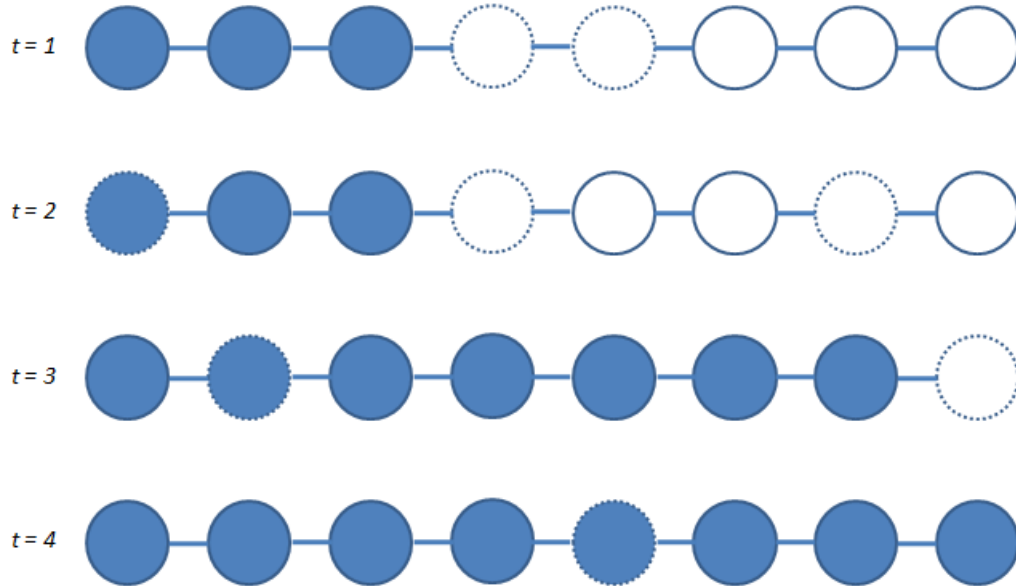


Figure 2.3: An example for the active node diffusion with infinite speed on a one-dimensional network with 8 nodes and 7 links. The active nodes are circled with solid lines and the adopted nodes are represented as filled circles. The active node probability p is 0.8. At time $t = 1$, the initial adopter (the leftmost node) diffuses to its active neighbor node. Likewise, the active neighbor node diffuses to its active neighbor on the right. The diffusion continues until there is no active neighbor node. At time $t = 2$, there is no diffusion, since the 4th node is not active. It takes 4 time clicks for the diffusion to cover all the nodes in the network.

We will use a different approach to find the exact solution for time to cover the entire network. This time the calculations will not focus on the average cluster size, instead we will focus on the average number of nodes that become adopted at each time click. At any time click there will be only one node that becomes adopted if there is an active node (neighbor to the rightmost node of the adopters) followed by a not active node. Likewise there will be only two nodes that become adopted, if there are two consecutive active nodes followed by a not active node. So, the average number of nodes that are diffused at any time click can be found

by the following formulation:

$$\begin{aligned}
\text{Diffusion Rate} &= p * (1 - p) * 1 + p^2 * (1 - p) * 2 + p^3 * (1 - p) * 3 + \dots \\
&= (1 - p) * \left(\sum_{i=0}^{\infty} p^i * i \right) \\
&= (1 - p) * \frac{p}{(1 - p)^2} \\
&= \frac{p}{(1 - p)}
\end{aligned} \tag{6}$$

The calculations suggest that there is no difference between the active link diffusion and the active node diffusion, in terms of the diffusion rate and the time to cover all the nodes in one-dimensional networks. In the following chapters, we will see that the diffusion rate and the time to cover the entire network in the node and link diffusion algorithms can vary depending on the network topology.

2.3 A Different Application: Infection and Vaccination

Now, we implement another application to our model. In the previous model, once the node is adopted, they stay as adopted in the rest of the time. We relax this assumption. In this model, the nodes can go back to the stage of not adopted, after they are adopted.

In the previous model the active link probabilities was given and fixed throughout the iterations. But, what if the nodes have control over the link probability and can alter the probability dynamically? It is interesting to think about what would change if the network adopts itself as the diffusion proceeds in the network.

In the most common epidemic models (e.g. SIR-SIS models [23] [24]), the nodes are either in one of the following stages, S when susceptible, I when infected and R when recovered. In recent years, the study for the epidemic models also included the news about the epidemics and the adaptation of the networks in order to prevent from the infectious diseases [25]. When the news about an epidemic reaches out to the whole network, the nodes start to adapt to the

new situation in order to reduce the risk of being infected. After the news, the nodes decrease the frequency of contacting with their neighbors. We implemented this idea to our diffusion model. We modeled an adaptation behavior to our network.

Let's consider a new link diffusion model in one-dimensional networks. In the new model, the innovation can be thought like an epidemic. Let's assume that when the epidemic diffuses to half of the population, the news about the epidemic spreads. For simplicity, we assume that the news spread immediately to all the nodes in the network. When the news reaches out, the nodes that are not infected decide to contact with their neighbors less often, so their active link probabilities decrease. Also, at the same time the vaccination starts to cure the infected nodes, starting from the very beginning (i.e. the leftmost node). The vaccination also diffuses as the epidemic but since the infected nodes do not change their active link probabilities, the vaccination diffuses faster than the epidemic. An example can be seen in the Figure 2.4.

The research question that we pursue in this model is the following: Is there a critical active link probability p' so that the vaccination catches the epidemic before it reaches to the last node in the network? In other words, what is the maximum contact frequency for the not infected nodes so that the epidemic can be stopped before it reaches to the end of the network?

We will use the analytical solution that we found previously. After news, there are two time to cover all formulas. First is for the epidemic, the second is for the vaccination.

$$\text{Epidemic Time to Cover All} = \frac{N * (1 - p')}{2 * p'} \quad (7)$$

$$\text{Vaccination Time to Cover All} = \frac{N * (1 - p)}{p} \quad (8)$$

Since we need vaccination should be faster than the epidemic, vaccination time should be smaller than the epidemic time. Hence, if we do the necessary calculation steps, we get the following formula.

$$p' < \frac{p}{2 - p} \quad (9)$$

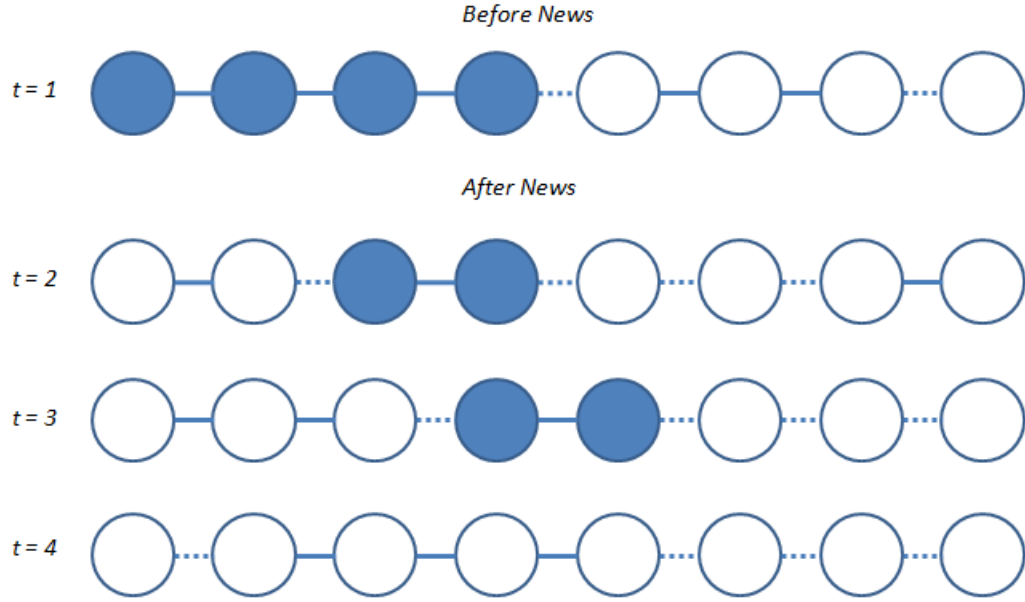


Figure 2.4: An example for the active link diffusion on a one-dimensional network with 8 nodes and 7 links. The active links are represented as solid line and the adopted nodes are represented as filled circles. Before news, the active link probability p is 0.8. After news, the active link probability p' is 0.2. At time $t = 1$, the initial adopter (the leftmost node) diffuses to its neighbor node. Since the epidemic diffuses to half of the nodes, the news about the epidemic decreases the active link probability for the rest of the nodes. At the same time, the cure for the infected nodes starts to spread. At time $t = 2$, there is no diffusion for the epidemic, but two of the nodes recovered from the infection. It takes 3 time clicks for the vaccination to cure all the infected nodes.

It means that after the news spread, not infected nodes should lower their contact frequency with their neighbors at most to the calculated number above. We simulated this diffusion model and the computational results are presented Table 2.2.

In Table 2.2, the percentage of the successful catch shows the percentage of simulations in which the vaccination reaches to the last node before the epidemic reaches. p' values are taken by equating the formula in 9. With these parameters of the active link probabilities, both epidemic and vaccination reach to the last node at the same time. As the network size N goes to infinity, the percentage should converge to 50% . Since on average both epidemic and vaccination have the same time to cover all, in half of the simulations epidemic would be the first to reach the last node, and in the other half the vaccination would be the first to reach the

end of the network. Therefore, the results are consistent with our model.

p	p'	N	<i>Percentage of the Successful Catch</i>
0.8	0.66	100	65 %
0.8	0.66	200	51 %
0.8	0.66	300	48 %
0.5	0.33	100	53 %

Table 2.2: The computational simulation results for the vaccination to catch the epidemic before it reaches to the last (the rightmost) node in a one-dimensional network with N nodes and $N - 1$ links. p is the before news active link probability, and p' is the after news active link probability. The results are the averages of 1000 simulations. The percentage of the successful catch will converge to 50% as N goes to infinity. This is consistent with our parameters, since with these settings of link probabilities both vaccination and epidemic reaches to the end of the network with the same time. In the half of the simulations, epidemic will reach to the end faster than vaccination, and in the other half, vaccination will reach to the end faster than epidemic.

2.4 Future Models

In the future, we will use asymmetric time clicks and active link change. In the previous models, we assumed that as time changes, the active link distributions also change. In other words, the active link distributions and the time are synchronized. In the future models, we will implement a different approach in which the time and links are not in harmony. The research question for this model can be summarized as the following: Which of the time click distribution and link distribution couple can minimize the time to cover all and maximize the diffusion rate?

3 Diffusion of Innovation with Infinite Speed in Two-Dimensional Networks

In this chapter, we will focus on the two dimensional regular lattice networks in which every node has 4 neighbors except the nodes on the sides. Firstly, we will implement the active link diffusion with infinite speed method to this network. Then, we will simulate the method and show the experimental results for the time to reach to the target node.

Consider a two-dimensional network that is explained above. In this network, let's consider the active link diffusion model with infinite speed that was discussed in the previous chapter. Assume that the starting point (i.e. the initial adopter) is fixed and it is the upper left node. Let us also assume that the aim is to reach to the bottom right node (i.e. the target node). In the first step model, we focus on the time to reach to the target node, instead of time to cover the entire network. Therefore, we use the term move, instead of diffuse. In our model, the movements are either to the node on the right or the node on the bottom. The active link probability p is given and constant throughout the time clicks. The question we pursue is that how many time clicks are needed to reach to the target node. Be aware that, instead of time to cover the entire network, here we only want to reach to the bottom right node.

There is another issue that we need to discuss. The answer for the time to reach to the target node depends on the knowledge of the active link paths. For instance, if a node can see all the active links in the network, then the node would follow the path which gets the node

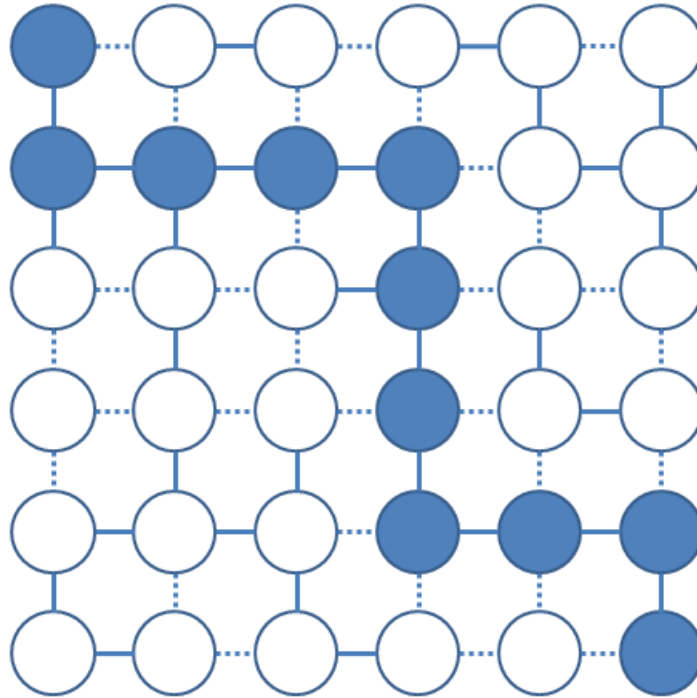


Figure 3.1: An illustrative example of the diffusion model in two-dimensional network with 36 nodes. The active link probability is 0.5 and there are 30 active links at each time click on average. The active links are represented as solid lines. There is a path with active links from the upper left node (initial adopter) to lower right node (target node). If there is perfect information about the active links, it can take only one time click to reach the target node

closer to the target node. On the other hand, if the node can only see the active links with the nearest neighbors, then the node would choose one of the active links randomly and hope to end up very close to the target node. Imagine that you are in a maze and try to go to the exit. If you know which way to follow, then you would exit very fast. However, if your sight is limited and can only see one road at a time and decide which way to go after you finish the road, then the time to exit would be very large compared to the previous example. Figures 3.1 and 3.2 are illustrative examples to this discussion.

As the two figures suggest, the time to reach to the target node depends on the limitation of the information about the active links. Unless there is perfect information about the active link distribution, it usually takes more than one time click.

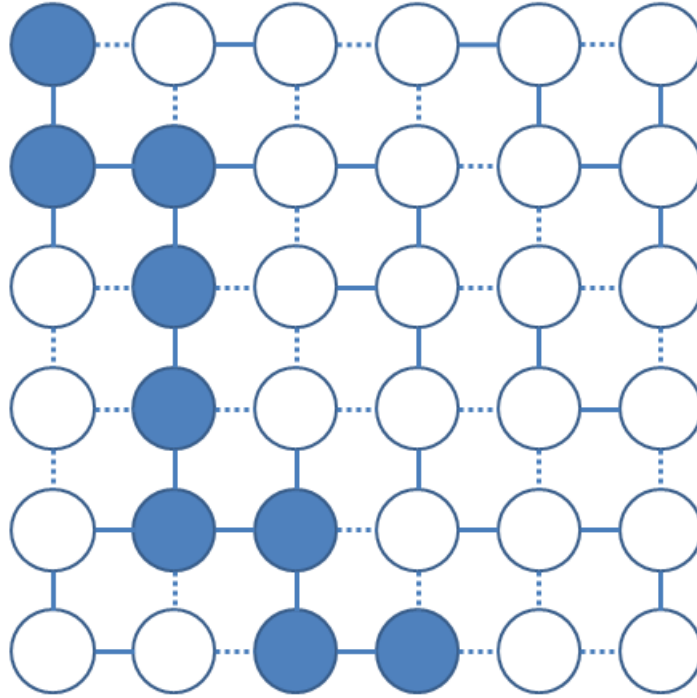


Figure 3.2: An illustrative example of the diffusion model in two-dimensional network with 36 nodes. The active link probability is 0.5 and there are 30 active links at each time click on average. The active links are represented as solid lines. There is a path with active links from the upper left node (initial adopter) to lower right node (target node). If there is limited information about the active links, it can take more than one time clicks to reach the target node.

3.1 Active Link Diffusion with Limited Information

In the first step, we will assume that only the active links with the nearest neighbors are visible. Therefore, if there are two active links that connects the node to its neighbors, one of them will be selected arbitrarily. As indicated before, the only movements are to the node on the right or to the node at the bottom. When the links with these two neighbors are not active, then there is no way to move. In that case, the time click is increased by one and the active links are redistributed. In the next time click, the node who is the closest to the target node chooses an active link to move closer to the target node. An example is shown in Figure 3.3.

Be aware that in this model, there are intermediate steps within the time clicks. At each intermediate step, there is only one active link to be chosen. However, the time is constant

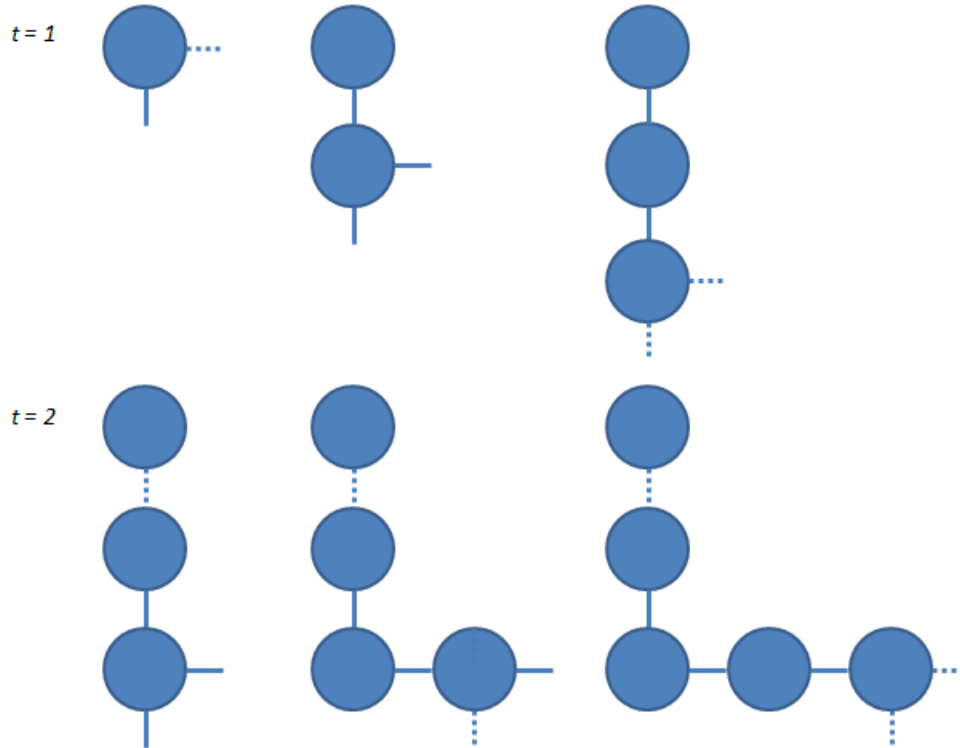


Figure 3.3: An example of the active link diffusion model in two-dimensional networks with limited information. At $t = 1$, in the first step the diffusion has only one way to go (to the bottom node). In the second step, there are two ways and again the bottom node is chosen. However, in the next step there is no way to go. After the redistribution of the active links at $t = 2$, the node who is closest to the target node (i.e. the bottom node at the beginning of $t = 2$) has two options and the right node is chosen arbitrarily. In the second step, there is no option and the right node is chosen again. In the third step, there is no active link to move. The same process goes on and on until the target node is reached

when the intermediate steps are executed. After all the intermediate steps are executed and there is no active links to move, then the time click is increased by one and the active links are redistributed. Also, this model is not exactly diffusion. The only goal here is that to reach to the target node. Therefore, only the node that is the closest to the target node moves forward.

Now that we explained the model and clarified several points, let's start finding the answer for the time to reach to the target node. We simulated the algorithm in the computational environment, and the results are presented in Table 3.1.

It is interesting to see that as the network size grows exponentially, the time to reach the target

p	N	<i>Time to Reach the Target Node</i>
0.8	625	5.16
0.8	2500	7.74
0.8	5625	10.31
0.8	10000	12.41

Table 3.1: The computational results for the time to reach to the target node in two dimensional regular lattice networks. The results are the averages of 1000 simulations at each settings.

node increases linearly. However, this is not surprising. Since this is not time to cover all the network. In order to reach the target node, there are exactly $2 * \sqrt{\text{Network Size } (N)}$ links to move. Hence, the table makes sense in terms of time and link counts.

In the following chapter, we will focus on the diffusion process on the two dimensional networks.

4 The Diffusion of Innovation in Complex Networks

In this chapter, we will analyze the diffusion of innovation in complex networks. In the previous chapters, we presented the simulation results and analytical solutions for the time to cover all the network and the time to reach to the target node in one dimensional and two dimensional networks. Now, we focus on more complex networks like small world, random and scale free networks. At the same time, another diffusion algorithm namely neighbor diffusion will also be discussed and simulated in these networks. We will provide insights about how different types of networks react on different diffusion algorithms. Also, we will give several explanations about the reasons behind.

Suppose that we are given an undirected network (graph) G , with adjacency matrix A , node set U , link weights W , and behaviors B . In this network, an innovation diffuses and the nodes dynamically decide whether to adopt the innovation. The innovation can be an adaptation of a new behavior, the decision to buy a new model of a smart phone, or voting for a party that you have never voted before, etc. In this chapter, the innovation is assumed to be adoption of a new behavior. We define an adoption mechanism as follows: If a particular node u has adopted to behavior b then $X_u^b = 1$, if not adopted then $X_u^b = 0$. However, not every node is willing to adopt. Each node has a potential (willingness) to adopt the behavior. If a node u has potential do adopt then $X_u^{p,b} = 1$, if the node has no potential then $X_u^{p,b} = 0$.

Now, let us assume that an initial adoption occurred in this network. In the previous chapters, the initial adopter was assumed to be a fixed node. In this model we relax this assumption. The initial adopter can be chosen randomly or using a specific selection process (e.g. probability distribution, among some certain candidates, etc.). Also, the initial adoption can consist of a certain number of nodes, namely I . Let us define the time as discrete time ticks $t = 1, t = 2, t = 3$, and so on. The initial adoption occurs at the beginning of time $t = 1$. At time $t = 1$, the initial adopters start to spread the news about the behavior via their links. The link neighbors of the initial adopters choose to adopt the behavior depending on the diffusion algorithm. The decision to adopt the behavior can be dependent on the node potentials that we defined earlier. Otherwise, we can use a model that the adoption depends on the neighbors' potentials. At each time tick, the diffusion can be limited to a certain number of nodes. Like in the previous chapters, the diffusion can have infinite speed; it can diffuse to every node available in the blink of an eye. These definitions are for the general model that we will use in this chapter. Along the chapter, we will use several assumptions and predefined variables that we will briefly explain.

As time goes to infinity, there can be two resulting equilibria. The first equilibrium is a fully adopted network in which every node has adopted the behavior. In this case, the fraction of adopters in the network will be equal to 1. The second equilibrium is an intermediate stage where there are both adopters and non adopters concurrently. In the second equilibrium, there might be potential nodes who have not adopted. This is mainly because of the nature of the diffusion algorithm. No matter how many time ticks pass, the news about the behavior will never reach to some potential nodes. In some networks although there is a huge potential, the fraction of the adopters remains relatively low. But in some other networks, most of the potential nodes adopts in the equilibrium.

There are a number of research goals that we aim to pursue in this model. The main questions can be summarized as follows. (a) How do the different diffusion models affect the fraction of adopters? (b) How do different network topologies affect the fraction of adopters? (c) Is there a threshold for the fraction of potential nodes above which the behavior becomes widespread?

(d) Do hubs (nodes with high number of connections) always play an important role for the diffusion? (e) Which one is more efficient: Few hubs with more connections or more hubs with fewer connections? (f) Does clustering coefficient (i.e. the percentage of the neighbors who are also neighbors with their selves) affect the diffusion? (g) Which one is more effective for diffusion: High clustered networks (i.e. tightly-knit networks) or low clustered networks? (h) On top of that, is there a middle ground for the clustering coefficient that maximizes the diffusion?

We will use three diffusion models, namely node diffusion, link diffusion and neighbor diffusion. For each model, we set the network parameters as follows.

$$Network\ Topology = \begin{cases} 2D\ Regular\ Lattice \\ Ring\ Lattice \\ Small\ World \\ Random \end{cases} \quad (10)$$

$$Number\ of\ Nodes\ (U) = 10,000 \quad (11)$$

$$Number\ of\ Links = 20,000 \quad (12)$$

$$Link\ Weights = 1 \quad (13)$$

$$X_u^b = \begin{cases} 1, & \text{if node } u \text{ is adopted the behavior } b \\ 0, & \text{if node } u \text{ is not adopted the behavior } b \end{cases} \quad (14)$$

$$X_u^{P,b} = \begin{cases} 1, & \text{if node } u \text{ has potential to adopt the behavior } b \\ 0, & \text{if node } u \text{ has no potential to adopt the behavior } b \end{cases} \quad (15)$$

$$Number\ of\ Behaviors\ (B) = 1 \quad (16)$$

$$Number\ of\ Initial\ Adopters\ (I) = 1 \quad (17)$$

4.1 Diffusion Models

4.1.1 Node Diffusion

The first model is the node diffusion model which is similar to the active node diffusion model that was explained in the previous chapter. This model is also very similar to Campbell's word-of-mouth model [26] (see also [27] for the implementation of Campbell's model). In the active node model, the active nodes are chosen with respect to a probability p . In this model, the nodes have an interest to a behavior. Some nodes might be more interested to adopt the behavior, whereas some other nodes might not be interested at all. The key thing is that the interest depends on the behavior. For instance, you might want to buy a new model of a smart phone only if the price is affordable. Therefore, the price plays an important role for the node's interest. The interest is modeled as the potential to adopt as explained previously. However, depending on the behavior (e.g. price of a smart phone) the fraction of potential nodes might change.

In this model, the fraction of potential nodes is fixed and the distribution of the potential nodes follows a uniform distribution $U(0, 1)$. If a node is potential, then it stays potential for the rest of the simulation. In this model, we assume that the diffusion spreads through only one link at a time click. An initial adopter is chosen arbitrarily to start the diffusion. At time $t = 1$, the link-neighbors of the initial adopter decide whether to adopt the behavior or not. Only the adopted nodes spread the news. At time $t = 2$, the adopted neighbors of the initial adopter spread the news. Then their neighbors decide to adopt, and so on.

The decision process is as follows: If a node has potential and the news about the behavior reaches to the node, then it adopts. The key here is that in order to adopt there should be at least one adopted neighbor so that the news about the behavior reaches to the node. The conditional

probability of a node adopting the new behavior is shown in the formula below.

$$Prob(X_u^b = 1 \mid X_y^b = 1 : \exists y \in N(u)) = \begin{cases} 1, & \text{if } X_u^{p,b} = 1 \\ 0, & \text{if } X_u^{p,b} = 0 \end{cases} \quad \forall u \in U \quad (18)$$

In the formula 18, $N(u)$ represents the set of neighbors of node u .

In this model, not every node with potential adopts the behavior. In order to adopt the behavior, the potential node needs to hear about the behavior from a neighbor that has already adopted. The condition for the above formula is that there should be at least one adopted neighbor. Hence, when the network reaches to equilibrium, there might be some nodes with potential who have not adopted. A basic example can be found in Figure 4.1 for the illustrative purposes.

Now, let's analyze how the network topology affects the diffusion. Imagine two networks with N nodes and $2N$ links. In these networks not only the number of nodes and links are the same, but also the degree distribution is the same. Every node in these networks has four neighbors. If you take a node from one network and replace it to the other network, the node wouldn't notice the difference. The networks are namely two-dimensional regular lattice and ring lattice. Figure 4.2 illustrate the 2D lattice and ring lattice.

We simulated the node diffusion model in 2D regular lattice and ring lattice using Netlogo [28]. We keep track of the fraction of adopters with respect to the fraction of potential nodes in the equilibrium. The results can be seen in Figure 4.3.

Figure 4.3 shows that the fraction of adopters depends on the network topology, in other word how connected the network is. Even though the number of nodes, the number of links and the degree distributions are the same the fraction of adopters can be very different. The adjacency matrix of the network is very critical for the diffusion.

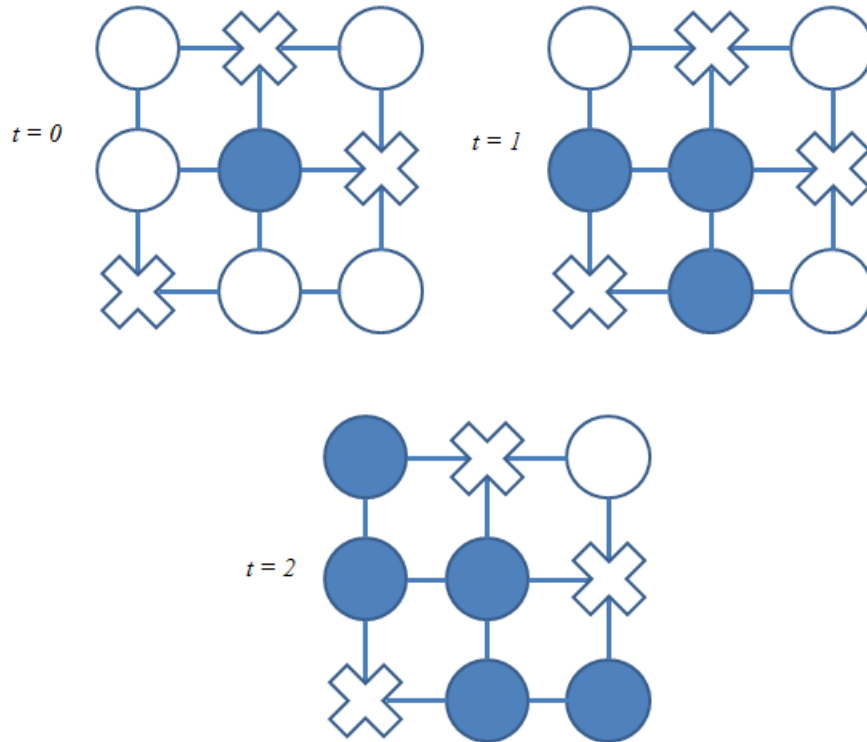


Figure 4.1: An example for the node diffusion on two dimensional network. The circles represent that the node has susceptibility higher than the price. The X's represent that the node has potential. At time $t = 0$, the filled node in the middle is the initial adopter. At time $t = 1$, only the potential neighbors of the initial adopter can adopt. Likewise, at time $t = 2$ there are two more potential neighbors to adopt. The key thing here is that those who adopted at time $t = 2$ could not have adopted at $t = 1$, since their neighbors had not adopted yet. The network reaches to equilibrium after $t = 2$. Although there is one more node who has potential, it can not adopt since its neighbors will never adopt, no matter how many time clicks pass.

Some potential nodes cannot adopt because they are not aware of the innovation. Although there is potential in the network, the behavior does not become widespread for some settings. The fraction of potential nodes plays an important role for the diffusion to spread widely in the networks. There arises a critical threshold for the fraction of potential nodes above which the behavior becomes widespread. In Figure 4.3, the critical threshold for the 2D lattice is 0.60. Above that threshold, the fraction of adopters are the same as the fraction of potential nodes. It shows that whenever there is a potential node, the news have reached to them. For the ring lattice, the critical threshold is 0.95. Below the threshold, most of the potential nodes do not become adopted in the equilibrium.

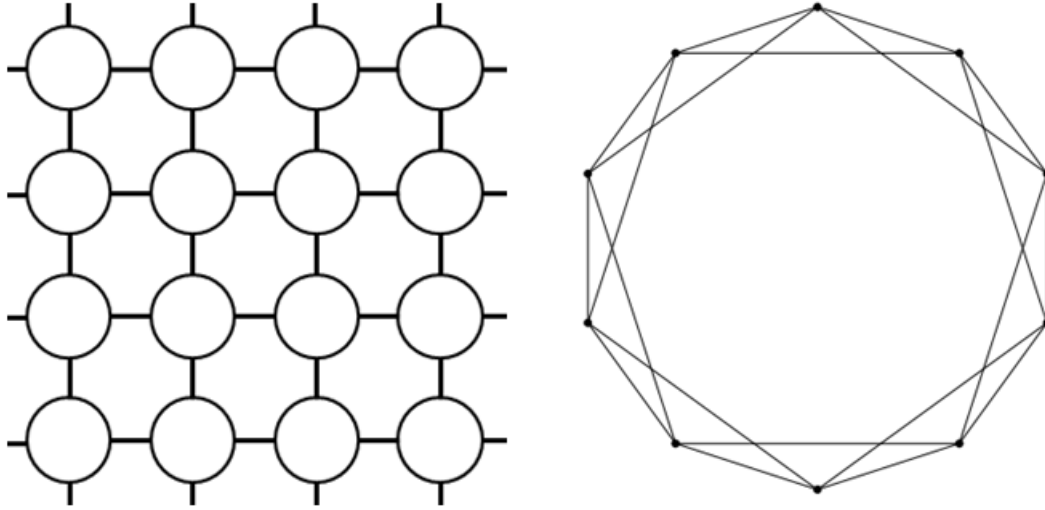


Figure 4.2: 2D regular lattice and ring lattice with N nodes and $2N$ links.

In the future chapters, we will dig into the reasons that lead to this difference in the fraction of adopters. We will analyze the effect of clustering coefficient and average path length. We will use the rewiring probability to create small world networks and random networks. Thus, we will be able to analyze the effect of clustering coefficient and average path length in the diffusion models.

4.1.2 Link Diffusion

The link diffusion is almost the same as the previous active link diffusion. The links are active or not depending on a probability distribution. The behavior diffuses via the active links. The percentage of the active links is predefined. The active links are fixed and remain active throughout the time clicks. Thus, unlike the previous active link diffusion, there is no redistribution of the active links. Starting from a random initial adopter, if there is an active link that ties the initial adopter with other nodes, then these nodes also become adopted. Now, let's simulate the link diffusion model in 2D regular lattice and ring lattice networks. The results for the fraction of adopters with respect to the fraction of active links are illustrated in Figure 4.4. Like in the node diffusion model, link diffusion also suggests that the network topology plays role for determining the fraction of adopters.

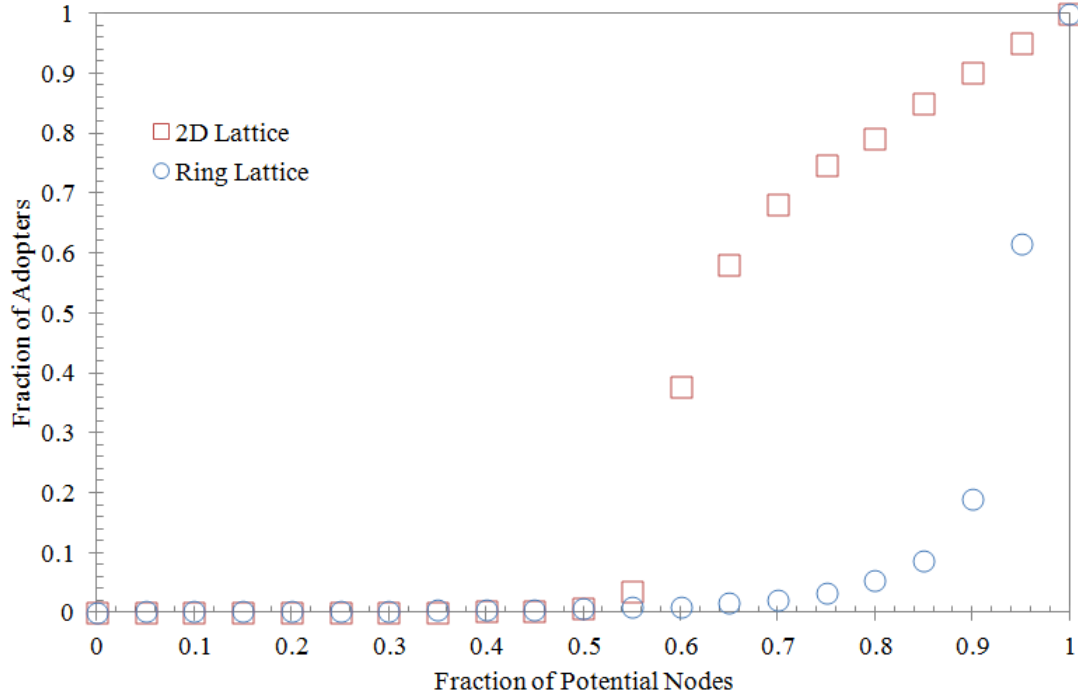


Figure 4.3: The fraction of adopters with respect to the fraction of potential nodes using the node diffusion in 2D and ring lattice networks. The parameters of the networks are explained in the formulas from 10 to 17. The results are averages over a thousand simulations. The networks are connected. Although the number of nodes, the number of links and the degree distributions are the same, the fraction of adopters can be very different. The fraction of adopters depends on the network topology in node diffusion.

4.1.3 Neighbor Diffusion

The neighbor diffusion model takes its root from the 19th chapter of Easley and Kleinberg’s textbook [1] in which they present the best response model. However, it is different from their model, and the modifications are made because our diffusion models have only one initial adopter. From one initial adopter, it is almost impossible for the innovation to reach the other nodes. The algorithm is as follows: Once the news about the behavior reaches to a node, the node decides to adopt depending on its neighbors. In addition to the previous settings in the formulas from 10 to 17, we also assume that every node can observe its neighbors’ potential. Therefore, the decision to adopt the behavior is completely based on the neighbors. If the proportion of the neighbors with potential is higher than a certain threshold, then the node

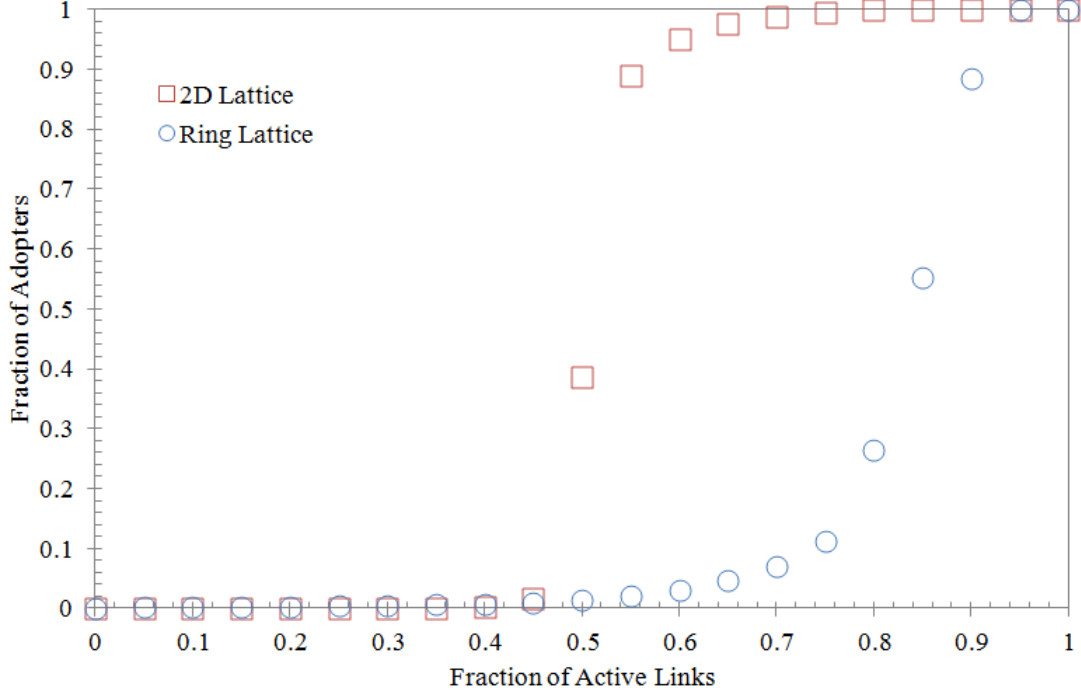


Figure 4.4: The fraction of adopters with respect to the fraction of active links in link diffusion in 2D regular lattice and ring lattice. The parameters of the networks are explained in the formulas from 10 to 17. Additionally, the fraction of potential nodes is 1 throughout the simulations. The results are averages over a thousand simulations. Although the number of nodes, the number of links and the degree distributions are the same, the fraction of adopters can be very different. The fraction of adopters depends on the network topology in link diffusion.

adopts. The conditional probability to adopt the behavior becomes the following.

$$\text{Prob}(X_u^b = 1 \mid X_y^b = 1 : \exists y \in N(u)) = \begin{cases} 1, & \text{if } \sum_{y \in N(u)} X_y^{p,b} / \sum_{y \in N(u)} w_{u,y} \geq \text{threshold} \\ 0, & \text{if } \sum_{y \in N(u)} X_y^{p,b} / \sum_{y \in N(u)} w_{u,y} < \text{threshold} \end{cases} \quad (19)$$

In the equation 19 $w_{u,y}$ represents the weight of links (which is by default 1) between node u and node y .

Above formulation means that if the number of potential neighbors divided by the number of all neighbors is higher than a predefined threshold, then the node adopts the behavior. Be aware that the adoption is nothing to do with the node's own potential. There can be a node with no potential but adopted as a result of its neighbors' potential. The simulation results of

the neighbor diffusion can be seen in Figure 4.5.

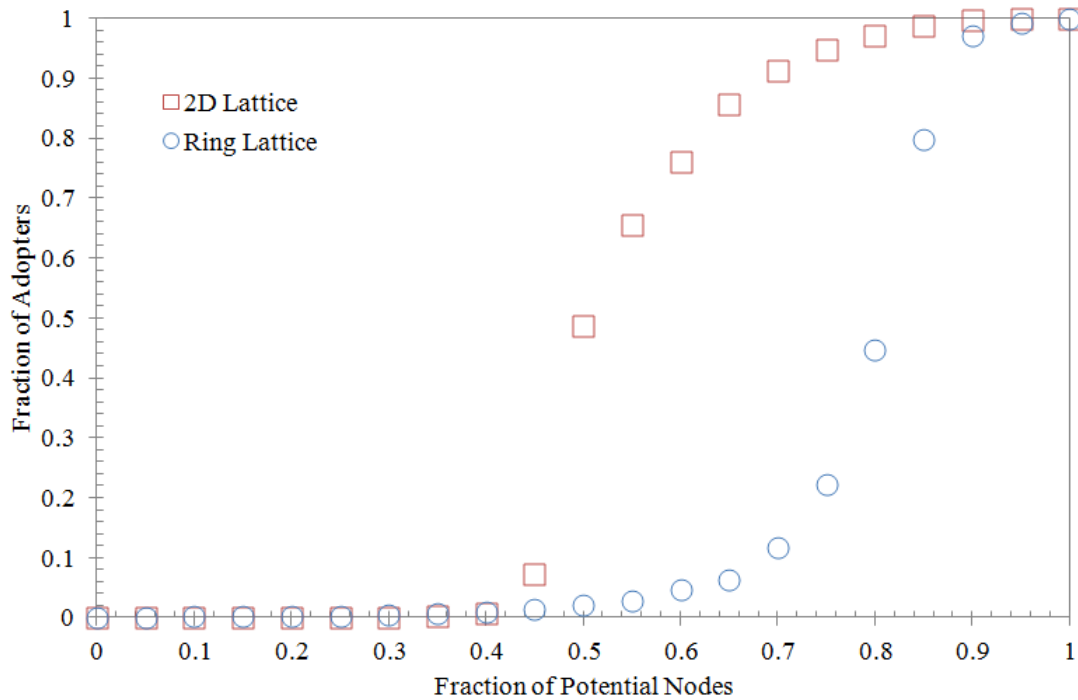


Figure 4.5: The fraction of adopters with respect to the fraction of potential nodes in neighbor diffusion in 2D regular lattice and ring lattice. The parameters of the networks are explained in the formulas from 10 to 17. The results are averages over a thousand simulations. The networks are connected. Although the number of nodes, the number of links and the degree distributions are the same, the fraction of adopters can be very different. The fraction of adopters depends on the network topology in neighbor diffusion.

According to the simulation results that can be viewed in Figure 4.5, the network topology plays role for determining the fraction of adopters in neighbor diffusion.

4.2 The Effect of Clustering Coefficient and Average Path Length in Diffusion

In the previous sections of this chapter, we showed that the topology of the network is very crucial for the diffusion. Even though the number of nodes, links and the degree distribution are the same, how connected the links are is very important for the spread of the behavior. In this section, we will look more closely to the network topology and in particular we will

analyze the effect of clustering coefficient and average path length using Watts-Strogatz model [29]. A node's clustering coefficient is basically the percentage of its neighbors who are also neighbors with each other. For instance, if a node has three neighbors and only one couple of the neighbors are linked, then the node's clustering coefficient is $1/3$. The average path length is another measure for the networks. It is basically the average of the number of steps in the shortest paths of all the nodes in a network.

Firstly, we will use ring lattice to create small world networks by rewiring the links. The ring lattice with all the nodes have four neighbors, has the clustering coefficient of 0.5. The average path length depends on the size of the network. However, these two properties can be altered by rewiring the links. By rewiring the links randomly with fixed probabilities, the clustering coefficient and the average path length decreases. As the rewiring link probability reaches to 1, the network becomes a random network, similar to the Erdos-Renyi random graphs [30]. The decrease is in the logarithmic scale and can be seen in the Figure 4.6.

Figure 4.6 shows that the small world, where the rewiring probability is 0.01, and ring lattice have almost the same clustering coefficient, but the average path length is significantly lower in small worlds compared to ring lattice. On the other hand, the small world and random network have almost the same average path length, but the clustering coefficient of the small world is significantly higher than the random network. These results suggest that in order to understand the effect of the average path length, we need to compare the simulation results for small world and ring lattice. Likewise, in order to analyze the effect of the clustering coefficient, the comparison of the small world and the random network is needed. Now, let us simulate the diffusion algorithms in these networks.

The results in Figure 4.9, Figure 4.11, and Figure 4.13 can be summarized as follows: The network topology has an impact over the diffusion and the experiments suggest that the low level of clustering coefficient and the less average path length is better for the diffusion. Although in some diffusion algorithms, the difference can be huge for decreasing the clustering coefficient (such as in the node diffusion), in other diffusion for instance in neighbor diffusion, lowering the clustering coefficient has a relatively little impact over the fraction of adopters.

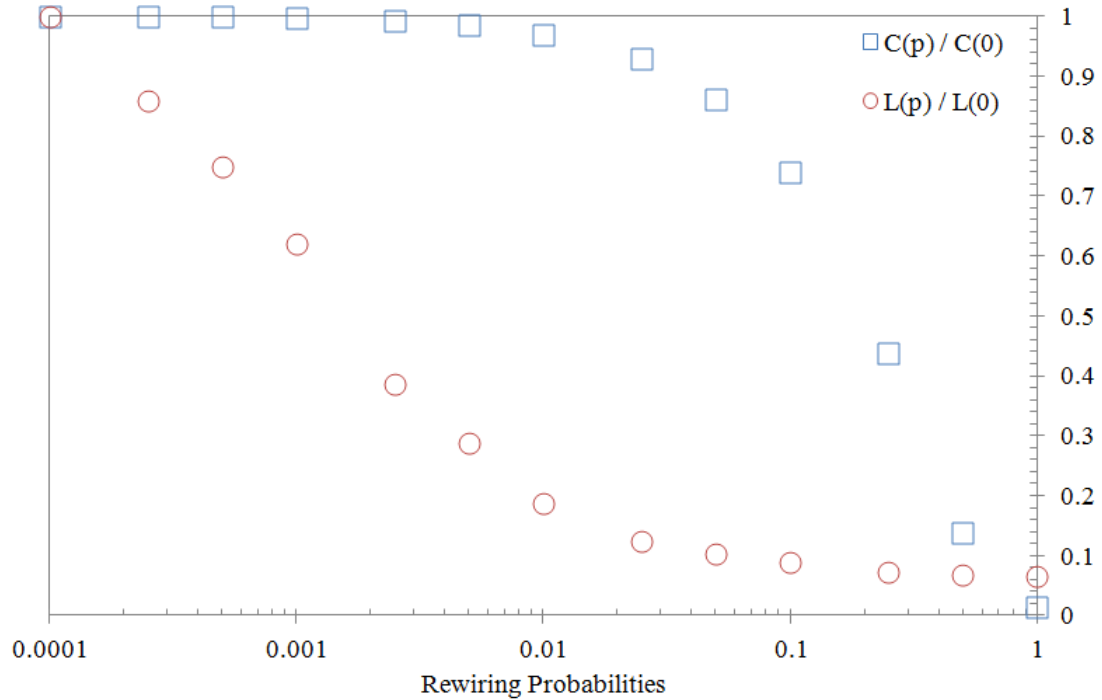


Figure 4.6: The change in the clustering coefficient and average path length with respect to the rewiring probabilities. C represents the clustering coefficient and L represents the average path length and p represents the rewiring probabilities. The change in the rewiring probabilities has different consequences in the clustering coefficient and average path length. Rewiring probability being zero represents the ring lattice, and rewiring probability being 1 represents the random network. When the rewiring probability is equal to 0.01, the average path length is as low as the random network but the clustering coefficient is as high as the ring lattice. This network is called the small world.

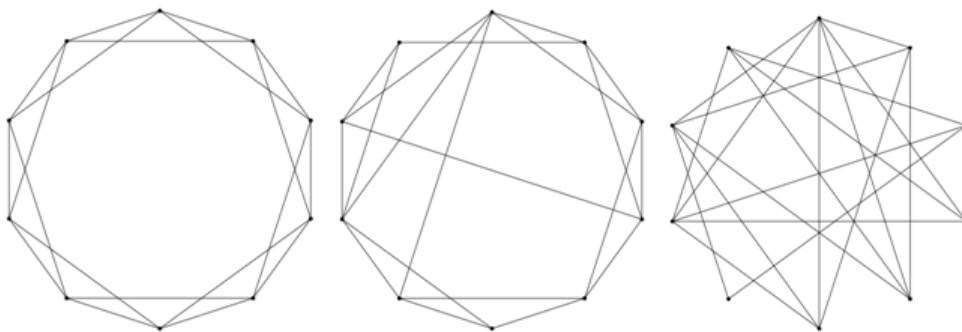


Figure 4.7: Ring lattice with rewiring probability = 0 (left), the small world with rewiring probability = 0.01 (center) and random network with rewiring probability = 1 (right).

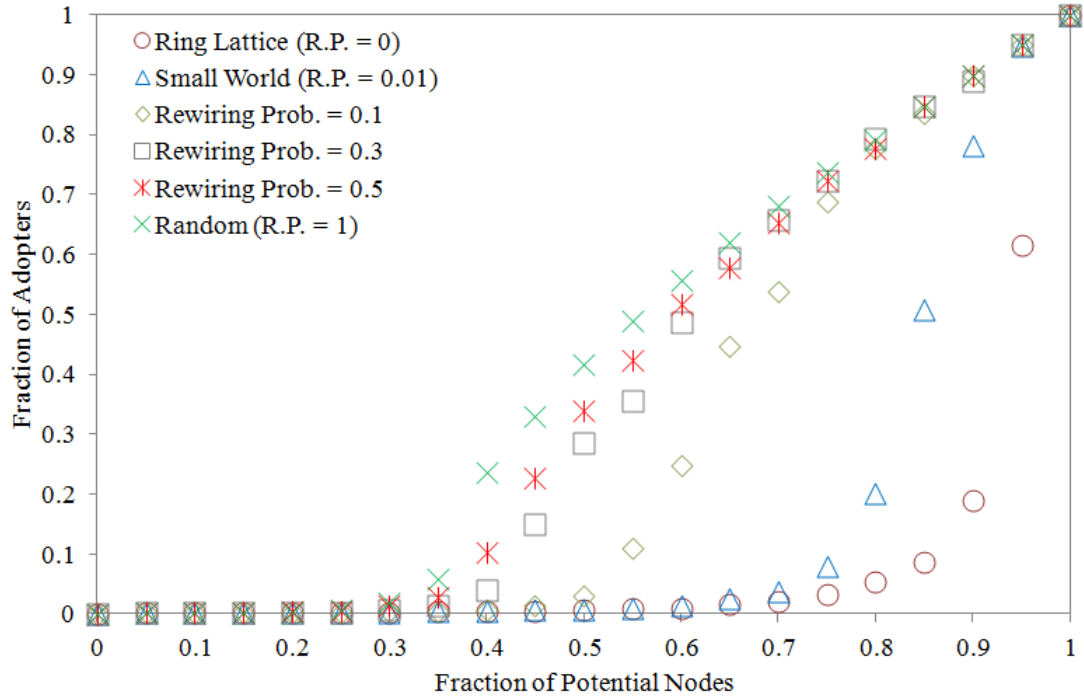


Figure 4.8: The effect of rewiring probabilities in node diffusion.

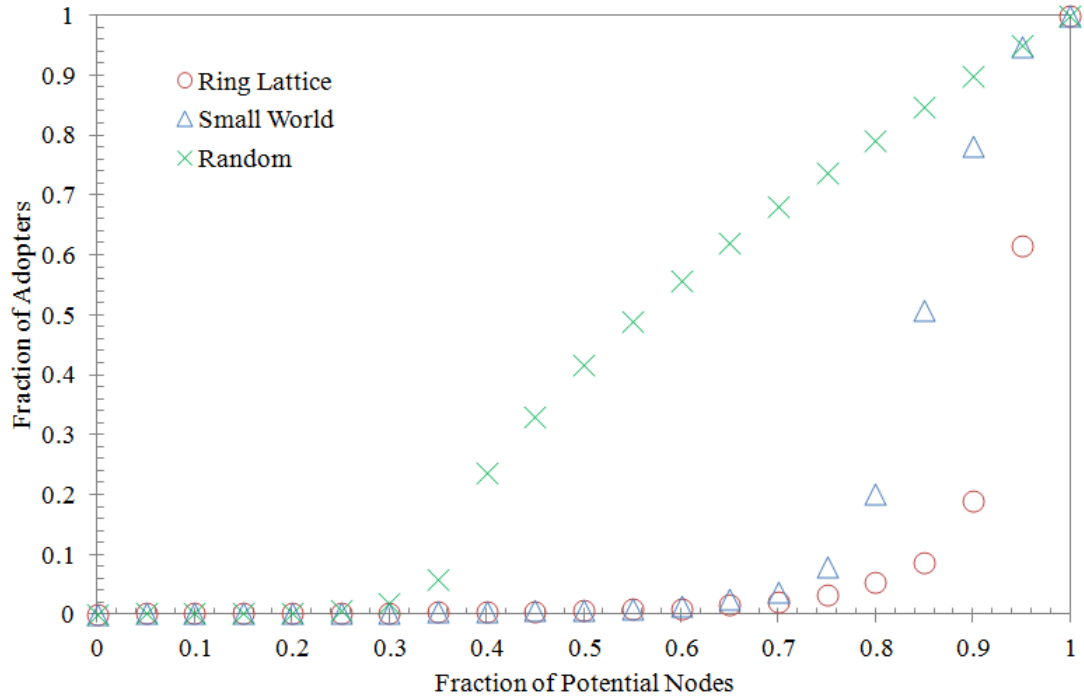


Figure 4.9: The effect of clustering coefficient and the average path length in node diffusion. The node diffusion is higher when the paths are smaller. Additionally, the node diffusion is higher when the clustering is lower.

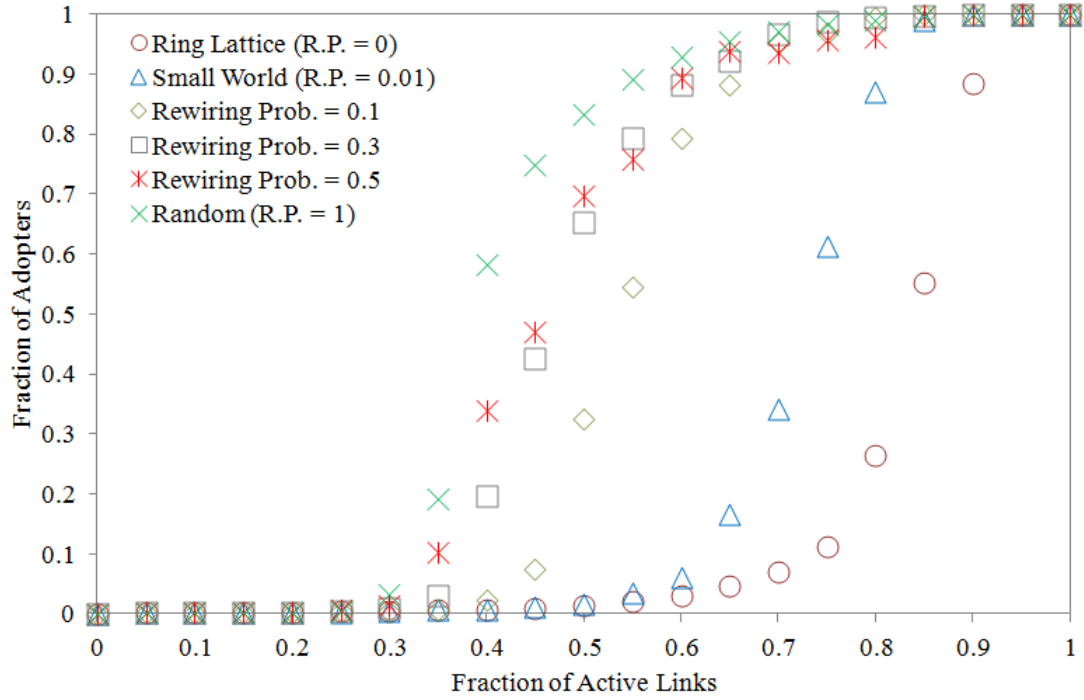


Figure 4.10: The effect of rewiring probabilities in link diffusion.

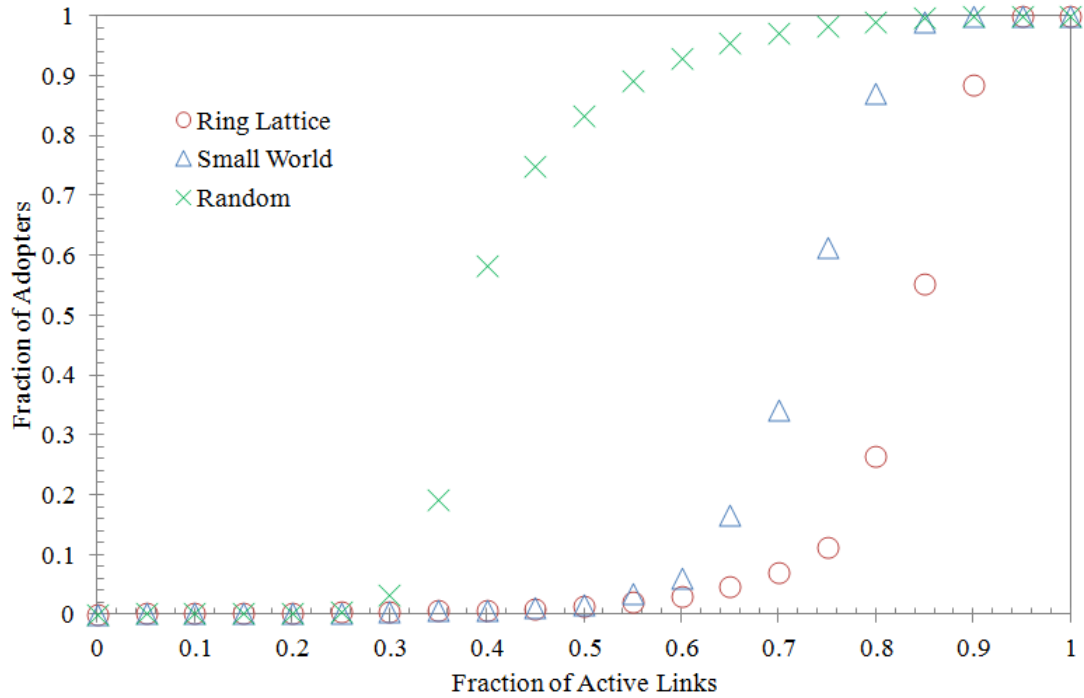


Figure 4.11: The effect of clustering coefficient and the average path length in link diffusion. The results are the same as the node diffusion. Firstly, the less the average path length, the more the link diffusion. Secondly, low the clustering coefficient suggests better link diffusion.

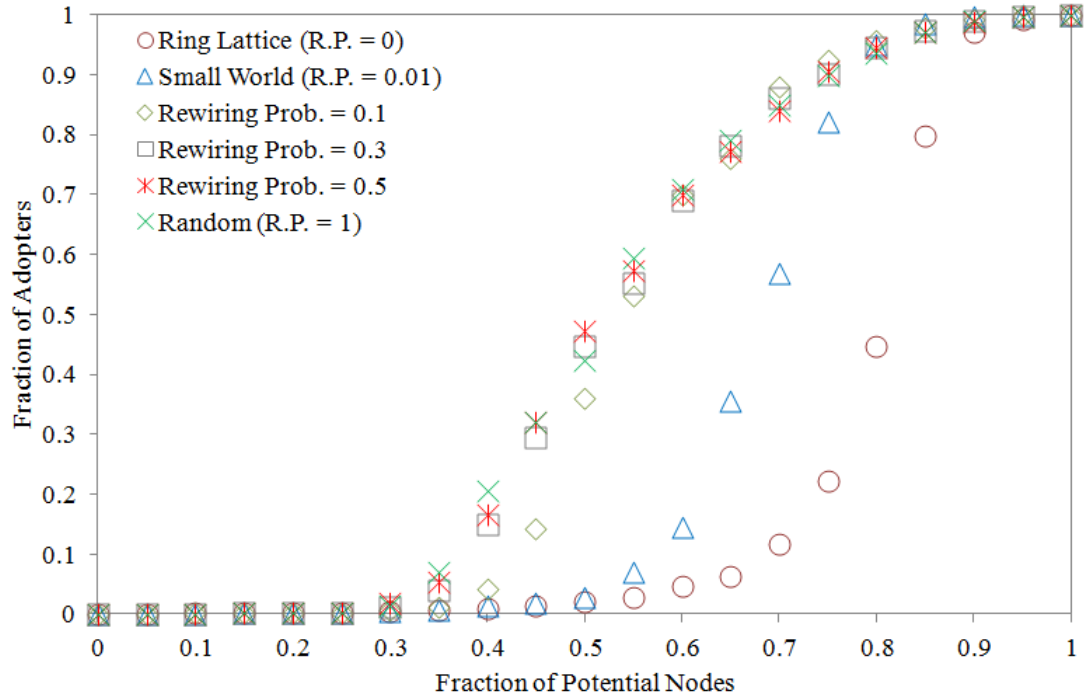


Figure 4.12: The effect of rewiring probabilities in neighbor diffusion.

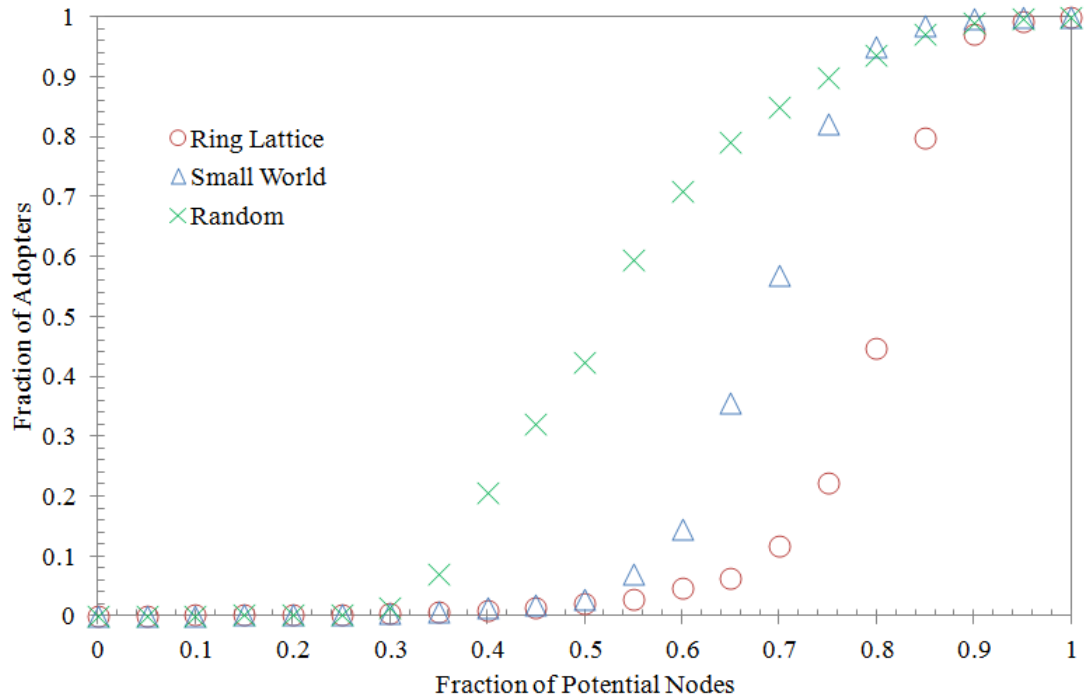


Figure 4.13: The effect of clustering coefficient and the average path length in neighbor diffusion. The results are the same as the link and node diffusion. The more the clustering coefficient, the less diffusion. The less the average path length, the more nodes to adopt.

4.3 The Effect of Hubs in Diffusion

In this section, we will analyze the effect of hubs (i.e. nodes with high number of connections) in diffusion. The hubs are present in today's networks, like the airline network, the world wide web, etc. [31]. In these networks, there are usually several hubs on which most of the shortest path relies. In this section, we will analyze how efficient the hubs are for the diffusion.

There are several algorithms for creating hub networks. The most common model is called rich get richer model [32]. In the rich get richer model, the new nodes will be linked to some node with a probability that is proportional to the number of connections of the particular node. Therefore, each new node is more likely to form a link with a node that has high number of connections. As a result, the network that is created using the rich get richer model will have several hubs. The network has a degree distribution which is called power law distribution. The degree distribution is linear in the logarithmic scale.

In the rich get richer model, new nodes are linked to a node in the network with a probability proportional to the number of connections. For instance, if there are only two nodes in the network with one link that ties them, the new node will be equally likely to form a link between one of the two existing nodes. After that, when the fourth node is added, the probabilities to create a link with a node depends on the number of links (connections) of the nodes. The networks generated with rich get richer model are called Scale Free networks. The following figure explains the rich get richer model.



Figure 4.14: The rich get richer model is illustrated. The first five images show the initial steps sequentially. After several iterations, the last image is an example for the hub networks.

The number of hubs in the rich get richer network is very small relative to the network size. However, the number of nodes with only one connection is more than the half of the network.

The number of nodes with n neighbors compared to the number of node with $n - 1$ neighbors linearly decreases in the logarithmic scale. Therefore, the degree distribution is called the power law distribution. The fraction of nodes with k connections can be seen in the formula below:

$$\text{Fraction of nodes with } k \text{ connections, } P(k) \sim k^{-\alpha} \quad (20)$$

We present another network that is similar with the rich get richer network in terms of number of nodes and links. However, in this network the hubs are not present. We generate this network using another model opposite of the rich get richer model, namely the poor get richer model. In this model, the new nodes are linked with a node proportional to the *number of connections*⁻¹. Therefore, as the number of connections increases the likelihood of creating new connections decreases. As a result, the poor get richer network has no hubs. Instead, there are lots of nodes with relatively small number of connections. We will call this network as the Robin Hood network. An illustrative example of the Robin Hood network is presented in Figure 4.15.

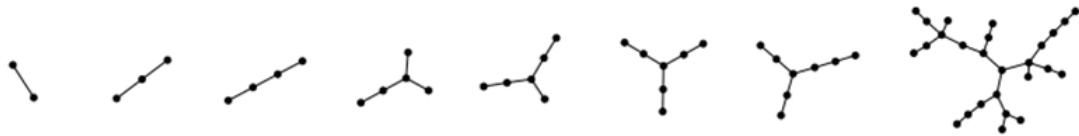


Figure 4.15: The poor get richer model is illustrated. The first seven images show the initial steps sequentially. After several iterations, the last image is an example for the Robin Hood networks.

There is one more network model that we would like to introduce that is in between the two networks presented in this section, i.e. Scale Free and Robin Hood networks. In this model, the new nodes are linked to a node arbitrarily. The model is called random get richer. One random node in the network links itself with the new node. The random get richer model has relatively more nodes with higher number of connections and relatively less nodes with fewer number of connections when compared to the poor get richer model. Figure 4.16 illustrates the poor get richer network.

The degree distribution of the Scale Free networks, Robin Hood networks and random get



Figure 4.16: The random get richer model is illustrated. The first seven images show the initial steps sequentially. After several iterations, the last image is an example for the random get richer networks.

richer networks are illustrated in Figure 4.17, Figure 4.18, and Figure 4.19, respectively. The random get richer model has relatively more nodes with higher number of connections and relatively less nodes with fewer number of connections when compared with the poor get richer model.

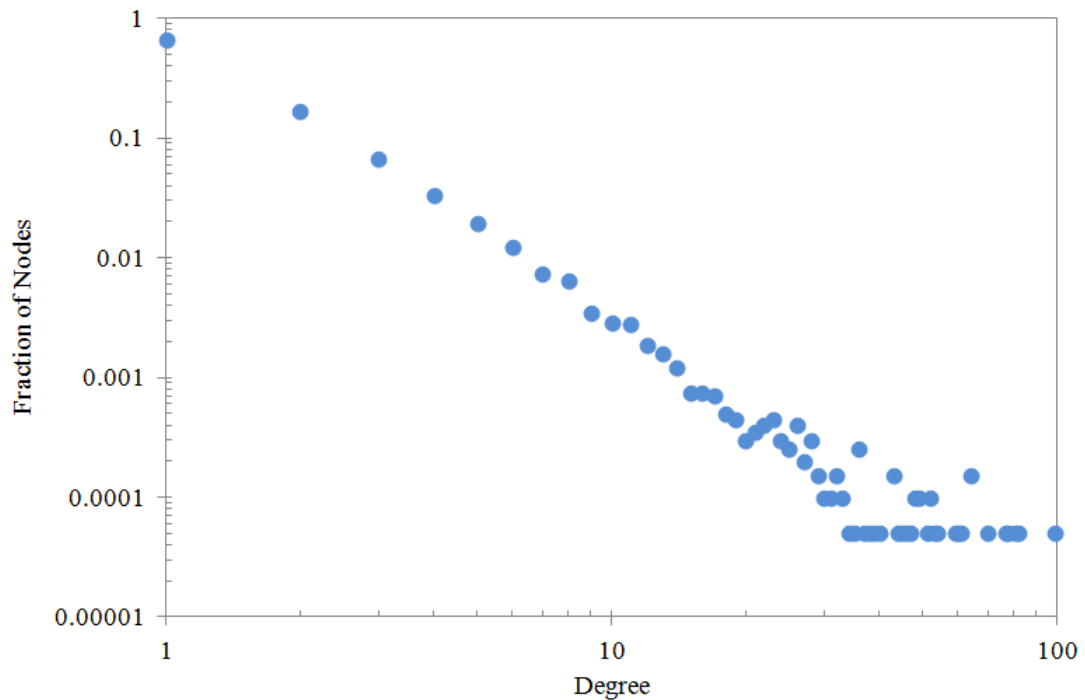


Figure 4.17: The degree distribution of rich get richer (Scale Free network) model in logarithmic scale.

Now, let us analyze the effect of hubs in diffusion. We created three networks that were described above with 1000 nodes and 999 links. We simulated the node, link and neighbor

diffusion models with these three networks and calculated the fraction of adopters that is averaged over one hundred simulations. The results are presented in the figures 4.20, 4.21 and 4.22. The experiments suggest that hubs have a positive effect in all the diffusion models. Also, there exists a domination in terms of the fraction of adopters in favor of the random get richer model when compared to the poor get richer model.

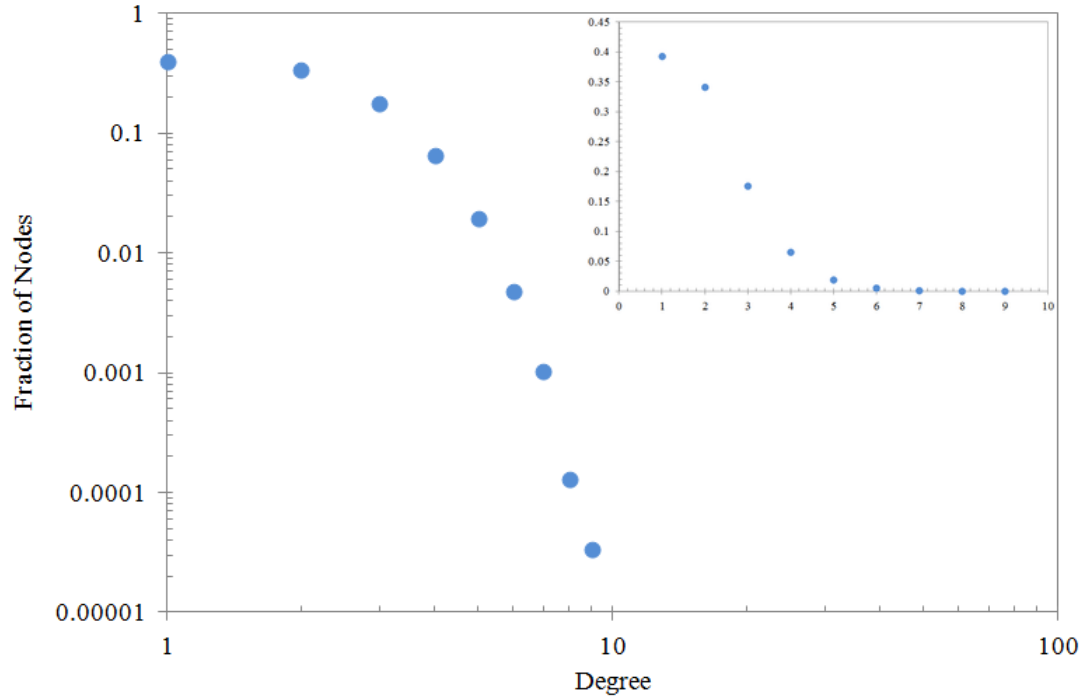


Figure 4.18: The degree distribution of poor get richer (Robin Hood network) model in logarithmic scale (in the large figure) and in regular scale (in the small figure).

The results show that the hubs are effective for the diffusion models. Also, more nodes with relatively higher connections are also effective for the diffusion. Since in the comparative results between the random get richer and poor get richer models, there exists a domination in terms of the fraction of adopters in favor of the random get richer model.

4.4 Time to Equilibrium in Diffusion

Unlike the previous chapter’s diffusion with infinite pace, the innovation diffuses to the closest link neighbors in one time click for the node, link and neighbor diffusion models. Therefore,

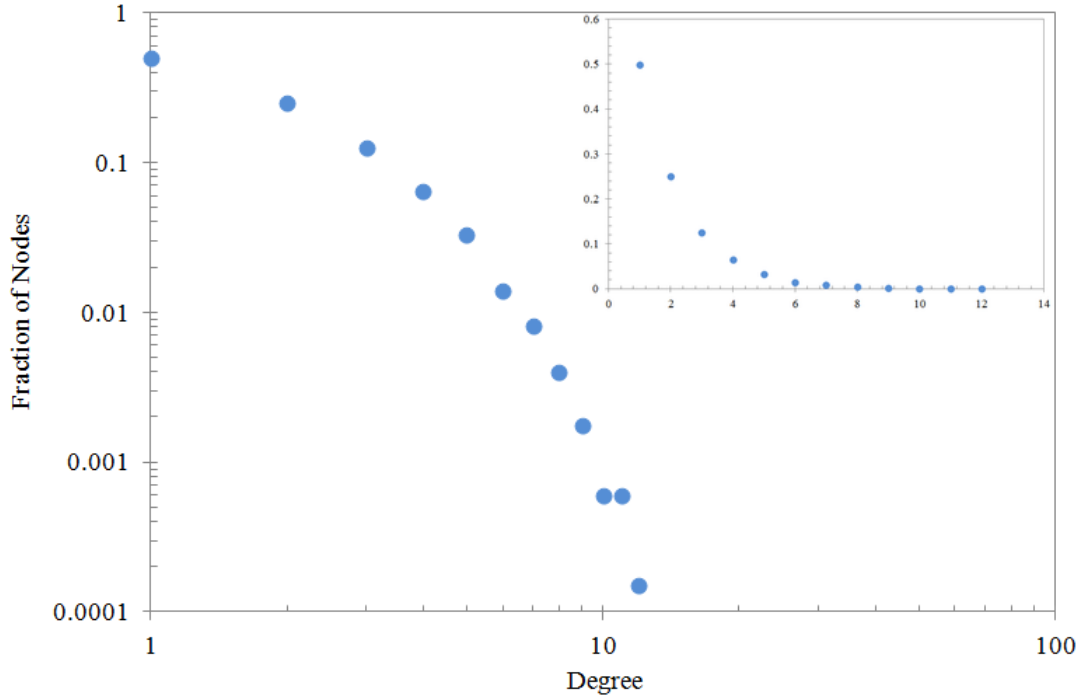


Figure 4.19: The degree distribution of random get richer (uniform distribution) model in logarithmic scale (in the large figure) and in regular scale (in the small figure).

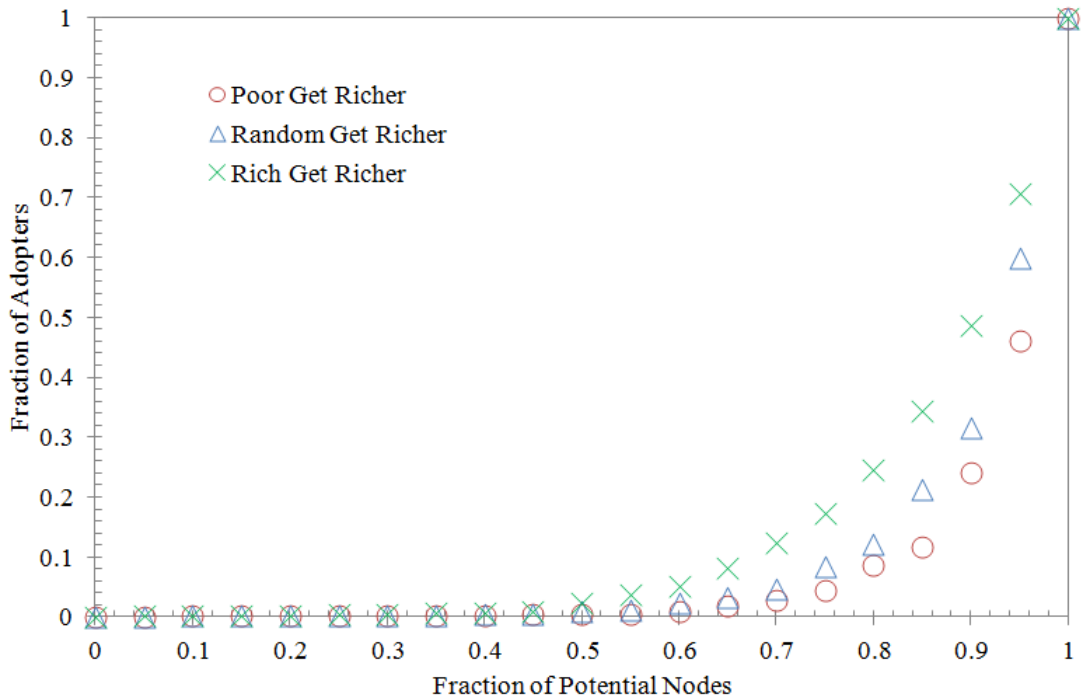


Figure 4.20: The effect of hubs in the node diffusion model.

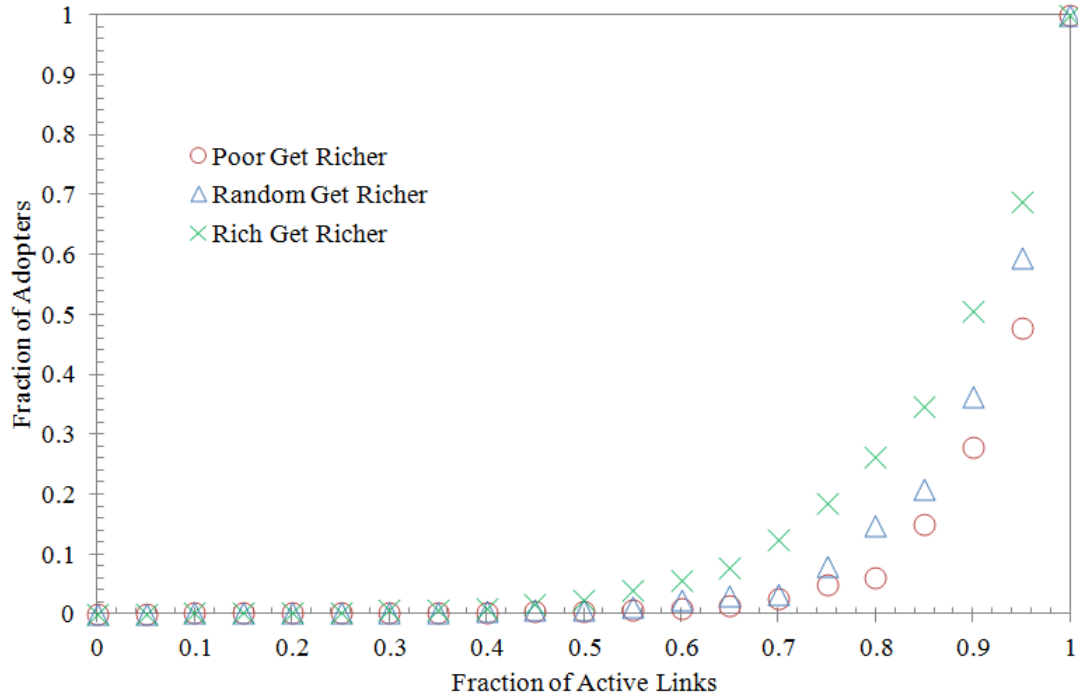


Figure 4.21: The effect of hubs in the link diffusion model.

reaching to the equilibrium takes time. We analyzed the time to equilibrium and the figures from 4.23 to 4.28 present the results.

The results for the node diffusion suggest that the clustering plays role in time as well as the average path length. Comparative results for the small world and random networks presents that clustering is also important for the time to equilibrium. On the other hand, the hubs make the diffusion faster.

In the link diffusion, the results are similar to the node diffusion. In summary, the hubs, lower clustering and smaller paths make the diffusion process quicker. However, the time to equilibrium reaches to maximum between 80% and 90% of the fraction of adopters . Above 0.9, although the fraction of adopters increase, the time to equilibrium decreases. In other words, it is faster to reach more nodes, which is an interesting result.

The results in neighbor diffusion are mostly the same as link diffusion and node diffusion. The hubs, lower clustering and smaller paths accelerate the diffusion process. Although the

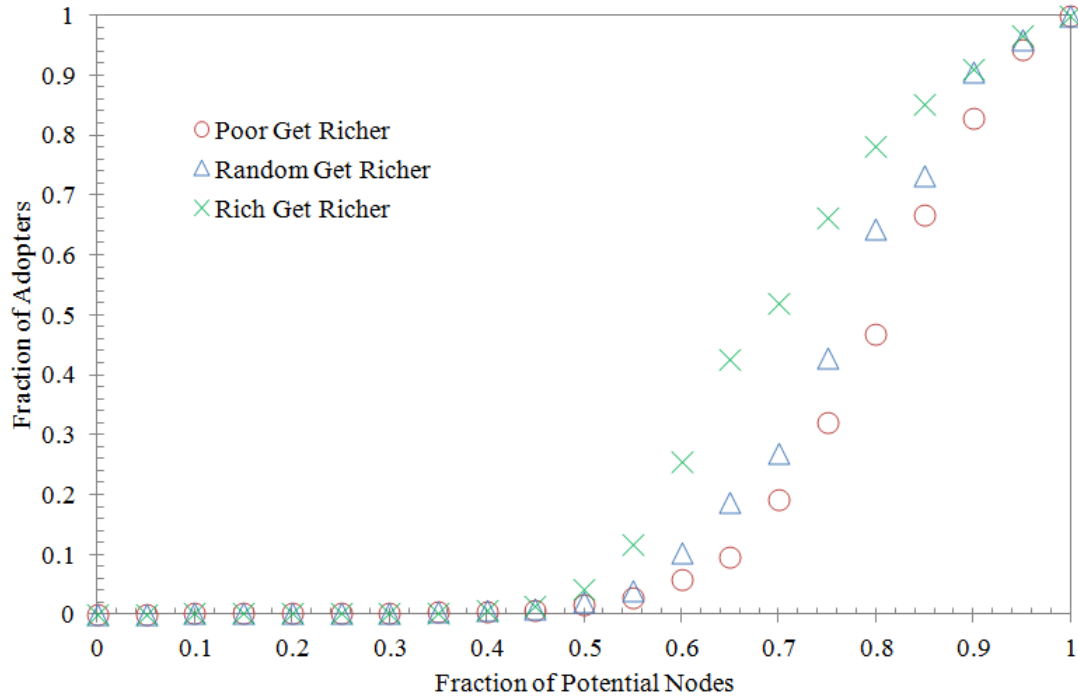


Figure 4.22: The effect of hubs in the neighbor diffusion model. The results are averaged over a hundred simulations. The experiments suggest that hubs have a positive effect in all the diffusion models. Also, there exists a domination in terms of the fraction of adopters in favor of the random get richer model when compared to the poor get richer model.

maximum time to equilibrium is around 0.95 for the ring lattice, the maximum time is around 0.85 for small world networks. Likewise, the time is maximum around 0.95 for the Robin Hood network, and random get richer model. However, as the fraction of adopters increases, the time also increases in the rich get richer model.

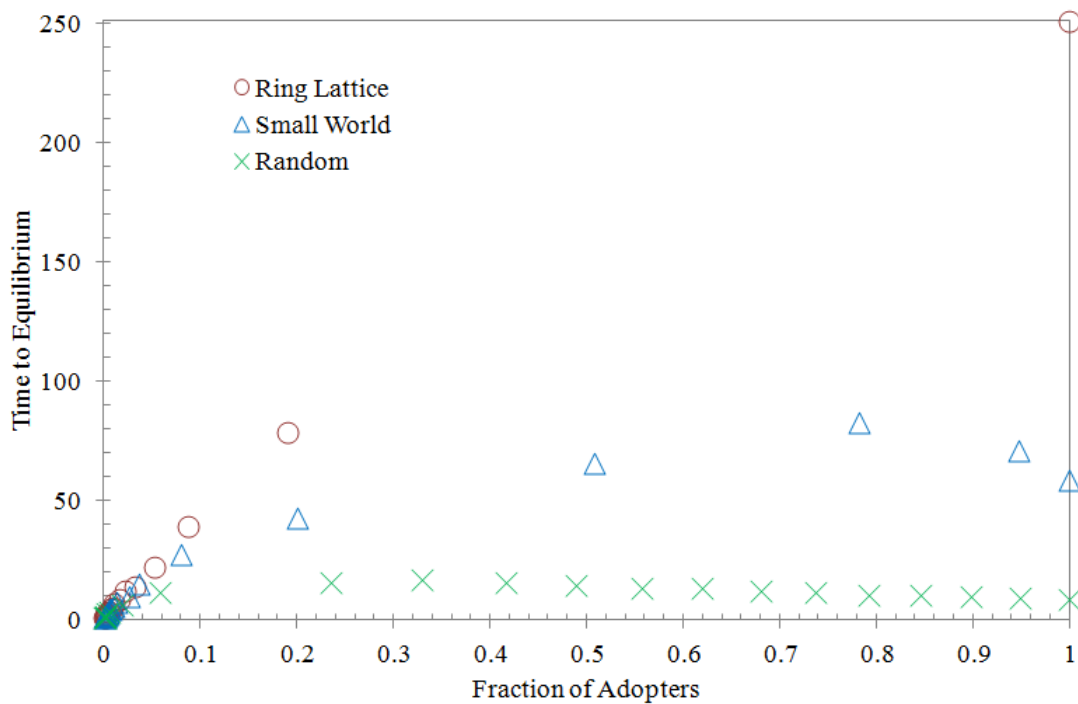


Figure 4.23: The time to equilibrium in node diffusion for ring lattice, small world and random networks. The results are averaged over a hundred simulations, and the number of nodes is 1,000.

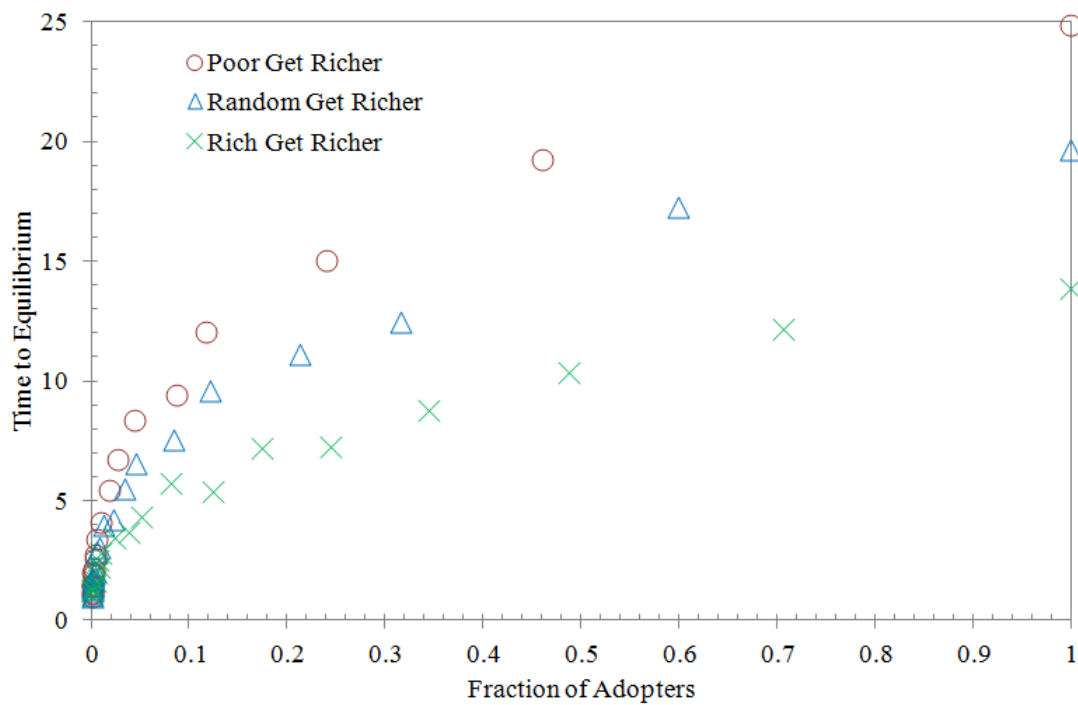


Figure 4.24: The time to equilibrium in node diffusion for hub networks. The results are averaged over a hundred simulations, and the number of nodes is 1,000.

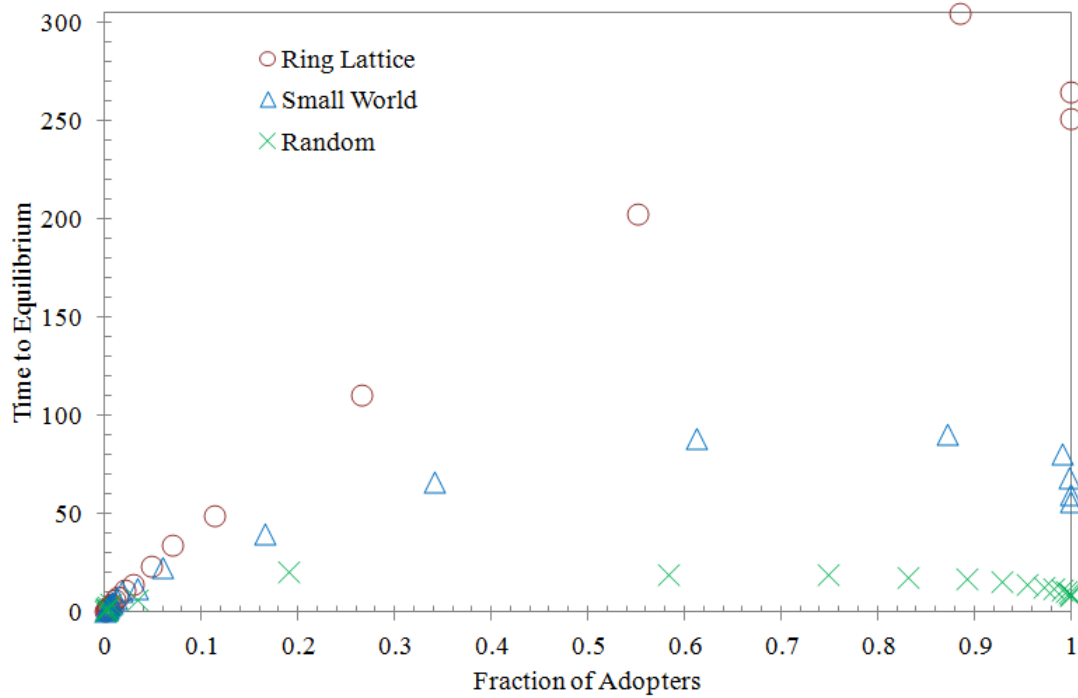


Figure 4.25: The time to equilibrium in link diffusion for ring lattice, small world and random networks. The results are averaged over a hundred simulations, and the number of nodes is 1,000.

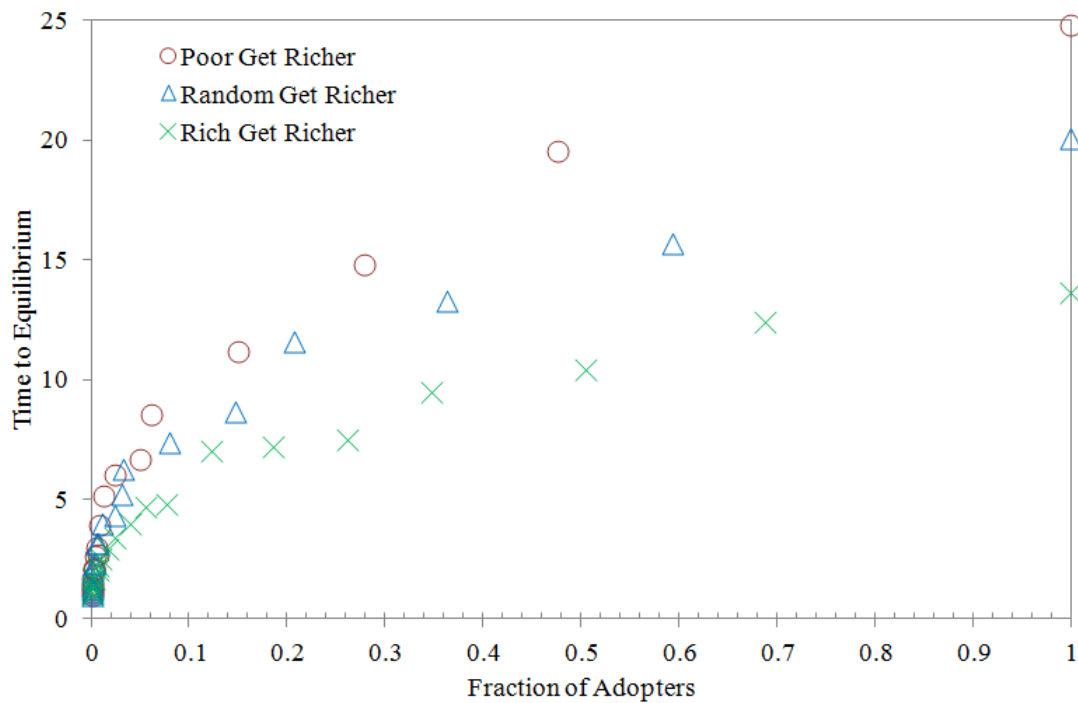


Figure 4.26: The time to equilibrium in link diffusion for hub networks. The results are averaged over a hundred simulations, and the number of nodes is 1,000.

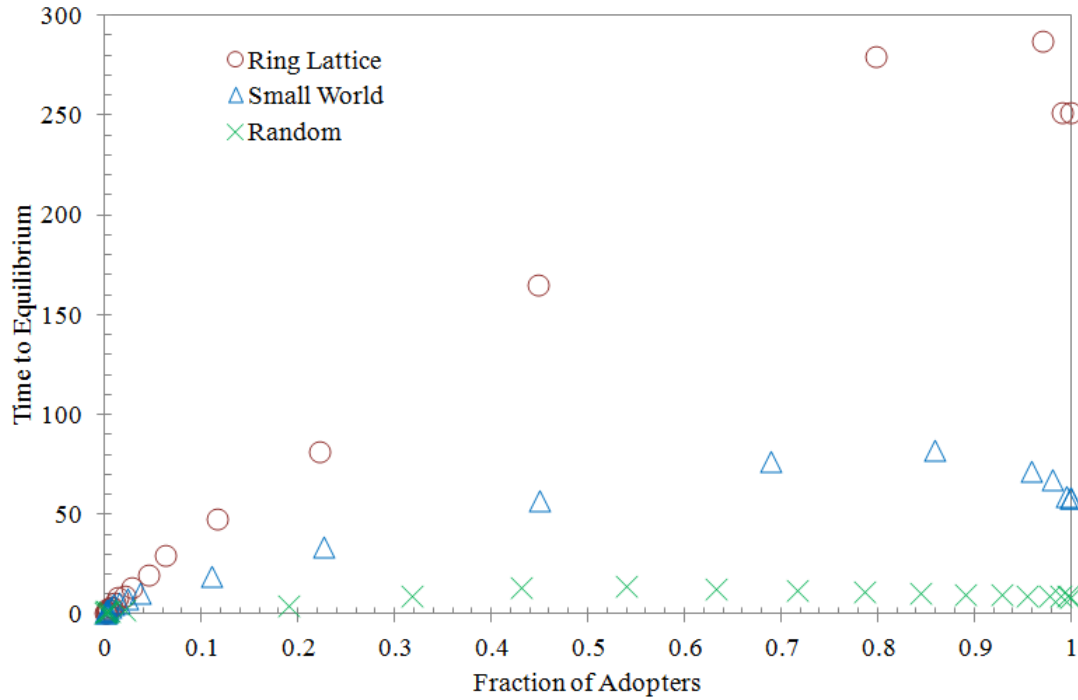


Figure 4.27: The time to equilibrium in neighbor diffusion for ring lattice, small world and random networks. The results are averaged over a hundred simulations, and the number of nodes is 1,000.

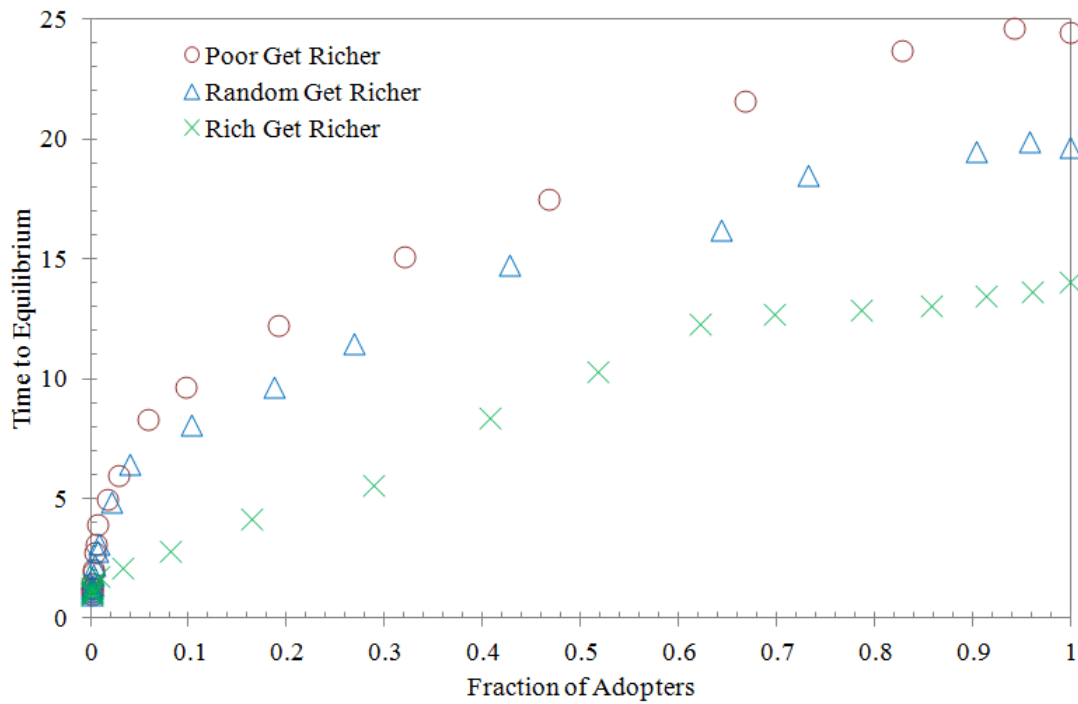


Figure 4.28: The time to equilibrium in neighbor diffusion for hub networks. The results are averaged over a hundred simulations, and the number of nodes is 1,000.

5 Collective Action

In this chapter, we will define the collective action model. The previous diffusion models were based on node's potential, link's active status or neighbor's potential. In this model, we will consider a collective action model, in which the nodes will adopt only if a certain fraction of its neighbors also adopt. Therefore, not only the neighbors are important for the decision to adopt, but also the neighbors of the neighbors play role in the decision. In fact, by the same logic, the entire network plays a role in the collective decision making process. So, the position of the node in the network will be as important as the network topology in collective action. In this chapter, we will focus on the position of the nodes and in some cases we will define several node motifs in networks that allow the collective action.

5.1 Definitions

Consider a social network, $G = (N, A)$, in which there are $N = \{1, 2, \dots, n\}$ nodes and the links are represented by an adjacency matrix, namely A . Nodes are labeled as the following two states; potentials (P), non-potentials (NP), based on a certain probability distribution.

Let a random node, say r , in the potentials set be an initial adopter. Let us also define the giant cluster of potential nodes, name it S , and initially S has only one element which is the random initial adopter (i.e. $S = \{r\}$). We define another set of nodes, namely C , be the collective subset which is initially empty. If the node u is in the collective subset, then $X_u^{c,b} = 1$. Finally, we define discrete time clicks as t .

In the collective action model, we aim to find a collective subset of potential nodes. Each node decides to adopt based on their neighbors. Each node is a part of the collective subset if the node has potential and a certain fraction of its neighbors are also in the collective subset. We assume that the nodes know whether their link neighbors' willingness to adopt. Therefore collective action is not only dependent on the node itself; it also depends on the node's position in the network.

$$Prob(X_u^{c,b} = 1 | u \in S) = \begin{cases} 1, & \text{if } \sum_{y \in N(u)} X_y^{c,b} / \sum_{y \in N(u)} w_{u,y} \geq \text{threshold} \\ 0, & \text{if } \sum_{y \in N(u)} X_y^{c,b} / \sum_{y \in N(u)} w_{u,y} < \text{threshold} \end{cases} \quad (21)$$

If a node is in the collective subset and the node knows for sure that a certain fraction of its neighbors are also in the collective subset, then the node become adopted. The nodes that become adopted will be the elements of the collective action.

5.2 Collective Action Algorithm

5.2.1 Finding the Giant Cluster of Potentials

The first step of the collective action algorithm is to find the giant cluster of potential nodes to which the initial adopter belongs. At each time click, the potential neighbors of all the potential nodes in the set S will be added to S . The iterations stop when there are no potential neighbors that are outside of the set S . The time clicks are also increased by 1 when a neighbor set is added to the set S . In fact, finding the giant cluster of potentials is exactly the node diffusion model. An example of this step of the algorithm is presented in Figure 5.1.

5.2.2 Finding the Collective Subset

In the definitions section, we defined a collective subset as C which was initially empty. Now, we will use the collective subset to find the adopted nodes in collective action. A node is in the collective subset if the node has potential and the node knows for sure that a certain fraction of its neighbors are also in the collective subset. But, how does the node know the neighbors

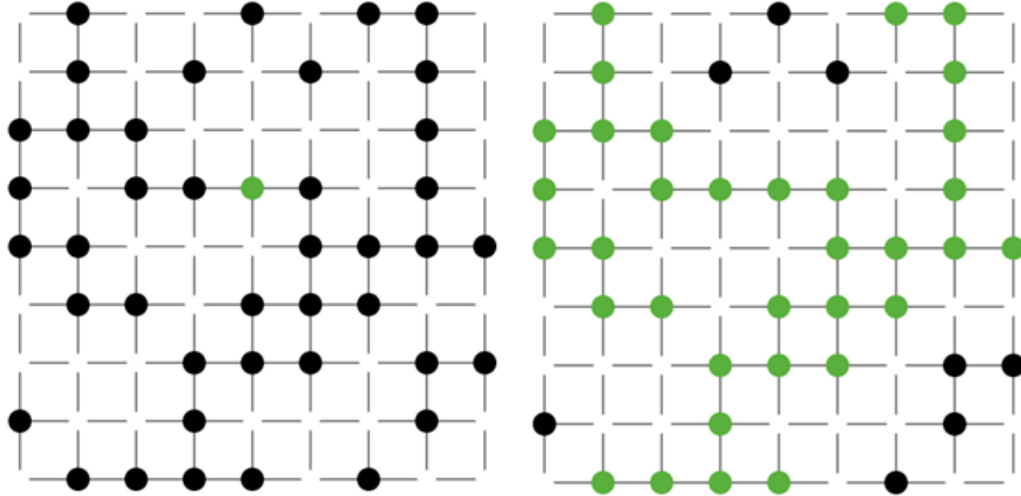


Figure 5.1: Finding the giant cluster of potential nodes in 2D regular lattice with 81 nodes and 144 links. The potential nodes are drawn as black circles and the non-potentials are drawn as white circles which are not visible. The potential status of the nodes are determined by a uniform random distribution $U \sim (0, 1)$. The green node in the left figure represents the initial adopter. All the potential nodes that is connected with the initial adopter are illustrated as green circles in the right figure. There are some potential nodes that are not added to the set S in the equilibrium on the right. Actually, this step is the node diffusion model defined in the previous chapter.

of its neighbors are also in the collective subset for sure?

Firstly, in order a node to be in the collective subset, the node should have a potential. In other words, if the node has no potential, then the node is automatically not in the collective subset. So, we can start with the following idea: A node is not in the collective subset, if a certain fraction of its neighbors have no potential.

Secondly, we assume that the nodes know whether their neighbors are also in the collective subset. Therefore, after a certain time period if at least a fraction of the node's neighbors are still in the collective subset, then the node knows for sure that at least a fraction of its neighbors are also in the collective subset.

In the first iteration of finding the collective subset, all the nodes in the giant cluster of potential nodes (i.e. S) are added to the collective subset, C except that the nodes with a

certain fraction of its neighbors have no potential. Then, for each iteration, the nodes with a certain fraction of its neighbors are not in C will be deleted from the collective subset. The iterations will stop either when the collective subset is empty or there are no nodes to be deleted from C . We also increase the time ticks by 1 at the end of each iteration. The remaining nodes in the collective subset are the elements of the collective action.

An example for finding the collective subset is presented in the figure below.

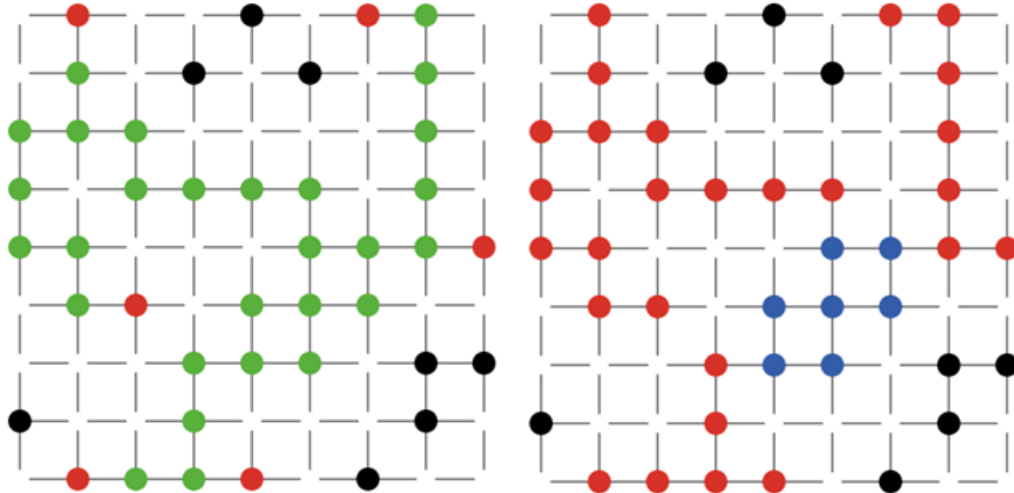


Figure 5.2: Collective action in 2D regular lattice with 81 nodes and 144 links. The potential nodes are drawn as visible circles and the non-potentials are drawn as white circles which are not visible. The potential states of the nodes are determined by a uniform random distribution. The threshold is the half fraction of the neighbors (at least 2 neighbors). The green nodes in the left figure represents the initial willing to adopt set. The red nodes in the left figure was a part of the giant cluster of potential nodes, but are excerpted from the collective subset since more than half of their neighbors have no potential (therefore not in the collective subset). After several iterations, the figure on the right is formed. The blue nodes in the right figure represents the collective subset.

5.3 The Effect of Clustering Coefficient and Average Path Length in Collective Action

In this subsection, we will analyze the effect of clustering coefficient and average path length in collective action. The network topology is worth testing for the collective action, and we

will analyze the effect of clustering, path lengths and hubs, like in the previous chapter's diffusion algorithms.

We simulated the ring lattice, small world and random networks to analyze the effect of clustering coefficient and average path lengths. The number of nodes are 1000 and the number of links are 2000 for these networks. The results are averaged over a thousand simulations. Also, we normalized the collective subset with the giant cluster of potential nodes. The normalization was necessary due to the difference in the giant cluster size for these networks. It would be unfair to compare the adopted nodes in collective action given that the sizes of giant potential nodes are different. The following figure shows the effect of clustering coefficient and average path length in collective action.

As explained previously, the ring lattice and small world have the same clustering coefficient (which is very high), whereas their average path lengths are different. On the other hand, the random network has the same path length with small world but random network has a lower clustering coefficient relative to the small world and ring lattice. The figure above suggests that the networks with tightly-knit small groups are effective in collective action. The effect of clustering coefficient is exactly the opposite compared to the diffusion algorithms. On the other hand, the path length has no effect in collective action. Small world and ring lattice has the same size of collective subsets, although their path lengths are significantly different from each other.

5.4 The Effect of Hubs in Collective Action

In this section, we will analyze the effect of hubs in collective action. The hub networks are explained in the previous chapter. We will use the same networks, namely poor get richer (Robin Hood), random get richer and rich get richer (Scale Free) networks. The effect of hubs in the collective action can be seen in the following figure.

As the figure suggests, the hubs have little impact (almost negligible) over the collective action. Although, it is interesting to see that below 0.5 the poor get richer (Robin Hood)

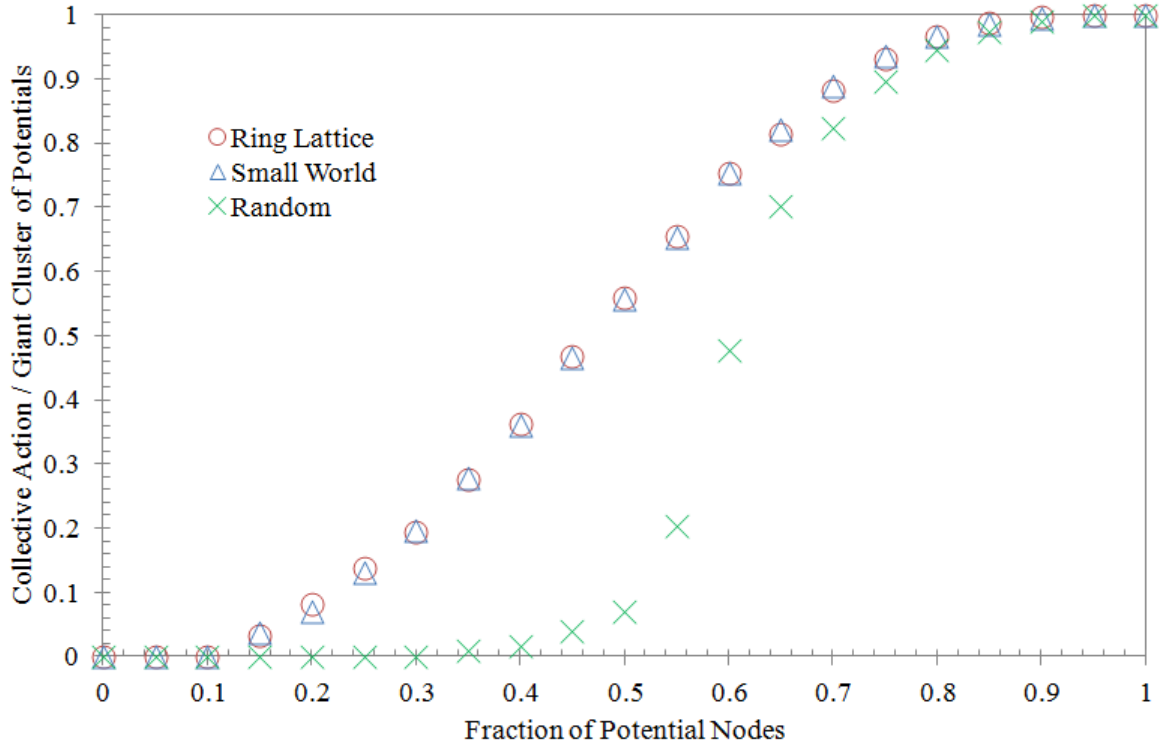


Figure 5.3: The effect of clustering coefficient and average path length in collective action. The number of nodes are 1000 and the number of links are 2000. The results are averaged over a thousand simulations. The y axis represents the number of nodes involving in collective action divided by the giant cluster of potential nodes. The normalization is needed to understand the effect of collective action, since the number of adopters can be biased due to the difference in the giant cluster of potentials. The average path length has no effect in collective action, however the clustering coefficient has a huge affect. The tightly-knit communities where the clustering coefficient is high are more effective in collective action.

network works relatively well, but above 0.5 the rich get richer (Scale Free) network works well compared to the other networks.

5.5 Conclusions

The first conclusion is that the collective subset is independent from the position of the initial adopter. In other words, the initial adopter does not affect the size of the collective subset.

Secondly, the collective subset is independent from the size of the giant cluster of potential nodes. It is possible that the size of the giant cluster of potentials in a network can be

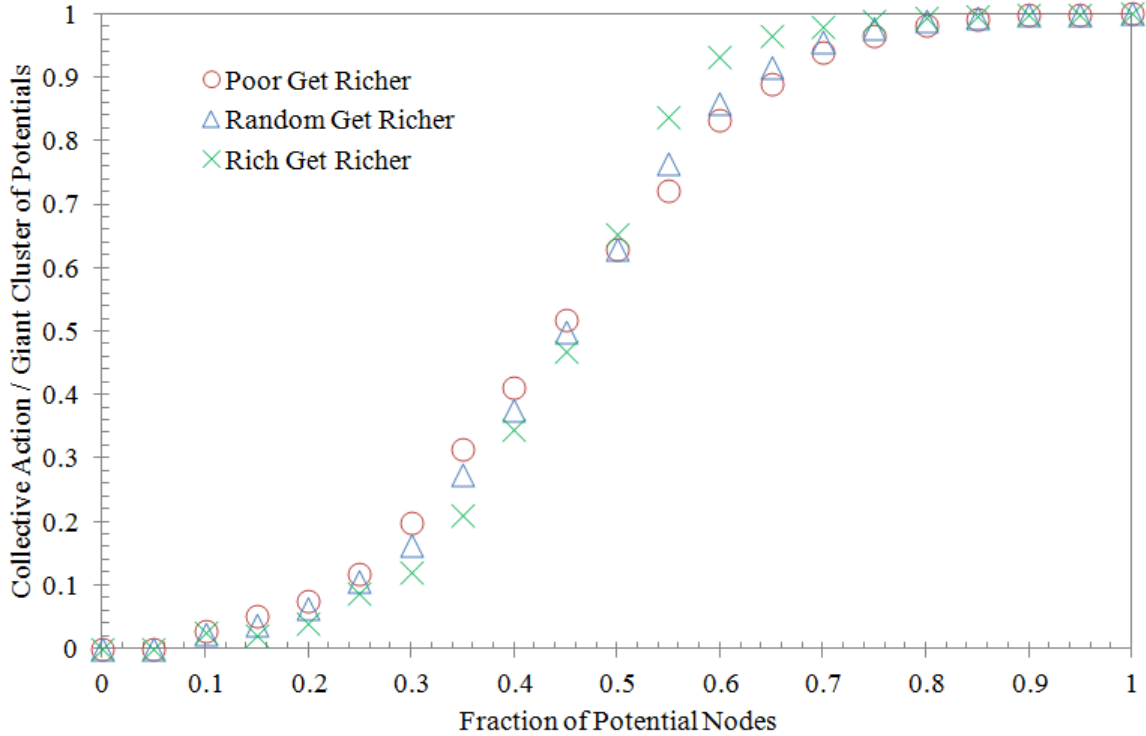


Figure 5.4: The effect of hubs in collective action. The number of nodes are 1000 and the number of links are 999. The results are averaged over a thousand simulations. The y axis represents the number of nodes involving the collective action divided by the giant cluster of potential nodes. As seen in the figure, the hubs have negligible impact in collective action.

larger than another network's size of giant cluster; however number of adopters can be smaller than the other. An example can be viewed below.

Thirdly, the collective subset depends on topology of the underlying network. The clustering coefficient does have an impact for the size of the collective action. The tightly-knit communities are very effective in collective action. On the other hand, the average path length and hubs do not effect the collective action.

The fourth conclusion is that the collective subset is subject to the position of the potential nodes in the network. Even when a couple of nodes with opposite potential status change their positions, the number of adopters changes dramatically. To illustrate this, below figure is presented.

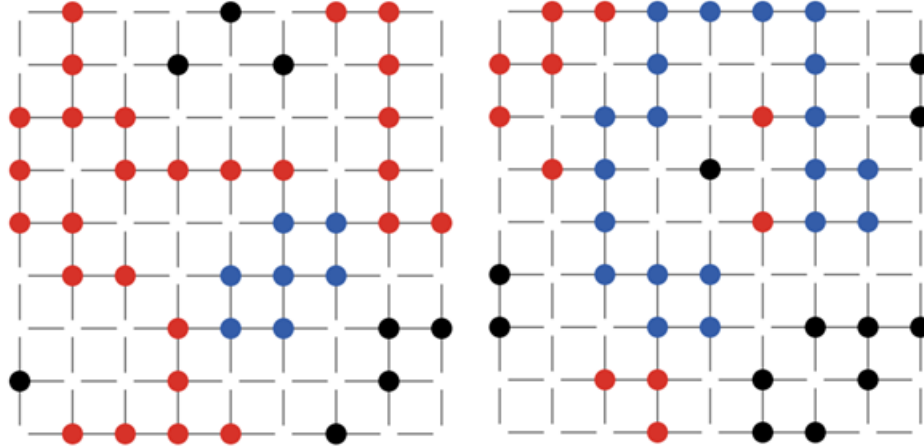


Figure 5.5: An illustrative example that the collective subset is independent from the size of the giant cluster of potential nodes in 2D regular lattice with 81 nodes and 144 links. The potential nodes are drawn as visible circles and the non-potentials are drawn as white circles which are not visible. The potential states of the nodes are determined by a uniform random distribution $U \sim (0, 1)$. The threshold is the 0.5 fraction of the neighbors (at least 2 neighbors). The blue nodes represent the collective subset. The figure on the left has the giant cluster size of 34, but the number nodes involving in collective action is 7. On the other hand, the right figure has the giant cluster size of 31, but the number nodes involving in collective action is 20.

The fifth conclusion is that finding the collective subset takes time. The time for finding the collective subset is bounded above by two times the diameter of the network (i.e. the largest shortest path). Finding the giant cluster of potential nodes takes at most the diameter time clicks. Since the initial adopter can only be far away from any node at most by the diameter. By the same logic, finding the collective subset from the giant cluster of potential nodes is bounded above by the diameter. Therefore, two times the diameters of the network is the upper bound for finding the collective subset.

The sixth conclusion is that the giant cluster of the potential nodes is the equilibrium of the node diffusion. This is already mentioned in the first step of the algorithm.

The last but not least conclusion that we will mention is that the collective subset is the largest subset in which every node in the giant cluster of potentials has at least a threshold of their neighbors are also in the subset. The last point is very important for understanding the

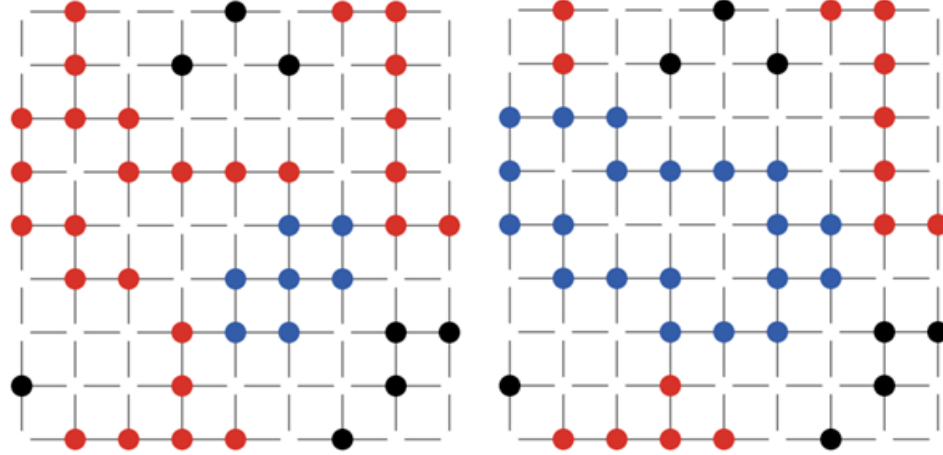


Figure 5.6: An illustrative example that the collective subset is subject to the position of the potential nodes in the network in 2D regular lattice with 81 nodes and 144 links. The potential nodes are drawn as visible circles and the non-potentials are drawn as white circles which are not visible. The potential states of the nodes are determined by a uniform random distribution $U \sim (0, 1)$. The threshold is the half fraction of the neighbors (i.e. at least 2 neighbors). The blue nodes represent the collective subset. Everything is the same between the two figures, except there is only a positional change of a potential and a non-potential node. Even with a single positional change, the number of nodes involving in collective action changes dramatically (from 7 to 20). The position of the potential nodes is very important in collective action.

collective action.

5.5.1 Motifs for Collective Subset in 2D Regular Lattice

In this section, we will analyze the collective subset motifs in the 2D regular lattice. There are certain motifs that can be used to find the collective subset in 2D regular lattice. As mentioned before, the collective subset is the largest subset of the giant cluster of potential nodes where every node in the collective subset has at least a threshold of their neighbors are also in the collective subset. Then, we can actually come up with some motifs in the 2D regular lattice. We will assume that the threshold is 0.5, in other words at least 2 neighbors is necessary for collective action.

First motif is that every cycle forms a collective subset. Since, at any cycle in 2D regular lattice every node has at least two neighbors in the cycle. Secondly, every path with cycles

6 Conclusion & Future Research

In this thesis, we first analyzed the diffusion of innovation in one-dimensional networks. Two diffusion algorithms were defined, namely active link diffusion and active node diffusion. We analytically solved the time to cover the entire network. Then, we have compared the analytical solution and computational results. The comparison of the analytical solution and simulation results suggested that analytical solution and computational results were consistent.

In the consecutive chapter, we changed our focus to two-dimensional networks. We simulated the aforementioned algorithms for the diffusion of innovation in two-dimensional networks. We defined the concept of limited information and analytically solved the time to reach to the target node. The simulation results showed that although the network size grows exponentially, the time to reach to the target node increases linearly.

In the next chapter, we analyzed the diffusion algorithms in complex networks. The main purpose was to analyze the effect of topology of the network, especially the clustering coefficient, average path length and degree distribution. We further introduced another diffusion algorithm, namely neighbor diffusion. Three diffusion algorithms were implemented into several well known networks. In particular, ring lattice, the small world and random networks were used for the simulations. The results suggested that lower clustering coefficient leads to more diffusion. Likewise, the average path length affects the final adopters. If the average path length is large, then the diffusion will not be effective relative to the smaller average path length. Finally, we showed that the hubs (the nodes with high number of connections) play a significant role in the diffusion processes.

After the diffusion algorithms, we focused on the collective decision making. In collective decision making, the nodes' decisions were dependent on the neighbors' decisions. If at least a certain threshold of neighbors adopts the behavior, then the particular node also adopts. The collective decision making algorithm was presented. The collective decision making consisted of two steps. The first step was basically the node diffusion. The second step consisted of several intermediate steps. It started with the giant cluster resulted from the node diffusion. Then the collective subset was computed in which every node in the subset had at least a certain fraction of its neighbors also in the collective subset. Finally, the effect of clustering coefficient, the average path length and the hubs also analyzed for the collective action. Despite of the diffusion results, higher clustering coefficient led bigger collective subset. Also, the average path length did not affect the collective subset. Lastly, the hubs didn't seem to play a significant role for the size of the collective subset.

The main purpose for this thesis was to present the importance of the network structure for the diffusion and collective decision making. For both the diffusion and collective action, the results suggested that network structure plays an important role. However, the results for the diffusion and the collective action were partially the opposite. In the diffusion models the tightly-knit communities where the clustering coefficient is high did not lead to large number of adopters. However, in the collective action the tightly-knit communities led to bigger sized collective subsets. In order to maximize the final adopters in the diffusion models and the size of the collective subset, our results suggest that the clustering coefficient needs to be in an intermediate level. Higher clustering coefficient leads to higher number of adopters for the diffusion, but lower collective subset size. For the average path length and hubs, collective action is indifferent, but the maximum adopters for the diffusion models prefer lower average path length and several hubs in the networks.

For the future research, we will analyze the effect of the initial adopter. The initial adopter was taken randomly in this thesis; however a specific selection might lead to different outcomes. Also, the time for the collective action was shown to be bounded above by two diameters.

However, further analysis on the time for the collective action remains a direction of the future research. Furthermore, the limitation for the information about the neighbors might be interesting to analyze. For instance, only a fraction of the neighbors might be visible for a particular node, and the node might have to guess or reason the remaining neighbors. Therefore, the decision whether to adopt the behavior can depend on how well the guess or the reasoning is.

References

- [1] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.
- [2] Charles Andrew Myers and George P Shultz. *The Dynamics of a Labor Market: A Study of the Impact of Employment Changes on Labor Mobility, Job Satisfactions, and Company and Union Policies*. Prentice-Hall, New York, 1951.
- [3] Albert Rees and George Schultz. *Workers in an Urban Labor Market*. University Of Chicago Press, Chicago, 1970.
- [4] Mark S Granovetter. The strength of weak ties. *American Journal of Sociology*, pages 1360–1380, 1973.
- [5] Yannis M Ioannides and Linda Datcher Loury. Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, pages 1056–1093, 2004.
- [6] Matthew O Jackson and Leeat Yariv. Diffusion, strategic interaction, and social structure. *Handbook of Social Economics*, 1:645–678, 2011.
- [7] James Samuel Coleman, Elihu Katz, and Herbert Menzel. *Medical Innovation: A Diffusion Study*. Bobbs-Merrill, Indianapolis, 1966.
- [8] Bryce Ryan and Neal C Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural Sociology*, 8(1):15–24, 1943.
- [9] Zvi Griliches. Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, pages 501–522, 1957.
- [10] Elihu Katz and Paul Felix Lazarsfeld. *Personal Influence: The Part Played by People in the Flow of Mass Communications*. The Free Press, 1955.
- [11] Paul Felix Lazarsfeld, Bernard Berelson, and Hazel Gaudet. *The Peoples Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Columbia University Press, New York, 1948.

- [12] Lawrence F Feick and Linda L Price. The market maven: A diffuser of marketplace information. *The Journal of Marketing*, pages 83–97, 1987.
- [13] Catherine Tucker. Identifying formal and informal influence in technology adoption with network externalities. *Management Science*, 54(12):2024–2038, 2008.
- [14] Timothy G Conley and Giorgio Topa. Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, 17(4):303–327, 2002.
- [15] Nicholas A Christakis and James H Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007.
- [16] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.
- [17] Edward L Glaeser, Bruce Sacerdote, and José A Scheinkman. Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–548, 1996.
- [18] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, pages 1420–1443, 1978.
- [19] Michael Suk-Young Chwe. Communication and coordination in social networks. *The Review of Economic Studies*, 67(1):1–16, 2000.
- [20] Michael Suk-Young Chwe. Structure and strategy in collective action. *American Journal of Sociology*, 105(1):128–156, 1999.
- [21] Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1):1–27, 2013.
- [22] Dietrich Stauffer and Ammon Aharony. *Introduction to Percolation Theory*. CRC press, 1994.
- [23] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical*,

- Physical and Engineering Sciences*, volume 115, pages 700–721. The Royal Society, 1927.
- [24] Daryl J Daley, Joe Gani, and Joseph Mark Gani. *Epidemic Modelling: An Introduction*, volume 15. Cambridge University Press, 2001.
- [25] Leah B Shaw and Ira B Schwartz. Fluctuating epidemics on adaptive networks. *Physical Review E*, 77(6):066101, 2008.
- [26] Arthur Campbell. Word-of-mouth communication and percolation in social networks. *The American Economic Review*, 103(6):2466–2498, 2013.
- [27] Paolo Zeppini, Koen Frenken, and Luis R Izquierdo. Innovation diffusion in networks: the microeconomics of percolation. Technical report, 2013.
- [28] Uri Wilensky. Netlogo itself. *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston*, <http://ccl.northwestern.edu/netlogo>, 1999.
- [29] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.
- [30] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
- [31] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [32] Robert K Merton. The matthew effect in science. *Science*, 159(3810):56–63, 1968.